Edited by

George P. Patrinos | Wilhelm Ansorge

Molecular Diagnostics

Second Edition

Molecular Diagnostics

This page intentionally left blank

Molecular Diagnostics

Second Edition

Edited by

George P. Patrinos
Wilhelm J. Ansorge

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
32 Jamestown Road, London NW1 7BY, UK
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA

First edition 2005
Second edition 2010

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-12-374537-8

Typeset by Macmillan Publishing Solutions (www.macmillansolutions.com)

Printed and bound in the United States of America

10 11 12 13 14 15 10 9 8 7 6 5 4 3 2 1

Contents

Note: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

This page intentionally left blank

Prof. Dr. Farid E. Ahmed Professor, Department of Radiation Oncology, Leo W. Jenkins Cancer Center, East Carolina University, The Brody School of Medicine, 600 Moye Boulevard, LSB 014, Greenville, NC 27858, USA.

Dr. Claus L. Andersen Associate Professor, Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, 8200-Aarhus N, Denmark.

Prof. Dr. Wilhelm J. Ansorge Visiting Professor, Ecole Polytechnique Federal Lausanne, PH D2 445 (Bâtiment PH) Station 3, Lausanne, CH-1015, Switzerland.

Prof. Dr. Aglaia Athanassiadou Professor of General Biology, University of Patras, Faculty of Medicine, School of Health Sciences, Department of General Biology, Asclepiou Street, Panepistimioupolis, GR-26110, Patras, Greece.

Prof. Dr. Francisco E. Baralle Head, Molecular Pathology Laboratory; Director, International Centre for Genetic Engineering and Biotechnology (ICGEB), Padriciano, 99, 34012 Trieste, Italy.

Dr. Matthew P. G. Barnett Senior Research Scientist, Food, Metabolism and Microbiology, Food and Textiles Group, AgResearch Limited, Palmerston North, New Zealand.

Dr. Sina Bavari Chief of Immunology, Target Identification, and Translational Research, United States Army Medical Research Institute of Infectious Diseases, Bacteriology Division, Ft. Detrick, MD, 21702, USA.

Prof. Dr. Alex van Belkum Professor of Molecular Microbiology, Erasmus MC, Department of Medical Microbiology and Infectious Diseases, Unit Research and Development, Dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands.

Dr. Walter Bell Associate Professor of Pathology, Division of Anatomic Pathology, University of Alabama at Birmingham, North Pavilion – Room 3540, University Station, Birmingham, AL 35294-7331, USA.

Dr. Fernando Benavides Associate Professor and Director, Genetic Services, Department of Carcinogenesis, The University of Texas MD Anderson Cancer Center, 1808 Park Road 1C – PO Box 389, Smithville, TX 78957, USA.

Dr. Karin Birkenkamp-Demtroder Associate Professor, Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, 8200-Aarhus N, Denmark.

Lauryn Blakesley Research Assistant, Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111-2497, USA.

Prof. Dr. Anja Katrin Bosserhoff Professor of Molecular Pathology, University of Regensburg, Institute of Pathology, Franz-Josef-Strauß-Allee 11, D-93042 Regensburg, Germany.

Dr. Bruce Budowle Senior Scientist, Federal Bureau of Investigation, Laboratory Division, 2501 Investigation Parkway, Quantico, Virginia, 22135, USA.

Dr. Chinh Thien Bui Senior Research Officer, Walter and Eliza Hall Research Institute, 4 Research Ave, Latrobe Research and Development Park, Bundoora, Victoria 3086, Australia.

Dr. David Burnett Consultant in Quality and Accreditation Systems, Lindens Lodge, Bradford Place, Penarth, CF64 1LA, United Kingdom.

Dr. Rowan S. Campbell Research Associate, University of North Texas, School of Medicine, PO Box 311277, Denton, Texas 76203, USA.

Prof. Dr. Timothy Caulfield Canada Research Chair in Health Law and Policy, Research Director, Health Law Institute; Professor, Faculty of Law, Faculty of Medicine and Dentistry, 461 Law Centre, University of Alberta, Edmonton, Alberta, T6G 2H5, Canada.

Maria Chiotis Research Assistant, Genomic Disorders Research Centre, 7th Floor, Daly Wing, St. Vincent's Hospital, 41 Victoria Parade, Fitzroy 3065, Melbourne, Victoria, Australia.

Tim Conze Graduate Student, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Prof. Dr. Daniel Corach Professor, Department of Genetics and Molecular Biology, University of Buenos Aires, School of Pharmacy and Biochemistry; Director, DNA Fingerpring Service (Servicio de Hiellas Digitales Genéticas-SHDG); Principal Research Fellow, National Research Council (CONICET), Junin 956, 1113-Buenos Aires, Argentina.

Prof. Dr. Richard G.H. Cotton Director, Genomic Disorders Research Centre, 7th Floor, Daly Wing, St. Vincent's Hospital, 41 Victoria Parade, Fitzroy 3065, Melbourne, Victoria, Australia.

Dr. Karina Dalsgaard-Sørensen Associate Professor, Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, 8200-Aarhus N, Denmark.

Dr. Carol A. Delaney Clinical Scientist in Molecular Cytogenetics, SE Scotland Cytogenetics Laboratory, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, United Kingdom.

Dr. Martine de Rycke Senior Scientist, Centre for Medical Genetics, UniversitairZiekenhuis Brussel, Department of Embryology and Genetics, Vrije Universiteit Brussel, Laarbeeklaan 101, B-1090 Brussels, Belgium.

Johan T. den Dunnen Leiden Genome Technology Center, Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.

Dr. Lars Dyrskjøt Associate Professor, Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, 8200-Aarhus N, Denmark.

Dr. Arthur J. Eisenberg Associate Professor and Director, DNA Identity Laboratory, Department of Pathology and Anatomy, University of North Texas Health Science Center, 3500 Camp Bowie Boulevard, Ft. Worth, Texas 76107, USA.

Dr. Vitor Faca Post-Doctoral Fellow, Department of Molecular Diagnostics, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, M5-C800, Seattle, WA 98109, USA.

Jared S. Farrar Laboratory Assistant, University of Utah School of Medicine, Department of Pathology, 50 N Medical Drive, Salt Lake City, UT 84132, USA.

Prof. Dr. Lynnette R. Ferguson Programme Leader, Nutrigenomics New Zealand; Professor and Head, Discipline of Nutrition, Faculty of Medical and Health Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand.

Prof. Dr. Paolo Fortina Professor of Medicine, Thomas Jefferson University; Director of Genomics and Diagnostics Center for Translational Medicine, Department of Medicine, 408 College Building, 1025 Walnut Street, Philadelphia, PA 19107, USA.

Dr. Jerry Fredenburgh Chief Executive Officer, Fredenburgh Consultants, 3403 Cypress Drive, Spring Grove, IL 60081-8617, USA.

Prof. Dr. Stefan Fruehauf Professor, Center for Tumor Diagnostics and Therapy, Paracelsus-Klinik, Am Natruper Holz 69, D-49090 Osnabrück, Germany.

Dr. Anna-Paula Giulietti Scientific Researcher, Laboratory for Experimental Medicine and Endocrinology (LEGENDO), Catholic University of Leuven, UZ Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.

Dr. Andrew K. Godwin Member, Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111-2497, USA.

Dr. Bert Gold Staff Scientist, Human Genetics Section, Laboratory of Genomic Diversity, National Cancer Institute at Frederick, Building 560, Room 21-21, Frederick, MD, 21702, USA.

Jenny Göransson Graduate Student, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Prof. Dr. William E. Grizzle Professor of Pathology, Department of Pathology, University of Alabama at Birmingham, Zeigler Research Building – Room 408, 703 South 19th Street Birmingham, AL 35294-0007, USA.

Dr. Piotr Grodzinski Center for Strategic and Scientific Initiative, National Cancer Institute, Bethesda, MD, USA.

Ida Grundberg Graduate Student, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Dr. Jean-Louis Guénet Emeritus Scientist, Institut Pasteur, France.

Dr. Ivo G. Gut Assistant Director and Department Head, Department for Translational Research, Centre National de Génotypage, CEA – Institut de Génomique, 2 rue Gaston Crémieux, F-91000 Evry Cedex, France.

Prof. Dr. Samir Hanash Professor and Head, Department of Molecular Diagnostics, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, M5-C800, Seattle, WA 98109, USA.

Dr. Claus Hellerbrand Assistant Professor, University Hospital Regensburg, Department of Internal Medicine I, Franz-Josef-Strauß-Allee 11, D-93042 Regensburg, Germany.

Sara Henriksson Graduate Student, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Dr. Federico Innocenti Assistant Professor, University of Chicago, Department of Medicine, Section of Hematology/Oncology, 5841 South Maryland Avenue, MC 2115, Chicago, IL 60637, USA.

Magnus Isaksson Graduate Student, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Malin Jarvius Graduate Student, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Prof. Dr. Jens L. Jensen Professor, Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, 8200-Aarhus N, Denmark

Yann Joly Research Scientist, University of Montreal, Public Law Research Center, Montréal, Québec, Canada.

Dr. Masood Kamali-Moghaddam Associate Professor, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Dr. Kylene Kehn-Hall Assistant Research Professor, Department of Microbiology, Immunology, and Tropical Medicine, The George Washington University, 2300 I Street, NW Ross Hall, Room 551, Washington, DC 20037, USA.

Thomas Köcher IMP – Research Institute of Molecular Pathology, Vienna, Austria.

Kaarel Krjutskov Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, Tartu, Estonia and Asper Biotech, Tartu, Estonia.

Dr. Mogens Kruhøffer Associate Professor, Molecular Diagnostic Laboratory, Department of Clinical

Biochemistry, Aarhus University Hospital, Skejby, 8200-Aarhus N, Denmark.

Ants Kurg Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, Tartu, Estonia and Asper Biotech, Tartu, Estonia.

Andreana Lambrinakos Ph.D. Scholar, Genomic Disorders Research Centre, 7th Floor, Daly Wing, St. Vincent's Hospital, 41 Victoria Parade, Fitzroy 3065, Melbourne, Victoria, Australia.

Prof. Dr. Ulf Landegren Professor, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Chatarina Larsson Graduate Student, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Dr. Stephanie Laufs Research Scientist, Department of Experimental Surgery, Medical Faculty Mannheim, University of Heidelberg, and Department of Molecular Oncology of Solid Tumors Unit, German Cancer Research Center (DKFZ), Heidelberg, Germany; Deutsches Krebsforschungszentrum (DKFZ), G360 Klinische Kooperationseinheit Molekulare Onkologie Solider Tumoren, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany.

Merike Leego Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, Tartu, Estonia and Asper Biotech, Tartu, Estonia.

Karl-Johan Leuchowius Graduate Student, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Ilona Lind Asper Biotech, Tartu, Estonia and Tartu University Hospital, Estonia.

Dr. Robin Liu Molecular Diagnostics, Osmetech Molecular Diagnostics, Pasadena, CA, USA.

Dr. Patrick W. Maier Research Scientist, Department of Radiation Oncology, Mannheim Medical Centre, University of Heidelberg; Universitätsklinik für Strahlentherapie und Radioonkologie, Universitätsklinikum Mannheim, Universität Heidelberg, Theodor-Kutzer-Ufer 1-3, D-68167 Mannheim, Germany.

Sharon Marsh Génome Québec and Montreal Heart Institute Pharmacogenomics Centre, Montreal, Quebec, Canada.

Dr. Chantal Mathieu Associate Professor of Medicine, Laboratory for Experimental Medicine and Endocrinology (LEGENDO), Catholic University of Leuven, UZ Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.

Dr. Panayiotis G. Menounos Director, Laboratory of Research, Nursing Military Academy, SAKETA "A" barrack, Vironas GR-16201, Athens, Greece.

Andres Metspalu Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, Tartu, Estonia, Asper Biotech, Tartu, Estonia and Estonian Genome Project, University of Tartu.

Dr. Emmanuelle Nicolas Research Associate, Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111-2497, USA.

Dr. Mats Nilsson Associate Professor, Uppsala University, Department of Genetics and Pathology, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Eneli Oitmaa Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, Tartu, Estonia and Asper Biotech, Tartu, Estonia.

Prof. Dr. Torben F. Ørntoft Professor, Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, 8200-Aarhus N, Denmark.

Dr. Lut Overbergh Scientific Researcher, Laboratory for Experimental Medicine and Endocrinology (LEGENDO), Catholic University of Leuven, UZ Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.

Dr. Franco Pagani Staff Scientist, International Centre for Genetic Engineering and Biotechnology (ICGEB), Padriciano, 99, 34012 Trieste, Italy.

Dr. Adamandia Papachatzopoulou Assistant Professor, University of Patras, Faculty of Medicine, School of Health Sciences, Department of General Biology, Asclepiou Street, Panepistimioupolis, GR-26110, Patras, Greece.

Dr. Katerina Pardali Post-Doctoral Fellow, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Dr. Andrea Farkas Patenaude Director of Psycho-Oncology Research and Assistant Professor of Psychology, Division of Pediatric Oncology, Dana-Farber Cancer Institute; Department of Psychiatry, Harvard Medical School, 44 Binney Street, Boston MA 02445, USA.

Dr. George P. Patrinos Assistant Professor and Head, Pharmacogenomics Group, University of Patras, School of Health Sciences, Department of Pharmacy University Campus, GR-26504, Rion, Patras, Greece; Adjunct Assistant Professor, Erasmus University Medical Center, Faculty of Medicine and Health Sciences, Department of Bioinformatics, Room Ee1579, PO Box 2040, 3000 CA, Rotterdam, the Netherlands.

Dr. Minoli Perera Instructor, University of Chicago, Department of Medicine, Section of Genetic Medicine, 5841 South Maryland Ave, Room TS 651, MC 6091, Chicago IL, 60637, USA.

Dr. Hartmut Peters Research Scientist, Institute of Medical Genetics, Charité-Universitätsmedizin, Humboldt University Berlin, Augustenburger Platz 1, D-13353, Berlin, Germany.

Dr. Mary Petrou University College London, Institute of Women's Health, Haemoglobinopathy Genetics Centre (Perinatal Centre), University College London Hospitals NHS Foundation Trust, London, UK.

Dr. Martin Philpott Senior Research Fellow, Discipline of Nutrition, Faculty of Medical and Health Sciences,

University of Auckland, Private Bag 92019, Auckland, New Zealand.

Dr. John V. Planz Assistant Professor and Associate Director, DNA Identity Laboratory, Department of Pathology and Anatomy, University of North Texas Health Science Center, 3500 Camp Bowie Boulevard, Ft. Worth, Texas 76107, USA.

Janne Pullat Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, Tartu, Estonia.

Dr. Ji Qiu Senior Scientist, Department of Molecular Diagnostics, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N, M5-C800, Seattle, WA 98109, USA.

Dr. Gudrun H. Reed Senior Research Specialist, University of Utah School of Medicine, Department of Pathology, 50 N Medical Drive, Salt Lake City, UT 84132, USA.

Kylee Rees Ph.D. Scholar, Genomic Disorders Research Centre, 7th Floor, Daly Wing, St. Vincent's Hospital, 41 Victoria Parade, Fitzroy 3065, Melbourne, Victoria, Australia.

Dr. Peter N. Robinson Staff Scientist, Institute of Medical Genetics, Charité-Universitätsmedizin, Humboldt University Berlin, Augustenburger Platz 1, D-13353, Berlin, Germany.

Georgina Sallmann Research Assistant, Genomic Disorders Research Centre, 7th Floor, Daly Wing, St. Vincent's Hospital, 41 Victoria Parade, Fitzroy 3065, Melbourne, Victoria, Australia.

Jan P. Schouten MRC-Holland BV, Amsterdam, The Netherlands.

Luisa Schubert Student, Innovative Cancer Diagnostic and Therapy Unit, German Cancer Research Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany.

Kathryn E. Scott Molecular Research Specialist, Thomas Jefferson University, Department of Medicine, Center for Translational Medicine Genomics, 415 College Building, 1025 Walnut Street, Philadelphia, PA 19107, USA.

Eline M. Sepers MRC-Holland BV, Amsterdam, The Netherlands.

Prof. Dr. Karen Sermon Professor, Centre for Medical Genetics, UniversitairZiekenhuis Brussel, Department of Embryology and Genetics, Vrije Universiteit Brussel, Laarbeeklaan 101, B-1090 Brussels, Belgium.

Dr. Ola Söderberg Associate Professor, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Dr. Marie Stampe-Ostenfeld Post-Doctoral Fellow, Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, 8200-Aarhus N, Denmark.

Dr. Eleana Stavrou Research Scientist, University of Patras, Faculty of Medicine, School of Health Sciences, Department of General Biology, Asclepiou Street, Panepistimioupolis, GR-26110, Patras, Greece.

Dr. Johan Stenberg Post-Doctoral Fellow, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Giulio Superti-Furga Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria.

Prof. Dr. Saul Surrey Professor of Medicine and Associate Director of Research, Cardeza Foundation for Hematologic Research and Division of Hematology, Thomas Jefferson University, 703 Curtis Building, 1015 Walnut Street, Philadelphia, PA 19107, USA.

Prof. Dr. Graham. R. Taylor Head of Genomic Services, Regional Genetics Laboratory and Cancer Research UK, Mutation Detection Facility, St. James's University Hospital, DNA laboratory, Room 6.2 Clinical Science Building, Leeds, Yorkshire LS9 7TF, United Kingdom.

Dr. Kasper Thorsen Post-Doctoral Fellow, Molecular Diagnostic Laboratory, Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, 8200-Aarhus N, Denmark.

Neeme Tonisson Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, Tartu, Estonia and Tartu University Hospital, Estonia.

Dr. Holger Tönnies Research Scientist, Cytogenetics and Molecular Cytogenetics Laboratory, Institute of Human Genetics, Christian-Albrechts-University, Kiel University Hospital, Schleswig-Holstein, D-24105 Kiel, Germany.

Dr. Jörg Tost Group Leader, Laboratory for Epigenetics, Centre National de Génotypage, CEA – Institut de Génomique, Bâtiment G2, 2 rue Gaston Crémieux, CP 5721, F-91057 Evry Cedex, France.

Dr. Ian Trounce Associate Director, Genomic Disorders Research Centre, 7th Floor, Daly Wing, St. Vincent's Hospital, 41 Victoria Parade, Fitzroy 3065, Melbourne, Victoria, Australia.

Dirk Valckx Research Associate, Laboratory for Experimental Medicine and Endocrinology (LEGENDO), Catholic University of Leuven, UZ Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.

Dr. Michael Ward Invitrogene Corp, Eugene, OR, USA

Irene Weibrecht Graduate Student, Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85, Uppsala, Sweden.

Prof. Dr. Carl T. Wittwer Professor of Pathology, University of Utah School of Medicine, Department of Pathology, ARUP Institute for Clinical and Experimental Pathology, 50 N Medical Drive, Salt Lake City, UT 84132, USA.

Dr. Anthony T. Yeung Director, Fannie E. Rippel Biochemistry and Biotechnology Facility; Member, Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111-2497, USA.

Prof. Dr. W. Jens Zeller Group Leader, Innovative Cancer Diagnostic and Therapy Unit, German Cancer Researh Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany.

Preface – First Edition

This book is the result of the initiative to edit a reference publication dealing exclusively with all the aspects of modern molecular diagnostics. The need for such a book derives from the intellectual revolution in biomedical science and the realization that human diseases, both congenital and inherited, are rapidly becoming susceptible to molecular quantitative analysis. Our effort has been assisted by many world-leading scientists, experts in their field, who have kindly accepted our invitation to compile the 48 chapters of this book and share with us and our readers their expertise, experience, and results, sometimes even unpublished ones.

In the post-genomic era at the dawn of the twenty-first century, the deciphering of the information encrypted into the human genome sequence, together with the expansion of our knowledge on the molecular basis of human inherited disorders, brought up the need to provide diagnosis at the molecular level for inherited diseases easily, accurately, and quickly. As a consequence, an increasing number of diagnostic and research laboratories, both academic and private in the USA and in Europe, are preparing to deal with the demands and challenges this new era sets forth for medicine and health sciences. They move in that direction not only by establishing innovative approaches for mutation detection, or improving existing ones, but also by implementing novel technologies for the benefit of the patient.

Despite the fact that molecular diagnostics has already entered its golden era, the scientific literature lacks a specialized reference book, dealing exclusively with the technology and techniques available for the detection of the sequence variations leading to inherited diseases. We consider this reference book, together with the underlying framework, to be fundamental for the fruitful synthesis of all aspects of modern molecular diagnostics to result in a well-orchestrated entity. There are only a few books available in the literature that allude to these issues in the topics they deal with, and none is dedicated exclusively to diagnosis at the molecular level. On the other hand, a great number of internet sites are scattered throughout the World Wide Web, which serve as tools, for example, for implementing mutation screening methodologies. Unfortunately, in their vast majority, these sites possess the inherent dangers of being outdated, unclear, and hence confusing, and are sometimes difficult and time consuming to find. Also, they often lack the important aspect of being compiled and based on many years of relevant experience of well-known expert researchers, leaders in their fields.

The innovative aspect of this book is its structure. It contains an expert introduction to each subject, next to only few protocols and technical details, which can be found in comprehensive reference lists at the end of each chapter. The contents of this book are divided in two parts. The first part is dedicated to the battery of the most widely used molecular biology techniques. Their arrangement is in a more or less chronologic order of their development. Starting from the first ones with relatively low throughput, like single-strand conformation polymorphism analysis, allele-specific amplification, temperature and denaturing gradient gel electrophoresis, and others, to more modern techniques, such as fluorescence *in situ* hybridization, denaturing high-performance liquid chromatography, two-dimensional gene scanning, microarrays, mass spectrometry, proteomic analysis, and others. These techniques are still evolving fast, becoming more cost effective, and are characterized by high throughput. Each of these chapters includes the principle and a brief description of the technique, followed by examples from the area of expertise from the selected contributor. The selected contributors are well-known experts in their fields and come from a variety of disciplines, so that the book covers efficiently not only a great number of techniques applied to molecular diagnosis but also their applications to a variety of inherited disorders.

The second part attempts to integrate the previously mentioned techniques to the different aspects of molecular diagnosis, such as identification of genetically modified organisms, stem cells, pharmacogenetics, modern forensic science, genetic quality of laboratory animals, molecular microbiology, and preimplantation genetic diagnosis. We believe that issues such as personalized medicine, pharmacogenetics, and integration of diagnostics and therapeutics will be subjects for debate in the following years. In addition, various everyday issues in a diagnostic laboratory are discussed, such as the establishment of mutation databases to store and organize the continuously growing information

on gene variation, genetic counseling, patenting of genes and of genetic tests, safety and quality management, and various ethical considerations and psychological issues pertaining to diagnostics. We feel that the inclusion of the latter issues in this reference book have great relevance to our society and the aim is to assist in finding answers to some of the problematic questions that undoubtedly will arise, since application of molecular diagnostics may precipitate an important ethical crisis that physicians and the communities they serve will be confronted with.

The intended audience of this book is university postgraduate students from various life sciences disciplines, physicians, scientists in human molecular genetics and medicine, professionals working in diagnostic laboratories in academia or industry, academic institutions, hospital libraries, biotechnology and pharmaceutical companies. We believe that this book will be of help to decision-making advisors in medical insurance companies. In addition, undergraduate medical and life science students will find very useful the description and explanation of recent modern techniques in life sciences. A major concern was to formulate the book contents in such a way that the notions described therein are clear and explained in a simple language and terminology. The numerous illustrations of the book are comprehensive and self-descriptive, and the glossary at the end of this book provides a brief explanation of the most commonly found terms.

We expect that some points in this book can be further improved. We would welcome comments and criticism from the attentive readers, which will contribute to the improvement of the content of this book in its future editions as well as helping to establish it as a reference publication in the field of molecular diagnostics.

Without the support and contributions of many people, the completion of this book would not be possible. We wish to thank the anonymous referees who supported our proposal and subsequently the people at Elsevier, in particular Dr. Tessa Picknett and her department, whose interest in this title has encouraged us to proceed.

We are also grateful to the editors, Drs. Claire Minto and Tari Paschall, and the senior production manager, Sarah Hajduk, at Elsevier, who helped us in a close collaboration to solve and overcome encountered difficulties. Their efficiency, pleasant manner, and patience added immensely to the smooth completion of this project from the very start.

We also express our gratitude to all contributors for delivering outstanding compilations that summarize their experience and many years of hard work in their field of research. We are indebted to Julio Esperas who was responsible for the design and the cover of this book and to the copy editor, Adrienne Rebello, who has refined the final manuscript prior to going into production. We also thank our university colleagues, in particular Prof. Miel Ribbe from the Medical School of the University of Amsterdam, for discussions of the content of this book and ensuring its greatest possible relevance. We owe our thanks to the academic reviewers for their constructive criticisms on the chapters.

We wish also to thank our families, from whom we have taken considerable amounts of time to dedicate to this work and whose patience and support have been conducive to the successful completion of this project.

George P. Patrinos
Erasmus Medical Centre
Rotterdam, The Netherlands

Wilhelm J. Ansorge
European Molecular
Biology Laboratory
Heidelberg, Germany

June 2005

Preface – Second Edition

Four years after the publication of the first edition of *Molecular Diagnostics*, we are pleased to deliver the second edition of this textbook. Molecular diagnostics is a rapidly evolving field, particularly after the deciphering of the human genome sequence. The need for a second edition of such a book derives not only from the recent discoveries and intellectual revolution in biomedical science but also from the rapid technological advancements applicable to low-cost and high-throughput molecular diagnostics, particularly pertaining to nanotechnology.

Over these last four years that this textbook has been available to the scientific community, it has been adopted as key reference in the field. This is clearly demonstrated from (a) the large number of copies sold worldwide in a relatively short time, (b) the various post-graduate and specialist training courses on molecular diagnostics that have been used as a syllabus, (c) the adoption by universities as a textbook for related undergraduate courses and curricula, which also led to its translation in 2008, (d) the various positive reviews obtained not only from external reviewers, e.g. Doody's and scientific journals, but also from fellow academics and students. The above have encouraged us and prompted Elsevier/Academic Press to move forward with the compilation of a second edition. As with the previous edition, our effort has been assisted by many internationally renowned experts in their field, who have kindly accepted our invitation to compile the 40 chapters of this book and share with us and our readers their expertise, experience, and results.

The second edition kept the original structure of the first edition, which was one of its main innovative aspects. Each chapter contains an expert introduction to each subject, next to few technical details and many applications for molecular genetic testing, which can be found in comprehensive reference lists at the end of each chapter. The contents of this book are divided in two parts. The first part is dedicated to the battery of the most widely used molecular biology techniques. Their arrangement is in a more or less chronologic order of their development. In order to keep pace with the recent developments, few chapters from the first edition have been omitted, others have been merged while a large number of chapters pertaining to high-throughput molecular diagnostic approaches, e.g. oligonucleotide microarrays, next-generation sequencing, mass spectrometry, MLPA, etc., have been included. The remaining chapters have been updated compared to the previous edition, to include not only technology innovations but also novel diagnostic applications. This resulted in the book being completely updated with over half of its content being new compilations. As with the first edition, each of these chapters includes the principle and a brief description of the technique, followed by examples from the area of expertise from the selected contributor.

The second part attempts to integrate the previously analyzed technology to the different aspects of molecular diagnostics, such as identification of genetically modified organisms, stem cells, pharmacogenomics, modern forensic science, genetic quality of laboratory animals, molecular microbiology, and preimplantation genetic diagnosis. Again, new chapters on areas where molecular genetic testing became relevant in the recent years, such as biodefense, victim identification in mass disasters, nutritional genomics (nutrigenomics), etc., have been included in this second edition. Finally, various everyday issues in a diagnostic laboratory, from locus-specific and national/ethnic mutation databases to gene patents and genetic counseling and related ethical and psychological issues to safety and quality management, are discussed. As with the first edition, we feel that the inclusion of the latter issues in this reference book has still great relevance to our society.

In addition, we made an effort to formulate the book contents such that the notions described therein are explained in a simple language and terminology for the book to be useful not only to experienced physicians and health care specialists and academics but also to undergraduate medical and life science students, and the numerous self-explanatory illustrations and glossary clearly contribute to this end. Next, we tried to be consistent with the official gene and genetic variation nomenclature throughout the compilation.

We are grateful to those colleagues who constructively criticized the first edition and identified deficiencies

that have been, hopefully, rectified in this second edition. However, we expect that some points in this book can still be further improved. Therefore, we would welcome comments and criticisms from attentive readers, which will contribute to improve the contents of this book even further in its future editions. We are also grateful to the editors, Tari Broderick, April Graham, and Janice Audet at Elsevier, who helped us in close collaboration to overcome encountered difficulties. We also express our gratitude to all contributors for delivering outstanding compilations that summarize their experience and many years of hard work in their field of research. We are indebted to Greg Harris who was responsible for cover design of this book and to the Elsevier production team (Alan Everett, Joe Howarth, Deena Burgess and Kim Lander) who refined the final manuscript. We owe our thanks to the academic reviewers for their constructive criticisms on the chapters and their positive evalution of our proposal for the second edition.

Last, but not least, we wish to cordially thank our families for their patience and continuous support over the years, from whom we have taken a considerable amount of time to devote to this project.

George P. Patrinos
University of Patras, Faculty of Health Sciences, Department of Pharmacy
Patras, Greece
Erasmus MC, Faculty of Medicine and Health Sciences, Department of Bioinformatics
Rotterdam, The Netherlands

Wilhelm J. Ansorge
Ecole Polytechnique Federal Lausanne,
Lausanne, Switzerland

September 2009

Foreword – First Edition

The recent pace of discoveries in human genetics and molecular biology is stunning. Completion of the human genome sequence analysis, successive development of diverse high-throughput analytical approaches (genomics, proteomics, transcriptomics), and most recently, developments in nano-technology have permitted innovative enhancement of genetic testing. Elucidation of the molecular bases of numerous human diseases, combined with rapidly evolving techniques in molecular biology, have brought molecular diagnostics to a high stage of sophistication which is still evolving.

This textbook on molecular diagnostics is ambitious in its scope and intended usage. It addresses not only scientists already in the field, but also those who intend to enter it. It provides a much-needed comprehensive reference publication on current and developing methods. Importantly, it also includes an innovative treatment of diverse molecular diagnostic techniques as they apply to the actual practice of clinical microbiology, forensic analysis, development of pharmaceuticals, and other fields.

The first section of the book covers the principles and applications of numerous genetic testing methodologies, in approximately the chronological order of discovery. This section includes chapters on both targeted and high-throughput diagnosis using genomic, transcriptomic, and proteomic approaches.

The second section is devoted to the application of molecular diagnostics in a variety of fields. It illustrates the vast potential as well as the versatility of molecular diagnostics, which are applicable not only to modern medical practice but also to the identification of genetically modified organisms, forensics, quality assessment of laboratory animals, pharmacogenomics, and other fields. Of special interest in this section are the discussions of recently emerged issues, such as gene patenting, ethics, safety, and the psychological context of genetic testing.

The editors succeeded in attracting more than 100 well-qualified authors to compile the 48 chapters of this book. The result is an up-to-date comprehensive treatment of molecular diagnostics, which aims to be a standard reference resource for biomedical scientists from a wide range of disciplines. It will be invaluable for clinicians and others who come into close contact with genetic testing for diagnostic purposes. It will also promote a wider understanding of molecular diagnostics among physicians, postgraduate students, researchers in the academic and corporate world, and even policymakers.

The self-explanatory nature of the book, assisted by a glossary and numerous illustrations, will also appeal to undergraduate students in the medical and life sciences. The editors can be congratulated for putting together such a valuable resource, for an area of the biomedical sciences that is of major and ever-increasing importance.

Prof. Fotis C. Kafatos
Director-General
European Molecular Biology Laboratory
Heidelberg, Germany

August 2004

This page intentionally left blank

Chapter 1

Molecular Diagnostics: Past, Present, and Future

George P. Patrinos[1,2] and Wilhelm J. Ansorge[3]

[1] *Department of Pharmacy, School of Health Sciences, University of Patras, Patras, Greece;*

[2] *Erasmus University Medical Center, Faculty of Medicine and Health Sciences, Department of Bioinformatics, Rotterdam, The Netherlands;*

[3] *Ecole Polytechnique Federal Lausanne, EPFL, Lausanne, Switzerland*

1.1 INTRODUCTION

Molecular or nucleic acid-based diagnosis of human disorders is referred to as the detection of the various pathogenic mutations in DNA and/or RNA samples in order to facilitate detection, diagnosis, subclassification, prognosis, and monitoring response to therapy. Molecular diagnostics combines laboratory medicine with the knowledge and technology of molecular genetics and has been enormously revolutionized over the last decades, benefiting from the discoveries in the field of molecular biology (see Table 1.1). The identification and fine characterization of the genetic basis of the disease in question is vital for accurate provision of diagnosis. Gene discovery provides invaluable insights into the mechanisms of disease, and gene-based markers allow physicians not only to assess disease predisposition but also to design and implement improved diagnostic methods. The latter is of great importance, as the plethora and variety of molecular defects demands the use of multiple rather than a single mutation detection platform. Molecular diagnostics is currently a clinical reality with its roots deep into the basic study of gene expression and function.

1.2 HISTORY OF MOLECULAR DIAGNOSTICS: INVENTING THE WHEEL

In 1949, Pauling and his coworkers introduced the term *molecular disease* into the medical vocabulary, based on their discovery that a single amino acid change at the β-globin chain leads to sickle cell anemia, characterized mainly by recurrent episodes of acute pain due to vessel occlusion. In principle, their findings have set the foundations of molecular diagnostics, although the big revolution occurred many years later. At that time, when molecular biology was only hectically expanding, the provision of molecular diagnostic services was inconceivable and technically not feasible. The first seeds of molecular diagnostics were provided in the early days of recombinant DNA technology, with many scientists from various disciplines working in concert. cDNA cloning and sequencing were at that time invaluable tools for providing the basic knowledge on the primary sequence of various genes. The latter provided a number of DNA probes, allowing the analysis via Southern blotting of genomic regions, leading to the concept and application of restriction fragment length polymorphism (RFLP) to track a mutant allele from heterozygous parents to a high-risk pregnancy. In 1976, Kan and coworkers carried out, for the first time, prenatal diagnosis of α-thalassemia, using hybridization on DNA isolated from fetal fibroblasts. Also, Kan and Dozy, in 1978, implemented RFLP analysis to pinpoint sickle cell alleles of African descent. This breakthrough provided the means of establishing similar diagnostic approaches for the characterization of other genetic diseases, such as phenylketonurea (Woo *et al.*, 1983), cystic fibrosis (Farrall *et al.*, 1986), and so on.

At that time, however, a significant technical bottleneck had to be overcome. The identification of the disease causing mutation was possible only through the construction of a genomic DNA library from the affected individual, in order first to clone the mutated allele and then determine its nucleotide sequence. Again, many human globin gene mutations were among the first to be identified through such approaches (Busslinger *et al.*, 1981; Treisman *et al.*, 1983). In 1982, Orkin and his coworkers showed that a number of sequence variations were linked to specific β-globin gene mutations. These groups of RFLPs, termed *haplotypes* (both intergenic and intragenic), have provided a

TABLE 1.1 The timeline of the principal discoveries in the field of molecular biology, which influenced the development of molecular diagnostics.

Date	Discovery
1949	Characterization of sickle cell anemia as a molecular disease
1953	Discovery of the DNA double helix
1958	Isolation of DNA polymerases
1960	First hybridization techniques
1969	*In situ* hybridization
1970	Discovery of restriction enzymes and reverse transcriptase
1975	Southern blotting
1977	DNA sequencing
1983	First synthesis of oligonucleotides
1985	Restriction fragment length polymorphism analysis
1985	Invention of PCR
1986	Development of fluorescent *in situ* hybridization (FISH)
1988	Discovery of the thermostable DNA polymerase – Optimization of PCR
1992	Conception of real-time PCR
1993	Discovery of structure-specific endonucleases for cleavage assays
1996	First application of DNA microarrays
2001	First draft versions of the human genome sequence
2001	Application of protein profiling in human diseases
2005	Introduction of the high-throughput next-generation sequencing technology

first-screening approach in order to detect a disease-causing mutation. Although this approach enabled researchers to predict which β-globin gene would contain a mutation, significantly facilitating mutation screening, no one was in the position to determine the exact nature of the disease-causing mutation, as many different β-globin gene mutations were linked to a specific haplotype in different populations (further information is available at http://globin.bx.psu.edu/hbvar; Hardison *et al.*, 2002; Patrinos *et al.*, 2004; Giardine *et al.*, 2007).

At the same time, in order to provide a shortcut to DNA sequencing, a number of exploratory methods for pinpointing mutations in patients' DNA were developed. The first methods involved mismatch detection in DNA/DNA or RNA/DNA heteroduplexes (Myers *et al.*, 1985a, b) or differentiation of mismatched DNA heteroduplexes using gel electrophoresis, according to their melting profile (Myers *et al.*, 1987). Using this laborious and time-consuming approach, a number of mutations or polymorphic sequence variations have been identified, which made possible the design of short synthetic oligonucleotides that were used as allele-specific probes onto genomic Southern blots. This experimental design was quickly implemented for the detection of β-thalassemia mutations (Orkin *et al.*, 1983; Pirastu *et al.*, 1983).

Despite the intense efforts from different laboratories worldwide, diagnosis of inherited diseases on the DNA level was still underdeveloped and therefore still not ready to be implemented in clinical laboratories for routine analysis of patients due to the complexities, costs, and time requirements of the technology available. It was only after a few years that molecular diagnosis entered its golden era with the discovery of the most powerful molecular biology tool since cloning and sequencing, the polymerase chain reaction (PCR).

1.3 THE PCR REVOLUTION: GETTING MORE OUT OF LESS

The discovery of PCR (Saiki *et al.*, 1985; Mullis and Faloona, 1987) and its quick optimization, using a thermostable *Taq* DNA polymerase from *Thermus aquaticus* (Saiki *et al.*, 1988) has greatly facilitated and in principle revolutionized molecular diagnostics. The most powerful feature of PCR is the large amount of copies of the target sequence generated by its exponential amplification (see Fig. 1.1), which allows the identification of a known mutation within a single day, rather than months. Also, PCR has markedly decreased or even diminished the need for radioactivity for routine molecular diagnosis. This has allowed molecular diagnostics to enter the clinical laboratory for the provision of genetic services, such as carrier or population screening for known mutations, prenatal diagnosis of inherited diseases, or in recent years, identification of unknown mutations, in close collaboration with research laboratories. Therefore, being moved to their proper environment, the clinical laboratory, molecular diagnostics could provide the services for which they have been initially conceived.

The discovery of PCR has also provided the foundations for the design and development of many mutation detection schemes, based on amplified DNA. In general, PCR either is used for the generation of the DNA fragments to be analyzed, or is part of the detection method. The first attempt was the use of restriction enzymes (Saiki *et al.*, 1985) or oligonucleotide probes, immobilized onto membranes or in solution (Saiki *et al.*, 1986) in order to detect the existing genetic variation, in particular the sickle cell disease-causing

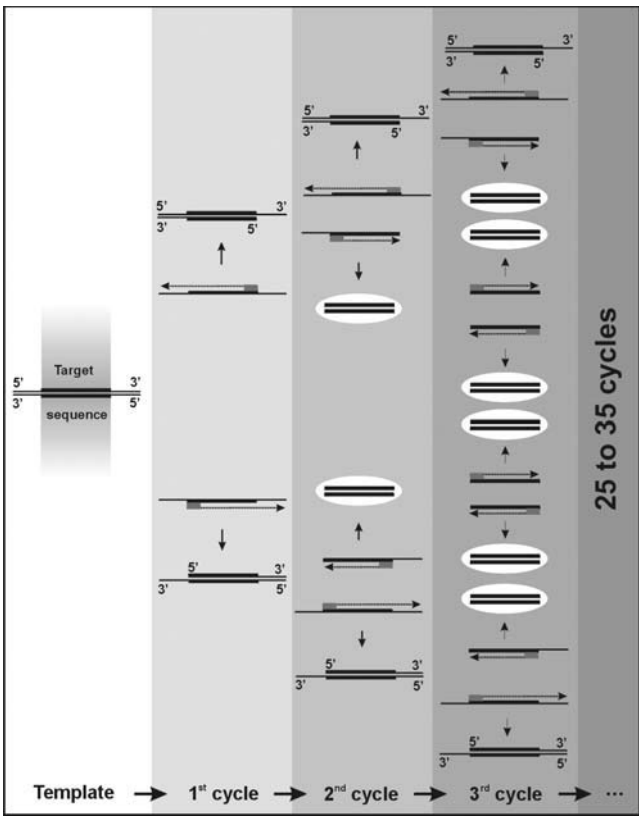


FIGURE 1.1 The PCR principle. Thick and thin black lines correspond to the target sequence and genomic DNA, respectively; gray boxes correspond to the oligonucleotide primers, and the correct size PCR products are included in the white ellipses. Dashed lined arrows depict the elongation of the template strand.

mutation. In the following years, an even larger number of mutation detection approaches have been developed and implemented. These techniques can be divided roughly into three categories, depending on the basis for discriminating the allelic variants:

1. *Enzymatic-based methods*. RFLP analysis was historically the first widely used approach, exploiting the alterations in restriction enzyme sites, leading to the gain or loss of restriction events (Saiki *et al.*, 1985). Subsequently, a number of enzymatic approaches for mutation detection have been conceived, based on the dependence of a secondary structure on the primary DNA sequence. These methods exploit the activity of resolvase enzymes T4 endonuclease VII, and, more recently, T7 endonuclease I to digest heteroduplex DNA formed by annealing wild-type and mutant DNA (Mashal *et al.*, 1995). Digestion fragments indicate the presence and the position of any mutations. A variation of the theme involves the use of chemical agents for the same purpose (Saleeba *et al.*, 1992; see also Chapter 3). Another enzymatic approach for mutation detection is the oligonucleotide ligation assay (Landegren *et al.*, 1988). In this technique, two oligonucleotides are hybridized to complementary DNA stretches at sites of possible mutations. The oligonucleotides' primers are designed such that the 3′ end of the first primer is immediately adjacent to the 5′ end of the second primer. Therefore, if the first primer matches completely with the target DNA, then the primers can be ligated by DNA ligase. On the other hand, if a mismatch occurs at the 3′ end of the first primer, then no ligation products will be obtained.
2. *Electrophoretic-based techniques*. This category is characterized by a plethora of different approaches designed for screening of known or unknown mutations, based on the different electrophoretic mobility of the mutant alleles, under denaturing or non-denaturing conditions. Single strand conformation polymorphism (SSCP) and heteroduplex (HDA) analyses (Orita *et al.*, 1989; see Chapter 4) were among the first methods designed to detect molecular defects in genomic loci. In combination with capillary electrophoresis (see Chapter 5), SSCP and HDA analysis now provide an excellent, simple, and rapid mutation detection platform with low operation costs and, most interestingly, the potential of easily being automated, thus allowing for high-throughput analysis of patients' DNA. Similarly, denaturing and temperature gradient gel electrophoresis (DGGE and TGGE, respectively) can be used equally well for mutation detection (see Chapter 6). In this case, electrophoretic mobility differences between a wild-type and mutant allele can be ''visualized'' in a gradient of denaturing agents, such as urea and formamide, or of increasing temperature. Finally, an increasingly used mutation detection technique is the two-dimensional gene scanning, based on two-dimensional electrophoretic separation of amplified DNA fragments, according to their size and base pair sequence. The latter involves DGGE, following the size separation step.
3. *Solid phase-based techniques*. This set of techniques consists of the basis for most of the present-day mutation detection technologies, since they have the extra advantage of being easily automated and hence are highly recommended for high-throughput mutation detection or screening. A fast, accurate, and convenient method for the detection of known mutations is reverse dot-blot, initially developed by Saiki and coworkers (1989) and implemented for the detection of β-thalassemia mutations. The essence of this method is the utilization of oligonucleotides, bound to a membrane, as hybridization targets for amplified DNA. Some of this technique's advantages is that one membrane strip can be used to detect many different known mutations in a single individual (a one strip/one patient type of assay), the potential of automation, and the ease of interpretation of the results, using a classical avidin-biotin system. However, this technique cannot be used

for the detection of unknown mutations. Continuous development has given rise to allele-specific hybridization of amplified DNA (PCR-ASO, Chapter 2) on filters and recently extended on DNA oligonucleotide microarrays (see Chapters 16 and 17) for high throughput mutation analysis (Gemignani *et al.*, 2002; Chan *et al.*, 2004). In particular, oligonucleotides of known sequence are immobilized onto appropriate surfaces and hybridization of the targets to the microarray is detected, mostly using fluorescent dyes.

The choice of the mutation detection method is dependent upon a number of variables, including the mutation spectrum of a given inherited disorder, the available infrastructure, and the number of tests performed in the diagnostic laboratory, and recently with issues of intellectual properties (see also section 1.5.1 and Chapter 36). Most of the clinical diagnostic laboratories have not invested in an expensive high technology infrastructure, since the test volumes, that is, the number of tests expected to be performed, have not been large enough to justify the capital outlay. Therefore, simple screening tests such as SSCP and HDA were and still are the methods of choice for many clinical laboratories, as they allow for rapid and simultaneous detection of different sequence variations at a detection rate of close to 100%. Although PCR has significantly facilitated the expansion of molecular diagnostics, it nonetheless has a number of limitations. First of all, amplification of CG repeat-rich regions can be problematic for *Taq* polymerase, which sometimes leads to the classic alternative of Southern blot analysis. Also, *Taq* polymerase is error prone at a range of 10^{-4} to 10^{-5} per nucleotide, which is strongly influenced by the conditions of the amplification reaction, such as magnesium or deoxyribonucleotide concentration, pH, temperature, and so on. Polymerase errors can contribute to unspecific background, depending on the detection method, resulting in limiting the detection level. To overcome these technical problems, positive results should be confirmed by alternative methods or by using high fidelity thermostable polymerases.

Finally, it needs to be stressed that despite the wealth of mutation detection methodologies, DNA sequencing is still considered the gold standard and the definitive experimental procedure for mutation detection. However, the costs for the initial investment and the difficulties for standardization and interpretation of ambiguous results have restricted its use only to basic research laboratories.

1.4 MOLECULAR DIAGNOSTICS IN THE POST-GENOMIC ERA

In February 2001, with the announcement of the first draft sequence of the human genome (International Human Genome Sequencing Consortium, 2001; Venter *et al.*, 2001) and subsequently with the genomic sequence of other organisms, molecular biology has entered into a new era with unprecedented opportunities and challenges. These tremendous developments put pressure on a variety of disciplines to intensify their research efforts to improve by orders of magnitude the existing methods for mutation detection, to make available data sets with genomic variation and analyze these sets using specialized software, to standardize and commercialize genetic tests for routine diagnosis, and to improve the existing technology in order to provide state-of-the-art automated devices for high-throughput genetic analysis.

The biggest challenge, following the publication of the human genome draft sequence, was to improve the existing mutation detection technologies to achieve robust cost-effective, rapid, and high-throughput analysis of genomic variation. In the last couple of years, technology has improved rapidly and new mutation-detection techniques have become available, whereas old methodologies have evolved to fit into the increasing demand for automated and high-throughput screening. The chromatographic detection of polymorphic changes of disease-causing mutations using denaturing high-performance liquid chromatography (DHPLC; for review, see Xiao and Oefner, 2001) is one of the new technologies that emerged. DHPLC reveals the presence of a genetic variation by the differential retention of homo- and heteroduplex DNA on reversed phase chromatography under partial denaturation.

Single-base substitutions, deletions, and insertions can be detected successfully by UV or fluorescence monitoring within two to three minutes in unpurified PCR products as large as 1.5 kilo bases. These features, together with its low cost, make DHPLC one of the most powerful tools for mutational analysis. Also, pyrosequencing, a non-gel-based genotyping technology, provides a very reliable method and an attractive alternative to DHPLC (Chapter 8). Pyrosequencing detects *de novo* incorporation of nucleotides based on the specific template. The incorporation process releases a pyrophosphate, which is converted to ATP and followed by luciferase stimulation. The light produced, detected by a charge coupled device camera, is ''translated'' to a pyrogram, from which the nucleotide sequence can be deducted (Ronaghi *et al.*, 1998).

The use of the PCR in molecular diagnostics is considered the gold standard for detecting nucleic acids and it has become an essential tool in the research laboratory. Real-time PCR (Holland *et al.*, 1991) has engendered wider acceptance of the PCR due to its improved rapidity, sensitivity, and reproducibility (see Chapter 7). The method allows for the direct detection of the PCR product during the exponential phase of the reaction, therefore combining amplification and detection in one single step. The increased speed of real-time PCR is due largely to reduced cycles, removal of post-PCR detection procedures, and the use of fluorogenic labels and sensitive methods of detecting

their emissions. Therefore, real-time PCR is a very accurate and sensitive methodology with a variety of applications in molecular diagnostics, allows a high throughput, and can easily be automated and performed on very small volumes, which makes it the method of choice for many modern diagnostic laboratories.

Above all, the DNA microarray-based genotyping approach offers simultaneous analysis of many polymorphisms and sequence alterations (see Chapters 16 and 17). Microarrays consist of hundreds of thousands of oligonucleotides attached on a solid surface in an ordered array. The DNA sample of interest is PCR amplified and then hybridized to the microarray. Each oligonucleotide in the high-density array acts as an allele-specific probe and therefore perfectly matched sequences hybridize more efficiently to their corresponding oligonucleotides on the array. The hybridization signals, obtained from allele-specific arrayed primer extension (AS-APEX) (Pastinen *et al.*, 2000), are quantified by high-resolution fluorescent scanning and analyzed by computer software, resulting in the identification of DNA sequence alterations. Therefore, using a high-density microarray makes possible the simultaneous detection of a great number of DNA alterations, hence facilitating genome-wide screening. Several arrays have been generated to detect variants in the HIV genome (Kozal *et al.*, 1996; Wen *et al.*, 2000), human mitochondria mutations (Erdogan *et al.*, 2001), β-thalassemia (Chan *et al.*, 2004; Cremonesi *et al.*, 2007), and glycose-6-phosphate dehydrogenase (G-6-PD) deficiency mutations (Gemignani *et al.*, 2002), and so on.

In recent years, there has been a significant development of proteomics, which has the potential to become an indispensable tool for molecular diagnostics. A useful repertoire of proteomic technologies is available, with the potential to undergo significant technological improvements, which would be beneficial for increased sensitivity and throughput while reducing sample requirement (see Chapters 18 and 21). The improvement of these technologies is a significant advance toward the need for better disease diagnostics. The detection of disease-specific protein profiles goes back to the use of two-dimensional protein gels (Hanash, 2000), when it was demonstrated that leukemias could be classified into different subtypes based on the different protein profile (Hanash *et al.*, 2002). Nowadays, mass spectrometers are able to resolve many protein and peptide species in body fluids, being virtually set to revolutionize protein-based disease diagnostics (see Chapter 21). The robust and high-throughput nature of the mass spectrometric instrumentation is unparalleled and imminently suited for future clinical applications, as elegantly demonstrated by many retrospective studies in cancer patients (reviewed in Petricoin *et al.*, 2002). Also, high-throughput protein microarrays, constructed from recombinant, purified, and yet functional proteins, allow the miniaturized and parallel analysis of large numbers of diagnostic markers in complex samples. The first pilot studies on disease tissues are already starting to emerge, such as assessing protein expression profiles in tissue derived from squamous cell carcinomas of the oral cavity (Knezevic *et al.*, 2001), or the identification of proteins that induce an acute antibody response in autoimmune disorders, using auto-antigen arrays (Robinson *et al.*, 2002). These findings indicate that proteomic pattern analysis ultimately might be applied as a screening tool for cancer in high-risk and general populations.

The development of state-of-the-art mutation detection techniques has not only a positive impact on molecular genetic testing of inherited disorders, but also provides the technical means to other disciplines. Mutation detection schemes are applicable for the identification of genetically modified (GM) products, which may contaminate non-GM seeds, or food ingredients containing additives and flavorings that have been genetically modified or have been produced from GM organisms (see Chapter 29). The same techniques can ascertain the genotype of an animal strain (see Chapter 31). Another research area that benefits from the continuous development of mutation detection strategies is pharmacogenetics and pharmacogenomics (see Chapter 22), referred to as the effort to define the inter-individual variations that are expected to become integral for treatment planning, in terms of efficacy and adverse effects of drugs. This approach uses the technological expertise from high-throughput mutation detection techniques, genomics, and functional genomics to define and predict the nature of the response of an individual to a drug treatment, and to rationally design newer drugs or improve existing ones. Ultimately, the identified genomic sequence variation is organized and stored into specialized mutation databases, enabling a physician or researcher to query upon and retrieve information relevant to diagnostic issues (see Chapter 25).

Finally, and for the last 20 years, DNA analysis and testing has also significantly revolutionized the forensic sciences. The technical advances in molecular biology and the increasing knowledge of the human genome have had a major impact on forensic medicine (see Chapter 26). Genetic characterization of individuals at the DNA level enables identity testing from a minimal amount of biological specimen, such as hair, blood, semen, bone, and so forth, in cases of sexual assault, homicide, and unknown human remains, and paternity testing is also changing from the level of gene products to the genomic level. DNA testing is by far more advantageous over the conventional forensic serology, and over the years has contributed to the acquittal of falsely accused people (saving most of them even from death row) and the identification of the individual who had committed criminal acts (Cohen, 1995), and even helped to specify identities of unknown human remains, such as those from the victims at Ground Zero in New York, or from the skeletons of the Romanov family members (Gill *et al.*, 1994).

1.5 FUTURE PERSPECTIVES: WHAT LIES BEYOND

As an intrinsic part of DNA technology, molecular diagnostics are rooted in the April 1953 discovery of the DNA double helix. Today, it is clear that they embody a set of notable technological advances allowing for thousands of diagnostic reactions to be performed at once and for a range of mutations to be simultaneously detected. The reasons for this dramatic increase are two-fold. First of all, the elucidation of the human genomic sequence, as well as that of other species such as bacterial or viral pathogens, has led to an increased number of diagnostically relevant targets. Second, the molecular diagnostic testing volume is rapidly increasing. This is the consequence of a better understanding of the basis of inherited diseases, therefore allowing molecular diagnostics to play a key role in patient or disease management.

Presently, a great number of blood, hair, semen, and tissue samples are analyzed annually worldwide in both public and private laboratories, and the number of genetic tests available is steadily increased year by year. Taking these premises into account, we can presume that it is only a matter of time before molecular diagnostic laboratories become indispensable in laboratory medicine. In the post-genomic era, genetic information will have to be examined in multiple health care situations throughout people's lives. Currently, newborns can be screened for phenylketonurea and other treatable genetic diseases (Yang *et al.*, 2001). It is also possible that in the not-so-distant future, children at high risk from coronary artery disease will be identified and treated to prevent changes in their vascular walls during adulthood. Similarly, parents will have the option of being informed about their carrier status for many recessive diseases before they decide to start a family. Although not widely accepted, this initiative has already started to be implemented in Cyprus, where a couple at risk for thalassemia syndrome have been advised to undergo a genetic test for thalassemia mutations before their marriage (see also Chapter 37). Also, for middle-aged and older populations, scientists will be able to determine risk profiles for various late-onset diseases, preferably before the appearance of symptoms, which at least could be partly prevented through dietary or pharmaceutical interventions. In the near future, the monitoring of individual drug response profiles throughout life, using genetic testing for the identification of their individual DNA signature, will be part of the standard medical practice. Soon, genetic testing will comprise a wide spectrum of different analyses with a host of consequences for individuals and their families, which is worth emphasizing when explaining molecular diagnostics to the public (see also Chapter 38). All these issues are discussed in detail next. However, and in order to be more realistic, many of these expectations still are based on promises, though quite optimistic ones. Thus, some of the new perspectives of the field could be a decade away, and several challenges remain to be realized.

1.5.1 Commercializing Molecular Diagnostics

Currently, clinical molecular genetics is part of mainstream health care worldwide. Almost all clinical laboratories have a molecular diagnostic unit or department. Although in recent years the notion of molecular diagnostics has increasingly gained interest, genetic tests are still not generally used for population screening, but rather for diagnosis, carrier screening, and prenatal diagnosis, and only on a limited basis. Therefore, and in order to make molecular diagnostics widely available, several obstacles and issues need to be taken into consideration and resolved in the coming years.

The first important issue is the choice of the mutation detection platform. Despite the fact that there are over 50 different mutation detection and screening methods, there is no single platform or methodology that prevails for genetic testing. Genotyping can be done using different approaches, such as filters, gels, microarrays, microtiter plates; different amplification-based technologies; different separation techniques, such as blotting, capillary electrophoresis, microarrays, mass spectroscopy; and finally different means for labeling, such as radioactive, fluorescent, chemiluminescent, or enzymatic substances. The variety of detection approaches makes it not only difficult but also challenging to determine which one is better suited for a laboratory setting. Generally speaking, DNA sequencing is the gold standard for the identification of causative or non-DNA sequence variations, particularly with the advent of the next-generation sequencing technologies (see also Chapter 24). The initial investment costs and the expected test volume are some of the factors that need to be taken into consideration prior to choosing the detection technique. Related issues are also the costs of the hardware and software, testing reagents, and kits. The latter is of great importance, since the fact that most of the diagnostic laboratories today are running ''home-brew'' assays – for example, not using well-standardized genetic testing kits due to cost barriers, which brings to surface the issue of quality control of the reagents (see Chapter 40) and of safety (see Chapter 39). Currently, there are several clinical and technical recommendations for genetic testing for monogenic disorders that have been issued by several organizations (see Table 1.2).

Another very important issue is training the personnel of a molecular diagnostic laboratory, reflecting in the quality and the correct interpretation of the results. Continuous education of the personnel of the diagnostic laboratory is crucial for the accuracy of the results provided (see also Chapter 40). Many times, such as in the case of prenatal or pre-implantation

TABLE 1.2 Indicative clinical and technical recommendations for genetic testing for monogenic disorders. ACMG: American College of Medical Genetics, ASHG: American Society of Human Genetics.

Disease/syndrome	Gene	References
Alzheimer's disease	*ApoE*	ACMG (1995)
Canavan disease	*ASPA*	ACMG (1998)
Cystic fibrosis	*CFTR*	Dequeker *et al.* (2000), Grody *et al.* (2001a)
Thrombophilia	Factor V Leiden	Grody *et al.* (2001b)
Fragile X syndrome	*FMR1*	Maddalena *et al.* (2001)
Prader–Willi/Angelman syndrome	15q11-q13	ASHG/ACMG (1996)
Multiple endocrine neoplasia	*MEN1/MEN2*	Brandi *et al.* (2001)
Tuberous sclerosis	*TSC1/TSC2*	Roach *et al.* (1999)
Breast cancer	*BRCA1*	Sorscher and Levonian (1997)

diagnosis, irrevocable decisions need to be made, most of the time based on a simple test result. In the past decade, there has been a significant reduction in the number of incorrect genotypes diagnosed, as a result of continuous training and proficiency testing schemes (http://www.eurogentest.org). In the USA, there is a voluntary biannual proficiency testing for molecular diagnostic laboratories, while in Europe, the EuroGenTest European Network of Excellence (http://www.eurogentest.org) has been founded to promote quality in molecular genetic testing through the provision of external quality assessment (proficiency testing schemes) and the organization of best practice meetings and publication of guidelines. It is generally true that many geneticists and non-geneticist physicians would benefit from continuous education regarding the appropriate use of molecular diagnostic tests, which is necessary to evaluate the method pre-analytically and to interpret results.

The legal considerations and the ethical concerns are also hurdles that need to be overcome in the coming years. One issue is reimbursing of the diagnosis costs. At present, there are no insurance companies that reimburse the costs for molecular testing to the people insured; the necessary regulatory and legal framework remains to be established. ''Legalizing'' molecular testing, by the adoption of the relevant regulations, would probably result in an increase of the test volume and at the same time it can pose an immense barrier to uncontrolled genetic testing. Similarly, the need to obtain an informed consent from the patient to be analyzed is also of great importance and should be encouraged and facilitated by the diagnostic laboratory.

On the other hand, the issue of intellectual properties hampers the wide commercialization of molecular diagnostics. Almost all the clinically relevant genes have been now patented and the terms that the patent holders offer vary considerably (see Chapter 36). Among the difficulties that this issue imposes is the limiting choice of mutation detection platforms, the large loyalties for reagent use, and the exclusive sublicenses that many companies grant to clinical laboratories, leading eventually to monopolies. Since one of the biggest challenges that the clinical laboratory is facing is patent and regulatory compliance, partnerships and collaborations may be envisaged in order to take the technology licenses to the diagnostic laboratory that will subsequently develop, standardize, and distribute the assays. These will partly alleviate some of the intellectual properties issues. Finally, the issue of the medical genetics specialty is more urgent than ever. In the USA, medical genetics has been formally recognized as a medical specialty only within the past 15 years, and in Europe, medical genetics only recently has been formally recognized as a specialty (http://www.eshg.org). The implementation of this decision is still facing substantial difficulties (http://www.eshg.org/geneticseurope.htm), which will probably take years to bypass. With the completion of the Human Genome Project, genetics has become the driving force in medical research and is now poised for integration into medical practice. An increase in the medical genetics workforce, including geneticists and genetic counselors, will be necessary in the coming years. After all, the Human Genome Project has made information of inestimable diagnostic and therapeutic importance available and therefore the medical profession now has the obligation to rise to both the opportunities and challenges that this wealth of genetic information presents.

1.5.2 Personalized Medicine

The term ''personalized medicine'' refers to the practice of medicine where patients receive the most appropriate medical treatment, fitting dosage, and combination of drugs based on their genetic background. Some of the reasons for many types of adverse drug reactions are already known and often related to polymorphic gene alleles of drug metabolizing enzymes (Nebert and Menon, 2001; Risch *et al.*, 2002). The application of high-throughput genotyping tools for the identification and screening of single nucleotide polymorphisms (SNPs) eventually can lead to the determination of the unique molecular signature of an individual in a relatively short period of time. This way, individual drug responses can be predicted from predetermined genetic variances correlated with a drug effect. In other words, this will allow the physician to provide the patient with a selective drug prescription (see Chapter 22). A handful of pharmaceutical companies are developing a precise haplotyping scheme to identify individuals/patients who will derive optimal benefit from drugs currently under development. Clinicians will facilitate this effort by importing clinical data into this haplotyping system for a complete patient analysis and drug evaluation. In addition to these efforts, there is a growing need to incorporate this increasingly complex body of knowledge to standard medical practice. Incorporating pharmacogenomics-related courses in the standard curriculum of medical schools potentially can ensure that the forthcoming generation of clinicians and researchers will be familiar with the latest developments in that field and will be capable of providing patients with the expected benefits of personalized medicine.

Similarly, nutrigenomics (or nutritional genomics) investigates the interactions between nutrition and an individual's genome, and the consequent downstream effects on their phenotype with the aim of providing tailored nutritional advice or developing specialist food products (see Chapter 23). In other words, nutrigenomics recognizes that specific dietary advice that can be beneficial for one individual may be inappropriate, or actually harmful, to another. Although comparable to pharmacogenomics, nutrigenomics is still considered as an emerging science contrary to pharmacogenomics, which is considered to have ''come of age'' (Allison, 2008).

However, there are growing concerns regarding the ethical aspects of personalized medicine. First of all, equality in medical care needs to be ensured, when genetics foretell clinicians which patients would be less likely to benefit from a particular drug treatment. Second, it will become increasingly vital to devise operational tools for the prevention of stigmatization and discrimination of different populations, in particular on ethnic grounds (van Ommen, 2002), and therefore every precaution should be taken to eliminate all lingering prejudice and bias associated with the study of human genetic variation. Other dilemmas include the right to deny an available treatment from specific patient populations according to genetic-derived indications, as currently is the case with prenatal diagnosis (see also Chapter 37). Appropriate guidelines will be crucially needed for the successful implementation of pharmacogenomics into clinical practice.

1.5.3 Personal Genomics

The ultimate goal in health care over the next decades will be the efficient integration of molecular diagnostics with therapeutics. With the advent of next generation sequencing in 2005 (Margulies *et al.*, 2005) and the avalanche of developments in this field since then (see Chapter 24), experts believe that reasonably soon, people will be able to have their own genomes sequenced for under $1,000. This is going to involve sequencing technology that is much cheaper and faster than today's machines and several efforts are currently under way, often encouraged by major funding bodies (e.g. the European Commission-funded READNA consortium; http://www.cng.fr/READNA). When that point is reached, this can ultimately be translated in a patient being able to carry a smart card, like an ordinary credit card, providing secure access to his or her genetic information. So far, the entire DNA sequence of a handful of individuals has been sequenced, such as Craig Venter (Levy *et al.*, 2007), Jim Watson (Wheeler *et al.*, 2008), and so on, while ''the 1000 genomes'' project is an ambitious venture that involves sequencing the genomes of approximately 1,200 people from around the world, with the overall goal ''…to create the most detailed and medically useful picture to date of human genetic variation'' (http://www.1000genomes.org). In the future, a person may appear at the clinic for treatment, ''carrying'' its entire genome at hand or, alternatively, nanotechnology could eventually enable DNA analysis with a portable DNA sequencing device.

Even though the expectations are high and companies are currently using these new technologies to provide information to individuals to predict health and disease outcome, even behavioral traits, it is generally premature to make promises for clinically useful information from genomic analyses. Next to that, there is an inherent danger of overestimating the usefulness of the various personalized genomic tests that can be ordered directly by consumers. Unlike other genetic analyses, these tests provide sheer amounts of genetic information, but their diagnostic or prognostic value remains uncertain because of (a) the lack of information about the influence of environmental and other factors, and (b) the weak association for the vast majority of genetic loci with disease. In a carefully conducted study by Janssens and coworkers (2008), the scientific evidence supporting gene–disease associations for genes included in genomic tests offered commercially

to consumers online was assessed. These authors concluded that the synthetic odds ratios for the ''predictive genomic'' tests offered by seven of these companies were ranging from 0.54 to 0.88 for protective variants and from 1.04 to 3.2 for risk variants. Furthermore, genes in genomic profiles assessing the risk for developing cardiovascular disease were more frequently associated with non-cardiovascular diseases, and the same was true for associations of genomic profiles with bone diseases. These findings clearly demonstrate that, currently, there is insufficient scientific evidence to conclude that genomic profiles are useful in measuring genetic risk for common diseases or in developing personalized diet and lifestyle recommendations for disease prevention.

It is puzzling how these companies use their clients' genetic profiles to tailor individualized diets and lifestyle recommendations. Also, it is noteworthy that some of the companies that provide personal genomic tests have no physicians involved in ordering these tests, with the argument that ''...patients deserve direct access to their health information without a physician intermediary'' (http://www.nytimes.com/2008/06/26/business/26gene.html). Overall, although these tests could provide value to customers by offering tools for social networking or genealogy, there are questions whether and how to regulate these tests, about the extent to which they provide (useful) medical information and the risks from misinterpreting them (Magnus *et al.*, 2009).

1.6 CONCLUSIONS

In the coming years, molecular diagnostics will continue to be of critical importance to public health worldwide. Molecular genetic testing will facilitate the detection and characterization of disease, as well as monitoring of the drug response, and will assist in the identification of genetic modifiers and disease susceptibility. A wide range of molecular-based tests is available to assess DNA variation and changes in gene expression. However, there are major hurdles to overcome before the implementation of these tests in clinical laboratories, such as which test to employ, the choice of technology and equipment, and issues such as cost effectiveness, accuracy, reproducibility, personnel training, reimbursement by third-party payers, and intellectual property. At present, PCR-based testing predominates; however, alternative technologies aimed at exploring genome complexity without PCR are anticipated to gain momentum in the coming years. Furthermore, development of integrated chip devices (''lab-on-a-chip'') should facilitate genetic readouts from single cells and molecules. Together with proteomic-based testing, these advances will improve molecular diagnostics and will present additional challenges for implementing such technology in public or private research units, hospitals, clinics, and pharmaceutical industries.

REFERENCES

Allison, M. (2008). Is personalized medicine finally arriving? *Nat. Biotechnol.* 26, 509–517.

American College of Medical Genetics (1998). Position statement on carrier testing for canavan disease. www.faseb.org/genetics/acmg/pol-31.htm.

American College of Medical Genetics consensus statement on factor V Leiden mutation testing (1995). *Genet. Med*. 3, 139–148.

American College of Medical Genetics/American Society of Human Genetics Working Group on ApoE and Alzheimer disease (1995). Statement on use of apolipoprotein E testing for Alzheimer disease. *JAMA* 274, 1627–1629.

American College of Medical Genetics/American Society of Human Genetics (1996). Diagnostic testing for Prader–Willi and Angleman syndromes: report of the ASHG/ACMG test and technology transfer committee. *Am. J. Hum. Genet.* 58, 1085–1088.

Brandi, M.L., Gagel, R.F., Angeli, A., Bilezikian, J.P., Beck- Peccoz, P., Bordi, C., Conte-Devolx, B., Falchetti, A., Gheri, R.G., Libroia, A., Lips, C.J., Lombardi, G., Mannelli, M., Pacini, F., Ponder, B.A., Raue, F., Skogseid, B., Tamburrano, G., Thakker, R.V., Thompson, N.W., Tomassetti, P., Tonelli, F., Wells, S.A., Jr., and Marx, S.J. (2001). Guidelines for diagnosis and therapy of MEN type 1 and type 2. *J. Clin. Endocrinol. Metab.* 86, 5658–5671.

Busslinger, M., Moschonas, N., and Flavell, R.A. (1981). Beta + thalassemia: aberrant splicing results from a single point mutation in an intron. *Cell* 27, 289–298.

Chan, K., Wong, M.S., Chan, T.K., and Chan, V. (2004). A thalassaemia array for Southeast Asia. *Br. J. Haematol.* 124, 232–239.

Cohen, J. (1995). Genes and behavior make an appearance in the O. J. trial. *Science* 268, 22–23.

Cremonesi, L., Ferrari, M., Giordano, P., Harteveld, C.L., Kleanthous, M., Papasavva, T., Patrinos, G.P., and Traeger-Synodinos, J. (2007). An overview of current microarray-based human globin gene mutation detection methods. *Hemoglobin* 31, 289–311.

Dequeker, E., Cuppens, H., Dodge, J., Estivill, X., Goossens, M., Pignatti, P.F., Scheffer, H., Schwartz, M., Schwarz, M., Tummler, B., and Cassiman, J.J. (2000). Recommendations for quality improvement in genetic testing for cystic fibrosis. European concerted action on cystic fibrosis. *Eur. J. Hum. Genet.* 8(Suppl. 2), S2–S24.

Erdogan, F., Kirchner, R., Mann, W., Ropers, H.H., and Nuber, U.A. (2001). Detection of mitochondrial single nucleotide polymorphisms using a primer elongation reaction on oligonucleotide microarrays. *Nucleic Acids Res.* 29, E36.

Farrall, M., Rodeck, C.H., Stanier, P., Lissens, W., Watson, E., Law, H.Y., Warren, R., Super, M., Scambler, P., Wainwright, B., and Williamson, R. (1986). First-trimester prenatal diagnosis of cystic fibrosis with linked DNA probes. *Lancet* 327, 1402–1405.

Gemignani, F., Perra, C., Landi, S., Canzian, F., Kurg, A., Tonisson, N., Galanello, R., Cao, A., Metspalu, A., and Romeo, G. (2002). Reliable detection of beta-thalassemia and G6PD mutations by a DNA microarray. *Clin. Chem.* 48, 2051–2054.

Giardine, B., van Baal, S., Kaimakis, P., Riemer, C., Miller, W., Samara, M., Kollia, P., Anagnou, N.P., Chui, D.H., Wajcman, H., Hardison, R.C., and Patrinos, G.P. (2007). HbVar database of human hemoglobin variants and thalassemia mutations: 2007 update. *Hum. Mutat.* 28, 206.

Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., and Sullivan, K. (1994). Identification of the remains of the Romanov family by DNA analysis. *Nat. Genet.* 6, 130–135.

Grody, W.W., Cutting, G.R., Klinger, K.W., Richards, C.S., Watson, M. S., and Desnick, R.J. (2001a). Subcommittee on Cystic Fibrosis Screening, Accreditation of Genetic Services Committee, ACMG. American College of Medical Genetics. Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. *Genet. Med.* 3, 149–154.

Grody, W.W., Griffin, J.H., Taylor, A.K., Korf, B.R., and Heit, J.A.; ACMG Factor V. Leiden Working Group (2001b).

Hanash, S.M. (2000). Biomedical applications of two-dimensional electrophoresis using immobilized pH gradients: current status. *Electrophoresis* 21, 1202–1209.

Hanash, S.M., Madoz-Gurpide, J., and Misek, D.E. (2002). Identification of novel targets for cancer therapy using expression proteomics. *Leukemia* 16, 478–485.

Hardison, R.C., Chui, D.H., Giardine, B., Riemer, C., Patrinos, G.P., Anagnou, N., Miller, W., and Wajcman, H. (2002). HbVar: a relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. *Hum. Mutat.* 19, 225–233.

Holland, P.M., Abramson, R.D., Watson, R., and Gelfand, D.H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of Thermus aquaticus. *Proc. Natl. Acad. Sci. USA* 88, 7276–7280.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Janssens, A.C., Gwinn, M., Bradley, L.A., Oostra, B.A., van Duijn, C.M., and Khoury, M.J. (2008). A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. *Am. J. Hum. Genet.* 82, 593–599.

Kan, Y.W., Golbus, M.S., and Dozy, A.M. (1976). Prenatal diagnosis of alpha-thalassemia. Clinical application of molecular hybridization. *N. Engl. J. Med.* 295, 1165–1167.

Kan, Y.W., and Dozy, A.M. (1978). Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proc. Natl. Acad. Sci. USA* 75, 5631–5635.

Knezevic, V., Leethanakul, C., Bichsel, V.E., Worth, J.M., Prabhu, V.V., Gutkind, J.S., Liotta, L.A., Munson, P.J., Petricoin, E.F., 3rd, and Krizman, D.B. (2001). Proteomic profiling of the cancer microenvironment by antibody arrays. *Proteomics* 1, 1271–1278.

Kozal, M.J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T.C., Richman, D.D., Morris, D., Hubbell, E., Chee, M., and Gingeras, T.R. (1996). Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nat. Med.* 2, 753–759.

Landegren, U., Kaiser, R., Sanders, J., and Hood, L. (1988). A ligase-mediated gene detection technique. *Science* 241, 1077–1080.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., Lin, Y., MacDonald, J.R., Pang, A.W., Shago, M., Stockwell, T.B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S.A., Busam, D.A., Beeson, K.Y., McIntosh, T.C., Remington, K.A., Abril, J.F., Gill, J., Borman, J., Rogers, Y.H., Frazier, M.E., Scherer, S.W., Strausberg, R.L., and Venter, J.C. (2007). The diploid genome sequence of an individual human. *PLoS Biol.* 5, e254.

Maddalena, A., Richards, C.S., McGinniss, M.J., Brothman, A., Desnick, R.J., Grier, R.E., Hirsch, B., Jacky, P., McDowell, G.A., Popovich, B., Watson, M., and Wolff, D.J. (2001). Technical standards and guidelines for fragile X: the first of a series of disease-specific supplements to the standards and guidelines for clinical genetics laboratories of the American College of Medical Genetics. Quality Assurance Subcommittee of the laboratory practice committee. *Genet. Med.* 3, 200–205.

Magnus, D., Cho, M.K., and Cook-Deegan, R. (2009). Direct-to-consumer genetic tests: beyond medical regulation? *Genome Med.* 1, 17.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., and Rothberg, J.M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.

Mashal, R.D., Koontz, J., and Sklar, J. (1995). Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat. Genet.* 9, 177–183.

Mullis, K.B., and Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 155, 335–350.

Myers, R.M., Larin, Z., and Maniatis, T. (1985a). Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* 230, 1242–1246.

Myers, R.M., Lumelsky, N., Lerman, L.S., and Maniatis, T. (1985b). Detection of single base substitutions in total genomic DNA. *Nature* 313, 495–498.

Myers, R.M., Maniatis, T., and Lerman, L.S. (1987). Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol.* 155, 501–527.

Nebert, D.W., and Menon, A.G. (2001). Pharmacogenomics, ethnicity, and susceptibility genes. *Pharmacogenomics J.* 1, 19–22.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* 86, 2766–2770.

Orkin, S.H., Kazazian, H.H., Jr., Antonarakis, S.E., Goff, S.C., Boehm, C.D., Sexton, J.P., Waber, P.G., and Giardina, P.J. (1982). Linkage of beta-thalassaemia mutations and beta-globin gene polymorphisms with DNA polymorphisms in human betaglobin gene cluster. *Nature* 296, 627–631.

Orkin, S.H., Markham, A.F., and Kazazian, H.H., Jr. (1983). Direct detection of the common Mediterranean beta-thalassemia gene with synthetic DNA probes. An alternative approach for prenatal diagnosis. *J. Clin. Invest* 71, 775–779.

Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L., and Syvanen, A.C. (2000). A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res.* 10, 1031–1042.

Patrinos, G.P., Giardine, B., Riemer, C., Miller, W., Chui, D.H., Anagnou, N.P., Wajcman, H., and Hardison, R.C. (2004). Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. *Nucleic Acids Res.* 32, D537–D541.

Pauling, L., Itano, H.A., Singer, S.J., and Wells, I.C. (1949). Sickle cell anemia, a molecular disease. *Science* 110, 543–548.

Petricoin, E.F., Zoon, K.C., Kohn, E.C., Barrett, J.C., and Liotta, L.A. (2002). Clinical proteomics: translating benchside promise into bedside reality. *Nat. Rev. Drug. Discov.* 1, 683–695.

Pirastu, M., Kan, Y.W., Cao, A., Conner, B.J., Teplitz, R.L., and Wallace, R.B. (1983). Prenatal diagnosis of beta-thalassemia. Detection of a single nucleotide mutation in DNA. *N. Engl. J. Med.* 309, 284–287.

Risch, N., Burchard, E., Ziv, E., and Tang, H. (2002). Categorization of humans in biomedical research: genes, race and disease. *Genome Biol.* 3, 1–12.

Roach, E.S., DiMario, F.J., Kandt, R.S., and Northrup, H. (1999). Tuberous Sclerosis Consensus Conference: recommendations for diagnostic evaluation. National Tuberous Sclerosis Association. *J. Child Neurol.* 14, 401–407.

Robinson, W.H., DiGennaro, C., Hueber, W., Haab, B.B., Kamachi, M., Dean, E.J., Fournel, S., Fong, D., Genovese, M.C., de Vegvar, H.E., Skriner, K., Hirschberg, D.L., Morris, R.I., Muller, S., Pruijn, G.J., van Venrooij, W.J., Smolen, J.S., Brown, P.O., Steinman, L., and Utz, P.J. (2002). Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat. Med.* 8, 295–301.

Ronaghi, M., Uhlen, M., and Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* 281, 363–365.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230, 1350–1354.

Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 324, 163–166.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988). Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487–491.

Saiki, R.K., Walsh, P.S., Levenson, C.H., and Erlich, H.A. (1989). Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86, 6230–6234.

Saleeba, J.A., Ramus, S.J., and Cotton, R.G. (1992). Complete mutation detection using unlabeled chemical cleavage. *Hum. Mutat.* 1, 63–69.

Sorscher, S., and Levonian, P. (1997). BCRA 1 testing guidelines for high-risk patients. *J. Clin. Oncol.* 15, 1711.

Treisman, R., Orkin, S.H., and Maniatis, T. (1983). Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. *Nature* 302, 591–596.

van Ommen, G.J. (2002). The Human Genome Project and the future of diagnostics, treatment and prevention. *J. Inherit. Metab. Dis.* 25, 183–188.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). The sequence of the human genome. *Science* 291, 1304–1351.

Wen, W.H., Bernstein, L., Lescallett, J., Beazer-Barclay, Y., Sullivan-Halley, J., White, M., and Press, M.F. (2000). Comparison of TP53 mutations identified by oligonucleotide microarray and conventional DNA sequence analysis. *Cancer Res.* 60, 2716–2722.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C.L., Irzyk, G.P., Lupski, J.R., Chinault, C., Song, X.Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D.M., Margulies, M., Weinstock, G.M., Gibbs, R.A., and Rothberg, J.M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872–876.

Woo, S.L., Lidsky, A.S., Guttler, F., Chandra, T., and Robson, K.J. (1983). Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. *Nature* 306, 151–155.

Xiao, W., and Oefner, P.J. (2001). Denaturing high-performance liquid chromatography: a review. *Hum. Mutat.* 17, 439–474.

Yang, Y., Drummond-Borg, M., and Garcia-Heras, J. (2001). Molecular analysis of phenylketonuria (PKU) in newborns from Texas. *Hum. Mutat.* 17, 523.

This page intentionally left blank

Section I

Molecular Diagnostic Technology

This page intentionally left blank

Chapter 2

Allele-Specific Mutation Detection

Aglaia Athanassiadou[1], Eleana F. Stavrou[1], Adamandia Papachatzopoulou[1] and George P. Patrinos[2,3]

[1]*Department of General Biology, Faculty of Medicine, School of Health Sciences, University of Patras, Patras, Greece;*

[2]*Department of Pharmacy, School of Health Sciences, University of Patras, Patras, Greece;*

[3]*Erasmus University Medical Center, Faculty of Medicine and Health Sciences, Department of Bioinformatics, The Netherlands*

2.1 INTRODUCTION

A primary function of molecular diagnostics is the detection of mutations and single nucleotide polymorphisms (SNPs) that are associated with particular phenotypes. This chapter provides a description of three relatively simple polymerase chain reaction (PCR)-based techniques that can be applied to detect known nucleotide variations in DNA.

The first approach, known as the amplification refractory mutation system, or PCR-ARMS, is based on the principle that a mismatch between the 3′ nucleotide of a PCR primer and the template reduces or prevents primer extension by *Taq* polymerase. A variety of strategies have been developed using primers that are complementary to, and allow for the specific amplification of, individual alleles.

The second approach described in this chapter is based on hybridization of PCR products to allele-specific oligonucleotide probes, or PCR-ASO. This method can be applied in two formats: the forward ASO approach where PCR products are immobilized on membrane and hybridized to labeled ASO probes; and the reverse ASO approach, where ASO probes are immobilized on the membrane and hybridized to labeled PCR products.

Finally, the third approach, competitive oligopriming, is a strategy based on allele-specific amplification, by the use of ASO. However, there is a fundamental difference between the usual methods involving allele-specific amplification and competitive oligonucleotide priming. In the former, mismatching between primer and template DNA prevents extension of DNA synthesis, whereas in the competitive oligopriming system mismatching prevents primer annealing.

Allele-specific mutation detection has been widely used in research and molecular diagnostics since their initial development in the late 1980s. The attraction of these methods lies in their simplicity and applicability to the analysis of virtually any known point mutation or SNP. Moreover, these methods do not require expensive and sophisticated instrumentation.

2.2 PCR-ARMS

In 1989, several independent groups described a PCR-based approach for analyzing known point mutations in DNA and distinguishing between normal, heterozygous, and homozygous mutant genotypes (Newton *et al.*, 1989; Nichols *et al.*, 1989; Okayama *et al.*, 1989; Sommer *et al.*, 1989; Wu *et al.*, 1989). The method is most commonly referred to as PCR-ARMS or ARMS (amplification refractory mutation system) (Newton *et al.*, 1989). It has also been referred to as PASA (PCR amplification of specific alleles) by Sommer and coworkers (1989) and as ASPCR (allele-specific PCR) by Wu and coworkers (1989). The initial reports demonstrated the applicability of this approach to analyze known disease mutations associated with α1-antitrypsin deficiency, amyloidotic polyneuropathy, phenylketonuria, and sickle cell disease. Subsequent validation studies quickly established that this strategy is applicable to the analysis of any known point mutation or SNP (Sommer *et al.*, 1992).

2.2.1 Basic Principles

PCR-ARMS is based on the observation that DNA amplification is inefficient or completely refractory if there is a mismatch between the 3′ terminal nucleotide of a PCR primer and the corresponding template (Newton *et al.*, 1989). *Taq* DNA polymerase lacks a 3′ to 5′ exonuclease

activity, and therefore cannot correct mismatches at the 3′ terminus of the primer. As such, complementary base-pairing at the 3′ end of the primer is required for efficient amplification by *Taq* DNA polymerase and is a strong determining factor of template specificity. Amplification of the normal allele, and not that of the mutant, is accomplished using a primer that is complementary to the normal allele and has a mismatch between the 3′ residue and the mutant allele. Conversely, only the mutant will be amplified if the 3′ residue of the primer is complementary to the mutant allele and not the normal allele. The specificity or discriminating power of the 3′ terminal nucleotide can be enhanced further by incorporating an additional mismatch positioned near the 3′ nucleotide (Newton *et al.*, 1989). The basic concept is illustrated in Fig. 2.1.

Various studies have attempted to quantify the inhibitory effect of different 3′ mismatches on PCR amplification (Kwok *et al.*, 1990; Sarkar *et al.*, 1990; Huang *et al.*, 1992; Ayyadevara *et al.*, 2000). Although some trends have emerged, the results were remarkably discordant. Sarkar and coworkers (1990) concluded that PCR is inhibited by mismatches between the template and the 3′ or 3′ penultimate nucleotide of the primer. Under relatively relaxed stringency conditions for primer-template annealing, Kwok and coworkers (1990) demonstrated a 20-fold reduction in amplification efficiency with A:A (primer-template) mismatches, 100-fold reductions with A:G, G:A, and C:C mismatches, and little or no reduction with any other mismatches. Under higher stringency conditions, Huang and coworkers (1992) showed some degree of inhibition with every combination of 3′ mismatch. The weakest inhibition, about 100-fold reduction, was associated with C:T mismatches. There was approximately 103-fold reduction with A:C, C:A, G:T, and T:G mismatches; 103- to 104-fold reduction with T:C and T:T mismatches; and at least 106-fold reduction with A:A, G:A, A:G, G:G, and C:C mismatches. Ayyadevara and coworkers (2000) systematically varied both the 3′ terminal and penultimate nucleotides of primers, all under relatively high stringency conditions. Their study indicated that primers ending with 3′ A are moderately inferior to those ending in other nucleotides. Allele-specific amplification had 40- to 100-fold reduction when the mismatched primer had T, G, or C at the 3′ terminal position. They also concluded that the penultimate 3′ nucleotide plays a minor role in mismatch discrimination, and that amplification efficiency is reduced when A (and to a lesser extent T) occupies the penultimate 3′ position.

The design and optimization of PCR-ARMS protocols is primarily a function of the target sequence and the nucleotide differences that define the alleles. In addition to mismatches between the 3′ terminal base of the primer and the target, single mismatches should be incorporated at several positions from the 3′ terminus. Apart from the theoretical considerations relating to the 3′ terminal position of the allele-specific primers, the design and optimization of PCR-ARMS primers follows the same considerations used for any other type of PCR. Primers are chosen to have comparable theoretical melting temperatures (T_m). Primer lengths are generally 20 nucleotides or longer, although the length is less important than the T_m. Primers should not have self-complementary sequences of 4 nucleotides or more, nor should they have more than 4 nucleotide complementarity between their 3′ ends.

As with any PCR-based strategy, false negative results could result from the presence of sequence variations that negatively impact on the primer annealing and/or amplification. This potential problem can be overcome by targeting the opposite strand for amplification, or by incorporating a degenerate nucleotide into the primer. For single mutation ARMS, the PCR conditions can be established by titrating the $MgCl_2$ concentration and/or primer concentrations, at constant annealing temperature. For multiplex ARMS, the first step is to optimize the PCR conditions for sensitive and specific detection of each allele. The objective is to define the PCR cycling parameters and $MgCl_2$ concentration under which all of the alleles will be amplified in an efficient and specific manner. It may be necessary to redesign one or more primer pair to achieve allele-specific amplification under one set of PCR conditions. Once the PCR parameters have been established, the primer pairs can be combined to evaluate the performance of the multiplex ARMS assay. Primer concentrations should be adjusted such that each of the alleles is amplified to a comparable degree. The specificity of the ARMS assay should be evaluated using samples from normal controls and known carriers of the mutations. Specificity can also be tested using serial dilutions of mutant DNA mixed with normal DNA (e.g. mutant:normal = 1:1, 1:2, 1:4, 1:8, etc.). Uniform and specific amplification of each allele may require further manipulation of the cycling parameters or the concentration of one or more reagents (primers, *Taq* polymerase, $MgCl_2$).

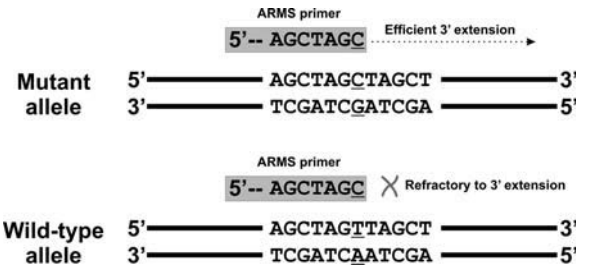


FIGURE 2.1 Schematic of PCR-ARMS. Schematic representation of a PCR-ARMS assay for the detection of a single-base mutation (underlined). The 3′ terminal nucleotide of the ARMS primer is complementary to the mutant allele. The ARMS primer has an additional mismatch positioned three bases from the 3′ terminal nucleotide (not shown).

2.2.2 Single and Multiplex PCR-ARMS

A common application of PCR-ARMS is the detection of individual point mutations in DNA. Primers are designed that will preferentially amplify the mutant allele, while being refractory to amplification of the normal allele. Included in the reaction mix is a second set of primers that are specific for a heterologous locus that serves as a positive control for PCR amplification. Conventional agarose or polyacrylamide gel electrophoresis systems are used to resolve the control amplicon from the mutant amplicon. Since the efficiency of amplification is inversely proportional to the length of the amplicon, the control amplicon should be larger or close in size to the mutant amplicon. Figure 2.2 illustrates the use of this approach to detect the most common *DHCR7* gene mutation associated with the Smith–Lemli–Opitz syndrome (Nowaczyk *et al.*, 2001). This assay distinguishes between samples that are positive for the IVS 8-1 G > C mutation (heterozygous or homozygous) and those that are negative for the mutation. This is sufficient for applications where it is not necessary to distinguish between heterozygotes and homozygotes (e.g. mutation analysis of unaffected carriers of a recessive disorder). For other applications, such as mutation analysis for individuals affected with a recessive disorder, it may be desirable to genotype individuals who are positive for the mutation. This can be accomplished by screening all positive samples with a second ARMS assay that is specific for the normal allele and refractory for the mutant allele. Heterozygotes will be positive for both ARMS assays, whereas homozygotes will be positive for only the mutant ARMS assay. The specificity of PCR-ARMS is such that pools of samples can be screened to identify rare carriers of specific mutations. This approach is capable of detecting a single positive sample in pools of 30 or more samples. This eliminates the need to test large numbers of samples individually to establish population frequencies for individual mutant alleles (Nowaczyk *et al.*, 2001; Waye *et al.*, 2002).

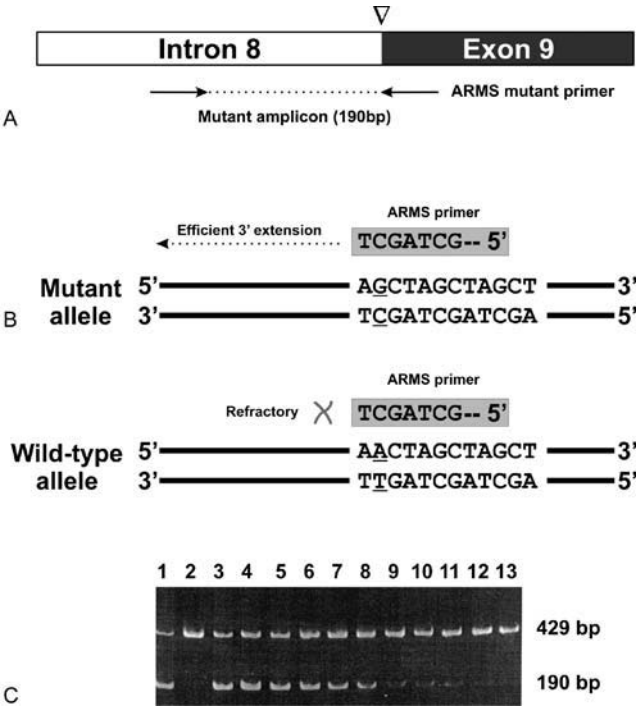


FIGURE 2.2 PCR-ARMS detection of a single *DHCR7* mutation. **A.** Schematic representation of the boundary between intervening sequence 8 and exon 8 of the *DHCR7* gene showing the position of a common mutation causing Smith–Lemli–Opitz syndrome. The mutation alters the canonical splice acceptor sequence (AG > AC) and can be detected using an ARMS primer specific for the mutant allele. **B.** PCR-ARMS detection of the IVS 8-1 GÆC mutation (underlined). The 3′ terminal nucleotide of the ARMS primer is complementary to the mutant allele, and an additional mismatch is incorporated three bases from the 3′ nucleotide (not shown). **C.** Analysis of PCR-ARMS products by non-denaturing polyacrylamide gel electrophoresis, visualized by ethidium bromide staining and UV fluorescence. The 190 bp fragment is specific for the IVS 8-1 G > C mutant allele. The 429 bp fragment corresponds to a region of the *HFE* gene that serves as a positive internal control for PCR amplification. Lane 1: IVS 8-1 G > C heterozygote; lane 2: normal control; lane 3: 1:1 mixture of carrier:normal DNA; lanes 4–13: 1:2n mixture (N = 1–10) of carrier:normal DNA, respectively (1:2 to 1:1024).

Multiplex PCR-ARMS assays can easily be developed for single-tube detection of four to six different mutations (Old *et al.*, 1990). Multiplexing PCR-ARMS assays is often necessary, since many genetic disorders are characterized by a small number of common mutations that account for a significant proportion, and in some instances, the majority of mutant alleles represented in a given population (see also Chapter 25). Once the spectrum of mutations and the frequencies of individual alleles have been established, PCR-ARMS can be used to simultaneously screen for the most common mutations. Figure 2.3 shows two multiplex PCR-ARMS panels that detect 11 common *DHCR7* gene mutations. Collectively, these mutations account for more than 85% of the mutant alleles detected in North American Smith-Lemli-Opitz syndrome patients.

2.2.3 Genotyping with PCR-ARMS

PCR-ARMS can be used to determine genotypes for individual mutations or SNPs, using two separate ARMS assays: one specific for the mutant allele and the other specific for the normal allele. Alternatively, PCR-ARMS systems have been developed for single-tube genotyping of mutations or SNPs. The simplest system involves bidirectional amplification, with the normal allele amplified using one strand as the template and the mutant allele amplified off the complementary strand (Waterfall and Cobb, 2001; Ye *et al.*, 2001). The primers are designed such that the lengths of the amplicons can easily be resolved by conventional gel electrophoresis. This strategy has a built-in

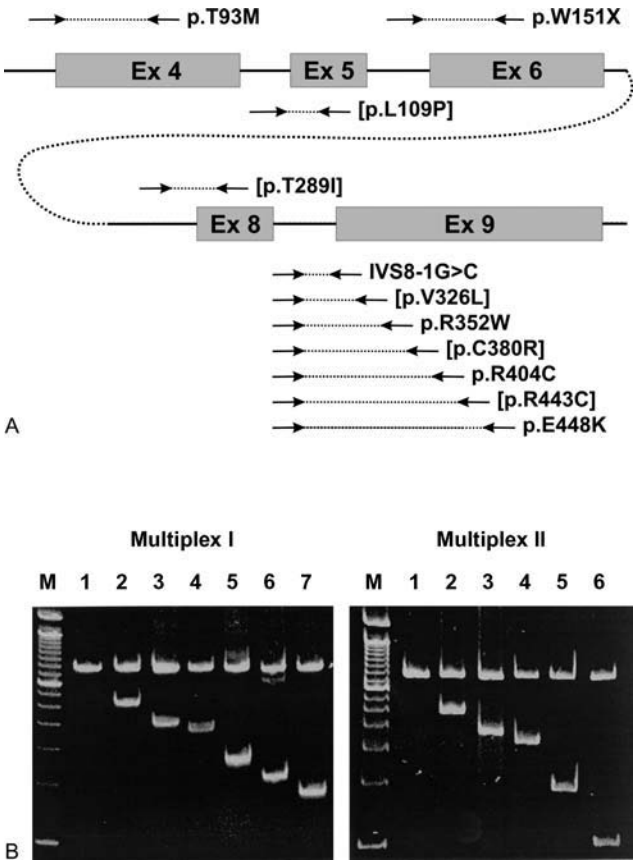


FIGURE 2.3 Multiplex PCR-ARMS detection of 11 *DHCR7* gene mutations. **A.** Schematic representation of the *DHCR7* gene (introns not drawn to scale) showing the seven coding exons and PCR-ARMS strategies for detecting 11 point mutations associated with Smith–Lemli–Opitz syndrome. The mutations are detected in two multiplex PCR-ARMS assays (multiplex II mutations indicated in brackets). **B.** Analysis of PCR-ARMS multiplex products. Each multiplex includes an internal control amplicon from the *HFE* gene. Multiplex I: lane 1: normal control; lane 2: p.E448K carrier; lane 3: p.R404C; lane 4: p.W151X; lane 5: p.R352W; lane 6: p.T93M; lane 7: IVS 8-1 G > C. Multiplex II: lane 1: normal control; lane 2: p.R443C carrier; lane 3: p.T289I; lane 4: p.C380R; lane 5: p.V326L; lane 6: p.L109P.

positive control resulting from amplification between the outermost primers of the normal and mutant amplicons. Figure 2.4 shows the application of PCR-ARMS to genotype samples for the most common mutation associated with hereditary hemochromatosis (*HFE* p.C282Y).

PCR-ARMS can be also used to establish haplotypes of individuals in the absence of samples from relatives. This is particularly useful for haplotyping SNPs that are located within distances that are amenable to PCR amplification. Consider the case of adjacent bi-allelic SNPs, where the alleles are designated Aa and Bb. PCR-ARMS using the four possible combinations of ARMS primers specific for the SNP A and SNP B alleles (AB, Ab, aB, ab) can be used to establish the haplotype (Eitan and Kashi, 2002).

2.2.4 Advantages and Limitations

PCR-ARMS assays are ideally suited for many molecular diagnostic applications, particularly those requiring detection of relatively small numbers of point mutations and having low-to-moderate throughputs. The primary advantage of PCR-ARMS is the ease with which multiplex assays can be developed, validated, and implemented. Moreover, the tests are non-radioactive and do not require expensive and sophisticated detection systems.

The most significant limitation of PCR-ARMS is that it can be used to detect only known mutations and polymorphisms. As such, it is usually necessary to combine PCR-ARMS with other molecular diagnostic strategies (e.g. sequencing) to provide comprehensive mutation detection. In its simplest formats, such as those described in this chapter, PCR-ARMS may be impractical for applications involving large numbers of mutations or high throughput. For such applications, consider using ARMS assays with allele detection strategies that are more amenable to automation. One such approach is to use ARMS primers labeled with fluorescent dyes.

2.3 PCR-ASO

The analysis of point mutations in DNA using hybridization with ASO probes is based on the principle that even single nucleotide mismatches between a probe and its target can destabilize the hybrid. ASO probes can be designed to be complementary and specific for the various alleles, thus providing a simple methodology to detect any known mutation or SNP.

The use of ASO probes actually predates PCR, and was a commonly used approach to analyze cloned DNA. Radioactively labeled ASO probes have even been used to diagnose genetic disease using non-amplified genomic DNA that has been immobilized on a membrane after restriction endonuclease digestion and electrophoretic separation (Conner *et al.*, 1983; Orkin *et al.*, 1983; Pirastu *et al.*, 1983). With the advent of PCR amplification (Saiki *et al.*, 1985, 1988a, b), PCR-ASO became one the first approaches used to analyze known point mutations within amplified DNA fragments (Saiki *et al.*, 1986).

2.3.1 Basic Principles

The design of ASO probes is largely dependent on the sequence of the region being targeted for analysis. ASO probes are generally short oligonucleotides (15- to 17-mers) with 30–50% G + C content, designed with the discriminating nucleotide located near the middle of the probe. Longer probes can be used to compensate for regions that have low G + C content. G:T and G:A mismatches are slightly destabilizing, whereas the effect is significantly

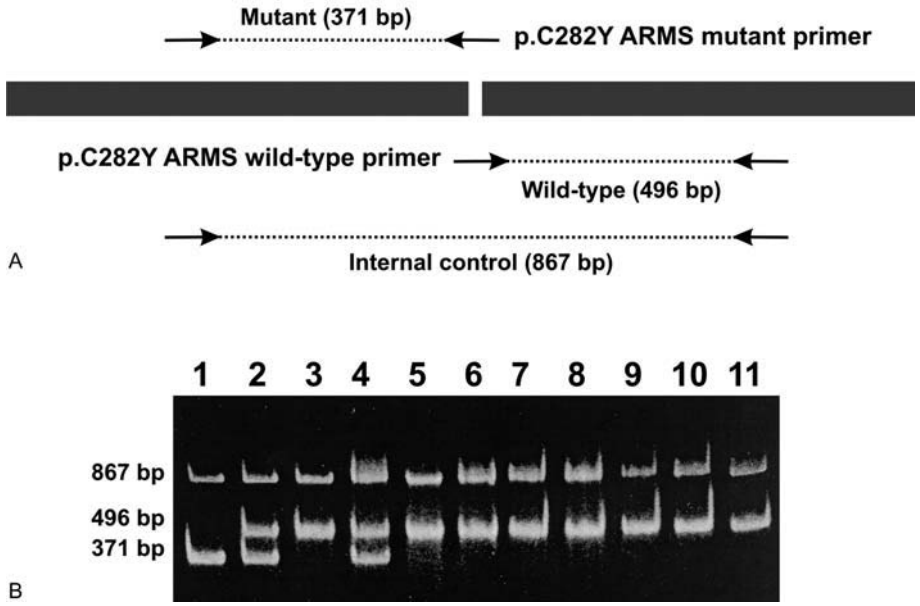


FIGURE 2.4 Genotyping of *HFE* p.C282Y mutation. **A.** Schematic representation of a portion of the *HFE* gene showing the most common mutation associated with hereditary hemochromatosis (p.C282Y). The multiplex analysis contains primers for two different PCR-ARMS assays, run off opposite strands. One PCR-ARMS assay is specific for the mutant allele, whereas the other assay is specific for the normal allele. The PCR products for the mutant and normal alleles have different lengths and can easily be distinguished by gel electrophoresis. **B.** Analysis of the PCRA-RMS products. Lane 1, p.C282Y homozygote; lane 2, p.C282Y heterozygote; lane 3, normal control; lanes 4–11, samples.

greater for A:A, T:T, C:T, and C:A mismatches (Ikuta *et al.*, 1987). Therefore, the choice of the sense or antisense strand may affect specificity (e.g. a C:A mismatch is often easier to discriminate than a G:T mismatch).

The choice of ASO primer sequences can involve a considerable amount of trial-and-error testing of candidate probes. Candidate probes should be evaluated individually using known positive and negative control samples, all under fixed hybridization and wash stringency conditions. For some mutations, it may be necessary to synthesize several versions of a given probe to attain allele specificity under the assay conditions. Care should be taken to avoid sequences that are associated with polymorphisms. False negative results could occur if a polymorphism impairs annealing and/or extension of a primer used for amplification, or if a polymorphism lies near the mutation being tested and destabilizes the ASO/target hybrid.

2.3.2 Forward and Reverse ASO Formats

The forward ASO format involves immobilizing PCR-amplified DNA fragments onto a nylon membrane, and hybridizing the membrane to a labeled oligonucleotide probe that is complementary and specific for a given sequence. The membrane is then washed at the appropriate stringency to dissociate any probe molecules that are not perfectly matched to the target. The first-generation PCR-ASO protocols utilized oligonucleotide probes that were phosphorylated at their 5′ termini with [γ-^{32}P], and exposure to X-ray film to detect the membrane bound probe-target hybrids (Saiki *et al.*, 1986). Subsequent protocols employed ASO probes that have biotin conjugated to their 5′ termini (Saiki *et al.*, 1988b). Following stringency washes, the probe-target hybrid is detected using streptavidin conjugated with horseradish peroxidase (HRP). The HRP activity can then be detected using a colorimetric detection with tetramethylbenzidine and hydrogen peroxide. Alternatively, the HRP activity can be detected using chemiluminescent substrates.

The forward ASO format is most useful when large numbers of samples are being screened for a small number of mutant alleles. Figure 2.5a shows examples of forward dot-blots for the two most common *ARSA* gene mutations in patients with metachromatic leukodystrophy (Polten *et al.*, 1991). However, the forward ASO format requires separate labeled probes and hybridization cycles for each allele being tested, making this technique overly cumbersome for applications involving multiple mutations.

Reverse ASO comes as a solution to this problem. In particular, reverse ASO entails immobilization of an array of ASO probes to a membrane strip and hybridization of the strip to labeled PCR-amplified DNA. The original reverse ASO format, or reverse dot-blot, employed probes that had poly(dT) tails added to their 3′ termini and were immobilized on the nylon membranes by UV cross-linking

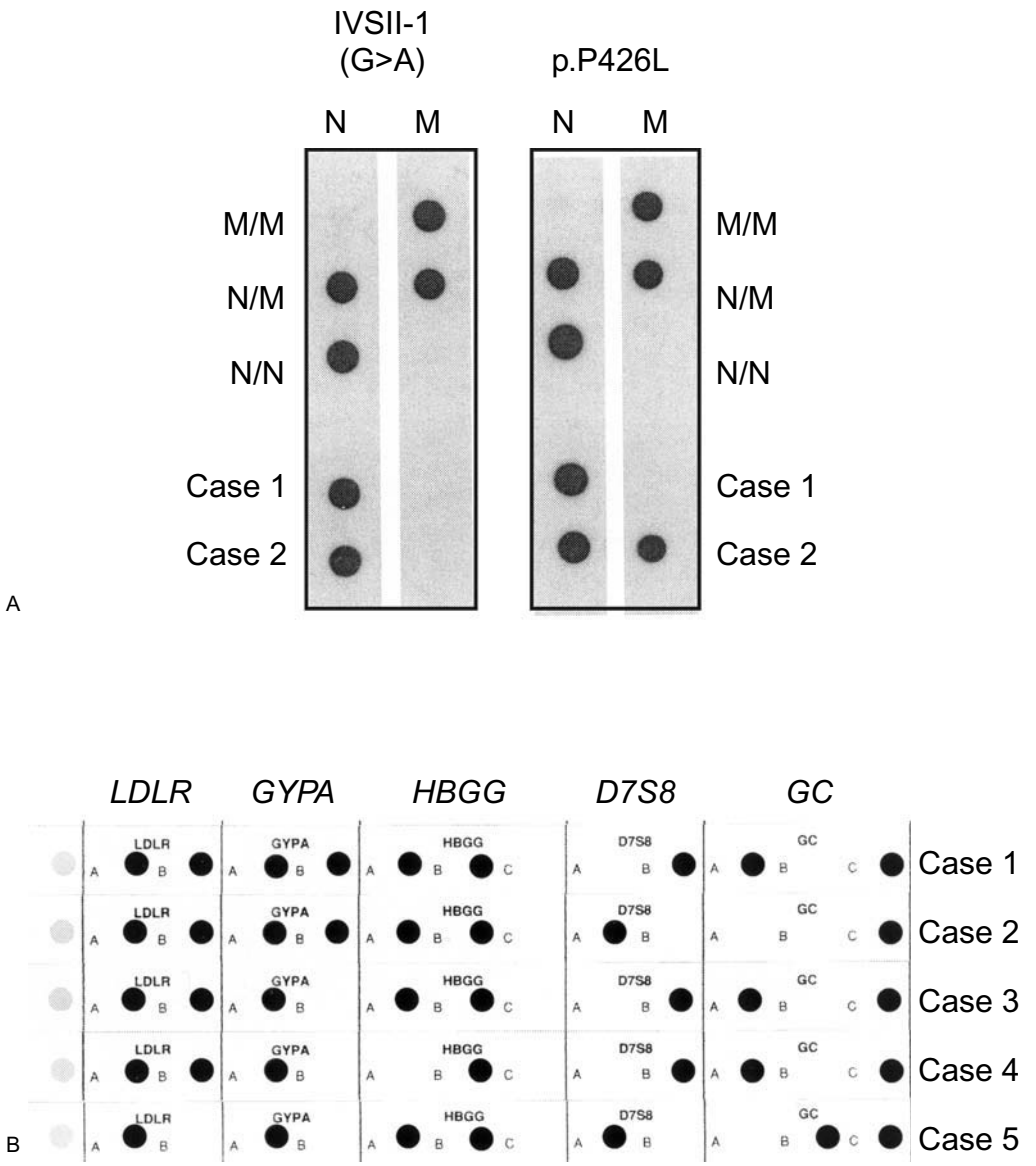


FIGURE 2.5 **A.** Forward PCR-ASO of ARSA gene mutations. PCR-ASO analysis of the two most common ARSA mutations (IVS2 + 1 G > A and p.P426L) causing metachromatic leukodystrophy. Corresponding regions of the ARSA were amplified by PCR and immobilized to nylon membranes. The membranes were hybridized to non-radioactively labeled ASO probes that are complementary to the normal or mutant alleles. Probe/target hybrids were detected by chemiluminescence. Each test strip contains amplified DNA from known controls (homozygous mutant (M/M), heterozygous mutant (N/M), normal (N/N)) and two samples. **B.** Reverse PCR-ASO of SNPs. PCR-ASO analysis of SNPs at five independent loci using the AmpliType PolyMarker™ reverse dot-blot system (Perkin–Elmer). The corresponding regions of the genome are amplified as a multiplex PCR using biotinylated primer sets. The PCR products are then hybridized to membrane strips containing ASO probes that are complementary to the various alleles. Biotinylated PCR products that hybridize to the membrane-bound ASO probes are detected using a non-radioactive colorimetric reaction. SNP pro-files are shown for five unrelated individuals. A, B, C on the strips correspond to the different alleles at each locus.

(Saiki *et al.*, 1989). The method subsequently was improved by covalent binding of the ASO probes to membranes via 5′ amino linkers (Zhang *et al.*, 1991; Chehab and Wall, 1992).

Over the past decade, the reverse ASO format has become a widely used tool for routine screening of genes that have numerous mutant alleles (reviewed in Gold, 2003; see also Chapter 34). Reverse PCR-ASO test kits have evolved for mutation screening for several genetic diseases, including α-thalassemia (Chan *et al.*, 1999; Foglietta *et al.*, 2003), β-thalassemia (Chehab, 1993; Cai *et al.*, 1994), and cystic fibrosis (Chehab and Wall, 1992; Makowski *et al.*, 2003). Also, several commercial kits have been developed for α- and β-thalassemia mutation detection (reviewed in Patrinos *et al.*, 2005), while test kits also have been developed for genotyping SNPs used for forensic identity testing (Budowle *et al.*, 1995; see also Chapter 26). Figure 2.5a shows commercial ASO reverse dot-blots for SNPs associated with five independent loci.

2.3.3 Advantages and Limitations

The PCR-ASO method, and particularly the reverse ASO format, provides a simple approach for simultaneous genotyping of large numbers of mutations and polymorphisms. The method can be applied to any known sequence variation, is non-radioactive, and does not require specialized instrumentation to detect the alleles. A potential drawback to the PCR-ASO strategy is the amount of developmental work needed to identify a panel of oligonucleotide probes that are allele specific under the same hybridization and wash stringency conditions. For small laboratories with limited resources, this initial investment may preclude the development of in-house PCR-ASO assays. For the same reasons, it is unlikely that commercial PCR-ASO will become available for rare diseases having limited market potential.

2.4 THE COMPETITIVE OLIGOPRIMING ASSAY

Competitive oligonucleotide priming (COP) of DNA synthesis has been described for the first time by Gibbs and coworkers in 1989. As previously mentioned, COP is a strategy for the detection of known sequence variations, based on allele-specific amplification, using ASO. However, there are two fundamental differences between the usual methods involving allele-specific amplification (Nollau and Wagener, 1997) and competitive oligonucleotide priming: in the former, mismatching between primer and template DNA prevents extension of DNA synthesis, whereas in the competitive oligopriming system mismatching prevents primer annealing (Table 2.1).

2.4.1 Basic Principles

Differential primer annealing in competitive oligopriming is achieved by the use of three primers of DNA synthesis instead of two, in one polymerase chain reaction. A pair of competitive oligonucleotides is used as forward primers and the third primer is serving as a common reverse one, for the respective PCR. The mismatching is formed within the forward primer, usually in the middle, but in some cases is formed at the 3′ end of the forward primer, as is the case of ARMS but in the competitive oligopriming setting, that is by using two competitive forward primers.

The forward primers that detect DNA sequence alterations in competitive oligopriming are a pair of synthetic short DNA sequences (competitive oligoprimers or COP primers), carrying the mutant or the normal configuration at the mutation site in the middle of their sequence and capable of discriminating between mutated and normal template DNA for annealing. Thus, competitive oligopriming is a system of allele-specific amplification, through differential primer annealing. In competitive oligopriming allele-specific amplification occurs for both alleles in the same reaction, which is the important feature of the method. Once the competitive amplification of DNA has been completed, identification of the mutant versus the normal amplified allele among the products of a competitive PCR

TABLE 2.1 Genetic loci for which the competitive oligopriming approach for mutation detection was applied. The position of the mismatch within COP primers, and the potential of multiplexing and/or high-throughput screening are recorded. *HPRT*: hypoxanthine phosphoribosyltransferase; *BCHE*: butyrylcholinesterase; *ALDH2*: aldehyde dehydrogenase; *PON*: paraoxonase/arylesterase.

Gene	Mismatch position	Multiplexing	High throughput	References
HPRT	Middle	No	No	Gibbs *et al.* (1989)
HBA2/HBA1, HBB	Middle	Yes	No	Chehab and Kan (1989)
HBB	Middle	Yes	No	Athanassiadou *et al.* (1995)
PON – Apolipo-protein B	3′ end	Yes	Yes	Germer and Higuchi (1999)
ALDH2	3′ end	No	No	Koch *et al.* (2000)
	3′ end	No	Yes	McClay *et al.* (2002)
MCAD Factor V	3′ end	No	No	Giffard *et al.* (2001)
Various SNPs	3′ end	No	Yes	Myakishev *et al.* (2001)
BCHE	Middle	No	No	Yen *et al.* (2003)

is rendered crucial. To this end a number of approaches have been applied, involving differential labeling of the competitive primers prior to their use in competitive PCR.

Thus, the competitive oligopriming assay (COP assay) is carried out in two consecutive stages: the first stage involves the DNA amplification by competitive oligopriming; the second stage involves the detection of the genetic identity – mutant versus normal – of the amplified material.

2.4.2 Genotyping with COP Assay

2.4.2.1 DNA Amplification In Vitro by COP

In COP PCR, both allele-specific oligonucleotides – the mutated and the normal one – are used in the same reaction and through the competition that arises between them for annealing on the target DNA sequence, the binding of the perfectly matched primer to the template is strongly favored relative to the primer differing by a single base.

Only one reverse primer is used in a competitive oligopriming PCR, to serve synthesis with both allele-specific primers, which therefore must derive from the same – opposite of that of the reverse primer – strand of DNA sequence. One nevertheless can formulate a competitive PCR, in which the competitive oligoprimers are of opposite direction, but this necessitates the use of two reverse primers, too, and most probably elaborate optimization procedures.

In COP PCR all three primers, namely the two forward primers for the mutation site, one (m) and the other (wt) and the common reverse primer, are used in one reaction (Fig. 2.6) with a given template DNA.

In this setting, competitive primers (m) and (wt) anneal only or mainly with the mutated and normal DNA, respectively, promoting correct priming and not vice versa. Results fall into three categories, depending on the genotype of the template DNA:

- If the amplified DNA contains only (or mainly) the normal primer (wt), the template DNA used is homozygous wild-type for this site.
- If the amplified DNA contains only (or mainly) the mutated primer (m), the template DNA used is homozygous mutant for this site.
- If the amplified DNA contains both (m) and (wt) primers in equal (or almost equal) amounts, the template DNA used is heterozygous for this site.

The efficiency of COP PCR depends on a number of parameters but limited data exist on this issue:

A) The *primers' length*: competition experiments between oligoprimers of different length have shown that 12-mers bear a greater potential for correct priming than 20-mers (Gibbs *et al*., 1989). However, the precise relationship between the length of competitive primers and their discriminatory potential depends on their overall sequence, as well as on the individual base mismatch (McClay *et al.*, 2002).

B) The *annealing temperature*: effective competition occurs at low stringency for shorter types of primers (12-mers) but annealing temperatures close to T_m (Yen *et al.*, 2003) or fairly stringent (McClay *et al.*, 2002) to high stringent condition (Athanassiadou *et al.*, 1995) gave good results with longer oligoprimers (16- to 20-mers).

C) The *primer concentration*: an excess primer to template ratio, promoters correct priming in low as well as in high stringency reactions (Athanassiadou *et al.*, 1995).

D) The *nature of the mismatch*: successful competition has been shown for a number of mismatches between primer and template DNA, as documented in only one work (Gibbs *et al*., 1989). Finally, no data exist for the possible effect on COP PCR of the exact position of the mismatch within the COP oligoprimers.

2.4.2.2 Genotype Detection Systems

The size of the PCR products in a COP PCR is the same in all three cases A, B, and C (Fig. 2.6), that is irrespective of the genotype of the PCR products and therefore a detection system of their genotype is necessary. This relies on the possibility for differential detection of the genetic identity (mutant vs normal) of the incorporated competitive primer and can be achieved by differential labeling of the two competitive primers. Various approaches have been employed so far: a color complementation assay was developed by Chehab and Kan (1989), which allows discrimination between fluorescent oligonucleotide primers. The addition of a 5′ GC tail to one of the primers so that it can be distinguished by T_m shift was employed by Germer and Higuchi (1999), and allele-specific PCR with universal energy transfer labeled primers was developed by Myakishev and coworkers (2001). Differential end labeling of the competitive oligoprimers with compounds that are recognized by different antibodies has also been applied (Athanassiadou *et al.*, 1995).

This approach was used for the differential identification of β-thalassemia mutation. In this system the detection of the identity (normal vs mutant) of the COP primers incorporated in each case is carried out by means of differential 5′ labeling of the COP primers (mutant primer with dansylchloride and normal primer with FITC) that were subsequently recognized by specific antibodies on a solid support. The common reverse primer was biotinylated so as to facilitate the formation of a conjugate on the amplified DNA.

Three sets of results are shown (see Fig. 2.7c), two sets for mutation *HBB*:c.93 − 21G > A (IVS I-110 G > A; I and II) and one for mutation *HBB*:c.92 + 6T > C (IVS I-6 T > C; III). The two sets of mutation *HBB*:c.93 − 21G > A

differ from each other only by 1°C in the annealing temperature; that is, case I is carried out at 1°C and case II at 2°C below high stringency. It is noticeable that a very low degree of mispriming persists for the n primer for mutation *HBB*:c.93 − 21G > A, like the practically negligible level of cross-hybridization of these primers when used in ASO hybridization, but this does not interfere with interpretation of results.

The detection of β-thalassemia and other mutations is efficient and reliable with the COP assay applied and this leads to the conclusions that the COP assay, as exemplified here, is a robust and reliable method of mutation detection.

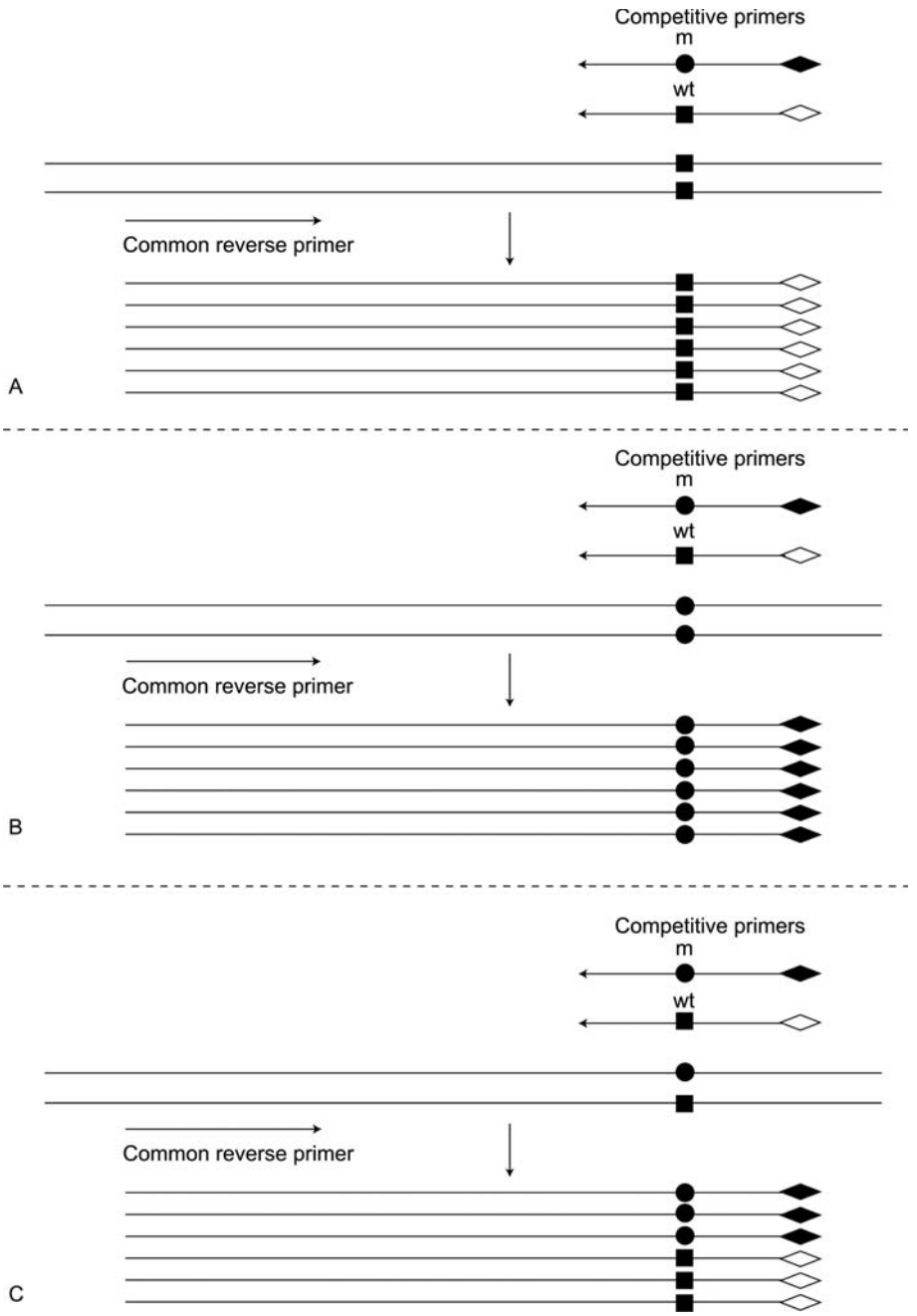


FIGURE 2.6 The three possible outcomes of a competitive PCR using a pair of allele-specific competitive primers: dark dots (m): mutant, and dark squares (wt): normal; as well as a common reverse primer, with template DNA **A.** Wild type (dark squares only), **B.** Homozygous mutant (dark dots only), and **C.** Heterozygous (dark squares and dark dots). Competitive PCR products in each case show correct priming as a result of efficient competition between the two primers for correct annealing on the template DNA. Dark and empty romvoids at the end of primers and PCR products represent differential labeling for the discrimination between normal and mutated amplified DNA.

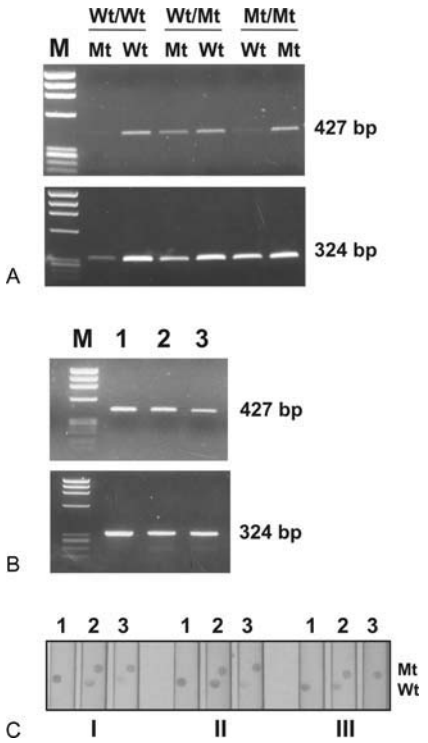


FIGURE 2.7 **A.** Results of single COP primer for two β-thalassemia mutations, *HBB*:c.93 − 21G > A (upper panel) and *HBB*:c.92 + 6T > C (lower panel), both at high stringency; m: mutant, n: normal. Intense correct priming is evident but also a variable degree of mismatch priming. **B.** Competitive PCRs with (1) wild-type, (2) heterozygous, and (3) homozygous mutant template DNA, for mutations *HBB*:c.93 − 21G > A (left) and *HBB*:c.92 + 6T > C (right). **C.** Detection of the mutant (m) and normal (n) products of the competitive PCRs for mutation *HBB*: c.93 − 21G > A (I and II) and *HBB*:c.92 + 6T > C (III), all at primer annealing temperatures at or close to T_m (high stringency).

2.4.3 Development of COP Assays with a 3′ End Mismatch

A combination of COP array and ARMS has also been applied in some cases. This is a COP system in which the two forward primers (mutant and normal) are designed to form the mismatch at their 3′ end, thus promoting differential DNA synthesis in one reaction, rather than differential primer annealing. Such a system was used for the purpose of allele association studies (Koch *et al.*, 2000) and the genotyping COP PCR products was by fluorescent labeling of the competitive primers at the 5′ end, consequently detected on an automated fluorescent DNA sequencer.

Another system of this category, developed by Germer and Higuchi (1999), is termed T_m-shift genotyping, as the detection of primer identity within COP PCR products is based on the addition of a GC-tail in one of the allele-specific primers. This results in a shift in the T_m of the PCR product that has incorporated the primer with the GC tail.

2.4.4 Multiplexing and High-Throughput COP Assays

Simultaneous DNA amplification of more than one genetic site on one DNA sample can be carried out by multiplex COP assay (Gibbs *et al.*, 1989) and it is dependent on the system by which the genetic identity of the COP PCR products is determined. When the different pairs of COP primers used produce COP PCR products of different length, it is possible to determine their identity by gel electrophoresis. However, a less complex detection method than gel electrophoresis is usually desirable.

A detection system based on color complementation assay was developed by Chehab and Kan (1989) for the simultaneous amplification of DNA loci for two β-thalassemia mutations and an α-globin gene deletion (five primers in total) that were simultaneously analyzed in one-tube reaction.

When a discrimination system for various PCR products, generated in a one-tube multiplex PCR, is not available, multiplex COP assay is possible only as a system of a number of COP PCR tubes, designed to operate in one PCR machine with the same PCR program.

In both these multiplex assays there is a requirement for the same annealing temperature for efficient competition of all COP primer pairs used. Shorter primers are expected to compete efficiently in not just one annealing temperature value, but rather in a spectrum of temperatures (albeit in lower stringency). The multiplex system in this case would be more tolerant with respect to the annealing temperatures of the individual COP PCRs, than with the longer COP primers. Evidently, every disease entity caused by multiple mutations needs its own analysis, which will determine the conditions for simultaneous detection of as many mutations as possible.

Regarding high-throughput approaches, that is testing many samples for one SNP, two COP PCRs have been developed, both using competitive oligoprimers with the mismatch at their 3′ end. One of them (Myakishev *et al.*, 2001) involved PCR amplification of genomic DNA with two-tailed allele-specific primers, which introduce priming sites for universal energy-transfer labeled primers. The accuracy of the method is greater in conditions of competitive oligopriming than when single allele-specific primer is used, with little if any non-specific signal recorded, especially with small amounts (0.4 ng) of genomic DNA required per reaction. In addition, it is cost effective as one set of energy-transfer primers can be used for all SNPs analyzed. A second approach is using a fluorescent label at their 5′ end, so that PCR products can be detected by conventional automatic sequencer (McClay *et al.*, 2002). The SniPTag used in this method is capable of discriminating among all combinations of nucleotide substitutions. A very interesting feature of this approach is that it is possible to increase throughput by loading the PCR products

for several markers' loci in a single lane on the automated sequencer, provided that these products have been designed to be of different length.

2.4.5 Advantages and Limitations of the COP Assay

The main advantage of COP assay is that it does not need heavy equipment or complex procedures and the bare essential of molecular biology analysis is enough in most cases. Additionally, it is cost effective as most labeling is commercially available for routine work. Thus the development of COP aims at the formulation of mutation detection systems that can be used almost anywhere, supporting molecular diagnosis in geographical areas that are not close to hospitals. Another advantage is the possibility for multiplexing, as the primer competition for correct annealing to template DNA is possible for temperatures of medium to low stringency, thus allowing for the accommodation of the annealing temperatures for several mutations in one tube or in one PCR run. High-throughput approaches are also possible, but do not include a setting for true multiplexing at the same time, due to the narrow range of optimum annealing temperature for each SNP.

The main disadvantage of the COP assay arises in the detection system, as there is a need to differentiate between equal size DNA molecules, but of different genetic identity, produced in one PCR. This difficulty has been interfering with the widespread use of COP. Second, as with the other allele-specific mutation detection methods, COP assay is applicable exclusively with DNA changes whose precise nature is known in advance, so that the COP oligoprimers can be designed methods in which prior knowledge of the nature of the mutation is not an absolute requirement, like denaturing gradient gel electrophoresis (DGGE, see Chapter 6), or single strand conformation polymorphism (SSCP) analysis (see Chapter 4), which can be applied for known and unknown DNA sequence changes. Finally, an aspect that makes COP assay difficult to apply is the rather long and laborious optimization procedure that is required, in order for every mutation, to specify the exact conditions for a COP PCR. This includes the elimination of misprimming and of improper differential detection of PCR products, which may interfere with interpretation of the results.

2.5 CONCLUSIONS

Allele-specific mutation detection is feasible and quite straightforward using all three above-mentioned methodologies, namely PCR-ARMS, PCR-ASO, and COP assay. All methods have a comparably high degree of accuracy and specificity. Although PCR-ARMS and PCR-ASO enjoy broad applicability in low-throughput laboratories, COP's potential has not been fully exploited yet, presumably because the detection systems are somewhat elaborate and need special equipment.

The question of the efficiency, in terms of accuracy and specificity, of the competitively primed versus the conventional single allele-specific DNA amplification was addressed by Giffard and coworkers (2001). In a simulation study, it has been argued that although conventional allele-specific amplification has somewhat higher inherent specificity than competitive oligopriming reaction, it actually may be easier to optimize the latter ones to offer greater reproducibility and tolerance to alterations in target amounts without any significant loss of specificity. This was supported by the work of Myakishev *et al.* (2001), mentioned earlier, where the accuracy of the method is greater in conditions of competitive oligopriming than when single allele-specific primer is used.

It is noteworthy that allele-specific mutation detection set the standards for the development of high-throughput systems, such as microarrays for the purpose of large-scale SNP genotyping that concern genome scans and DNA diagnostics of genetic, acquired, and infectious diseases (see also Chapters 16 and 17). A methodology, which resembles PCR-ASO, has also been employed to screen for *HBB* mutations in an array format (Foglieni *et al.*, 2004). A microelectronic DNA chip was constructed that contained PCR-amplified fragments from a large number of β-thalassemia homozygous and compound heterozygous samples, using primer pairs, one of which was biotinylated at its 5′ end (Santacroce *et al.*, 2002). Following denaturation, the biotinylated strand was electronically spotted to discrete sites on streptavidin-coated gel pad surfaces, and allele-specific dye-labeled oligonucleotide probes were used to detect wild-type and nine different mutants. For this array-based genotyping platform, sensitivity and specificity of 90–100% have been reported, which together with the low-cost reagents and short processing times strongly indicate a potential use for this technology in screening programs for large populations with a high incidence of different types of β-thalassemia. However, this technology has not yet been implemented for routine globin gene mutation detection.

REFERENCES

Athanassiadou, A., Papachatzopoulou, A., and Gibbs, R.A. (1995). Detection of genetic analysis of beta-thalassemia mutations by competitive oligopriming. *Hum. Mutat.* 6, 30–35.

Ayyadevara, S., Thaden, J.J., and Shmookler Reis, R.J. (2000). Discrimination of primer 3′-nucleotide by Taq DNA polymerase during polymerase chain reaction. *Anal. Biochem.* 284, 11–18.

Budowle, B., Lindsey, J.A., DeCou, J.A., Koons, B.W., Giusti, A.M., and Comey, C.T. (1995). Validation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and GC (PM loci), and HLA-DQa using a multiplex amplification and typing procedure. *J. Forensic Sci.* 40, 45–54.

Cai, S.P., Wall, J., Kan, Y.W., and Chehab, F.F. (1994). Reverse dot blot probes for the screening of b-thalassemia mutations in Asians and American Blacks. *Hum. Mutat.* 3, 59–63.

Chan, V., Yam, I., Chen, F.E., and Chan, T.K. (1999). A reverse dot-blot for rapid detection of non-deletion a thalassaemia. *Br. J. Haematol.* 104, 513–515.

Chehab, F.F. (1993). Molecular diagnostics: past, present, and future. *Hum. Mutat.* 2, 331–337.

Chehab, F.F., and Kan, Y.W. (1989). Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA,* 86, 9178–9182.

Chehab, F.F., and Wall, J. (1992). Detection of multiple cystic fibrosis mutations by reverse dot blot hybridization: a technology for carrier screening. *Hum. Genet.* 89, 163–168.

Conner, B.J., Reyes, A.A., Morin, C., Itakurs, K., Teplitz, R.L., and Wallace, R.B. (1983). Detection of sickle cell S-globin allele by hybridization with synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA,* 80, 278–282.

Eitan, Y., and Kashi, Y. (2002). Direct micro-haplotyping by multiple double PCR amplifications of specific alleles (MD-PASA). *Nucleic Acids Res.* 30, e62.

Foglieni, B., Cremonesi, L., Travi, M., Ravani, A., Giambona, A., Rosatelli, M.C., Perra, C., Fortina, P., and Ferrari, M. (2004). Beta-thalassemia microelectronic chip: a fast and accurate method for mutation detection. *Clin. Chem.* 50, 73–79.

Foglietta, E., Bianco, I., Maggio, A., and Giambona, A. (2003). Rapid detection of six common Mediterranean and three non-Mediterranean a-thalassemia point mutations by reverse dotblot analysis. *Am. J. Hematol.* 74, 191–195.

Germer, S., and Higuchi, R. (1999). Single-tube genotyping without oligonucleotide probes. *Genome Res.* 9, 72–78.

Gibbs, R.A., Nguyen, P.-N., and Caskey, T.C. (1989). Detection of single DNA base differences by competitive oligonucleotide priming. *Nucleic Acids Res.* 17, 2437–2448.

Giffard, P.M., McMahon, J.A., Gustafson, H.M., Barnard, R.T., and Voisey, J. (2001). Comparison of competitively primed and conventional allele-specific nucleic acid amplification. *Anal. Biochem.* 292, 207–215.

Gold, B. (2003). Origin and utility of the reverse dot-blot. *Expert Rev. Mol. Diagn.* 3, 143–152.

Huang, M.-M., Arnheim, N., and Goodman, M.F. (1992). Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res.* 20, 4567–4573.

Ikuta, S., Tagaki, K., Wallace, R.B., and Itakura, K. (1987). Dissociation kinetics of 19 base paired oligonucleotide-DNA duplexes containing different single mismatched base pairs. *Nucleic Acids Res.* 15, 797–811.

Koch, H.G., McClay, J., Loh, E.-W., Higuchi, S., Zhao, J.-H., Sham, P., Ball, D., and Craig, I.W. (2000). Allele association studies with SSR and SNP markers at known physical distances within a 1Mb region embracing the ALDH2 locus in Japanese, demonstrates linkage disequilibrium extending up to 400 kb. *Hum. Mol. Genet.* 9, 2993–2999.

Kwok, S., Kellogg, D.E., McKinney, N., Soasic, D., Goda, L., Levenson, C., and Sninsky, J.J. (1990). Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.* 18, 999–1005.

Makowski, G.S., Nadeau, F.L., and Hopfer, S.M. (2003). Single tube multiplex PCR detection of 27 cystic fibrosis mutations and 4 polymorphisms using neonatal blood samples collected on Guthrie cards. *Ann. Clin. Lab. Sci.* 33, 243–250.

McClay, J.L., Sugden, K., Koch, H.G., Higuchi, S., and Craig, I.W. (2002). High-throughput single-nucleotide polymorphism genotyping by fluorescent competitive allele-specific polymerase chain reaction (SniPTag). *Anal. Biochem.* 301, 200–206.

Myakishev, M.V., Khripin, Y., Hu, S., and Hamer, D.H. (2001). High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers. *Genome Res.* 11, 163–169.

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C., and Markham, A.F. (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* 17, 2503–2516.

Nichols, W.C., Liepieks, J.J., McKusick, V.A., and Benson, M.D. (1989). Direct sequencing of the gene for Maryland/German familial amyloidotic polyneuropathy type II and genotyping by allele-specific enzymatic amplification. *Genomics* 5, 535–540.

Nollau, P., and Wagener, C. (1997). Methods for detection of point mutations: performance and quality assessment. *Clin. Chem.* 43, 1114–1128.

Nowaczyk, M.J., Nakamura, L.M., Eng, B., Porter, F.D., and Waye, J.S. (2001). Frequency and ethnic distribution of the common DHCR7 mutation in Smith–Lemli–Opitz syndrome. *Am. J. Med. Genet.* 102, 383–386.

Okayama, H., Curiel, D.T., Brantly, M.L., Holmes, M.D., and Crystal, R.D. (1989). Rapid nonradioactive detection of mutations in the human genome by allele-specific amplification. *J. Lab. Clin. Med.* 114, 105–113.

Old, J.M., Varawalla, N.Y., and Weatherall, D.J. (1990). Rapid detection and prenatal diagnosis of b-thalassaemia: studies in Indian and Cypriot populations in the UK. *Lancet* 336, 834–837.

Orkin, S.H., Markham, A.F., and Kazazian, H.H., Jr. (1983). Direct detection of the common Mediterranean beta-thalassemia gene with synthetic DNA probes: an alternate approach for prenatal diagnosis. *J. Clin. Invest.* 71, 775–779.

Patrinos, G.P., Kollia, P., and Papadakis, M.N. (2005). Molecular diagnosis of thalassemia and hemoglobinopathies: from RFLP analysis to microarrays. *Hum. Mutat.* 26, 399–412.

Pirastu, M., Kan, Y.W., Cao, A., Conner, B.J., Teplitz, R.L., and Wallace, R.B. (1983). Prenatal diagnosis of b-thalassemia: detection of a single nucleotide mutation in DNA. *N. Engl. J. Med.* 309, 284–287.

Polten, A., Fluharty, A.L., Fluharty, C.B., Kappler, J., von Figura, K., and Gieselmann, V. (1991). Molecular basis of different forms of metachromatic leukodystrophy. *N. Engl. J. Med.* 324, 18–24.

Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B., and Erlich, H. A. (1986). Analysis of enzymatically amplified b-globin and HLA-DQa DNA with allele-specific oligonucleotide probes. *Nature* 324, 163–166.

Saiki, R.K., Chang, C.-A., Levenson, C.H., Warren, T.C., Boehm, C.D., Kazazian, H.H., Jr., and Erlich, H.A. (1988b). Diagnosis of sickle cell anemia and b-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. *N. Engl. J. Med.* 319, 537–541.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G. T., Mullis, K.B., and Erlich, H.A. (1988a). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487–491.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985). Enzymatic amplification of b globin sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230, 1350–1354.

Saiki, R.K., Walsh, P.S., Levenson, C.H., and Erlich, H.A. (1989). Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA,* 86, 6230–6234.

Santacroce, R., Ratti, A., Caroli, F., Foglieni, B., Ferraris, A., Cremonesi, L., Margaglione, M., Seri, M., Ravazzolo, R., Restagno, G., Dallapiccola, B., Rappaport, E., Pollak, E.S., Surrey, S., Ferrari, M., and Fortina, P. (2002). Analysis of clinically relevant single-nucleotide polymorphisms by use of microelectronic array technology. *Clin. Chem.* 48, 2124–2130.

Sarkar, G., Cassady, J., Bottema, C.D.K., and Sommer, S.S. (1990). Characterization of polymerase chain reaction amplification of specific alleles. *Anal. Biochem.* 186, 64–68.

Sommer, S.S., Cassady, J.D., Sobell, J.L., and Bottema, C.D. (1989). A novel method for detecting point mutations or polymorphisms and its application to population screening for carriers of phenylketonuria. *Mayo Clin. Proc.* 64, 1361–1372.

Sommer, S.S., Groszbach, A.R., and Bottema, C.D. (1992). PCR amplification of specific alleles (PASA) is a general method for rapidly detecting known single-base changes. *Biotechniques* 12, 82–87.

Waterfall, C.M., and Cobb, B.D. (2001). Single tube genotyping of sickle cell anaemia using PCR-based SNP analysis. *Nucleic Acids Res.* 29, e119.

Waye, J.S., Nakamura, L.M., Eng, B., Hunnisett, L., Chitayat, D., Costa, T., and Nowaczyk, M.J. (2002). Smith–Lemli–Opitz syndrome: carrier frequency and spectrum of DHCR7 mutations in Canada. *J. Med. Genet.* 39, e31.

Wu, D.Y., Ugozzoli, L., Pal, B.K., and Wallace, R.B. (1989). Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA,* 86, 2757–2760.

Ye, S., Dhillon, S., Ke, X., Collins, A.R., and Day, I.N.M. (2001). An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res.* 29, e88.

Yen, T., Hightingale, N., Burns, J.C., Sullivan, D.R., and Stewart, P. M. (2003). Butyrylcholinesterase (BCHE) genotyping for post-succinylcholine apnea in an Australian population. *Clin. Chem.* 49, 1297–1308.

Zhang, Y., Coyne, M.Y., Will, S.G., Levenson, C.H., and Kawasaki, E.S. (1991). Single-base mutational analysis of cancer and genetic diseases using membrane-bound modified oligonucleotides. *Nucleic Acids Res.* 19, 3929–3933.

This page intentionally left blank

Chapter 3

Enzymatic and Chemical Cleavage Methods to Identify Genetic Variation

Chinh T. Bui[1], Emmanuelle Nicolas[2], Georgina Sallmann[1], Maria Chiotis[1], Andreana Lambrinakos[1], Kylee Rees[1], Ian Trounce[1], Richard G.H. Cotton[1], Lauryn Blakesley[2], Andrew K. Godwin[2] and Anthony T. Yeung[2]

[1] *Genomic Disorders Research Centre and The University of Melbourne, Fitzroy, Melbourne, Australia;*

[2] *Fox Chase Cancer Center, Philadelphia, PA, USA*

3.1 INTRODUCTION

Although there are several methods for the high sensitivity detection of known mutations, unknown mutations are more difficult to uncover. Recently, the latter methods have been improved significantly and utilized in practical applications. These include the establishment of a single nucleotide polymorphism (SNP) map for a specific part of the genome for a specific animal population or the screening for unknown mutations in important genes, such as cancer susceptibility genes (Neuhausen and Ostrander, 1997; Warner *et al*., 1999; Oleykowski *et al*., 1998; Hartge *et al*., 1999). These genes are large, with many exons, and thus hundreds of possible mutations that affect the functions of those proteins can be found. Also, these methods can serve as tools for reverse genetics to screen for chemically induced point mutations in specific regions of specific genes (Colbert *et al*., 2001) and for screening for specific mutations in the genome of emerging pathogenic microorganisms (Sokurenko *et al*., 2001). In this chapter, two of these mutation detection methods will be discussed in detail.

A mutation is by definition a change in the DNA sequence as compared to the population at large. Historically, mutations are detected by phenotypic changes. At the DNA level, it is often, but not always, revealed by direct DNA sequencing. Indeed, most researchers involved in mutation screening have encountered mutations that are not detectable by Sanger dideoxy-DNA sequencing, but obvious to complementary approaches that detect changes in the physical properties of the DNA helix when a mismatch is present. In mutation screening methods not based on enzymatic DNA polymerization, the mutated DNA helix often is first converted to mismatch heteroduplexes when the polymerase chain reaction (PCR; Mullis and Faloona, 1987) products of two alleles of the gene of interest are amplified by PCR, mixed, denatured, and rehybridized. Next, the chemical or enzymatic properties of a mismatched base are exploited to lead to a break in the DNA strand near the mismatch. Finally, a suitable fragment analysis method is used to visualize the shortened DNA fragments, in either single-stranded form or the double-stranded form, that are produced by the DNA break near the mismatch residue. The suitability of a method for each laboratory depends on the existing skills in that laboratory and the available analysis platform.

The mixing of two alleles of a gene of interest in PCR results in heteroduplexes at the end of a denaturation and renaturation step. However, homozygous mutations and hemizygous mutations will not lead to the formation of mismatch heteroduplexes. Thus, it is important that a mutation detection method be sufficiently sensitive to detect a mutation in a pool of multiple normal alleles so that statistically a mutation will not be missed.

3.2 CHEMICAL PROPERTIES OF MISMATCHES

Mismatches can be either insertions/deletions or base substitutions. The latter include A/A, G/G, C/C, T/T, A/G, A/C, G/T, C/A, and C/T. It is important to realize that no two mismatches have the same chemistry and structure, therefore most mutation detection methods may recognize either one or more of the lesions with high specificity but not all of them, or conversely, recognize all the lesions at only moderate specificity.

A mismatched base pair presents minimal information to the mutation detection system as compared to the

recognition of multiple base pair palindrome sequences by restriction enzymes. In a mismatch, often only two opposing bases in the two DNA strands are involved with modification in the hydrogen bonding, leading to enhanced chemical reactivity. Alternatively, a mismatched base may flip out of the helix at some frequency for enzyme recognition. The mismatched base may also destabilize local regions of the DNA helix to create single-strandedness. The single-strandedness of insertion/deletions can also be exploited (Burdon and Lees, 1985). However, the natural tendency of DNA helix to undergo breathing at AT-rich sequences and to exhibit secondary structures like hairpins and cloverleaf structures at some sequences can create false positive signals for mismatch detection systems. Other sequences have structures that can hide a mismatch and cause some mismatch detection methods to give false negative results.

3.3 CHEMICAL CLEAVAGE OF MISMATCH METHOD FOR MUTATION DETECTION

Detection of unknown mutations is a complex and expensive task, particularly for screening kilobase lengths of DNA sequence for a single base change and/or small insertions and deletions. Chemical Cleavage of Mismatch (CCM) technology, developed by Cotton and coworkers (1988), theoretically establishes simple and cost-effective chemical means to detect all types of mismatches, and thus mutations at this point in time (Cotton *et al*., 1988). The protocol employs two commercially available chemicals, hydroxylamine (NH_2OH) and potassium permanganate ($KMnO_4$) (or osmium tetroxide (OsO_4) in an earlier version) to react with mismatched cytosine and thymine residues, respectively (Ramus and Cotton, 1996; Cotton, 1999). The modified mismatched DNA becomes highly susceptible to cleavage by piperidine. The resulting DNA fragments are simply analyzed by denaturing polyacrylamide gel or capillary electrophoresis (Ren, 2001) to identify the mutation sites. Due to its simplicity in manipulation and high sensitivity, the method has been continuously improved and documented in very simple protocols by many users (see Table 3.1). For example, $KMnO_4$ was first introduced in 1990 to replace the hazardous chemical OsO_4. Radioactive-labeled (^{32}P and ^{35}S) DNA probes were replaced with fluorescence-labeled primers. DNA was also incorporated onto streptavidin-coated magnetic beads to bypass multiple washing steps. Both chemical reactions (with $KMnO_4$ and NH_2OH) were simplified by being carried out in a single tube. Recently, the DNA samples were

TABLE 3.1 Protocols used in chemical cleavage of mismatches.

Milestones	Protocols	References
1988	The first described CCM technique (hydroxylamine and osmium tetroxide)	Cotton *et al*. (1988)
1990	$KMnO_4$/TMAC (tetramethyl ammonium chloride) was introduced to replace the hazardous chemical OsO_4	Gogos *et al*. (1990)
1991	^{32}P- and ^{35}S-labeled probes for PCR products	Saleeba and Cotton (1991), Cotton *et al*. (1988)
1995	Use of fluorescent primers	Verpy *et al*. (1994), Haris *et al*. (1994)
1996	Single tube CCM method	Ramus and Cotton (1996)
1998	Development of solid phase CCM	Rowley *et al*. (1995), Roberts *et al*. (1997)
1999	$KMnO_4$/TEAC (tetraethyl ammonium chloride) condition was applied	Roberts *et al*. (1997), Lambrinakos *et al*. (1999)
1999	Cleavage reactions with amine-bases	Block (1999)
2001	Solid phase CCM method (silica bead)	Bui *et al*. (2003a)
2001	Piperidine in cleavage loading dye solution	Cotton and Bray (2001)
2001	Capillary electrophoresis with laser-induced fluorescence detection	Ren (2001)
2003	Spectroscopic method (oxidized thymine and cytosine were followed by measuring absorbance at 420 nm)	Bui *et al*. (2003b)
2006	Spectroscopic method (using plate reader)	Tabone *et al*. (2006)

adsorbed onto silica beads under high salt concentrations (3 M TEAC solution) before undergoing the chemical modification steps (Bui *et al*., 2003a). The post-cleavage washing steps also were eliminated by incubation of piperidine together with the loading dye buffer. In addition, various types of amine-bases were also reported to improve cleavage reactions used in CCM methods (Block, 1999). All relevant protocols are suitable for analyzing mismatches located on long stretches of DNA (up to 2.0 kb). Recently, the cleavage step with piperidine was omitted in CCM as the permanganate oxidation of mismatched thymine and cytosine could be followed up via UV-visible spectroscopy (Bui *et al*., 2003b). This technique was very simple and suitable for short DNA heteroduplexes (up to 300 bp). For the purpose of practical guidance, this chapter describes the solid phase, liquid phase, and spectroscopic protocols which are most commonly used in our laboratories and others (Gogos *et al*., 1990; Rowley *et al*., 1995; Roberts *et al*., 1997; Lambrinakos *et al*., 1999; Tabone *et al*., 2006).

The formation of heteroduplexes can be performed by mixing and reannealing equimolar amounts of wild-type and mutant DNA, thereby resulting in mismatched base pairs. In principle, the imperfect duplexes are different from their corresponding perfect duplexes in terms of their local conformational changes, physical and chemical properties (Bui *et al*., 2002). These changes are reflected at lower melting temperatures in the imperfect duplexes (Patel *et al*., 1982), less stability as indicated by thermodynamic constants (Patel *et al*., 1982) and extra-helical or "flip-out" phenomena of mismatched bases (Kao *et al*., 1993; Roberts and Cheng, 1998). However, such discrepancies induced by mismatches are so small that the mismatched sites can be recognized and cleaved only by enzymatic means and not by chemical reagents (Kennard, 1988). For this reason, the mismatched sites require further modification by chemicals before the chemical cleavage reaction (piperidine) can take place. Two common chemicals, $KMnO_4$ and NH_2OH, are the most effective for modifying thymine and cytosine mismatches, respectively, in this regard. The former reaction leads to the formation of a mixture of a thymine glycol and a ketone analog (some evidence indicates that $KMnO_4$ also reacts with cytosine but to a lesser extent; Bui and Cotton, 2002) and the latter gives rise to a modified cytosine containing a hydroxylamine moiety (see Fig. 3.1). Physical data established on the model for short mismatched oligonucleotides (38 bp) indicate that the differences in melting temperatures and gel mobility of the chemically modified heteroduplex samples are significant compared to the unmodified one (Bui *et al*., 2003b). The results suggest that the destabilized mismatched site favors the site-selective cleavage reaction of piperidine and the mismatch can be pinpointed on a denaturing gel as cleaved bands. Mechanism of the modification reaction by $KMnO_4$ is fully elucidated at molecular levels. $KMnO_4$ reacts with thymine and cytosine to form the complexes, which gave strong absorption at 420 nm. The formation of color intermediate was confirmed on solid support and formed a basis for a simple spectroscopic technique to follow up the oxidation of mismatched DNA (Bui *et al*., 2004). Proof of concept was carried out on synthetic 39 bp DNA fragments containing single base mismatches and the oxidation results with $KMnO_4$ clearly showed the color change, which could be detected by measuring the absorbance at 420 nm after 30 min reaction (end-point analysis) or by continuously scanning the reaction mixture over a period of time (scanning analysis). The mutation detection via spectroscopy has been successfully applied in longer mismatched DNA (up to 300 bp) using UV/Vis microplate reader (Tabone *et al*., 2006). The assay was validated as 100% effective in detection of single point mutation using a blind manner for all mismatched DNA samples, derived from the mouse β-globin gene promoter region, containing single point mutations as

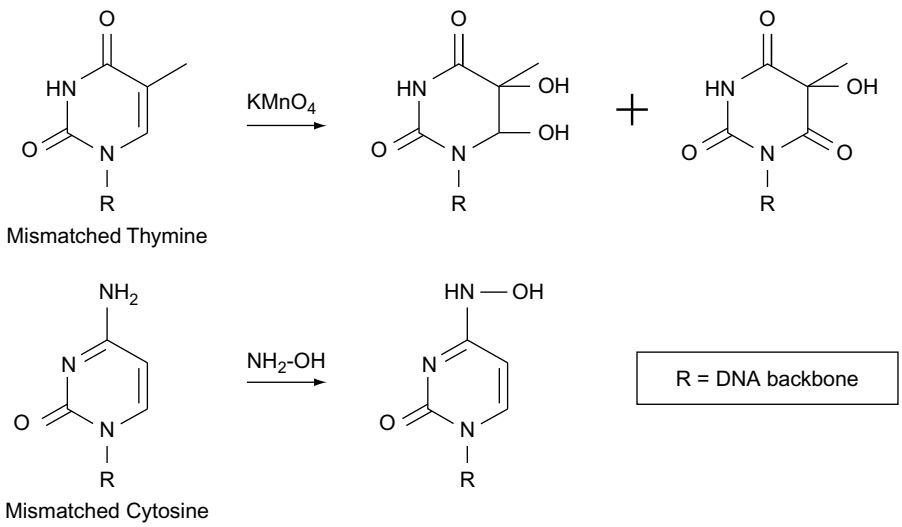


FIGURE 3.1 Chemical reactions involved in the CCM method. $KMnO_4$ selectively reacts with mismatched thymine to afford a mixture of thymine glycol and a ketone analog, and hydroxylamine reacts with mismatched cytosine to afford a single monosubstituted product under the described conditions.

well as patients who were previously screened for mutations by other methods (sequencing, SSCP or DHPLC analysis).

In general, CCM and new versions (solid phase or spectroscopic technique) are considered as the method of choice as it can detect all key types of mismatch (T/G, T/C, C/C, A/C, and T/T), which represent all eight possible mispairs that can be generated from the heteroduplex formation (see Fig. 3.2). The other mismatches (A/G, G/G, and A/A) can be detected via the complementary heteroduplexes (i.e. A/A will be detected by T/T). It is also emphasized that some neighboring matched bases also respond to the reactions due to instability of the whole region near the mismatched site. In addition, when both mutant and wild-type DNA are labeled, the chance of detecting mutations will be doubled (see Fig. 3.2).

3.3.1 Liquid Phase Protocol

The standard liquid phase CCM protocol consists of six steps: chemical modification with $KMnO_4$ and hydroxylamine, termination, separation, washing, cleavage, and gel-electrophoresis (see Table 3.2). The wild-type and mutant DNA samples are amplified (by PCR) using the fluorescence-labeled primers (6-FAM and HEX at 5′A and 3′A ends,

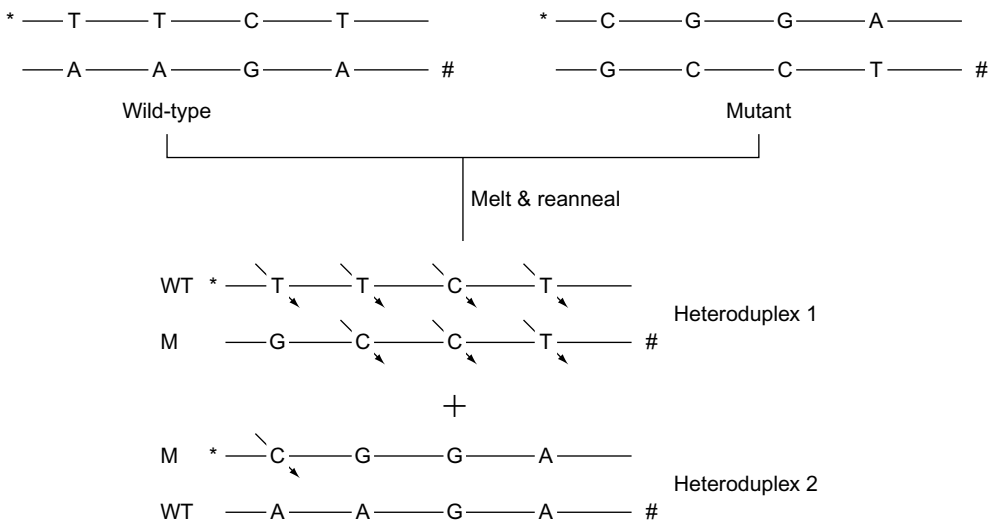


FIGURE 3.2 Formation of heteroduplexes from the wild-type (WT) and mutant (M) homoduplexes. * indicates 5′A-FAM and # indicates 3′A-HEX fluorescence-labeled DNA strands.

TABLE 3.2 The liquid phase protocol.

Steps	Procedure[a]
$KMnO_4$ reaction	0.2 mL of 100 mM $KMnO_4$ + 19.8 ml of 3 M TEAC. Incubate at 25°C for 10 min.
Hydroxylamine reaction	20 mL of 4.2 M hydroxylamine solution (pH = 6.0 with triethylamine). Incubate at 37°C for 40 min.
Termination	Add 200 ml of STOP buffer (25 mg/ml tRNA, 0.1 mM EDTA, 0.3 M sodium acetate, pH 5.2).
Separation	Add 750 ml of ice-cold 100% ethanol (at −20°C for at least 30 min). Centrifuge for 20 min at 14,000 rpm to collect the precipitates.
Washing	Rewash the pellet with 200 ml of 70% ethanol. Centrifuge it again for 10 min (14,000 rpm). Air-dry the pellet for 10 min.
Cleavage	Add 10 ml of cleavage loading dye solution (20 ml piperidine, 64 ml formamide and 16 ml dye (50 mg blue dextran in 1 ml distilled water)) to the DNA pellets and incubate at 90°C for 30 min.
Gel separation[b]	Load samples onto a denaturing gel and analyze on an ABI 377 DNA sequencer (2 ml of sample is needed for each well).

[a]6 ml of homoduplex and heteroduplex DNA (0.6 mg DNA) in TE buffer (10 mM Tris-HCl, 1 mM EDTA, ethylene-diamine-tetra-acetic acid, pH 8.0) is used for each of the reactions.
[b]The denaturing polyacrylamide gel 4.25% (acrylamide: bis-*acrylamide, 19:1), 6 M urea gel in the TBE buffer (16.2 g Tris-base, 8.1 g boric acid and 1.12 g EDTA in 1,500 ml distilled water, pH = 8.0).*

respectively). Subsequently, the amplified DNA samples are purified, using either a commercially available purification kit (Stratagene™ PCR Purification Kit, CA, USA) or agarose gel electrophoresis. The resulting wild-type and mutant DNA samples are mixed in equal amounts to form the heteroduplexes prior to the assay. In the first step of the described protocol, the reactions of heteroduplex DNA with $KMnO_4$ and hydroxylamine are usually carried out in separate tubes. It is also noted that both reactions can be optimized and carried out in a single tube protocol (but this will not be discussed in this chapter). When the reactions are completed, the DNA samples are separated by ethanol precipitation and washed carefully before the next cleavage step. To simplify the protocol, the gel loading dye (blue dextran) is added to the cleavage solution and the reaction mixtures can be directly loaded onto a denaturing gel immediately after the cleavage step. The DNA fragments are analyzed by an ABI 377-DNA sequencer without further purification (only 2 ml of sample is needed for loading). Two types of size standards (Tamra 500 and Tamra 2500) are added to the gel. Therefore, analysis with the ABI sequencer allows identification of the positions of the mismatches without the use of sequencing.

3.3.2 Solid Phase Protocol

In order to bypass the separation and washing steps, the solid phase protocols have been developed by immobilizing DNA on silica solid supports (see Table 3.3). Attachment of DNA on silica in a high salt solution has been established and well practiced as an effective purification technique (Bui *et al*., 2003a). In the described protocol, the commercially available silica beads are used (MO BIO Laboratories Inc. CA, USA) to bind to the DNA, and the DNA-attached beads are then sequentially treated with chemicals, washing solutions, and piperidine for cleavage. In the last cleavage step, the DNA fragments are cleaved by piperidine and released simultaneously from the solid supports. The resulting supernatant is isolated and loaded directly onto a denaturing polyacrylamide gel.

Mismatch detection is based on the comparison of the homoduplex and heteroduplex traces. A mutation is identified by cleavage peaks present in the trace of heteroduplex sample but not in the control homoduplex sample.

In the authors' laboratories, both liquid phase and solid phase protocols are routinely used for mismatch detection, and some typical examples are described next.

The liquid and solid phase protocols were successfully carried out on 547 bp DNA fragments derived from the cloned mouse β-globin gene promoter DNA to detect T/C and T/G mismatches, respectively (see Figs 3.3 and 3.4).

Detection of single base insertion and deletion (C base) was carried out with DNA fragments (893 bp in Fig. 3.5 and 660 bp in Fig. 3.6) derived from human mitochondrial DNA by using the liquid phase protocol with hydroxylamine reaction.

3.3.3 Spectroscopic Protocol

Spectroscopic analysis is based on the increased oxidation level of the reaction of mismatched thymine and cytosine with potassium permanganate. In a typical protocol,

TABLE 3.3 The solid phase protocol.

Steps	Procedure
Loading DNA onto solid supports	Mix 3 ml of Ultra-bind bead suspension with 1 ml of DNA samples (0.1 to 0.2 mg of homoduplex or heteroduplex DNA) on shaker for 1 hour at 25°C. Centrifuge at 14,000 rpm, the pellets are collected. Wash the beads by resuspending in the Ultra-wash solutions (2 × 500 mL). Centrifuge the mixture, discard the supernatant, and air-dry the beads at 25°C for 15 min.
$KMnO_4$ reaction	Mix the beads with 0.3 ml of 100 mM $KMnO_4$ in 29.7 ml of 3 M TEAC solution. Allow to stand at 25°C for 5 min.
Hydroxylamine reaction	Mix the beads with 15 ml of 4.2 M hydroxylamine solution in 15 ml of 3 M TEAC solution. Allow to stand at 37°C for 40 min.
Washing	The beads are separated by centrifugation and washed twice with the Ultra-wash solution (200 mL per wash) and the pellets are air-dried at 25°C for 15 min.
Cleavage	Add 10 mL of the cleavage dye solution to each reaction tube and vortex well. Incubate the tubes at 90°C for 30 min. Mix tubes by flicking occasionally. Cool the tubes on ice and the supernatant is separated by centrifugation.
Gel separation	Load samples onto a denaturing gel and analyze on an ABI 377 DNA sequencer (2 ml of sample is needed for each well).

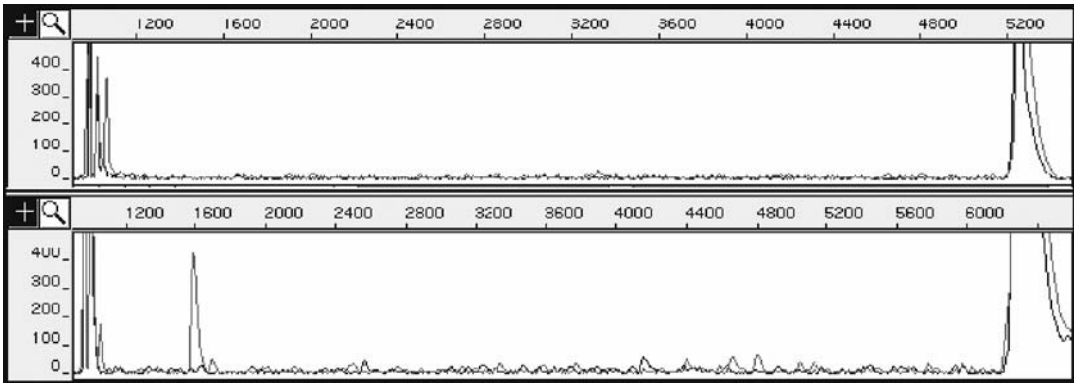


FIGURE 3.3 Detection of T/C mismatch by the liquid phase protocol. The heteroduplex DNA trace (bottom) displays a strong cleavage peak at the mismatched C base of the 3′A-HEX sequence, induced by hydroxylamine/piperidine. The control trace (top) shows no cleavage peak.

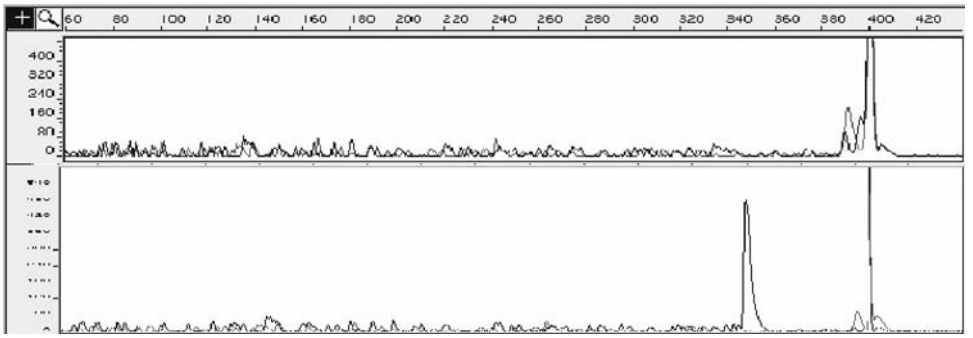


FIGURE 3.4 Detection of T/G mismatch by the solid phase protocol. The heteroduplex DNA trace (bottom) displays a strong cleavage peak at the mismatched T base of the 5′A-FAM sequence, induced by $KMnO_4$/piperidine. The control trace (top) shows no cleavage peak.

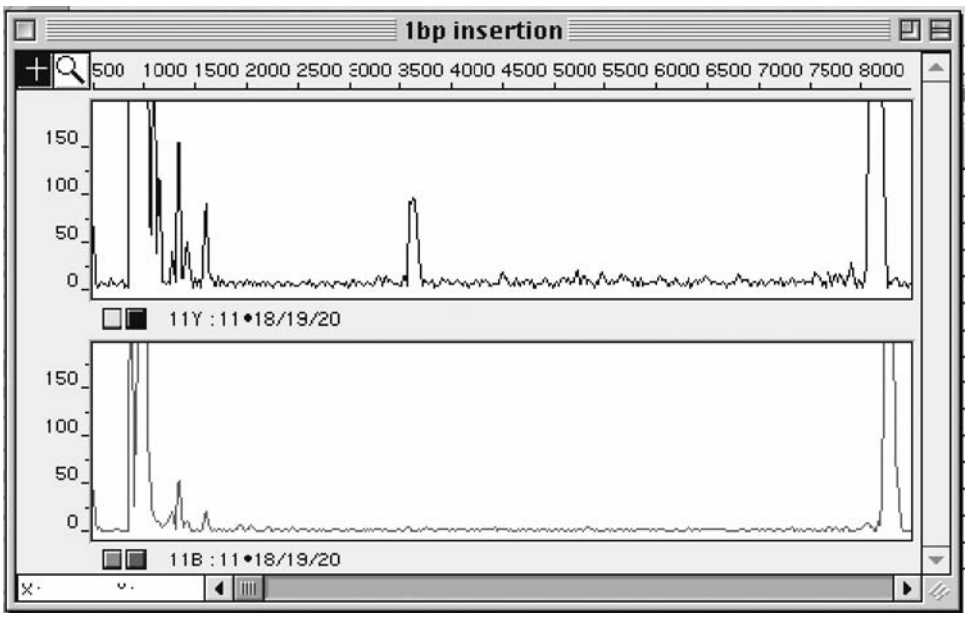


FIGURE 3.5 Detection of a single base insertion (C-base) by the liquid phase protocol. The DNA trace (top) displays a strong cleavage peak at the insertion position (C base) of the 5′A-FAM sequence, induced by hydroxylamine/piperidine. The control trace (bottom) shows no cleavage peak.

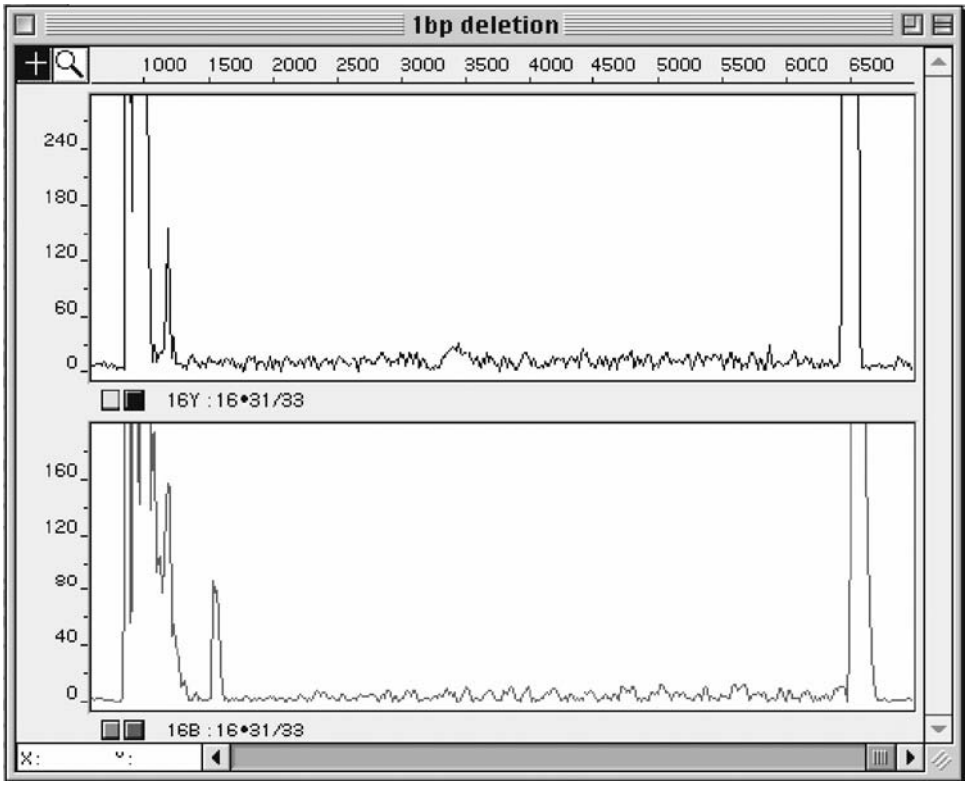


FIGURE 3.6 Detection of a single base deletion (C-base) by the liquid phase protocol. The DNA trace (bottom) displays a strong cleavage peak at the deletion position (C base) of the 3′A-HEX sequence, induced by hydroxylamine/piperidine. The control trace (top) shows no cleavage peak.

synthetic DNA heteroduplexes (mismatched) and homoduplexes (control) (20 nmol each, in two separate reaction vessels) are mixed with 10 μl of $KMnO_4$ (100 nmol in distilled H_2O) in 0.97 ml of 3 M TEAC solution at 25°C. The reaction mixture is carried out in a quartz cuvette and the absorbance is measured at 420 nm after 30 min reaction time. The absorbance value of the mismatched DNA sample is higher than the one obtained from mismatched DNA (control). A typical example is outlined in Table 3.4.

The spectroscopic protocol can be applied for PCR-amplified DNA samples (100–300 bp) to detect mismatches (Tabone *et al*., 2006). The differences in absorbances are measured after 30 min by spectrophotometer or microtiter plate reader. DNA sample concentrations as low as 0.5 to 1.5 μg were found to be suitable for the spectroscopic assays with 0.1 mM $KMnO_4$ (final concentration) in the above TEAC solution.

3.4 ADVANTAGES AND LIMITATIONS

The major advantage of the solid phase protocol is that it is fast and simple in manipulation compared to the liquid phase protocol. The protocol bypasses the washing steps and DNA does not require separation by ethanol precipitation techniques. However, the solid phase protocol is most suitable for short DNA sequences (up to around 500 bp). Adsorption of long DNA on beads is not sufficient due to limit of surface area of the beads. Expensive cost of the silica beads and the washing liquid are also taken into account.

In addition, no false positive and negative results have been reported in the CCM method so far. Usage of 5′-FAM and 3′-HEX fluorescence-labeled primers will offer two chances of a mutation being detected. A typical example is the detection of a C/C mismatch of 547 bp DNA fragment by using hydroxylamine. The mismatched DNA trace (heteroduplex) displays two strong cleavage peaks of the 5′-FAM sequence and 3′-HEX sequence at the mismatched C base (see Fig. 3.7).

Optimization is recommended when starting to use the method. Prolonged incubation leads to multiple cleavage peaks on the background due to overreactions. No cleavage peaks may be the result of short incubations. Controls are needed to assure that the chemicals are active. Adjacent matched bases appear to react with $KMnO_4$/piperidine and the trace may display multiple peaks. These unusual cases have been observed in a recent study, suggesting the strong destacking phenomenon (bubble formation) within the mismatched site enhanced the reactivity (Lambrinakos *et al*., 1999).

AT-rich sequences are also susceptible to multiple modification/cleavage reactions and a peak with a plateau shape may be observed. This was observed on the homoduplex

TABLE 3.4 The spectroscopic protocol.

DNA heteroduplexes	A420 nm (mismatched)	A420 nm (control)[a]	% Increase in oxidation level
Mismatch T-C at the 19th base from 5′ end	0.358	0.300	16%
Mismatch T-T at the 19th base from 5′ end	0.530	0.230	57%
Mismatch T-G at the 19th base from 5′ end	0.326	0.230	30%

[a]*Homoduplex (control) 5′-GGAAGAAGGCATACGGGTGAACTAGGGCAGCGGACAAT-3′ 3′-CCTTCTTCCGTATGCCCACTTGATCCCGTCGCCTGTTA-5′.*

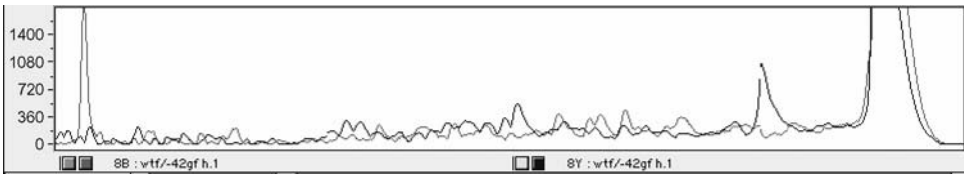


FIGURE 3.7 Detection of C/C mismatch by the solid phase protocol. The heteroduplex DNA trace displays two strong cleavage peaks of the 3-HEX and 5′A-FAM sequences induced by hydroxylamine/piperidine reactions. The control trace is not shown.

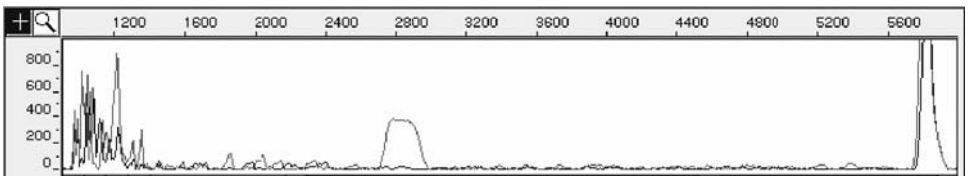


FIGURE 3.8 Multiple cleavage reactions at the AT-rich sequence by $KMnO_4$/piperidine. The homoduplex DNA trace displays a plateau peak due to cleavage reactions at the AT-rich sequence.

500 bp DNA sample containing 18 consecutive A/T matched pairs (see Fig. 3.8). $KMnO_4$ susceptibility to AT-rich sequences was also suggested due to the local destacking nature (curvature) of the repeated A-track region (De Santis *et al*., 1990).

Finally, to improve the quality of the result, PCR products of mutant and wild-type DNA usually are purified prior to any subsequent treatment, and the $KMnO_4$ solution should be freshly made before use because an aging solution turns brown–yellow with the precipitation of MnO2 after a few days. The new CCM version (spectroscopic technique) now becomes an even simpler and inexpensive assay for detection of mutations and polymorphisms as it does not require expensive and toxic chemicals as well as separation steps.

3.5 ENZYMATIC CLEAVAGE OF MISMATCH METHODS

3.5.1 The mutHLS System of Mismatch Detection

The post-replicative mismatch repair enzyme systems were explored as potential tools for mismatch detection (Lu and Hsu, 1992; Lishanski *et al*., 1994; Wagner *et al*., 1995). Because the MutHLS system is tightly coupled to the DNA replication machinery *in vivo*, it is not meant to screen the genome as a whole for the presence of mismatches. The crystal structure further illuminated that MutS bends the DNA only slightly, allowing it to recognize almost all of the mismatches, but not with very high specificity or affinity

(Natrajan *et al*., 2003). As a result, mutation screening based on MutS protein has thus far been unsatisfactory.

3.5.2 DNA N-glycosylase Approaches of Mismatch Detection

The DNA N-glycosylases have evolved nucleotide-binding pockets with very tight fit and nucleotide specificity, including the ability to insert a peptide loop into the base that has been displaced. Unfortunately, nature needs only two mismatches to be recognized by DNA glycosylases. The thymine DNA N-glycosylase recognizes the T/G mismatch that results from the deamination of 5-methylcytosine (Hardeland *et al*., 2001), and the MutY and its homologs recognize the A/G mismatch that occurs from the misincorporation of 8-oxo guanine across from the A residue (Sanchez *et al*., 2003). Thus, methods of mutation detection that use DNA glycosylases do not utilize all the mismatch heteroduplexes that are present during the PCR amplification of two alleles. Moreover, the DNA N-glycosylases do not detect insertion/deletion mutations, or mutations that do not form T/G and A/G mismatches. Because a DNA N-glycosylase only creates an apurinic site but will not break the DNA chain, the detection of DNA truncation at the mismatch site requires either the addition of an apurinic endonuclease or a base treatment step such as piperidine at elevated temperatures, and DNA cleanup steps. Double-stranded DNA breaks are not produced, thus requiring the use of denaturing conditions for the shortened single-stranded products to be analyzed. Successful application of the DNA glycosylase methods have been documented in numerous publications (Zhang *et al*., 2002).

3.5.3 The Resolvase Approach of Mismatch Detection

Another enzymatic approach involves the use of DNA resolvases that recognize the DNA distortion created by mismatches mimicking the DNA recombination intermediate structures. Successful applications of this approach use the T4 endonuclease VII system (Mashal *et al*., 1995; Youil *et al*., 1995, 1996) and the related T7 endonuclease I system (Babon *et al*., 2003). These enzymes cut the mismatch duplexes within a few nucleotides of the mismatch sites, and usually lead to double-stranded breaks in the heteroduplex. The resolvases also lead to non-specific DNA cutting at some unknown DNA sequences and imperfect products of PCR reactions (Norberg *et al*., 2001).

3.5.4 The Endonuclease V Plus Ligation Method

Although endonuclease V itself is not very mismatch specific, its incision within one or two nucleotides of a mismatch base allows a DNA ligase to distinguish between the mismatch product nicks that cannot be ligated and the nicks made at non-mismatch sites that can be repaired by ligation (Huang *et al*., 2002). Some sequences may contain mismatches that are not detectable by endo V or ligatable due to mismatch slippage. In common with other approaches, there is non-specific nicking at AT-rich regions, and the ligation repair reaction is lengthy.

3.5.5 The Plant Mismatch Endonuclease Method

The enzymatic mutation detection approach that currently shows much potential for exploitation in a number of applications is the plant mismatch endonucleases exemplified by the CEL I endonuclease of celery (Oleykowski *et al*., 1998, 1999; Yang *et al*., 2000; Kulinski *et al*., 2000). Recently, the CEL I method of mutation detection was successfully applied to the whole mitochondria genome (Bannwarth *et al*., 2006, 2008), to the screening of epidermal growth factor receptor mutations (Janne *et al*., 2006), to 25 human mutations (Tsuji and Niida, 2007), for variants in *ATM*, *TGFB1*, *XRCC1*, *XRCC3*, *SOD2*, and *hHR21* (Ho *et al*., 2006), for *ATRX* gene mutation (Wada *et al*., 2006), for *TP53* (Poeta *et al*., 2007), for *Sult1a1* (Greber *et al*., 2005), and for TILLING in rice (Till *et al*., 2007) and Drosophila (Winkler *et al*., 2005). These plant nucleases apparently belong to a subgroup in the S1 nuclease family and are induced during plant senescence and remodeling. Mutation detection is extremely simple, applicable on various fragment analysis platforms, whereas the detection of the products of DNA truncation at the mismatch site in either the single-stranded form or the double-stranded forms gives the potential of multiple detection formats. Moreover, the high precision of the CEL I nuclease to cut at the 3′ phosphodiester bond immediately next to the mismatched base allows DNA ligation repair to be used more effectively if desired. The latter was used in the early experiments during the development of the CEL I platform, but is not necessary for most CEL I applications. The high mismatch specificity of CEL I nuclease is believed to come from the enzyme binding to both bases of a base substitution mismatch at the same time.

Current CEL I mutation detection assay is exemplified in Fig. 3.9, in which a DNA heteroduplex contains a mismatch. The ability of CEL I to form a single-stranded DNA nick in short incubations but convert to double-strand DNA truncation mode under conditions of longer incubations or enzyme excess has allowed two powerful assay approaches to be developed.

3.5.5.1 Single-Stranded DNA Truncation Assay

This assay mode is widely used in fragment analysis platforms like the ABI-377 slab gel system, the LiCor infrared

slab gel system, the ABI-3100/3730 capillary DNA sequencers, and the Beckman CEQ8000 infrared DNA sequencer. In this assay, the PCR primers for a given target region, often under 600 bp long, are labeled with one color for the forward primer and a second color for the reverse primer. The color combination used for the 377 system is 6-FAM/TET, for the ABI 3100/3730 is 6-FAM/HEX, and for the Beckman CEQ 8000 is Cy5/Cy 5.5 or D3/D4. For the latter, the Cy5/Cy5.5 combination is easier in DNA synthesis and purification. Purification of the infrared dye primers is not necessary because the presence of reporterless primers enhance the PCR efficiency without diminishing the mutation detection sensitivity. As illustrated in Fig. 3.9, the two strands will be differentially labeled after PCR and heteroduplexes will be formed, either during PCR or thereafter by denaturation and renaturation. The heteroduplex is treated with CEL I, without further purification, for about 5–30 min. The cut DNA can be loaded onto a DNA sequencer/fragment analyzing system without further purification. A purification step enhances the performance of some fragment analysis platforms like the CEQ 8000 and the LiCor sequencer. The CEL I truncation bands of two colors are measured on the fragment analysis system. The sum of the lengths of the bands of the two colors correlated to the same SNP is equal to the full-length PCR primer plus one nucleotide, or more if the insertion involves more than one base. CEL I cuts an insertion at the phosphodiester bond at the 3′ end of the loop, and then shortens the single-stranded region slowly thereafter. This assay is simple, sensitive, and easily automated for high throughput. The two color cuts each originate from a different DNA molecule in the case of the single-strand cut assay, and thus their presence represents independent confirmation of the presence of the mismatch. This assay is used routinely in *BRCA1* and *BRCA2* genetic screening at the Fox Chase Cancer Center, and in the TILLING reverse genetics procedure (Colbert *et al*., 2001) which offers another method of targeted gene knockout in plants, zebra fish, mice, ES cells, and other organisms.

An important point in the single-strand mismatch nicking assay is that the nicked DNA may be a minor population in the case of a weak mismatch substrate like the base substitutions T/T and G/T mismatches. Although the insertion/deletions produce restriction-enzyme-like strong signals with CEL I single-strand nicking assay, many base substitutions require rescaling of the fluorescence intensity axis (see Figs 3.10 and 3.11). Experiments in Figs 3.11 and 3.12 use a 500 bp PCR product from the exon 11.4 of the BRCA1 gene (Oleykowski *et al*., 1998). This human genomic fragment, containing two TZC base substitutions and one GZA base substitution, is a very demanding standard routinely used in the authors' laboratory for comparison of the mutation detection performance of different assay systems and conditions. The CEL I mismatch-specific signals may be small when the PCR product and the primer peaks are expressed in full scale (see Fig. 3.11, insert), but they are still many times higher in signal than the background peaks. Figure 3.11 illustrates that even in the case of three mismatches being present in the same PCR fragment, they are decisively identified in the CEL I mismatch detection assay on the CEQ 8000. The same performance was previously reported for the ABI 377 platform (Kulinski *et al*., 2000), and also routinely obtained for the ABI 3100 capillary sequencing system (see Fig. 3.12).

A PCR product from two alleles form heteroduplexes of two combinations for any given base substitution (e.g. a CZT transition produces both C/A and T/G mismatches and four recognition sites for CEL I). Because CEL I mismatch nuclease can recognize all mismatches and can cut at either strand or both strands of a mismatch, all mismatches have multiple chances to be detected with this system. When used carefully, this system does not produce false positives and false negatives.

Table 3.5 shows the assay conditions used for the CEQ 8000 system (Yeung laboratory) and the conditions routinely used for BRCA1/BRCA2 mutation detection on the ABI 3100 system (Godwin laboratory), respectively. The CEQ 8000 system apparently produces slightly sharper peaks than the ABI 3100, but is less tolerant to salt and primer contaminations. Therefore, an ethanol precipitation step is needed prior to loading of the CEL I reaction products onto the CEQ 8000. On the contrary, the ABI 3100

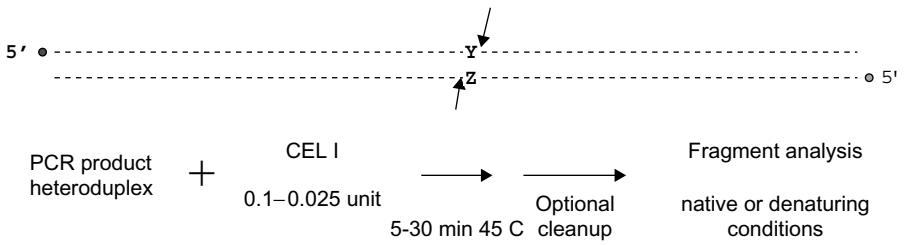


FIGURE 3.9 Schematic of the CEL I mismatch nuclease mutation detection procedure. The top panel illustrates that in a DNA heteroduplex containing a mismatch Y/Z, CEL I makes either one or two incisions in the same DNA molecule denoted by the two arrows. The bottom panel illustrates the simplicity of the CEL I mutation detection method. Adenaturing fragment analysis method may use 5′ termini radioactive labeling or fluorescence two-color labeling followed by resolution of an automated DNA sequencer. The non-denaturing fragment analysis methods may use SYBR-green fluorescence staining and resolution on an agarose gel or PAGE or an Agilent BioAnalyzer lab-on-a-chip.

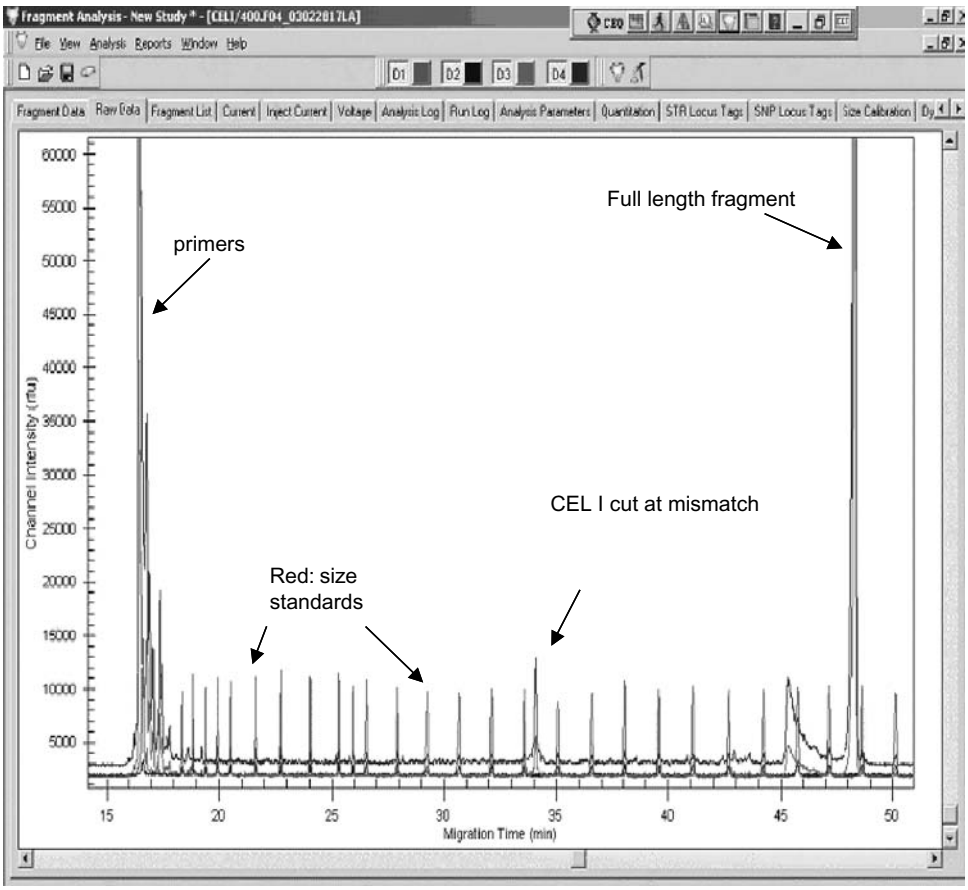


FIGURE 3.10 CEL I mutation detection of a CZT base substitution mutation in a 500 bp PCR product of human BRCA1 gene. A screen shot of the unprocessed data of the CEL I mutation detection product resolved in the infrared capillary sequencer Beckman CEQ 8000 is shown.

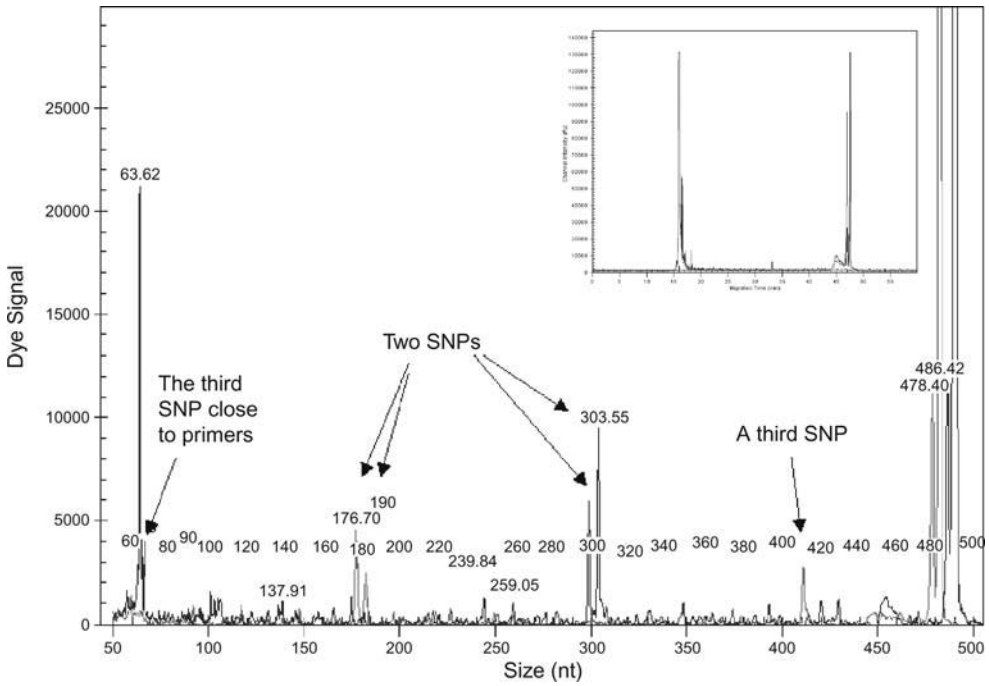


FIGURE 3.11 CEL I mutation detection of three SNPs in the same fragment of BRCA1 exon 11.4. CEQ 8000 analysis close-up of CEL I cut at three SNPs, two T to C, and one G to A base substitutions in the same fragment. The sizes of the standard ladder are shown by numbers in the figure. Incubation was 0.025 units of CEL I for 30 min at 45°C. The insert represents the unprocessed chromatogram in full scale display, illustrating that the CEL I cuts at multiple SNPs are small signals between tall peaks of full length PCR product and the PCR primer peaks.

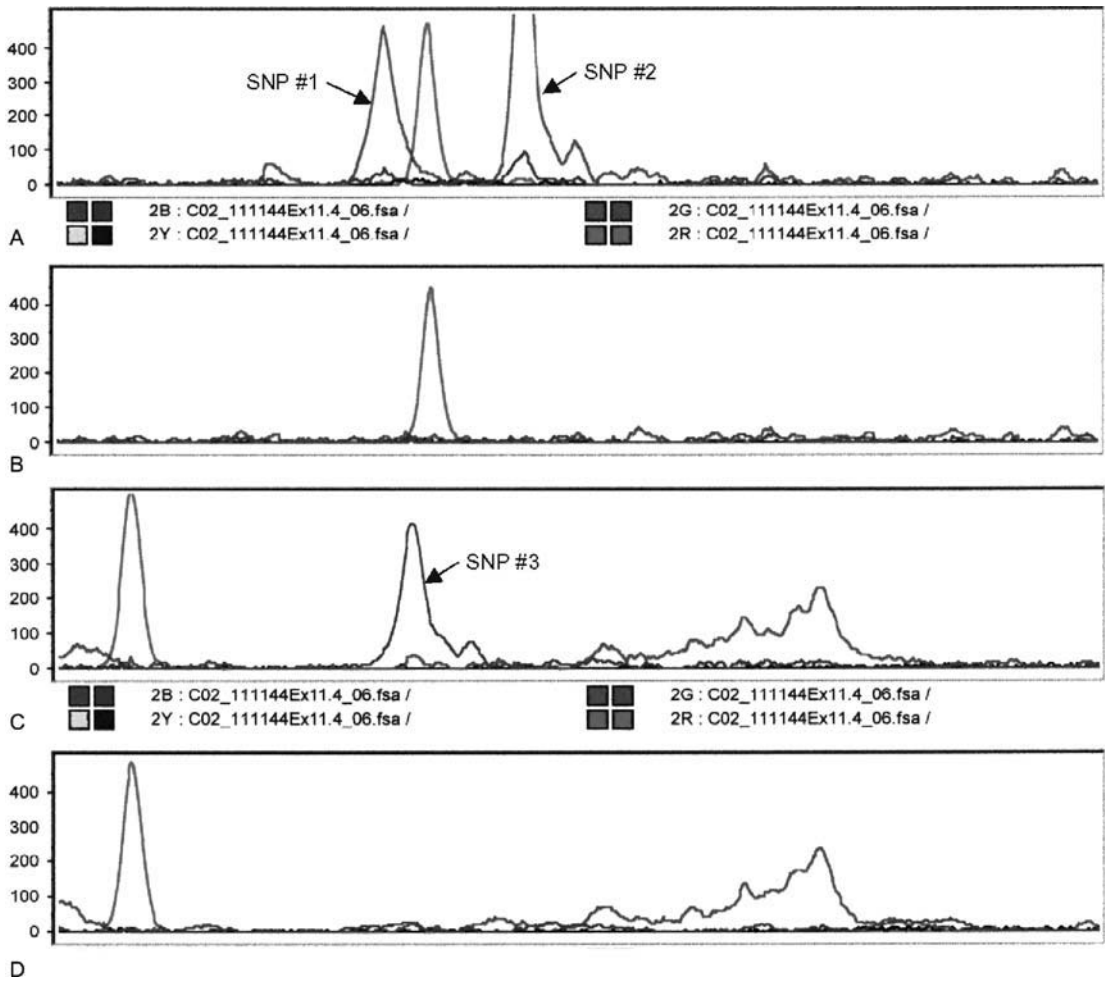


FIGURE 3.12 CEL I mismatch nuclease mutation detection using the Genescan® procedure in the ABI 3100 automated capillary DNA sequencer. The same PCR product fragment containing three SNPs used in the Fig. 3.10 experiment is shown. The forward and reverse primers are labeled at the 5′ termini with 6-FAM and HEX, respectively. The assay conditions are specified in Table 3.2 (under Godwin laboratory procedures). **A.** The two SNPs detected by the forward primer. **B.** The size marker and the low background noise. **C.** The third SNP detected by the reverse primer. **D.** Size marker and background noise corresponding to panel C.

apparently can load the CEL I reaction product with no sample purification. Gel filtration desalting is not suitable for the CEQ 8000 system in which the primers and PCR products contain very hydrophobic infrared cyanine dyes that lead to binding to the Sephadex used in the gel filtration columns. Addition of acetonitrile to the buffers diminishes the sample loss, but also decreases the gel filtration performance.

3.5.5.2 Double-Stranded DNA Truncation Assay

The above single-stranded end-labeled primer CEL I assay uses partial mismatch digestion to minimize the fluorescence label loss to the 5′ exonuclease activity of CEL I. This activity appears to cut within a few nucleotides of the 5′ termini slowly, and limits the accumulation of the mismatch incision products with the fluorescent reporter still attached. This 5′ endonuclease activity is minimized in the presence of 0.5 units of Taq DNA polymerase in the CEL I incubation, or the presence of a cyanine dye reporter at the 5′ end of the PCR product. This partial digestion, on the positive side, allows multiple mutations to be detected in the same PCR fragment, even when they are three nucleotides apart. The use of initial kinetics in the mismatch digestion also leads to high signal-to-noise ratio, and a broad window of enzyme quantity or digestion time of at least 20-fold and still producing useful mutation detection data. This 5′ exonuclease effect can be avoided when a DNA staining procedure is used instead of the 5′ termini label. For example, using SYBR-green staining of the double-stranded DNA fragments, the 5′ exonuclease-processed mismatch incision fragments can still be visualized to produce much stronger signal-to-noise ratio in the mutation detection assay than the single-strand assay method (see Fig. 3.13). This is one of the principles developed in the Giraff method of CEL I genomic scanning for bacterial mutations (Sokurenko *et al*., 2001). In Fig. 3.13, the CEL I-incised double-stranded DNA is analyzed in a

TABLE 3.5 CEL I mismatch nuclease single-strand mismatch nicking protocols for capillary DNA sequencers.

Capillary sequencer CEL protocols	Yeung laboratory, Beckman CEQ 8000	Godwin laboratory, ABI 3100
PCR reaction	20 μL	20 μL
Human genomic DNA	60–75 ng	75–150 ng
Reaction buffer, 10×	$MgCl_2$, 2 mM final	$MgCl_2$
DMSO	5% final	6.7% final
dNTP	200 μM	80 μM
Primers, each	1.5 pmol	4 pmol
DNA polymerase	1U AmpliTaq Gold	1U AmpliTaq
PCR cycles	94°C for 4 min	94°C for 4 min
	97°C for 1 min	20 cycles of:
Exons that are prone to deletion are checked with agarose gels		94°C for 5 sec
		65°C for 1 min, touchdown at −0.5°C/cycle
		72°C for 1 min
	30 cycles of:	32 cycles of:
	94°C for 10 sec	94°C for 5 sec
	55°C for 20 sec	55°C for 1 min
	72°C for 45 sec	72°C for 1 min
	72°C for 4 min, then 4°C	72°C for 5 min, then 4°C
Denaturation/renaturation step	94°C 1 min then cooled to 4°C over 30 min	None
CEL I reaction	20 μL rxn	10 μL rxn
PCR reaction product	5 μL	5 μL
Water	12 μL	3.1 μL
10× CEL I buffer	2 μL	0.9 μL
CEL I enzyme	0.1 units	0.025 units
Incubation	45°C for 7 min	45°C for 1 hr
Stopping CEL I reaction	5 μL of (0.1 mM EDTA, 1.2 M Na-acetate, 15 v/v glycogen, 1.7% v/v of NF co-precipitant) to the 20 μL CEL I rxn	Add 3 μL 5 mM o-phenantholine or just add formamide tracking dye. No ethanol ppt or centrifugation
	60 μL of 95% ethanol, to a final volume of 85 μL. Centrifuge for 15 min 4°C at 21,000 × g	
	Wash pellet 2x with 70% ethanol, dry in speedvac	
Preparation for capillary	Resuspend pellet in 40 μL of sample loading solution (SLS)	5 μL of stopped enzyme rxn is added to 15 μL of marker mix
Sequencer loading	Add 1.5 μL of suspension to 28.5 mL of SLS. Add 10 μL marker mix	
	Load all 40 μL to the CEQ8000	Load all 20 μL to the ABI 3100
	Sample well ~0.25 μL of the original PCR reaction	Sample well approximately 1.92 μL of the reaction original PCR

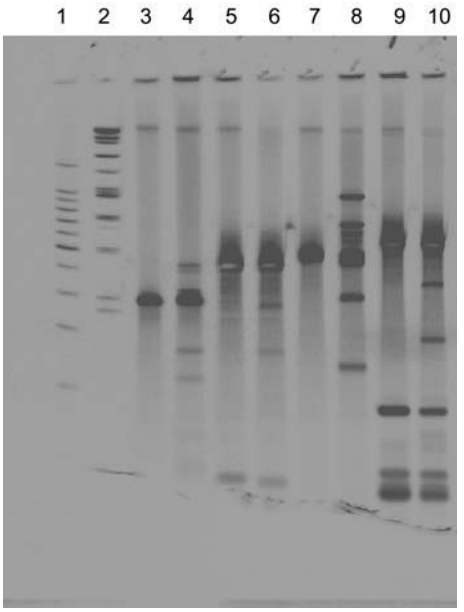


FIGURE 3.13 CEL I mutation detection double-strand DNA truncation assay of SNPs in the BRCA1 gene. Each PCR product was incubated with 0.025 units of CEL I for 17 hours at 37°C and resolved on 8% native PAGE in 1X TBE buffer. DNA bands were visualized with SYBR-green fluorescence staining. Lanes 1, 100 base marker; 2, 1 kb marker; 3, Exon 2 control; 4, Exon 2 AG deletion; 5, Exon 20 control; 6, Exon 20 C insertion; 7, Exon 11.9 control; 8, Exon 11.9 4 bp deletion; 9, Exon 11.4 control; 10, Exon 1 1.4 G A base substitution.

native PAGE visualized with SYBR-green fluorescence staining. A second benefit of this native gel approach is that any single-stranded non-specific nicking by CEL I in the double-stranded DNA does not contribute to background unless it leads to double-stranded truncation. Double-stranded truncation at a mismatch by CEL I can be obtained by using the same amount of CEL I as in the single-strand nicking assay but extending the incubation for 16 hours. The use of a higher concentration of CEL I for a shorter incubation produces much higher signal-to-noise ratio. The new CEL nuclease SURVEYOR™ agarose gel kit supplied by Transgenomic Inc. has been further optimized for this assay and can detect base substitutions in long PCR products (i.e. 4 kb in length). For shorter fragments, a pool depth of one mutant in 40 normal alleles can be obtained. For biological DNA, the Giraff approach has demonstrated mismatch detection in heteroduplexes longer than 10 kb (Sokurenko *et al*., 2001).

The full exploitation of the agarose/acrylamide gel assay format is still in progress. For example, the use of the Agilent BioAnalyzer lab-on-a-chip system for this assay would produce automation and very sensitive detection of PCR products up to 12 kb, provides options for background subtraction, and takes less than an hour. Because the proofreading DNA polymerase Optimase is used in the SURVEYOR™ kit, the PCR products can be longer, while it allows trans-intronic PCR reactions to be performed for inbred strains. Besides leading to higher screening efficiency, longer PCR products also allow the positioning of the query region near the center of the PCR product, thereby making this assay easy and efficient even when an ordinary inexpensive agarose gel apparatus is used. This mutation detection format is truly fast, convenient, and amenable to automation because no harsh chemicals are used, all enzyme steps can be done in the same tube, it takes little over an hour, and processing of the CEL nuclease incised DNA is unnecessary prior to loading onto the native condition fragment analysis systems.

3.5.5.3 Endonucleolytic Mutation Analysis by Internal Labeling (EMAIL)

The 5′ exo/endonuclease activity of the CEL family of nucleases is a major limitation on the ease of use of the CEL nuclease mutation detection method when 5′ reporter labeling approaches are used. Recently, internally labeling by polymerase to incorporate a small amount of fluorescently labeled dNTP during the PCR reaction was shown to lead to a DNA heteroduplex product in which the fluorescence reporter is protected by the CEL nuclease and thus lead to higher assay sensitivity (Cross *et al*., 2008). The mutation detection assay was conveniently performed on a 3730 DNA Analyzer (Applied Biosystems).

3.5.5.4 The s-RT-MELT Method Combining CEL Nuclease Enzymatic Selection with Real-Time DNA-Melting Curve Detection

Most CEL nuclease mutation methods are gel-based. Recently, a method that does not involve the use of gel electrophoresis has been developed (Li *et al*., 2007). This method, s-RT-MELT, scores for the appearance of CEL nuclease incision in heteroduplex DNA by the effect of the incision on the melting curve of the DNA duplex by using a real-time PCR instrument to perform a melting curve analysis. Besides the possibility of high throughput by using the microtiter plate formats, the assay employs ligation-mediated PCR reaction to amplify the CEL nuclease cut fragment, with PCR primer designs that enrich for the CEL nuclease-incised fragments. This cleaver assay format is compatible with robotics and thus has the potential to greatly enhance the throughput of enzymatic mutation detection.

3.6 CONCLUSIONS

There has been dramatic improvement in the methods for screening for unknown mutations. These improvements have already translated into exciting applications in numerous research centers. For both the chemical and the enzymatic methods of mutation detection, PCR artefacts

and DNA polymerization errors often limit the sensitivity these assays can deliver. One approach that potentially can remove these errors is the use of hairpin primers in PCR that effectively link each pair of semiconservative DNA replication strands as one duplex (Kaur and Makrigiorgos, 2003). Any duplex that contains a PCR error can be removed with some mismatch-specific method to be determined so that error-free PCR products can be collected and used in high-sensitivity mutation detection. It is expected that more improvements will be forthcoming so that effective mutational screening will become a routine research and diagnostic tool for modern genetics.

REFERENCES

Babon, J.J., McKenzie, M., and Cotton, R.G. (2003). The use of resolvases T4 endonuclease VII and T7 endonuclease I in mutation detection. *Mol. Biotechnol.* 23, 73–81.

Bannwarth, S., Procaccio, V., and Paquis-Flucklinger, V. (2006). Rapid identification of unknown heteroplasmic mutations across the entire human mitochondrial genome with mismatch-specific Surveyor Nuclease. *Nat. Protoc.* 1, 2037–2047.

Bannwarth, S., Procaccio, V., Rouzier, C., Fragaki, K., Poole, J., Chabrol, B., Desnuelle, C., Pouget, J., Azulay, J.P., Attarian, S., Pellissier, J.F., Gargus, J.J., Abdenur, J.E., Mozaffar, T., Calvas, P., Labauge, P., Pages, M., Wallace, D.C., Lambert, J.C., and Paquis-Flucklinger, V. (2008). Rapid identification of mitochondrial DNA (mtDNA) mutations in neuromuscular disorders by using surveyor strategy. *Mitochondrion* 8, 136–145.

Block, W. (1999). Improved detection of mutations in nucleic acids by chemical cleavage. Patent Appl. No. PCT/ US98/16385.

Bui, C.T., and Cotton, R.G.H. (2002). Comparative study of permanganate oxidation reactions of nucleotide bases by spectroscopy. *Bioorg. Chem.* 30, 133–137.

Bui, C.T., Rees, K., Lambrinakos, A., Bedir, A., and Cotton, R.G.H. (2002). Site selective reactions of imperfectly matched DNA with small chemical molecules: applications in mutation detection. *Bioorg. Chem.* 30, 216–232.

Bui, C.T., Lambrinakos, A., Babon, J.J., and Cotton, R.G.H. (2003a). Chemical cleavage reactions of DNA on solid support: application in mutation detection. *BMC Chem. Biol.* 3, 1.

Bui, C.T., Lambrinakos, A., and Cotton, R.G.H. (2003b). Spectroscopic study in permanganate oxidation reactions of oligonucleotides containing single base mismatches. *Biopolymers* 70, 628–636.

Bui, C.T., Sam, L.A., and Cotton, R.G.H. (2004). UV-visible spectral identification of the solution- phase and solid-phase permanganate oxidation reactions of thymine acetic acid. *Bioorg. Med. Chem. Lett.* 14, 1313–1315.

Burdon, M.G., and Lees, J.H. (1985). Double-strand cleavage at a 2-base deletion mismatch in a DNA heteroduplex by Nuclease-S1. *Bioscience Reports* 5, 627–632.

Colbert, T.G., Till, B., Tompa, R., Reynolds, S.H., Steine, M., Yeung, A.T., McCallum, C.M., Comai, L., and Henikoff, S. (2001). High-throughput screening for induced point mutations. *Plant Physiol.* 126, 480–484.

Cotton, R.G.H., Rodrigues, H.R., and Campbell, R.D. (1988). Reactivity of cytosine and thymine in single base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl. Acad. Sci. USA,* 85, 4397–4401.

Cotton, R.G.H. (1999). Detection of mutations in DNA and RNA by chemical cleavage. In *The Nucleic Acid Protocols Handbook Methods in Molecular Biology Series* (Rapley, R., ed.), pp. 685–693. John Wiley & Sons, Inc.

Cotton, R.G.H., and Bray, P.J. (2001). Using CCM and DHPLC to detect mutations in the glucocorticoid receptor in atherosclerosis: a comparison. *J. Biochem. Biophys. Methods,* 47, 91–100.

Cross, M.J., Waters, D.L., Lee, L.S., and Henry, R.J. (2008). Endonucleolytic mutation analysis by internal labeling (EMAIL). *Electrophoresis*. in press.

De Santis, P., Palleschi, A., Savino, M., and Scipioni, A. (1990). Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of DNA curvature. *Biochemistry* 29, 9269–9273.

Gogos, J.A., Karayiorgou, M., Aburatani, H., and Kafatos, F.C. (1990). Detection of single base mismatches of thymine and cytosine residues by potassium permanganate and hydroxylamine in the presence of tetralkylammonium salts. *Nucleic Acids Res.* 18, 6807–6812.

Greber, B., Tandara, H., Lehrach, H., and Himmelbauer, H. (2005). Comparison of PCR-based mutation detection methods and application for identification of mouse Sult1a1 mutant embryonic stem cell clones using pooled templates. *Hum. Mutat.* 25, 483–490.

Hardeland, U., Bentele, M., Lettieri, T., Steinacher, R., Jiricny, J., and Schar, P. (2001). Thymine DNA glycosylase. *Prog. Nucleic Acid Res. Mol. Biol.* 68, 235–253.

Haris, I.I., Green, P.M., Bentley, D.R., and Giannelli, F. (1994). Mutation detection by fluorescent chemical cleavage: application to hemophilia B. *PCR Meth. Appl.* 3, 268–271.

Hartge, P., Struewing, J.P., Wacholder, S., Brody, L.C., and Tucher, M.A. (1999). The prevalence of common BRCA1 and BRCA2 mutations among Ashkenazi Jews. *Am. J. Hum. Genet.* 64, 963–970.

Ho, A.Y., Atencio, D.P., Peters, S., Stock, R.G., Formenti, S.C., Cesaretti, J.A., Green, S., Haffty, B., Drumea, K., Leitzin, L., Kuten, A., Azria, D., Ozsahin, M., Overgaard, J., Andreassen, C.N., Trop, C.S., Park, J., and Rosenstein, B.S. (2006). Genetic predictors of adverse radiotherapy effects: the Gene-PARE project. *Int. J. Radiat. Oncol. Biol. Phys.* 65, 646–655.

Huang, J., Kirk, B., Favis, R., Soussi, T., Paty, P., Cao, W., and Barany, F. (2002). An endonuclease/ligase based mutation scanning method especially suited for analysis of neoplastic tissue. *Oncogene* 21, 1909–1921.

Janne, P.A., Borras, A.M., Kuang, Y., Rogers, A.M., Joshi, V.A., Liyanage, H., Lindeman, N., Lee, J.C., Halmos, B., Maher, E.A., Distel, R.J., Meyerson, M., and Johnson, B.E. (2006). A rapid and sensitive enzymatic method for epidermal growth factor receptor mutation screening. *Clin. Cancer Res.* 12, 751–758.

Kao, J.Y., Goljer, I., Phan, T.A., and Bolton, P.H. (1993). Characterization of the effects of a thymine glycol residue on the structure, dynamics and stability of duplex DNA by NMR. *J. Biol. Chem.* 268, 17787–17793.

Kaur, M., and Makrigiorgos, G.M. (2003). Novel amplification of DNA in a hairpin structure: towards a radical elimination of PCR errors from amplified DNA. *Nucleic Acids Res.* 31, e26.

Kennard, O. (1988). Structural studies of base pair mismatches and their relevance to theories of mismatches formation and repair. In *Structure and Expression, Vol. 2 in DNA and its Drugs Complexes* (Sarma, R. H., and Sarma, M.H., eds), pp. 1–25. Adenine Press.

Kulinski, J., Besack, D., Oleykowski, C.A., Godwin, A.K., and Yeung, A.T. (2000). The CEL I Enzymatic Mutation Detection Assay. *Biotechniques* 29, 44–48.

Lambrinakos, A., Humphrey, K.E., Babon, J.J., Ellis, T.P., and Cotton, R.G.H. (1999). Reactivity of potassium permanganate and tetraethylammonium chloride with mismatched bases and a simple mutation detection protocol. *Nucleic Acids Res.* 27, 1866–1874.

Li, J., Berbeco, R., Distel, R.J., Janne, P.A., Wang, L., and Makrigiorgos, G.M. (2007). s-RT-MELT for rapid mutation scanning using enzymatic selection and real time DNA-melting: new potential for multiplex genetic analysis. *Nucleic Acids Res.* 35, e84.

Lishanski, A., Ostrander, E.A., and Rine, J. (1994). Mutation detection by mismatch binding-protein, MutS, in amplified DNA – application to the cystic-fibrosis gene. *Proc. Natl. Acad. Sci. USA*. 91, 2674–2678.

Lu, A.L., and Hsu, I.C. (1992). Detection of single DNA-base mutations with mismatch repair enzymes. *Genomics* 14, 249–255.

Mashal, R.D., Koontz, J., and Sklar, J. (1995). Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat. Genet.* 9, 177–183.

Mullis, K.B., and Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 155, 335–350.

Natrajan, G., Lamers, M.H., Enzlin, J.H., Winterwerp, H.H., Perrakis, A., and Sixma, T.K. (2003). Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. *Nucleic Acids Res.* 31, 4814–4821.

Neuhausen, S.L., and Ostrander, E.A. (1997). Mutation testing of early-onset breast cancer genes BRCA1 and BRCA2. *Genet. Test.* 1, 75–83.

Norberg, T., Klaar, S., Lindqvist, L., Lindahl, T., Ahlgren, J., and Bergh, J. (2001). Enzymatic mutation detection method evaluated for detection of p53 mutations in cDNA from breast cancers. *Clin. Chem.* 47, 821–828.

Oleykowski, C.A., Bronson Mullins, C.R., Godwin, A.K., and Yeung, A.T. (1998). Mutation detection using a novel plant endonuclease. *Nucleic Acids Res.* 26, 4597–4602.

Oleykowski, C.A., Bronson Mullins, C.R., Chang, D.W., and Yeung, A.T. (1999). Incision at nucleotide insertions/deletions and basepair mismatches by the SP Nuclease of spinach. *Biochemistry* 38, 2200–2205.

Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Dallas, J., Itakura, K., and Breslauer, K.J. (1982). Structure, dynamics, and energetics of deoxyguanosine. Thymidine wobble base pair formation in the self-complementary d(CGTGAATTCGCG) duplex in solution. *Biochemistry* 21, 437–444.

Poeta, M.L., Manola, J., Goldwasser, M.A., Forastiere, A., Benoit, N., Califano, J.A., Ridge, J.A., Goodwin, J., Kenady, D., Saunders, J., Westra, W., Sidransky, D., and Koch, W.M. (2007). TP53 mutations and survival in squamous-cell carcinoma of the head and neck. *N. Engl. J. Med.* 357, 2552–2561.

Ramus, S.J., and Cotton, R.G.H. (1996). Single-tube chemical cleavage of mismatch: successive treatment with hydroxylamine and osmium tetroxide. *Biotechniques* 21, 216–220.

Ren, J. (2001). Chemical mismatch cleavage analysis by capillary electrophoresis with laser-induced fluorescence detection. In *Methods in Molecular Biology* (Mitchelson, K.R., and Cheng, J., eds), Vol. 163, pp. 231–239. Humana Press Inc, Totowa, NJ.

Roberts, R.J., and Cheng, X. (1998). Base flipping. *Annu. Rev. Biochem.* 67, 181–198.

Roberts, E., Deeble, V.J., Woods, C.G., and Taylor, G.R. (1997). Potassium permanganate and tetraethylammonium chloride are a safe and effective substitute for osmium tetroxide in solidphase fluorescent chemical cleavage of mismatch. *Nucleic Acids Res.* 25, 3377–3378.

Rowley, G., Saad, S., Giannelli, F., and Green, P.M. (1995). Ultra rapid mutation detection by multiplex, solid-phase chemical cleavage. *Genomics* 30, 574–582.

Saleeba, J.A., and Cotton, R.G.H. (1991). ^{35}S-labelled probes improve detection of mismatched base pairs by chemical cleavage. *Nucleic Acids Res.* 19, 127–128.

Sanchez, A.M., Volk, D.E., Gorenstein, D.G., and Lloyd, R.S. (2003). Initiation of repair of A/G mismatches is modulated by sequence context. *DNA Repair* 2, 863–878.

Sokurenko, E.V., Tchesnokova, V., Yeung, A.T., Oleykowshi, C.A., Trintchina, E., Hughes, K.T., Rashid, R.A., Brint, J.M., Moseley, S.L., and Lory, S. (2001). Detection of simple mutations and polymorphisms in large genomic regions. *Nucleic Acids Res.* 29, e111.

Tabone, T., Sallmann, G., Webb, E., and Cotton, R.G.H. (2006). Detection of 100% of mutations in 124 individuals using a standard UV/Vis microplate reader: a novel concept for mutation scanning. *Nucleic Acids Res.* 34, e45.

Till, B.J., Cooper, J., Tai, T.H., Colowit, P., Greene, E.A., Henikoff, S., and Comai, L. (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC Plant. Biol.* 7, 19.

Tsuji, T., and Niida, Y. (2008). Development of a simple and highly sensitive mutation screening system by enzyme mismatch cleavage with optimized conditions for standard laboratories. *Electrophoresis*. in press.

Verpy, E., Biasotto, M., Meo, T., and Tosi, M. (1994). Efficient detection of point mutations on color-coded strands of target DNA. *Proc. Natl. Acad. Sci. USA,* 91, 1873–1877.

Wada, T., Fukushima, Y., and Saitoh, S. (2006). A new detection method for ATRX gene mutations using a mismatch-specific endonuclease. *Am. J. Med. Genet.* 140, 1519–1523.

Wagner, R., Debbie, P., and Radman, M. (1995). Mutation detection using immobilized mismatch binding-protein (MutS). *Nucleic Acids Res.* 23, 2948–3944.

Warner, E., Foulkes, W., Goodwin, P., Meschino, W., Blondal, J., Paterson, C., Ozcelik, H., Gross, P., Allingham-Hawkins, D., Hamel, N., Di Prospero, L., Contiga, V., Serruya, C., Klein, M., Moslehi, R., Honeyford, J., Liede, A., Glendon, G., Brunet, J.S., and Narod, S. (1999). Prevalence and penetrance of BRCA1 and BRCA2 gene mutations in unselected Ashkenazi Jewish women with breast cancer. *J. Natl. Cancer Inst.* 91, 1241–1247.

Winkler, S., Schwabedissen, A., Backasch, D., Bökel, C., Seidel, C., Bönisch, S., Fürthauer, M., Kuhrs, A., Cobreros, L., Brand, M., and González-Gaitán, M. (2005). Target-selected mutant screen by TILLING in Drosophila. *Genome Res.* 15, 718–723.

Yang, B., Wen, X., Oleykowski, C.A., Kodali, N.A., Miller, C.G., Kulinski, J., Besack, D., Yeung, J.A., Kowalski, D., and Yeung, A.T. (2000). Purification, cloning and characterization of the CEL I nuclease. *Biochemistry* 39, 3533–3541.

Youil, R., Kemper, B.W., and Cotton, R.H. (1995). Screening for mutations by enzyme mismatch cleavage with T4 endonuclease-7. *Proc. Natl. Acad. Sci. USA,* 92, 87–91.

Youil, R., Kemper, B., and Cotton, R. (1996). Detection of 81 of 81 known mouse b-globin promoter mutations with T4 endonuclease VII – the EMC method. *Genomics* 32, 431–435.

Zhang, Y., Kaur, M., Price, B.D., Tetradis, S., and Makrigiorgos, G.M. (2002). An amplification and ligation-based method to scan for unknown mutations in DNA. *Hum. Mutat.* 20, 139–147.

Chapter 4

Mutation Detection by Single Strand Conformation Polymorphism and Heteroduplex Analysis

Panayiotis G. Menounos[1] and George P. Patrinos[2,3]

[1]Nursing Military Academy, Laboratory of Research, Athens, Greece;

[2]Department of Pharmacy, School of Health Sciences, University of Patras, Patras, Greece;

[3]Erasmus University Medical Center, Faculty of Medicine, and Health Sciences, Department of Bioinformatics, Rotterdam, The Netherlands

4.1 INTRODUCTION

Single strand conformation polymorphism (SSCP) and heteroduplex analysis (HDA) are two of the most popular electrophoresis-based mutation detection methods. Coupled to DNA amplification of the sequence to be analyzed, these techniques have become the methods of choice for a number of molecular diagnostic laboratories. This can be explained mainly by the numerous advantages, namely their technical simplicity and relatively high specificity for the detection of sequence variations, the low operation costs, and the potential for automation for high-throughput mutation analysis. If fluorescently labeled primers are employed during DNA amplification, SSCP analysis can be also performed in gel- or capillary electrophoresis-based automated sequencers (F-SSCP, CE-SSCP), hence allowing for precise, reproducible and high-throughput analysis of the genomic variation. There are several factors that influence sensitivity, and therefore need to be taken into account in order to obtain reproducible results as well as to maximize the sensitivity of mutation detection. In the following pages, the theory and practice of both SSCP and HDA will be discussed. In particular, emphasis will be given to the principle, the parameters influencing the sensitivity and reproducibility of the results, the available detection schemes, and the limitations of both techniques. Finally, a number of applications for screening genomic loci in order to investigate the underlying molecular heterogeneity are also discussed.

4.2 PRINCIPLES OF SINGLE STRAND CONFORMATION POLYMORPHISM ANALYSIS

The analysis of single strand conformation polymorphism (SSCP) has been established by Orita and colleagues (1989) as a simple, efficient, and reliable method for the detection of sequence alterations in genomic loci. Based on polymerase chain reaction (PCR), SSCP was developed soon after the introduction of PCR technology, and relied on the fact that relatively short single-stranded DNA fragments can migrate in a non-denaturing gel as a function not only of their size but also of their sequence. In other words, following amplification of any given DNA sequence, the amplified DNA fragments are subjected to denaturation with either heat or chemical agents, such as formamide. Subsequently, the denatured DNA fragments are electrophoresed through a native (non-denaturing) polyacrylamide gel. During electrophoresis, single-stranded DNA fragments adopt a specific three-dimensional shape according to their nucleotide sequence, and hence exhibit a unique conformation. Therefore, their electrophoretic mobility is dependent upon the previously mentioned three-dimensional shape (see Fig. 4.1). Based on these principles, it is well understood that even a single base difference between a DNA fragment being tested and its wild-type counterpart is sufficient to adopt a different conformation and thus to migrate at a different position during electrophoresis (see Fig. 4.2).

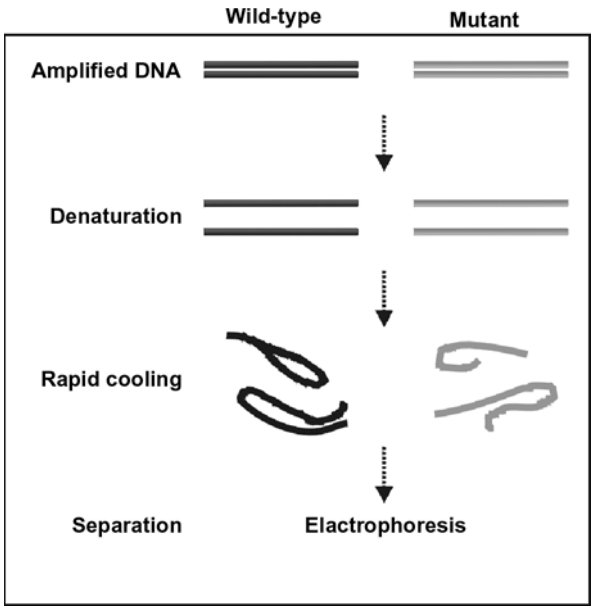


FIGURE 4.1 Schematic representation of the SSCP principle. Amplified DNA for both the wild-type and mutant alleles is subjected to denaturation and then to immediate cooling, where the denatured single-stranded alleles adopt a specific conformation. As the single-stranded wild-type and mutant alleles have different conformations, they can be distinguished when electrophoresed through a native polyacrylamide gel.

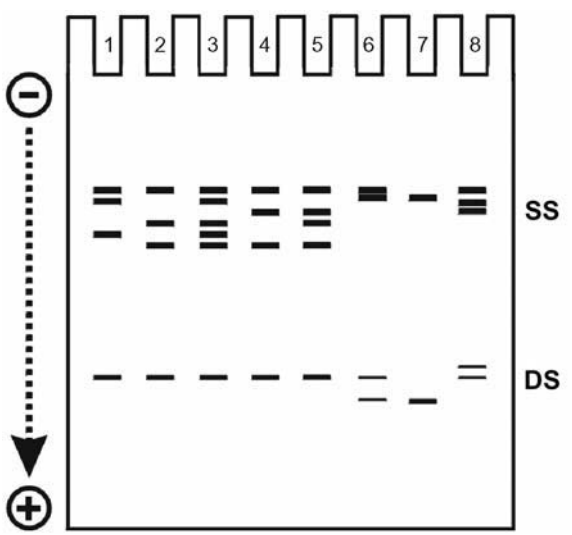


FIGURE 4.2 Schematic drawing of a typical result from an SSCP analysis. After staining, the gel can be divided into two parts: the part with the bands corresponding to the single-stranded alleles (SS) and the other part with the double-stranded bands (DS), as complete denaturation is sometimes not feasible. In this example, lane 1 corresponds to the wild-type pattern for any given locus analyzed. If lane 2 corresponds to a homozygous case for mutation A, then lane 3 corresponds to a heterozygous case for the same mutation, and lanes 4 and 5 correspond to a homozygous case for mutation B and a compound heterozygous case for mutations A and B, respectively. Lanes 6 to 8 are characteristic examples of the electrophoretic pattern of small deletions or insertions, which can also be distinguished at the DS part of the gel (see also Fig. 4.5b). Lanes 6 and 7, heterozygous and homozygous cases for a small deletion, respectively. Lane 8, heterozygous case for a small insertion.

SSCP analysis is extremely advantageous for fast mutation screening of known loci because it is easy to use and can preclude the use of radioactive substances for detection (see also sections 4.3 and 4.4). On the other hand, SSCP is not considered the method of choice for the analysis of unknown sequences, as there is no theoretical background established so far that could enable one to predict the exact electrophoretic mobility of a given DNA fragment according to its sequence (as is the case with denaturing gradient gel electrophoresis (DGGE; see also Chapter 6)). It is noteworthy that SSCP has already enjoyed an enormous success in mutation screening and it is still the method of choice in many molecular genetic laboratories.

4.3 FLUORESCENT SINGLE STRAND CONFORMATION POLYMORPHISM ANALYSIS

Fluorescent SSCP (F-SSCP) is a non-radioactive high-resolution PCR-SSCP method, in which fluorescently labeled PCR products are electrophoresed and detected by an automated DNA sequencer (Makino *et al.*, 1992). A procedure for fluorescence labeling of the PCR product and detection of bands of double-stranded DNA in polyacrylamide gel illuminated with UV-light has been described previously (Chehab and Kan, 1989). However, the sensitivity of this method was not high enough for SSCP analysis. Therefore, coupling SSCP analysis with an automated DNA sequencer enables highly sensitive fragment detection. Use of an automated DNA sequencer also permits strict control of any desired temperature, a fundamental requirement of SSCP analysis to obtain reproducible results. This way, fragment separation patterns can be highly reproducible at any ambient condition. In certain cases, variations in the mobility patterns have been reported (Makino *et al.*, 1992). A solution to this problem is normalization between lanes by including an internal standard fragment labeled with a different fluorescent dye in each lane (Ellison *et al.*, 1993; Iwahana *et al.*, 1994). Co-electrophoresis of an internal size standard DNA labeled with a dye different from that of the sample DNA allows the relative fragment mobility of sample DNAs to be reproducibly determined using mobility of the internal standard DNA as a reference.

F-SSCP results can be also quantitatively interpreted. This is feasible since the bands are detected as peaks in the fluorogram and their heights are proportional to the intensity of the fluorescence in a wide dynamic range (Fig. 4.3). The direct entry of data into a computer allows objective interpretation of the results, while data from multiple samples can be processed for further analysis, using specialized analysis software with a reduced possibility of error. These features can be particularly important for pedigree analysis and linkage studies of polymorphic DNA markers detected by SSCP analysis.

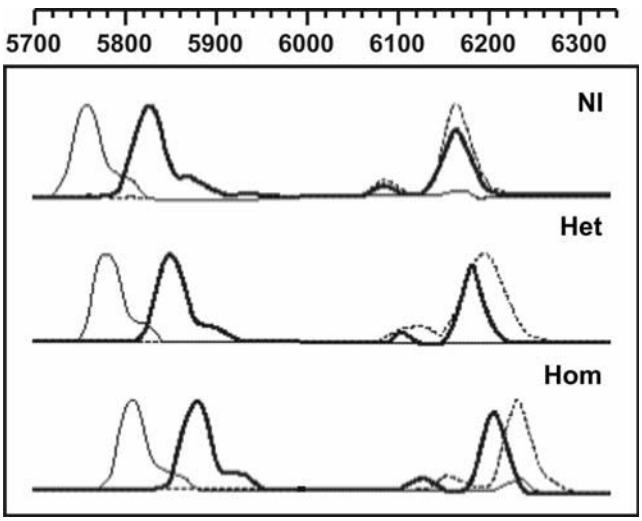


FIGURE 4.3 Principle of F-SSCP analysis. Wild-type DNA is represented with the thick line and run in each lane as a control, along with the normal (Nl), heterozygous (Het) and homozygous (Hom) samples (thin line). Primers tagged with different fluorescent dyes were employed.

The first generation F-SSCP was employing gel-based DNA sequencers. The Pharmacia ALF DNA sequencer was the instrument of choice as, at that time, it was the only commercially available fluorescence-based DNA sequencer equipped with an electrophoresis apparatus that allowed accurately controlled electrophoresis temperature. With the advent of capillary electrophoresis (see also next chapter), multicolor fluorescently labeled DNA fragments were PCR amplified in a single tube reaction and directly subjected to SSCP analysis using an automatically operated capillary electrophoretic system (CE-SSCP; Inazuka *et al.*, 1997). In principle, any automated capillary or gel electrophoresis system can be used for F-SSCP, although the Applied Biosystems (Foster City, CA, USA) and Pharmacia (Uppsala, Sweden) ALF automated sequencers are more frequently employed.

F-SSCP advantages include:

1. Non-radioactive labeling of the PCR products using different fluorescent dyes, allowing for highly sensitive and reproducible electrophoretic pattern detection and direct data storage and processing using specialized software;
2. Automated reloading of samples, allowing their electrophoresis in multiple temperatures which results in higher sensitivity;
3. Precise temperature control, resulting in higher sensitivity and greater reproducibility;
4. Fragment mobility normalization with the use of internal lane control, allowing more accurate detection of mutations resulting in minor mobility alterations;
5. Lower running costs than conventional SSCP analysis, particularly if a large number of samples are to be analyzed in the long term with a throughput of several hundred samples per day, depending on the number of capillaries of the automated DNA sequencer;
6. Multiplexing, resulting in an increase in the analysis throughput. Multiplexing should include PCR products differing by at least 15 bp in size (Ellison, 1996), and provided that the latter fragments are previously analyzed individually in separate lanes to ensure that their electrophoretic patterns do not overlap.

Minor disadvantages include the requirement of an expensive infrastructure, namely the automated DNA sequencer, while the size range for sensitive mutation detection has not been significantly extended. Therefore, similar to conventional SSCP analysis, PCR products only up to 500 bp can be efficiently analyzed using F-SSCP with a recommended size of 100–350 bp.

4.4 PARAMETERS INFLUENCING SINGLE STRAND CONFORMATION POLYMORPHISM ANALYSIS

There are several parameters of SSCP analysis influencing the pattern of single strand conformation that will eventually affect electrophoretic migration. These are DNA amplification (such as the size of PCR products), denaturation, and the electrophoretic conditions (the length and duration of gel run, the ionic strength in the buffer being used, the temperature as well as the gel matrix composition). Another important aspect of SSCP analysis is the visualization of the single-stranded DNA fragments, using a number of detection methods. All these aspects will be discussed in detail in the following paragraphs.

4.4.1 DNA Amplification

Certain key aspects during amplification of the DNA fragment in question are of great importance in the performance of the SSCP analysis. Preferably, the size of the DNA fragment to be amplified should be between 150 and 350 bp. With certain exceptions, where mutation detection results have been reported for fragments of approximately 550 bp, this range is relatively safe and has reached a consensus between different laboratories (Hayashi, 1991; Hayashi and Yandell, 1993). Typically, one should start to screen DNA fragments that will lie within the range of 100 to 350 bp, although this is often dependent on the sequence of the DNA fragment, the GC content, as well as the various electrophoretic conditions being used. However, if longer PCR fragments are to be screened, it is advisable to digest the fragment with restriction enzymes. The choice of restriction enzymes may affect considerably the resulted conformation and hence the performance of the SSCP assay. In general, the longer the fragments, the harder it is for any

given single nucleotide change to have an effect in the conformation of the fragment, although under certain circumstances detection efficiency is not uniformly decreased with increasing DNA fragment length. It is noteworthy that sensitivity of SSCP can be greatly improved even for fragments as big as 800 bp, by running the electrophoresis in low pH buffer systems and at a fixed temperature (Kukita *et al.*, 1997).

After amplification, all PCR products should be screened on an agarose gel for the desired product length. If, despite considerable efforts, undesired side products are difficult to eliminate, a PCR product purification protocol should be applied. In addition, a negative and, where possible, a positive control of each PCR product should be included during each amplification step, since this will be the only indirect evidence that the screening assay is performed on the desired PCR products. PCR products can be stored on DNase-free tubes at 4°C. However, prolonged storage should be avoided and the subsequent steps of SSCP protocol should be performed as soon as possible.

4.4.2 Denaturation

In SSCP analysis, it is important to achieve complete and as much irreversible denaturation of the DNA strands as possible. Incomplete denaturation, partial folding, and reannealing to double-stranded DNA will greatly reduce the amount of single-stranded DNA in the assay and will subsequently affect detection of the single-stranded molecules. Usually, denaturation of the PCR products is carried out by incubation to high temperature, that is, 95°C for 5 to 7 min, and immediately after are chilled on ice for approximately 10 min. Alternatively, there are numerous denaturing agents, such as formamide, methyl-mercuric hydroxide, sodium hydroxide, and urea, which seem to perform well (Humphries *et al.*, 1997). Formamide is the most commonly used denaturing agent. In certain conditions, it is advisable to use 5–10% glycerol prior to loading the samples on the gel. This strategy was previously shown to produce sharper DNA bands, which greatly facilitate subsequent interpretation of results. Notably, an alternative method to increase the concentration of the single-stranded DNA molecules is asymmetric PCR. Following the first amplification step, an aliquot of the PCR product is used as template for nested PCR using only one of the primers employed in the previous amplification round (Lázaro and Estivill, 1992). This approach overcomes the incomplete DNA denaturation and reannealing.

4.4.3 Electrophoretic Parameters

Prior to reviewing the various electrophoretic parameters that may influence the SSCP analysis, it is worthwhile mentioning that currently there is neither adequate theory nor any physicochemical model available that could allow one to predict the three-dimensional structure of any given single-stranded DNA fragment, and as a result, its electrophoretic mobility. Apart from the size of the DNA fragment and its GC content, the following parameters have been empirically found to affect the sensitivity of SSCP analysis: the gel matrix composition, the buffer composition (ionic strength, the pH and buffer supplements, such as glycerol), the duration of gel run, the gel length, the DNA concentration, and the electrophoresis temperature. The effect of these factors on SSCP resolution and sensitivity is outlined next.

1. *Gel matrix composition*. For conventional SSCP analysis, the most common and widely accepted matrix is a cross-linked acrylamide polymer (8–12%). The small pore size of acrylamide-derived matrices makes it ideal for enhanced resolution and discrimination even at the nucleotide level. Higher resolution can be achieved upon addition of 10–15% of sucrose or glycerol. It has been previously shown that the mutation detection enhancement (MDE) gel (FMC Bioproducts) had a mutation detection rate of approximately 95% (Ravnik-Glavac *et al.*, 1994). Although there is a considerable variability in the percentage of mutations detected with this commercially available gel matrix, it should be noted that in a considerable and rather growing number of studies for which SSCP has been implemented, MDE gels were used instead of the standard acrylamide. For CE-SSCP analysis, data suggest that a sensitivity of 98–99% can be obtained using a 10% long chain poly-*N*,*N*-dimethylacrylamide polymer (LPDMA; Jespersgaard *et al.*, 2006).
2. *Buffer composition*. So far, the Tris-borate buffer is the buffer of choice for most investigators in SSCP analysis. However, in certain instances HEPES buffer has been demonstrated to offer an alternative solution that may increase the sensitivity of the SSCP. The addition of 5% glycerol has been shown to lower the pH and to decrease the electrostatic repulsion between the negatively charged phosphates in the nucleic acid backbone, resulting in a higher resolution between the mutant and wild-type DNA fragments. Conformational structures can also be more compacted by increasing the salt concentration. Finally, buffer systems with low pH have been shown to increase the sensitivity for mutation detection in larger DNA fragments (Kukita *et al.*, 1997).
3. *Gel length and duration of gel run*. There is a considerable variation in the duration of the gel run that has been adopted by the various diagnostic laboratories. It is inevitable that the time of electrophoresis is dependent on both the length of the gel as well as the applied voltage. It is preferable to start the electrophoresis with a relatively moderate voltage and increase it as soon as the PCR fragments have been migrated into the gel.

The length could vary between 10 and 40 cm. In general the bigger the gel length, the better the resolution, since at several occasions the conformational changes of the wild-type and mutant single-stranded alleles are so minor that they may migrate at very close proximity to each other.

4. *Temperature*. Several laboratories have demonstrated in the past the importance of temperature on the conformational changes of the DNA fragments. It is conceivable that the known temperature effect on the stabilization of the secondary structure of single-stranded DNA fragments may affect to a varying degree (depending on the primary sequence) the SSCP results. In addition to a lower pH, decreasing the temperature to 4°C has been shown to enhance the stability of the conformation for any given single-stranded DNA fragment. It is advisable that one should try a gradual temperature decrease, starting from 15°C and descending with an increment of 2–4°C at a time. For CE-SSCP, Jespersgaard and coworkers (2006) concluded that a sensitivity of 98–99% could be obtained when using the LPDMA polymers and a single electrophoresis temperature of 27°C, by analyzing a temperature range between 18 and 35°C.
5. *DNA concentration*. In the past, several investigators have used high DNA concentrations in order to enhance detection. Unfortunately, this has often led to a decrease in the specific concentration of the single-stranded DNA. Even after the addition of formamide, it has been empirically shown that at high concentrations, the two single-stranded molecules tend to reanneal and form a double-stranded DNA. It is preferable, therefore, to keep the DNA concentration relatively low in the loading buffer. Also, gel overloading can sometimes result in abnormal migration of the bands, leading to decreased resolution. Similarly, in F-SSCP a critical step includes dilution of the PCR products. If the sample is too concentrated, then peaks will not be sharp enough (also known as peak broadening) to allow unambiguous detection, leading to false positive results. On the other hand, if the sample is overdiluted, the poor signal-to-noise ratio will also result in false negative results. Therefore, the appropriate dilution factor should be estimated for each PCR, ranging from 1/10 to 1/80.

4.4.4 Detection

Most molecular diagnostic laboratories are well adapted in the use of radioactivity and have the equipment and expertise required for this type of protocol. Typically, autoradiography requires immobilization of the gel, a drying step, and finally film exposure. However, this approach is more time consuming than some of the more recently described protocols that utilize silver staining or detection with fluorescent dyes. Alternatively, in F-SSCP, analysis can be done on automated sequencers with fluorescent dye-labeled DNA fragments. In relative terms, and although ethidium bromide staining can be considered an attractive alternative to the use of fluorescent substances (Yap and McGee, 1992a; Lázaro and Estivill, 1992), the silver staining approach comprises one of the most straightforward, fast, as well as sensitive methods for the visualization of bands on conventional SSCP gels. In brief, after electrophoresis, the polyacrylamide gels are first fixed with 10% acetic acid for approximately 30 min at room temperature and subsequently washed with water. Depending on the concentration of silver nitrate, incubation with the silver nitrate solution can last for approximately 60 min (in a 0.001% $AgNO_3$, 0.036% formaldehyde solution). This incubation step is performed in the dark, while avoiding any contamination with protein-containing solution (proteins are stained extensively with silver nitrate leading to immense background). Subsequently, the polyacrylamide gels are washed with water and color development is performed by incubating the gel for 5 to 10 min with a color development solution (containing 2.5% Na_2CO_3, 0.036% formaldehyde, and 0.002% sodium thiosulfate). Color development can be stopped with a solution containing a chelating agent (such as 1.5% EDTA). Gels can be subsequently fixed with 30% ethanol and 4% glycerol. The stained gels are transferred to a vacuum dryer and are immobilized to a porous paper. Results are interpreted visually or can be analyzed by means of an image analysis system.

4.5 HETERODUPLEX ANALYSIS FOR MUTATION DETECTION

The principle of heteroduplex analysis (HDA) is simple and closely related to that of SSCP. In brief, heteroduplexes are formed between different DNA alleles; for example, by mixing wild-type and mutant amplified DNA fragments, followed by denaturation at 95°C and slow reannealing to room temperature (White *et al.*, 1992). If the target DNA consists of different alleles already – for example, a heterozygous case – then heteroduplexes are formed automatically during the amplification step. The result is the formation of two homoduplexes and two heteroduplexes, which are retarded during electrophoresis in native polyacrylamide gels (see Fig. 4.4). There are two types of heteroduplex molecules, depending on the type of the mutation, which, in turn, reflects on their stability. In other words, small deletions or insertions create stable heteroduplexes, termed *bulge-type* heteroduplexes, which have been verified by electron microscopy (Wang *et al.*, 1992). On the other hand, single base substitutions form the so-called *bubble-type* heteroduplexes, which are much more difficult to visualize, and optimization of the experimental conditions is required to achieve optimal resolution of this type of heteroduplex. A number of molecular diagnostic

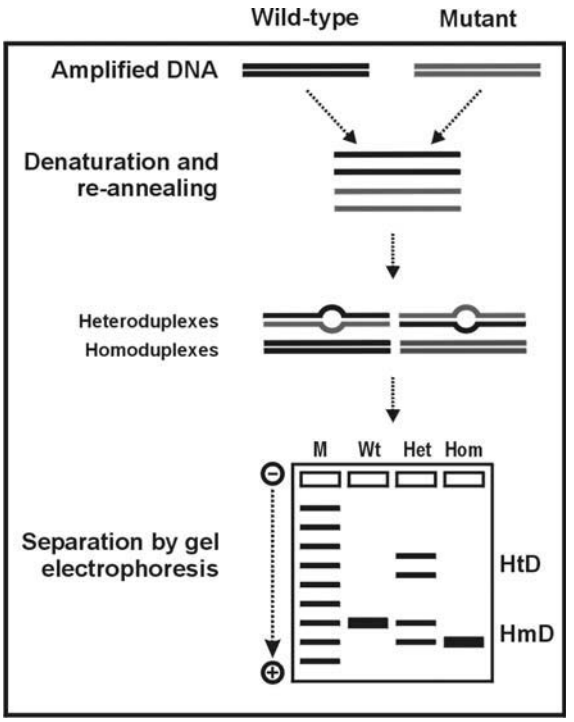


FIGURE 4.4 The HDA principle. Unlike SSCP analysis, amplified DNA is now subjected to denaturation followed by slow reannealing of the denatured alleles, leading to both homoduplexes (HmD) and heteroduplexes (HtD). The latter migrate more slowly during gel electrophoresis, due to their sequence mismatch(es) and therefore heterozygous (Het) and homozygous (Hom) cases can be easily distinguished from the wild type (Wt) based on their electrophoretic pattern. M: size marker.

techniques, based on heteroduplex formation, are reported in the literature, such as enzymatic or chemical cleavage (see Chapter 3) and denaturing gradient gel electrophoresis (DGGE; see Chapter 6), but HDA appears to be the most attractive one, as it can be performed rapidly on short gels without the need of specialized equipment and the use of radioactivity. The most typical example of the use of HDA for mutation screening is the rapid detection of the 3 bp p.F508del deletion in the *CFTR* gene, leading to cystic fibrosis (Wang *et al.*, 1992).

HDA can be also adapted to a fluorescent capillary system, such as the currently available automated DNA sequencers. This is possible by labeling the primers with fluorescent dyes, such as 6-carboxyfluorescein (FAM), hexachloro-6-carboxyfluorescein (HEX), etc. Like F-SSCP and CE-SSCP, capillary-based HDA (CE-HDA) also has the advantage of multiplexing, which together with the use of different fluorescent dyes can significantly increase the analysis throughput (Kozlowski and Krzyzosiak, 2001). Use of multi-capillary systems, e.g. 96-capillary platforms, can also reduce the time needed for processing each sample. For example, up to 10 PCR products can be analyzed in one capillary per run, indicating that the entire analysis of a patient's *BRCA1* and *BRCA2* coding regions can be completed in a single run within 1.5 h (Esteban-Cardeñosa *et al.*, 2004). In brief, CE-HDA is a high-throughput, sufficiently sensitive and cost-effective methodology and can be easily implemented to offer reliable genetic analysis in molecular diagnostic laboratories with large sample volumes.

4.6 SENSITIVITY AND LIMITATIONS

Several factors can affect the sensitivity of both SSCP and HDA analysis and their optimization is highly empirical, as there is no adequate theoretical basis or type of algorithm (as in DGGE; Myers *et al.*, 1987) that would enable researchers to predict the three-dimensional conformation of the single-stranded DNA fragments under specific experimental procedures. Those elements that frequently affect the sensitivity of both methods will be discussed in this section.

4.6.1 Optimizing SSCP Sensitivity

In general, most of the single-stranded DNA fragments will have a more compacted conformation at lower temperatures or in the presence of higher salt concentrations. It follows, however, that by increasing the salt concentration, the conductivity is also increased, thus having a rather significant effect on temperature. Depending on these parameters, which largely have been determined empirically, it is suggested that electrophoresis be performed at room temperature with 5–10% glycerol or at 4°C without addition of glycerol or salt; this is considered a good starting point. In addition, it is highly suggestive that an empirical estimate of SSCP sensitivity will be based on the likelihood of detecting known sequence variations under controlled conditions. This is expected to provide a valuable means by which the effect on electrophoretic mobility of known sequence variations will be determined. Under no circumstances should a perfect overall detection efficiency be anticipated. Hayashi and Yandell (1993) have conducted a valuable study to determine which are the most effective sets of parameters influencing the detection of a single mutation. Given that the latter is largely dependent on DNA fragment's size, it is expected that for most fragments shorter than 200 nucleotides, more than 90% of the sequence variations will be detected. Gradually, as the size of the PCR fragment increases to 300–350 nucleotides, a safe prediction is that more than 80% of mutations will be detected. It should be noted, however, that within this range, Sheffield and colleagues (1993) have shown that the overall sensitivity of the SSCP is not affected by the position of a given mutation. On the contrary, Ellis and coworkers (2000) have shown that the increase in fragment size did not have a substantial effect on F-SSCP sensitivity, by analyzing *ABCC7* exon 11 amplicons of 190 bp and 490 bp, respectively. Alternatively, a more quantitative method for estimating the sensitivity of

the SSCP detection method based on statistical arguments has been developed. This method is based on the fact that the chance of any given strand to exhibit a mobility shift is independent from the other strand. On this assumption, the probability of observing shifts in both strands (P2), in one strand (P1), or in none of the strands (P0), is equal to x^2, $2x(1-x)$ and $(1-x)^2$, respectively, where x is the sensitivity of the technique when only one of the strands is labeled. The ratio of P2 to P1 is equal to the observed number of the mutations on both strands over the number of the observed mutations on the single strand [$r = (x/2(1-x)$]. The useful sensitivity of the technique can be calculated as being equal to 1 − P0. Based on this probabilistic theory, the estimates from previous studies are that the sensitivity for 100–200 bp fragments is approximately 96%, regardless of the presence or absence of glycerol (Hayashi and Yandell, 1993). However, by adding glycerol, the sensitivity is still high for fragments ranging from 200 to 300 bp, but is decreasing when glycerol is not used. The latter may indicate the inability of the former calculation to depict the actual electrophoretic mobility, which to some extent is expected. In addition, it may also explain the fact that substitutions, which induce significant conformational changes, are most likely to have an effect, to a variable extent, on the other strand. Despite this discrepancy, it can be safely concluded that decreasing the fragment size will greatly enhance the sensitivity. It seems rational that the overall effect of a given mutation is displayed more efficiently when the total number of nucleotides surrounding that particular mutation is less. It is also profound that glycerol greatly increases the overall sensitivity when electrophoresis is performed at room temperature.

Practical observations suggest that any fragment that exhibits a differential mobility will often migrate very close to the reference fragment. However, the overall fragment number is not always predictable in advance. Any given number of conformations may be supported to a variable extent by the applied electrophoretic parameters. Furthermore, the band intensity is irrelevant to allelic differences, as it is strictly dependent on the different conformations. Therefore, SSCP is not a safe method to predict gene-dosage effects. In addition, the simultaneous detection of more than one mutation in a single DNA sample is not easily predictable, as previous data have shown that the electrophoretic pattern may vary considerably within different experiments. In general, although highly reproducible, the mobility of single-stranded DNA conformers cannot be predicted in advance from sequence information. Such attempts to predict SSCP mobility changes by modeling single-stranded DNA conformations using the structure prediction program Mfold (http://frontend.bioinfo.rpi.edu/applications/mfold/) have not provided consistent results. Improvements in modeling the structures of single-stranded DNA would make it possible to more accurately predicting SSCP mobility (Nakabayashi and Nishigaki, 1996), the idiosyncratic nature of SSCP remains its main weakness as a diagnostic tool (Liu *et al.*, 1999).

Attempts to further improve the sensitivity of the SSCP have led to the development of the RNA-SSCP approach (Sarkar *et al.*, 1992). Here, although the method is essentially the same, the double-stranded DNA is converted to the corresponding single-stranded RNA by means of one of the two primers that has phage promoter sequences on its 5′ end. This method has shown a higher sensitivity compared to the previously described SSCP methodology. Its sensitivity may rely on the fact that the *in vitro* transcribed strand has no complementary strand to reanneal with. Therefore, sufficient amounts of the *in vitro* transcribed product can be electrophoresed and easily analyzed even by ethidium bromide staining (Sarkar *et al.*, 1992).

So far, there has been much discussion concerning the false negative results of this assay and possible ways to minimize them. However, false positive results also affect the net outcome of the SSCP analysis. In order to minimize the frequency of reporting false positive results, it is advisable to perform repeated SSCP electrophoretic runs (particularly when SSCP data are of clinical interest). An additional way is to determine the minimum mobility variation, which is detectable within the context of laboratory SSCP conditions. In practical terms, mobility differences of 3 mm are generally clear, but a detection difference of only 2 mm requires excellent gel running conditions and is often subjective. Therefore, any difference smaller than or equal to 2 mm can be considered only with reservation.

Reproducibility is the last and perhaps most essential parameter applicable to most techniques in molecular diagnostics. In general, if conditions are kept constant then the resulting reproducibility is usually high. Nevertheless, Hayashi (1991) has previously suggested that DNA sequences may have different stable conformations. The latter is thus interpreted as a variable that may compromise SSCP reproducibility. It relies on the possibility that when the free energy difference between different conformations is small, an oscillation between different structures of comparable energies may be observed.

4.6.2 Sensitivity of HDA

So far, the sensitivity of the HDA methodology has not been determined to the extent of the SSCP analysis. Rossetti and colleagues (1995) compared directly both SSCP and HDA assays for the detection of known mutations in a panel of four genes. Despite the fact that none of the assays was performed with 100% efficiency, HDA detected slightly more mutations than SSCP in the same samples. Interestingly, these authors suggested that both techniques could be used in concert to detect all mutations.

Another aspect to improve the performance of HDA is the gel matrix. As in SSCP, the use of MDE (derived from HydroLink D5000™; Keen *et al.*, 1991) has basically made

HDA a valuable mutation detection technique. Today, the majority of the diagnostic laboratories, in which HDA is the method of choice to detect genomic variation, are employing MDE gels.

Finally, as already mentioned in section 4.5, the *bubble-type* heteroduplexes, which are formed due to the presence of single base substitutions, are much more difficult to visualize compared to the *bulge-type* heteroduplexes. In order to overcome this bottleneck, and based on the observation that heteroduplexes are much easier to visualize when a deletion or insertion mutation is involved, the Universal Heteroduplex Generator (UHG) was conceived (Wood *et al.*, 1993a, b). In brief, the UHG consists of a synthetic DNA fragment, which bears a small (that is, 2 to 5 bp) deletion. This synthetic fragment is amplified by the use of the same oligonucleotide primers as the DNA under study. After amplification, the test amplicon is mixed with the amplified UHG, denatured, and then slowly reannealed, followed by electrophoresis. If no mutation is present in the test DNA, then only a bulge-type heteroduplex will be present, slightly retarded compared to the homoduplex. If, however, a single base substitution is also present, then the resulting heteroduplex will have two mismatches: a bulge and a bubble type, which will result in the heteroduplex migrating significantly lower, compared to the simple bulge-type heteroduplex. The use of UHG has been reported for the detection of known mutations within a number of loci, such as von Willebrand disease (Wood *et al.*, 1995), phenylketonurea (Wood *et al.*, 1993a), and for prenatal determination of blood group alleles (Stoerker *et al.*, 1996).

4.7 DETECTION OF THE UNDERLYING GENOMIC VARIATION USING SSCP AND HDA

Due to their numerous advantages, SSCP and HDA analyses are nowadays the methods of choice in a growing number of private or public molecular diagnostic laboratories to either interrogate known mutations or scan for known or unknown mutations in short stretches of DNA and in relatively short time. Similarly, SSCP and HDA have also been proven to be invaluable tools for basic science, enabling both the identification of causative genes for human hereditary diseases and mapping of genomic loci.

A short summary of the existing applications of SSCP and HDA follows in the next paragraphs, which is only indicative for the applicability of these techniques in almost every genomic locus.

4.7.1 Applications in Basic Science

The utilization of SSCP in basic science as a tool for genomic DNA analysis is well established. Mutations in several key candidate genes implicated in various cell processes have now been identified, and the extent of those mutations as well as their frequency has given an insight into the role of these molecules in the relevant processes.

Damage to DNA is considered to be the main initiating event by which genotoxins cause hereditary effects and cancer. An accumulation of mutations throughout the genome will eventually result in cell death or in a cascade of events, which in turn may initiate malignant transformation. Therefore, it is not surprising that SSCP analysis was first used in screening candidate genes in tumorigenesis (Suzuki *et al.*, 1990; Yap and McGee, 1992b). Since then, several tissue specimens have been examined and nearly all possible tumor types, isolated from a variety of tissue resources, have been considered for mutations in several key suspect genes. A difficulty often was the fact that many genes such as the *RB* or *BRCA1* and *BRCA2* genes have an enormous size, comprising several exons, and their analysis was often cumbersome if not impossible. Nevertheless, a number of research groups, including ours, have identified a multitude of p53 and nm-23 genomic alterations in almost every tumor type, such as in breast, colorectal (see Fig. 4.5), prostate, ovarian, and so on. In effect, this methodology proved to be

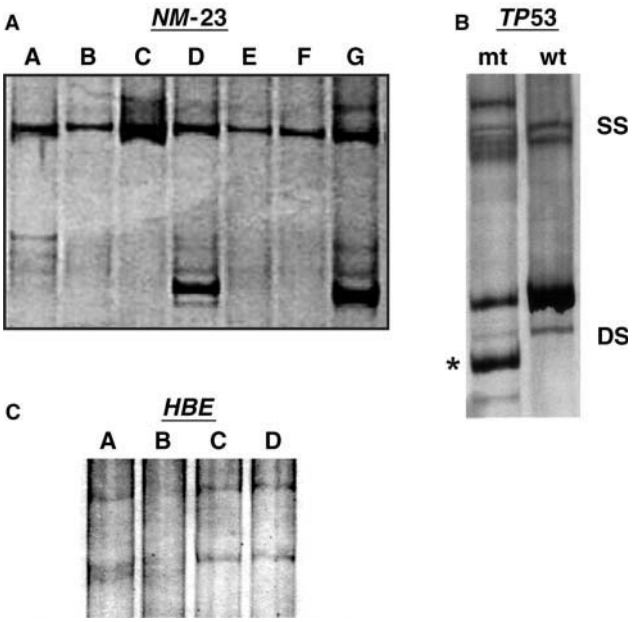


FIGURE 4.5 Typical examples of silver stained SSCP gels, in which different genomic loci, responsible for tumorigenesis, are analyzed. **A.** Analysis of the nm-23 gene. Lanes A, D, and G correspond to three different heterozygous cases; lanes B, C, E, and F are normal individuals (Garinis, G. and Menounos, P.G., unpublished). **B.** Analysis of the human p53 locus. Different electrophoretic mobility of both the single-stranded (SS) and double-stranded (DS) alleles, due to a small deletion (depicted by an asterisk) in a heterozygous breast cancer affected individual (mt), compared to the wild-type (wt) electrophoretic pattern (adapted from Patrinos *et al.*, 1999, with permission). **C.** Analysis of the human *HBE* (ε-globin) gene. Lanes A and B: Heterozygote and homozygous cases for the presence of the ε/HincII polymorphism respectively. Lanes C, D: Homozygous cases for the absence of the ε/HincII polymorphism. (Adapted from Papachatzopoulou *et al.*, 2006.)

particularly useful in revealing, in a stepwise approach, that the altered expression of several cell cycle regulatory molecules either at the genomic or transcriptional and protein levels may exert a synergetic effect on tumor growth and chromosomal instability on breast cancer and non-small cell lung and colorectal carcinomas.

In addition to the utilization of SSCP methodology as a screening tool, several investigators have previously employed this technique for gene mapping in mouse genes (Beier, 1993 and references therein). The methodology is based on the fact that a given polymorphism readily can be found in non-coding regions of genes such as the 3′ untranslated regions or introns, between alleles of mouse species, and in several occasions between inbred strains as well. The segregations of these polymorphisms can be analyzed with recombinant inbred or interspecific crosses and the strain distribution pattern obtained can be compared with that for other markers and analyzed by standard linkage analysis algorithms. For instance, SSCP has previously been employed to localize 39 mouse-specific sequence-tagged sites (STSs), generated from mouse–hamster somatic cell hybrids. These were subsequently integrated with other markers to generate a high-density map of mouse chromosome 1 containing over 100 markers typed on a single interspecific backcross (Watson *et al.*, 1992). SSCP analysis also has been applied in relatively fewer cases for the linkage analysis of human genes (Nishimura *et al.*, 1993; Avramopoulos *et al.*, 1993).

4.7.2 Molecular Diagnostic Applications

Both conventional and fluorescent or capillary-based SSCP and HDA can be successfully used for the detection of known mutations in any genomic locus. A brief summary of the numerous applications of SSCP analysis for various human genes is given in Table 4.1. When SSCP analysis is coupled to non-radioactive detection schemes, then it most certainly becomes the method of choice for routine molecular diagnostic analysis (see Fig. 4.6 for representative examples from *G6PD* mutation screening (Menounos *et al.*,

TABLE 4.1 Summary of the majority of genomic loci, for which an SSCP mutation analysis strategy is designed and implemented.

Genes	Disease/syndrome	References
Ras *TP53*	Various types of cancer	Suzuki *et al.* 1990 Sheffield *et al.* (1993), Kutach *et al.* (1999), Makino *et al.* (2000)*
BRCA1 *BRCA2*	Susceptibility to breast cancer	Castilla *et al.* (1994) Phelan *et al.* (1996)
RB1	Retinoblastoma	Hogg *et al.* (1992), Shimizu *et al.* (1994)
APC	Adenomatous polyposis coli	Groden *et al.* (1993), Varesco *et al.* (1993)
CYP21	Congenital adrenal hyperplasia	Bobba *et al.* (1997)
PAH	Phenylketonurea	Dockhorn-Dworniczak *et al.* (1991), Labrune *et al.* (1991)
G6PD	Glycose-6-dehydrogenase deficiency	Calabrò *et al.* (1993)
HFE	Hereditary hemochromatosis	Hertzberg *et al.* (1998)
F7	Hemophilia A	Economou *et al.* (1992)
F9	Hemophilia B	David *et al.* (1993)
SPTA1 *SPTB*	Hereditary elliptocytosis and spherocytosis	Maillet *et al.* (1996)
ANK1 *SLC4A1* (Band-3)	Hereditary spherocytosis	Eber *et al.* (1996) Jarolim *et al.* (1996)
HBA2, HBA1	α-thalassemia	Harteveld *et al.* (1996)

(*Continued*)

TABLE 4.1 (Continued)

Genes	Disease/syndrome	References
HBB	β-thalassemia	Takahashi-Fujii *et al.* (1994)
FAA	Fanconi anemia	Levran *et al.* (1997)
CFTR	Cystic fibrosis	Claustres *et al.* (1993)
PKD1	Polycystic kidney disease 1	Afzal *et al.* (1999)
PKD2	Polycystic kidney disease 2	Veldhuisen *et al.* (1997)
LDLR	Familial hypercholesterolemia	Day *et al.* (1995), Sözen *et al.* (2005)*
NF1	Neurofibromatosis type 1	Upadhyaya *et al.* (1995)
NF2	Neurofibromatosis type 2	MacCollin *et al.* (1994)
MECP2	Rett syndrome	Zappella *et al.* (2003)
GBA	Gaucher disease	Kawame *et al.* (1992)
DMD	Duchenne/Becker muscular dystrophy	Tuffery *et al.* (1993)
WFS1	Wolfram syndrome	Strom *et al.* (1998)
GJB2 *GJA1*	Autosomal recessive non-syndromic hearing loss	Scott *et al.* (1998) Liu *et al.* (2001)
RHO	Retinitis pigmentosa	Dryja *et al.* (1991)
FBN1	Marfan syndrome	Hayward *et al.* (1994)
TTR	Familial amyloid polyneuropathy	Saeki *et al.* (1992)
WIT1	Wilms tumor	Clarkson *et al.* (1993)
EXT1, EXT2	Hereditary multiple exostoses	Dobson-Stone *et al.* (2000)*
GCK	Maturity-onset diabetes of the young (MODY)	Thomson *et al.* (2003)*

**F-SSCP analysis.*

2000) and *HBE* genetic analysis (Papachatzopoulou *et al.*, 2006)). Additionally, HDA strategies have been designed for the detection of genetic defects in a number of human genes, such as *CFTR* (Fig. 4.7; Rommens *et al.*, 1990), *NF1* (Shen *et al.*, 1993) and *NF2* (Sainz *et al.*, 1994), *APC* (Paul *et al.*, 1993), *PKU* (Wood *et al.*, 1993a), *BRCA1* and *BRCA2* (using the CE-HDA approach; Kozlowski and Krzyzosiak, 2001; Esteban-Cardeñosa *et al.*, 2004), and so on.

Furthermore, SSCP and HDA can be used equally well for molecular typing in clinical microbiology (see also Chapter 30). Recently, Nair and coworkers (2002) have investigated the possibility of using SSCP analysis for the detection of nucleotide variation at the groEL gene, in order to differentiate Salmonella strains both at the interserovar and the intraserovar levels. In this study, SSCP analysis has exhibited the potential to complement classic typing methods such as serotyping and phage typing for the identification of salmonella serovars, due to its rapidity and simplicity. SSCP analysis of the groEL genes of carious Salmonella serovars produced various SSCP profiles, which indicates the potential of this technique to differentiate various Salmonella serovars. Also, Peters and coworkers (1997) reported successful differentiation between hepatitis C virus (HVC) quasispecies, using F-SSCP analysis of RT-PCR-generated fragments from HCV RNA, containing hypervariable region I (HVRI) of the HCV genome. Provided that the virus titer is sufficient, these authors showed that F-SSCP is a quick and reliable method for HCV quasispecies analysis. Finally, F-SSCP

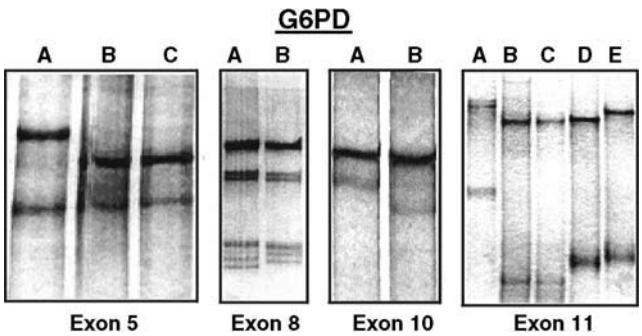


FIGURE 4.6 Analysis of the coding region of the human G6PD locus (adapted from Menounos *et al.*, 2000, with permission). Exon 5: Lanes B and C correspond to the wild-type pattern and lane A corresponds to a heterozygous case for a point mutation. Exon 8: Lane A: normal individual, lane B: heterozygous case. Exon 10: Lane A: heterozygous case for a point mutation, lane B: normal individual. Exon 11: Lanes B and C correspond to the wild-type pattern, lanes A and D correspond to heterozygous cases for different mutations, and lane E corresponds to a heterozygous case with two point mutations in cis, one of which is the same as in lane D (note the minor mobility difference of the bands for the single-stranded alleles between lanes D and E).

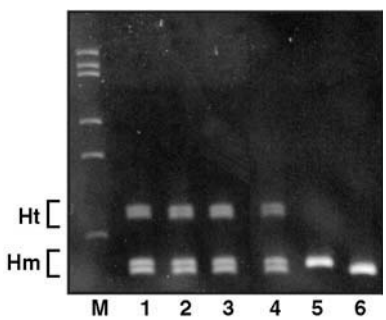


FIGURE 4.7 Mutation screening in the Exon 10 of the *CFTR* gene for the p.F508del mutation, leading to cystic fibrosis, using HDA. Electrophoresis is performed in a non-denaturant 8% acrylamide gel. The amplification product of the wild-type allele is 97 bp and that of the mutant allele is 94 bp. Lanes 1 to 4 correspond to p.F508del heterozygotes, lane 6 corresponds to a p.F508del homozygote, and lane 5 to a wild-type individual. The electrophoretic mobility of the heteroduplexes (Ht) is retarded compared to the homoduplexes (Hm). M: ϕX174/HaeIII size marker (photo courtesy of Dr. Angeliki Balassopoulou, Athens, Greece).

has been proven a particularly useful method in human platelet antigen genotyping (Quintanar *et al.*, 1998).

4.8 CONCLUSIONS AND FUTURE ASPECTS

As previously mentioned, SSCP and HDA are simple, reliable, and sensitive methods for the detection of nucleotide sequence changes in genomic loci. The methods can detect single nucleotide substitutions, insertions, or deletions of a short nucleotide sequence accurately and in a relatively short time, and therefore they can be used for DNA analysis of human cancers and other genetic disorders. Almost all the polymorphic base substitutions, thought to be present every few hundred base pairs in genomic DNA, also can be detected and used as genetic markers. In addition, the identification of known mutations and polymorphisms can readily be standardized, and a role of SSCP and/or HDA analysis in routine laboratory screening procedures has been already demonstrated.

Finally, the generation of the draft human genome sequence has imposed an increasing need for methods and technology that can be used for high-throughput mutation screening in a large number of DNA samples. Initially, improvements in SSCP included sample handling, allowing mutation analysis in an extremely rapid manner (Whittall *et al.*, 1995). In particular, a 96-well format was employed for storing DNA and subsequently for amplification and mutation screening, requiring no tube labeling and reducing the probability of contamination by using a multipipette. In addition, higher throughput was also achieved by modification of the gel system. With the advent of F-SSCP and CE-SSCP analyses, unprecedented opportunities emerged for significantly increasing the throughput and sensitivity of the analysis and, at the same time, reducing the analysis time and costs. In certain cases, F-SSCP was implemented for the development of single-cell CFTR genetic tests in preimplantation genetic diagnosis (PGD, see also Chapter 33; Blake *et al.*, 2001). Moreover, performance of capillary electrophoresis/SSCP in a microarray format also has been shown to decrease the overall analysis time, without any significant loss in resolution (Medintz *et al.*, 2000, 2001). These advances will be discussed extensively in the following chapter. One way or another, one can safely predict that SSCP and HDA will continue to be extremely popular mutation detection methodologies and they will substantially contribute to our growing knowledge of the human genetic variation as well as of other organisms.

REFERENCES

Afzal, A.R., Hand, M., Ternes-Pereira, E., Saggar-Malik, A., Taylor, R., and Jeffery, S. (1999). Novel mutations in the 3 region of the polycystic kidney disease 1 (PKD1) gene. *Hum. Genet.* 105, 648–653.

Avramopoulos, D., Cox, T., Kraus, J.P., Chakravarti, A., and Antonarakis, S.E. (1993). Linkage mapping of the cystathionine beta-synthase (CBS) gene on human chromosome 21 using a DNA polymorphism in the 3′A untranslated region. *Hum. Genet.* 90, 566–568.

Beier, D.R. (1993). Single-strand conformation polymorphism (SSCP) analysis as a tool for genetic mapping. *Mamm. Genome* 4, 627–631.

Bobba, A., Iolascon, A., Giannattasio, S., Albrizio, M., Sinisi, A., Prisco, F., Schettini, F., and Marra, E. (1997). Characterisation of CAH alleles with non-radioactive DNA single strand conformation polymorphism analysis of the CYP21 gene. *J. Med. Genet.* 34, 223–228.

Blake, D.L., Dean, N.L., Knight, C., Tan, S.L., and Ao, A. (2001). Direct comparison of detection systems used for the development of single-cell genetic tests in preimplantation genetic diagnosis. *J. Assist. Reprod. Genet.* 18, 557–565.

Calabrò, V., Mason, P.J., Filosa, S., Civitelli, D., Cittadella, R., Tagarelli, A., Martini, G., Brancati, C., and Luzzatto, L. (1993). Genetic

heterogeneity of glucose-6-phosphate dehydrogenase deficiency revealed by single-strand conformation and sequence analysis. *Am. J. Hum. Genet.* 52, 527–536.

Castilla, L.H., Couch, F.J., Erdos, M.R., Hoskins, K.F., Calzone, K., Garber, J.E., Boyd, J., Lubin, M.B., Deshano, M.L., Brody, L.C., Collins, F.S., and Weber, B.L. (1994). Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nat. Genet.* 8, 387–391.

Chehab, F.F., and Kan, Y.W. (1989). Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* 86, 9178–9182.

Clarkson, P.A., Davies, H.R., Williams, D.M., Chaudhary, R., Hughes, I.A., and Patterson, M.N. (1993). Mutational screening of the Wilms's tumour gene, WT1, in males with genital abnormalities. *J. Med. Genet.* 30, 767–772.

Claustres, M., Laussel, M., Desgeorges, M., Giansily, M., Culard, J.F., Razakatsara, G., and Demaille, J. (1993). Analysis of the 27 exons and flanking regions of the cystic fibrosis gene: 40 different mutations account for 91.2% of the mutant alleles in southern France. *Hum. Mol. Genet.* 2, 1209–1213.

David, D., Rosa, H.A., Pemberton, S., Diniz, M.J., Campos, M., and Lavinha, J. (1993). Single-strand conformation polymorphism (SSCP) analysis of the molecular pathology of hemophilia B. *Hum. Mutat.* 2, 355–361.

Day, I.N., Whittall, R., Gudnason, V., and Humphries, S.E. (1995). Dried template DNA, dried PCR oligonucleotides and mailing in 96-well plates: LDL receptor gene mutation screening. *Biotechniques* 18, 981–984.

Dobson-Stone, C., Cox, R.D., Lonie, L., Southam, L., Fraser, M., Wise, C., Bernier, F., Hodgson, S., Porter, D.E., Simpson, A.H., and Monaco, A.P. (2000). Comparison of fluorescent single-strand conformation polymorphism analysis and denaturing high-performance liquid chromatography for detection of EXT1 and EXT2 mutations in hereditary multiple exostoses. *Eur. J. Hum. Genet.* 8, 24–32.

Dockhorn-Dworniczak, B., Dworniczak, B., Brommelkamp, L., Bulles, J., Horst, J., and Bocker, W.W. (1991). Non-isotopic detection of single strand conformation polymorphism (PCRSSCP): a rapid and sensitive technique in diagnosis of phenylketonuria. *Nucleic Acids Res.* 19, 2500.

Dryja, T.P., Hahn, L.B., Cowley, G.S., McGee, T.L., and Berson, E.L. (1991). Mutation spectrum of the rhodopsin gene among patients with autosomal dominant retinitis pigmentosa. *Proc. Natl. Acad. Sci. USA* 88, 9370–9374.

Eber, S.W., Gonzalez, J.M., Lux, M.L., Scarpa, A.L., Tse, W.T., Dornwell, M., Herbers, J., Kugler, W., Ozcan, R., Pekrun, A., Gallagher, P.G., Schroter, W., Forget, B.G., and Lux, S.E. (1996). Ankyrin-1 mutations are a major cause of dominant and recessive hereditary spherocytosis. *Nat. Genet.* 13, 214–218.

Economou, E.P., Kazazian, H.H., Jr., and Antonarakis, S.E. (1992). Detection of mutations in the factor VIII gene using single-stranded conformational polymorphism (SSCP). *Genomics* 13, 909–911.

Ellis, L.A., Taylor, C.F., and Taylor, G.R. (2000). A comparison of fluorescent SSCP and denaturing HPCL for high throughput mutation scanning. *Hum. Mutat.* 15, 556–564.

Ellison, J.S. (1996). Fluorescence-based mutation detection. Single-strand conformation polymorphism analysis (F-SSCP). *Mol. Biotechnol.* 5, 17–31.

Ellison, J., Dean, M., and Goldman, D. (1993). Efficacy of fluorescence-based PCR-SSCP for detection of point mutation. *BioTechniques* 15, 684–691.

Esteban-Cardeñosa, E., Duran, M., Infante, M., Velasco, E., and Miner, C. (2004). High-throughput mutation detection method to scan BRCA1 and BRCA2 based on heteroduplex analysis by capillary array electrophoresis. *Clin. Chem.* 50, 313–320.

Groden, J., Gelbert, L., Thliveris, A., Nelson, L., Robertson, M., Joslyn, G., Samowitz, W., Spirio, L., Carlson, M., Burt, R., Leppert, M., and White, R. (1993). Mutational analysis of patients with adenomatous polyposis: identical inactivating mutations in unrelated individuals. *Am. J. Hum. Genet.* 52, 263–272.

Harteveld, K.L., Heister, A.J., Giordano, P.C., Losekoot, M., and Bernini, L.F. (1996). Rapid detection of point mutations and polymorphisms of the alpha-globin genes by DGGE and SSCA. *Hum. Mutat.* 7, 114–122.

Hayashi, K., and Yandell, D.W. (1993). How sensitive is PCRSSCP? *Hum. Mutat.* 2, 338–346.

Hayashi, K. (1991). PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. *PCR Methods Appl.* 1, 34–38.

Hayward, C., Rae, A.L., Porteous, M.E., Logie, L.J., and Brock, D.J. (1994). Two novel mutations and a neutral polymorphism in EGF-like domains of the fibrillin gene (FBN1): SSCP screening of exons 15–21 in Marfan syndrome patients. *Hum. Mol. Genet.* 3, 373–375.

Hertzberg, M.S., McDonald, D., and Mirochnik, O. (1998). Rapid diagnosis of hemochromatosis gene Cys282Tyr mutation by SSCP analysis. *Am. J. Hematol.* 57, 260–261.

Hogg, A., Onadim, Z., Baird, P.N., and Cowell, J.K. (1992). Detection of heterozygous mutations in the RB1 gene in retinoblastoma patients using single-strand conformation polymorphism analysis and polymerase chain reaction sequencing. *Oncogene* 7, 1445–1451.

Humphries, S.E., Gudnason, V., Whittall, R., and Day, I.N. (1997). Single-strand conformation polymorphism analysis with high throughput modifications, and its use in mutation detection in familial hypercholesterolemia. International Federation of Clinical Chemistry Scientific Division: Committee on Molecular Biology Techniques. *Clin. Chem.* 43, 427–435.

Inazuka, M., Wenz, H.M., Sakabe, M., Tahira, T., and Hayashi, K. (1997). A streamlined mutation detection system: multicolor post-PCR fluorescence labeling and single-strand conformational polymorphism analysis by capillary electrophoresis. *Genome Res.* 7, 1094–1103.

Iwahana, H., Yoshimoto, K., Mizusawa, N., Kudo, E., and Itakura, M. (1994). Multiple fluorescence-based PCR-SSCP analysis. *Biotechniques* 16, 296–305.

Jarolim, P., Murray, J.L., Rubin, H.L., Taylor, W.M., Prchal, J.T., Ballas, S.K., Snyder, L.M., Chrobak, L., Melrose, W.D., Brabec, V., and Palek, J. (1996). Characterization of 13 novel band 3 gene defects in hereditary spherocytosis with band 3 deficiency. *Blood* 88, 4366–4374.

Jespersgaard, C., Larsen, L.A., Baba, S., Kukita, Y., Tahira, T., Christiansen, M., Vuust, J., Hayashi, K., and Andersen, P.S. (2006). Optimization of capillary array electrophoresis single-strand conformation polymorphism analysis for routine molecular diagnostics. *Electrophoresis* 27, 3816–3822.

Kawame, H., Hasegawa, Y., Eto, Y., and Maekawa, K. (1992). Rapid identification of mutations in the glucocerebrosidase gene of Gaucher disease patients by analysis of single-strand conformation polymorphisms. *Hum. Genet.* 90, 294–296.

Keen, J., Lester, D., Inglehearn, C., Curtis, A., and Bhattacharya, S. (1991). Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. *Trends Genet.* 7, 5.

Kozlowski, P., and Krzyzosiak, W.J. (2001). Combined SSCP/duplex analysis by capillary electrophoresis for more efficient mutation detection. *Nucleic Acids Res.* 29, E71.

Kukita, Y., Tahira, T., Sommer, S.S., and Hayashi, K. (1997). SSCP analysis of long DNA fragments in low pH gel. *Hum. Mutat.* 10, 400–407.

Kutach, L.S., Bolshakov, S., and Ananthaswamy, H.N. (1999). Detection of mutations and polymorphisms in the p53 tumor suppressor gene by single-strand conformation polymorphism analysis. *Electrophoresis* 20, 1204–1210.

Labrune, P., Melle, D., Rey, F., Berthelon, M., Caillaud, C., Rey, J., Munnich, A., and Lyonnet, S. (1991). Single-strand conformation polymorphism for detection of mutations and base substitutions in phenylketonuria. *Am. J. Hum. Genet.* 48, 1115–1120.

Lázaro, C., and Estivill, X. (1992). Mutation analysis of genetic diseases by asymmetric-PCR SSCP and ethidium bromide staining: application to neurofibromatosis and cystic fibrosis. *Mol. Cell. Probes* 6, 357–359.

Levran, O., Erlich, T., Magdalena, N., Gregory, J.J., Batish, S.D., Verlander, P.C., and Auerbach, A.D. (1997). Sequence variation in the Fanconi anemia gene FAA. *Proc. Natl. Acad. Sci. USA* 94, 13051–13056.

Liu, Q., Feng, J., Buzin, C., Wen, C., Nozari, G., Mengos, A., Nguyen, V., Liu, J., Crawford, L., Fujimura, F.K., and Sommer, S.S. (1999). Detection of virtually all mutations-SSCP (DOVAM-S): a rapid method for mutation scanning with virtually 100% sensitivity. *Biotechniques* 26, 932.

Liu, X.Z., Xia, X.J., Adams, J., Chen, Z.Y., Welch, K.O., Tekin, M., Ouyang, X.M., Kristiansen, A., Pandya, A., Balkany, T., Arnos, K.S., and Nance, W.E. (2001). Mutations in GJA1 (connexin 43) are associated with non-syndromic autosomal recessive deafness. *Hum. Mol. Genet.* 10, 2945–2951.

MacCollin, M., Ramesh, V., Jacoby, L.B., Louis, D.N., Rubio, M.P., Pulaski, K., Trofatter, J.A., Short, M.P., Bove, C., Eldridge, R., Parry, D.M., and Gusella, J.F. (1994). Mutational analysis of patients with neurofibromatosis 2. *Am. J. Hum. Genet.* 55, 314–320.

Maillet, P., Alloisio, N., Morle, L., and Delaunay, J. (1996). Spectrin mutations in hereditary elliptocytosis and hereditary spherocytosis. *Hum. Mutat.* 8, 97–107.

Makino, R., Yazyu, H., Kishimoto, Y., Sekiya, T., and Hayashi, K. (1992). F-SSCP: fluorescence-based polymerase chain reaction-single-strand conformation polymorphism (PCR-SSCP) analysis. *PCR Methods Applic.* 2, 10–13.

Makino, R., Kaneko, K., Kurahashi, T., Matsumura, T., and Mitamura, K. (2000). Detection of mutation of the p53 gene with high sensitivity by fluorescence-based PCR-SSCP analysis using low-pH buffer and an automated DNA sequencer in a large number of DNA samples. *Mutat. Res.* 452, 83–90.

Medintz, I., Wong, W.W., Sensabaugh, G., and Mathies, R.A. (2000). High speed single nucleotide polymorphism typing of a hereditary haemochromatosis mutation with capillary array electrophoresis microplates. *Electrophoresis* 21, 2352–2358.

Medintz, I.L., Paegel, B.M., Blazej, R.G., Emrich, C.A., Berti, L., Scherer, J.R., and Mathies, R.A. (2001). High-performance genetic analysis using microfabricated capillary array electrophoresis microplates. *Electrophoresis* 22, 3845–3856.

Menounos, P., Zervas, C., Garinis, G., Doukas, C., Kolokithopoulos, D., Tegos, C., and Patrinos, G.P. (2000). Molecular heterogeneity of the glucose-6-phosphate dehydrogenase deficiency in the Hellenic population. *Hum. Hered.* 50, 237–241.

Myers, R.M., Maniatis, T., and Lerman, L.S. (1987). Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol.* 155, 501–527.

Nair, S., Lin, T.K., Pang, T., and Altwegg, M. (2002). Characterization of Salmonella serovars by PCR-single-strand conformation polymorphism analysis. *J. Clin. Microbiol.* 40, 2346–2351.

Nakabayashi, Y., and Nishigaki, K. (1996). Single-strand conformation polymorphism (SSCP) can be explained by semistable conformation dynamics of single-stranded DNA. *J. Biochem. (Tokyo)* 120, 320–325.

Nishimura, D.Y., Purchio, A.F., and Murray, J.C. (1993). Linkage localization of TGFB2 and the human homeobox gene HLX1 to chromosome 1q. *Genomics* 15, 357–364.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* 86, 2766–2770.

Papachatzopoulou, A., Menounos, P.G., Kolonelou, C., and Patrinos, G.P. (2006). Mutation screening in the human epsilon-globin gene using single-strand conformation polymorphism analysis. *Am. J. Hematol.* 81, 136–138.

Patrinos, G.P., Garinis, G., Kounelis, S., Kouri, E., and Menounos, P. (1999). A novel 23-bp deletion in exon 5 of the p53 tumor suppressor gene. *J. Mol. Med.* 77, 686–689.

Paul, P., Letteboer, T., Gelbert, L., Groden, J., White, R., and Coppes, M.J. (1993). Identical APC exon 15 mutations result in a variable phenotype in familial adenomatous polyposis. *Hum. Mol. Genet.* 2, 925–931.

Peters, T., Schlayer, H.J., Hiller, B., Rösler, B., Blum, H., and Rasenack, J. (1997). Quasispecies analysis in hepatitis C virus infection by fluorescent single strand conformation polymorphism. *J. Virol. Methods.* 64, 95–102.

Phelan, C.M., Lancaster, J.M., Tonin, P., Gumbs, C., Cochran, C., Carter, R., Ghadirian, P., Perret, C., Moslehi, R., Dion, F., Faucher, M.C., Dole, K., Karimi, S., Foulkes, W., Lounis, H., Warner, E., Goss, P., Anderson, D., Larsson, C., Narod, S.A., and Futreal, P.A. (1996). Mutation analysis of the BRCA2 gene in 49 site-specific breast cancer families. *Nat. Genet.* 13, 120–122.

Quintanar, A., Jallu, V., Legros, Y., and Kaplan, C. (1998). Human platelet antigen genotyping using a fluorescent SSCP technique with an automatic sequencer. *Br. J. Haematol.* 103, 437–444.

Ravnik-Glavac, M., Glavac, D., and Dean, M. (1994). Sensitivity of single-strand conformation polymorphism and heteroduplex method for mutation detection in the cystic fibrosis gene. *Hum. Mol. Genet.* 3, 801–807.

Rommens, J., Kerem, B.S., Greer, W., Chang, P., Tsui, L.C., and Ray, P. (1990). Rapid nonradioactive detection of the major cystic fibrosis mutation. *Am. J. Hum. Genet.* 46, 395–396.

Rossetti, S., Corra, S., Biasi, M.O., Turco, A.E., and Pignatti, P.F. (1995). Comparison of heteroduplex and single-strand conformation analyses, followed by ethidium fluorescence visualization, for the detection of mutations in four human genes. *Mol. Cell. Probes* 9, 195–200.

Saeki, Y., Ueno, S., Takahashi, N., Soga, F., and Yanagihara, T. (1992). A novel mutant (transthyretin Ile-50) related to amyloid polyneuropathy. Single-strand conformation polymorphism as a new genetic marker. *FEBS Lett.* 308, 35–37.

Sainz, J., Huynh, D.P., Figueroa, K., Ragge, N.K., Baser, M.E., and Pulst, S.M. (1994). Mutations of the neurofibromatosis gene and lack of the gene product in vestibular schwannomas. *Hum. Mol. Genet.* 3, 885–891.

Sarkar, G., Yoon, H.S., and Sommer, S.S. (1992). Screening for mutations by RNA single-strand conformation polymorphism (rSSCP): comparison with DNA-SSCP. *Nucleic Acids Res.* 20, 871–878.

Scott, D.A., Kraft, M.L., Carmi, R., Ramesh, A., Elbedour, K., Yairi, Y., Srisailapathy, C.R., Rosengren, S.S., Markham, A.F., Mueller, R.F., Lench, N.J., Van Camp, G., Smith, R.J., and Sheffield, V.C. (1998). Identification of mutations in the connexin 26 gene that cause autosomal recessive nonsyndromic hearing loss. *Hum. Mutat.* 11, 387–394.

Sheffield, V.C., Beck, J.S., Kwitek, A.E., Sandstrom, D.W., and Stone, E.M. (1993). The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* 16, 325–332.

Shen, M.H., Harper, P.S., and Upadhyaya, M. (1993). Neurofibromatosis type 1 (NF1): the search for mutations by PCR heteroduplex analysis on Hydrolink gels. *Hum. Mol. Genet.* 2, 1861–1864.

Shimizu, T., Toguchida, J., Kato, M.V., Kaneko, A., Ishizaki, K., and Sasaki, M.S. (1994). Detection of mutations of the RB1 gene in retinoblastoma patients by using exon-by-exon PCRSSCP analysis. *Am. J. Hum. Genet.* 54, 793–800.

Sözen, M.M., Whittall, R., Oner, C., Tokatli, A., Kalkanoğlu, H.S., Dursun, A., Coşkun, T., Oner, R., and Humphries, S.E. (2005). The molecular basis of familial hypercholesterolaemia in Turkish patients. *Atherosclerosis* 180, 63–71.

Stoerker, J., Hurwitz, C., Rose, N.C., Silberstein, L.E., and Highsmith, W.E. (1996). Heteroduplex generator in analysis of Rh blood group alleles. *Clin. Chem.* 42, 356–360.

Strom, T.M., Hortnagel, K., Hofmann, S., Gekeler, F., Scharfe, C., Rabl, W., Gerbitz, K.D., and Meitinger, T. (1998). Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (wolframin) coding for a predicted transmembrane protein. *Hum. Mol. Genet.* 7, 2021–2028.

Suzuki, Y., Orita, M., Shiraishi, M., Hayashi, K., and Sekiya, T. (1990). Detection of ras gene mutations in human lung cancers by single-strand conformation polymorphism analysis of polymerase chain reaction products. *Oncogene* 5, 1037–1043.

Takahashi-Fujii, A., Ishino, Y., Kato, I., and Fukumaki, Y. (1994). Rapid and practical detection of beta-globin mutations causing beta-thalassemia by fluorescence-based PCR-single-stranded conformation polymorphism analysis. *Mol. Cell. Probes* 8, 385–393.

Thomson, K.L., Gloyn, A.L., Colclough, K., Batten, M., Allen, L.I. Beards, F., Hattersley, A.T., and Ellard, S. (2003). Identification of 21 novel glucokinase (GCK) mutations in UK and European Caucasians with maturity-onset diabetes of the young (MODY). *Hum. Mutat.* 22, 417.

Tuffery, S., Moine, P., Demaille, J., and Claustres, M. (1993). Base substitutions in the human dystrophin gene: detection by using the single-strand conformation polymorphism (SSCP) technique. *Hum. Mutat.* 2, 368–374.

Upadhyaya, M., Maynard, J., Osborn, M., Huson, S.M., Ponder, M., Ponder, B.A., and Harper, P.S. (1995). Characterisation of germline mutations in the neurofibromatosis type 1 (NF1) gene. *J. Med. Genet.* 32, 706–710.

Varesco, L., Gismondi, V., James, R., Robertson, M., Grammatico, P., Groden, J., Casarino, L., De Benedetti, L., Bafico, A., Bertario, L., Sala, P., Sassatelli, R., Ponz de Leon, M., Biasco, G., Allergetti, A., Aste, H., De Sanctis, S., Rossetti, C., Illeni, M.T., Sciarra, A., Del Porto, G., White, R., and Ferrara, G.B. (1993). Identification of APC gene mutations in Italian adenomatous polyposis coli patients by PCR-SSCP analysis. *Am. J. Hum. Genet.* 52, 280–285.

Veldhuisen, B., Saris, J.J., de Haij, S., Hayashi, T., Reynolds, D.M., Mochizuki, T., Elles, R., Fossdal, R., Bogdanova, N., van Dijk, M.A., Coto, E., Ravine, D., Norby, S., Verellen-Dumoulin, C., Breuning, M.H., Somlo, S., and Peters, D.J. (1997). A spectrum of mutations in the second gene for autosomal dominant polycystic kidney disease (PKD2). *Am. J. Hum. Genet.* 61, 547–555.

Wang, Y.H., Barker, P., and Griffith, J. (1992). Visualization of diagnostic heteroduplex DNAs from cystic fibrosis deletion heterozygotes provides an estimate of the kinking of DNA by bulged bases. *J. Biol. Chem.* 267, 4911–4915.

Watson, M.L., D'Eustachio, P., Mock, B.A., Steinberg, A.D., Morse, H.C., 3rd, Oakey, R.J., Howard, T.A., Rochelle, J.M., and Seldin, M.F. (1992). A linkage map of mouse chromosome 1 using an interspecific cross segregating for the gld autoimmunity mutation. *Mamm. Genome* 2, 158–171.

White, M.B., Carvalho, M., Derse, D., O'Brien, S.J., and Dean, M. (1992). Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* 12, 301–306.

Whittall, R., Gudnason, V., Weavind, G.P., Day, L.B., Humphries, S.E., and Day, I.N. (1995). Utilities for high throughput use of the single strand conformational polymorphism method: screening of 791 patients with familial hypercholesterolaemia for mutations in exon 3 of the low density lipoprotein receptor gene. *J. Med. Genet.* 32, 509–515.

Wood, N., Tyfield, L., and Bidwell, J. (1993a). Rapid classification of phenylketonuria genotypes by analysis of heteroduplexes generated by PCR-amplifiable synthetic DNA. *Hum. Mutat.* 2, 131–137.

Wood, N., Standen, G., Hows, J., Bradley, B., and Bidwell, J. (1993b). Diagnosis of sickle-cell disease with a universal heteroduplex generator. *Lancet* 342, 1519–1520.

Wood, N., Standen, G.R., Murray, E.W., Lillicrap, D., Holmberg, L., Peake, I.R., and Bidwell, J. (1995). Rapid genotype analysis in type 2B von Willebrand's disease using a universal heteroduplex generator. *Br. J. Haematol.* 89, 152–156.

Yap, E.P., and McGee, J.O. (1992a). Nonisotopic SSCP detection in PCR products by ethidium bromide staining. *Trends Genet.* 8, 49.

Yap, E.P., and McGee, J.O. (1992b). Nonisotopic SSCP and competitive PCR for DNA quantification: p53 in breast cancer cells. *Nucleic Acids Res.* 20, 145.

Zappella, M., Meloni, I., Longo, I., Canitano, R., Hayek, G., Rosaia, L., Mari, F., and Renieri, A. (2003). Study of MECP2 gene in Rett syndrome variants and autistic girls. *Am. J. Med. Genet.* 119B, 102–107.

Chapter 5

Capillary Electrophoresis

Anja Bosserhoff[1] and Claus Hellerbrand[2]

[1]*Institute of Pathology, University of Regensburg, Regensburg, Germany;*

[2]*Department of Internal Medicine I, University Hospital Regensburg, Regensburg, Germany*

5.1 INTRODUCTION

Diagnosis of inherited diseases or cancer predisposition frequently involves analysis of specific mutations or polymorphisms. The number of characterized monogenetic and polygenetic diseases is rising significantly every year. In consequence, molecular diagnostics is faced with an increasing number of patient samples and with a rising complexity of genetic diagnostics. In order to apply genetic analyses for large groups of patients or population screening, automation of a sensitive and precise method is highly desirable. With capillary electrophoresis (CE), analytical techniques were developed, which can rapidly process large numbers of patient samples in an automated fashion. Further, in the area of proteomics separation, analysis and characterization of proteins by CE is a challenging task to understand the molecular basis of biological processes and to discover new biomarkers and drug targets.

5.2 HISTORY, PRINCIPLE, AND POTENTIAL APPLICATIONS OF CAPILLARY ELECTROPHORESIS

5.2.1 History

Electrophoretic separation of molecules in a glass tube and subsequent detection of the separated compounds by ultraviolet absorption was first described by Hjerten (1967). In his communication, separation of serum proteins, inorganic and organic ions, peptides, nucleic acids, viruses, and bacteria are described. However, electrophoresis in a tube did not become popular until 1981, when Jorgenson and Lucaks (1981) demonstrated the high-resolution power of capillary zone electrophoresis (CZE). In the late 1980s, the first commercial CE instrument was offered on the market. Since then, many advances and applications have taken place with tremendous impact on the progress of science. Thus, the acceleration and rapid success of the Human Genome Project (HGP), leading to the generation of the draft human genomic sequence in an astonishing short period of time (Lander *et al.*, 2001; Venter *et al.*, 2001) was only possible due to the introduction of CE-based sequencers allowing the production of 14.9 billion base pairs of sequence in just nine months. These analyses built the basis for genomics, e.g. systematic analysis of gene interactions. In addition, systemic analysis of proteins and metabolic pathways received increasing attention, and understanding the complex direct and indirect interaction of biological components as genes proteins and metabolites is the goal of system biology. Also here, high-throughput and dynamic analysis of biological data by CE-based techniques play a central role enabling such system biology approaches.

5.2.2 Principle

In principle, all CE techniques are carried out using the same equipment: (1) a high-voltage power supply, (2) the anode and cathode buffer reservoirs with corresponding electrodes, (3) the separation chamber, i.e. the capillary tube, (4) the injection system, and (5) the detector (Fig. 5.1).

A capillary tube is filled with buffer, placed between the two buffer reservoirs, and an electric field is applied by means of a high power supply. Then, defined volume is introduced into the capillary by replacing one buffer reservoir with the sample vial. A detector is located at the opposite side of the capillary at the injection site. The sample data are collected and stored by computer, and analyzed using dedicated software.

CE is based on the movement of molecules in an electric field. Initially, simple buffer solutions were employed to separate ionic water-soluble solutes. However, in contrast to classical electrophoresis, CE is not restricted to the separation of large molecules based on size and charge. Several modifications of CE have been developed to separate molecules

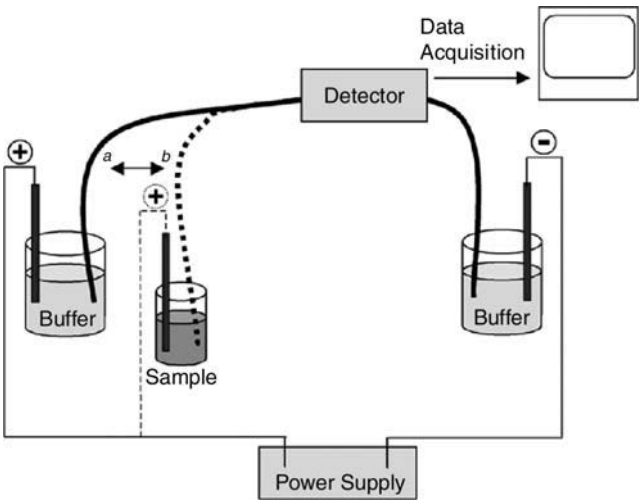


FIGURE 5.1 Principle of capillary electrophoresis. A capillary tube is placed between the two buffer reservoirs, and an electric field is applied by means of a high power supply (a). Then, a defined volume of the analysate is introduced into the capillary by replacing one buffer reservoir with the sample vial (b). The result of the electrophoretic separation is measured by a detector, and the sample data are collected and stored on a computer.

TABLE 5.1 Subtypes of the CE methodology.

Subtype	Abbreviation
Capillary zone electrophoresis	CZE
Capillary gel electrophoresis	CGE
Micellar electrokinetic chromatography	MEKC
Capillary isoelectric focusing	CIEF
Capillary isotachophoresis	CITP
Constant denaturant capillary electrophoresis	CDCE

with low molecular weight or neutral charge (some of them summarized in Table 5.1).

Capillary zone electrophoresis (CZE) is the simplest form of CE. A capillary is filled with a relatively low viscosity buffer and the analysates migrate from one end of the capillary to the other with velocities determined by the charge-to-mass ratio to form discrete peaks (zones). In contrast, *capillary gel electrophoresis* (CGE) separates analytes, such as DNA or proteins, based on molecular weight. A variety of compounds, such as bis-polyacrylamide, agarose, or methylcellulose, have been used to act as a molecular sieve to cause separation. Terabe (2004) introduced *micellar electrokinetic chromatography* (MEKC) for the separation of neutral compounds by the addition of a micelle sodium dodecyl sulfate (SDS) to the buffer solution. Similar to classical isoelectric focusing, in *capillary isoelectric focusing* (CIEF) a pH gradient is created inside a capillary to separate peptides and proteins on the basis of their isoelectric point. In *capillary isotachophoresis* (CITP) a combination of two buffers with different mobilities is used causing the analytes of interest to be concentrated in zones between the leading and terminating constituents.

Constant denaturant capillary electrophoresis (CDCE) permits high-resolution separation of single-base variations occurring in an approximately 100 bp isomelting DNA sequence based on their differential melting temperatures (Li-Sucholeiki *et al.*, 1999). It requires different degrees of partial denaturation at a given temperature in the presence of denaturing agents (Khrapko *et al.*, 1994).

5.2.3 Applications

CE has been successfully used in a wide range of application areas. Routine methods have been established in a number of application areas including the analysis of pharmaceuticals, DNA, chiral compounds, proteins, peptides, clinical and forensic samples, metal ions, and inorganic anions.

For clinical diagnostics, CE is widely used for analyzing proteins in physiological matrices, such as serum, urine, and cerebrospinal fluid. Furthermore, CE is used in clinical and molecular biology settings for DNA analysis and also for forensic and therapeutic drug screening. In addition to clinical applications, CE finds wide use in the analysis of pesticides, food content and composition, and pollutants in water and soil samples.

More and more, CE is challenging the use of traditional gel electrophoresis methods. CE especially benefits from high resolution, simplicity, versatility, low operating costs and the possibility of direct sample injection without complex sample pretreatment.

CE is characterized by high sensitivity, allowing analysis of small amounts of samples with a minimal amount of reagents. The rugged nature of the CE capillary allows it to be rinsed and cleaned between injections with relatively harsh solutions, such as NaOH. This can allow direct injection of biofluids with possible reductions in sample pretreatment requirements, which are attractive in clinical chemistry where sample throughput is high. High-throughput screening involves testing of a large number of samples in a rapid and often automated fashion. Due to the possibility for automation and parallelization, CE allows fast, large-scale and high-throughput analysis. Sophisticated CE autosamplers are available that allow quantitative injection and analysis of a large number of samples in an unattended sequence.

Furthermore, in molecular diagnostics, CE has several advantages over classical techniques like slab polyacrylamide gel electrophoresis (PAGE), Southern blotting, sequencing or conventional gel electrophoresis. CE has the ability to analyze not only non-ideal tissue samples such as archival paraffin embedded formalin fixed tissue, but also automated digital imaging capabilities that can use either peak height or peak area in a semi-quantitative or quantitative manner.

5.3 CAPILLARY ELECTROPHORESIS IN MOLECULAR DIAGNOSTICS

Nucleotides have been quantified in different matrices, including tissue and cell extracts and several DNA and RNA sources. Therapeutic antisense oligonucleotides are of interest for many applications and CE can be used to characterize and quantify these materials (Righetti and Gelfi, 1998).

DNA separations are generally conducted by CGE, in which gel- or polymer-filled capillaries act as sieving media to resolve the different lengths of DNA. DNA of different lengths has similar electrophoretic mobility as each increase in size is accompanied by a corresponding increase in the number of negative charges. Therefore, the separation by mobility differences is not the preferred approach and the majority of separations are achieved using a sieve mechanism. The capillary is filled with a matrix of synthetic or natural polymer. The various DNA fragments migrate through this matrix and become entangled with or trapped in the matrix. The migration of the larger DNA fragment is retarded to a greater extent, which results in a size-based separation mechanism.

A variety of compounds, such as bis-polyacrylamide, agarose, or methylcellulose, have been used to act as a molecular sieve to cause separation (Ren, 2000). One of the most active research areas for compound separation has been the investigation of alternative sieving matrices to replace gels (Cottet *et al.*, 1998; Magnusdottir *et al.*, 1998). Recently, a nanostructured copolymer matrix has been successfully used to separate oligonucleotides with high resolution by CE using a very short separation channel which simulates real microchip conditions (Zhang *et al.*, 2006).

Generally, it is difficult to prepare gel-filled capillaries manually due to bubble formation. Further, gels have a limited lifetime, since they are easily destroyed by high current and joule heating. CE requires a liquid medium, and the best choice appears to be a semi-dilute solution of polymers, which is very much similar to gel, since the molecules are entangled with one another in a way that they create a fine mesh (Sunada and Blanch, 1997). Entangled polymer solutions (ESCE, entangled polymer solutions CE) can be replaced after each run and are therefore much easier to operate. Further, *capillary non-gel sieving electrophoresis* (CNGSE) has been employed in the biological sciences for the size-based separation of macromolecules such as nucleic acids and facilitates the application of CE for detection of mutations and polymorphisms in human molecular diagnostics.

5.4 MODES OF APPLICATION

Besides gene sequencing most protocols adapted to automated CE represent analyses of DNA fragment length or DNA restriction patterns (RFLP), analyses of single-strand conformation polymorphism (SSCP, see also previous chapter), and microsatellite analyses.

5.4.1 Sequencing

Sequence analysis is the gold standard for molecular genetic testing. It yields the greatest amount of information since it identifies the order of each deoxynucleotide base of a particular target, usually amplified DNA or cDNA. Although PCR techniques made sequencing routine, detection and analysis of the resulting DNA fragments became the major bottleneck for large-scale DNA sequencing (Dovichi, 1997; Schmalzing *et al.*, 1999). Although the identification of the human genome sequence is completed, the importance of DNA sequencing has not been diminished. Sequencing of other organisms' genomes, comparative genomics or screening for human genetic defects are still challenging fields of application for the future.

DNA sequencing using dye-labeled didesoxynucleotides and the method developed by Sanger and coworkers (1992) were adapted to CE. The single-base resolving capability of CE permits all four fluorochrome products to be separated simultaneously compared to classical sequencing with PAGE, where each chain terminator has to be loaded in separate lanes. The result is a multicolored ladder where each color represents a different base. The order of the bases is then analyzed using secondary software that characterizes the sequence in terms of identity, and relatedness to prototypical sequences in a database. Today CE serves as the routine high-throughput technique for sequencing. CE systems with multi-array capillaries (Tan and Yeung, 1998; Wu *et al.*, 2008; Zhang *et al.*, 1999) will further enhance high throughput.

5.4.2 Analyses of DNA Fragment Length or Restriction Patterns and Microsatellites

Cleavage of DNA products and analysis of the DNA cleavage pattern by CE is currently widely used to detect gene mutations (Andersen *et al.*, 1998; Sell and Lugemwa, 1999). The method is applicable in cases where, due to mutations or polymorphisms, restriction sites are lost or gained.

In contrast to sequencing, fragment analysis incorporates one fluorochrome-labeled PCR primer in a standard PCR reaction after which the product(s) are separated and visualized. Fragment analysis is more rapid and less expensive because the fluorochrome is incorporated in the initial PCR reaction thus eliminating the need for sequencing reactions that incorporate the dye terminating bases. Interpretation of the fragment analysis data is also easier than sequence analysis because the fragments are readily identified by the imaging software and the need for secondary sequence software analysis is not required.

Using this technique, some research groups established protocols for high-speed separation that allow distinguishing between wild-type and mutant PCR products extremely rapidly (Chan *et al.*, 1996) with a short effective length capillary and high field strength.

While less expensive and more rapid, fragment analysis does not provide exact, detailed sequence data. In addition,

PCR fragments can result from erroneous amplification and thus verification using sequencing techniques or independent probes is sometimes recommended for quality assurance purposes. In these cases, sequencing is often used to confirm the identity of the PCR product. Because of problems due to incomplete cleavage the method is being replaced by SSCP where possible.

Furthermore, microsatellite analysis was originally performed using gel electrophoresis and radioactively labeled PCR products of defined areas of the genome. Application on CE made this method faster and more reliable (see paragraph below).

5.4.3 Analyses of Single-Strand Conformation Polymorphism

Genetic diagnosis of an inherited disease or cancer predisposition often involves search for unknown point mutations in several genes. SSCP, a method exploiting visualization of different secondary structures under native conditions combined with CE, was shown by many groups to be a rapid and automated technique for processing large numbers of samples (see also previous chapter). Larsen and coworkers (1999) used the method to detect point mutations associated with the inherited cardiac disorders long QT syndrome (LQTS) and hypertrophic cardiomyopathy (HCM). Sensitivity has been reported to be almost 100%, when 34 different point mutations were analyzed, while 10 previously unknown variants were found. These results clearly demonstrate that the method has high resolution, good reproducibility, and is very robust. Based on the possibility of automation and short time of analysis, the method should be suitable for high-throughput applications such as genetic screening of large populations. Several other groups have already established this method for specific screening of mutations in other diseases (see below).

5.4.4 Other Applications

CDCE has been combined with high-fidelity DNA amplification and automated multicapillary instrument with fraction collection allowing sensitive, automated, and cost-effective analysis of healthy human tissues and population screening for disease-associated single nucleotide polymorphisms (SNPs) in large pooled samples (Muniappan and Thilly, 1999; Li *et al.*, 2005). Further applications of CDCE include studies of somatic mitochondrial mutations with respect to aging and measurement of mutational spectra of nuclear genes. The technique of CDCE was also shown to be applicable for detection of mutations in the K-ras and N-ras genes (Kumar *et al.*, 1995; Bjorheim *et al.*, 1998).

Kuypers and coworkers (1996) developed a method for online melting of double-stranded DNA for SSCP-CE, while they further improved the integration of PCR and CE by composing a contamination-free, automated PCR in the CE apparatus (Kuypers *et al.*, 1998).

Further, RT-PCR followed by CE had been used to quantify the expression levels of specific gene products (Borson *et al.*, 1998; Butler, 1998; Odin *et al.*, 1999). Bor and coworkers (2000) described a protocol for simultaneous quantification of several mRNA species, after calibrating RT-PCR with CE. In a study of Schummer and coworkers (1999), the clinical relevance of multidrug resistance-associated protein (MRP) gene expression was correlated with chemoresistance in prostate carcinoma. De Cremoux and coworkers (1999) compared the *c-erbB-2* gene amplification in breast cancer with the expression of c-erbB-2 protein evaluated by immunohistochemistry and found a concordance of 91% between the two techniques. They concluded that *c-erbB-2* gene amplification can be accurately quantitated by competitive PCR followed by CE and that this method is also suitable for small, fixed tissue specimens. Stanta and Bonin (1998) also used CE to quantify specific RT-PCR products and were able to show that the quantitation by CE is more reliable than by dot-blot.

Finally, Kuypers and coworkers (1994) compared the use of CE with that of slab gel electrophoresis for quantification of chromosomal translocations in lymphoma, showing that the results of both methods are comparable.

5.5 SPECIFIC DIAGNOSTIC APPLICATIONS

5.5.1 Diagnosis of Neoplastic Disorders

In the diagnosis of neoplastic disorders, CE has been used for the detection of chromosomal aberrations, microsatellite instability, clonality assays, and the detection of several other cancer relevant genetic mutations, like SNPs.

Colorectal cancer is the leading cause of death related to cancer in western countries (Weir *et al.*, 2003). Molecular biology studies have led to the identification of two broad categories of molecular alterations in colorectal cancer. Loss of heterozygosity (LOH) represents 80% of colorectal cancers and is characterized by aneuploidy or allelic losses. The second group displays phenotypic microsatellite instability (MSI-positive tumors), has a near-diploid karyotype and a relatively low frequency of allelic losses. It accounts for 15% of all colorectal cancers. As a consequence of these two phenomena, other specific genetic events occur at high frequency. These include inactivation of tumor suppressor genes by genetic (deletion or mutation) or epigenetic events (Garinis *et al.*, 2002), activation of proto-oncogenes by mutation, and dysregulated expression of several other molecules known to be involved in the development of colorectal cancer.

5.5.1.1 Loss of Heterozygosity

Aneuploidy indicates gross losses or gains in chromosomal DNA and is often seen in many human primary tumors and

premalignant conditions. Loss of one allele at a chromosomal locus may imply the presence of a tumor suppressor gene at that site. Loss of both alleles at a given locus (homozygous deletion) is an even stronger indicator of the existence of a tumor suppressor gene. Many of these loci are already associated with one or more known candidate tumor suppressor genes, including 17p (*p53* gene), 5q21 (*APC* gene), 3p21 (β-catenin (*CTNNB1*) gene), 9p (*p16* and *p15* genes), and 13q (*RB* gene).

LOH analysis was originally performed using gel electrophoresis and radioactively labeled PCR products of defined areas to the genome. The use of CE made this analysis more reliable and faster (Canzian *et al.*, 1996). CE was used for the detection of the loss of several tumor suppressor genes in micro-dissected cancerous tissue (Marsh *et al.*, 2003). Gene fragments were amplified using PCR with flanking oligonucleotides bearing fluorescent labels and subsequently separated and analyzed by CE.

The *p53* gene locus is the commonest site demonstrating loss of heterozygosity. TP53 is a DNA binding protein, which is a transcriptional activator and can cause cell cycle arrest in response to DNA damage.

A second tumor suppressor gene adenomatous polyposis coli (*APC*) is inactivated in more than 80% of early colorectal cancers. An important function of the *APC* gene is to prevent the accumulation of molecules associated with cancer, such as catenins. Familial adenomatous polyposis (FAP) is an autosomal dominant disease caused by germline mutations in the *APC* gene. FAP is a rare condition in which hundreds or thousands of polyps develop along the length of the colon, and, if left untreated, lead to colon cancer. Figure 5.2 depicts LOH of the *APC* locus in tumor compared to normal tissue.

An additional application of LOH assays is the identification of chromosomal differences between normal and tumor tissue. This is useful in distinguishing between tumor recurrence vs. *de novo* cancer formation (Rolston *et al.*, 2001; Sasatomi *et al.*, 2002).

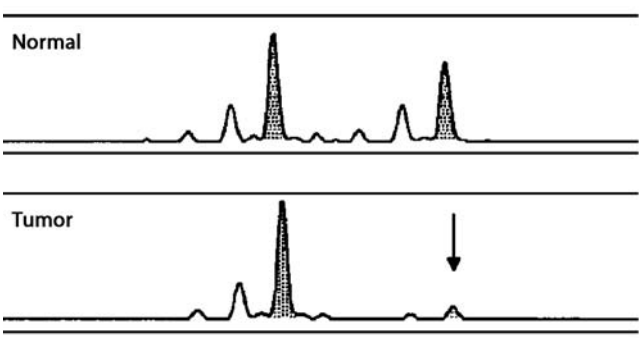


FIGURE 5.2 Analysis for LOH by CE. Genomic DNA is isolated from normal or tumor tissue, amplified with specific primers for defined chromosomal loci and the PCR products are subsequently separated by CE. The two alleles are represented by two peaks in DNA isolated from normal tissue, automatically marked with gray color by the analysis software. In contrast, in DNA from the tumor tissue, one peak is missing (arrow).

5.5.1.2 Microsatellite Instability

Microsatellite instability (MSI), or replication error, comprises length alterations of oligonucleotide repeat sequences that occur somatically in human tumors. They are the manifestation of genomic instability where tumor cells have a decreased overall ability to faithfully replicate DNA.

MSI is a frequent, if not obligatory, surrogate marker of underlying functional inactivation of one of the human DNA mismatch repair genes (*MLH1*, *MSH2*, *MSH6*, *PMS1*, *PMS2* (reviewed in Jacob and Praz, 2002)). DNA mismatch repair enzymes normally remove misincorporated single or multiple nucleotide bases as a result of random errors during recombination or replications. Functional loss of mismatch repair occurs due to biallelic inactivation via combination of gene mutation, LOH and/or promoter methylation (Herman *et al.*, 1998; Jacob and Praz, 2002).

Germline mutation of a mismatch repair gene has been shown to be the autosomal dominant genetic defect in most hereditary non-polyposis colorectal cancer (HNPCC) patients (Marra and Boland, 1995; Saletti *et al.*, 2001). A second hit incurred in tumor cells in HNPCC individuals results in biallelic inactivation of the specific *MMR* gene. This results in loss of faithful replication of microsatellite DNA in tumor (Saletti *et al.*, 2001). Bethesda criteria (Henson *et al.*, 1995) have been outlined to guide the identification of this syndrome. However, finding MSI positive tumors was shown to be the best predictor of germline mutation (Liu *et al.*, 1999). Although implicating a germline defect in HNPCC patients, MSI is also found in 15–20% of sporadic colon cancer (Goel *et al.*, 2003), where the finding reflects an overall increase in genomic instability.

CE has been used for microsatellite analysis as well as for sequencing of DNA mismatch repair genes. Figure 5.3 illustrates mutation detection in the human *MSH-2* gene, using CE-based DNA sequencing, while Fig. 5.4 depicts the result of the analysis of two marker regions of MSI, comparing normal and tumor tissue.

Many different microsatellite markers and loci have been used to identify MSI in tumors. In 1997, the US National Cancer Institute (NCI) recommended a panel of microsatellite markers for use in colorectal cancer MSI testing, and the continued use of the MSI marker panel is still recommended (Boland *et al.*, 1998; Muller *et al.*, 2004). Berg and coworkers (2000) described a fluorescent multiplex PCR-capillary electrophoresis (FM-CE) assay that permits the simultaneous detection of all five loci proposed by the NCI. Further, a protocol for SSCP-CE screening for alterations in the exon of the MSH2 gene leading to the possibility of high-throughput screening has been published (Merkelbach-Bruse *et al.*, 2000). This protocol significantly reduces time and expenses, compared to conventional sequence determination.

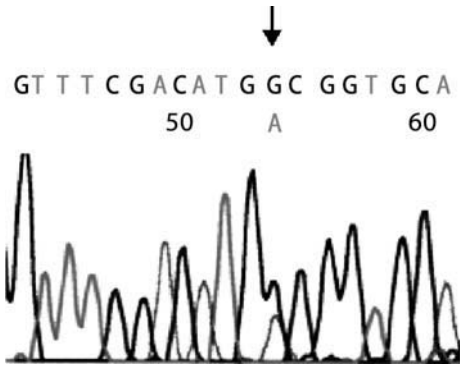


FIGURE 5.3 Mutation analysis by sequencing using CE. Genomic DNA is amplified, followed by a second round of amplification performed with deoxynucleotides along with limiting amounts of chain terminating dideoxynucleotides, labeled with different fluorophores on each base. This second PCR reaction results in a series of differently colored, truncated DNA products due to a random incorporation of the dideoxynucleotides in competition with the normal deoxynucleotides. Subsequently, these products are separated by CE. The result of the sequence analysis of 19 bp of the coding region of the *MSH2* gene is shown. Clearly, at one position (arrow) two peaks appear, indicating the presence of a mutation (G to A in one allele) in heterozygous state. (Courtesy of W. Dietmaier, University of Regensburg, Germany.)

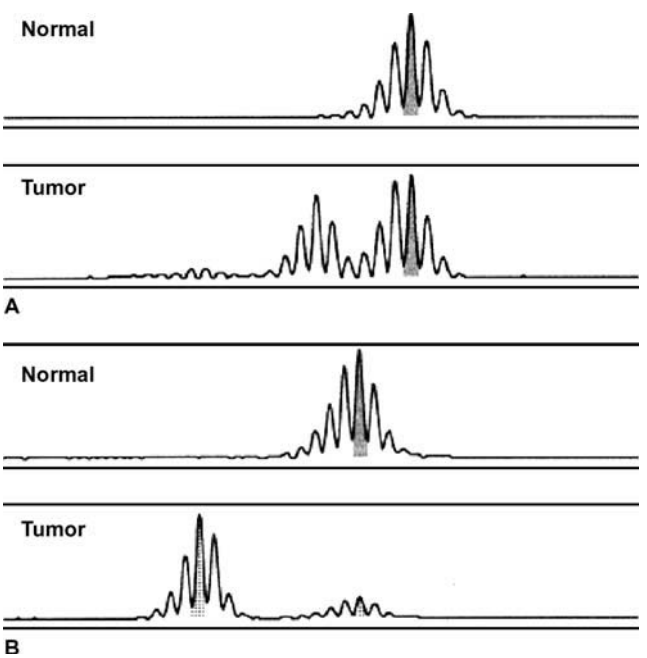


FIGURE 5.4 Analysis of microsatellite instability by CE. Detection of microsatellite instability is performed using genomic DNA isolated from normal or tumor tissue. Defined chromosomal loci are amplified by PCR and the PCR products are separated by CE. Panels A and B depict the results of the analysis of two different genetic loci. In both cases, abnormalities in the displayed CE pattern of the tumor are detectable, in comparison to the normal tissue. (Courtesy of W. Dietmaier, University of Regensburg, Germany.)

5.5.1.3 Monoclonality Assays

Rearrangements of antigen receptor genes in B- and T-cells generate products of unique length and sequence. To identify clonal lymphocyte populations in the majority of clonal B- and T-cell malignancies, B- and T-cell clonality assays have been developed. Monoclonal populations of B-cells are detected through analysis of changes in the Ig heavy chain or kappa chain. T-cell monoclonality can be detected by amplifying T-cell gamma or beta receptors.

Multiple primer sets are needed to detect clonal rearrangements or chromosomal translocations within these antigen receptor loci. Using standard methods, products of the individual reactions must be analyzed separately, and small clonal populations remain difficult to identify. Miller and coworkers (1999) developed an integrated fluorescence-based approach using CE to increase amplicon resolution, analytical sensitivity, and overall assay throughout. The newly developed method showed 94% agreement with individual B- and T-cell PCR assays, and had an overall monoclonal detection rate of almost 100%. Munro and coworkers (1999) also reported a microarray-based CE assay for T- and B-cell proliferative disorders, which again dramatically reduced time without loss of diagnostic accuracy.

5.5.1.4 Analysis of Tumor-Related Mutations

In addition to chromosomal and allelic alteration, CE can be used also to detect other genetic alterations with implications on cancer development and prognosis. Wenz and coworkers (1998) have proposed to use CE for detection of p53 mutations and have shown in a small study that all 10 samples could be correctly identified as mutated or not. Furthermore, different kinds of mutations could be separated. Also, Atha and coworkers (1998) and Kuypers and coworkers (1993) identified known mutations in the *p53* gene by SSCP-CE, while Ekstrom and coworkers (1999) used the system of CDCE to identify mutations in exon 8 of *p53* by heteroduplex analysis (see also previous chapter).

5.5.1.5 Single Nucleotide Polymorphisms

Detection of SNPs in cancer relevant genes can have important clinical implications. For example, TP53, which is associated with a "normal" genotype, has been shown to predict prolonged survival of patients with certain tumors (Bandoh *et al.*, 2002; Kandioler *et al.*, 2002).

Furthermore, determination of SNPs can help to identify the most appropriate treatment of certain types of cancers, as shown for the treatment with a specific tyrosine kinase inhibitor in gastrointestinal stromal tumors expressing mutant c-kit (Miettinen *et al.*, 2002). In addition, analysis of SNPs is used for assessing the risk of developing cancer. *BRCA1* and *BRCA2* variants can be used to help identify women with enhanced risk for developing breast or ovarian cancer (Nicoletto *et al.*, 2001).

Matrix metalloproteinases (MMPs) are a family of closely related enzymes that degrade the extracellular matrix. MMPs are implicated in connective tissue destruction during cancer invasion and metastasis of tumor cells. A guanosine

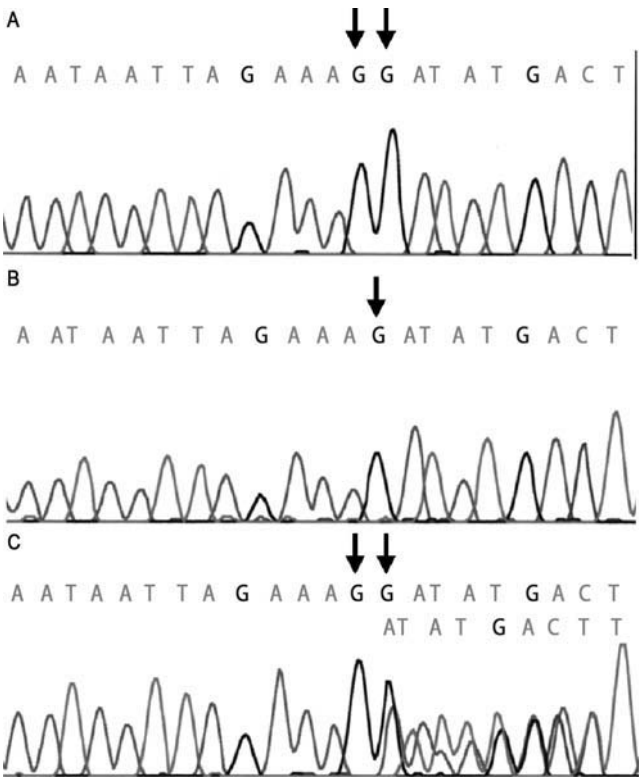


FIGURE 5.5 Analysis of polymorphisms using CE. The results of three different CE analyses for a known SNP (insertion of a guanine (G) nucleotide at −1607 bp in the *MMP-1* promoter sequence) is shown. In panel **A.** Homozygosity for the insertion of a G is shown. In panel **B.** The wild-type sequence with only one G is depicted. Finally, in panel **C.** The analysis reveals a heterozygous genotype. Note the frameshift in the heterozygous sample, as depicted by the appearance of double peaks immediately adjacent to the insertion site (indicated by the arrows).

insertion/deletion polymorphism within the promoter region of *MMP-1* influences the transcription of the gene (the insertion-type (2G) promoter possesses higher transcriptional activity than the deletion-type (G) promoter). The 2G-genotype has been found to be a genetic risk factor for the development of several cancers, including colon cancer (Ghilardi *et al.*, 2001), ovarian carcinoma (Kanamori *et al.*, 1999), and lung cancer (Zhu *et al.*, 2001), and was found to be associated with a bad prognosis in several tumors. Different *MMP-1* genotypes can be identified by CE-based DNA sequencing, as shown in Fig. 5.5.

5.5.2 Diagnosis of Hereditary Diseases and Prenatal Testing

Mutation detection using CE-based approaches is nowadays widespread in diagnostic laboratories. For obvious reasons, only a fraction of the available mutation detection approaches will be briefly outlined below and in Table 5.2.

Fragile X syndrome is a common form of inherited mental retardation with an incidence of 1 in 4,000 to

TABLE 5.2 Indicative list of a number of molecular diagnostic applications of CE.

Diagnostic applications	References
1. Neoplastic disorders	
LOH detection	Canzian *et al.* (1996), Miller *et al.* (1999)
Microsatellite instability	Berg *et al.* (2000), Merkelbach-Bruse *et al.* (2000)
Analysis of monoclonality	Miller *et al.* (1999), Munro *et al.* (1999)
Detection of tumor-related mutations (SNPs)	Nicoletto *et al.* (2001), Bandoh *et al.* (2002), Kandioler *et al.* (2002)
2. Diagnosis of hereditary diseases and prenatal testing	
Hereditary hemochromatosis	Lupski *et al.* (1991), Bosserhoff *et al.* (1999)
Fragile X syndrome	Barta *et al.* (2001)
Thalassemia	Geisel *et al.* (1999)
Congenital adrenal hyperplasia	Trent *et al.* (1998)
Charcot-Marie-Tooth 1A	Tsui *et al.* (1992a), Thormann *et al.* (1999)
Cystic fibrosis	Tsui *et al.* (1992b)
Huntington disease	Williams *et al.* (1999)
Risk for coronary heart diseases or thrombosis	Baba *et al.* (1995), Benson *et al.* (1999), Sell and Lugemwa (1999)
3. Diagnosis of infectious diseases	
Bacterial infections	
Mycobacterial species	Hernandez *et al.* (1999)
Pseudomonas sp./gram-negative non-fermenting bacteria	Ghozzi *et al.* (1999)
Listeria sp.	Sciacchitano *et al.* (1998)
Viral diseases	
Herpes simplex	Pancholi *et al.* (1997)
Hepatitis C virus	Pancholi *et al.* (1997), Doglio *et al.* (1998)
HIV-1	Kolesar *et al.* (1997)
4. Identity testing	
Forensic applications	La Fountain *et al.* (1998), Pouchkarev *et al.* (1998)
Paternity testing	
Identifying of suitable recipients of organ transplantation	
Bone marrow engraft analysis	

5,000 males. Almost all cases are caused by expansion of a $(CGG)_n$ trinucleotide repeat within the 5′ untranslated region of the *FMR1* (fragile X mental retardation) gene transcript. Until today the disease was reliably diagnosed by Southern blotting, requiring large samples and high input of time. Larsen and coworkers (1997) employed automated CE for accurate and high-throughput analysis of the FRAXA $(CGG)_n$ region in the normal and permutation range. Their method is based on PCR amplification of extracted genomic DNA followed by automated CE and detection of multicolor fluorescence. The method proved to be useful in both research and clinical mutation screening when a large number of samples, predominantly in the normal range of amplification, are to be analyzed. Most recently, Strom and coworkers (2007) developed a new method called capillary Southern analysis that allows automated high-throughput screening for *FMR1* alleles. Initially, samples are analyzed by a multiplex PCR that contains an internal control to establish gender. Only females homozygous at the *FMR1* locus are further analyzed by capillary Southern analysis. Theoretically, this method can detect expansion as high as 2,000 CGG repeats, although Strom and coworkers (2007) found the largest non-mosaic *FMR1* present in their series was 950 CGG repeats.

Also, Huntington disease (HD) belongs to the group of neurodegenerative disorders characterized by unstable expanded trinucleotide repeats. In the case of HD the expansion of a CAG repeat occurs in the IT15 gene. Williams and colleagues have established CE analysis for sizing CAG repeats and showed that it enables confident use in sizing HD alleles (Williams *et al.*, 1999).

Further applications of CE involve multi-allelic-specific amplification in the analysis of the 21-hydroxylase gene for patients with *congenital adrenal hyperplasia* (Barta *et al.*, 2001), and prenatal diagnosis of β-thalassemia, one of the most common recessive inherited disorders where many different mutations need to be detected (Trent *et al.*, 1998). Recently, a chip-based capillary electrophoresis detecting system was described that facilitates the rapid and sensitive prenatal diagnosis of β-thalassemia (Hu *et al.*, 2008).

Geisel and coworkers (1999) described an SSCP-CE method to screen for unknown mutations in the low-density lipoprotein (*LDL*) receptor gene. They PCR-amplified the promoter region as well as all 18 exons and tested the accuracy of the developed technique by reproducing 61 known genetic variations by a distinct abnormal SSCP pattern.

The c.C677T mutation of the methylentetrahydrofolate reductase (*MTHFR*) gene is a nutrient-oriented mutation that is associated with elevated levels of homocysteine and an increased risk of *coronary heart disease*. An optimized assay for automated PCR-RFLP genotyping of the MTHFR gene was established by Sell and Lugemwa (1999). Following amplification, the resulting PCR product was digested by the *HinfI* restriction enzonuclease, and the resulting fragments were analyzed by CE. The method was shown to be suitable for high-throughput screening and will support the screening of large sample sizes. Saffroy and coworkers (2002) introduced a multiplex analysis of mutations in factor V Leiden (c.G1691A), prothrombin (c.G20210A), and 5,10-methylenetetrahydrofolate reductase (c.C677T), which have been associated with an enhanced risk of thrombosis, using a CE setting. This technique can be applied to specimens from large clinical trials and epidemiological surveys.

Hereditary hemochromatosis (HH) represents an autosomal recessive disease in which increased iron absorption causes iron overload and irreversible tissue damage. Two point mutations in the *HFE* gene on chromosome 6p have been found to be associated with HH and led to the possibility of patient screening before the onset of irreversible tissue damage. Jackson and coworkers (1997) have developed a heteroduplex analysis using capillary electrophoresis for the detection of the p.C282Y mutation. An SSCP-CE approach has been recently adapted for the detection of point mutations in codon 63 or 282 of HH patients (Bosserhoff *et al.*, 1999), indicating that SSCP-CE is a reliable, cost-effective, sensitive, and rapid method for genotyping *HFE* mutations (Fig. 5.6). Nevertheless, CE can be performed equally well for RFLP analysis in order to provide diagnosis of *HFE* gene mutations (Fig. 5.7).

The hereditary Charcot Marie Tooth 1 A neuropathy (CMT1A) and hereditary neuropathy with liability to pressure palsies (HNPP) are caused by a duplication and a deletion, respectively, at chromosome 17p11.2–p12 encompassing the peripheral myelin protein 22 (*PMP22*) gene. Multiplex PCR followed by capillary electrophoresis provides a rapid and reliable detection system for duplications/deletions of the *PMP22* gene (Lin *et al.*, 2006).

Finally, several commercially available kits have been developed for the diagnosis of mutations in the cystic fibrosis gene applying CE base techniques (Tomaiuolo *et al.*, 2003). One is based on multiplex oligonucleotide ligation assay and allows the screening of 31 different mutations in the *CFTR* gene, corresponding to approximately 95% of the mutated *CFTR* alleles (Zielenski *et al.*, 2002).

5.5.3 Diagnosis of Infectious Diseases

Diagnosis of infectious diseases is a fast growing field for CE applications. In general, molecular methods do not require the presence of viable organisms permitting the identification of bacteria, viruses, and fungi that are difficult if not impossible in culture. Molecular identification of infectious agents can be used for both diagnostic and therapeutic purposes and CE has the main advantages of including higher throughput and sensitivity than conventional methods.

Chronic hepatitis C virus (HCV) infection is a worldwide public health problem with a global prevalence of 2%. Attallah and coworkers (2004) proved that CZE provides

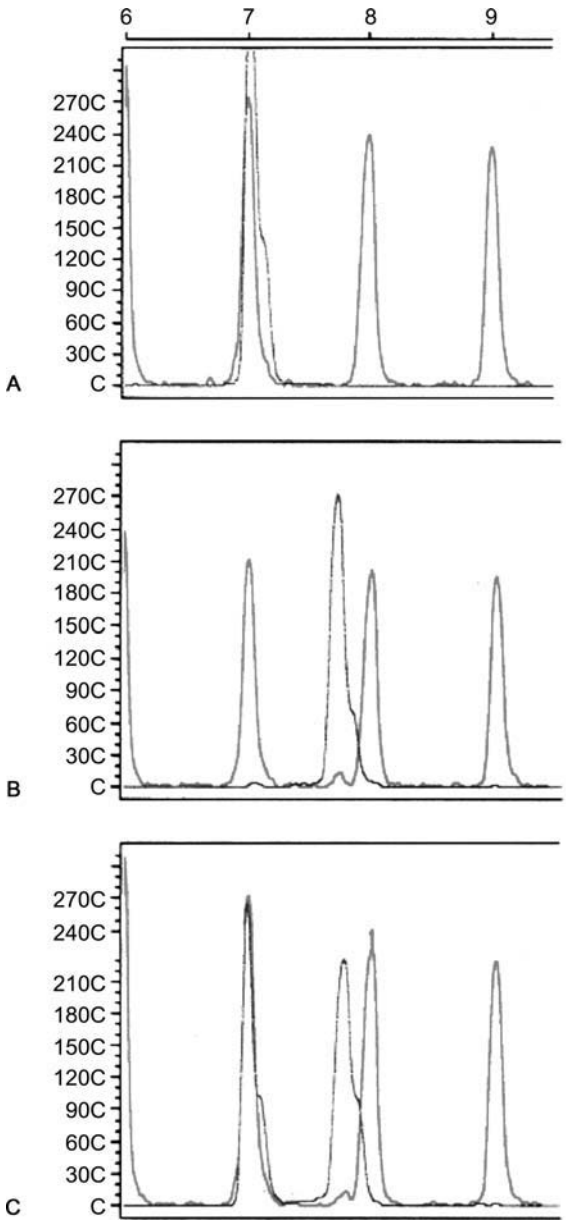


FIGURE 5.6 Analysis of codon 282 of the *HFE* gene in HH patients with a combined SSCP-CE analysis. Three typical SSCP profiles of codon 282 PCR fragments are shown. Panel **A.** Shows the analysis of a normal individual. Panel **B.** Displays the result of a patient with the mutation at codon 282 of the *HFE* gene in the homogyzous state (one peak at a different position than in the wild-type sequence). In panel **C.** A heterozygous genotype is seen (two peaks). Together with the PCR products analyzed, the HD-400-ROX standard was applied, clearly allowing the identification of the characteristic peaks of the individual genotypes.

a rapid and inexpensive method for diagnosis and mass screening of a large number of HCV-infected individuals. Further, since it has been shown that different HCV genotypes are associated with distinct profiles of pathogenicity and responses to antiviral treatment, demand for HCV genotyping has increased. Doglio and coworkers (1998) developed a CE-based detection mode, which in combination with direct cycle sequencing provides a simple and rapid method for routine HCV genotyping. More recently, temperature gradient capillary electrophoresis (TGCE) was introduced as a rapid and inexpensive method for genotyping of HCV that does not require sequencing (Margraf *et al.*, 2004).

Kolesar and coworkers (1997) used CE coupled to a laser-induced fluorescence technique (CE-LIF) to directly quantify HIV-1 RNA. They developed a fluorescent-labeled DNA probe with optimal stability and sensitivity for RNA hybridization of HIV-1 RNA isolated from plasma and showed that as little as 19fg of HIV RNA could be reliably and quantitatively detected.

The technique of enterobacterial repetitive intergenic consensus (ERIC)-PCR produces genomic DNA fingerprints that allow discrimination between bacterial species and strains. Sciacchitano (1998) applied this technique coupled to CE to differentiate *Listeria monocytogenes*, an important foodborne pathogen implicated in numerous cases of listeriosis.

Hernandez and coworkers (1999) developed a scheme for rapid identification of *Mycobacterium* sp. based on a combined restriction enzyme-CE analysis of PCR-amplified DNA and showed this detection method to be comparable to conventional methods for identification of mycobacteria.

SSCP-CE was used by Ghozzi and coworkers (1999) to rapidly identify *Pseudomonas aeruginosa* and other gram-negative non-fermenting bacilli from patients with cystic fibrosis. These authors have shown that this approach is suited for rapid identification of the main gram-negative non-fermenting bacteria.

In addition to diagnosis of infectious diseases, CE-based techniques such as micellar electrokinetic chromatography are used for the analysis of antibiotics and antiviral drugs.

5.5.4 Identity Testing and Forensic Applications

In 1985, multi- and single-locus DNA probes for the detection of RFLPs were first applied to identify specific individuals. Later, the introduction of PCR for specific amplification of short tandem repeat (STR) polymorphisms (microsatellites) represented a major breakthrough, and STR typing methods are widely used today for human identity testing including forensic DNA analysis. Following multiplex PCR amplification, DNA samples containing the length-variant STR alleles are typically separated by capillary electrophoresis and genotyped by comparison to an allelic ladder (Butler, 2007).

Currently, 6–16 fluorescently labeled STR loci are analyzed simultaneously with a single PCR amplification. As part of multiplex PCR kits, sex determination of forensic samples can be obtained using CE-based analysis of the X–Y homologous gene amelogenin (La Fountain *et al.*, 1998; Pouchkarev *et al.*, 1998). The analysis of 6–10 STRs

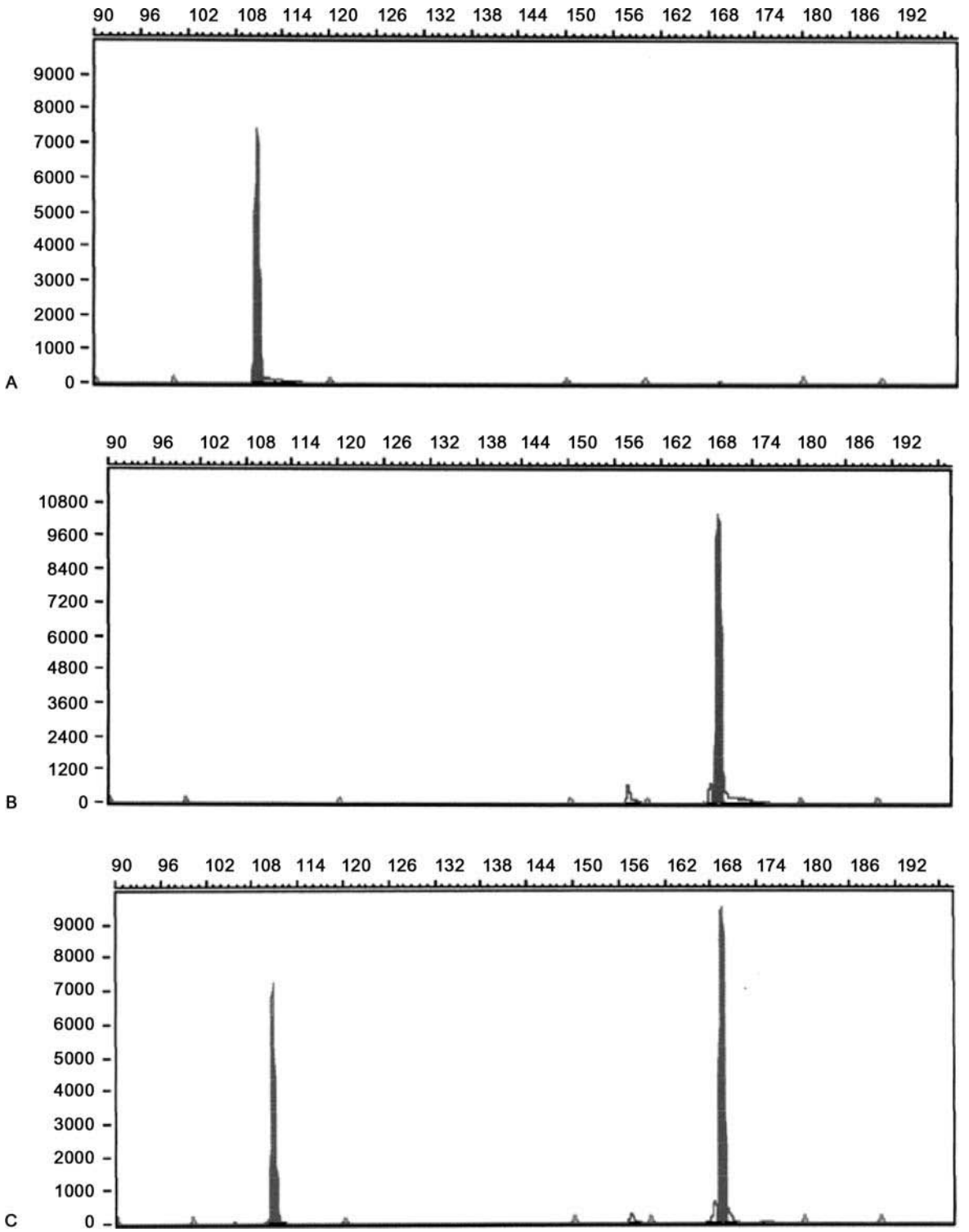


FIGURE 5.7 RFLP analysis using CE. RFLP analysis of genomic DNA by classical gel electrophoresis can be replaced by CE as mentioned in the text. PCR reaction and digestion of the PCR product is performed following the standard protocol but using one fluorescent-labeled primer. Subsequent analysis by CE reveals characteristic patterns for the individual genotypes. As an example, analysis of the *HFE* gene is shown. The *HFE* mutation at codon 282 generates a new restriction site. Digestion with the according restriction enzyme results in two shorter DNA fragments (only the one with the aligning fluorescent-labeled primer at one end can be detected by CE). In panels **A.** and **B.** Only one peak at different positions is seen, corresponding to the wild-type (digested) and mutated sequence of the *HFE* gene (undigested), respectively. In panel **C.** The heterozygous genotype is displayed (two peaks). (Courtesy of A. Hartmann, University of Regensburg, Germany.)

provides a random match probability of approximately one in five billion. Applications range from identifying suitable recipients for organ transplantation and bone marrow engraft analysis to paternity testing and forensic testing.

Also, in the past years, sequencing of mitochondrial DNA (mtDNA) has also become a routine forensic casework application. This method is used when only limited quantities or poor quality of DNA are available for testing. It can be applied to hair shafts, bones, teeth, and other samples that are not suitable for routine STR analysis (Carey and Mitnik, 2002; Pouchkarev *et al.*, 1998). The two hypervariable regions HV1 and HV2 of the polymorphic control region of mitochondrial DNA are often used in forensic applications to differentiate among individuals within a population. The analysis of mitochondrial DNA sequences currently provides a power of discrimination of approximately 1 in 3,000 due to the limited size of mitochondrial sequence databases. Butler and colleagues (Butler *et al.*, 1998) successfully attempted to replace expensive sequencing of the amplified PCR products by the use of restriction endonuclease digestion followed by CE to separate and size the PCR-RFLP fragments. This approach offers a rapid alternative method for screening of polymorphisms.

DNA analysis has become a key element in forensic applications and the use of CE-based DNA typing as scientific evidence has been accepted in US courts of law (Marchi and Pasacreta, 1997) as well as in many other countries. CE as a rapid, cheap, and high-throughput high-resolution method for the analysis of amplified DNA fragments is currently being used in many forensic laboratories for casework applications and paternity testing.

5.6 FUTURE IMPROVEMENTS

Undoubtedly, the future direction of CE will lie in improved instrumentation and enhanced method development possibilities, leading to faster but smaller devices for high-throughput screening.

One of the most promising processes has been made in the area of micro-fabricated capillary array electrophoresis (CAE). CAE represents a new means to perform CE even faster ($<$160 s) and to analyze more different samples in parallel (Behr *et al.*, 1999; Gao *et al.*, 1999; Woolley *et al.*, 1997). Genetic mapping and DNA sequencing can be completed more rapidly by using this means with up to 48–96 capillaries running simultaneously. As described by Mansfield and coworkers (1997) the CAE system has the capacity to generate up to 5.5 million genotypes per year. For detecting HH mutations, CAE was already shown to be feasible (Simpson *et al.*, 1998).

To provide even smaller micro-machined analytic systems, non-electronic chips will be developed in future. Generally, entire CE systems are constructed on glass or plastic chips. Gel-filled plastic channels in the microchips can be used for DNA analysis. These devices contain buffer and sample channels and have a connection to a detection system. The first systems established are microchip-based, which analyze blood parameters in clinical diagnostics. Integration of enzymatic DNA digestion and CE on-chip was first performed by Jacobson and Ramsey (1996). The DNA sample and the restriction enzyme reacted on the device in a 0.7 nl reaction chamber for 10 s followed by resolution on a 1% cellulose gel within 3 minutes. The fragments were detected by laser-induced fluoresence. Also, outstanding progress towards the ''lab-on-chip'' concept was made by the functional integration of a PCR heating chamber coupled to a CE separation channel on a chip (Woolley *et al.*, 1996). Recently, an on-chip electric field driven electrochemical detection system was introduced using a microchannel with gold microband electrodes (Ordeig *et al.*, 2008). Ultimately, this approach creates new opportunities for application in inexpensive portable chip-based CE devices.

In essence, CE and its variations offer great possibilities in the area of molecular diagnostics, and their further improvement will lead to faster and inexpensive test equipment for clinical analysis. Further, in the area of "omics" CE appears one of the analytical backbones enabling system biology approaches.

REFERENCES

Andersen, P.S., Larsen, L.A., Kanters, J., Havndrup, O., Bundgaard, H., Brandt, N.J., Vuust, J., and Christiansen, M. (1998). Mutation detection by cleavase in combination with capillary electrophoresis analysis: application to mutations causing hypertrophic cardiomyopathy and Long-QT syndrome. *Mol. Diagn.* 3, 105–111.

Atha, D.H., Wenz, H.M., Morehead, H., Tian, J., and O'Connell, C.D. (1998). Detection of p53 point mutations by single strand conformation polymorphism: analysis by capillary electrophoresis. *Electrophoresis* 19, 172–179.

Attallah, A.M., Abdel Malak, C.A., Elghawalby, N.A., Shehatta, A.S., Abdel-Raouf, M., and Shiha, G.E. (2004). Identification of a specific marker for hepatitis C virus infection using capillary zone electrophoresis. *Clin. Chim. Acta* 346, 171–179.

Baba, Y., Tomisaki, R., Sumita, C., Morimoto, I., Sugita, S., Tsuhako, M., Miki, T., and Ogihara, T. (1995). Rapid typing of variable number of tandem repeat locus in the human apolipoprotein B gene for DNA diagnosis of heart disease by polymerase chain reaction and capillary electrophoresis. *Electrophoresis* 16, 1437–1440.

Bandoh, N., Hayashi, T., Kishibe, K., Takahara, M., Imada, M., Nonaka, S., and Harabuchi, Y. (2002). Prognostic value of p53 mutations, bax, and spontaneous apoptosis in maxillary sinus squamous cell carcinoma. *Cancer* 94, 1968–1980.

Barta, C., Ronai, Z., Sasvari-Szekely, M., and Guttman, A. (2001). Rapid single nucleotide polymorphism analysis by primer extension and capillary electrophoresis using polyvinyl pyrrolidone matrix. *Electrophoresis* 22, 779–782.

Behr, S., Matzig, M., Levin, A., Eickhoff, H., and Heller, C. (1999). A fully automated multicapillary electrophoresis device for DNA analysis. *Electrophoresis* 20, 1492–1507.

Benson, J.M., Ellingsen, D., Renshaw, M.A., Resler, A.G., Evatt, B.L., and Hooper, W.C. (1999). Multiplex analysis of mutations in four genes using fluorescence scanning technology. *Thromb. Res.* 96, 57–64.

Berg, K.D., Glaser, C.L., Thompson, R.E., Hamilton, S.R., Griffin, C.A., and Eshleman, J.R. (2000). Detection of microsatellite instability by fluorescence multiplex polymerase chain reaction. *J. Mol. Diagn.* 2, 20–28.

Bjorheim, J., Lystad, S., Lindblom, A., Kressner, U., Westring, S., Wahlberg, S., Lindmark, G., Gaudernack, G., Ekstrom, P., Roe, J., Thilly, W.G., and Borresen-Dale, A.L. (1998). Mutation analyses of KRAS exon 1 comparing three different techniques: temporal temperature gradient electrophoresis, constant denaturant capillary electrophoresis and allele specific polymerase chain reaction. *Mutat. Res.* 403, 103–112.

Boland, C.R., Thibodeau, S.N., Hamilton, S.R., Sidransky, D., Eshleman, J.R., Burt, R.W., Meltzer, S.J., Rodriguez-Bigas, M.A., Fodde, R., Ranzani, G.N., and Srivastava, S. (1998). A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* 58, 5248–5257.

Bor, M.V., Sorensen, B.S., and Nexo, E. (2000). Simultaneous quantitation of several mRNA species by calibrated reverse transcription polymerase chain reaction and capillary electrophoresis: analysis of the epidermal growth factor receptor and its activating ligands EGF, TGF-alpha, and HB-EGF in rat liver. *Lab. Invest* 80, 983–986.

Borson, N.D., Strausbauch, M.A., Wettstein, P.J., Oda, R.P., Johnston, S.L., and Landers, J.P. (1998). Direct quantitation of RNA transcripts by competitive single-tube RT- PCR and capillary electrophoresis. *Biotechniques* 25, 130–137.

Bosserhoff, A.K., Seegers, S., Hellerbrand, C., Scholmerich, J., and Buttner, R. (1999). Rapid genetic screening for hemochromatosis using automated SSCP-based capillary electrophoresis (SSCP-CE). *Biotechniques* 26, 1106–1110.

Butler, J.M. (1998). Rapid assessment of PCR product quality and quantity by capillary electrophoresis. *Methods Mol. Biol.* 98, 39–47.

Butler, J.M. (2007). Short tandem repeat typing technologies used in human identity testing. *Biotechniques* 43, ii–iv.

Butler, J.M., Wilson, M.R., and Reeder, D.J. (1998). Rapid mitochondrial DNA typing using restriction enzyme digestion of polymerase chain reaction amplicons followed by capillary electrophoresis separation with laser-induced fluorescence detection. *Electrophoresis* 19, 119–124.

Canzian, F., Salovaara, R., Hemminki, A., Kristo, P., Chadwick, R.B., Aaltonen, L.A., and de la Chapelle, A. (1996). Semiautomated assessment of loss of heterozygosity and replication error in tumors. *Cancer Res.* 56, 3331–3337.

Carey, L., and Mitnik, L. (2002). Trends in DNA forensic analysis. *Electrophoresis* 23, 1386–1397.

Chan, K.C., Muschik, G.M., Issaq, H.J., Garvey, K.J., and Generlette, P.L. (1996). High-speed screening of polymerase chain reaction products by capillary electrophoresis. *Anal. Biochem.* 243, 133–139.

Cottet, H., Gareil, P., and Viovy, J.L. (1998). The effect of blob size and network dynamics on the size-based separation of polystyrenesulfonates by capillary electrophoresis in the presence of entangled polymer solutions. *Electrophoresis* 19, 2151–2162.

de Cremoux, P., Martin, E.C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Liva, S., Pouillart, P., Sastre-Garau, X., and Magdelenat, H. (1999). Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* 83, 157–161.

Doglio, A., Laffont, C., Thyss, S., and Lefebvre, J.C. (1998). Rapid genotyping of hepatitis C virus by direct cycle sequencing of PCR-amplified cDNAs and capillary electrophoresis analysis. *Res. Virol.* 149, 219–227.

Dovichi, N.J. (1997). DNA sequencing by capillary electrophoresis. *Electrophoresis* 18, 2393–2399.

Ekstrom, P.O., Borresen-Dale, A.L., Qvist, H., Giercksky, K.E., and Thilly, W.G. (1999). Detection of low-frequency mutations in exon 8 of the TP53 gene by constant denaturant capillary electrophoresis (CDCE). *Biotechniques* 27, 128–134.

Gao, Q., Pang, H.M., and Yeung, E.S. (1999). Simultaneous genetic typing from multiple short tandem repeat loci using a 96-capillary array electrophoresis system. *Electrophoresis* 20, 1518–1526.

Garinis, G.A., Patrinos, G.P., Spanakis, N.E., and Menounos, P.G. (2002). DNA hypermethylation: when tumour suppressor genes go silent. *Hum. Genet.* 111, 115–127.

Geisel, J., Walz, T., Bodis, M., Nauck, M., Oette, K., and Herrmann, W. (1999). Fluorescence-based single-strand conformation polymorphism analysis of the low density lipoprotein receptor gene by capillary electrophoresis. *J. Chromatogr. B. Biomed Sci. Appl.* 724, 239–247.

Ghilardi, G., Biondi, M.L., Mangoni, J., Leviti, S., DeMonti, M., Guagnellini, E., and Scorza, R. (2001). Matrix metalloproteinase-1 promoter polymorphism 1G/2G is correlated with colorectal cancer invasiveness. *Clin. Cancer Res.* 7, 2344–2346.

Ghozzi, R., Morand, P., Ferroni, A., Beretti, J.L., Bingen, E., Segonds, C., Husson, M.O., Izard, D., Berche, P., and Gaillard, J.L. (1999). Capillary electrophoresis-single-strand conformation polymorphism analysis for rapid identification of Pseudomonas aeruginosa and other gram-negative nonfermenting bacilli recovered from patients with cystic fibrosis. *J. Clin. Microbiol.* 37, 3374–3379.

Goel, A., Arnold, C.N., Niedzwiecki, D., Chang, D.K., Ricciardiello, L., Carethers, J.M., Dowell, J.M., Wasserman, L., Compton, C., Mayer, R.J., Bertagnolli, M.M., and Boland, C.R. (2003). Characterization of sporadic colon cancer by patterns of genomic instability. *Cancer Res.* 63, 1608–1614.

Henson, D.E., Fielding, L.P., Grignon, D.J., Page, D.L., Hammond, M.E., Nash, G., Pettigrew, N.M., Gorstein, F., and Hutter, R.V. (1995). College of American Pathologists Conference XXVI on clinical relevance of prognostic markers in solid tumors. Summary. Members of the Cancer Committee. *Arch. Pathol. Lab. Med.* 119, 1109–1112.

Herman, J.G., Umar, A., Polyak, K., Graff, J.R., Ahuja, N., Issa, J.P., Markowitz, S., Willson, J.K., Hamilton, S.R., Kinzler, K.W., Kane, M.F., Kolodner, R.D., Vogelstein, B., Kunkel, T.A., and Baylin, S.B. (1998). Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. USA* 95, 6870–6875.

Hernandez, S.M., Morlock, G.P., Butler, W.R., Crawford, J.T., and Cooksey, R.C. (1999). Identification of Mycobacterium species by PCR-restriction fragment length polymorphism analyses using fluorescence capillary electrophoresis. *J. Clin. Microbiol.* 37, 3688–3692.

Hjerten, S. (1967). Free zone electrophoresis. *Chromatogr. Rev.* 9, 122–219.

Hu, H., Li, C., Xiong, Q., Gao, H., Li, Y., Chang, Q., and Liang, Z. (2008). Prenatal diagnosis of beta-thalassemia by chip-based capillary electrophoresis:. *Prenat. Diagn.* 28, 222–229.

Jackson, H.A., Bowen, D.J., and Worwood, M. (1997). Rapid genetic screening for haemochromatosis using heteroduplex technology. *Br. J. Haematol.* 98, 856–859.

Jacob, S., and Praz, F. (2002). DNA mismatch repair defects: role in colorectal carcinogenesis. *Biochimie* 84, 27–47.

Jacobson, S.C., and Ramsey, J.M. (1996). Integrated microdevice for DNA restriction fragment analysis. *Anal. Chem.* 68, 720–723.

Jorgenson, J.W., and Lukacs, K.D. (1981). Free-zone electrophoresis in glass capillaries. *Clin. Chem.* 27, 1551–1553.

Kanamori, Y., Matsushima, M., Minaguchi, T., Kobayashi, K., Sagae, S., Kudo, R., Terakawa, N., and Nakamura, Y. (1999). Correlation between

expression of the matrix metalloproteinase-1 gene in ovarian cancers and an insertion/deletion polymorphism in its promoter region. *Cancer Res.* 59, 4225–4227.

Kandioler, D., Zwrtek, R., Ludwig, C., Janschek, E., Ploner, M., Hofbauer, F., Kuhrer, I., Kappel, S., Wrba, F., Horvath, M., Karner, J., Renner, K., Bergmann, M., Karner-Hanusch, J., Potter, R., Jakesz, R., Teleky, B., and Herbst, F. (2002). TP53 genotype but not p53 immunohistochemical result predicts response to preoperative short-term radiotherapy in rectal cancer. *Ann. Surg.* 235, 493–498.

Khrapko, K., Hanekamp, J.S., Thilly, W.G., Belenkii, A., Foret, F., and Karger, B.L. (1994). Constant denaturant capillary electrophoresis (CDCE): a high resolution approach to mutational analysis. *Nucleic Acids Res.* 22, 364–369.

Kolesar, J.M., Allen, P.G., and Doran, C.M. (1997). Direct quantification of HIV-1 RNA by capillary electrophoresis with laser-induced fluorescence. *J. Chromatogr. B. Biomed Sci. Appl.* 697, 189–194.

Kumar, R., Hanekamp, J.S., Louhelainen, J., Burvall, K., Onfelt, A., Hemminki, K., and Thilly, W.G. (1995). Separation of transforming amino acid-substituting mutations in codons 12, 13 and 61 the N-ras gene by constant denaturant capillary electrophoresis (CDCE). *Carcinogenesis* 16, 2667–2673.

Kuypers, A.W., Willems, P.M., van der Schans, M.J., Linssen, P.C., Wessels, H.M., de Bruijn, C.H., Everaerts, F.M., and Mensink, E.J. (1993). Detection of point mutations in DNA using capillary electrophoresis in a polymer network. *J. Chromatogr.* 621, 149–156.

Kuypers, A.W., Meijerink, J.P., Smetsers, T.F., Linssen, P.C., and Mensink, E.J. (1994). Quantitative analysis of DNA aberrations amplified by competitive polymerase chain reaction using capillary electrophoresis. *J. Chromatogr. B. Biomed Appl.* 660, 271–277.

Kuypers, A.W., Linssen, P.C., Willems, P.M., and Mensink, E.J. (1996). On-line melting of double-stranded DNA for analysis of single-stranded DNA using capillary electrophoresis. *J. Chromatogr. B. Biomed Appl.* 675, 205–211.

Kuypers, A.W., Linssen, P.C., Lauer, H.H., and Mensink, E.J. (1998). Contamination-free and automated composition of a reaction mixture for nucleic acid amplification using a capillary electrophoresis apparatus. *J. Chromatogr. A.* 806, 141–147.

La Fountain, M., Schwartz, M., Cormier, J., and Buel, E. (1998). Validation of capillary electrophoresis for analysis of the X–Y homologous amelogenin gene. *J. Forensic Sci.* 43, 1188–1194.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, L.W., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, M., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la N Dedhia, B.M., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., Szustakowki, J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., and Choi, S. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Larsen, L.A., Gronskov, K., Norgaard-Pedersen, B., Brondum-Nielsen, K., Hasholt, L., and Vuust, J. (1997). High-throughput analysis of fragile X (CGG)n alleles in the normal and premutation range by PCR amplification and automated capillary electrophoresis. *Hum. Genet.* 100, 564–568.

Larsen, L.A., Christiansen, M., Vuust, J., and Andersen, P.S. (1999). High-throughput single-strand conformation polymorphism analysis by automated capillary electrophoresis: robust multiplex analysis and pattern-based identification of allelic variants. *Hum. Mutat.* 13, 318–327.

Li-Sucholeiki, X.C., Khrapko, K., Andre, P.C., Marcelino, L.A., Karger, B.L., and Thilly, W.G. (1999). Applications of constant denaturant capillary electrophoresis/high-fidelity polymerase chain reaction to human genetic analysis. *Electrophoresis* 20, 1224–1232.

Li, Q., Deka, C., Glassner, B.J., Arnold, K., Li-Sucholeiki, X.C., Tomita-Mitchell, A., Thilly, W.G., and Karger, B.L. (2005). Design of an automated multicapillary instrument with fraction collection for DNA mutation discovery by constant denaturant capillary electrophoresis (CDCE). *J. Sep. Sci.* 28, 1375–1389.

Lin, C.Y., Su, Y.N., Lee, C.N., Hung, C.C., Cheng, W.F., Lin, W.L., Chen, C.A., and Hsieh, S.T. (2006). A rapid and reliable detection system for the analysis of PMP22 gene dosage by MP/DHPLC assay. *J. Hum. Genet.* 51, 227–235.

Liu, T., Tannergard, P., Hackman, P., Rubio, C., Kressner, U., Lindmark, G., Hellgren, D., Lambert, B., and Lindblom, A. (1999). Missense mutations in hMLH1 associated with colorectal cancer. *Hum. Genet.* 105, 437–441.

Lupski, J.R., de Oca-Luna, R.M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B.J., Saucedo-Cardenas, O., Barker, D.F., Killian, J.M., and Garcia, C.A. (1991). DNA duplication associated with Charcot-Marie-Tooth disease type 1 A. *Cell* 66, 219–232.

Magnusdottir, S., Viovy, J.L., and Francois, J. (1998). High resolution capillary electrophoretic separation of oligonucleotides in low-viscosity, hydrophobically end-capped polyethylene oxide with cubic order. *Electrophoresis* 19, 1699–1703.

Mansfield, E.S., Vainer, M., Harris, D.W., Gasparini, P., Estivill, X., Surrey, S., and Fortina, P. (1997). Rapid sizing of polymorphic microsatellite markers by capillary array electrophoresis. *J. Chromatogr. A.* 781, 295–305.

Marchi, E., and Pasacreta, R.J. (1997). Capillary electrophoresis in court: the landmark decision of the People of Tennessee versus Ware. *J. Capillary Electrophor.* 4, 145–156.

Margraf, R.L., Erali, M., Liew, M., and Wittwer, C.T. (2004). Genotyping hepatitis C virus by heteroduplex mobility analysis using temperature gradient capillary electrophoresis. *J. Clin. Microbiol.* 42, 4545–4551.

Marra, G., and Boland, C.R. (1995). Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. *J. Natl. Cancer Inst.* 87, 1114–1125.

Marsh, J.W., Finkelstein, S.D., Demetris, A.J., Swalsky, P.A., Sasatomi, E., Bandos, A., Subotin, M., and Dvorchik, I. (2003). Genotyping of hepatocellular carcinoma in liver transplant recipients adds predictive power for determining recurrence-free survival. *Liver Transpl.* 9, 664–671.

Merkelbach-Bruse, S., Kose, S., Losen, I., Bosserhoff, A.K., and Buettner, R. (2000). Detection of HNPCC patients by multiple use of capillary electrophoresis techniques. *Combin. Chem. High Throughput Screen* 3, 519–524.

Miettinen, M., Majidi, M., and Lasota, J. (2002). Pathology and diagnostic criteria of gastrointestinal stromal tumors (GISTs): a review. *Eur. J. Cancer* 38, S39–S51.

Miller, J.E., Wilson, S.S., Jaye, D.L., and Kronenberg, M. (1999). An automated semiquantitative B and T cell clonality assay. *Mol. Diagn.* 4, 101–117.

Muller, A., Edmonston, T.B., Dietmaier, W., Buttner, R., Fishel, R., and Ruschoff, J. (2004). MSI-testing in hereditary non-polyposis colorectal carcinoma (HNPCC). *Dis. Markers* 20, 225–236.

Muniappan, B.P., and Thilly, W.G. (1999). Application of constant denaturant capillary electrophoresis (CDCE) to mutation detection in humans. *Genet. Anal.* 14, 221–227.

Munro, N.J., Snow, K., Kant, J.A., and Landers, J.P. (1999). Molecular diagnostics on microfabricated electrophoretic devices: from slab gel- to capillary- to microchip-based assays for T- and B-cell lymphoproliferative disorders. *Clin. Chem.* 45, 1906–1917.

Nicoletto, M.O., Donach, M., De Nicolo, A., Artioli, G., Banna, G., and Monfardini, S. (2001). BRCA-1 and BRCA-2 mutations as prognostic factors in clinical practice and genetic counseling. *Cancer Treat. Rev.* 27, 295–304.

Odin, E., Wettergren, Y., Larsson, L., Larsson, P.A., and Gustavsson, B. (1999). Rapid method for relative gene expression determination in human tissues using automated capillary gel electrophoresis and multicolor detection. *J. Chromatogr. B. Biomed Sci. Appl.* 734, 47–53.

Ordeig, O., Godino, N., del Campo, J., Munoz, F.X., Nikolajeff, F., and Nyholm, L. (2008). On-chip electric field driven electrochemical detection using a poly(dimethylsiloxane) microchannel with gold microband electrodes. *Anal. Chem.* 80, 3622–3632.

Pancholi, P., Oda, R.P., Mitchell, P.S., Landers, J.P., and Persing, D.H. (1997). Diagnostic detection of herpes simplex and Hepatitis C viral amplicons by Capillary Electrophoresis: Comparison with Southern blot detection. *Mol. Diagn.* 2, 27–37.

Pouchkarev, V.P., Shved, E.F., and Novikov, P.I. (1998). Sex determination of forensic samples by polymerase chain reaction of the amelogenin gene and analysis by capillary electrophoresis with polymer matrix. *Electrophoresis* 19, 76–79.

Ren, J. (2000). High-throughput single-strand conformation polymorphism analysis by capillary electrophoresis. *J. Chromatogr. B Biomed. Sci. Appl.* 741, 115–128.

Righetti, P.G., and Gelfi, C. (1998). Recent advances in capillary zone electrophoresis of DNA. *Forensic Sci. Int.* 92, 239–250.

Rolston, R., Sasatomi, E., Hunt, J., Swalsky, P.A., and Finkelstein, S.D. (2001). Distinguishing de novo second cancer formation from tumor recurrence: mutational fingerprinting by microdissection genotyping. *J. Mol. Diagn.* 3, 129–132.

Saffroy, R., Lemoine, A., Haas, P., Tindiliere, F., Marion, S., and Debuire, B. (2002). Rapid automated simultaneous screening of (G1691A) Factor V, (G20210A) prothrombin, and (C677T) methylenetetrahydrofolate reductase variants by multiplex PCR using fluorescence scanning technology. *Genet. Test.* 6, 233–236.

Saletti, P., Edwin, I.D., Pack, K., Cavalli, F., and Atkin, W.S. (2001). Microsatellite instability: application in hereditary non-polyposis colorectal cancer. *Ann. Oncol.* 12, 151–160.

Sanger, F., Nicklen, S., and Coulson, A.R. (1992). DNA sequencing with chain-terminating inhibitors. 1977 [classical article]. *Biotechnology* 24, 104–108.

Sasatomi, E., Finkelstein, S.D., Woods, J.D., Bakker, A., Swalsky, P.A., Luketich, J.D., Fernando, H.C., and Yousem, S.A. (2002). Comparison of accumulated allele loss between primary tumor and lymph node metastasis in stage II non-small cell lung carcinoma: implications for the timing of lymph node metastasis and prognostic value. *Cancer Res.* 62, 2681–2689.

Schmalzing, D., Koutny, L., Salas-Solano, O., Adourian, A., Matsudaira, P., and Ehrlich, D. (1999). Recent developments in DNA sequencing by capillary and microdevice electrophoresis. *Electrophoresis* 20, 3066–3077.

Schummer, B., Siegsmund, M., Steidler, A., Toktomambetova, L., Kohrmann, K.U., and Alken, P. (1999). Expression of the gene for the multidrug resistance-associated protein in human prostate tissue. *Urol. Res.* 27, 164–168.

Sciacchitano, C.J. (1998). DNA fingerprinting of Listeria monocytogenes using enterobacterial repetitive intergenic consensus (ERIC) motifs-polymerase chain reaction/capillary electrophoresis. *Electrophoresis* 19, 66–70.

Sell, S.M., and Lugemwa, P.R. (1999). Development of a highly accurate, rapid PCR-RFLP genotyping assay for the methylenetetrahydrofolate reductase gene. *Genet. Test.* 3, 287–289.

Simpson, P.C., Roach, D., Woolley, A.T., Thorsen, T., Johnston, R., Sensabaugh, G.F., and Mathies, R.A. (1998). High-throughput genetic analysis using microfabricated 96-sample capillary array electrophoresis microplates. *Proc. Natl. Acad. Sci. USA* 95, 2256–2261.

Stanta, G., and Bonin, S. (1998). RNA quantitative analysis from fixed and paraffin-embedded tissues: membrane hybridization and capillary electrophoresis. *Biotechniques* 24, 271–276.

Strom, C.M., Huang, D., Li, Y., Hantash, F.M., Rooke, J., Potts, S.J., and Sun, W. (2007). Development of a novel, accurate, automated, rapid, high-throughput technique suitable for population-based carrier screening for Fragile X syndrome. *Genet. Med.* 9, 199–207.

Sunada, W.M., and Blanch, H.W. (1997). Polymeric separation media for capillary electrophoresis of nucleic acids. *Electrophoresis* 18, 2243–2254.

Tan, H., and Yeung, E.S. (1998). Automation and integration of multiplexed on-line sample preparation with capillary electrophoresis for high-throughput DNA sequencing. *Anal. Chem.* 70, 4044–4053.

Terabe, S. (2004). Micellar electrokinetic chromatography. *Anal. Chem.* 76, 241A–246A.

Tomaiuolo, R., Spina, M., and Castaldo, G. (2003). Molecular diagnosis of cystic fibrosis: comparison of four analytical procedures. *Clin. Chem. Lab. Med.* 41, 26–32.

Trent, R.J., Le, H., and Yu, B. (1998). Prenatal diagnosis for thalassaemia in a multicultural society. *Prenat. Diagn.* 18, 591–598.

Tsui, L.C. (1992a). Mutations and sequence variations detected in the cystic fibrosis transmembrane conductance regulator (CFTR) gene: a report from the Cystic Fibrosis Genetic Analysis Consortium. *Hum. Mutat.* 1, 197–203.

Tsui, L.C. (1992b). The spectrum of cystic fibrosis mutations. *Trends Genet.* 8, 392–398.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di, F., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., and Nodell, M. (2001). The sequence of the human genome. *Science* 291, 1304–1351.

Weir, H.K., Thun, M.J., Hankey, B.F., Ries, L.A., Howe, H.L., Wingo, P.A., Jemal, A., Ward, E., Anderson, R.N., and Edwards, B.K. (2003). Annual report to the nation on the status of cancer, 1975–2000, featuring the uses of surveillance data for cancer prevention and control. *J. Natl. Cancer Inst.* 95, 1276–1299.

Wenz, H.M., Ramachandra, S., O'Connell, C.D., and Atha, D.H. (1998). Identification of known p53 point mutations by capillary electrophoresis using unique mobility profiles in a blinded study. *Mutat. Res.* 382, 121–132.

Williams, L.C., Hedge, M.R., Herrera, G., Stapleton, P.M., and Love, D.R. (1999). Comparative semi-automated analysis of (CAG) repeats in the Huntington disease gene: use of internal standards. *Mol. Cell. Probes* 13, 283–289.

Woolley, A.T., Hadley, D., Landre, P., deMello, A.J., Mathies, R.A., and Northrup, M.A. (1996). Functional integration of PCR amplification and capillary electrophoresis in a microfabricated DNA analysis device. *Anal. Chem.* 68, 4081–4086.

Woolley, A.T., Sensabaugh, G.F., and Mathies, R.A. (1997). High-speed DNA genotyping using microfabricated capillary array electrophoresis chips. *Anal. Chem.* 69, 2181–2186.

Wu, R., Hu, L., Wang, F., Ye, M., and Zou, H. (2008). Recent development of monolithic stationary phases with emphasis on microscale chromatographic separation. *J. Chromatogr. A* 1184, 369–392.

Zhang, J., Voss, K.O., Shaw, D.F., Roos, K.P., Lewis, D.F., Yan, J., Jiang, R., Ren, H., Hou, J.Y., Fang, Y., Puyang, X., Ahmadzadeh, H., and Dovichi, N.J. (1999). A multiple-capillary electrophoresis system for small-scale DNA sequencing and analysis. *Nucleic Acids Res.* 27, e36.

Zhang, J., Burger, C., and Chu, B. (2006). Nanostructured polymer matrix for oligonucleotide separation. *Electrophoresis* 27, 3391–3398.

Zhu, Y., Spitz, M.R., Lei, L., Mills, G.B., and Wu, X. (2001). A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter enhances lung cancer susceptibility. *Cancer Res.* 61, 7825–7829.

Zielenski, J., Aznarez, I., Onay, T., Tzounzouris, J., Markiewicz, D., and Tsui, L.C. (2002). CFTR mutation detection by multiplex heteroduplex (mHET) analysis on MDE gel. *Methods Mol. Med.* 70, 3–19.

This page intentionally left blank

Chapter 6

Temperature and Denaturing Gradient Gel Electrophoresis

Hartmut Peters and Peter N. Robinson
Institute of Medical Genetics, Charité-Universitätsmedizin Berlin, Germany

6.1 INTRODUCTION

Temperature gradient gel electrophoresis (TGGE) and the related method denaturing-gradient gel electrophoresis (DGGE) are both based on the principle that the electrophoretic mobility of double-stranded DNA fragments is significantly reduced by their partial denaturation. Owing to the sequence dependence of the melting properties of DNA fragments, sequence variations can be detected. Although the sensitivity of TGGE and DGGE in detecting point mutations in genetic disorders and other settings has been reported to be close to 100%, these methods have never become as popular as other mutation detection methods such as SSCP (see Chapter 4), which may be related to the perception that it is difficult to design adequate PCR primers and set up the assays. In this chapter, the basic principles of TGGE/DGGE will be discussed and procedures for setting up assays will be described, including how to design and test PCR primers suitable for TGGE/DGGE analysis. Furthermore, studies on the sensitivity of TGGE/DGGE for mutation analysis of genetic disorders will be reviewed and an overview of variations on the basic TGGE/DGGE method will be provided. TGGE and DGGE are robust and highly sensitive methods for mutation screening of genetic disorders that have many advantages, which counterbalance the extra effort required in establishing the method.

6.2 THE THEORY OF TEMPERATURE-GRADIENT GEL ELECTROPHORESIS

6.2.1 Melting Behavior of Short Double-Stranded DNA Fragments

Myers and coworkers (1985b) originally developed a method of separating DNA fragments differing by single nucleotide substitutions in denaturing gradient gels. The method was based on the notion that the denaturation (melting) of DNA fragments can be regarded as an equilibrium for each base pair (bp) between two distinct states: (1) double helical, and (2) a more random state in which bases are neither paired nor stacked on adjacent bases in any orderly way (Myers *et al.*, 1987). The change from the first to the second state is caused by increasing temperature or increasing concentration of denaturing agents.

In the case of single-nucleotide substitutions, the replacement of an A:T bp (two hydrogen bonds) by a G:C pair (three hydrogen bonds) will generally be expected to increase the temperature at which the corresponding DNA sequence melts. The context of the nucleotide substitution also plays a role, and substitutions of A:T by T:A pairs, or G:C by C:G pairs, can also affect the temperature at which a DNA sequence dissociates.

Furthermore, a DNA fragment dissociates in a stepwise fashion as the temperature is gradually increased. Dissociation occurs nearly simultaneously in distinct, approximately 50 to 300 nucleotide long regions, termed "melting domains". All nucleotides in a given melting domain dissociate in an all-or-nothing manner within a narrow temperature interval.

The melting temperature (T_m) indicates the temperature at which 50% of the individual molecules are dissociated in the given melting domain, and 50% are double helical. As indicated above, the T_m is strongly dependent on the individual DNA sequence and can be significantly altered by small changes in the DNA sequence including single-nucleotide substitutions.

6.2.2 Electrophoretic Mobility and the Melting State of DNA Fragments

TGGE is based on detecting differences in the electrophoretic mobility between molecules that may differ only at a single position. DNA fragments produced by

the polymerase chain reaction (PCR) are subjected to electrophoresis through a linearly increasing gradient of temperature (or concentration gradient of denaturing agents such as urea and formamide for DGGE). Nucleotide substitutions and other small changes in the DNA sequence are associated with additional bands following TGGE.

The electrophoretic mobility of DNA fragments differs according to whether the fragment is completely double helical, if one or more melting domains has dissociated, or if complete dissociation to two single-stranded molecules has occurred. Each of these states can be visualized using a perpendicular TGGE experiment, as will be discussed further below (see section 6.3.2).

The electrophoretic mobility of a double helical (nondenatured) DNA fragment is not significantly altered by single-nucleotide substitutions within it, but is primarily dependent on the length and perhaps the curvature of the fragment (Haran *et al.*, 1994). Therefore, assuming that PCR products contain a mixture of two DNA fragments that differ at a single position, as would be the case for a heterozygous point mutation, both fragments will initially progress through the gel at the same speed.

When the molecules reach that point in the gel where the temperature equals their T_m, the molecules will experience a decrease in mobility owing to a transition from a completely duplex (double helical) conformation to a partially denatured one. Dissociation of the first or first few melting domains generally results in a dramatic reduction in the mobility of the DNA fragment, because the fragment takes on a complex, branched conformation.

Due to the strong sequence dependence of the melting temperature, branching (dissociation) and consequent retardation of electrophoretic mobility occurs at different levels of the temperature gradient associated with bands at different positions in the gel (Myers *et al.*, 1987). In addition to the two homoduplex molecules (wt/wt and mt/mt), two different heteroduplex molecules (wt/mt and mt/wt) can be formed by dissociating and reannealing DNA fragments containing a heterozygous mutation prior to performing TGGE (Fig. 6.1). In practice, it is also possible to perform 40 cycles of PCR; the activity of the *Taq* polymerase is exhausted in the final cycles of PCR, such that heteroduplices are formed as efficiently as if one performed denaturation and reannealing following PCR. Heteroduplex fragments then contain unpaired bases or "bulges" in the otherwise double helical DNA, resulting in a significant reduction in the T_m of the affected melting domain (Ke and Wartell, 1995). The melting temperatures of the two heteroduplex molecules are generally different from one another, so that each heteroduplex is separately visible in the gel. A heterozygous point mutation will thus be visualized by the appearance of four bands: a band representing the normal allele (homoduplex), a band representing the mutant homoduplex that will lie above or underneath the wild-type homoduplex band, depending on the effect of the mutation on the T_m, and two heteroduplex bands that are always above the homoduplex bands (Fig. 6.2; Myers *et al.*, 1987). Mutant and wild-type homoduplex bands are separated by 2–10 mm in a typical polyacrylamide gel, and the heteroduplex bands are often three or more cm above the homoduplex bands.

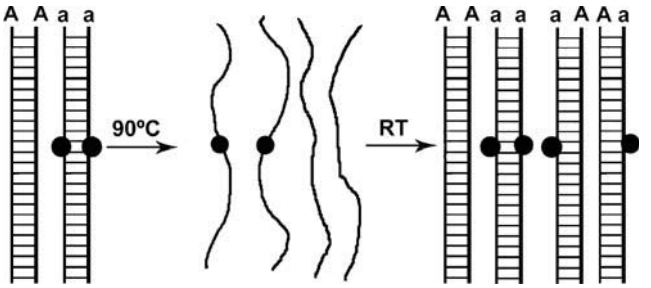


FIGURE 6.1 Mechanism of heteroduplex formation. In the case of heterozygous point mutations in genetic disorders, PCR produces two alleles differing only at the position of the point mutation. A wild-type (AA) and a mutant (aa) molecule are present at an approximately 1:1 ratio. Denaturation followed by reannealing of these molecules produces both wild-type (AA) and mutant (aa) homoduplex molecules, as well as two heteroduplex molecules, consisting of a wild-type and a mutant strand (Aa and aAa).

6.2.3 Mutations are only Detectable in the Lowest Melting Domain(s)

In the discussion above, a significant issue is that mutations are detectable only in the melting domain(s) with the lowest melting temperature. If, however, a DNA molecule contains several melting domains with different melting temperatures, it is generally not possible to visualize mutations located elsewhere than in the melting domain with the lowest T_m. Once the DNA fragments reach the temperature at which the first melting domain dissociates, the mobility of the fragment is greatly reduced so that it may not reach temperatures relevant for the higher T_m domains under the conditions of the experiment. Also, dissociation of the highest T_m domain results in complete dissociation of the DNA fragment into two single-stranded DNA molecules. Single-stranded DNA, like completely double helical DNA, does not demonstrate differences in electrophoretic mobility, owing to small sequence changes, and hence there is no possibility of distinguishing two sequences once complete dissociation has occurred.

The consequence of these observations is that only mutations in the lowest T_m domain can be reliably detected by TGGE or DGGE (Myers *et al.*, 1987).

6.2.4 GC- and Psoralen Clamps Extend the Usefulness of TGGE

Myers and coworkers (1985a) presented an extension of the original DGGE protocol that allowed mutations in every region of the DNA fragment under analysis to be detected.

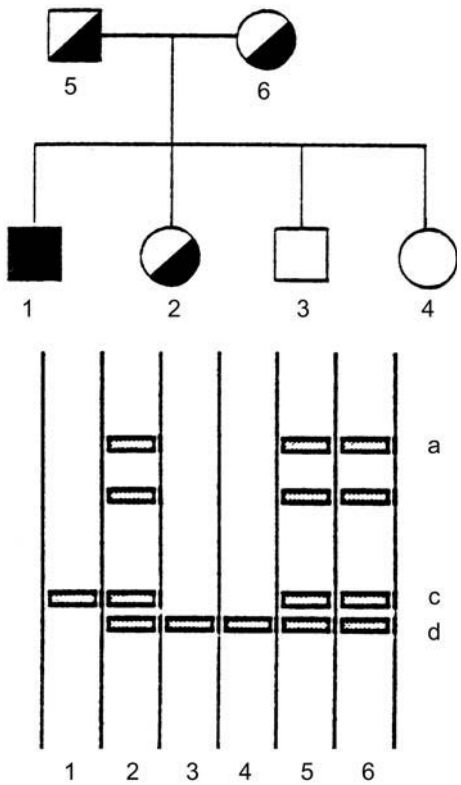


FIGURE 6.2 Parallel TGGE/DGGE. Mutation screening is generally performed with the temperature or denaturing gradient parallel to the direction of electrophoresis. In this example, results of electrophoresis from top to bottom for a hypothetical family segregating an autosomal recessive disorder are shown. Cases 3 and 4 are normal, carrying only the wild-type allele ("d"). Cases 2, 5, and 6 are heterozygous for a point mutation resulting in the appearance of an additional homoduplex band ("c"), as well as two additional heteroduplex bands ("a" and "b"). Case 1, which is homozygous for the mutation, shows just the mutant homoduplex band ("c").

These researchers attached a 135 bp, GC-rich sequence, known as "GC-clamp", to the β-globin promoter region in which mutations were being sought. The β-globin promoter region was found to contain two melting domains; without the GC-clamp, only mutations in the domain with the lower T_m could be visualized in the gel. Owing to its high GC content, the GC-clamp has a significantly higher melting temperature than most naturally occurring sequences. The attachment of the GC-clamp was found to significantly alter the melting properties of the β-globin sequence and mutations in the entire β-globin sequence could be experimentally detected (Myers *et al.*, 1985a). By adding a 40 nt G + C-rich sequence to one of the two PCR primers, a GC-clamp can be conveniently added to any DNA fragment produced by PCR (Sheffield *et al.*, 1989). It is also possible to use a universal GC-clamp that is incorporated into amplified DNA fragments during PCR, thereby avoiding the expense of synthesizing long primers (Top, 1992).

Psoralen-modified PCR primers are an alternative to GC-clamps. One of the two PCR primers is 5′ modified by 5-(ω-hexyloxy)-psoralen. The 5′ terminus of the primer should have two adenosine residues; if the natural sequence does not have AA, this sequence should be appended to the specific DNA sequence of the primer. Psoralens are bifunctional photoreagents that can form covalent bonds with pyrimidine bases (especially thymidine). If intercalated at 5′-TpT in double helical DNA (this will be the complementary sequence of the 3′ terminus of the other strand following PCR), psoralen forms a covalent bond with thymidine after photoinduction (Costes *et al.*, 1993b). Photoinduction can be performed by exposing the PCR products to a source of UV light (365) for 5 to 15 minutes, which can be conveniently done in the original PCR tubes or 96-well plate.

In general, psoralen clamping provides comparable results to GC clamping, except that cross-linking of the PCR fragments is only approximately 85% efficient, so that one observes single-stranded, denatured DNA fragments running below the main bands in the TGGE. Psoralen clamping is sometimes preferred over GC-clamping because the PCR is often easier to optimize, and bipolar clamping is possible if necessary (see below). Psoralen modification of primers is available from many commercial oligonucleotide sources.

6.3 THE PRACTICE OF TEMPERATURE GRADIENT GEL ELECTROPHORESIS

Detailed protocols for TGGE and DGGE are available elsewhere (Kang *et al.*, 1995; Murdaugh and Lerman, 1996). In the following paragraphs, the most important issues concerning how to set up TGGE or DGGE assays successfully are discussed, including especially the issues related to primer design and optimization procedures. Several points that apply only to DGGE are discussed in section 6.4.

6.3.1 Primer Design for TGGE/DGGE

One of the first and most widely used computer programs to design primers for TGGE was the Melt87 package by Lerman and Silverstein (1987). An updated version of this program (Melt94) is available at http://web.mit.edu/osp/www/melt.html. The Melt87 program calculates the T_m for each bp in the DNA fragment, i.e. the temperature at which 50% of the individual molecules are double helical and 50% of the molecules are in a fully disordered, melted state. The results of such a calculation are termed "melting map" (Fig. 6.3). One notices that DNA fragments are typically divided into distinct melting domains of about 50 to 300 bp in length, in which all base pairs have nearly identical T_m. The melting map demonstrates the lowest melting domain in the DNA fragment; as mentioned above, only mutations in this region will be visible by TGGE analysis.

A further useful program in the Melt87 package is SQHTX. This program calculates the expected displacement in the gradient for a single-nucleotide mismatch (as would be the case for a heteroduplex molecule with a single-nucleotide

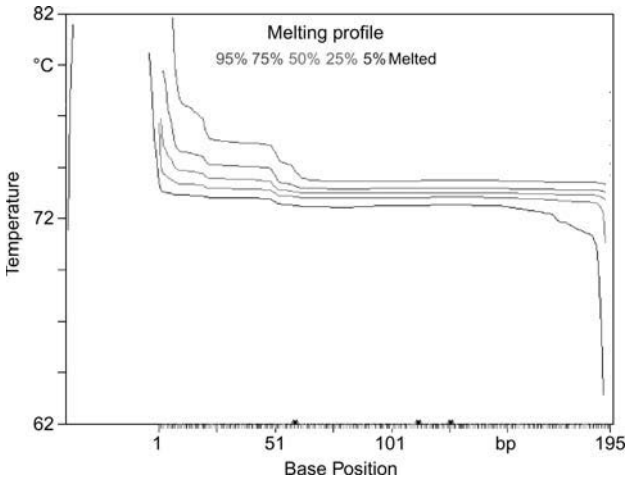


FIGURE 6.3 Melting map. This graphic represents a fragment from exon 14 of the *NF1* gene and was produced using TGGE-Star. Each tick on the x-axis represents a base pair. The base pairs are numbered from 1 to 195. The y-axis shows the temperature where the probability for a bp to be melted has the value 0.95, 0.75, 0.5, 0.25, and 0.05, respectively. The 5′-terminus of the fragment corresponds to a GC-clamp. Additionally, one can distinguish two further melting domains: from the 5′-terminus to the 50th bp and from the 50th bp to the 3′-terminus. The difference between these two melting domains is small and the sensitivity of TGGE is not disturbed. If the difference between these two plateaus in the curve were higher, both regions would need to be tested in two different PCR-TGGE steps. Mutations were detected in both regions of this fragment: three asterisks above the x-axis mark positions of mutations detected with this assay.

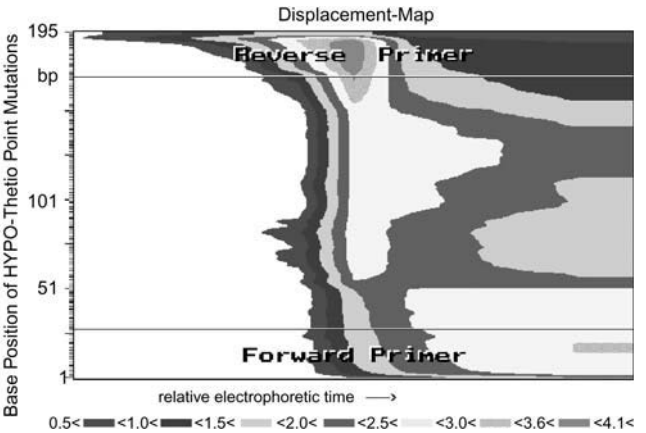


FIGURE 6.4 Displacement maps calculating using the program SQHTX, and graphic created with TGGE-Star. In the case of a heterozygous mutation, two heteroduplex bands occur. Heteroduplices do not migrate as far as the wild-type fragments because they "melt" at lower temperatures. The distance of heteroduplex bands and wild-type bands depends on the electrophoretic duration (x-axis) and the base position (y-axis). A mutation can only be detected, when the displacement is higher than the resolution of the gel. The color codes indicate different electrophoretic times, and the width of each band of color indicates the expected displacement (in arbitrary units) in the gel for a point mutation at the corresponding position in the sequence. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

substitution) at every position in the fragment. This analysis provides the clearest indication of the position in the fragment, where mutations will be detectable by TGGE analysis (Lerman and Silverstein, 1987). Figure 6.4 provides an example of a displacement map calculated with SQHTX.

The Melt87 programs are DOS-based and difficult to use for those with little experience with DOS and menu-based programs. Melt87 has no graphic capabilities of its own, and users need to process its output with a graphics program of their choice. For this reason, several freely available and proprietary programs have become available, which are significantly easier to use (Table 6.1).

Users should load a DNA sequence encompassing the DNA fragment to be analyzed (e.g. an exon with flanking intron sequences) together with about 100 nucleotides "extra" sequence to either side of the fragment of interest. The above-mentioned programs can be used to find primers that result in a DNA fragment with melting properties adequate for TGGE or DGGE. In general, some amount of trial and error is needed to find optimal primers for any given sequence. Users need to decide both the position of the forward and reverse primers as well as whether the GC-clamp is to be placed on

TABLE 6.1 Programs for the design of PCR primers for use in TGGE/DGGE.

Name	Comment	URL
Melt94	DOS-based	http://web.mit.edu/osp/www/melt.html
TGGE-Star	DOS-based, freely available user-friendly wrapper for Melt87 (Gille and Gille, 2002)	http://www.charite.de/bioinf/tgge
Poland	Server-based implementation of Poland's algorithm (Steger, 1994)	http://www.biophys.uni-duesseldorf.de/local/POLAND/poland.html
MELTingeny	A commercial, Java-based GUI program with flexible routines for designing DGGE/TGGE primers	http://www.ingeny.com
WinMelt, MacMelt	Commercial GUI programs for melting profile analysis	http://www.medprobe.com

the 5′ or 3′ PCR primer or both (see below for discussion of bipolar clamping). Programs such as TGGE-Star and MELTingeny facilitate this process by allowing users to easily shift primer positions and recalculate the melting maps. It should be mentioned that a 40-nucleotide GC clamp can be "substituted" for a psoralen clamp in the computer analysis.

6.3.2 Perpendicular TGGE for the Determination of the T_m

In most cases in which TGGE is used for mutation analysis, parallel electrophoresis with simultaneous analysis of multiple samples will be performed. For each such assay, the optimal temperature gradient and run time must be determined experimentally. The procedures used for this purpose are described in this and the following paragraph.

The optimization process begins with a perpendicular TGGE experiment, in which electrophoresis is performed perpendicularly to the temperature gradient (Fig. 6.5). Perpendicular TGGE is used to verify the reversible melting behavior of the DNA fragment and to determine its T_m under the experimental conditions. Perpendicular TGGE is run with a gradient of 20°C–60°C, which will be adequate for the vast majority of PCR fragments. Electrophoresis is initially performed at room temperature for 10–15 minutes to run the sample into the gel. Then, electrophoresis is stopped while a temperature gradient of 20°C–60°C is established, after which electrophoresis should be continued for 90–120 minutes. Figure 6.5 demonstrates the use of this analysis to determine the T_m of the DNA fragment being analyzed.

6.3.3 Travel Schedule Experiments

Up to three novel bands are observed upon TGGE/DGGE analysis of a heterozygous mutation or polymorphism. The separation will begin to become apparent when the heteroduplex molecules have reached their T_m, as their mobility will be retarded by partial denaturation. Separation of the homoduplex molecules will occur in a region of the gradient surrounding the T_m of the lowest melting domain of the DNA fragment. Therefore, TGGE assays are set up to avoid a long running time before the samples reach the effective range of separation. One should choose the temperature gradient such that the effective range of separation is approximately in the middle or somewhat above the middle of the gel, and that the upper and lower temperature ranges are separated by about 15°C from the T_m of the DNA fragment.

Once an appropriate temperature gradient has been chosen, the optimal running time can be determined by a travel-schedule experiment, which is a parallel TGGE experiment in which samples are applied every 30 minutes for 3 hours (or longer), such that the last sample to be loaded has run 30 minutes, and the first sample 3 hours. Usually, one will see a reduction on electrophoretic mobility of samples after a certain period of time (generally 60–90 minutes if the temperature gradient was chosen correctly). Samples often do not continue to wander in the gel with any significant velocity once their melting temperature has been reached. These gels are generally run for about 30 minutes longer than the time determined in this manner (Fig. 6.6). Different choices of the range and starting point of the temperature gradient affect both the range in the gel at which mutations will be visible as well as the optimal running time (Fig. 6.7).

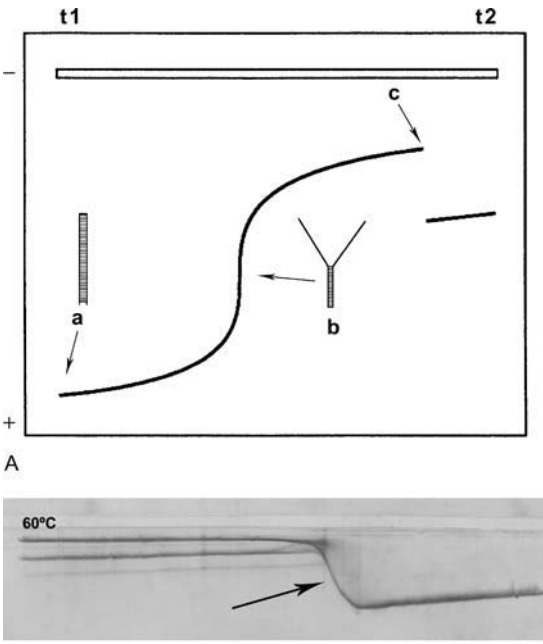


FIGURE 6.5 **A.** Schematic drawing of a perpendicular TGGE/DGGE gel. A temperature gradient from t1 (e.g. 20°C) to t2 (e.g. 60°C) is established perpendicularly to the direction of electrophoresis (indicated by " − " and " + "). Fragments at lower temperatures remain completely double helical and have a relatively high electrophoretic mobility ("a"). Once the melting temperature of the lowest-temperature melting domain is reached, partial denaturation of the DNA fragment ("b") causes a significant reduction of electrophoretic mobility. The temperature at which 50% of individual molecules are melted is denoted as the "melting temperature" (T_m), and is indicated by the arrow in the figure ("b"). A reversible denaturation step is observed as a continuous transition (curve). Once the temperature of the highest-melting domain is reached, irreversible melting occurs, causing a discontinuous transition in the melting curve ("c"). **B.** Perpendicular TGGE gel. In this example, a PCR fragment corresponding to *NF1* gene exon 14 was analyzed. PCR product was applied and run into the gel at 10°C for 15 minutes. Then, a temperature gradient from 20°C to 60°C was established perpendicularly to the direction of electrophoresis, which was then performed for an additional 60 minutes. One observes a high electrophoretic mobility in portions of the gel with temperatures below the T_m of the fragment. The gradual decrease in mobility around the middle of the gel indicates reversible melting of the lowest-temperature melting domain. In portions of the gel with temperatures above the T_m of the fragment, partial denaturation of the fragment leads to a significantly reduced electrophoretic mobility. The arrow at the midway point of the curve indicates the T_m of the fragment under the experimental conditions (approximately 39°C).

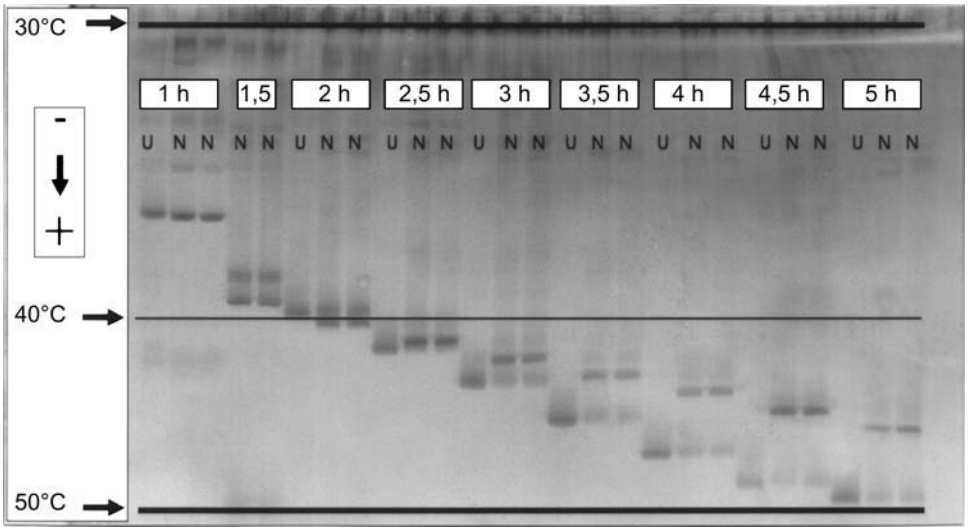


FIGURE 6.6 Travel schedule experiment. This experiment is used to determine the optimal running time of a TGGE experiment. Fragments, corresponding to exon 19a of the *NF1* gene, in which one of the primers was modified with psoralen (see section 6.2.4), were applied at intervals of 30 minutes, such that the first fragments had a total running time of 5 h, and the last fragments to be applied had a running time of 60 minutes. Lanes labeled "U" contain PCR fragments that were not UV irradiated to effect psoralen-mediated crosslinking, and lanes labeled "N" (two lanes were loaded for each timepoint) contain irradiated PCR fragments. One sees that the fragments initially completely double-helical (1 h), such that irradiated and non-irradiated fragments display the same band pattern. Starting at the T_m of this fragment (40°C), the non-irradiated fragments ("U") undergo complete dissociation so that only a single-strand band running well below the main band of the irradiated (cross-linked) fragments is visible (compare the time points at 2.5 and 3 hours). Additionally, the irradiated fragments ("N") show a large reduction in electrophoretic mobility following partial denaturation at about 40°C. Under these conditions, an optimal running time would be 3 hours, although the running time could be reduced by adjusting the temperature gradient (see also Fig. 6.7).

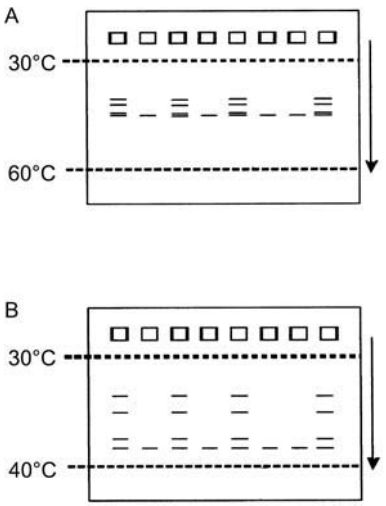


FIGURE 6.7 The effect of different temperature gradients on separation of samples. A relatively wide gradient (A) leads to a relatively small separation of homoduplex and heteroduplex bands in the case of a heterozygous mutation, while a shallower gradient (B) increases the separation. The lower temperature of the gradient can also be adjusted to reduce the amount of time needed before the T_m of the lowest melting domain is reached.

6.3.4 Bipolar Clamping

Occasionally, TGGE analysis will result in fuzzy bands that are difficult to evaluate, despite apparently adequate melting behavior, as predicted by Melt94 or other programs. Bipolar clamping of PCR products, by means of attaching a psoralen clamp to each of the two PCR primers rather than just one, is an effective method to improve melting characteristics of PCR fragments that are otherwise not amenable to TGGE/DGGE analysis (Gille *et al.*, 1998). Bipolar clamping is a simple procedure that can significantly improve results of TGGE analysis in cases where analysis with only one clamp has yielded suboptimal results. Programs such as TGGE-Star (Gille and Gille, 2002) offer the possibility of computer analysis with two clamps, and may suggest the use of bipolar clamping for amplicons whose predicted melting properties are otherwise not satisfactory.

6.4 DENATURING GRADIENT GEL ELECTROPHORESIS (DGGE)

The theory of DGGE/TGGE is described in detail in the first part of this chapter. Parallel DGGE is a form of polyacrylamide gel electrophoresis in which a double-stranded DNA fragment migrates into a gradient of linearly increasing denaturing conditions. The denaturing gradient is functionally equivalent to the temperature gradient of TGGE. The denaturants used are heat (a constant temperature of generally 60°C) and a fixed ratio of formamide (ranging from 0 to 40%) and urea (ranging from 0 to 7 M). The temperature of 60°C was empirically chosen to exceed the melting temperature of an AT-rich DNA fragment in the absence of a denaturant. For extremely GC-rich DNA

sequences higher temperatures (e.g. 75°C) can be used. To achieve a uniform temperature distribution the electrophoresis unit is attached to a circulating water bath.

6.4.1 Optimization of Gel Running Conditions

The computer programs (e.g. Melt94) described above reduce the number of preliminary experiments required for optimization of the gel running conditions. However, it is still necessary to run some preliminary gels to determine the optimal electrophoresis conditions and running times and to confirm that the optimal denaturing gradient has been chosen. The aim of these travel schedule gels is to have well separated bands (normal and mutation positive control are simultaneously loaded on the gels), which are "focused" by the gradient. PCR products with two low-melting domains require different gel conditions for the analysis of each domain.

The choice of the denaturant concentration range can be determined as follows. The differences in gradient depth (the displacement) between a fragment and the same fragment with a change at a specified bp are calculated by the program SQHTX (Lerman and Silverstein, 1987) as described in section 6.3.1. SQHTX calculates the displacement as the difference in temperature at which the wild-type homoduplex and the heteroduplex molecules partially melt (Fig. 6.4). To convert between the temperature values and the denaturant concentration, a difference of 1°C is converted to a difference of 3% denaturant concentration (approximately equivalent to 1 cm distance within a 20% urea gradient gel). An experimental determination of gradient behavior can be achieved by perpendicular gel electrophoresis. Data from the perpendicular gels help to estimate the denaturant concentration range to use in parallel gel electrophoresis. For parallel gels, the gradient should be initially chosen with a 25% to 30% difference in denaturant concentration centered around the melting temperature of the domain (Myers *et al.*, 1987). Once optimized gel running conditions have been established, the method can be used for mutation screening.

6.5 THE USE OF TGGE/DGGE FOR MUTATION DETECTION

TGGE and DGGE have been used to investigate a large number of disease genes, some of which are listed in the following sections. Due to the relative ease of detecting heterozygous mutations owing to the occurrence of up to three novel bands, TGGE and DGGE have been particularly useful for disorders characterized by heterozygous mutations or frequent *de novo* mutations (reviewed in Fodde and Losekoot, 1994). In light of the effort involved in designing primers and optimizing conditions, TGGE or DGGE is generally reserved for situations when large numbers of samples are to be screened for mutations.

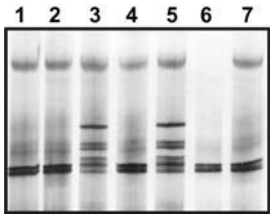


FIGURE 6.8 A heterozygous point mutation in exon 31 of the *NF1* gene detected by TGGE. The third and fifth lanes display a classical four-band appearance due to the presence of the heterozygous mutation g.5839C > T.

Most mutation screening protocols involve the simultaneous analysis of 24 or more samples on one parallel TGGE/DGGE. In general, altered band patterns are easy to spot. The classical appearance of heterozygous mutations (Figs 6.2 and 6.8) is due to the appearance of three additional bands. With some mutations, only one or two additional bands are seen. In the authors' experience, the specificity of TGGE/DGGE is exquisitely high. In other words, a false-positive four-band pattern has never been observed.

6.6 DETECTION RATE AND SENSITIVITY

By using DGGE, Myers and coworkers (1985b) detected an estimated 40% of the sequence variants in a DNA fragment up to 500 bp in their initial study. The use of GC-clamps, psoralen clamps, or bipolar clamping, which aids the formation of uniform low melting domains, significantly improved the detection rate of TGGE/DGGE, which in many cases approaches nearly 100%.

The sensitivity of TGGE/DGGE for detecting known mutations is generally reported to be nearly 100%, generally performing as well or better than other mutation detection methods (Abrams *et al.*, 1990; Ferec *et al.*, 1992; Gelfi *et al.*, 1997; Gejman *et al.*, 1998; Tchernitchko *et al.*, 1999; Zschocke *et al.*, 2000; Breton *et al.*, 2003). In one study with a panel of known mutations, DGGE detected 201 of 201 known mutations in the CFTR gene (Macek *et al.*, 1997). The reasons for lower reported detection rates of unknown mutations in some studies has been speculated to be due to genetic heterogeneity (Ferec *et al.*, 1999), clinical overdiagnosis (Katzke *et al.*, 2002) or location of mutations in intronic or promoter regions that were not included in the screening program. Optimization of the TGGE/DGGE assay conditions and primers, perhaps including the use of bipolar clamping (Gille and Gille, 2002), may increase sensitivity. In summary, the sensitivity of TGGE/DGGE, when properly used, is close to 100%.

TGGE/DGGE has also been shown to be very sensitive in the detection of mutations in situations where the

mutation sequence is present in proportions less than 50% (as is generally the case when heterozygous mutations are sought in genomic DNA). This has proved useful in detection of heteroplasmy in mitochondrial disorders with heteroplasmic proportions as low as 1% (Tully *et al.*, 2000), as well as in testing for residual disease in cancer (Ahnhudt *et al.*, 2001; Alkan *et al.*, 2001).

6.7 RELATED TECHNIQUES AND VARIANTS

A wide range of improvements and further developments of the principles underlying DGGE and TGGE have appeared in the last decade, the most important of which are briefly summarized below.

In *Broad range DGGE*, a single gel and a single set of conditions are used to screen all the exons of one gene (Guldberg and Guttler, 1994; Hayes *et al.*, 1999).

In *Multiplex DGGE*, several exons are simultaneously analyzed in one DGGE gel (Costes *et al.*, 1993a).

In *genomic DGGE (gDGGE)*, genomic DNA is digested with a restriction enzyme, electrophoresed by DGGE, transferred to nylon membrane and hybridized to a unique DNA probe (Borresen *et al.*, 1988).

In *constant DGGE (cDGGE)*, gels contain constant concentrations of denaturants. This allows an increased resolution of mutant fragments since they will constantly migrate with a different electrophoretic mobility through the whole length of the gel (Hovig *et al.*, 1991).

In *constant denaturant capillary electrophoresis (CDCE)*, DNA migrates through a 30 cm quartz capillary of 75 μm inner diameter, filled with a viscous polyacrylamide solution. A 10 cm part of the capillary, prior to the detector, is heated to a temperature permitting partial melting (see also previous chapter). Usually, the DNA is fluorescein labeled and detected by laser-induced fluorescence (Khrapko *et al.*, 1994). Separation of DNA fragments is achieved by the differential velocity of partly melted DNA in a medium with uniform denaturant concentration. Chip-based variants of temperature gradient capillary electrophoresis have recently been developed (Zhang *et al.*, 2007).

In *temporal temperature gradient gel electrophoresis (TTGE)*, a constant concentration of urea or formamide is used as in cDGGE but the temperature during the run is gradually increased (Yoshino *et al.*, 1991; Wiese *et al.*, 1995). The denaturant concentration (usually 6–8% urea) used in TTGE can be determined either from the theoretical melting curve or experimentally from a perpendicular DGGE.

In *microtemperature-gradient gel electrophoresis (μTGGE)*, a minimized gel (20 × 20 × 0.5 mm) leads to the reduction of the amount of DNA required and to shorter running times (approximately 12 min at 100 V, 10 mA). The method was used in microbial ecology and epidemiology (Biyani and Nishigaki, 2001; Tominaga, 2007).

In *double-gradient, denaturing gradient gel electrophoresis (DG-DGGE)*, in addition to the chemical denaturing gradient (formamide and urea) a second sieving gradient (e.g. 6–12% polyacrylamide gradient) is used (Cremonesi *et al.*, 1997).

The *two-dimensional DNA fingerprinting/two-dimensional gene scanning (TDGS, 2D-DNA typing)*, combines size separation of DNA fragments in the first dimension with their sequence-specific separation through DGGE in the second dimension (see also next chapter).

Denaturing HPLC (dHPLC) uses an ion-pair chromatography separation principle, combined with a precise control of the column temperature and optimized mobile phase gradient for separation of mutant DNA molecules (reviewed in Xiao and Oefner, 2001).

In *cycling gradient capillary electrophoresis (CGCE)*, DNA sequence variants are detected based on their differential migration in a polymer-filled capillary system. A cycling (oscillating) temporal temperature gradient is applied. This improvement enables utilization of a multiple injection technique, in which multiple samples are injected into the same capillary (or set of capillaries) separated by predefined time intervals of partial electrophoresis. A 96-capillary system is able to screen over 15,000 samples in 24 h (Minarik *et al.*, 2003).

6.8 TECHNICAL EQUIPMENT FOR TGGE/DGGE

In general, for DGGE, pre-existing vertical electrophoresis equipment with buffer-tank and combined heater/stirrer thermostat can be adapted. For TGGE, special equipment to achieve a constant temperature gradient is necessary.

The *Biometra TGGE* (Goettingen, Germany, http://www.biometra.de) system uses a temperature block powered by Peltier technology, which enables a strictly linear gradient that may allow more reproducible conditions than with conventional chemical gradients or temperature gradients using water baths. The Biometra TGGE system is available in two formats: a TGGE ''mini'' system operates small gels and is therefore suitable for fast, serial experiments and a TGGE maxi system provides a large separation distance and allows high parallel sample throughput.

The *DCode mutation detection system* (Bio-Rad Laboratories, Hercules USA, http://www.biorad.com) can be used to screen mutations by DGGE, TGGE, CDGE, TTGE, and by other techniques. The system performs TTGE by controlling the buffer temperature during the electrophoresis run. A temperature control module regulates the rate of temperature increase in a uniform and linear fashion.

Sooner Scientific (Garvin, USA, http://www.soonersci.com) offers five different sized DGGE Systems variants (for 2, 4 or 8 smaller gels or one large gel).

The *INGENYphorU system* (Ingeny International, GP Goes, The Netherlands, http://www.ingeny.com) is suitable for DGGE, TGGE, CDGE and other techniques.

6.9 APPLICATIONS OF TGGE/DGGE AND RELATED METHODS

TGGE/DGGE has been applied in an increasing number of studies. A recent search in PubMed database found over 2,200 citations (assessed in July 2008). The following applications have been described:

- Screening for polymorphisms in human genes: e.g. *COL1A2* gene (Borresen *et al.*, 1988), *SERPINA1* (alpha-1-antitrypsin; Hayes, 2003), *HBG1*/*HBG2* (human γ-globin genes, Patrinos *et al.*, 1998, 2001; Fig. 6.9).
- Mutation detection in human genes: e.g. p53 (Pignon *et al.*, 1994), *FBN1* (Tiecke *et al.*, 2001; Katzke *et al.*, 2002; Robinson *et al.*, 2002), *NF1* (Peters *et al.*, 1999; Fahsold *et al.*, 2000), Dystrophin gene (Hofstra *et al.*, 2004), *HBD* (human δ-globin, Papadakis *et al.*, 1997) and *HBB* (β-globin) genes (Losekoot *et al.*, 1990; Fig. 6.9), multiple endocrine neoplasia type 1 (Balogh *et al.*, 2004), and Y-chromosomal microdeletions (Bienvenu *et al.*, 2003).
- Mutation and single nucleotide polymorphism (SNP) detection with chip-based temperature gradient capillary electrophoresis (Zhang *et al.*, 2007).
- Mutation and polymorphism detection in mitochondrial DNA (Hanekamp *et al.*, 1996; Chen *et al.*, 1999).
- Analysis in microbial ecology, determination of bio-diversity of bacterial populations in soil, fresh or salt water (Muyzer and Smalla, 1998; van Elsas *et al.*, 2002), in rumen microbial populations (Deng *et al.*, 2008), and detection of microbes in food (Ercolini, 2004).
- Genome profiling and provisional microbial species identification on the basis of random PCR and TGGE (Watanabe *et al.*, 2002).
- Determining of bio-diversity in fecal or intestinal microflora (Tannock, 2002), and in endodontic infections (Siqueira *et al.*, 2005).
- HLA typing (Uhrberg *et al.*, 1994; Knapp, 2005).
- Analysis of proteins and antibody binding (Riesner *et al.*, 1991; Arakawa *et al.*, 1993).
- Clonality analysis of T-cell or T-cell receptors (Plonquet *et al.*, 2002; Lukowsky, 2003).
- Mutation detection and detection of variation between genomes of viral strains (Lu *et al.*, 2002; Motta *et al.*, 2002).
- Analysis of bio-diversity and polymorphisms in plants (Gomes *et al.*, 2003; Nikolcheva *et al.*, 2003).
- Examination of the fidelity of DNA polymerases (Keohavong and Thilly, 1989).
- High-throughput discovery of SNPs and other genetic polymorphisms (Hsia *et al.*, 2005; Maher *et al.*, 2006).

6.10 CONCLUSIONS

TGGE/DGGE and related methods provide a very high sensitivity and are relatively easy and cheap to perform once the assays have been designed and optimized. The main advantages consist in the high detection rate and specificity and improved heterozygote detection. The methodology is simple, non-radioactive, and relatively non-toxic. The disadvantages of TGGE and DGGE include mainly the limitation of PCR fragment length to about 500 nucleotides, the difficulties of analyzing GC-rich fragments, and the need for computer analysis of potential PCR fragments (which on the other hand can save time and money by eliminating the use of inadequate primers). However, once primers and conditions have been chosen, TGGE/DGGE is a robust and easy to perform mutation screening method. It is particularly well suited for the detection of known and unknown mutations in large genes, where high sensitivity is required and when large numbers of samples are to be tested.

ACKNOWLEDGEMENTS

The authors would like to thank Christoph Gille, Anja Klose, Angelika Pletschacher, and Horst Schlechte.

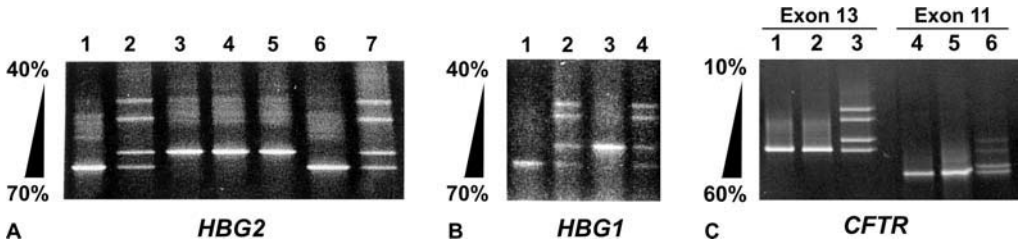


FIGURE 6.9 Mutation detection using DGGE analysis. **A.** Screening for the Gγ-158 C > T polymorphism in the promoter region of the human *HBG2* (Gγ-globin) gene. Lanes 3, 4, and 5 correspond to homozygous samples for that polymorphism; lanes 2 and 7 correspond to heterozygous samples; and lanes 1 and 6 correspond to samples that do not carry this polymorphism on either of the two alleles. **B.** DGGE analysis of the promoter region of the human *HBG1* (Aγ-globin) gene. Lanes 2 and 4 correspond to heterozygous cases for the g.-117G > A mutation, leading to the Greek type of non-deletional Hereditary Persistence of Fetal Hemoglobin (nd-HPFH); lane 3 corresponds to a homozygous case for the same mutation; lane 1 corresponds to a wild-type control (photo courtesy of George P. Patrinos). **C.** Mutation analysis of exons 11 and 13 of the CFTR gene. Lanes 1, 2, 4, and 5 correspond to the wild-type cases; lane 3 corresponds to a heterozygous case for the p.E822X nonsense mutation; lane 6 corresponds to a heterozygous case for the p.G542X nonsense mutation, leading to cystic fibrosis (photo courtesy of Dr. Angeliki Balassopoulou, Athens, Greece). The gradient of denaturing agents is depicted at the left side of each gel.

REFERENCES

Abrams, E.S., Murdaugh, S.E., and Lerman, L.S. (1990). Comprehensive detection of single base changes in human genomic DNA using denaturing gradient gel electrophoresis and a GC clamp. *Genomics* 7, 463–475.

Ahnhudt, C., Muche, J.M., Dijkstal, K., Sterry, W., and Lukowsky, A. (2001). An approach to the sensitivity of temperature-gradient gel electrophoresis in the detection of clonally expanded T-cells in cutaneous T-cell lymphoma. *Electrophoresis* 22, 33–38.

Alkan, S., Cosar, E., Ergin, M., and Hsi, E. (2001). Detection of T-cell receptor-gamma gene rearrangement in lymphoproliferative disorders by temperature gradient gel electrophoresis. *Arch. Pathol. Lab. Med.* 125, 202–207.

Arakawa, T., Hung, L., Pan, V., Horan, T.P., Kolvenbach, C.G., and Narhi, L.O. (1993). Analysis of the heat-induced denaturation of proteins using temperature gradient gel electrophoresis. *Anal. Biochem.* 208, 255–259.

Balogh, K., Patócs, A., Majnik, J., Rácz, K., and Hunyady, L. (2004). Genetic screening methods for the detection of mutations responsible for multiple endocrine neoplasia type 1. *Mol. Genet. Metab.* 83, 74–81.

Bienvenu, T., Patrat, C., and Jouannet, P. (2003). Molecular detection of Y chromosome microdeletions: a new approach based on the denaturing gradient gel electrophoresis. *Gynecol. Obstet. Fertil.* 31, 639–646.

Biyani, M., and Nishigaki, K. (2001). Hundredfold productivity of genome analysis by introduction of microtemperature-gradient gel electrophoresis. *Electrophoresis* 22, 23–28.

Borresen, A.L., Hovig, E., and Brogger, A. (1988). Detection of base mutations in genomic DNA using denaturing gradient gel electrophoresis (DGGE) followed by transfer and hybridization with gene-specific probes. *Mutat. Res.* 202, 77–83.

Breton, J., Sichel, F., Abbas, A., Marnay, J., Arsene, D., and Lechevrel, M. (2003). Simultaneous use of DGGE and DHPLC to screen TP53 mutations in cancers of the esophagus and cardia from a European high incidence area (Lower Normandy, France). *Mutagenesis* 18, 299–306.

Chen, T.J., Boles, R.G., and Wong, L.J. (1999). Detection of mitochondrial DNA mutations by temporal temperature gradient gel electrophoresis. *Clin. Chem.* 45, 1162–1167.

Costes, B., Fanen, P., Goossens, M., and Ghanem, N. (1993a). A rapid, efficient, and sensitive assay for simultaneous detection of multiple cystic fibrosis mutations. *Hum. Mutat.* 2, 185–191.

Costes, B., Girodon, E., Ghanem, N., Chassignol, M., Thuong, N.T., Dupret, D., and Goossens, M. (1993b). Psoralen-modified oligonucleotide primers improve detection of mutations by denaturing gradient gel electrophoresis and provide an alternative to GC-clamping. *Hum. Mol. Genet.* 2, 393–397.

Cremonesi, L., Firpo, S., Ferrari, M., Righetti, P.G., and Gelfi, C. (1997). Double-gradient DGGE for optimized detection of DNA point mutations. *Biotechniques* 22, 326–330.

Deng, W., Xi, D., Mao, H., and Wanapat, M. (2008). The use of molecular techniques based on ribosomal RNA and DNA for rumen microbial ecosystem studies: a review. *Mol. Biol. Rep.* 35, 265–274.

Ercolini, D. (2004). PCR-DGGE fingerprinting: novel strategies for detection of microbes in food. *J. Microbiol. Methods* 56, 297–314.

Fahsold, R., Hoffmeyer, S., Mischung, C., Gille, C., Ehlers, C., Kucukceylan, N., Abdel-Nour, M., Gewies, A., Peters, H., Kaufmann, D., Buske, A., Tinschert, S., and Nurnberg, P. (2000). Minor lesion mutational spectrum of the entire NF1 gene does not explain its high mutability but points to a functional domain upstream of the GAP-related domain. *Am. J. Hum. Genet.* 66, 790–818.

Ferec, C., Audrezet, M.P., Mercier, B., Guillermit, H., Moullier, P., Quere, I., and Verlingue, C. (1992). Detection of over 98% cystic fibrosis mutations in a Celtic population. *Nat. Genet.* 1, 188–191.

Ferec, C., Raguenes, O., Salomon, R., Roche, C., Bernard, J.P., Guillot, M., Quere, I., Faure, C., Mercier, B., Audrezet, M.P., Guillausseau, P.J., Dupont, C., Munnich, A., Bignon, J.D., and Le Bodic, L. (1999). Mutations in the cationic trypsinogen gene and evidence for genetic heterogeneity in hereditary pancreatitis. *J. Med. Genet.* 36, 228–232.

Fodde, R., and Losekoot, M. (1994). Mutation detection by denaturing gradient gel electrophoresis (DGGE). *Hum. Mutat.* 3, 83–94.

Gejman, P.V., Cao, Q., Guedj, F., and Sommer, S. (1998). The sensitivity of denaturing gradient gel electrophoresis: a blinded analysis. *Mutat. Res.* 382, 109–114.

Gelfi, C., Righetti, S.C., Zunino, F., Della Torre, G., Pierotti, M.A., and Righetti, P.G. (1997). Detection of p53 point mutations by double-gradient, denaturing gradient gel electrophoresis. *Electrophoresis* 18, 2921–2927.

Gille, C., Gille, A., Booms, P., Robinson, P.N., and Nurnberg, P. (1998). Bipolar clamping improves the sensitivity of mutation detection by temperature gradient gel electrophoresis. *Electrophoresis* 19, 1347–1350.

Gille, C., and Gille, A. (2002). TGGE-STAR: primer design for melting analysis using PCR gradient gel electrophoresis. *Biotechniques* 32, 264, 266, 268.

Gomes, N.C., Fagbola, O., Costa, R., Rumjanek, N.G., Buchner, A., Mendona-Hagler, L., and Smalla, K. (2003). Dynamics of fungal communities in bulk and maize rhizosphere soil in the tropics. *Appl. Environ. Microbiol.* 69, 3758–3766.

Guldberg, P., and Guttler, F. (1994). ''Broad-range'' DGGE for single-step mutation scanning of entire genes: application to human phenylalanine hydroxylase gene. *Nucleic Acids Res.* 22, 880–881.

Hanekamp, J.S., Thilly, W.G., and Chaudhry, M.A. (1996). Screening for human mitochondrial DNA polymorphisms with denaturing gradient gel electrophoresis. *Hum. Genet.* 98, 243–245.

Haran, T.E., Kahn, J.D., and Crothers, D.M. (1994). Sequence elements responsible for DNA curvature. *J. Mol. Biol.* 244, 135–143.

Hayes, V.M. (2003). Genetic diversity of the alpha-1-antitrypsin gene in Africans identified using a novel genotyping assay. *Hum. Mutat.* 22, 59–66.

Hayes, V.M., Wu, Y., Osinga, J., Mulder, I.M., van der Vlies, P., Elfferich, P., Buys, C.H., and Hofstra, R.M. (1999). Improvements in gel composition and electrophoretic conditions for broad-range mutation analysis by denaturing gradient gel electrophoresis. *Nucleic Acids Res.* 27, e29.

Hofstra, R.M., Mulder, I.M., Vossen, R., de Koning-Gans, P.A., Kraak, M., Ginjaar, I.B., van der Hout, A.H., Bakker, E., Buys, C.H., van Ommen, G.J., van Essen, A.J., and den Dunnen, J.T. (2004). DGGE-based whole-gene mutation scanning of the dystrophin gene in Duchenne and Becker muscular dystrophy patients. *Hum. Mutat.* 23, 57–66.

Hovig, E., Smith-Sorensen, B., Brogger, A., and Borresen, A.L. (1991). Constant denaturant gel electrophoresis, a modification of denaturing gradient gel electrophoresis, in mutation detection. *Mutat. Res.* 262, 63–71.

Hsia, A.P., Wen, T.J., Chen, H.D., Liu, Z., Yandeau-Nelson, M.D., Wei, Y., Guo, L., and Schnable, P.S. (2005). Temperature gradient capillary electrophoresis (TGCE) – a tool for the high-throughput discovery and mapping of SNPs and IDPs. *Theor. Appl. Genet.* 111, 218–225.

Kang, J., Kühn, J.E., Schäfer, P., Immelmann, A., and Henco, K. (1995). Quantification of DNA and RNA by PCR. In *PCR 2: A Practical Approach* (McPherson, M.J., Hames, B.D., and Taylor, G.R., eds), pp. 119–133. IRL Press, Oxford.

Katzke, S., Booms, P., Tiecke, F., Palz, M., Pletschacher, A., Turkmen, S., Neumann, L.M., Pregla, R., Leitner, C., Schramm, C., Lorenz, P., Hagemeier, C., Fuchs, J., Skovby, F., Rosenberg, T., and Robinson, P.N. (2002). TGGE screening of the entire FBN1 coding sequence in 126 individuals with Marfan syndrome and related fibrillinopathies. *Hum. Mutat.* 20, 197–208.

Ke, S.H., and Wartell, R.M. (1995). Influence of neighboring base pairs on the stability of single base bulges and base pairs in a DNA fragment. *Biochemistry* 34, 4593–4600.

Keohavong, P., and Thilly, W.G. (1989). Fidelity of DNA polymerases in DNA amplification. *Proc. Natl. Acad. Sci. USA* 86, 9253–9257.

Khrapko, K., Hanekamp, J.S., Thilly, W.G., Belenkii, A., Foret, F., and Karger, B.L. (1994). Constant denaturant capillary electrophoresis (CDCE): a high resolution approach to mutational analysis. *Nucleic Acids Res.* 22, 364–369.

Knapp, L.A. (2005). Denaturing gradient gel electrophoresis and its use in the detection of major histocompatibility complex polymorphism. *Tissue Antigens* 5, 211–219.

Lerman, L.S., and Silverstein, K. (1987). Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. *Methods Enzymol.* 155, 482–501.

Losekoot, M., Fodde, R., Harteveld, C.L., van Heeren, H., Giordano, P.C., and Bernini, L.F. (1990). Denaturing gradient gel electrophoresis and direct sequencing of PCR amplified genomic DNA: a rapid and reliable diagnostic approach to beta thalassaemia. *Br. J. Haematol.* 76, 269–274.

Lu, Q., Hwang, Y.T., and Hwang, C.B. (2002). Mutation spectra of herpes simplex virus type 1 thymidine kinase mutants. *J. Virol.* 76, 5822–5828.

Lukowsky, A. (2003). Clonality analysis by T-cell receptor gamma PCR and high-resolution electrophoresis in the diagnosis of cutaneous T-cell lymphoma (CTCL). *Methods Mol. Biol.* 218, 303–320.

Macek, M., Jr., Mercier, B., Mackova, A., Miller, P.W., Hamosh, A., Ferec, C., and Cutting, G.R. (1997). Sensitivity of the denaturing gradient gel electrophoresis technique in detection of known mutations and novel Asian mutations in the CFTR gene. *Hum. Mutat.* 9, 136–147.

Maher, P.M., Chou, H.H., Hahn, E., Wen, T.J., and Schnable, P.S. (2006). GRAMA: genetic mapping analysis of temperature gradient capillary electrophoresis data. *Theor. Appl. Genet.* 113, 156–162.

Minarik, M., Minarikova, L., Bjorheim, J., and Ekstrom, P.O. (2003). Cycling gradient capillary electrophoresis: a low-cost tool for high-throughput analysis of genetic variations. *Electrophoresis* 24, 1716–1722.

Motta, F.C., Rosado, A.S., and Couceiro, J.N. (2002). Standardization of denaturing gradient gel electrophoresis for mutant screening of influenza A (H3N2) virus samples. *J. Virol. Methods* 101, 105–115.

Murdaugh, S.E., and Lerman, L.S. (1996). DGGE. Denaturing gradient gel electrophoresis and related techniques. In *Laboratory Protocols for Mutation Detection* (Landegren, U., ed.), pp. 33–37. Oxford University Press, Oxford.

Muyzer, G., and Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek* 73, 127–141.

Myers, R.M., Fischer, S.G., Lerman, L.S., and Maniatis, T. (1985a). Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Res.* 13, 3131–3145.

Myers, R.M., Lumelsky, N., Lerman, L.S., and Maniatis, T. (1985b). Detection of single base substitutions in total genomic DNA. *Nature* 313, 495–498.

Myers, R.M., Maniatis, T., and Lerman, L.S. (1987). Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol.* 155, 501–527.

Nikolcheva, L.G., Cockshutt, A.M., and Barlocher, F. (2003). Determining diversity of freshwater fungi on decaying leaves: comparison of traditional and molecular approaches. *Appl. Environ. Microbiol.* 69, 2548–2554.

Papadakis, M., Papapanagiotou, E., and Loutradi-Anagnostou, A. (1997). Scanning method to identify the molecular heterogeneity of delta-globin gene especially in delta-thalassemias: detection of three novel substitutions in the promoter region of the gene. *Hum. Mutat.* 9, 465–472.

Patrinos, G.P., Kollia, P., Loutradi-Anagnostou, A., Loukopoulos, D., and Papadakis, M.N. (1998). The Cretan type of non-deletional hereditary persistence of fetal hemoglobin [Agamma-158C- > T] results from two independent gene conversion events. *Hum. Genet.* 102, 629–634.

Patrinos, G.P., Kollia, P., Papapanagiotou, E., Loutradi-Anagnostou, A., Loukopoulos, D., and Papadakis, M.N. (2001). Agamma-haplotypes: a new group of genetic markers for thalassemic mutations inside the 5′ regulatory region of the human Agamma-globin gene. *Am. J. Hematol.* 66, 99–104.

Peters, H., Hess, D., Fahsold, R., and Schulke, M. (1999). A novel mutation L1425P in the GAP-region of the NF1 gene detected by temperature gradient gel electrophoresis (TGGE). Mutation in brief no. 230. Online. *Hum. Mutat.* 13, 337.

Pignon, J.M., Vinatier, I., Fanen, P., Jonveaux, P., Tournilhac, O., Imbert, M., Rochant, H., and Goossens, M. (1994). Exhaustive analysis of the P53 gene coding sequence by denaturing gradient gel electrophoresis: application to the detection of point mutations in acute leukemias. *Hum. Mutat.* 3, 126–132.

Plonquet, A., Gherardi, R.K., Creange, A., Antoine, J.C., Benyahia, B., Grisold, W., Drlicek, M., Dreyfus, P., Honnorat, J., Khouatra, C., Rouard, H., Authier, F.J., Farcet, J.P., Delattre, J.Y., and Delfau-Larue, M.H. (2002). Oligoclonal T-cells in blood and target tissues of patients with anti-Hu syndrome. *J. Neuroimmunol.* 122, 100–105.

Riesner, D., Henco, K., and Steger, G. (1991). Temperature-gradient gel electrophoresis: a method for the analysis of conformational transitions and mutations in nucleic acids and proteins. In *Advances in Electrophoresis* (Chrambach, A., Dunn, M.J., and Radola, B.J., eds), Vol. 4, pp. 169–250. VCH, Verlagsgesellschaft, Weinheim.

Robinson, P.N., Booms, P., Katzke, S., Ladewig, M., Neumann, L., Palz, M., Pregla, R., Tiecke, F., and Rosenberg, T. (2002). Mutations of FBN1 and genotype-phenotype correlations in Marfan syndrome and related fibrillinopathies. *Hum. Mutat.* 20, 153–161.

Sheffield, V.C., Cox, D.R., Lerman, L.S., and Myers, R.M. (1989). Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc. Natl. Acad. Sci. USA* 86, 232–236.

Siqueira, J.F., Jr., Rôças, I.N., and Rosado, A.S. (2005). Application of denaturing gradient gel electrophoresis (DGGE) to the analysis of endodontic infections. *J. Endod.* 31, 775–782.

Steger, G. (1994). Thermal denaturation of double-stranded nucleic acids: prediction of temperatures critical for gradient gel electrophoresis and polymerase chain reaction. *Nucleic Acids Res.* 22, 2760–2768.

Tannock, G.W. (2002). Analysis of the intestinal microflora using molecular methods. *Eur. J. Clin. Nutr.* 56(Suppl. 4), S44–S49.

Tchernitchko, D., Lamoril, J., Puy, H., Robreau, A.M., Bogard, C., Rosipal, R., Gouya, L., Deybach, J.C., and Nordmann, Y. (1999).

Evaluation of mutation screening by heteroduplex analysis in acute intermittent porphyria: comparison with denaturing gradient gel electrophoresis. *Clin. Chim. Acta* 279, 133–143.

Tiecke, F., Katzke, S., Booms, P., Robinson, P.N., Neumann, L., Godfrey, M., Mathews, K.R., Scheuner, M., Hinkel, G.K., Brenner, R.E., Hovels-Gurich, H.H., Hagemeier, C., Fuchs, J., Skovby, F., and Rosenberg, T. (2001). Classic, atypically severe and neonatal Marfan syndrome: twelve mutations and genotype-phenotype correlations in FBN1 exons 24–40. *Eur. J. Hum. Genet.* 9, 13–21.

Tominaga, T. (2007). Rapid determination of multi-locus sequence types of Listeria monocytogenes by microtemperature-gradient gel electrophoresis. *J. Microbiol. Methods* 70, 471–478.

Top, B. (1992). A simple method to attach a universal 50-bp GC-clamp to PCR fragments used for mutation analysis by DGGE. *PCR Methods Appl.* 2, 83–85.

Tully, L.A., Parsons, T.J., Steighner, R.J., Holland, M.M., Marino, M. A., and Prenger, V.L. (2000). A sensitive denaturing gradient-Gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. *Am. J. Hum. Genet.* 67, 432–443.

Uhrberg, M., Hinney, A., Enczmann, J., and Wernet, P. (1994). Analysis of the HLA-DR gene locus by temperature gradient gel electrophoresis and its application for the rapid selection of unrelated bone marrow donors. *Electrophoresis* 15, 1044–1050.

van Elsas, J.D., Garbeva, P., and Salles, J. (2002). Effects of agronomical measures on the microbial diversity of soils as related to the suppression of soil-borne plant pathogens. *Biodegradation* 13, 29–40.

Watanabe, T., Saito, A., Takeuchi, Y., Naimuddin, M., and Nishigaki, K. (2002). A database for the provisional identification of species using only genotypes: web-based genome profiling. *Genome Biol.* 3. RESEARCH0010.

Wiese, U., Wulfert, M., Prusiner, S.B., and Riesner, D. (1995). Scanning for mutations in the human prion protein open reading frame by temporal temperature gradient gel electrophoresis. *Electrophoresis* 16, 1851–1860.

Xiao, W., and Oefner, P.J. (2001). Denaturing high-performance liquid chromatography: a review. *Hum. Mutat.* 17, 439–474.

Yoshino, K., Nishigaki, K., and Husimi, Y. (1991). Temperature sweep gel electrophoresis: a simple method to detect point mutations. *Nucleic Acids Res.* 19, 3153.

Zhang, H.D., Zhou, J., Xu, Z.R., Song, J., Dai, J., Fang, J., and Fang, Z.L. (2007). DNA mutation detection with chip-based temperature gradient capillary electrophoresis using a slantwise radiative heating system. *Lab Chip* 9, 1162–1170.

Zschocke, J., Quak, E., Guldberg, P., and Hoffmann, G.F. (2000). Mutation analysis in glutaric aciduria type I. *J. Med. Genet.* 37, 177–181.

Chapter 7

Real-Time Polymerase Chain Reaction

Lut Overbergh, Anna-Paula Giulietti, Dirk Valckx and Chantal Mathieu
Laboratory for Experimental Medicine and Endocrinology (LEGENDO), Catholic University of Leuven, Leuven, Belgium

7.1 HISTORY OF PCR

The polymerase chain reaction (PCR) is a technique based on the exponential amplification of DNA by the thermostable *Thermus aquaticus* (Taq) polymerase. The method uses a pair of synthetic oligonucleotide primers, each hybridizing to one strand of a double-stranded DNA (dsDNA) target, with the pair spanning a region that will be exponentially amplified. The annealed primers act as a substrate for the Taq DNA polymerase, creating a complementary DNA strand via sequential addition of deoxynucleotides. The process classically consists of three steps:

1. A denaturation step at 94 or 95°C.
2. Primer annealing to the ssDNA strands at 60°C.
3. Primer extension at 72°C.

The technique was first described by Kary Mullis in the 1980s (Mullis *et al.*, 1986), for which he received the Nobel Prize in 1993. This classic end-point PCR is a yes/no reaction because it measures DNA product formation after a fixed number of cycles, that is, in the plateau phase of the reaction, thus resulting in qualitative information on the presence or absence of a certain gene or mRNA. This PCR technique has become one of the most influential tools in the biological and medical sciences (Guyer and Koshland, 1989).

Over the years, numerous adaptations and applications to this classic end-point PCR have been described, including semiquantitative PCR, quantitative competitive PCR, and its latest innovation, real-time PCR. A first adaptation measures PCR product accumulation during the exponential phase of the reaction, resulting in semiquantitative data. This method necessitates interruption of the PCR reaction after an experimentally determined number of cycles. Furthermore, samples from a single experimental setup can be analyzed only over a relatively small linear range.

Alternatively, competitive PCR has been developed, resulting in quantitative data. This method, however, needs extensive optimization, since it requires the coamplification of an internal cDNA or RNA control (competitor) with the unknown sample in the same tube. Quantification is performed by titrating an unknown amount of target template against a dilution series of known amounts of the standard. The internal control consists of target DNA or RNA that has been slightly modified. Thus, one set of primers is designed that coamplifies the target and the competitor, with the same efficiency, although they can be distinguished from each other (by, for instance, difference in length or restriction sites; Clementi *et al.*, 1994). This method provides a strategy for accurate quantification, but the construction of internal standards is technically sophisticated and labor intensive.

For the detection of PCR products using either of these methods several detection techniques can be used, which all require excessive post-PCR manipulations. The most classically used are agarose gel electrophoresis with ethidium bromide or SYBR Green staining, fluorescent labeling and analysis using polyacrylamide gels, radioactive labeling, and Southern blotting or detection by phosphorimaging. Major drawbacks using these classical detection systems are the use of hazardous chemicals and the potential risk of laboratory contamination. Moreover, all these post-PCR manipulations are very time consuming.

The development of a new procedure in the mid-1990s for the analysis and quantification of DNA or RNA, based on fluorescence-kinetic RT-PCR, enabled quantification of the PCR product in real-time (Higuchi *et al.*, 1993; Heid *et al.*, 1996; Gibson *et al.*, 1996). This sensitive and accurate technique permits quantification of PCR product during the exponential phase of the PCR reaction. This is in full contrast to the classic end-point assays, as they are designed to provide information as rapidly as the amplification process itself, thus requiring no post-PCR manipulations. The development of this real-time PCR again had a revolutionary impact on molecular research and diagnostics.

7.2 PRINCIPLE OF REAL-TIME PCR

7.2.1 Real-Time PCR Using Hydrolysis Probes: The Classic TaqMan System

Real-time PCR was first described using hydrolysis probes (Heid *et al.*, 1996; Gibson *et al.*, 1996). The technique is based on the coupling of two important processes. First, the construction of dual-labeled oligonucleotide probes, also called hydrolysis or TaqMan probes, which emit a fluorescent signal upon cleavage, based on the principle of fluorescence resonance energy transfer (FRET, Stryer, 1978; Cardullo *et al.*, 1988). Second, the discovery that the Taq DNA polymerase possesses a $5' > 3'$ exonuclease activity, which can be exploited to degrade the fluorescent labeled probe (Holland *et al.*, 1991). The oligonucleotide probe used in this assay is non-extendable at its $3'$ end and is dual labeled, with a reporter fluorochrome, for example FAM (6-carboxyfluorescein), and a quencher fluorochrome, for example TAMRA (6-carboxy-tetramethylrhodamine). It is designed to anneal to the target sequence internally of the primers, during the annealing and extension phase of the PCR reaction. In its free, intact form no fluorescent emission can be measured, because fluorescent emission of the reporter dye is absorbed by the quenching dye. However, upon annealing of the probe to one of the target strands, the probe will become degraded by the $5'$–$3'$ exonuclease activity of the Taq polymerase. Consequently, the reporter and quencher dye become separated and the reporter dye emission is no longer transferred to the quenching dye (no more FRET), resulting in an increase of reporter fluorescent emission (e.g. for FAM at 518 nm). This process occurs in every cycle and does not interfere with the exponential accumulation of PCR product. The increase in fluorescence is measured cycle by cycle and directly correlates with the amount of PCR product formed (Heid *et al.*, 1996; Gibson *et al.*, 1996; Fig. 7.1).

Apart from the classically used fluorescent reporter dye FAM, other reporter dyes are available. These include, for instance, TET (tetrachloro-6-carboxyfluorescein), JOE (2,7,-dimethoxy-4,5-dichloro-6-carboxyfluorescein), HEX (hexacholoro-6-carboxyfluorescein), VIC, Texas Red or Cy5. The choice of different reporter dyes, with a minor overlap in fluorescent emission spectra, makes it possible to perform multiplex PCR reactions, thus simultaneously amplifying different DNA targets. Similarly, there is a choice between different quencher dyes. The most classically used quencher dye is TAMRA. DABCYL (4-(4′A-dimethylaminophenylazol) benzoic acid) can also be used as a quencher dye, but its use is much more prevalent in the molecular beacons probes (see section 7.2.4). An advantage of using DABCYL is its reduced auto-fluorescence compared to TAMRA. More recently, several dark quencher fluorochromes have become available. These absorb the energy that is emitted by the reporter dye and release it as heat rather than as fluorescence. This results in a lower background signal and thus a higher sensitivity. It is therefore to be expected that dark quenching dyes will become the standard, replacing the TAMRA quenchers in the near future.

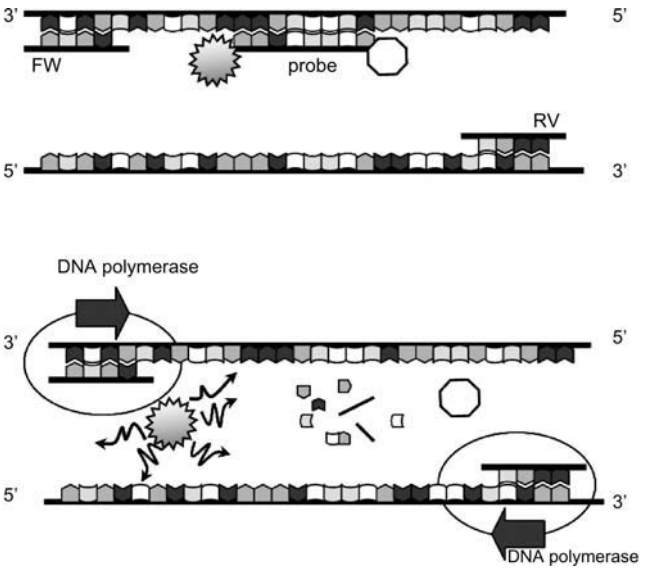


FIGURE 7.1 Hydrolysis or TaqMan probes. The dual-labeled TaqMan probe is cleaved by the $5'$ exonuclease activity of Taq DNA polymerase, during the extension step of the PCR reaction. The quencher and fluorophore are brought apart, which results in an increase in fluorescent emission. FW, forward primer; RV, reverse primer; gray 16-star, fluorophore; white hexagon, quencher.

7.2.2 Real-Time PCR Using DNA Intercalating Dyes

Another widely used real-time PCR technique is based on the detection and quantification of PCR products using fluorescent DNA intercalating dyes. The principle of this technique was first described by Higuchi (1993), who monitored the increase in ethidium bromide fluorescence using a charged coupled devise camera, a method that was referred to as kinetic PCR. More recently, SYBR Green I, which is less toxic than ethidium bromide, is widely used as a dye, which incorporates into dsDNA (see Fig. 7.2). Other asymmetric cyanine dyes have become more recently available, such as SYBR Green Save, BEBO (4-[(3-methyl-6-(benzothiazol-2-yl)-2,3-dihydro-(benzo-1,3-thiazole)-2-methylidene)]-1-methyl-pyridinium iodide) (Martin *et al.*, 2003) or BOXTO (4-[6-(benzoxazole-2-yl-(3-methyl-)-2,3-dihydro-(benzo-1,3-thiazole)-2-methylidene)]-1-methyl-quinolinium chloride) (Ahmad, 2007). These dyes incorporate into the minor groove of dsDNA, by which its fluorescence is greatly enhanced. During the PCR reaction, the amount of double-stranded target will increase exponentially, paralleled by an increase in SYBR Green I incorporation and fluorescent emission. In each cycle the fluorescent emission will increase gradually during the extension phase of the reaction, and will be low or absent during the denaturation phase.

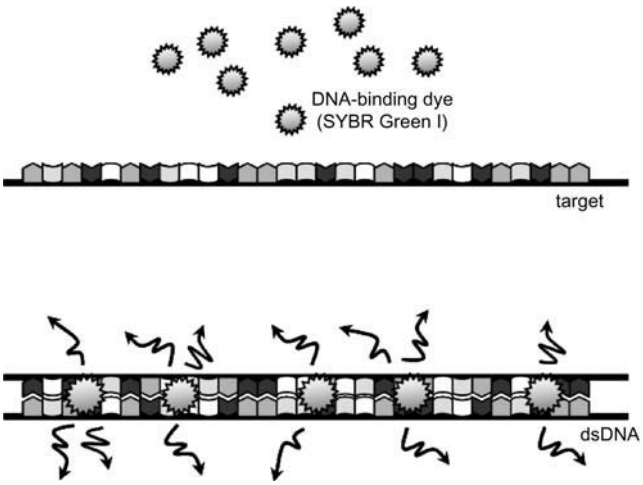


FIGURE 7.2 DNA intercalating dyes. DNA intercalating dyes incorporate into the minor groove of dsDNA. When this occurs, the dye will emit fluorescent light.

The greatest advantage of this method as opposed to the use of fluorescent-labeled probes is that it can be used with any pair of primers for any target. It is a cheaper alternative and requires less specialist knowledge than the design of fluorescent-labeled probes. Consequently, specificity is diminished due to the risk of amplifying non-specific PCR products or primer/dimers. To discriminate between specific and non-specific PCR products, a melting curve analysis has to be performed, after termination of the PCR reaction (Ririe *et al.*, 1997). In this way, the fraction of fluorescence originating from the specific target can be distinguished from that originating from primer/dimers or non-specific amplification products. This analysis is performed by slowly increasing the temperature from 40°C to 95°C, during which fluorescent emission is monitored continuously. Fluorescent emission will be low at low temperatures – when all PCR products are double stranded – and it will increase dramatically around the melting temperature of the PCR product. The rationale behind this melting curve analysis is based on the fact that PCR products of different length will have different melting temperatures, which will result in distinct peaks when plotting the first negative derivative of the fluorescence versus the temperature. Because of their shorter length, primer dimers will, for instance, have a lower melting temperature as compared to the specific PCR product. This will consequently lead to a fluorescent peak at a lower temperature, when plotting the first negative derivative of the fluorescence versus the temperature (see Fig. 7.3).

7.2.3 Real-Time PCR Using Hybridization Probes

Real-time PCR analysis with hybridization probes uses two juxtaposed sequence-specific probes, also known as HybProbes. The development of this system was first described in the mid-1980s (Heller and Morrison, 1985). Each probe has a single fluorescent reporter, one a donor fluorophore at its 3′ end and the other an acceptor fluorophore at its 5′ end. The sequences of the two probes are designed to anneal to the target sequences in very close proximity to each other (i.e. within 1–5 nucleotides), in a head-to-tail arrangement, bringing the two dyes very close to each other (see Fig. 7.4). As long as the probe is in its free, unbound form, the fluorescent signal from the reporter dye is not detected. However, during the annealing phase of the PCR reaction, the probes anneal to the target sequence and the two fluorophores come in close proximity to each other. This will result in emission of light from the donor fluorochrome, which will excite the acceptor fluorochrome, a process referred to as resonance energy transfer. The dye in one of the probes transfers energy, allowing the other one to dissipate fluorescence at a different wavelength. The amount of fluorescence emitted can be measured during the annealing phase of the PCR reaction, and is directly proportional to the amount of target DNA generated during the PCR process (Bernard and Wittwer, 2000).

A variant of these classic hybridization probes uses a fluorescent-labeled primer/probe combination. In this case, a 5′-labeled hybridization probe is designed to anneal to the PCR strand in close proximity to one of the PCR primers, which has a fluorophore at its 3′ end. This method requires that the fluorescent-labeled primer be positioned near the probe, usually within 5 base pairs, to allow adequate resonance energy transfer with the complementary probe (von Ahsen *et al.*, 2000).

Hybridization probes are usually constructed with FAM as the 3′ donor fluorophore, and a range of different acceptor fluorophores are commonly used (e.g. ROX, Cy5, LC Red640, LC Red705) as the 5′ acceptor fluorophore.

7.2.4 Real-Time PCR Using Molecular Beacons

Molecular beacons are probes that contain a stem-and-loop structure in their intact, unbound form. They are dual-labeled, with a fluorophore linked to one end of the molecule and a quencher linked to the other end (Tyagi and Kramer, 1996). Fluorescence is quenched when the probe is in its hairpin-like structure due to the proximity between quencher and fluorophore, resulting in complete absorption of any photons emitted by the fluorophore. When the probe sequence in the loop anneals to a complementary target sequence, a conformational change allows the formation of a linear structure whereby fluorescent energy transfer is no more occurring, resulting in an increase in fluorescence emission (see Fig. 7.5). Molecular beacons are especially suitable for identifying point mutations. They can distinguish targets that differ by only a single nucleotide and they are significantly more specific than conventional hydrolysis probes of equivalent length.

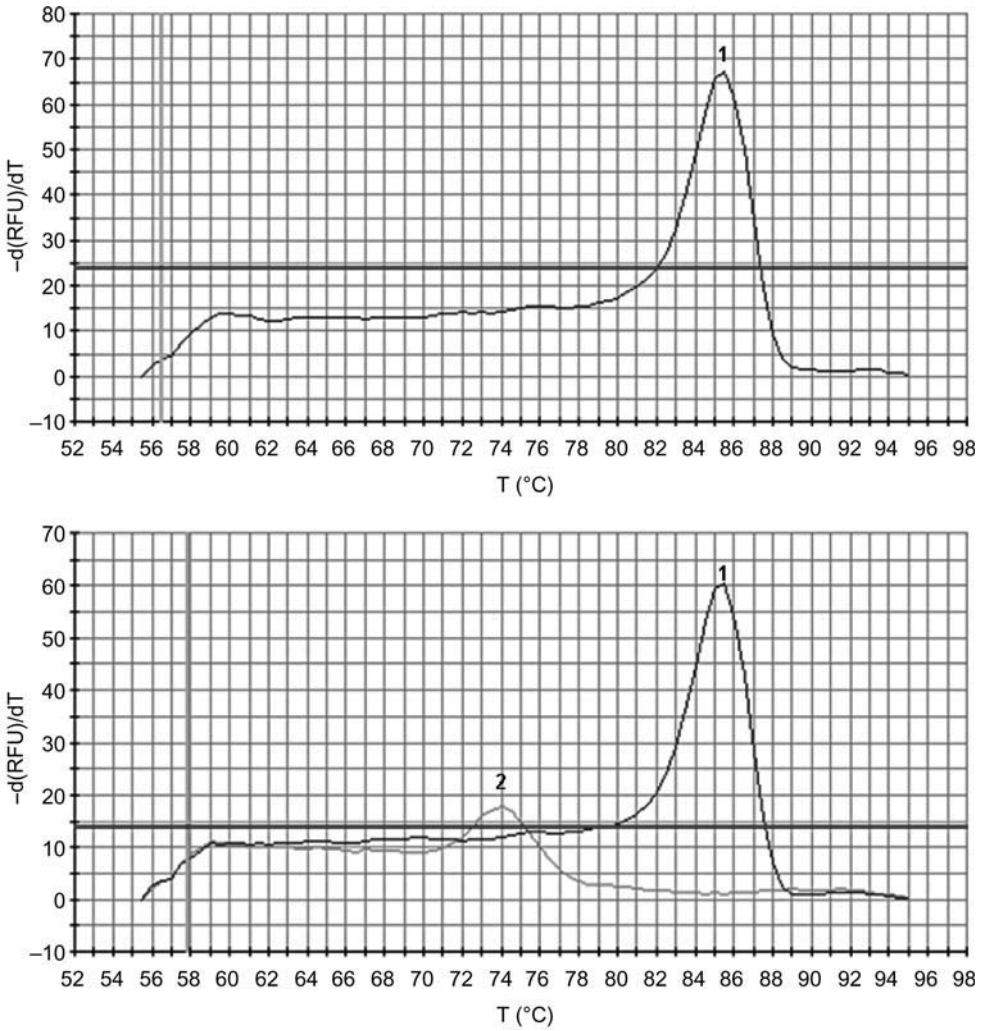


FIGURE 7.3 Post-PCR melting curve analysis when using DNA intercalating dyes. The first negative derivative of the fluorescence (y-axis) is plotted against the temperature (x-axis). In case of specific DNA amplification a single peak (peak 1) is observed (panel A); while in case of primer/dimer formation a second peak at lower temperature (peak 2) is observed (panel B).

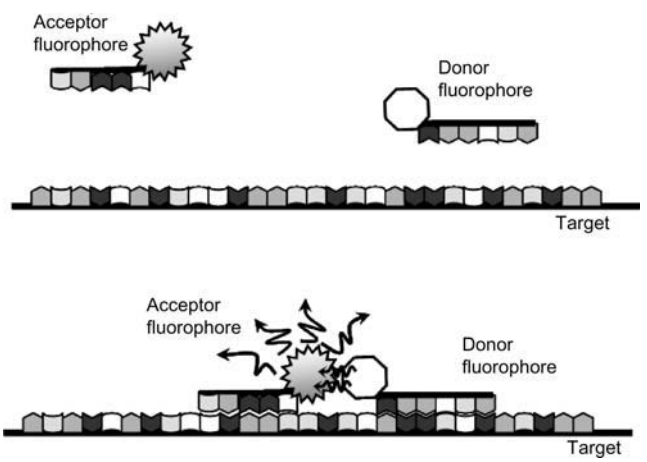


FIGURE 7.4 Hybridization probes. Two probes are used, one carrying an acceptor fluorophore and the other one carrying a donor fluorophore. When both fluorophores are brought into close proximity, i.e. when the probes anneal to the target sequence, the donor is able to excite the acceptor through FRET and fluorescence emission will occur.

7.2.5 Real-Time PCR Using Scorpions

Real-time PCRs using the scorpions system is carried out with two oligonucleotides: a primer and a fluorescent molecule that combine the primer and probe function (Whitcombe *et al.*, 1999). In the primer/probe combination, a primer sequence is linked to a specific probe sequence that is held in a hairpin-loop form. The stem-loop tail is separated from the primer by a PCR blocker to prevent the Taq DNA polymerase from amplifying the stem-loop sequence. This configuration brings the fluorophore in close proximity to the quencher and avoids fluorescence. As soon as annealing between the primer/probe and the target occurs, the hairpin is opened and the fluorophore and quencher are separated, resulting in an increase in fluorescence emission (see Fig. 7.6). Scorpions differ from molecular beacons and hydrolysis probes, in that their structure promotes a unimolecular probing mechanism. This results in a stronger fluorescent signal, especially under fast

cycling conditions. Scorpions represent a relatively new chemistry, validated for mutation detection, but most likely it will be adapted to other assays (Thelwell *et al.*, 2000).

7.2.6 Real-Time PCR Using Other Detection Chemistries

More recently, other sophisticated detection chemistries, which will not be discussed in further detail here, have been developed for real-time PCR. Like hybridization probes, molecular beacons, and scorpions, these systems all rely on the FRET principle, although without the need for hydrolysis by the nuclease activity of the Taq DNA polymerase. The list of possible detection chemistries is continuously growing; some examples are ResonSense probes, Hy-beacon probes, Light-up probes, Simple probes, Lion probes, AllGlo probes, and Displacement probes.

In most cases, these labeled probes are based on the classical nucleic acids. For specific applications also synthetic analogs are used. This is, for instance, the case in the locked nucleic acid (LNA) probes. These are typically very short probes, resulting in a very high binding efficiency to their complementary DNA, making them particularly suitable for single nucleotide polymorphism detection (Johnson *et al.*, 2004).

Another specific example of special interest is minor groove-binding probes (MGB). These are, like the TaqMan probes, hydrolysis probes with a minor groove-binding molecule attached to the end of the probe, again with the aim of allowing construction of shorter probe designs with increased specificity. These are more sensitive especially to single base mismatches, and are therefore ideally suitable for single nucleotide polyporphism (SNP) detection and allelic discrimination (Kutyavin *et al.*, 2000).

Next to specific modifications in the probes, many protocols have also been designed, making use of modified primers. Apart from the Scorpion primers, described above, some examples are LUX primers, Ampliflour primers, and the QZyme system.

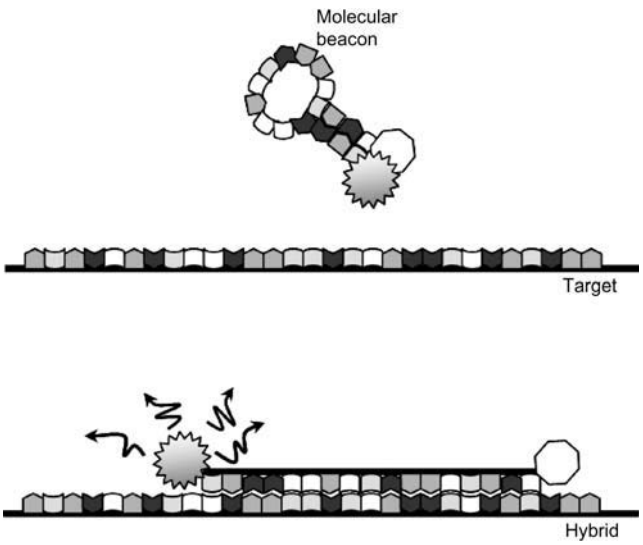


FIGURE 7.5 Molecular beacons. Molecular beacons are hairpin-shaped probes. The fluorophore and quencher are in close proximity when the probe is in its free, unbound state. When the probe anneals to the complementary target sequence, its conformation will change. Thereby the quencher and fluorophore are brought apart, causing fluorescence emission.

7.3 REAL-TIME THERMAL CYCLERS

Since the development of real-time PCR in the mid-1990s, instrumentation systems have already undergone an extensive evolution. The first available instrument, the 7700 SDS

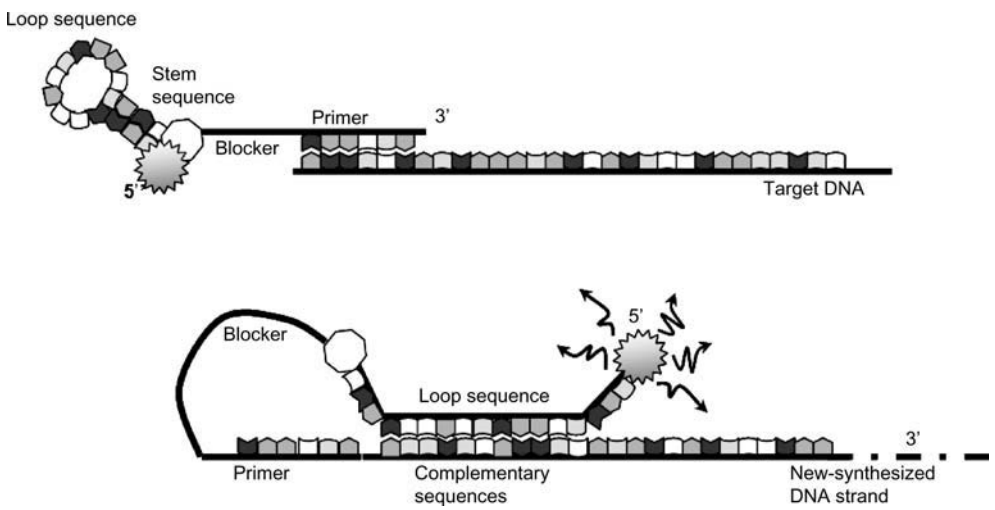


FIGURE 7.6 Scorpions. Scorpions are single-stranded dual-labeled fluorescent primer/probes with a hairpin-shaped structure. The primer/probe contains a 5′ end fluorophore and an internal quencher dye directly linked to the 5′ end of a PCR primer via a PCR blocker. In the unbound form, no fluorescent emission occurs. When binding to the target sequence, the hairpin shape unfolds and the loop region of the probe hybridizes intramolecularly to the newly synthesized target sequence. This results in fluorescent emission, because fluorophore and quencher are separated.

from Applied Biosystems, took up almost an entire bench area, made use of an expensive laser as the light source, and had to be placed in an air-conditioned laboratory. This instrument was unable to produce a true analysis in real-time, since data could be viewed only after termination of the PCR reaction. As the technology improved, systems became available that allowed detection of a PCR product from the moment it is formed, thus in pragmatic real-time. These newer machines are much smaller in size, and the laser has been changed for less precious tungsten/halogen or light-emitting diode lamps. Recently, instruments became available that can even be backpacked and used for on-site analysis. These instruments are able to give results in less then 30 min of analysis time, a rather important feature, for instance, for the on-site detection of pathogen outbreaks. Because of the wide range of instruments developed by different companies, prices have dropped significantly and instruments are becoming a standard tool for routine molecular and diagnostic laboratories.

Although all possible applications can be performed on all instruments, each one has specific advantages and disadvantages. The choice for a specific real-time thermal cycler therefore is dependent on the specific application one is focusing on. For quantification of gene expression, for instance, the length of a run is not the most important factor, although one would like to choose an instrument which can analyze a large number of samples simultaneously. Indeed, a high turnover of sample numbers will be performed more easily in 96-well or even 384-well systems. On the other hand, if the main application is pathogen identification, for example, velocity is the main issue and the best choice would be a rapid thermal cycler.

Most of the thermal cyclers are able to perform multiplex PCR, are sensitive and accurate, and increasingly present user-friendly software. In Table 7.1, the main features of the most commonly used available real-time PCR thermal cyclers are summarized.

7.4 HOW DATA ARE OBTAINED

When performing real-time PCR, the ability to monitor the amplification process of the PCR reaction in real-time revolutionizes the way in which the data are obtained. Typically, the reactions are characterized by the point in time during PCR cycling when amplification of a PCR product reaches a certain detection level, as opposed to end-point detection, where the amount of product formed is measured after a fixed number of cycles. Furthermore, quantification is based on the inherent property of a PCR reaction, that the more input DNA copies one starts with, the fewer cycles of PCR amplification it takes to make a specific number of amplification products. Finally, the fact that the formation of amplification product linearly correlates with the amount of fluorescence emission is exploited in the real-time PCR assay.

In practice, using any of the developed detection chemistries on any of the available instruments, the increase in fluorescence emission can be read by a sequence detector in real-time, during the course of the reaction, and is a direct consequence of target amplification during PCR. In Fig. 7.7, a typical amplification plot is shown, in which the terms and definitions routinely used in real-time quantitative PCR are illustrated. During the initial cycles of the PCR reaction, there is little or no change in fluorescence signal. This stage is the baseline of the amplification plot. The fluorescence emission of the product at each time point is measured during PCR cycling, and is defined as Rn^+. Analogously, Rn^- is the fluorescent emission of the baseline. The increase in fluorescence is calculated by the computer software program, and is plotted on the y-axis as the DRn value, using the equation $DRn = Rn^+ - Rn^-$. Thus, this value directly correlates with the probe degradation (in case of hydrolysis or TaqMan probes) during the PCR process, and consequently with the formation of specific PCR product. An arbitrary threshold is chosen, based on the variability of the baseline, usually determined as ten times the standard deviation of the baseline, set from cycle 3 to 15. This, of course, can be changed manually for each individual experiment if necessary. Threshold cycle (Ct) values are then calculated by determining the point at which the fluorescence exceeds this chosen threshold. Ct is reported as the cycle number at this point. Therefore, Ct values decrease linearly with increasing input target quantity. This is used as a quantitative measurement of the input target.

7.5 HOW DATA ARE QUANTIFIED

To quantify the results obtained by real-time PCR, two different methods are commonly used: the standard curve method and the comparative threshold method. More recently, a third method has been introduced, named the Pfaffl method.

7.5.1 The Standard Curve Method

In the standard curve method a sample with known concentration is used to make a dilution series, which is used as a standard curve. Samples, which can be used to construct such a dilution series, are purified plasmid dsDNA, *in vitro* transcribed RNA, *in vitro* synthesized ssDNA, or any cDNA sample expressing the target gene. The concentration of these DNA or RNA samples can be measured spectrophotometrically at 260 nm and converted to the number of copies using the molecular weight of the DNA or RNA. For absolute quantification of mRNA expression absolute standards have to be used; for instance, *in vitro* transcribed RNA. Because of the labor-intensive construction of these standard clones, this method is not widely used.

TABLE 7.1 Available real-time thermal cyclers. The most important features of all real-time thermal cyclers that are currently commercially available are listed. The information was retrieved from the suppliers' websites. CCD: charge-coupled device.

Instrument	Excitation	Detection	Sample format	Run time
Applied Biosystems – http://www.appliedbiosystems.com				
StepOne Real-Time PCR System	Light-emitting diode	Photodiode 3 emission filters	48-well	Fast <40 min Normal: 1 h 40 min
StepOnePlus Real-Time PCR System	Light-emitting diode	Photodiode 4 emission filters	96-well	Fast <40 min Normal: 1 h 40 min
7300 Real-Time PCR System	Tungsten halogen lamp 1 excitation filter	CCD camera 4 emission filters	96-well	Normal: 1 h 40 min 9600 emulation: <2 h
7500 Real-Time PCR System 7500 Fast Real-Time PCR System	Tungsten halogen lamp 5 excitation filters	CCD camera 5 emission filters	96-well	Normal: 1 h 40 min 9600 emulation: <2 h Fast <40 min
ABI Prism® 7900HT	Extended life argon-ion laser source	Spectrograph and CCD camera	96-well 384-well	Fast 96 well: 33 min Fast 384 well: 52 min Array or normal <2 h
BioRad – http://www.bio-rad.com				
MyiQ Single-Color Real-time PCR Detection System	Tungsten halogen lamp	CCD camera – 1 filter	96-well	Around 2 h
iCycler iQ5 Real-Time PCR Detection System	Tungsten halogen lamp	CCD camera – 5 filters	96-well	Around 2 h
MiniOpticon Real-Time PCR Detection System	48 blue-green light-emitting diodes	2 photodiodes	48-well	Around 2 h
Chromo4 Real-Time PCR Detection System	4 light-emitting diodes	4 photodiodes	96-well	Around 2 h
CFX96 Real-Time PCR Detection System	6 light-emitting diodes	6 photodiodes	96-well	40 min
Cepheid – http://www.cepheid.com				
Smart Cycler® System	4 optical channels Light-emitting diode	4 photodiodes	16 to 96 module units Up to 96 independent wells	20–40 min
Corbett Life Science – www.corbettlifescience.com				
Rotor-Gene 6000 series	Light-emitting diode	Photomultiplier tube Up to 6-plex	36 200 µl micro tubes 72 100 µl strip-tubes 100 30 µl tubes	40 min

(*Continued*)

TABLE 7.1 (Continued)

Instrument	Excitation	Detection	Sample format	Run time
Eppendorf – http://www.eppendorf.com				
Mastercycler ep realplex2 real-time PCR system	96 light-emitting diodes	1 channel photomultiplier tube	96-well	<2 h
Mastercycler ep realplex2 S real-time PCR system				40 min
Mastercycler ep realplex4 real-time PCR system	96 light-emitting diodes	2 channel photomultiplier tube	96-well	24 min
Mastercycler ep realplex4 S real-time PCR system				40 min
Idaho Technology – http://www.idahotech.com				
R.A.P.I.D.™ System	Light-emitting diode	Three-color optics module	32 composite glass/plastic vessel	≤35 min
R.A.P.I.D.™ LT System				
RAZOR® EX System	Light-emitting diode	One-color optics module	12 tubes	30 min
Roche – http://www.roche-applied-science.com				
LightCycler® 1.5 Instrument	Light-emitting diode	3 channels	32 glass capillaries	45 min
LightCycler® 2.0 Instrument	Light-emitting diode	6 channels	32 glass capillaries	45 min
LightCycler® 480 Real-Time PCR System	High-intensity xenon lamp	CCD camera	96- or 384-well	40 min
	5 excitation filters	6 emission filters		
Stratagene-http://www.stratagene.com				
Mx3000P® Real-Time PCR System	Quartz tungsten halogen source lamp	Single scanning photomultiplier tube	96-well plate	Around 2 h
	4 position excitation filter wheel	4 position detection filter wheel		
Mx3005P® Multiplex Quantitative PCR System	Quartz tungsten halogen source lamp	Single scanning photomultiplier tube	96-well plate	Around 2 h
	5 position excitation filter wheel	5 position detection filter wheel		

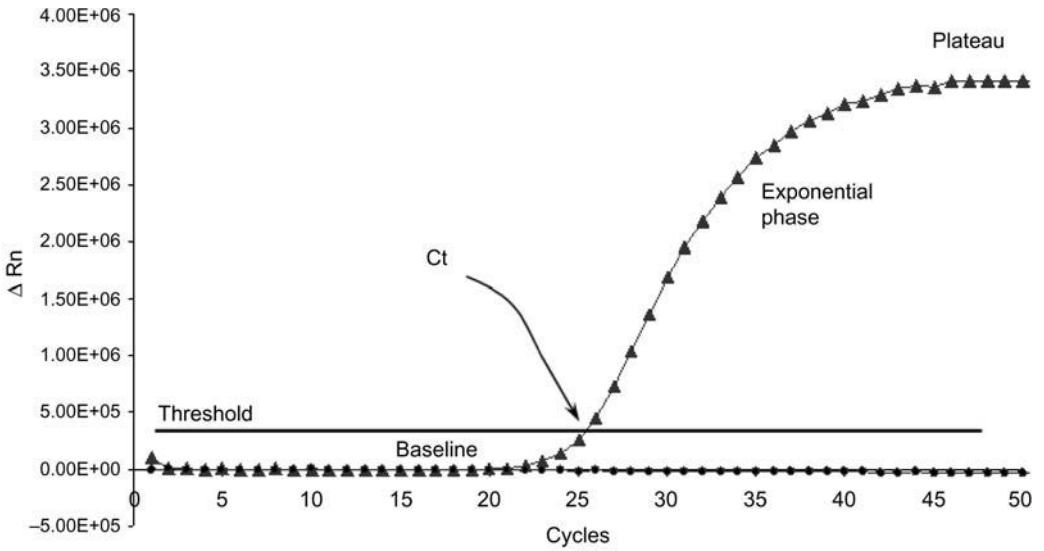


FIGURE 7.7 Amplification plot. Typical amplification plot of a real-time RT-PCR reaction, with a baseline, exponential phase, and plateau phase. Cycle number is plotted against DRn (the difference between the fluorescence detected at a certain point of the reaction and the initial fluorescence, or the fluorescent emission of the baseline). The threshold is chosen based on the variation of the baseline. Ct (threshold cycle) is the point where the detected fluorescence crosses this given threshold.

More often, cDNA plasmid standards are used for quantification. These are constructed by cloning a cDNA fragment into a suitable plasmid vector. For quantification of mRNA expression, however, this will result in only a relative quantification, because variations in efficiency of the reverse transcription step are not controlled.

When creating a standard curve by serial two-fold dilutions of the standard sample, two consecutive points will have a Ct difference of 1. Similarly, Ct values from ten-fold diluted samples will differ by 3.3. This, of course, is assuming that 100% PCR efficiency is reached. Consequently, the slope of the standard curve is a measure of the efficiency of the PCR reaction. For serial ten-fold dilutions, it should ideally be −3.3. In practice, standard curves with a slope between −3.0 and −3.6 are considered acceptable. Also the sensitivity of the PCR reaction is reflected in the standard curve, by the point at which the standard curve crosses the y-axis (y-intercept). Indeed, the lower the Ct value at this point, the higher the sensitivity of the PCR reaction (see Fig. 7.8). By plotting the Ct value of an unknown sample on the standard curve, the amount of input target sequence in the sample can be determined. This calculation is performed automatically by the software program of the real-time PCR instrument.

7.5.2 The Comparative Ct Method

An alternative method used for relative quantification is the comparative Ct method, or the DDCt method (Livak and Schmittgen, 2001). This method uses arithmetic formulas to calculate relative expression levels, compared to a calibrator. A non-treated control sample can be used, for instance, as a calibrator. Moreover, the value of the unknown target is normalized to an endogenous reference gene (e.g. housekeeping gene). The amount of target, relative to the calibrator and normalized to the reference gene, is measured by the equation $2_{-\mathrm{DDCt}}$, where $\mathrm{DDCt} = \mathrm{DCt}_{\mathrm{sample}} - \mathrm{DCt}_{\mathrm{calibrator}}$, and DCt is the Ct of the target gene subtracted by the Ct of the reference gene. The equation thus represents the normalized expression of the target gene in the unknown sample, relative to the normalized expression of the calibrator sample. Importantly, for the DDCt method to be applicable, the efficiency of PCR amplification for the target gene and the housekeeping gene must be approximately equal. For every real-time PCR assay that is being set up, this has to be tested, by determining how the $\mathrm{DCt}_{\mathrm{sample}}$ and $\mathrm{DCt}_{\mathrm{calibrator}}$ varies with template dilution. In case the efficiencies between PCR amplification for target and housekeeping genes are different, a new set of primer/probe combinations has to be designed. Alternatively, the standard curve method or the more recently developed Pfaffl method (described below) can be applied.

7.5.3 The Pfaffl Method

As an alternative to the comparative Ct method, more recently another mathematical model has been developed by M.W. Pfaffl (2001). In this method, the real-time PCR efficiency is also taken into account. Relative expression ratios (R) are calculated based on the efficiency of the real-time PCR assay (E) and the threshold cycle of the unknown sample (Ct) versus a calibrator (e.g. non-treated control

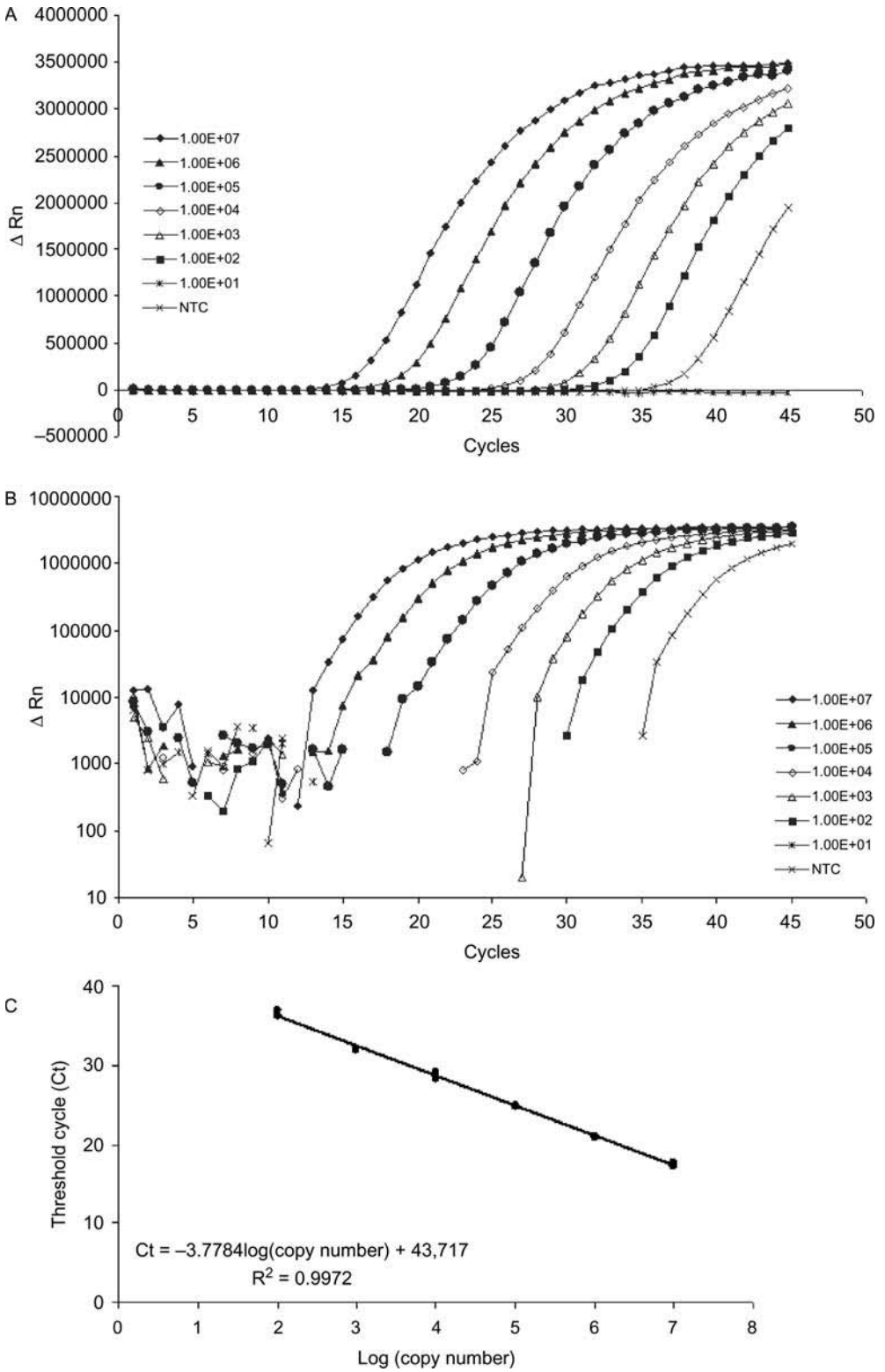


FIGURE 7.8 Standard curve. **A.** Linear amplification plot. Amplification plot of b-actin plasmid cDNA. Seven points of a ten-fold dilution series of a b-actin cDNA plasmid standard, amplified using the TaqMan system. **B.** Logarithmic amplification plot. The same amplification curve as in A, but shown in a logarithmic scale. **C.** Standard curve for b-actin, constructed by plotting threshold cycle (Ct) values against input cDNA copy numbers.

sample), and expressed in comparison to a reference housekeeping gene. This is expressed by the following equation:

$$\text{Ratio} = (1 + E_{\text{target}})^{\Delta Ct(\text{target})}/(1 + E_{\text{ref}})^{\Delta Ct(\text{ref})}$$

where E_{target} = PCR efficiency of the target gene, E_{ref} = PCR efficiency of the reference gene, ΔCt(target) = Ct(sample) − Ct(calibrator) of the target gene, and ΔCt(ref) = Ct(sample) − Ct(calibrator) of the reference gene.

In analogy with the $\Delta\Delta$Ct method, ΔCt is the Ct of the target gene subtracted by the Ct of the housekeeping gene. Because this method takes into account the intrinsic efficiency of each PCR reaction, it can be applied for any real-time PCR amplification assay, with enhanced accuracy.

7.6 MULTIPLEX REAL-TIME PCR

Multiplex real-time PCR can refer either to the simultaneous amplification and detection of different target genes in one tube, or to the use of multiple fluorogenic probes for the discrimination of different alleles. The term is therefore somewhat confusing, since it is not a mere extension of classical multiplexing known in conventional PCR, which was simply the amplification of multiple templates within one reaction, using different primers.

Since diagnostic analyses are often restricted by the limited availability of bioptic material, and one of the primary goals is to decrease analysis time, multiplex PCR is considered an attractive solution. However, it is still not routinely used today.

One difficulty is associated with limitations caused by interference of multiple sets of primers, which can reduce the dynamic range of sensitivity. In practice, a problem will arise when highly differentially expressed targets have to be quantified simultaneously, since the exponential phase of amplification of the most abundant target will not overlap with that of the less abundant target. Thus, in order to be able to simultaneously amplify these different targets, an extensive optimization is required, which most often makes use of primer limiting conditions (for the most abundant target).

Other obstacles arise from the limited number of different fluorophore reporters with a good spectral resolution. Real-time PCR instrumentations contain optimized filters to minimize the overlap of the emission spectra from the different available fluorophores. Despite this, the number of fluorophores that can be combined and clearly distinguished is limited when compared with the resolution in conventional multiplex PCR. Recent improvements in the design of other probe formats as well as novel combinations of fluorophores have aided greatly in the ability to simultaneously detect a larger number of targets. Also recent developments in real-time PCR instruments have significantly improved, allowing multiplexing of up to six different colors (see Table 7.1).

Another approach, which is used for the detection of human genetic diseases, is based on the discrimination between single or multiple nucleotide changes (between different alleles) by making use of the differences in melting point. In this approach, a single fluoroprobe can be used to distinguish between products, based on their distinct melting temperatures, which is reflected by the differences in thermodynamic stabilities of the perfectly complementary and the mismatched probe-target duplexes.

Finally, the combined use of multicolor fluorimetry and fluorescence melting curve analysis can greatly increase the number of targets that can be detected simultaneously (Wittwer *et al.*, 2001). A nice example, applying this sophisticated technique, is the simultaneous detection of all 27 possible base substitutions occurring in codons 12, 13, and 61 from the wild-type sequence of the *Nras* oncogene (Elenitoba-Johnson *et al.*, 2001).

7.7 APPLICATIONS IN MOLECULAR DIAGNOSTICS

7.7.1 Clinical Microbiology

Real-time PCR has been shown to be extremely useful for studies in the field of clinical virology, bacteriology, and fungal diagnosis (see also section 30.2.3). Most of the assays developed allow an increased frequency as well as enhanced speed of pathogen detection as compared to conventional culture techniques. Moreover, quantification of pathogen load is made possible. For real-time assay development one has to take into account that most infectious agents are characterized by a high mutation rate, which can dramatically influence the pathogen load estimation. This can be overcome by designing primers in highly conserved regions. Alternatively, of course, sequence variations can provide the basis for development of subtype-specific assays.

7.7.1.1 Virology and Bacteriology

Real-time PCR is increasingly used for the diagnosis of different diseases caused by DNA viruses, such as cytomegalovirus (Tanaka and Kimura, 2000), Epstein Barr virus (Ohga and Kubo, 2001), and hepatitis B virus (Brechtuehl and Whalley, 2001). It is a very reliable technique to titer viral genome density in clinical specimens, to monitor patients with virus infections, or to monitor reactivation in patients with latent infections. Indeed, viral DNA copy numbers change proportionally with anti-viral therapy (Nitsche *et al.*, 1999). It should thus be clear that the application of real-time PCR in this field resulted in a great improvement in following the impact of the therapy. Other applications in the field of virology for which the real-time PCR technology has been validated are the detection

of Bacillus anthrax (Makino and Cheun, 2003), human papillomavirus (Hubbard, 2003), influenza virus (Ellis and Zambon, 2002), and parpovirus (Bultmann *et al.*, 2003), to name just a few.

Similarly, real-time PCR methods are used increasingly for the detection of bacterial pathogens. Validated examples are, for instance, the detection and quantification of different *Mycobacterium* species (Miller *et al.*, 2002; Rondini *et al.*, 2003), *Helicobacter pylori* (Lascols *et al.*, 2003), and *Streptococcus pneumonia* (van Haeften *et al.*, 2003).

7.7.1.2 Fungal Diagnosis

In the detection of fungal infections it has become essential to develop methods that allow for a quick, sensitive, and specific detection. Indeed, fungal infections are a major cause of morbidity and mortality in immuno-compromised patients (for instance, after aggressive chemotherapy, organ transplantation, or in immuno-deficient patients). It is evident that early initiation of antifungal therapy, which critically depends on the early, fast, and accurate detection method, is essential in reducing the high mortality rates associated with fungaemia. The gold standard for detection of fungal infections has long been blood culture. But this technique is very time consuming and lacks sensitivity. Therefore the validation of real-time PCR in the detection of fungal infections essentially overcomes these limitations and is a promising tool to become the new gold standard.

Real-time PCR assays have been developed using broad-range fungal primers: for example, the fungal 18S rRNA gene. These amplify fungi-specific and highly conserved sequences of multicopy genes and are clinically useful for the detection of fungal infections. Additionally, PCR assays targeting species-specific sequences can be used to identify the pathogen. Alternatively, specific identification of the pathogen can be obtained by subsequent sequencing of the amplicon (Imhof *et al.*, 2003).

The development of these real-time PCR assays allows species determination of fungal DNA not only in human blood samples, but also in human tissue biopsies. This is a great step forward, since in this way, deep-seated fungal infections can be analyzed, which was not possible with the previous culture-based techniques (Imhof *et al.*, 2003). Pathogenic fungi, for which detection protocols by real-time PCR have been validated, for example, are *Candida*, *Aspergillus*, *Histoplasma*, and *Conidiobolus* species (Imhof *et al.*, 2003; White *et al.*, 2003).

7.7.2 Food Microbiology

Rapid and accurate detection of bacterial pathogens from food samples is important, both for food quality assurance and to trace outbreaks of bacterial pathogens within the food supply. Real-time PCR analysis meets these criteria, since it significantly shortens analysis time compared to conventional biochemical and serological identification methods, and because the technique can be applied directly on preenrichment media or food products. A specific problem when analyzing food pathogens is that foods are often complex matrices, which require selective enrichment steps to overcome problems of low pathogen numbers. To address this issue, a universal enrichment broth has been developed, enabling enrichment of multiple pathogens (Bailey and Cox, 1992). An additional problem that has to be faced in the case of PCR-based detection methods for food pathogens is to discriminate between life and death pathogenic cells. One approach used is to apply a living/dead dye that covalently binds to DNA and inhibits PCR amplification from dead cells (Rudi *et al.*, 2002). Alternatively, the development of mRNA-based real-time PCR assays as opposed to DNA-based assays can be applied, which has the advantage of serving as an accurate indicator of pathogen viability (Rijpens and Herman, 2002).

Protocols have been validated for the detection of *Escherichia coli*, *Salmonella* (Bhagwat *et al*., 2003), and *Listeria* strains (Norton, 2002) in food and environmental samples, to name just a few examples.

7.7.3 Clinical Oncology

7.7.3.1 Minimal Residual Disease

The detection of minimal residual disease (MRD) – that is, the detection of a very low number of malignant cells – significantly correlates with the clinical outcome in many hematological malignancies. Therefore, MRD monitoring is important for therapy guidance in clinical settings. Real-time PCR-based techniques for the detection of MRD have been developed, and already are used in clinical protocols. The major challenge when optimizing such assays is to reach a maximal sensitivity. Real-time PCR assays for MRD detection can be divided into three main categories, depending on the specific type of target envisioned:

1. Immunoglobulin and T-cell receptor gene rearrangements.
2. Breakpoint fusion regions of chromosome aberrations and fusion-gene transcripts.
3. Aberrant genes or aberrantly expressed genes.

The choice for a specific application is dependent on the type of disease. In hematological malignancies, real-time PCR assays have been optimized; for instance, for acute lymphoblastic leukemia (Pongers-Willemse *et al.*, 1998; Donovan *et al.*, 2000), acute myeloid leukemia (Jaeger and Kainz, 2003), chronic myeloid leukemia (Gonzalez *et al.*, 2003), multiple myeloma (Gerard *et al.*, 1998; Rasmussen *et al.*, 2000), and non-Hodgkin's lymphoma (Yashima *et al.*, 2003).

Importantly, in this field efforts are being made to meet standardized criteria and an international uniformity. Several European networks have been established, which have developed common universal guidelines for data analysis and reporting MRD data. Furthermore, these networks also perform quality control rounds, which are required to monitor the performance of the participating laboratories and to further improve and standardize real-time PCR assays (van der Velden *et al.*, 2003, 2007).

7.7.3.2 Single Nucleotide Polymorphisms

Single nucleotide polymorphisms (SNPs), present in the human genome, are used as genetic markers to follow the inheritance patterns of chromosomal regions from generation to generation. Moreover, they consist of a powerful tool in the study of genetic factors associated with human diseases (Johnson and Todd, 2000; Risch, 2000). In a clinical setting, detection of specific SNPs can be very useful for determining the relative percentages of donor and recipient cells after allogeneic bone marrow transplantation (Oliver *et al.*, 2000).

Real-time PCR assays have been developed for SNP detection. For this application the level of discrimination between target and non-target allele is the most important challenge when optimizing the technique. In this real-time PCR method, one set of primers and two allele-specific fluorescent-labeled probes are used, making use of different reporter dyes. The two alleles can thus be distinguished by the differential fluorescent emission of the two different reporter dyes (Livak, 1999). Alternatively, SNP detection is often making use of molecular beacons (Tyagi *et al.*, 1998) or scorpion probes (Whitcombe *et al.*, 1999). For each individual SNP assay, the level of discrimination has to be tested experimentally, since it is dependent on the mismatch.

7.7.3.3 Chromosomal Translocations

Chromosomal translocations, which are often taking place in tumor cells, can be employed as tumor-specific PCR targets. The primers are designed so that they anneal to opposite sides of the breakpoint in the fusion gene. These fusion genes are interesting targets for the design of a PCR method. Indeed, they are directly related to the oncogenic process and therefore stable throughout the disease course. Moreover, the breakpoint fusion sites at the DNA level differ in each patient so that patient-specific real-time PCR strategies can be applied. A disadvantage of patient-specific assays is, of course, the higher analysis costs. A real-time PCR assay has been developed for the t(14,18) translocation, involving the *BCL2* and *IGH* genes (Summers *et al.*, 2001). For the design of real-time PCR assays in this setting a major difficulty is the relatively large amplicon size often necessary to span the fusion breakpoints, especially if one wants to design an assay for different patients clustering in relatively small breakpoint areas.

Alternatively, real-time PCR can be performed on the transcripts of tumor-specific fusion genes. These are disease-specific transcripts located over chromosome breakpoints, leading to an in-frame RNA product (Rowley, 1998). Interestingly, these fusion transcripts can be identical in individual patients despite distinct breakpoints, because the breakpoints often are located in introns. An advantage of this approach is that the same set of primer/probes can be used for analysis of individual patients with the same fusion transcript, but different breakpoint translocations (van der Velden *et al.*, 2003). Examples of validated real-time PCRs using this approach are the BCR/ABL (Jones *et al.*, 2003), MLL/AF9 (Scholl *et al.*, 2003), CBFbeta/MYH11 fusion gene transcripts (Marcucci *et al.*, 2001), or the simultaneous detection of ten different fusion gene transcripts (Osumi *et al.*, 2002).

7.7.4 Gene Therapy

The primary goal of gene therapy is to specifically deliver the therapeutic gene to the target organ in a time- and dose-dependent manner and, most importantly, to avoid delivery to non-target organs, since this may result in toxic side effects. In this case, the therapeutic gene is delivered to the target cells through a vector system, most often a viral vector (Crystal, 1995). Two important parameters have to be analyzed when considering gene therapy as a drug delivery system:

1. Gene transfer estimation, which is the expression level of the therapeutic gene in regard to target tissue levels over time.
2. The bio-distribution, which is the distribution of the drug in different organs for different routes of administration.

Real-time PCR assays have been validated both in regard to gene transfer as well as bio-distribution of gene therapy vectors. The major challenge in optimizing these assays is the accuracy of quantification and the sensitivity of the assay. Real-time PCR assays using hybridization probes, for instance, have been validated for adenovirus gene transfer vectors and proven to be quantitative, reproducible, and sensitive (Senoo *et al.*, 2000; Hackett *et al.*, 2000).

7.7.5 Quantification of Gene Expression

Probably one of the most widely used applications of real-time PCR is the quantification of mRNA expression, or real-time reverse-transcriptase PCR (RT-PCR). Major challenges when optimizing real-time PCR assays for gene expression studies are the sensitivity of the assay, accurate

quantification, avoiding amplification of contaminating genomic DNA, and the choice of a relevant housekeeping gene as control. The use of real-time PCR in molecular diagnostics has been reported, for instance, for the analysis of tissue-specific gene expression (Bustin, 2000) and for analyzing cytokine mRNA expression profiles (Giulietti *et al.*, 2001). Currently, real-time RT-PCR assays are being combined with other sophisticated technologies, such as microarray analysis and laser capture micro-dissection. The availability of these high-throughput technologies enable the analysis of gene expression to measure alterations in molecular genetic events associated with the initiation and progression of a variety of diseases (Elkahloun *et al.*, 2002).

In the field of immunology, quantification of gene expression by real-time RT-PCR is widely used for the analysis of cytokines and chemokines. Cytokines are proteins, which are secreted by many different cell types, such as lymphocytes, monocytes, macrophages, dendritic cells, endothelial cells, and fibroblasts. They play a central role in modulating immune responses, including lymphocyte activation, proliferation, differentiation, survival, and apoptosis. Consequently, abnormal cytokine expression patterns will contribute to immune-mediated disorders, such as autoimmune, allergic, and infectious diseases (Borish and Steinke, 2003; Rabinovitch, 2003). In addition, they are commonly the focus of studies on immunosuppressive therapy during organ transplantation.

Chemokines, also known as chemotaxic cytokines, are secreted proteins that function as chemo-attractants, controlling the trafficking of specific subsets of leukocytes to sites of tissue damage to mediate specific immunological functions. In autoimmune diseases, the migration and accumulation of leukocytes in diseased target organs is a critical step in the development of the disease (Kunkel and Godessart, 2002).

Therefore, a reliable quantification of cytokine, cytokine receptor, chemokine, and chemokine receptor mRNA levels in diseased target organs are fundamental for our understanding of diverse pathological states and for monitoring therapeutic effects. Major challenges when optimizing real-time PCR assays for gene expression studies are the sensitivity of the assay, accurate quantification, and avoiding amplification of contaminating genomic DNA. The use of real-time RT-PCR has been validated for a large panel of cytokines and chemokines (Giulietti *et al.*, 2001; Overbergh *et al.*, 2003b, 2006). This greatly refined the role of cytokines and chemokines in numerous biological and clinical contexts; for instance, in type 1 diabetes (Gysemans *et al.*, 2003; Overbergh *et al.*, 2003a), in type 2 diabetes (Guilietti *et al.*, 2007) and experimental autoimmune encephalitis (van Etten *et al.*, 2003). According to the numerous publications using real-time PCR technique for quantification of cytokine and chemokine expression, it has definitely become the gold standard in this field.

7.8 CRITERIA FOR DEVELOPING REAL-TIME PCR ASSAYS

The choice of method for performing real-time PCR reactions depends on the specific application envisioned. An important distinction has to be made, for instance, depending on whether one is dealing with measuring mRNA expression levels or with DNA copy number. Other factors, such as the number of genes, the number of targets, the importance of fast screening, allelic discrimination, accurate quantification, sensitivity, and the costs, need careful consideration.

Criteria to be taken into account when optimizing real-time PCR assays for mRNA quantification, such as primer design, template preparation, and normalization of the results, are described next. These are based mainly on our own experience in the field of quantification of mRNA expression levels of cytokines, chemokines, or other immune related targets, in the context of gaining more insight into the mechanisms of disease outcome in type 1 diabetes (Overbergh *et al.*, 2000; Gysemans *et al.*, 2000; Giulietti *et al.*, 2001).

7.8.1 Primer and Probe Design

For the design of primers and probes, different software programs are available. The most commonly used are Primer Express (Applied Biosystems) and the LightCycler Probe Design (Roche), although several useful websites assisting in the design of primers and probes have also appeared recently. A widely used freely available software package for primer design is Primer 3 (http://frodo.wi.mit.edu/). As in classical PCR reactions, the primers should be free of direct and inverted repeats, as well as of homopolymeric runs. Furthermore, primers should not be able to form inter- and intramolecular dimers. Special attention should be given to the 3′ sequence: this sequence should not form dimers or hairpins and the binding should not be too strong in order to prevent non-specific extension. Particularly for real-time PCR detection, the amplicon length should be very short, with the common rule being ‘‘the shorter the better’’ (ideally 50–150 base pairs). For the design of probes (in our case TaqMan probes), a few important specifications have to be taken into account. Probes should have a melting temperature about 10°C higher than the primers in order to anneal to the target sequence during the extension phase of the PCR reaction (which is usually performed at 60°C). Furthermore, probes should not contain a guanosine at their 5′ end, and should have more cytidines than guanosines.

When designing primers/probes for the quantification of target mRNA expression, special care has to be taken to avoid coamplification of contaminating genomic DNA. An efficient way to do so is by designing primer/probe

combinations on exon–intron boundaries. Sometimes this strategy is not possible, for example, in the case of an intron-less gene or in case the genomic DNA sequence is not (yet) available. In these cases one is obliged to perform a DNase treatment on the RNA samples. Moreover, contradictory to what the suppliers claim, DNase treatment will not always result in complete removal of DNA (Bustin, 2002).

Since primer/probe design and experimental validation is time consuming, a free, publicly available RTPrimerDB, a real-time PCR primer, and probe database, which provides validated real-time PCR primer and probe sequences, has been developed. At present, it contains useful user-based validated primer/probe sequences for human, mouse, rat, fruit fly, zebrafish, and many other species, using all currently available detection chemistries. These real-time PCR primer/probe records are available at http://medgen.ugent.be/rtprimerdb/ (Pattyn *et al.*, 2003, 2006).

7.8.2 Gene Expression Normalization

A reliable quantitative real-time RT-PCR method requires correction for experimental variations in individual reverse transcriptase and PCR efficiencies. Indeed, differences in efficiency of the reverse transcriptase reaction will result in an amount of cDNA that does not correspond to the starting amount of RNA. Furthermore, because of the exponential nature of the PCR reaction, minor differences in PCR amplification efficiency will result in major differences in the PCR product.

Currently, the most widely applied method to correct for these variations is normalization to a housekeeping gene. In theory, an ideal housekeeping gene should be expressed at a constant level among different tissues of an organism, at all stages of development, and should not be affected by the experimental treatment itself. In practice, however, finding a gene with these characteristics is an almost impossible task. Indeed, the expression of a housekeeping gene can also be regulated by the experimental treatment or can be tissue dependent. In this regard, more and more evidence is accumulating suggesting that all genes are regulated at least under some conditions. Therefore, it has become a fact that a universal reference gene with a constant expression in all tissues simply does not exist. It is therefore important that prior to performing each specific set of experiments, a suitable housekeeping gene needs to be validated. The most commonly used housekeeping genes are b-actin, glyceraldehyde-3-phosphate-dehydrogenase, rRNA, and hypoxanthine guanine phosphoribosyl transferase. Other genes, such as cyclophilin, mitochondrial ATP synthase 6, and porphobilinogen deaminase, also are used. Recently, different panels of reference genes became commercially available for testing (http://www.tataa.com/referencepanels.htm) and specific software programs have been developed in the search for an optimal reference gene for a specific experimental setup (Vandesompele *et al.*, 2002; Pfaffl *et al.*, 2004). Using these systems, one has to keep in mind that reference genes with similar functional properties may have a co-regulated expression. It may therefore be advisable to choose reference genes from distinct functional classes (http://www.tataa.com/referencepanels.htm).

7.8.3 Real-Time PCR and Laser Capture Microdissection

The method of normalization is rather straightforward when the RNA samples originate from pure cell populations or cell-lines. However, in case RNA extractions and real-time PCR have to be carried out on whole tissue or tissue biopsies, an additional problem arises when normalizing the results to a housekeeping gene. In the field of immunology, for instance, some experimental treatments can change the size of an organ, for example by inducing splenomegaly, rendering the tissue unsuitable for comparison to a normal control tissue. Similarly, in transplanted or auto-immune attacked organs, the diseased organ will have significantly larger immune cell infiltrates compared to a normal organ. Other treatments, such as irradiation, can change the expression level of several housekeeping genes, again making normalization very difficult. Also, in cancer research the presence of normal cells in the immediately surrounding of cancer cells is a major difficulty when normalizing to a housekeeping gene. Furthermore, starting from whole tissues or tissue biopsies in these cases will result in the averaging of the expression of different cell types, and the expression profile of a specific cell type will be masked because of the bulk of surrounding cells. In addition, no information will be obtained as to which cell type is the actual producer of the measured target RNA.

An elegant approach to circumvent this problem of differences in the cell populations itself, which has received increased attention during the last couple of years, involves the combination of real-time PCR with laser capture microdissection (LCM). The technique of LCM was described for the first time in 1996 (Emmert-Buck *et al.*, 1996). It is a simple, reliable, and rapid technique, which allows the extraction of pure subpopulations of cells from heterogeneous tissue samples, with retention of cell morphology. This is performed by directing a brief laser pulse at cells within a tissue section placed on a glass slide. In this way, individual cells or groups of cells can be selected, without damaging the cells. Therefore, gene expression patterns, both at the RNA and at the protein level, are maintained, and samples are suitable for molecular analysis. The combination of LCM and real-time PCR is increasingly being used, especially in the genetic analysis of cancer cells (Elkahloun *et al.*, 2002; de Preter *et al.*, 2003). This technique has also been validated in other areas, for instance for the analysis

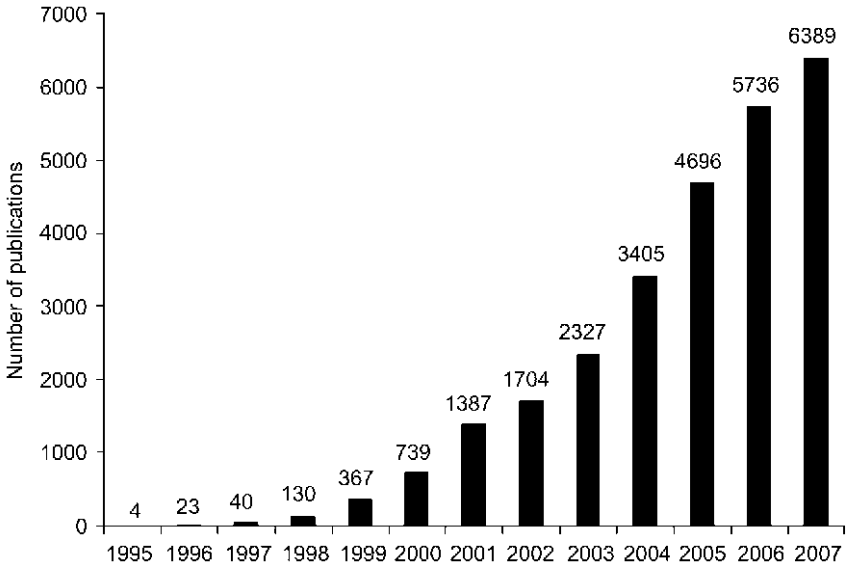


FIGURE 7.9 Real-time PCR publications. The keywords "real-time" and "PCR" were typed in the PubMed search engine (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi), the results were crossed with the years between 1996 (year of the first publication in real-time PCR) and 2007. The number of citations by year is shown. It is clear that the number of publications using the real-time PCR technique is increasing exponentially.

of gene expression upon *Mycobacterium tuberculosis* infection (Zhu *et al.*, 2003) or in atherosclerotic lesions (Trogan *et al.*, 2002).

7.9 CONCLUSIONS

The introduction of the real-time PCR technology has revolutionarily simplified the quantification of DNA and RNA. This has had a great impact in the field of molecular research and diagnostics, since enormous amounts of data can be obtained within a very short research time. The decreased costs for the thermal cyclers as well as for the necessary reagents for applying the technique have aided in its rapid increase in use. Indeed, it is clear from a Medline search for the keyword "real-time PCR" that its use increased exponentially over the last decade (see Fig. 7.9). Therefore, it seems that this technique is becoming the gold standard for the detection and quantification of DNA and RNA in the general research or diagnostic laboratory.

Although real-time PCR assays by themselves are characterized by high precision and reproducibility, the accuracy of the data obtained are largely dependent on several other factors. In order to design and analyze experiments using real-time PCR it is not sufficient to simply extend one's knowledge of classic end-point PCR. Indeed, many other controls are needed to be certain of the accuracy of the results when using real-time PCR assays. These factors, such as sample preparation, quality of the standard, choice of a housekeeping gene, and normalization of samples, need careful consideration and optimization. Furthermore, it is important that standardized criteria and international uniformity in experimental design and data analysis is reached, in order to be able to compare data between different laboratories.

New developments, such as the combination of real-time PCR with other sophisticated techniques such as LCM, make it possible to measure gene expression or DNA copy numbers in specific cell types from *in vivo* samples, which were previously very difficult to analyze. These make the possible applications of real-time PCR even more attractive for the future.

REFERENCES

Ahmad, A.I. (2007). BOXTO as a real-time thermal cycling reporter dye. *J. Biosci.* 32, 229–239.

Bailey, J.S., and Cox, N.A. (1992). Universal pre-enrichment broth for the simultaneous detection of Salmonella and Listeria in foods. *J. Food Prot.* 55, 256–259.

Bernard, P.S., and Wittwer, C.T. (2000). Homogeneous amplification and variant detection by fluorescent hybridization probes. *Clin. Chem.* 46, 147–148.

Bhagwat, A.A. (2003). Simultaneous detection of *Escherichia coli* O157:H7, *Listeria monocytogenes* and *Salmonella* strains by real-time PCR. *Int. J. Food Microbiol.* 84, 217–224.

Borish, L.C., and Steinke, J.W. (2003). Cytokines and chemokines. *J. Allergy Clin. Immunol.* 111, S460–S475.

Brechtuehl, K., and Whalley, S. (2001). A rapid real-time quantitative polymerase chain reaction for hepatitis B virus. *J. Virol. Methods* 93, 105–113.

Bultmann, B.D., Klingel, K., Sotlar, K., Bock, C.T., and Kandolf, R. (2003). Parvovirus B19: a pathogen responsible for more than heatologic disorders. *Virchows Arch.* 442, 8–17.

Bustin, S.A. (2000). Absolute quantification of mRNA using realtime reverse transcriptase polymerase chain reaction assays. *J. Mol. Endocrinol.* 25, 169–193.

Bustin, S.A. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol.* 29, 23–29.

Cardullo, R.A., Agrawal, S., Flores, C., Zamecnick, P.C., and Wolf, D.E. (1988). Detection of nucleic acid hybridization by nonradioactive fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. USA* 85, 8790–8794.

Clementi, M., Bagnarelli, P., Manzin, A., and Menzo, S. (1994). Competitive polymerase chain reaction and analysis of viral activity at the molecular level. *Genet. Anal. Tech. Appl.* 11, 1–6.

Crystal, R.G. (1995). Transfer of genes to humans: early lessons and obstacles to success. *Science* 270, 404–410.

De Preter, K., Vandesompele, J., Heimann, P., Kockx, M.M., Van Gele, M., Hoebeeck, J., De Smet, E., Demarche, M., Laueys, G., Van Roy, N., De Paepe, A., and Speleman, F. (2003). Application of laser capture microdissection in genetic analysis of neuroblastoma and neuroblastoma precursor cells. *Cancer Lett.* 197, 53–61.

Donovan, J.W., Ladetto, M., Zou, G., Neuberg, D., Poor, C., Bowers, D., and Gribben, J.G. (2000). Immunoglobulin heavychain consensus probes for real-time PCR quantification of residual disease in acute lymphoblastic leukemia. *Blood* 15, 2651–2658.

Elenitoba-Johnson, K.S.J., Bohling, S.D., Wittwer, C.T., and King, T.C. (2001). Multiplex PCR by multicolor fluorimetry and fluorescence melting curve analysis. *Nat. Med.* 7, 249–253.

Elkahloun, A.G., Gaudet, J., Robinson, G.S., and Sgroi, D.C. (2002). In situ gene expression analysis of cancer using laser capture microdissection, microarrays and real time quantitative PCR. *Cancer Biol. Ther.* 1, 354–358.

Ellis, J.S., and Zambon, M.C. (2002). Molecular diagnosis of influenza. *Rev. Med. Virol.* 12, 375–389.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A., and Liotta, L.A. (1996). Laser capture microdissection. *Science* 274, 998–1001.

Gerard, C.J., Olsson, K., Ramanathan, R., Reading, C., and Hanania, E.G. (1998). Improved quantitation of minimal residual disease in multiple myeloma using real-time polymerase chain reaction and plasmid-DNA complementarity determining region III standards. *Cancer Res.* 58, 3957–3964.

Gibson, U.E.M., Heid, C.A., and Williams, P.M. (1996). A novel method for Real-Time Quantitative RT-PCR. *Genome Res.* 6, 995–1001.

Giulietti, A., Overbergh, L., Valckx, D., Decallonne, B., Bouillon, R., and Mathieu, C. (2001). An overview of Real-Time Quantitative PCR: applications to quantify cytokine gene expression. *Methods* 25, 386–401.

Giulietti, A., van Etten, E., Overbergh, L., Stoffels, K., Bouillon, R., and Mathieu, C. (2007). Monocytes from type 2 diabetic patients have a pro-inflammatory profile. 1,25-Dihydroxyvitamin D(3) works as anti-inflammatory. *Diabetes Res. Clin. Pract.* 77, 47–57.

Gonzalez, D., Gonzalez, M., Alonso, M.E., Lopez-Perez, R., Balanzategui, A., Chillon, M.C., Silva, M., Garcia-Sanz, R., and San Miguel, J.F. (2003). Incomplete DJH rearrangements as a novel tumor target for minimal residual disease quantitation in multiple myeloma using real-time PCR. *Leukemia* 17, 1051–1057.

Guyer, R.L., and Koshland, D.E., Jr. (1989). The Molecule of the Year. *Science* 246, 1543–1546.

Gysemans, C.A., Waer, M., Valckx, D., Laureys, J.M., Mihkalsky, D., Bouillon, R., and Mathieu, C. (2000). Early graft failure of xenogeneic islets in NOD mice is accompanied by high levels of interleukin-1 and low levels of transforming growth factor beta mRNA in the grafts. *Diabetes* 49, 1992–1997.

Gysemans, C.A., Stoffels, K., Giulietti, A., Overbergh, L., Waer, M., Lannoo, M., Feige, U., and Mathieu, C. (2003). Prevention of primary non-function of islet xenografts in autoimmune diabetic NOD mice by anti-inflammatory agents. *Diabetologia* 46, 1115–1123.

Hackett, N.R., El Sawy, T., Lee, L.Y., Silva, I., O'Leary, J., Rosengart, T.K., and Crystal, R.G. (2000). Use of Quantitative TaqMan Real-Time PCR to track the time-dependent distribution of gene transfer vectors in vivo. *Mol. Ther.* 2, 649–656.

Heid, C.A., Stevens, J., Livak, K.J., and Williams, P.M. (1996). Real-Time Quantitative PCR. *Genome Res.* 6, 986–994.

Higuchi, R., Fockler, C., Dollinger, G., and Watson, R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology* 11, 1026–1030.

Holland, P.M., Abramson, R.D., Watson, R., and Geldfand, D.H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5′ to 3′ exonuclease activity of *Thermus aquaticus*. *Proc. Natl. Acad. Sci. USA* 88, 7276–7280.

Hubbard, R.A. (2003). Human papillomavirus testing methods. *Arch. Pathol. Lab. Med.* 127, 940–945.

Imhof, A., Schaer, C., Schoedon, G., Schaer, D.J., Walter, R.B., Schaffner, A., and Schneemann, M. (2003). Rapid detection of pathogenic fungi from clinical specimens using LightCycler real-time fluorescence PCR. *Eur. J. Clin. Microbiol. Infect. Dis.* 22, 558–560.

Jaeger, U., and Kainz, B. (2003). Monitoring minimal residual disease in AML: the right time for real-time. *Ann. Hematol.* 82, 139–147.

Johnson, G.C., and Todd, J.A. (2000). Strategies in complex disease mapping. *Curr. Opin. Genet. Dev.* 10, 330–334.

Johnson, M.P., Haupt, L.M., and Griffiths, L.R. (2004). Locked nucleic acid (LNA) single nucleotide polymorphism (SNP) genotype analysis and validation using real-time PCR. *Nucleic Acids Res.* 32, e55.

Jones, C.D., Yeung, C., and Zehnder, J.L. (2003). Comprehensive validation of a real-time quantitative bcr-abl assay for clinical laboratory use. *Am. J. Clin. Pathol.* 120, 42–48.

Kingsbury, D.T., and Falkow, S.Z. eds (1985). *Rapid Detection and Identification of Infectious Agents*, pp. 245–256. Academic Press, New York.

Kunkel, S.L., and Godessart, N. (2002). Chemokines in autoimmunity: from pathology to therapeutics. *Autoimmun. Rev.* 1, 313–320.

Kutyavin, I.V., Afonina, I.A., Mills, A., Gorn, V.V., Lukhtanov, E.A., Belousov, E.S., Singer, M.J., Walburger, D.K., Lokhov, S.G., Gall, A.A., Dempcy, R., Reed, M.W., Meyer, R.B., and Hedgpeth, J. (2000). 3′-Minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res.* 28, 655–661.

Lascols, C., Lamarque, D., Costa, J.M., Copie-Bergman, C., Le Glaunec, J.M., Deforges, L., Soussy, C.J., Petit, J.C., Delchier, J.C., and Tankovic, J. (2003). Fast and accurate quantitative detection of Helicobacter pylori and identification of Clarithromycin resistance mutations in H. pylori isolates from gastric biopsy specimens by real-time PCR. *J. Clin. Microbiol.* 41, 4573–4577.

Livak, K.J. (1999). Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. *Genet. Anal.* 14, 143–149.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2_{-\mathrm{DDCt}}$ method. *Methods* 25, 402–408.

Makino, S., and Cheun, H.I. (2003). Application of the real-time PCR for the detection of airborne microbial pathogens in reference to the anthrax spores. *J. Microbiol. Methods* 53, 141–147.

Marcucci, G., Caligiuri, M.A., Dohner, H., Archer, K.J., Schlenk, R.F., Dohner, K., Maghraby, E.A., and Bloomfield, C.D. (2001). Quantification of CBFbeta/MYH11 fusion transcript by real-time

RT-PCR in patients with INV(16) acute myeloid leukemia. *Leukemia* 15, 1072–1080.

Martin, B., Jonas, K., Gunnar, W., and Mikael, K. (2003). A new minor groove binding asymmetric cyanine reporter dye for real-time PCR. *Nucleic Acids Res.* 31, e45.

Miller, N., Cleary, T., Kraus, G., Young, A.K., Spruill, G., and Hnatyszyn, H.J. (2002). Rapid and specific detection of Mycobacterium tuberculosis from acid-fast bacillus smearpositive respiratory specimens and BacT/ALERT MP culture bottles by using fluorogenic probes and real-time PCR. *J. Clin. Microbiol.* 40, 4143–4147.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51, 263–273.

Nitsche, A., Steuer, N., Schmidt, C.A., Landt, O., and Siegert, W. (1999). Different real-time PCR formats compared for the quantitative detection of human cytomegalovirus DNA. *Clin. Chem.* 45, 1932–1937.

Norton, D.M. (2002). Polymerase chain reaction-based methods for detection of Listeria monocytogenes: toward real-time screening for food and environmental samples. *J. AOAC. Int.* 85, 505–515.

Ohga, S., and Kubo, E.N.A. (2001). Quantitative monitoring of circulating Epstein-Barr virus DNA for predicting the development of post-transplantation lymphoproliferative disease. *Int. J. Hematol.* 73, 323–326.

Oliver, D.H., Thompson, R.E., Griffin, C.A., and Eshleman, J.R. (2000). Use of single nucleotide polymorphisms (SNP) and real-time polymerase chain reaction for bone marrow engraftment analysis. *J. Mol. Diagn.* 2, 202–208.

Osumi, K., Fukui, T., Kiyoi, H., Kasai, M., Kodera, Y., Kudo, K., Kato, K., Matsuyama, T., Nait, K., Tanimoto, M., Hirai, H., Saito, H., Ohno, R., and Naoe, T. (2002). Rapid screening of leukemia fusion transcripts in acute leukemia by real-time PCR. *Leuk. Lymphoma* 43, 2291–2299.

Overbergh, L., Decallonne, B., Waer, M., Rutgeerts, O., Valckx, D., Casteels, K.M., Laureys, J., Bouillon, R., and Mathieu, C. (2000). 1alpha,25-dihydroxyvitamin D3 induces an autoantigen-specific T-helper1/T-helper2 immune shift in NOD mice immunized with GAD65 (p524–543). *Diabetes* 49, 1301–1307.

Overbergh, L., Decallonne, B., Branisteanu, D.D., Valckx, D., Kasran, A., Bouillon, R., and Mathieu, C. (2003a). Acute shock induced by antigen vaccination in NOD mice. *Diabetes* 52, 335–341.

Overbergh, L., Giulietti, A., Valckx, D., Decallonne, B., and Mathieu, C. (2003b). The use of real-time reverse transcriptase PCR for the quantification of cytokine gene expression. *J. Biomol. Tech.* 14, 33–43.

Overbergh, L., Gysemans, C., and Mathieu, C. (2006). Quantification of chemokines by real-time reverse transcriptase PCR: applications in type 1 diabetes. *Expert Rev. Mol. Diagn.* 6, 51–64.

Pattyn, F., Speleman, F., De Paepe, A., and Vandesompele, J. (2003). RTPrimerDB: the real-time PCR primer and probe database. *Nucleic Acids Res.* 31, 122–123.

Pattyn, F., Robbrecht, P., De Paepe, A., Speleman, F., and Vandesompele, J. (2006). RTPrimerDB: the real-time PCR primer and probe database, major update 2006. *Nucleic Acids Res.* 34, D684–D688.

Pfaffl, M.W. (2001). A new mathematical method for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29, e45.

Pfaffl, M.W., Tichopad, A., Prgomet, C., and Neuvians, T.P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* 26, 509–515.

Pongers-Willemse, M.J., Verhagen, O.J., Tibbe, G.J., Wijkhuijs, A.J., de Haas, V., Roovers, E., van der Schoot, C.E., and van Dongen, J.J. (1998). Real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia using junctional region specific TaqMan probes. *Leukemia* 12, 2006–2014.

Rabinovitch, A. (2003). Immunoregulation by cytokines in autoimmune diabetes. *Adv. Exp. Med. Biol.* 520, 159–193.

Rasmussen, T., Poulsen, T.S., Honore, T.S., and Johnson, H.E. (2000). Quantification of minimal residual disease in multiple myeloma using an allele-specific real-time PCR assay. *Exp. Hematol.* 28, 1039–1045.

Rijpens, N.P., and Herman, L.M. (2002). Molecular methods for identification and detection of bacterial food pathogens. *J. AOAC. Int.* 85, 984–995.

Ririe, K.M., Rasmussen, R.P., and Wittwer, C.T. (1997). Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal. Biochem.* 245, 154–160.

Risch, N.J. (2000). Searching for genetic determinants in the new millennium. *Nature* 405, 847–856.

Rondini, S., Mensah-Quainoo, E., Troll, H., Bodmer, T., and Pluschke, G. (2003). Development and application of real-time PCR assay for quantification of *Mycobacterium ulcerans* DNA. *J. Clin. Microbiol.* 41, 4231–4237.

Rowley, J.D. (1998). The critical role of chromosome translocations in human leukemias. *Annu. Rev. Genet.* 32, 495–519.

Rudi, K., Nogva, H.K., Moen, B., Nissen, H., Bredholt, S., Moretro, T., Naterstad, K., and Holck, A. (2002). Development and application of new nucleic acid-based technologies for microbial community analyses in foods. *Int. J. Food Microbiol.* 78, 171–180.

Scholl, C., Breitinger, H., Schlenk, R.F., Döhner, H., Fröhling, S., and Döhner, K. (2003). Development of a real-time RT-PCR assay for the quantification of the most frequent MLL/AF9 fusion types resulting from translocation t(9;11)(p22;q23) in acute myeloid leukemia. *Genes, Chromosomes Cancer* 38, 274–280.

Senoo, M., Matsubara, Y., Fujii, K., Nagasaki, Y., Hiratsuka, M., Kure, S., Uehara, S., Okamura, K., Yajima, A., and Narisawa, K. (2000). Adenovirus-mediated *in utero* gene transfer in mice and guinea pigs: tissue distribution of recombinant andenovirus determined by quantitative TaqMan-polymerase chain reaction assay. *Mol. Genet. Metab.* 69, 269–276.

Stryer, L. (1978). Fluorescence energy transfer as a spectropic ruler. *Annu. Rev. Biochem.* 47, 819–846.

Summers, K.E., Goff, L.K., Wilson, A.G., Gupta, R.K., Lister, T.A., and Fitzgibbon, J. (2001). Frequency of the Bcl-2/IgH rearrangement in normal individuals: implications for the monitoring of disease in patients with follicular lymphoma. *J. Clin. Oncol.* 19, 420–424.

Tanaka, N., and Kimura, H. (2000). Quantitative analysis of cytomegalovirus load using a real-time PCR assay. *J. Med. Virol.* 60, 455–462.

Thelwell, N., Millington, S., Solinas, A., Booth, J., and Brown, T. (2000). Mode of action and application of Scorpion primers to mutation detection. *Nucleic Acids Res.* 28, 3752–3761.

Trogan, E., Choudhury, R.P., Dansky, H.M., Rong, J.X., Breslow, J.L., and Fisher, E.A. (2002). Laser capture microdissection analysis of gene expression in macrophages from atherosclerotic lesions of apolipoprotein E-deficient mice. *Proc. Natl. Acad. Sci. USA* 99, 2234–2239.

Tyagi, S., and Kramer, F.R. (1996). Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* 14, 303–308.

Tyagi, S., Bratu, D.P., and Kramer, F.R. (1998). Multicolor molecular beacons for allele discrimination. *Nat. Biotechnol.* 16, 49–53.

Van der Velden, V.H., Hochhaus, A., Cazzaniga, G., Szczepanski, T., Gabert, J., and van Dongen, J.J.M. (2003). Detection of minimal residual disease in hematological malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia* 17, 1013–1034.

Van der Velden, V.H., Cazzaniga, G., Schrauder, A., Hancock, J., Bader, P., Panzer-Grumayer, E.R., Flohr, T., Sutton, R., Cave, H., Madsen, H.O., Cayuela, J.M., Trka, J., Eckert, C., Foroni, L., Zur Stadt, U., Beldjord, K., Raff, T., van der Schoot, C.E., and van Dongen, J.J.European Study Group on MRD detection in ALL (ESG-MRD-ALL) (2007). Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. *Leukemia* 21, 604–611.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3, R34.

Van Etten, E., Branisteanu, D.D., Overbergh, L., Bouillon, R., Verstuyf, A., and Mathieu, C. (2003). Combination of a 1,25-dihydroxyvitamin D(3) analog and a bisphosphonate prevents experimental autoimmune encephalitis and preserves bone. *Bone* 32, 397–404.

Van Haeften, R., Palladino, S., Kay, I., Keil, T., Heath, C., and Waterer, G.W. (2003). A quantitative LightCycler PCR to detect Streptococcus pneumoniae in blood and CSF. *Diagn. Microbiol. Infect. Dis.* 47, 407–414.

von Ahsen, N., Oellerich, M., and Schütz, E. (2000). Use of two reporter dyes without interference in a single-tube rapid-cycle PCR: alpha1-Antitrypsin genotyping by multiplex real-time fluorescence PCR with the LightCycler. *Clin. Chem.* 46, 156–161.

Whitcombe, D., Theaker, J., Guy, S.P., Brown, T., and Little, S. (1999). Detection of PCR products using self-probing amplicons and fluorescence. *Nat. Biotechnol.* 17, 804–807.

White, P.L., Shetty, A., and Barnes, R.A. (2003). Detection of seven *Candida* species using the Light-Cycler system. *J. Med. Microbiol.* 52, 229–238.

Wittwer, C.T., Herrmann, M.G., Cameron, N.G., and Elenitoba-Johnson, K.S. (2001). Real-time Multiplex PCR assays. *Methods* 25, 430–442.

Yashima, A., Maesawa, C., Uchiyama, M., Tarusawa, M., Satoh, T., Satoh, M., Enomoto, S., Sugawara, K., Numaoka, H., Murai, K., Utsugisawa, T., Ishida, Y., and Masuda, T. (2003). Quantitative assessment of contaminating tumor cells in autologous peripheral blood stem cells of B-cell non-Hodgkin lymphomas using immunoglobulin heavy chain gene allele-specific oligonucleotide real-time quantitative polymerase chain reaction. *Leukemia Res.* 27, 925–934.

Zhu, G., Xiao, H., Mohan, V.P., Tanaka, K., Tyagi, S., Tsen, F., Salgame, P., and Chan, J. (2003). Gene expression in the tuberculous granuloma: analysis by laser capture microdissection and real-time PCR. *Cell. Microbiol.* 5, 445–453.

This page intentionally left blank

Chapter 8

Pyrosequencing

Sharon Marsh
Génome Québec and Montreal Heart Institute Pharmacogenomics Centre, Montreal, Quebec, Canada

8.1 INTRODUCTION

Pyrosequencing® is based on sequencing by synthesis. The chemistry was developed in the late 1990s (Nyren, 2007) and has been expanded into a medium-throughput technology for genetic and epigenetic analysis.

The assay utilizes the natural release of pyrophosphate whenever a nucleotide is incorporated onto an open 3′ DNA strand. The released pyrophosphate is used in a sulfurylase reaction releasing ATP. The released ATP is used by luciferase in the conversion of luciferin to oxyluciferin. The reaction results in the emission of light, which is recorded by a charge-coupled device (CCD) camera in the form of peaks, known as a Pyrogram® (Marsh *et al*., 2005; Ronaghi, 2001; Ronaghi *et al*., 1998; Fig. 8.1). When a nucleotide is not incorporated into the reaction, no pyrophosphate is released and the unused nucleotide is removed from the system by degradation through apyrase. This process is performed in a closed system in a single well. The processing is simple and relatively fast (post-PCR ~20 minutes/96 samples for a typical assay) and the cost of the reaction is comparable to other medium-throughput technologies. One of the major advantages is the flexibility of the system in terms of its range of applications. Another advantage is that assays developed in one laboratory, including published assays, are readily transferable to another lab with a Pyrosequencing instrument with no or minimal optimization. Finally, Pyrosequencing assays are amenable to a range of source DNA, including degraded/low-quality DNA. The flexibility of Pyrosequencing assay design and analysis makes it an ideal medium-throughput genotyping system that meets most research laboratory needs.

8.2 TECHNOLOGY

8.2.1 Liquid-Phase

The original Pyrosequencing instrument, the PSQ™96, used streptavadin-coated magnetic beads and a magnetic prep tool with disposable covers to process the post-PCR samples (Rose *et al*., 2003). The available software was limited to single nucleotide polymorphism (SNP) detection with only simplex reads, and allele quantification. Software for short sequence reads was also available. Major limitations involved the inability to assess tri-allelic (SNPs) or insertions and deletions (indels). Methylation analysis and DNA pooling analysis could be performed using the allele quantification software. The assay limitations were rectified with the PSQ™MA, which introduced multiplex, tri-allelic, and indel screening. The sensitivity of the camera in the earlier systems was considerably lower than that used in later systems. Subsequent instruments with improved camera sensitivity use streptavadin-coated sepharose beads with a vacuum prep tool in place of the magnetic beads (Marsh *et al*., 2005), allowing a lowering of PCR volume and consequently reducing reagent costs. In addition, specialized software is now available to analyze methylation.

A video run-through of a typical Pyrosequencing genotyping experiment is available online (King and Scott-Horton, 2008). Typically, PCR amplification is performed to generate an amplicon (optimum size 300 bp or less) containing the variant(s) of interest. One of the PCR primers contains a 5′ biotin tag. The PCR product is mixed with streptavadin-coated sepharose beads. Streptavadin has a high binding affinity for biotin. The PCR/bead complex is captured onto filter tips using a vacuum preparation tool and subsequently removed with sodium hydroxide to denature the PCR product. This way, the non-biotinylated DNA strand is removed and single-stranded PCR product is left attached to the sepharose beads. The single-stranded PCR product/bead complex is deposited into a plate containing an internal primer that will anneal typically within 5 bp of the variant(s) of interest in the PCR amplicon. The plate is heated and cooled to room temperature to allow annealing to take place.

Once the samples have been amplified, the plate is run on a Pyrosequencing instrument. Enzyme and substrate mixtures and nucleotides are dispensed using an inkjet-style cartridge into each sample well. Subsequently, nucleotides

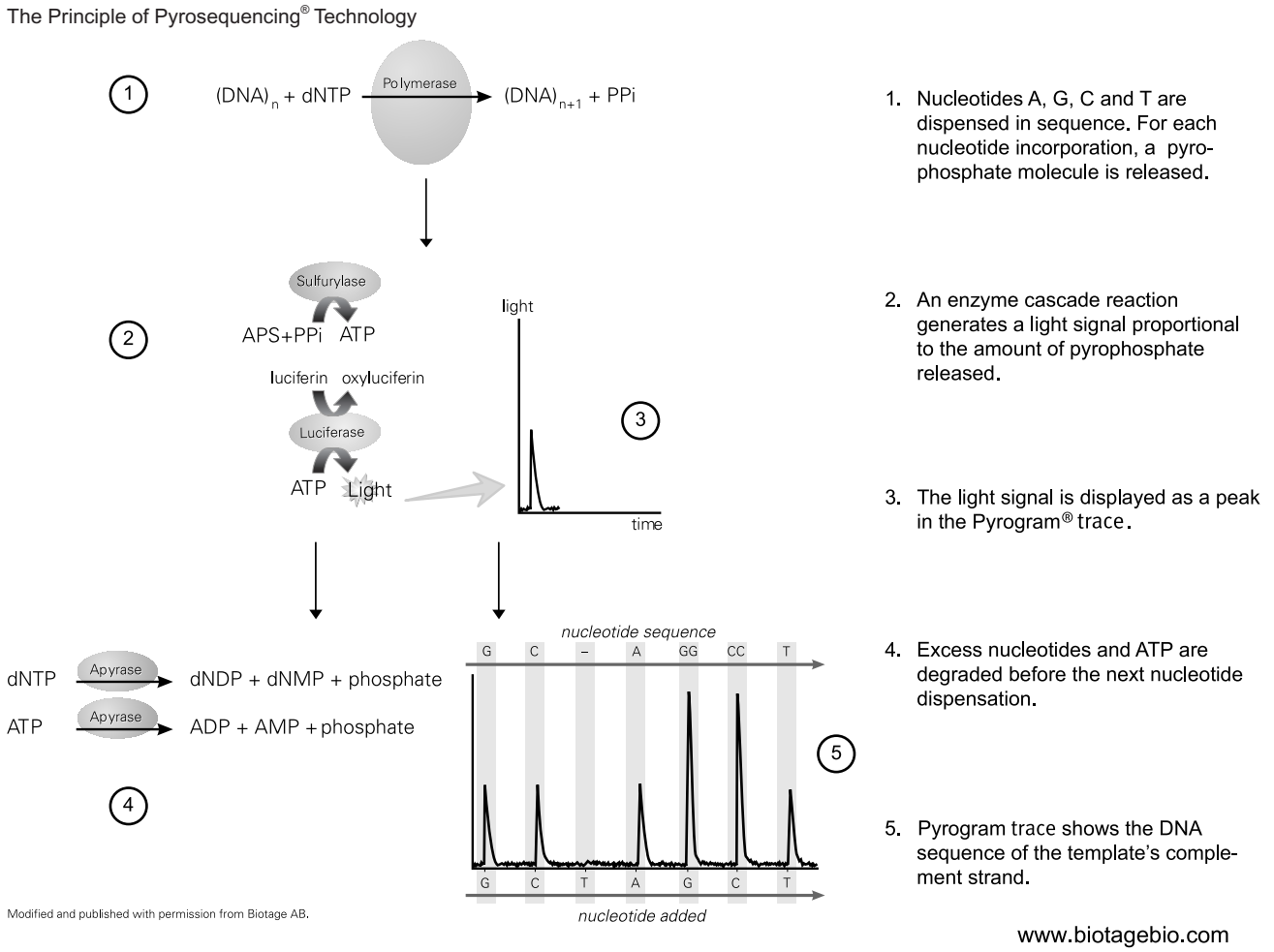
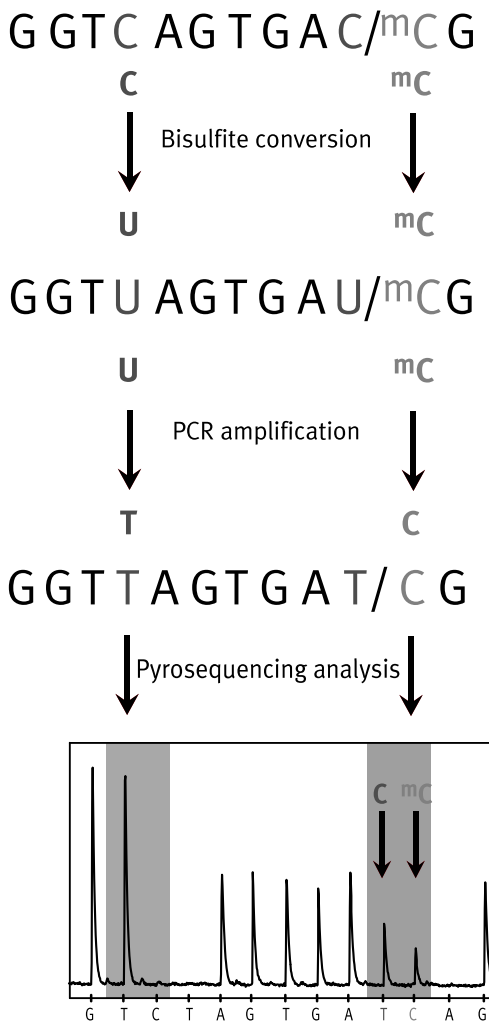


FIGURE 8.1 The principle of Pyrosequencing. Published with permission from Biotage AB.

are dispensed individually into each well and the Pyrosequencing reaction is allowed to take place (Fig. 8.1). Light emission relative to the incorporation of nucleotides is recorded using a CCD camera and visualized in the form of peaks known as Pyrograms™. A modified dATP (dATP-α-S) is used in place of adenine as dATP would be used directly in the luciferase reaction, leading to non-specific peaks for each dATP dispensation. Although the modified dATP does solve this problem, a higher background is typically seen for adenine positions on the Pyrogram, and this is automatically compensated for by the analysis software. The reaction can be monitored real-time in each well and typically takes one minute/nucleotide dispensation/96-well plate. A typical genotyping reaction to identify SNPs or indels will take 20 minutes per 96-well plate.

For polymorphism and methylation analysis (Fig. 8.2; Yu and Marsh, 2008) where the sequence is known in advance, the software can be programmed to dispense a specific nucleotide pattern corresponding to the sequence immediately adjacent to the 3′ end of the internal primer, running through the variant position and including a few extra bases after the variant. This provides an internal control where the pattern of the Pyrogram can be matched to the expected pattern based on the known sequence. Typically, the software automatically matches the expected and observed patterns and passes or fails individual wells based on the match (King and Scott-Horton, 2007). For sequencing analysis the nucleotides can be dispensed sequentially (e.g. A,C,G,T,A,C,G,T,A,C,G,T, etc.), and the resulting pattern of nucleotide incorporation is read in a similar manner to a sequencing chromatogram. One advantage with Pyrosequencing reads is that the peaks do not overlap as they may do in sequencing chromatograms and in some situations, e.g. for polymorphic dinucleotide repeats (Saeki *et al*., 2003) or regions where secondary structure is affecting the sequencing read, (Ronaghi *et al*., 1999) the results are easier to interpret.

Cost can be an issue with Pyrosequencing as biotin-labeled primers are expensive (especially compared to unmodified primers). This can be alleviated by the use of a universal primer (Guo and Milewicz, 2007). PCR primers are designed with one of them containing a 16 bp tag.



Unmethylated Cytosine C is transformed to Uracil U by the bisulfite treatment, whereas methylated Cytosine ^{m}C remains unchanged

The products of the PCR reaction arise from DNA that originally contained either an unmethylated Cytosine at a given position (the proportion of which is represented by T in the final product) or a methylated Cytosine (the proportion represented by C)

Any Cytosine not followed by a Guanine is unmethylated; therefore it should show full conversion in all copies to Thymine. This is a useful quality control to confirm that the transformation process went to completion (blue column). Reference peaks (all peaks outside the coloured columns) confirm that the assay was performed at the correct positions. Methylation at each CpG site is individually quantified (orange column)

An example of a DNA sequence and its conversion by bisulfite treatment and further amplification by PCR. By Pyrosequencing analysis, unmethylated Cytosine, C is measured as the relative content of T at the CpG site, and methylated Cytosine, ^{m}C, is measured as the relative content of C at the CpG site.

FIGURE 8.2 Detection of methylation by Pyrosequencing. Published with permission from Biotage AB. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

After PCR has been performed, the PCR product is further amplified with a biotinylated universal primer that primes from the tag. This allows the use of one biotinylated primer for all assays, rather than the need to purchase unique biotinylated primers for each assay.

Although the instrument analysis software does allow multiplex reactions to be run (either multiple variants on one PCR amplicon or up to three variants on separate amplicons in the same well) (Patel *et al*., 2007), assay design for multiplex reactions is complex and not currently automated, even within the commercially available Pyrosequencing Assay Design software. Consequently, this aspect of Pyrosequencing still has limited functionality.

The system is also restricted where homopolymers are involved in or close to the polymorphism. Two or more of the same nucleotide occurring adjacently will lead to double or more peak height, rather than separate peaks for each base. The signal will reach a plateau, meaning more than 5 or 6 of the same nucleotide in a row will not be accurately distinguishable (less in the case of adenine as the signal tends to run higher). Consequently, not every sequence is amenable to Pyrosequencing analysis. However, as Pyrosequencing allows for more flexibility in primer placement (both PCR and internal primers) than most other genotyping assays, it is often possible to design the assay such that the internal primer covers the homopolymer stretch, eliminating the need for this to be present in the Pyrogram.

8.2.2 Solid-Phase

One of the limitations with liquid-phase Pyrosequencing is that after multiple nucleotide dispensations the apyrase levels are depleted. Also, the sequential addition of nucleotides dilutes the enzyme concentrations. These factors lead

to signal drop-off, visualized by decreased peak heights, and increased background noise, and limit the length of sequence that can accurately be determined (Ronaghi *et al*., 2007).

454 Sequencing™ technology (see also sections 8.2.3.4, 8.3.3 and Chapter 24) uses emulsion-based clonal amplification with primers that are both target specific and contain a tag allowing the fragments to bind to beads. These primers also allow internal (sequencing) primer binding. Each bead contains one single-stranded DNA copy and is subject to emulsified PCR (em-PCR™), which results in millions of copies per bead (Fig. 8.3). A Pico Titer Plate (PTP™) is used to hold the beads (one bead per well), nucleotides are added across the plate and a typical Pyrosequencing reaction ensues. The resulting signal is recorded via a CCD camera and visualized in the form of a Flowgram and analyzed using the Genome Sequencer FLX™ software. The system is flushed between each nucleotide addition to reduce background. The current system allows sequencing reads up to 500 bp at a rate of about 1 billion bases per day. Accuracy when compared to traditional sequencing is high (Huse *et al*., 2007) and the throughput and efficiency is greatly enhanced. The use of clonal amplification and fluidics system removed the problems associated with longer reads from the classic Pyrosequencing technology.

In addition to sequencing up to 500 bp stretches of DNA, 454 Sequencing™ technology system can be used for transcriptome analysis, to identify transcription factor binding sites, methylation analysis (after bisulfate transformation of the sample), and SNP and indel identification.

8.2.3 Instruments

8.2.3.1 PyroMark™ ID

The PyroMark™ID instrument will perform the majority of Pyrosequencing applications including analysis of di-, tri- or tetra-allelic SNPs (simplex or multiplex), insertions, deletions, methylation analysis (with the Pyro Q-CpG™ software), and allele quantification. Ninety-six samples are processed simultaneously. In addition, the PyroMark™ID has the capability for short sequencing, which can be used for microbial typing, among other applications. The IdentiFire™ software automatically compares short sequence runs to local databases to confirm the microbial species identified in the assay.

The software will automatically analyze each well, scoring them for pass, check or fail. Check indicates that user intervention is required to confirm the genotype. All wells can be manually edited and the software registers if an edit has taken place. Data output can be in the form of text or Excel report sheets, which can be modified to upload and archive into appropriate databases (Van Booven, 2007).

8.2.3.2 PyroMark™ MD

The PyroMark™MD instrument lacks the short sequence application but otherwise has similar capabilities to the PyroMark™ID. It can also be purchased with an optional automated plate loader. The automated plate loader allows a run of up to ten 96-well plates to be performed in one batch, minimizing hands-on time. A typical 96-well

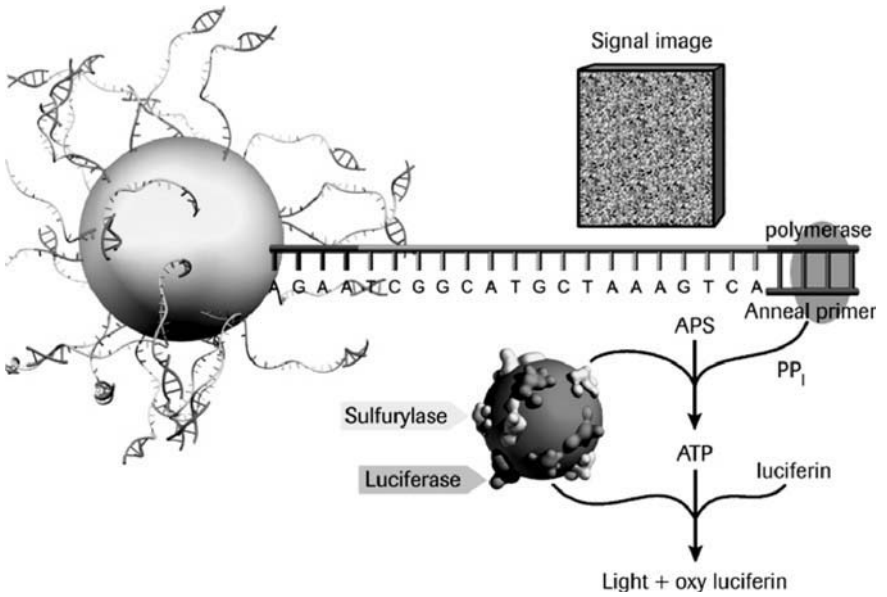


FIGURE 8.3 Sequencing reaction of the 454™ Genome Sequencer System. Millions of copies of a single clonal fragment are contained on each DNA Capture Bead. © Roche Diagnostics GmbH.

plate takes approximately 20 minutes; however, assays containing longer sequences will increase the run time.

8.2.3.3 PyroMark™ Q24 System

The PyroMark™Q24 instrument is currently approved for *in vitro* diagnostics in France, Germany, Ireland, Italy, Sweden, and the United Kingdom. This is a scaled down version of the PyroMark™MD. It assays 24 samples simultaneously (with no automated plate loading), allows SNP, indel and methylation analysis, and takes up a much smaller footprint than the PyroMark™MD or PyroMark™ID. Rather than having a linked computer to run the machine and analyze the data, data are available for download via a USB port and can be analyzed separately at a personal computer. This system is ideal for labs requiring a flexible genotyping system but lacking either the space or capacity for a 96-well system.

8.2.3.4 Genome Sequencer FLX™ Instrument

The GS FLX™ Standard Series produces an average of 250 bp of sequence, at a rate of approximately 100 megabases/7.5 hours. The updated GS FLX Titanium™ series can produce an average of 400 bp of sequence (mode 500 bp) at a rate of approximately 1 billion bases per day. The Titanium series has a redesigned PTP™ with a higher density of wells compared to the standard series, and improved reagents allowing longer sequencing stretches to be analyzed.

The software allows integration with LIMS and barcode scanning and the DNA sequence data are readily exportable to Excel. The GS *de novo* Assembler software aligns and assembles the DNA sequence generated; GS Reference Mapper software allows comparison between known DNA sequences and genomes; and the GS Amplicon Variant Analyzer Software analyzes methylation, SNP, and indel data, as well as analyzing sequencing data.

8.2.4 PyroMark™ tests

Pre-made kits (for research purposes) are available for several commonly studied polymorphisms, fungal and mycobacteria typing, and methylation (Table 8.1). An up-to-date list of currently available kits can be found at http://www.pyrosequencing.com/DynPage.aspx?id=14600. The assays contain optimized pre-tested reagents and primers, eliminating the need for assay design. This can be a useful time saver as optimizing PCR conditions and primer redesign are the main time-consuming bottlenecks when using Pyrosequencing.

8.3 APPLICATIONS

Single nucleotide polymorphisms or mutations (medical diagnostics and pharmacogenomics), indels, short tandem repeats, pooled allele frequencies, HLA typing, gene copy number, allelic imbalance, methylation status, viral typing, and short sequencing stretches are among the numerous applications of Pyrosequencing (Marsh, 2007b). In addition, massively parallel sequencing can be performed using the 454 Life Sciences™ system (see also Chapter 24). A comprehensive database of publications involving Pyrosequencing can be found at http://www.pyrosequencing.com/DynPage.aspx?id=8890&mn1=1388. A similar database for 454 Sequencing™ technology can be found at http://www.454.com/publications-and-resources/index.asp.

8.3.1 Diagnostics

8.3.1.1 Microbial Typing

Rapid identification of species of bacteria, viruses, molds and yeast is often essential for determining the source of infection and the relevant treatment (see also Chapter 30). Often short stretches of nucleotides are sufficient to unambiguously identify different species. This is beyond the majority of genotyping systems but standard sequencing techniques are an expensive solution. Pyrosequencing using the sequencing software has proven useful for multiple typing applications (Ronaghi and Elahi, 2002; Woodford *et al*., 2007). A comprehensive methodology for phage typing is available (Rahim, 2007) and pre-made kits and primers for bacteria and fungi typing are available from Pyrosequencing (Table 8.1). The PyroMark™ID Pyrosequencing instrument in conjunction with IdentiFire™ software offers a rapid typing package. For 153 yeast isolates the fungal kit from Pyrosequencing identified all test organisms, including both common and rare *Candida* species (Borman *et al*., 2008). Pyrosequencing was also used to accurately identify mycobacteria isolates with 93.5% accuracy (Heller *et al*., 2008).

8.3.1.2 HLA Typing

HLA is the major histocompatibility complex region on chromosome 6q21. This is the most polymorphic region of the human genome identified to date (Choo, 2007). The primary role is in immune response; however, HLA antigens are implicated for a range of biological roles. HLA cross-matching is important in transplantation to avoid transplant rejection. The HLA-DRB1 locus is also commonly associated with disease susceptibility, for example rheumatoid arthritis, Type I diabetes mellitus and ankylosing spondylitis are all associated with specific HLA types (Choo, 2007). The sequencing software for Pyrosequencing offers a high resolution solution to HLA typing, which can be difficult to interpret by standard sequencing (Ringquist *et al*., 2002, 2007). Pyrosequencing methods have been used to identify alleles within the HLA locus, including identification of HLA-DRB1*04 as a risk factor for development of alopecia areata, and HLA-DRB1*03 (in particular HLA-DRB1*0301) as a protective allele for the same disease (Entz *et al*., 2006).

TABLE 8.1 PyroMark tests available at time of publication (more information available at http://www.pyrosequencing.com/DynPage.aspx?id=14600 and http://www.pyrosequencing.com/DynPage.aspx?id=24914).

Kit	Test	Research use only	System
PyroMark™ MOTT	Identification of mycobacteria other than tuberculosis	Also available for use in CLIA laboratories	PyroMark™ID
PyroMark™ Fungi Test	Identification of commonly encountered fungi, including yeasts and molds	Also available for use in CLIA laboratories	PyroMark™ID
PyroMark™ *APOE*	Codon 112 and 158 polymorphisms	Yes	PyroMark™ID and PyroMarkMD
PyroMark™ *BRAF*	p.V600E mutation	Yes	PyroMark™ID and PyroMarkMD
PyroMark™ *HFE*	p.H63D and p.S65C mutations	Yes	PyroMark™ID and PyroMarkMD
PyroMark™ *KRAS*	Contiguous multi-variable mutations in exons 12, 13, and 61	Yes	PyroMark™ID and PyroMarkMD
PyroMark™ LINE-1	Methylation levels of LINE-1 transposons	Yes	PyroMark™ID and PyroMarkMD
PyroMark™ *MGMT*	Methylation levels of the *MGMT* promoter region	Yes	PyroMark™ID and PyroMarkMD
PyroMark™ *MLH1*	Methylation levels of the *MLH1* promoter region	Yes	PyroMark™ID and PyroMarkMD
PyroMark™ *MTHFR*	c.677C > T genotype	Yes	PyroMark™ID and PyroMarkMD
PyroMark™ *p16*	Methylation levels of the *p16* promoter region	Yes	PyroMark™ID and PyroMarkMD
PyroMark™ Prader Willi/ Angelman	Methylation levels of paternal and maternally derived chromosome 15	Yes	PyroMark™ID and PyroMarkMD
PyroMark™ Q24 *BRAF*	p.V600E mutation	Yes	PyroMark™Q24
PyroMark™ Q24 *KRAS*	Contiguous multi-variable mutations in exons 12, 13, and 61	Yes	PyroMark™Q24
PyroMark™ Q24 LINE-1	Methylation levels of LINE-1 transposons	Yes	PyroMark™Q24
PyroMark™ Q24 *MGMT*	Methylation levels of the *MGMT* promoter region	Yes	PyroMark™Q24
PyroMark™ Q24 *MLH1*	Methylation levels of the *MLH1* promoter region	Yes	PyroMark™Q24
PyroMark™ Q24 *p16*	Methylation levels of the *p16* promoter region	Yes	PyroMark™Q24

8.3.1.3 Forensics

Molecular forensic analysis (see also Chapter 26) can also benefit from Pyrosequencing technology. Genotyping for red blood cells, HLA antigens and platelets has been successfully and accurately performed with Pyrosequencing technology (Wu and Csako, 2006). In addition, in mixed samples, variations in mitochondrial DNA can be assessed by Pyrosequencing to identify individuals within the mixture (Andreasson *et al*., 2006). This application takes advantage of the allele quantification software, where a percentage of each peak in the variant site is given.

8.3.1.4 Copy Number

Copy number variants can be determined utilizing the allele quantification software. Typically, a comparison with a region known not to be amplified or deleted is

performed to determine the relative amplification or deletion of the gene in question (Pielberg and Andersson, 2007; Soderback *et al*., 2005). Copy number detection has been carried out for *BRAF*, a gene involved in cutaneous malignant melanomas. A mutation (p.V600E) can be readily assessed by a pre-made PyroMark™ kit (Table 8.1). Further assessment of *BRAF* copy number can determine the frequency of *BRAF* p.V600E mutation alleles in each individual (Spittle *et al*., 2007). Animal studies also make use of Pyrosequencing to detect gene copy number. A dominant white coat color in domestic pigs is caused by duplication of the *KIT* receptor gene, and this is readily assessed by Pyrosequencing (Pielberg and Andersson, 2007).

8.3.2 Research

8.3.2.1 DNA Pooling

The need for rapid and inexpensive detection of allele frequencies in case-control studies has led to the use of pools of DNA samples. This allows a quick comparison of overall allele frequency in the case pool compared to the control pool. The Pyrosequencing allele quantification software allows allele frequencies in pooled DNA samples to be determined (Wasson, 2007; Wasson *et al*., 2002). This has been done successfully and accurately in pools of up to 374 individuals with accuracy to detect 4% difference in allele frequency between populations (Neve *et al*., 2002). Sample sets assessed using this methodology include Parkinson's disease case-control pools where 230 SNPs were assessed using a universal biotinylated primer (Doostzadeh *et al*., 2008). In addition, a case-control study for lung cancer risk identified differences in allele frequency between pools of 50 cases and 50 controls (Spinola *et al*., 2007). The larger the number of DNA samples in the pool the lower the discrepancy between the average allele frequencies based on individual samples compared to the pooled frequency (Lavebratt *et al*., 2004). Presumably in smaller pools pipetting errors, etc. will be more obvious, whereas their contribution to the overall frequency will decrease in larger pools. The same allele quantification software can be utilized for other applications, such as copy number and allelic imbalance.

8.3.2.2 Allelic Imbalance

In situations where non-coding polymorphisms appear to have a functional effect there is the possibility that regulatory polymorphisms in linkage disequilibrium (e.g. promoter region variants affecting expression, or 3′ untranslated region polymorphisms affecting RNA stability) are responsible for altering gene expression. The assessment of allelic imbalance in heterozygous samples can give an indication if cis-acting variants are responsible and assist in narrowing down the search for functional polymorphisms in the gene. Pyrosequencing has been used to assess allelic imbalance for heterozygous SNPs in cDNA (Wang and Elbein, 2007). A very commonly studied polymorphism, *ABCB1* c.3435C > T appears to be functional even though it is a synonymous polymorphism (does not change the amino acid sequence). The search for possible function for this SNP has led to the identification of allelic imbalance. Variability in expression levels between the C and T alleles in individuals heterozygous for c.3435C > T was seen using Pyrosequencing and two possible cis-acting regulatory polymorphisms in the promoter and upstream of the gene have been identified (Loeuillet *et al*., 2007).

8.3.2.3 Gene Copy Number

Studies of copy number variations have increased substantially with the increased knowledge of their incidence in the human genome. In addition to the diagnostic potential (see above), research involving gene copy number variation can be substantially improved by the use of Pyrosequencing. For example, in addition to assessing the common polymorphisms within the *CYP2D6* gene, gene amplification of *CYP2D6* can be identified by Pyrosequencing using the *CYP2C8P* pseudogene as a comparison (Soderback *et al*., 2005). This allows the rapid detection of poor metabolizers (*CYP2D6*5*; gene deletion) and ultra-rapid metabolizers (*CYP2D6*2*xN; multiple copies of the *CYP2D6* gene) to be identified, allowing information about a patient's ability to metabolize many drugs to be known.

8.3.2.4 Pharmacogenomics

Pharmacogenomics encompasses the search for genetic determinants of drug response and toxicity (Evans and McLeod, 2003; Marsh, 2007a; see also Chapter 22). Pyrosequencing is widely applicable to pharmacogenomics research as it allows the analysis of SNPs, mutations, indels, copy number, and short tandem repeats on a single system (Marsh *et al*., 2005). The technology has been used to identify differences in allele frequencies of pharmacogenomic variants between populations (King *et al*., 2005; Lanfear *et al*., 2004a; Marsh *et al*., 2004; Oliveira *et al*., 2007), and also to determine associations between variants and toxicity (de Jong *et al*., 2004; Hoskins *et al*., 2008; Marsh *et al*., 2007a; Ranganathan *et al*., 2008) or outcome (Cresci *et al*., 2008; Hoskins *et al*., 2008; Marsh *et al*., 2007a, b) in patients receiving medications. Pyrosequencing has been used for pharmacogenomics research in many fields, including oncology (Marsh *et al*., 2007a, b), cardiology (Cresci *et al*., 2008; Lanfear *et al*., 2004a, b) and rheumatology (Ranganathan *et al*., 2004, 2008). In a study of the beta-adrenergic receptor (in the context of beta-blocker efficacy) a novel variant was identified through routine analysis of a common variant using Pyrosequencing (Lanfear *et al*., 2004b). This novel variant would have been missed by other commonly used genotyping techniques.

8.3.2.5 Methylation

In addition to standard genetic variations at the DNA level, in cancer pharmacogenomics aberrant methylation of the tumor genome may also play an important role in drug response and toxicity. This is also readily analyzed using Pyrosequencing technology (Tost and Gut, 2007; Fig. 8.2; see also Chapter 14) with the Pyro Q-CpG™ software, which allows the analysis of multiple CpG (methylation) sites within a single reaction. In a study of 52 tumors from colorectal cancer patients (with matched tumor samples for comparison) methylation status of the promoter region of three genes (*MLH1*, *DPYD*, and *CDKN2A*) was determined using Pyrosequencing. An association between increased *MLH1* methylation status and decreased *MLH1* RNA expression was identified (Yu *et al*., 2008).

8.3.3 Massively Parallel Sequencing

The availability of high-throughput sequencing at a reasonable cost, as provided by the 454 Sequencing™ technology, has opened the doors to many avenues of research (see also Chapter 14). Up to 100 million bases of DNA sequence can be produced within two days (Bushman *et al*., 2008). Microbial typing is an obvious application for massively parallel sequencing. The 16s rRNA gene can be sequenced to identify the source species for many bacteria. Using 454 Sequencing™ technology, a total of 100,000 to 200,000 sequence reads of ~100 bp were obtained per sample from eight samples with the aim of identifying vaginal bacteria species during pregnancy. These data may help to optimize treatment and better understand the influence of microorganisms on pregnancy complications (Sundquist *et al*., 2007). In addition, 454™ sequencing technology can be used to identify mutations in the HIV virus, leading to drug resistance. This is particularly useful as rare mutations can be identified via this method, as well as the commonly screened for mutations (Bushman *et al*., 2008).

In addition to microbial typing, human and animal genetic analysis can benefit from the availability of massively parallel sequencing. A famous application of 454 Sequencing™ has been to sequence the entire genome of James D. Watson (Wheeler *et al*., 2008), which was completed in a period of two months. Other high profile studies include the sequencing of the woolly mammoth genome (Miller *et al*., 2008) and the Neanderthal genome (Noonan *et al*., 2006).

In a cytogenetically normal patient with acute myeloid leukemia, massively parallel sequencing of the whole genome identified eight novel mutations when compared to the patient's germline genome. The percentage of the patient's tumor cells and relapse samples containing these mutations was determined using 454 Sequencing™ technology (Ley *et al*., 2008). This method also detected a small percentage of cells containing the mutations in the germline sample, providing information on the level of tumor contamination in the germline sample (Ley *et al*., 2008).

DNA pooling is also possible with this system. Pools of DNA amplicons were sequenced using 454 Sequencing™ technology to identify variants in known disease genes. From four pools of between 27 and 164 amplicons a total of 80 million base pairs of sequence was produced, and 412 variants were identified, including validation of known variants in the target genes (Bordoni *et al*., 2008). This application has the potential to improve the use of DNA pooling to identify novel variants in disease and other clinically relevant genes.

8.4 CONCLUSIONS

Pyrosequencing is a robust medium-throughput genotyping system capable of analyzing a wide range of DNA variation. The methodology is easy to perform and readily transferable to other laboratories. Applications vary widely from research to diagnostics. No system is perfect, and the main problems with Pyrosequencing include the need for expensive biotinylated primers, and the inability to accurately detect variants within long (>5 or 6 bp) homopolymer stretches. In addition, multiplexing, while possible (Lee *et al*., 2008; Okada *et al*., 2008; Patel *et al*., 2007), is difficult to design. The adaptation of Pyrosequencing technology to massively parallel sequencing further extends the utility, with applications for both prokaryotic and eukaryotic sequencing. The advent of 454 Sequencing™ technology will ensure that the chemistry behind Pyrosequencing has a lasting future.

ACKNOWLEDGEMENTS

The assistance of Maritha Lundman, Margery Doyle and Pyrosequencing, and Kate Montgomery and Roche Diagnostics is greatly appreciated. The author is funded by Génome Québec.

REFERENCES

Andreasson, H., Nilsson, M., Budowle, B., Frisk, S., and Allen, M. (2006). Quantification of mtDNA mixtures in forensic evidence material using pyrosequencing. *Int. J. Legal Med.* 120, 383–390.

Bordoni, R., Bonnal, R., Rizzi, E., Carrera, P., Benedetti, S., Cremonesi, L., Stenirri, S., Colombo, A., Montrasio, C., Bonalumi, S., Albertini, A., Bernardi, L.R., Ferrari, M., and De Bellis, G. (2008). Evaluation of human gene variant detection in amplicon pools by the GS-FLX parallel Pyrosequencer. *BMC Genomics* 9, 464.

Borman, A.M., Linton, C.J., Miles, S.J., and Johnson, E.M. (2008). Molecular identification of pathogenic fungi. *J. Antimicrob. Chemother.* 61(Suppl. 1), i7–i12.

Bushman, F.D., Hoffmann, C., Ronen, K., Malani, N., Minkah, N., Rose, H.M., Tebas, P., and Wang, G.P. (2008). Massively parallel pyrosequencing in HIV research. *AIDS* 22, 1411–1415.

Choo, S.Y. (2007). The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei Med. J.* 48, 11–23.

Cresci, S., Jones, P.G., Sucharov, C.C., Marsh, S., Lanfear, D.E., Garsa, A., Courtois, M., Weinheimer, C.J., Wu, J., Province, M.A., Kelly, D.P., McLeod, H.L., and Spertus, J.A. (2008). Interaction between PPARA genotype and beta-blocker treatment influences clinical outcomes following acute coronary syndromes. *Pharmacogenomics* 9, 1403–1417.

de Jong, F.A., Marsh, S., Mathijssen, R.H., King, C., Verweij, J., Sparreboom, A., and McLeod, H.L. (2004). ABCG2 pharmacogenetics: ethnic differences in allele frequency and assessment of influence on irinotecan disposition. *Clin. Cancer Res.* 10, 5889–5894.

Doostzadeh, J., Shokralla, S., Absalan, F., Jalili, R., Mohandessi, S., Langston, J.W., Davis, R.W., Ronaghi, M., and Gharizadeh, B. (2008). High throughput automated allele frequency estimation by pyrosequencing. *PLoS ONE* 3, e2693.

Entz, P., Blaumeiser, B., Betz, R.C., Lambert, J., Seymons, K., Eigelshoven, S., Hanneken, S., Kruse, R., Nurnberg, P., Nagy, M., and Nothen, M.M. (2006). Investigation of the HLA-DRB1 locus in alopecia areata. *Eur. J. Dermatol.* 16, 363–367.

Evans, W.E., and McLeod, H.L. (2003). Pharmacogenomics – drug disposition, drug targets, and side effects. *N. Engl. J. Med.* 348, 538–549.

Guo, D.C., and Milewicz, D.M. (2007). Universal primer applications for pyrosequencing. *Methods Mol. Biol.* 373, 57–62.

Heller, L.C., Jones, M., and Widen, R.H. (2008). Comparison of DNA pyrosequencing with alternative methods for identification of mycobacteria. *J. Clin. Microbiol.* 46, 2092–2094.

Hoskins, J.M., Marcuello, E., Altes, A., Marsh, S., Maxwell, T., Van Booven, D.J., Pare, L., Culverhouse, R., McLeod, H.L., and Baiget, M. (2008). Irinotecan pharmacogenetics: influence of pharmacodynamic genes. *Clin. Cancer Res.* 14, 1788–1796.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* 8, R143.

King, C.R., and Scott-Horton, T. (2007). Pyrosequencing: a simple method for accurate genotyping. *Methods Mol. Biol.* 373, 39–56.

King, C.R., and Scott-Horton, T.J. (2008). Pyrosequencing: a simple method for accurate genotyping. JoVE. 11, http://www.jove.com/index/Details.stp?ID=630, doi: 610.3791/3630

King, C.R., Yu, J., Freimuth, R.R., McLeod, H.L., and Marsh, S. (2005). Interethnic variability of ERCC2 polymorphisms. *Pharmacogenomics J.* 5, 54–59.

Lanfear, D.E., Marsh, S., Cresci, S., Spertus, J.A., and McLeod, H.L. (2004a). Frequency of compound genotypes associated with beta-blocker efficacy in congestive heart failure. *Pharmacogenomics* 5, 553–558.

Lanfear, D.E., Marsh, S., and McLeod, H.L. (2004b). Caution with beta1-adrenergic receptor genotyping. *Clin. Pharmacol. Ther.* 76, 185–186.

Lavebratt, C., Sengul, S., Jansson, M., and Schalling, M. (2004). Pyrosequencing-based SNP allele frequency estimation in DNA pools. *Hum. Mutat.* 23, 92–97.

Lee, S.S., Kim, W.Y., Jang, Y.J., and Shin, J.G. (2008). Duplex pyrosequencing of the TPMT3C and TPMT6 alleles in Korean and Vietnamese populations. *Clin. Chim. Acta,* 398, 82–85.

Ley, T.J., Mardis, E.R., Ding, L., Fulton, B., McLellan, M.D., Chen, K., Dooling, D., Dunford-Shore, B.H., McGrath, S., Hickenbotham, M., Cook, L., Abbott, R., Larson, D.E., Koboldt, D.C., Pohl, C., Smith, S., Hawkins, A., Abbott, S., Locke, D., Hillier, L.W., Miner, T., Fulton, L., Magrini, V., Wylie, T., Glasscock, J., Conyers, J., Sander, N., Shi, X., Osborne, J.R., Minx, P., Gordon, D., Chinwalla, A., Zhao, Y., Ries, R. E., Payton, J.E., Westervelt, P., Tomasson, M.H., Watson, M., Baty, J., Ivanovich, J., Heath, S., Shannon, W.D., Nagarajan, R., Walter, M.J., Link, D.C., Graubert, T.A., DiPersio, J.F., and Wilson, R.K. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456, 66–72.

Loeuillet, C., Weale, M., Deutsch, S., Rotger, M., Soranzo, N., Wyniger, J., Lettre, G., Dupre, Y., Thuillard, D., Beckmann, J.S., Antonarakis, S.E., Goldstein, D.B., and Telenti, A. (2007). Promoter polymorphisms and allelic imbalance in ABCB1 expression. *Pharmacogenet Genomics* 17, 951–959.

Marsh, S. (2007a). Pharmacogenomics. *Ann. Oncol.* 18(Suppl. 9), ix24–28.

Marsh, S. (2007b). Pyrosequencing applications. *Methods Mol. Biol.* 373, 15–24.

Marsh, S., King, C.R., Ahluwalia, R., and McLeod, H.L. (2004). Distribution of ITPA P32T alleles in multiple world populations. *J. Hum. Genet.* 49, 579–581.

Marsh, S., King, C.R., Garsa, A.A., and McLeod, H.L. (2005). Pyrosequencing of clinically relevant polymorphisms. *Methods Mol. Biol.* 311, 97–114.

Marsh, S., Paul, J., King, C.R., Gifford, G., McLeod, H.L., and Brown, R. (2007a). Pharmacogenetic assessment of toxicity and outcome after platinum plus taxane chemotherapy in ovarian cancer: the Scottish Randomised Trial in Ovarian Cancer. *J. Clin. Oncol.* 25, 4528–4535.

Marsh, S., Somlo, G., Li, X., Frankel, P., King, C.R., Shannon, W.D., McLeod, H.L., and Synold, T.W. (2007b). Pharmacogenetic analysis of paclitaxel transport and metabolism genes in breast cancer. *Pharmacogenomics J.* 7, 362–365.

Miller, W., Drautz, D.I., Ratan, A., Pusey, B., Qi, J., Lesk, A.M., Tomsho, L.P., Packard, M.D., Zhao, F., Sher, A., Tikhonov, A., Raney, B., Patterson, N., Lindblad-Toh, K., Lander, E.S., Knight, J.R., Irzyk, G.P., Fredrikson, K.M., Harkins, T.T., Sheridan, S., Pringle, T., and Schuster, S.C. (2008). Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456, 387–390.

Neve, B., Froguel, P., Corset, L., Vaillant, E., Vatin, V., and Boutin, P. (2002). Rapid SNP allele frequency determination in genomic DNA pools by pyrosequencing. *Biotechniques* 32, 1138–1142.

Noonan, J.P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Paabo, S., Pritchard, J.K., and Rubin, E.M. (2006). Sequencing and analysis of Neanderthal genomic DNA. *Science* 314, 1113–1118.

Nyren, P. (2007). The history of pyrosequencing. *Methods Mol. Biol.* 373, 1–14.

Okada, Y., Nakamura, K., Adachi, A., Watai, Y., Horiuchi, R., and Yamamoto, K. (2008). Development of a single-tube PCR-pyrosequencing method for the simultaneous and rapid detection of four variant alleles of CYP2C9 gene polymorphism. *J. Clin. Pharm. Ther.* 33, 187–192.

Oliveira, E., Marsh, S., van Booven, D.J., Amorim, A., Prata, M.J., and McLeod, H.L. (2007). Pharmacogenetically relevant polymorphisms in Portugal. *Pharmacogenomics* 8, 703–712.

Patel, P., Shiao, Y.H., and Fortina, P. (2007). Multiplex pyrosequencing for DNA variation analysis. *Methods Mol. Biol.* 373, 75–88.

Pielberg, G., and Andersson, L. (2007). Gene copy number detection in animal studies. *Methods Mol. Biol.* 373, 147–156.

Rahim, A.A. (2007). Pyrosequencing of phage display libraries for the identification of cell-specific targeting ligands. *Methods Mol. Biol.* 373, 135–146.

Ranganathan, P., Culverhouse, R., Marsh, S., Ahluwalia, R., Shannon, W.D., Eisen, S., and McLeod, H.L. (2004). Single nucleotide polymorphism profiling across the methotrexate pathway in normal

subjects and patients with rheumatoid arthritis. *Pharmacogenomics* 5, 559–569.

Ranganathan, P., Culverhouse, R., Marsh, S., Mody, A., Scott-Horton, T.J., Brasington, R., Joseph, A., Reddy, V., Eisen, S., and McLeod, H.L. (2008). Methotrexate (MTX) pathway gene polymorphisms and their effects on MTX toxicity in Caucasian and African American patients with rheumatoid arthritis. *J. Rheumatol.* 35, 572–579.

Ringquist, S., Alexander, A.M., Rudert, W.A., Styche, A., and Trucco, M. (2002). Pyrosequence-based typing of alleles of the HLA-DQB1 gene. *Biotechniques* 33, 166–170, 172, 174–165.

Ringquist, S., Styche, A., Rudert, W.A., and Trucco, M. (2007). Pyrosequencing-based strategies for improved allele typing of human leukocyte antigen loci. *Methods Mol. Biol.* 373, 115–134.

Ronaghi, M. (2001). Pyrosequencing sheds light on DNA sequencing. *Genome Res.* 11, 3–11.

Ronaghi, M., and Elahi, E. (2002). Pyrosequencing for microbial typing. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 782, 67–72.

Ronaghi, M., Nygren, M., Lundeberg, J., and Nyren, P. (1999). Analyses of secondary structures in DNA by pyrosequencing. *Anal. Biochem.* 267, 65–71.

Ronaghi, M., Shokralla, S., and Gharizadeh, B. (2007). Pyrosequencing for discovery and analysis of DNA sequence variations. *Pharmacogenomics* 8, 1437–1441.

Ronaghi, M., Uhlen, M., and Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* 281, 363, 365.

Rose, C.M., Marsh, S., Ameyaw, M.M., and McLeod, H.L. (2003). Pharmacogenetic analysis of clinically relevant genetic polymorphisms. *Methods Mol. Med.* 85, 225–237.

Saeki, M., Saito, Y., Jinno, H., Tohkin, M., Kurose, K., Kaniwa, N., Komamura, K., Ueno, K., Kamakura, S., Kitakaze, M., Ozawa, S., and Sawada, J. (2003). Comprehensive UGT1A1 genotyping in a Japanese population by pyrosequencing. *Clin. Chem.* 49, 1182–1185.

Soderback, E., Zackrisson, A.L., Lindblom, B., and Alderborn, A. (2005). Determination of CYP2D6 gene copy number by pyrosequencing. *Clin. Chem.* 51, 522–531.

Spinola, M., Leoni, V.P., Galvan, A., Korsching, E., Conti, B., Pastorino, U., Ravagnani, F., Columbano, A., Skaug, V., Haugen, A., and Dragani, T.A. (2007). Genome-wide single nucleotide polymorphism analysis of lung cancer risk detects the KLF6 gene. *Cancer Lett.* 251, 311–316.

Spittle, C., Ward, M.R., Nathanson, K.L., Gimotty, P.A., Rappaport, E., Brose, M.S., Medina, A., Letrero, R., Herlyn, M., and Edwards, R.H. (2007). Application of a BRAF pyrosequencing assay for mutation detection and copy number analysis in malignant melanoma. *J. Mol. Diagn.* 9, 464–471.

Sundquist, A., Bigdeli, S., Jalili, R., Druzin, M.L., Waller, S., Pullen, K.M., El-Sayed, Y.Y., Taslimi, M.M., Batzoglou, S., and Ronaghi, M. (2007). Bacterial flora-typing with targeted, chip-based Pyrosequencing. *BMC Microbiol.* 7, 108.

Tost, J., and Gut, I.G. (2007). Analysis of gene-specific DNA methylation patterns by pyrosequencing technology. *Methods Mol. Biol.* 373, 89–102.

Van Booven, D.J. (2007). Pyrosequencing genotype storage techniques. *Methods Mol. Biol.* 373, 177–186.

Wang, H., and Elbein, S.C. (2007). Detection of allelic imbalance in gene expression using pyrosequencing. *Methods Mol. Biol.* 373, 157–176.

Wasson, J. (2007). Allele quantification and DNA pooling methods. *Methods Mol. Biol.* 373, 63–74.

Wasson, J., Skolnick, G., Love-Gregory, L., and Permutt, M.A. (2002). Assessing allele frequencies of single nucleotide polymorphisms in DNA pools by pyrosequencing technology. *Biotechniques* 32, 1144–1150.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C.L., Irzyk, G.P., Lupski, J.R., Chinault, C., Song, X.Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D.M., Margulies, M., Weinstock, G.M., Gibbs, R.A., and Rothberg, J.M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872–876.

Woodford, N., North, S.E., and Ellington, M.J. (2007). Detecting mutations that confer oxazolidinone resistance in gram-positive bacteria. *Methods Mol. Biol.* 373, 103–114.

Wu, Y.Y., and Csako, G. (2006). Rapid and/or high-throughput genotyping for human red blood cell, platelet and leukocyte antigens, and forensic applications. *Clin. Chim. Acta,* 363, 165–176.

Yu, J., Freimuth, R.R., Culverhouse, R., Marsh, S., Watson, M.A., and McLeod, H.L. (2008). DNA methylotype analysis in colorectal cancer. *Oncol. Rep.* 20, 921–927.

Yu, J., and Marsh, S. (2008). SNP and DNA methylation analysis with Pyrosequencing. In *Biomarker Methods in Drug Discovery and Development: Methods and Protocols* (Wang, F., ed.), pp. 119–140. Humana Press.

Chapter 9

Application of Padlock and Selector Probes in Molecular Medicine

Mats Nilsson, Chatarina Larsson, Johan Stenberg, Jenny Göransson, Ida Grundberg, Magnus Isaksson, Tim Conze and Sara Henriksson

Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory, Se-751 85, Uppsala, Sweden

9.1 INTRODUCTION

One of the rate limiting factors in the progression of molecular medicine is the techniques available for molecular analysis. Technology improvements are often followed by scientific breakthrough, exemplified by techniques such as molecular cloning, DNA sequencing, and the polymerase chain reaction (PCR). A common objective in molecular medicine is to analyze the quantity and/or sequence of a large number of candidate genes or loci of genetic variation. The number of genes under study is often significantly larger than one and yet significantly fewer than all. This intermediary range in the analytical spectrum is technically challenging. Techniques that perform well when analyzing individual targets do not automatically work well when targeting multiple genes. Conversely, techniques that enable global analysis of virtually all genes are not cost effective and accurate enough to be directly implemented for targeted analyses, particularly when aiming for diagnostics. In this chapter some approaches for targeted multiplex analysis will be described.

Another important area of investigation is to study where in a complex tissue a particular gene or gene variant is expressed. Here, existing *in situ* hybridization techniques do not readily resolve single-nucleotide variants, or even splice variants of genes. New techniques are needed to improve the resolution of such single-cell analyses in the intact context of tissue samples.

In this chapter, we will describe assays based on padlock and selector probes, both employing DNA ligase-assisted/DNA circularization reactions. We will describe some unique properties of these circularization approaches and the advantages they bring. The assays have been used for large-scale genotyping and amplification of targeted parts of the genome, DNA copy-number analysis, infectious diagnostics, single-molecule detection, and *in situ* analysis with single-nucleotide resolution. The different applications and techniques will be described in the following sections.

9.2 PADLOCK AND SELECTOR PROBES

Both padlock and selector probe assays are based on highly specific DNA ligase-catalyzed DNA circularization reactions (Figs 9.1 and 9.2). Even if the overall architectures of the assays are similar, the outcome of the circularization reaction is quite different, since ligation of padlock probes is templated by the target sequence, while it is the other way around in the selector probe assay, as will be described below. The circularization can be performed by ligation only, after a gap-fill reaction, or after an invasive cleavage reaction (Figs 9.1 and 9.2).

9.2.1 Padlock Probes

The padlock probe concept was invented by Ulf Landegren in the early 1990s, as a development of his oligonucleotide ligation assay (Landegren *et al*., 1988), and was published in 1994 (Nilsson *et al*., 1994). A padlock probe consists of three segments: two target-complementary end sequences and a non-target-complementary element that links the end sequences together. The target-complementary segments are typically in the range of 15–22 nucleotides in length and the linking segment is typically longer than the combined length of the target-complementary segments. The linking segment can also comprise another polymer than DNA, e.g. poly-ethyleneglycole (PEG) (Nilsson *et al*., 1994). The padlock probe design is such that the ends of the probe hybridize head to tail on the single-stranded target molecule, forming a nicked DNA duplex that can be sealed by a DNA ligase (Fig. 9.1a). The linking segment

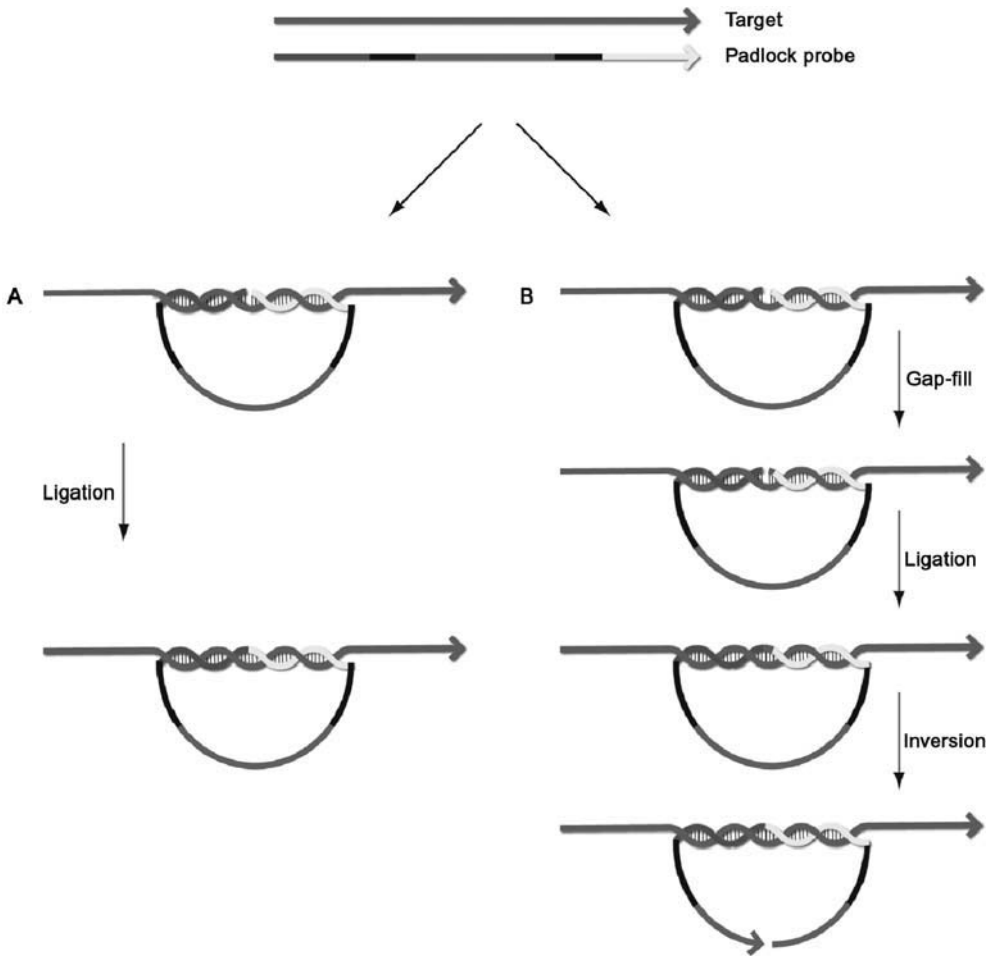


FIGURE 9.1 The architecture of the padlock and molecular inversion probe assays. Both assays employ probe molecules consisting of a 5′ target-complementary sequences (blue), a linking segment (black) including sequences for amplification and identification (red), and a 3′ target-complementary sequence (yellow). The ends bind head to tail when the probe hybridizes to a denatured single-stranded target molecule (gray), forming a structure that can be circularized through the action of one or several enzymatic reactions. **A.** In the original padlock probe assay, the ends are joined by a DNA ligase. **B.** In the molecular inversion probe assay, a single-nucleotide gap is filled by a DNA polymerase to create a suitable structure for the DNA ligase-assisted probe circularization. After an exonuclease clean-up of the sample, degrading linear target and probe molecules, thereby enriching for the circularized molecules, the probe molecules are opened by degradation of uracil residues in the linking segment. The linearization serves to release the probe molecules from the encircled target molecules. This improves the amplification efficiency, and may also reduce probe-dependent amplification artefacts. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

typically harbors sequences for amplification (primer motifs) and identification of the amplification product (tag sequences). Amplification of padlock probes can be done using PCR, rolling circle amplification (RCA), and exponential variants of RCA (hRCA and C2CA), or combinations thereof, which will be described in the following paragraphs.

9.2.2 Gap-Fill and Molecular Inversion Probes

In the gap-fill version of the padlock probe, the ends are designed to hybridize a distance apart on the target. The gap between the ends is then filled in by a polymerase prior to circularization by ligation (Fig. 9.2c; Akhras *et al*., 2007b; Hardenbol *et al*., 2003; Porreca *et al*., 2007). The gap-fill padlock probes will be described in more detail in section 9.6, where techniques for multiplexed targeted resequencing are discussed. In the molecular inversion probe (MIP) variant of gap-fill padlock probes, the gap consists of a single nucleotide at a single-nucleotide polymorphism (SNP) position. The SNP is queried in four separate tubes, each containing one of the dNTPs, a DNA polymerase, and a DNA ligase (Hardenbol *et al*., 2003). Circularization is used in conjunction with an enzymatic cleavage of the linking segment, resulting in the reacted probe being converted into a linear molecule, with the order of the segments inverted (Fig. 9.1b). Unreacted probes, at the same time, are cleaved in two parts, reducing the risk of background signal in a subsequent amplification reaction. The molecular inversion

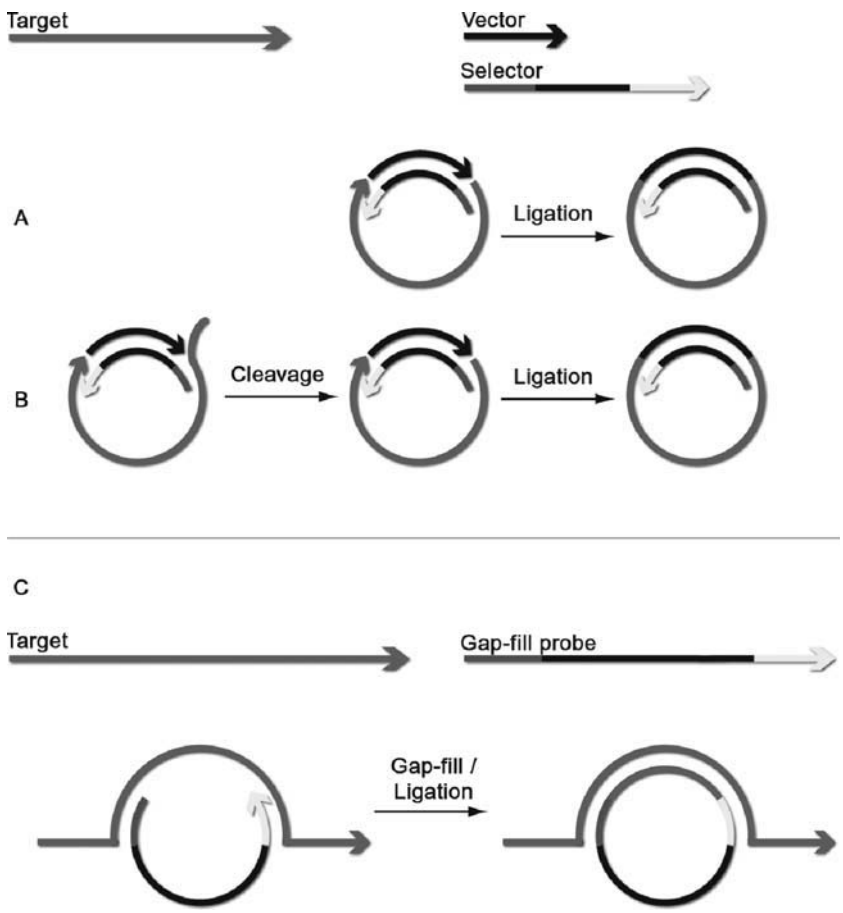


FIGURE 9.2 The architecture of the selector and gap-fill padlock probe assays. Similarly to padlock probes, selector probes consist of a 5′ target-complementary sequence (blue), a linking segment containing sequences for universal amplification reagents (black), and a 3′ target-complementary sequence (yellow). The linking segment is made double stranded by hybridization to a complementary vector oligonucleotide (black). The partially double-stranded selector probe acts as a template for a circularization of a specific genomic DNA fragment through two alternative routes. **A.** In the simplest format the selector probe hybridizes to the ends of a denatured single-stranded restriction fragment containing the target sequence, forming a structure that can be circularized by two ligation reactions. **B.** Alternatively, the 5′ end of a restriction fragment is cleaved in a structure-specific endonucleolytic ''invasive'' reaction, to minimize the non-target sequence content of a restriction fragment. From this reaction, a structure is formed that is identical to the one in panel A that is circularized by two ligation reactions. **C.** The gap-fill padlock probe ligation reaction acts through a completely different mechanism, where the target strand is first replicated to fill the gap between the two target-complementary probe sequences of the gap-fill probe, before a structure is formed that can be circularized through the action of a DNA ligase. This approach does not require restriction digestion, which simplifies assay design. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

probe has been developed to construct highly multiplexed targeted genotyping assays (see section 9.3).

9.2.3 Selector Probes

A selector probe resembles a padlock probe in that it consists of three segments, and that the end segments are complementary to a genomic target sequence. However, these ends are designed to hybridize to each of the ends of a single-stranded restriction fragment. Also, a general vector oligonucleotide is hybridized to the internal segment of the selector probe, allowing two ligation events to close the construct into a circle (Fig. 9.2a). The circle is thus composed of the targeted genomic sequence and the vector sequence. Alternatively, the 5′ part of the restriction fragment can be trimmed before ligation. This is performed by hybridizing the 5′ end of the selector probe to an internal part of the target fragment. Thereby a structure is formed that can be cut by a structure-specific endonuclease, e.g. Taq polymerase, as described by Lyamichev and coworkers (1993), whereby the protruding 5′ part of the fragment is released (Fig. 9.2b).

9.3 APPLICATION OF PADLOCK AND MOLECULAR INVERSION PROBES FOR GENOTYPING

Padlock probes are well suited for targeted multiplex analysis of genetic variation. The probes require dual recognition of the target sequence in order to become circularized,

providing sufficient specificity to probe unique sequences in the genome (Lizardi *et al*., 1998; Antson *et al*., 2000). Robust discrimination of single-nucleotide differences is achieved due to the stringent substrate requirement of the DNA ligase used in the circularization reaction (Landegren *et al*., 1988). Finally, the uni-molecular circularization approach renders padlock probe-based multiplexed assays less prone to problems with cross-reactivity than bi-molecular assays such as multiplexed PCR. A large number of circularization probes can therefore be combined in multiplexed assay without problems of interference (Banér *et al*., 2003; Hardenbol *et al*., 2003). The probes are equipped with sequence motifs in the linking segment that are used to sort the amplification product from a multiplex padlock probe assay on a microarray comprising tag sequences that are designed to be maximally divergent in sequence. The specificity of such tag-array hybridizations is typically good enough for genotyping, but when there are differences in the concentration of different targets, such as in transcription profiling, there are limitations in the dynamic range of this analysis platform. This limitation has recently been overcome by introducing a ligation-based tag-recognition coupled to a localized RCA signal amplification (Ericsson *et al*., 2008).

The padlock probe in its original design has been used in moderately multiplexed assays for infectious disease diagnostics (see section 9.4), for *in situ* genotyping and detection assays (see section 9.8), for genotyping (Banér *et al*., 2003; van Eijk, *et al*., 2004), and to determine which T-cell receptor is expressed in clonally expanded T-cell populations by probing for the expression of the entire repertoire of Vb genes (Banér *et al*., 2005).

9.3.1 Genotyping Using Padlock Probes and Hyper-Branched RCA

The somewhat surprising finding that very small DNA circles can also template rolling-circle DNA synthesis was shown independently by Fire and Xu (1995) and by Kool and coworkers (Liu *et al*., 1996). Here, the DNA polymerase replicates a circular DNA template over and over again, building a long single-stranded tandem repeated product. Padlock probes were early combined with this RCA mechanism (Banér *et al*., 1998; Lizardi *et al*., 1998; Zhang *et al*., 1998), which has been key to many of the applications discussed later in this chapter. Φ29 DNA polymerase has been shown to be particularly well suited for RCA, perhaps due to the combination of strand displacement activity, high processivity, and both single-stranded and double-stranded 3′ exonucleolytic (proof-reading) activity (Banér *et al*., 1998; Lizardi *et al*., 1998; Salas, 2004). One of the first applications of padlock probe RCA utilized an exponential version of the technique called hyper-branched RCA (hRCA) (Banér *et al*., 1998; Lizardi *et al*., 1998; Zhang *et al*., 1998). Here, the RCA is performed with a forward and a reverse primer, just like in PCR. But in this case every primer binding event gives rise to multiple copies of the tandem-repeated RCA product and the process is isothermal since the product strands are separated through the strand displacement activity of the DNA polymerase. The same mechanism, coupled with random hexamer primer libraries, is utilized in the multiple strand displacement amplification (MDA) used to uniformly amplify the content of DNA samples (Dean *et al*., 2002). The amplification is exponential, allowing for rapid detection of SNPs. Genotyping assays have been developed based on different incarnations of this technique, using PCR products or genomic DNA as template, and calling the alleles either in separate tubes using intercalating dyes, or in one tube using primers that emit fluorescence when consumed in the reaction (Thomas *et al*., 1999; Faruqi *et al*., 2001; Pickering *et al*., 2002; Alsmadi *et al*., 2003).

9.3.2 Massively Multiplex Targeted Genotyping Using Molecular Inversion Probes

The most significant application of circularization probes to date is perhaps the massively multiplexed targeted genotyping that has been achieved with the molecular inversion probes (MIP). The method was initially developed at Stanford by Hardenbol and coworkers (2003); and then to 10,000-plex assays at the company ParAllele (Hardenbol, *et al*., 2005), which was later acquired by Affymetrix (http://www.affymetrix.com). The assay is now commercialized by Affymetrix as their targeted genotyping product. The technique has been used in a variety of applications as briefly described below. A more detailed review of the MIP assay has been compiled by Absalan and Ronaghi (2007). The assay is available as kits of probes and tag microarrays to obtain genotyping results from 1,000 to 25,000 targeted SNPs. The technique has been used in the human and bovine HapMap projects (Gibbs *et al*., 2003; Altshuler *et al*., 2005; Frazer *et al*., 2007; Khatkar *et al*., 2007) to establish a map of the most common haplotype blocks in these populations, and similarly to investigate the extent of linkage disequilibrium in populations of wild mice (Laurie *et al*., 2007). The MIP assay has also been used in association studies to map genes that increase susceptibility to Type I diabetes (Smyth *et al*., 2006), prostate cancer (Zheng *et al*., 2006), and colorectal cancer (Zanke *et al*., 2007). The assay is also suitable for analysis of degraded DNA, such as found in paraffin-embedded formalin-fixed tissue samples, due to the short footprint of the probes and the fact that the hybridization and ligation reactions are probably more tolerant to DNA damages than techniques that require replication of the target strand, such as PCR (Wang *et al*., 2005).

These authors also demonstrated the technique's utility to detect copy-number variations in a sample.

9.4 APPLICATION OF PADLOCK PROBES FOR INFECTIOUS DISEASE DIAGNOSTICS

A common task in infectious disease diagnostics is to identify an infectious agent from a panel of candidate pathogens (see also Chapter 30). Padlock probe assays are well suited for this type of analyses due to their multiplexing ability and specificity. Compared to culture-based diagnostics, these assays accelerate diagnosis from weeks to hours (Kong *et al*., 2008). Padlock probes and similar technologies have been successfully applied for detection of a wide variety of pathogens in a variety of assay formats. The assays are directed to RNA or DNA sequences, and use real-time fluorescence, microarray, or luminescence readouts. Some yield sequence information, while others quantify pathogens. Padlock probes have also been applied for *in situ* analysis of infectious organisms which is described in section 9.8.4. Interestingly, many of the pathogen diagnostic applications with padlocks come from the field of agriculture and veterinary medicine, but there are also examples of human fungal and bacterial pathogen detection and subtyping assays (Tong *et al*., 2007; Kong *et al*., 2008).

9.4.1 Padlock Probe-Based Infectious Disease Diagnostics

The potential of padlock probe assays for infectious disease diagnostics was demonstrated by the detection and subtyping of three important viral livestock pathogens using multiplexed padlock ligation to cDNA, employing tag microarrays to analyze the RCA and PCR-amplified padlock probes (Banér *et al*., 2007). The padlock probes were designed against several highly conserved regions in each virus. This redundant testing with independent probes renders the assay tolerant to novel mutations. The same principle of multiplexed padlock ligation to cDNA was applied by Gyarmati and coworkers (2008), who developed a padlock probe assay to genotype all 16 hemagglutinin and all nine neuraminidase gene subtypes of avian influenza virus. All known isolates of the avian influenza virus were correctly subtyped in single reactions, reducing assay times from weeks, for the state-of-the-art culture method, to hours.

Van Doorn and coworkers (2007) combined multiplex padlock probe assays with qPCR readout in a novel assay format to enable an increased dynamic range compared to microarray readout. A pool of padlock probes directed to 11 plant pathogens were reacted with the samples in a single reaction and then reversibly bound to beads. The target sequences were then washed off from the beads, the probes were eluted, and non-reacted probes were removed by exonucleases. The remaining circular padlock probes were loaded onto an OpenArray qPCR platform that allows for 3,072 PCRs with up to 64 different primer sets in 33 nl wells (http://www.biotrove.com). Each pathogen-specific padlock probe was equipped with a specific PCR-primer pair motif and was analyzed in separate qPCR reactions. RNA viruses have also been detected with padlock probes without reverse transcription into cDNA (Millard *et al*., 2006).

Diagnostic applications outside a lab environment are enabled by biosensor devices that perform padlock probe ligation, isothermal amplification, and detection reactions with fluorescence readout (McCarthy *et al*., 2006, 2007). Other biosensor approaches that also have been applied for detection of pathogens are described in more detail in section 9.7.

9.4.2 Gap-Fill and MIP-Based Infectious Disease Diagnostics

Gap-fill probes have also been applied in pathogen detection and subtyping assay. Novais and coworkers (2008) developed an assay for sensitive detection of *Mycobacterium tuberculosis*. They amplified circularized MIPs by PCR and sequenced the product using Pyrosequencing. They further improved sensitivity to a limit of detection of 200 genomes by adding all dNTPs to the bioluminescence reaction performed in the Pyrosequencing instrument, instead of adding the four nucleotides sequentially, which is required to obtain sequence information. Akhras and coworkers (2007b) developed a gap-fill padlock probe assay, known as connector inversion probe (CIP) assay, similar to the one described in more detail in section 9.6.2. DNA circles are produced in an extensive gap-fill polymerization and ligation reaction, allowing for amplification and sequencing of DNA between the two target-complementary segments of a padlock probe. This technology was applied for antibiotic resistance screening in *Neisseria gonorrhoeae* and genotyping of human papillomavirus (HPV) samples. The same group also subtyped human papillomavirus sequences with 24 PCR amplified MIPs, sorted on a tag microarray (Akhras *et al*., 2007a).

9.5 TARGETED MULTIPLEX CNV ANALYSIS USING SELECTOR PROBES

Structural variations including deletions, insertions, and inversions have recently received a lot of attention in the field of genomics. The impact on mammalian phenotypic variation due to copy-number variation (CNV) appears to be more extensive than expected (Eichler *et al*., 2007; Kidd *et al*., 2008). New methods are required to detect

and analyze this variation which is quantitative in nature, compared to SNPs that basically only require qualitative binary information to score.

9.5.1 The Multiple Ligation-Dependent Genome Amplification Assay

Selector probes have recently been applied for interrogation of CNVs, using the multiplex ligation-dependent genome amplification (MLGA) assay (Isaksson *et al*., 2007). MLGA uses selector probes to equip specific genomic fragments with a common primer pair. The selected target fragments can be *in silico* designed to differ in length and thereby facilitate a capillary gel electrophoresis readout. The relative peak area of the amplification product represents target abundance and by normalizing peaks to reference loci, CNV can be analyzed between a sample and a reference (Fig. 9.3). The MLGA assay can be used to interrogate up to 30–40 fragments depending on the electrophoretic resolution. They can be used to resolve boundaries of CNVs by spreading the targets for the assay over a CNV candidate region. In a study by Salmon Hillbertz and coworkers (2007), selector probes were designed to cover a 750 kb region on canine chromosome 18 associated with the Ridge phenotype in Rhodesian Ridgeback dogs. Its name describes its most unique characteristic, a clearly defined symmetrical ridge running the length of its back, formed by hair growing in the opposite direction from the rest of its coat. In this duplication boundary mapping approach, a first set of MLGA probes were designed with 100 kb interval covering a region of 2 Mb. The results confirmed the CNV, and to further map the start and end position of the region, some additional MLGA probes were designed between duplicated and non-duplicated regions. The borders of the potential tandem repeat were then defined precisely enough to amplify the breakpoint by PCR. The 133.4 kb duplication involved five genes, whereof three FGF genes are strong candidates for causing the Ridge phenotype (Salmon Hillbertz *et al*., 2007).

9.6 HIGH-THROUGHPUT TARGETED SEQUENCING USING SELECTORS AND GAP-FILL PADLOCK PROBES

The last few years have seen an accelerating development within methodology and instrumentation for very-high-throughput sequencing (Mardis, 2008). This development will enable new types of genetic analyses, such as whole genome sequencing, deep sequencing of tumor samples, global expression analyses, sequencing of material from chromatin immune-precipitation experiments, and more.

One application that this new technology is well suited for is the resequencing of many selected parts of a genome, such as all exons from a large set of genes. This requires that the targeted parts of the genome are somehow enriched in the sample. Traditionally, this enrichment has been performed using PCR, but this is poorly suited for multiplexing, meaning that a very large number of separate reactions would be required. The ideal method for preparing samples as templates for resequencing would be one that can be performed in a single reaction, which enriches for all the regions of interest (and nothing else), and which yields uniform enrichment for all targeted regions. Attempts to achieve this have been demonstrated by using capture and release of selected parts of the genome on custom microarrays, followed by an adaptor-linker PCR (Albert *et al*., 2007; Okou *et al*., 2007). In this section, we will describe probe-based methods combining hybridization with enzymatic activity to select the desired parts of the genome.

9.6.1 Targeted Sequencing Using Selector Probes

Selector probes can be employed to amplify a large set of arbitrary restriction fragments in multiplex. These fragments can be selected so that certain desired sequences, e.g. all exons from a set of genes, are amplified. By using the endonucleolytic cleavage, as described in section 9.2.3 above (Fig. 9.2b), the restriction fragments can be trimmed to decrease the amount of undesired sequence that is also amplified. Dahl and coworkers (2007) showed that this method, coupled with massively parallel Pyrosequencing (Margulies *et al*., 2005), can be applied to the amplification of all coding sequence and flanking regions from ten genes implicated in colorectal cancer. Approximately 90% of all sequenced reads were from the targeted regions, and 93% of the desired sequence was sequenced at least once. The quality of sequence was sufficient to identify most previously known mutations in the analyzed cell-lines. In this work, the number of reads originating from the different amplified fragments was rather uneven. To maximize the portion of the region of interest that can be analyzed, this has to be further improved. Another drawback of this technique is the requirement of suitable restriction enzymes and restriction sites. In the described work, five parallel restriction reactions were required in order to generate suitable fragments for all regions. Also, these fragments cannot be exactly matched to desired regions, so there will always be some co-amplification of undesired sequence.

Fredriksson and coworkers (2007) modified the selector method to select fragments generated by a multiplex PCR, rather than the genomic DNA itself. One hundred and seventy primer pairs were designed to amplify all coding sequence of ten genes. Such a reaction will generate many undesired amplicons that are not products of a cognate primer pair. By designing probes (called gene collectors) that guide the circularization of products with cognate primers at the ends, all the intended fragments can be collected and further amplified

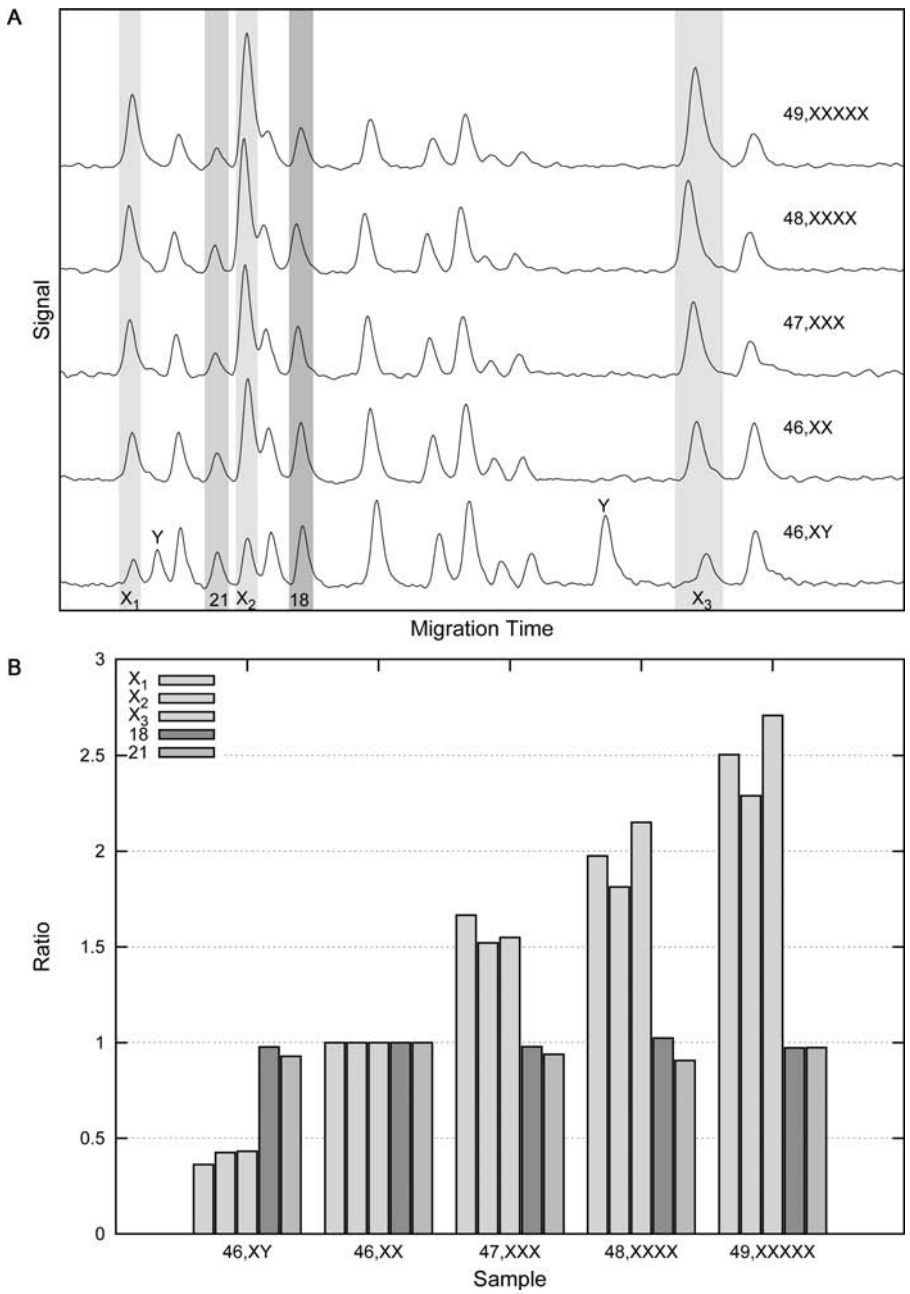


FIGURE 9.3 The MLGA assay for copy-number analysis. In the MLGA assay, selector probes are designed to generate amplification products of different size from loci in the genome to be probed for copy-number variation. The products from such multiplex amplifications are separated through electrophoresis and the relative amount of each amplification product is measured. Since the relative representation of all amplicons is very reproducible it is possible to measure relative copy-number deviations between samples. **A.** Elution diagrams from an Agilent Bioanalyzer instrument for five different samples from human cell-lines harboring different copy numbers of the X-chromosome. The products from loci located on the X-chromosome are highlighted green, and the products from two probes targeting chromosomes 18 and 21 are highlighted blue and red, respectively. Note the presence of two products from the Y-chromosome in the male (XY) sample only and the steady relative increase of the X-chromosomal peaks in the series of one (bottom trace) to five (top trace) X chromosomes. **B.** Analysis of the data shown in panel A to determine relative copy-number deviations. The areas of the highlighted peaks in panel A were normalized to the peak areas of the normal female sample (46, XX). Two copies of a locus will, thus, yield a ratio of 1, and three copies 1.5, etc. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

by hRCA. This has been shown to yield 90% coverage and a more even distribution than that achieved using the selector method. However, the specificity is lower, with ~60% of the amplified material representing the targeted regions. On the other hand, as this method does not require restriction enzymes, an assay can be designed more precisely to target only coding sequence of interest. Also, any collector probes from different designs may be combined in a reaction,

whereas for selector probes this can only be done with probes designed for the same restriction reaction.

9.6.2 Targeted Sequencing Using Gap-Fill Padlock Probes

In a related approach, gap-fill padlock probes were used having the ends hybridize a distance apart, spanning an entire exon (Porreca *et al*., 2007). A polymerase is then used to extend the spanned region, and a ligase joins the extended sequence with the far end of the probe (Fig. 9.2c). As with the selector probes, a set of circular DNA molecules, all containing sequence of interest along with a general, synthetic sequence, is generated. Again, this can be amplified in multiplex, in this case using RCA followed by PCR, and finally hRCA for the final library construction. An impressive 55,000 probes were deployed to target 6.7 Mbp of protein coding sequences in the genome. The probes were synthesized through enzymatic amplification of a probe library synthesized on a microarray (Tian *et al*., 2004). The product was then analyzed on the Illumina Genome Analyzer, and results showed that about 10,000 of the 55,000 targeted sequences were amplified and sequenced at least once.

In conclusion, there are several proposals for how to solve the issue of template preparation for targeted resequencing using gap-fill padlock probes. In time, it will likely become apparent which of these methods is most suitable, and most probably it will turn out that all these methods have their own niche to fill, based on their different characteristics on cost, degree of enrichment, and uniformity of amplification.

9.7 BIOSENSOR APPROACHES BASED ON ROLLING-CIRCLE AMPLIFIED PADLOCK PROBES

The combination of the strictly target-dependent padlock probe circularization reaction and the strictly circle-dependent RCA reaction offers opportunities for novel biosensor approaches enabling detection of even single molecules in multiplexed assays. Some of these approaches will be described in this section.

9.7.1 Homogeneous Amplified Single-Molecule Detection

Most methods used for molecular analyses measure an average of a molecular population, for instance based on fluorescence intensity of a reporter molecule. To reach ultimate precision and quantification, however, molecules should preferably be analyzed individually. Current techniques for single-molecule detection (SMD), especially ones with fluorescence readout of the results, demand highly advanced equipment and are often slow and not very robust.

An alternative approach is based on amplified SMD, where individual molecules are clonally amplified and detected. Detection of single templates with PCR was first demonstrated by Li and coworkers (1988), where single human sperm cells were added to individual reactions to allow detection of recombination events in the male germline. A decade later, Vogelstein and Kinzler (1999) used single-molecule PCRs to achieve digital quantification of somatic mutations. Instead of recording the degree of amplification, the reactions were treated as binary calls (''yes'' or ''no'' answer), and the technique was therefore called ''digital PCR''. Single DNA molecules were achieved by diluting the samples in microtiter plates to the extent that only every second well contained a template. The method has been further developed by replacing the well compartments with oil/water micelle microreactors containing microbeads. With this strategy, the locally amplified products are fluorescently labeled according to their genotype and analyzed with flow cytometry (Dressman *et al*., 2003). Emulsion PCR has, however, mainly found its application as an amplification method to create single molecule clones as templates for several sequencing strategies.

Another way of transforming a population of DNA molecules to amplified single molecules, and thus preserving its digital nature, is by using RCA, which is an amplification technique that creates inherently localized DNA clusters. Interestingly, the long DNA threads produced in the RCA collapse into a random coil with a diameter of about 700 nm (Melin *et al*., 2007). A local enrichment of fluorescence of up to a factor of one thousand can be achieved in a homogeneous format when hybridizing fluorescence-labeled oligonucleotide probes to the repeats in the rolling-circle product (RCP), resulting in bright fluorescent objects that are easily visualized using standard fluorescence microscopy (Blab *et al*., 2004). The high signal-to-noise ratio allows robust and rapid detection and counting of amplified molecules by pumping the RCP solution through a microfluidic channel mounted in a confocal microscope operating in a scanning mode across the width of the flow channel (Fig. 9.4; Jarvius *et al*., 2006).

A padlock probe is usually equipped with a tag, correlating the reacted probe to its target. Hence, different RCP populations can be labeled with different fluorescence and spectrally distinguished. In fact, amplification with RCA is very well suited for multiplexing since the RCPs interact very little with each other, and even fully complementary RCPs aggregate to a negligible degree (Melin *et al*., 2007). The approach of counting individual RCPs homogeneously has been used to selectively detect and quantify the bacteria *V. cholerae* and *V. fisheri*, in a padlock probe-based assay (Jarvius *et al*., 2006).

9.7.2 A Magnetic Biosensor Approach

Magnetic nanoparticles, commonly referred to as magnetic beads, are applied in many research fields including

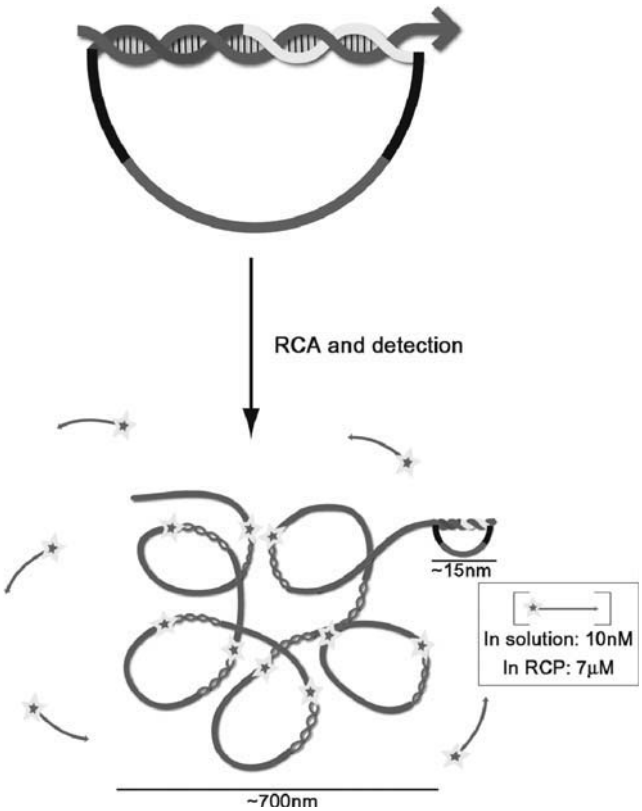


FIGURE 9.4 Padlock probe and RCA-based single-molecule detection. A padlock probe is circularized in a strictly target-dependent ligation reaction. Then a long tandem-repeated DNA product is formed in a strictly circle-dependent RCA reaction, thus forming a rolling-circle product (RCP) in a strictly target-dependent manner. The RCA product is labeled with fluorescence tagged short oligonucleotide probes (red with yellow star) that hybridize to a tag sequence in the product (red). The long product strand collapses into a submicron-sized object due to random coiling. The local concentration of fluorescence tags in a product comprising 1,000 copies of the circle is about 7 μM, compared to the 10 nM in the surrounding solution that is sufficient to saturate the RCPs with fluorescence tags. This difference in concentration makes it possible to detect single RCPs in solution without the need for washing away excess fluorescence tags. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

biotechnology, biomedicine, and drug discovery (Saiyed *et al*., 2003). Several bioassays have been developed based on the Brownian relaxation biosensor scheme. Here, the susceptibility of magnetic beads to reorient in an AC field is measured. The smaller the bead diameter, the more rapid is its response to a shift in the electric field. This response has a characteristic critical frequency, called the Brownian relaxation frequency. In these biosensor approaches, small magnetic beads are equipped with probes and upon interaction with a target a slight decrease in relaxation frequency can be observed as an effect of the increased hydrodynamic radius of the magnetic bead. The frequency shift can be recorded with a magnetometer, a superconducting quantum interference device (SQUID). In a study by Strömberg and coworkers (2008), the Brownian relaxation principle was used to detect padlock probe-generated RCPs. When the magnetic beads bind RCPs, a very dramatic shift in frequency occurs due to their large size, and the multivalent binding interactions between the beads and the RCPs. This effect can be easily observed as a loss of signal in the free-bead frequency domain, and the response can be quantified by measuring the decrease in amplitude which correlates well with the RCP concentration. This, in turn, corresponds to the number of reacted padlock probes and therefore the target concentration. With the magnetometer used in the study, the limit of detection was in the low picomolar range. By constructing a miniaturized magnetic biosensor with size closer to the size of the beads, it should be possible to increase the measurement sensitivity even further, thereby allowing for lower concentrations of RCPs and probe-tagged beads.

9.8 *IN SITU* GENOTYPING USING PADLOCK PROBES

Padlock probes were early recognized for being useful for analysis of DNA sequences *in situ* due to the strong link that is formed between the circularized probe and its target strand (Nilsson *et al*., 1994), opening up possibilities to observe genetic heterogeneity within cell populations, and also to study the localization of target molecules within individual cells. It was also shown that robust genotyping of repeated centromeric sequences differing in a single nucleotide position could be done on human metaphase chromosomes, revealing an unexpected distribution of these sequence variants among human chromosome 13 and 21 (Nilsson *et al*., 1997). The technique was developed further by using padlock probes synthesized by PCR to enable synthesis of longer and thus more strongly labeled probes, and the number of centromeric repeats was extended to include chromosomes 7, 12, and 15 (Antson *et al*., 2000, 2003). In these studies, the probes were directly labeled with haptens or fluorophores. With this approach, however, it was not feasible to detect single DNA target molecules since it was impossible to distinguish specifically bound probe molecules from probes non-specifically adsorbed to cells and the surface of the microscopy slide. This highlighted the need for a signal amplification scheme that is strictly depending on the molecular recognition reaction.

9.8.1 Single-Molecule Detection and Genotyping *In Situ*

Similar to the single-molecule detection approach with RCA described in section 9.7.1, individual rolling circle products can also be generated and visualized *in situ* in cell and tissue samples in a strictly circularization-dependent reaction. This was first demonstrated by Lizardi and

coworkers (1998), with the detection of nuclear sequences in cytological halo preparations. There are indications that the RCA reaction is slowed down when it occurs under topological inhibition, as when the padlock probe is not able to free itself from the target strand (Banér *et al*., 1998). In response to this, Christian and coworkers (2001) presented an *in situ* target denaturation scheme employing enzymatic preparation of single-stranded DNA at the location of the target sequence. In this method, DNA is digested 5′ upstream of the target sequence. This is followed by removal of the non-target strand using an exonuclease, leaving the target sequence single stranded and available for hybridization. They combined this denaturation with padlock probe detection of the target sequence followed by RCA. Since the enzymatic denaturation method creates a nearby free DNA end, the DNA polymerase is able to push off the padlock probe from the target strand when the topological constraint becomes too high. The drawback with the presented denaturation procedure is that as the padlock probe falls off the target strand, the link between the target sequence and the RCA product is lost, increasing the risk of losing the signal from the slide or creating false positive signals due to drifting of RCA products into nearby cells. A variant of this denaturation process was later described in a robust protocol for SNP genotyping of mitochondrial DNA *in situ* (Figs 9.5 and 9.6a; Larsson *et al*., 2004). Here, instead of digesting the DNA upstream of the target sequence, the cut takes place downstream of the target. This means that after exonucleolysis to make the target single stranded, and padlock circularization on the target sequence, the nearby 3′ end of the target strand itself can be used to prime the following RCA reaction. In this way, not only is the topological inhibition avoided, but also a covalent link is created between the target molecule and the positive detection signal, which minimizes the risk of signal misplacement.

Larsson and coworkers (2004) also showed that this method can be used for relative target quantification. By counting differentially labeled RCA products in individual cells, the proportion of mutant vs. wild-type mitochondrial (mt) DNA molecules differing in a single nucleotide could be determined. This ability to quantify different target molecules was further demonstrated in a more automated way, employing image analysis tools to perform digital counting of RCA products from detected mtDNA molecules in cultured cells (Jahangir Tafrechi *et al*., 2007). Dedicated software for RCA product identification and counting in digital images in a semi-automated fashion, BlobFinder, is now freely available for users (Allalou and Wählby, http://www.cb.uu.se/~amin/BlobFinder).

9.8.2 Nuclear DNA Detection with Padlock Probes *In Situ*

Following the early reports that padlock probes and RCA *in situ* could be used for detection of specific sequences in genomic DNA and interphase nuclei (Lizardi *et al*., 1998; Christian *et al*., 2001), further demonstrations of the technique applied for this purpose were not published for a considerable time. This was probably due to the fact that the efficiency of target detection with this technique was quite low, making it difficult for the regular researchers to apply it to their targets of interest. In fact, even when using the target-primed RCA for amplification of padlock probes, detection of mtDNA targets was estimated at about 10% of available targets (Larsson *et al*., 2004). Considering that nuclear DNA exhibits a much tighter packing and more complex organization than mtDNA, it is not difficult to imagine the difficulties in finding proper assay conditions for detecting genomic DNA in low copy numbers. Instead, nuclear target sequences of high copy number have recently been successfully detected using target-primed RCA or similar techniques, providing useful steps towards the ultimate goal of efficient detection of nuclear single copy sequences. Li and coworkers (2005) presented a proof-of-principle assay for detection of double strand breaks in mammalian cells utilizing RCA primed by the target strand. In this method, double strand breaks are introduced in cells by co-transfecting them with yeast HO endonuclease and its recognition site. When induced, HO endonuclease will perform digestion at integrated recognition sites, and site-specific cleavage is then detected by *in situ* RCA of padlock probes directed against the known sequence next to the double strand breaks. The RCA assay was also combined with immunofluorescence, demonstrating the potential to perform studies of signaling pathways triggered by DNA double strand breaks by this method.

Padlock probes and target-primed RCA *in situ* were also used in the comet assay for measuring DNA damage and repair (Shaposhnikov *et al*., 2006). In the comet assay, cells are cast in agarose gels on objective slides, which are then subjected to gel electrophoresis. This causes cell nuclei to adopt a comet-like shape, with the amount of DNA forming the tail of the comet reflecting the amount of DNA damage in the nucleus. *In situ* RCA is a preferred alternative to FISH in these types of cell preparations, owing to the fact that all steps can be executed at low temperatures and that the background is very low since signals are generated in a strictly target-dependent manner. For target detection with padlock probes in these preparations, the initial approach for target-primed RCA, with restriction digestion and exonucleolysis, was used for denaturing DNA to allow for detection of the nuclear 26 bp Alu core sequence with sequence-specific probes and RCA (Fig. 9.6). This sequence is present in high numbers in each nucleus, which was reflected in the amount of signal obtained in genomic DNA with the *in situ* assay. Also, probes directed against mtDNA sequences were used, providing evidence that mtDNA disperses from the gels during preparation of the comets.

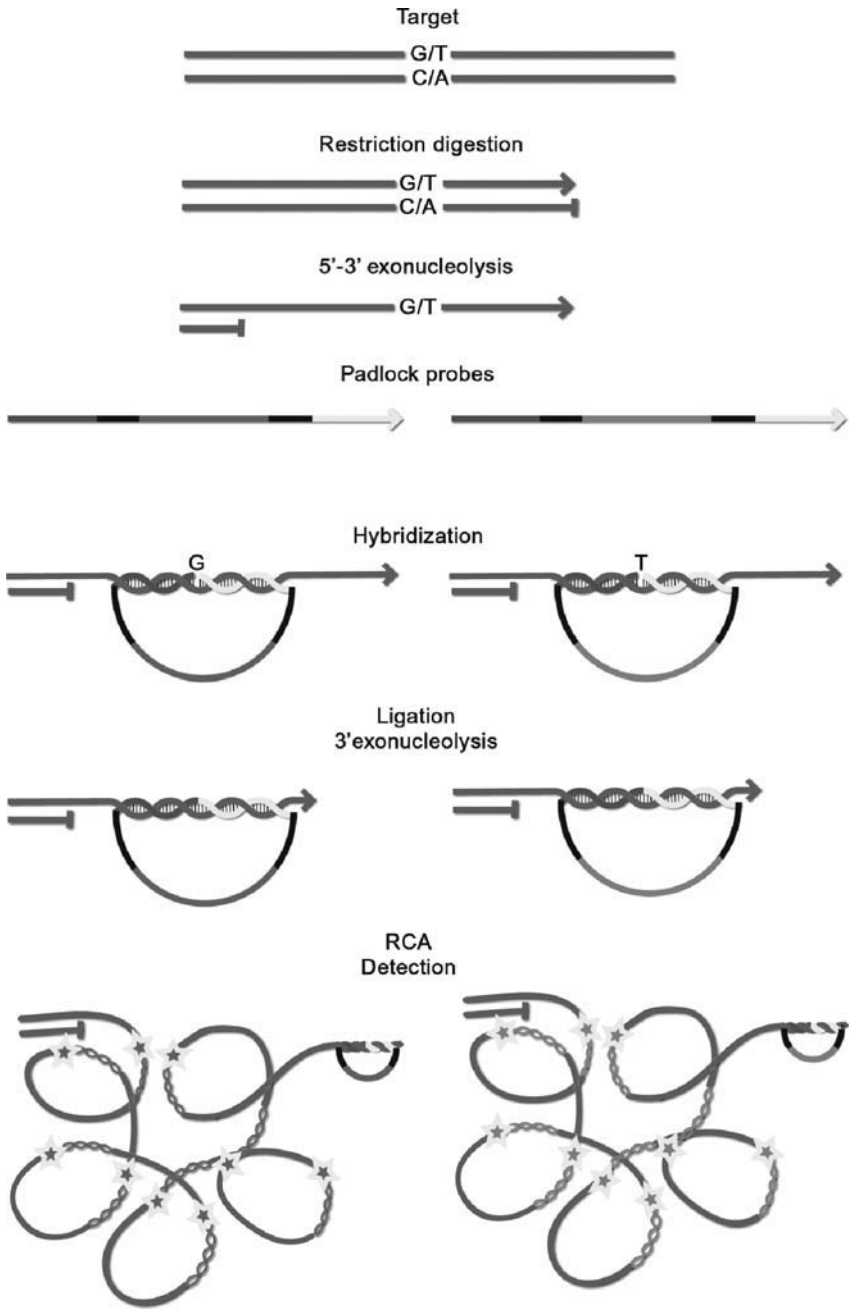


FIGURE 9.5 *In situ* genotyping using padlock probes and target-primed RCA. The first steps of the procedure serve to prepare the target strands for probing and priming through a series of enzymatic treatments of fixed and permeabilized cells. First, the target strands are digested at, or 3′ to, the target sequence, to generate an end close to the padlock probe binding site. Then the non-target strand is removed by 5′ exonucleolysis, making the target sequence single stranded and available for hybridization. A pair of allele-specific padlock probes is added, having target-complementary end sequences (blue and yellow) that are identical except for the 3′ position where the SNP is queried. The probes are equipped with two tag sequences (red and green) to differentiate the amplification products from the two allele-specific padlock probes in the final tag-hybridization and fluorescence detection steps. After a combined hybridization and ligation step, the probes are rolling-circle amplified. The DNA polymerase used has a very potent 3′ exonucleolytic activity that degrades any 3′ protruding end of the target strand, creating a target strand that primes RCA of the padlock probe. A RCP is created as an extension of the target strand, thereby ensuring that the signal will remain localized with the target strand. The products from the two allele-specific padlock probes are differentially labeled with two fluorescence-tagged detection probes (red and green with yellow stars), hybridizing to the tag-sequence motif in the non-target complementary part of the RCPs. The results of an *in situ* genotyping experiment are shown in Fig. 9.6. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

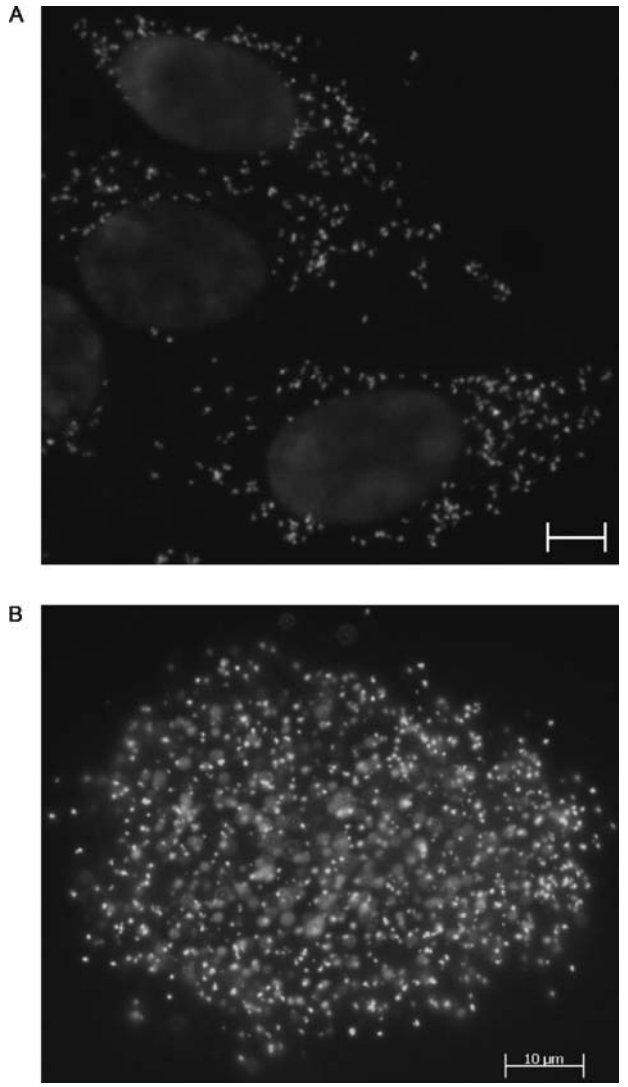


FIGURE 9.6 Detection of DNA *in situ.* The pictures show padlock probe and RCA-based detection of different DNA sequences *in situ*. Scale bars represent 10 μm and genomic DNA is stained blue using DAPI. **A.** Genotyping of the A3243G point mutation in human mtDNA following the scheme outlined in Fig. 9.5. Two human cell-lines were grown and fixed on a microscopy slide, one harboring the wild-type sequence of mtDNA, and the other close to 100% of the mutant version. A pair of allele-specific padlock probes was designed such that the mutant probe generates green RCPs in the detection step and the wild-type probe red RCPs. **B.** Detection of the 26 bp core Alu-repeat sequence in a comet preparation. In the comet assay, cells are immobilized in an agarose gel and subjected to alkali treatment to release the chromatin from the cell nucleus. Then an electric field is applied which will primarily move damaged DNA to the tail of the comet while intact chromatin remains in the head of the comet. The head is to the right in the picture and the tail extends to the left. The RCPs were labeled with green fluorescence. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

Standard assays for *in situ* detection of nuclear DNA sequences, such as FISH, are commonly executed on condensed chromosomes in metaphase spreads to be able to positively identify and assign targets to specific chromosome locations. Chromosomes in metaphase are even more condensed than interphase nuclei and, not surprisingly, early on unsuccessful attempts of using *in situ* RCA for detection of single copy target sequences on metaphase chromosomes were reported (Lizardi *et al*., 1998; Zhong *et al*., 2001). Using the target primed approach, Lohmann and coworkers (2007) demonstrated detection of repeated sequences on chromosomes Y and 6 on metaphase spreads. The detected targets were present in about 2,000 and 12–51 copies per chromosome, respectively. These authors reported labeling of all targeted Y chromosomes at an efficiency of 1–10% of available targets, while detection of the less repeated target on chromosome 6 was less successful with about 30% of spreads missing RCA products on one or two sister chromatids.

9.8.3 Padlock Probes for Detection of RNA *In Situ*

There is also a considerable interest in using padlock probes and RCA *in situ* to be able to perform expression profiling and transcript genotyping *in situ*. One obstacle for this has long been the decreased padlock probe ligation efficiency experienced when performing ligation on an RNA template (Nilsson *et al*., 2000, 2001). When considering the target-primed RCA, there has also been the question whether polymerases are able to initiate DNA synthesis using an RNA primer.

Stougaard and coworkers (2007) presented a way of circumventing the problems of padlock probe ligation on RNA molecules by introducing a new probe format, turtle probes, for the detection of non-polyadenylated RNA molecules in cultured cells. Turtle probes are similar to padlock probes in that they are also linear probes that are converted into circular molecules by ligation of the two probe ends. In contrast to padlock probes, though, ligation does not take place using the target strand as template. Instead, the turtle probe folds part of itself into a hairpin structure where an internal sequence of the probe serves as a template for joining of the probe ends. In this study, 5S rRNA, Epstein-Barr virus encoded RNA (EBER1) and the RNA template for human telomerase (hTR) were detected in formalin-fixed paraffin-embedded (FFPE) cells and tissue. The presented method answers the question of whether single-stranded RNA can prime RCA, but is so far limited to rather abundant non-polyadenylated RNA targets where a target site can be chosen very close to the 3′ end of the RNA. A more general approach for detection of RNA targets would thus be appreciated to be able to fully gain the advantages with using short amplified oligonucleotide probes compared to traditional FISH methods.

9.8.4 Padlock Probes for DNA Detection in Microbes

Interest has recently arisen to use padlock probes for detection of DNA sequences in bacterial cells. Maruyama and

coworkers (2005, 2006) first demonstrated how padlock probes could be used to visualize and count bacteria that had taken up free DNA from the environment in the form of plasmids. These authors also detected a single copy gene of the bacterial genome, demonstrating the possibilities of using the technique for detection of targets of lower abundance. The padlock probe technology is attractive for this type of work due to the fact that relatively short (<40 nt) DNA sequences are targeted, and because cells that are unavailable for analysis using standard methods, such as selective cultivation, can be analyzed.

Smolina and coworkers (2007, 2008) used padlock probes to detect specific bacteria by targeting single copy signature sites of different bacterial species. They used peptide nucleic acid (PNA) openers to locally denature double-stranded DNA and allow for padlock probe circularization. This approach avoids global DNA denaturation, which could be advantageous for some applications but also limits detection to specific sites in the genome, hence limiting the general applicability of this method.

9.9 CONCLUSIONS

The intramolecular DNA circularization assays have found, as described in this chapter, some important applications in molecular medicine. There is likely more to come since some of the techniques are yet not mature, but still very promising. The targeted multiplexed genome amplification approaches described in section 9.6, for example, have potential to become a standard pre-preparation step for high-throughput DNA sequencers – particularly in applications where the source of DNA is limiting, such as in tumor biopsies, and where really deep sequence coverage is desired.

The circularization techniques combined with RCA have some unique and promising properties for rapid diagnostics, such as outlined in section 9.7. Application areas may include infectious diagnostics and high-performance multiplexed quantitative expression analyses. Finally, the single-molecule *in situ* genotyping technique described in section 9.8 is unique and may find a broad application in biomedical research and diagnostics. One example of a relevant clinical pathology investigation is to find rare surviving malignant cells in tissue biopsies taken after therapy. Since the approach is multiplexible, it could also be applied in cell differentiation studies to determine the co-localization of several lineage-specific transcripts in developing embryos or organs, including tumor tissue, and thereby tracking the developmental history of individual cells in complex tissues.

To conclude, there is potential for many exciting developments of DNA circularization assays in addition to the ones mentioned above. All possibilities that come with this approach, such as inherent multiplexibility, single-molecule sensitivity, digital quantification, and localized detection, have not yet been fully explored.

REFERENCES

Absalan, F., and Ronaghi, M. (2007). Molecular inversion probe assay. *Methods Mol. Biol.* 396, 315–330.

Akhras, M.S., Thiyagarajan, S., Villablanca, A.C., Davis, R.W., Nyren, P., and Pourmand, N. (2007a). PathogenMip assay: a multiplex pathogen detection assay. *PLoS ONE* 2, e223.

Akhras, M.S., Unemo, M., Thiyagarajan, S., Nyren, P., Davis, R.W., Fire, A.Z., and Pourmand, N. (2007b). Connector inversion probe technology: a powerful one-primer multiplex DNA amplification system for numerous scientific applications. *PLoS ONE* 2, e915.

Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X., Richmond, T.A., Middle, C.M., Rodesch, M.J., Packard, C.J., Weinstock, G.M., and Gibbs, R.A. (2007). Direct selection of human genomic loci by microarray hybridization. *Nat. Methods* 4, 903–905.

Alsmadi, O.A., Bornarth, C.J., Song, W., Wisniewski, M., Du, J., Brockman, J.P., Faruqi, A.F., Hosono, S., Sun, Z., Du, Y., Wu, X., Egholm, M., Abarzua, P., Lasken, R.S., and Driscoll, M.D. (2003). High accuracy genotyping directly from genomic DNA using a rolling circle amplification based assay. *BMC Genomics* 4, 21.

Altshuler, D., Brooks, L.D., Chakravarti, A., Collins, F.S., Daly, M.J., and Donnelly, P. (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320.

Antson, D.O., Isaksson, A., Landegren, U., and Nilsson, M. (2000). PCR-generated padlock probes detect single nucleotide variation in genomic DNA. *Nucleic Acids Res.* 28, e58.

Antson, D.O., Mendel-Hartvig, M., Landegren, U., and Nilsson, M. (2003). PCR-generated padlock probes distinguish homologous chromosomes through quantitative fluorescence analysis. *Eur. J. Hum. Genet.* 11, 357–363.

Banér, J., Nilsson, M., Mendel-Hartvig, M., and Landegren, U. (1998). Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res.* 26, 5073–5078.

Banér, J., Isaksson, A., Waldenström, E., Jarvius, J., Landegren, U., and Nilsson, M. (2003). Parallel gene analysis with allele-specific padlock probes, and tag microarrays. *Nucleic Acids Res.* 31, e103.

Banér, J., Marits, P., Nilsson, M., Winqvist, O., and Landegren, U. (2005). Analysis of T-cell receptor Vβ gene repertoires after immune stimulation and in malignancy by use of padlock probes and microarrays. *Clin. Chem.* 51, 768–775.

Banér, J., Gyarmati, P., Yacoub, A., Hakhverdyan, M., Stenberg, J., Ericsson, O., Nilsson, M., Landegren, U., and Belak, S. (2007). Microarray-based molecular detection of foot-and-mouth disease, vesicular stomatitis and swine vesicular disease viruses, using padlock probes. *J. Vir. Methods* 143, 200–206.

Blab, G.A., Schmidt, T., and Nilsson, M. (2004). Sensitive and homogenous detection of single rolling-circle replication products. *Anal. Chem.* 76, 495–498.

Christian, A.T., Pattee, M.S., Attix, C.M., Reed, B.E., Sorensen, K.J., and Tucker, J.D. (2001). Detection of DNA point mutations and mRNA expression levels by rolling circle amplification in individual cells. *Proc. Natl. Acad. Sci. USA* 98, 14238–14243.

Dahl, F., Stenberg, J., Fredriksson, S., Welch, K., Zhang, M., Nilsson, M., Bicknell, D., Bodmer, W.F., Davis, R.W., and Ji, H. (2007). Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc. Natl. Acad. Sci. USA* 104, 9387–9392.

Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S.F., Egholm, M., and Lasken, R.S. (2002). Comprehensive human

genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* 99, 5261–5266.

Dressman, D., Yan, H., Traverso, G., Kinzler, K.W., and Vogelstein, B. (2003). Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. USA* 100, 8817–8822.

Eichler, E.E., Nickerson, D.A., Altshuler, D., Bowcock, A.M., Brooks, L.D., Carter, N.P., Church, D.M., Felsenfeld, A., Guyer, M., Lee, C., Lupski, J.R., Mullikin, J.C., Pritchard, J.K., Sebat, J., Sherry, S.T., Smith, D., Valle, D., and Waterston, R.H. (2007). Completing the map of human genetic variation. *Nature* 447, 161–165.

Ericsson, O., Jarvius, J., Schallmeiner, E., Howell, M., Nong, R., Reuter, H., Hahn, M., Stenberg, J., Nilsson, M., and Landegren, U. (2008). A dual-tag microarray platform for high-performance nucleic acid and protein analyses. *Nucleic Acids Res.* 36, e45.

Faruqi, A.F., Hosono, S., Driscoll, M.D., Dean, F.B., Alsmadi, O., Bandaru, R., Kumar, G., Grimwade, B., Zong, Q., Sun, Z., Du, Y., Kingsmore, S., Knott, T., and Lasken, R.S. (2001). High-throughput genotyping of single nucleotide polymorphisms with rolling circle amplification. *BMC Genomics* 2, 4.

Fire, A., and Xu, S.Q. (1995). Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci. USA* 92, 4641–4645.

Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M., Tsui, S.K., Xue, H., Wong, J.T., Galver, L.M., Fan, J.B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.F., Phillips, M.S., Roumy, S., Sallee, C., Verner, A., Hudson, T.J., Kwok, P.Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.C., Mak, W., Song, Y.Q., Tam, P.K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Schaffner, S.F., Sham, P.C., Varilly, P., Altshuler, D., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M. K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Tsunoda, T., Mullikin, J.C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C. D., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Gibbs, R.A., Belmont, J.W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Wheeler, D.A., Yakub, I., Gabriel, S.B., Onofrio, R.C., Richter, D.J., Ziaugra, L., Birren, B.W., Daly, M.J., Altshuler, D., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archeveque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., and Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861.

Fredriksson, S., Baner, J., Dahl, F., Chu, A., Ji, H., Welsh, K., and Davis, R.W. (2007). Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector. *Nucleic Acids Res.* 35, e47.

Gibbs, R.A., Belmont, J.W., Hardenbol, P., Willis, T.D., Yu, F., Yang, H., Ch'ang, L.Y., Huang, W., Liu, B., Shen, Y., Tam, P.K., Tsui, L.C., Waye, M.M., Wong, J.T., Zeng, C., Zhang, Q., Chee, M.S., Galver, L.M., Kruglyak, S., Murray, S.S., Oliphant, A.R., Montpetit, A., Hudson, T.J., Chagnon, F., Ferretti, V., Leboeuf, M., Phillips, M.S., Verner, A., Kwok, P.Y., Duan, S., Lind, D.L., Miller, R.D., Rice, J.P., Saccone, N.L., Taillon-Miller, P., Xiao, M., Nakamura, Y., Sekine, A., Sorimachi, K., Tanaka, T., Tanaka, Y., Tsunoda, T., Yoshino, E., Bentley, D.R., Deloukas, P., Hunt, S., Powell, D., Altshuler, D., Gabriel, S.B., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Aniagwu, T., Marshall, P.A., Matthew, O., Nkwodimmah, C., Royal, C.D., Leppert, M.F., Dixon, M., Stein, L.D., Cunningham, F., Kanani, A., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Donnelly, P., Marchini, J., McVean, G.A., Myers, S.R., Cardon, L.R., Abecasis, G.R., Morris, A., Weir, B.S., Mullikin, J.C., Sherry, S.T., Feolo, M., Daly, M.J., Schaffner, S.F., Qiu, R., Kent, A., Dunston, G.M., Kato, K., Niikawa, N., Knoppers, B.M., Foster, M.W., Clayton, E.W., Wang, V.O., Watkin, J., Sodergren, E., Weinstock, G.M., Wilson, R.K., Fulton, L.L., Rogers, J., Birren, B.W., Han, H., Wang, H., Godbout, M., Wallenburg, J.C., L'Archeveque, P., Bellemare, G., Todani, K., Fujita, T., Tanaka, S., Holden, A.L., Lai, E.H., Collins, F.S., Brooks, L.D., McEwen, J.E., Guyer, M.S., Jordan, E., Peterson, J.L., Spiegel, J., Sung, L.M., Zacharia, L.F., Kennedy, K., Dunn, M. G., Seabrook, R., Shillito, M., Skene, B., Stewart, J.G., Valle, D.L., Jorde, L.B., Cho, M.K., Duster, T., Jasperse, M., Licinio, J., Long, J. C., Ossorio, P.N., Spallone, P., Terry, S.F., Lander, E.S., Nickerson, D.A., Boehnke, M., Douglas, J.A., Hudson, R.R., Kruglyak, L., and Nussbaum, R.L. (2003). The international HapMap project. *Nature* 426, 789–796.

Gyarmati, P., Conze, T., Zohari, S., Leblanc, N., Nilsson, M., Landegren, U., Baner, J., and Belak, S. (2008). Simultaneous genotyping of all hemagglutinin and neuraminidase subtypes of avian influenza viruses using padlock probes. *J. Clin. Microbiol.* 46, 1747–1751.

Hardenbol, P., Baner, J., Jain, M., Nilsson, M., Namsaraev, E.A., Karlin-Neumann, G.A., Fakhrai-Rad, H., Ronaghi, M., Willis, T.D., Landegren, U., and Davis, R.W. (2003). Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat. Biotechnol.* 21, 673–678.

Hardenbol, P., Yu, F., Belmont, J., Mackenzie, J., Bruckner, C., Brundage, T., Boudreau, A., Chow, S., Eberle, J., Erbilgin, A., Falkowski, M., Fitzgerald, R., Ghose, S., Iartchouk, O., Jain, M., Karlin-Neumann, G., Lu, X., Miao, X., Moore, B., Moorhead, M., Namsaraev, E., Pasternak, S., Prakash, E., Tran, K., Wang, Z., Jones, H.B., Davis, R.W., Willis, T.D., and Gibbs, R.A. (2005). Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* 15, 269–275.

Isaksson, M., Stenberg, J., Dahl, F., Thuresson, A.C., Bondeson, M.L., and Nilsson, M. (2007). MLGA – a rapid and cost-efficient assay for gene copy-number analysis. *Nucleic Acids Res.* 35, e115.

Jahangir Tafrechi, R.S., van de Rijke, F.M., Allallou, A., Larsson, C., Sloos, W.C., van de Sande, M., Wahlby, C., Janssen, G.M., and Raap, A.K. (2007). Single-cell A3243G mitochondrial DNA mutation load assays for segregation analysis. *J. Histochem. Cytochem.* 55, 1159–1166.

Jarvius, J., Melin, J., Göransson, J., Stenberg, J., Fredriksson, S., Gonzalez-Rey, C., Bertilsson, S., and Nilsson, M. (2006). Digital quantification using amplified single-molecule detection. *Nat. Methods* 3, 725–727.

Khatkar, M.S., Zenger, K.R., Hobbs, M., Hawken, R.J., Cavanagh, J.A., Barris, W., McClintock, A.E., McClintock, S., Thomson, P.C., Tier, B., Nicholas, F.W., and Raadsma, H.W. (2007). A primary assembly of a bovine haplotype block map based on a 15,036-single-nucleotide polymorphism panel genotyped in holstein-friesian cattle. *Genetics* 176, 763–772.

Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., Haugen, E., Zerr, T., Yamada, N.A., Tsang, P., Newman, T.L., Tuzun, E., Cheng, Z., Ebling, H.M., Tusneem, N., David, R., Gillett, W., Phelps, K.A., Weaver, M., Saranga, D., Brand, A., Tao, W., Gustafson, E., McKernan, K., Chen, L., Malig, M., Smith, J.D., Korn, J.M., McCarroll, S.A., Altshuler, D.A., Peiffer, D.A., Dorschner, M., Stamatoyannopoulos, J., Schwartz, D., Nickerson, D.A., Mullikin, J.C., Wilson, R.K., Bruhn, L., Olson, M.V., Kaul, R., Smith, D.R., and Eichler, E.E. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature* 453, 56–64.

Kong, F., Tong, Z., Chen, X., Sorrell, T., Wang, B., Wu, Q., Ellis, D., and Chen, S. (2008). Rapid identification and differentiation of Trichophyton species, based on sequence polymorphisms of the ribosomal internal transcribed spacer regions, by rolling-circle amplification. *J. Clin. Microbiol.* 46, 1192–1199.

Landegren, U., Kaiser, R., Sanders, J., and Hood, L. (1988). A ligase-mediated gene detection technique. *Science* 241, 1077–1080.

Larsson, C., Koch, J., Nygren, A., Janssen, G., Raap, A.K., Landegren, U., and Nilsson, M. (2004). In situ genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes. *Nat. Methods* 1, 227–232.

Laurie, C.C., Nickerson, D.A., Anderson, A.D., Weir, B.S., Livingston, R.J., Dean, M.D., Smith, K.L., Schadt, E.E., and Nachman, M.W. (2007). Linkage disequilibrium in wild mice. *PLoS Genet.* 3, e144.

Li, H.H., Gyllensten, U.B., Cui, X.F., Saiki, R.K., Erlich, H.A., and Arnheim, N. (1988). Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 335, 414–417.

Li, J., Young, C.S., Lizardi, P.M., and Stern, D.F. (2005). In situ detection of specific DNA double strand breaks using rolling circle amplification. *Cell Cycle* 4, 1767–1773.

Liu, D., Daubendiek, S.L., Zillman, M.A., Ryan, K., and Kool, E.T. (1996). Rolling circle DNA synthesis: small circular oligonucleotides as efficient templates for DNA polymerases. *J. Am. Chem. Soc.* 118, 1587–1594.

Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C., and Ward, D.C. (1998). Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* 19, 225–232.

Lohmann, J.S., Stougaard, M., and Koch, J. (2007). Detection of short DNA sequence elements on metaphase chromosomes using padlock probes and target primed rolling circle DNA synthesis. *BMC Mol. Biol.* 8, 103.

Lyamichev, V., Brow, M.A., and Dahlberg, J.E. (1993). Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* 260, 778–783.

Mardis, E.R. (2008). The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24, 133–141.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., and Rothberg, J.M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 326–327.

Maruyama, F., Kenzaka, T., Yamaguchi, N., Tani, K., and Nasu, M. (2005). Visualization and enumeration of bacteria carrying a specific gene sequence by in situ rolling circle amplification. *Appl. Environ. Microbiol.* 71, 7933–7940.

Maruyama, F., Tani, K., Kenzaka, T., Yamaguchi, N., and Nasu, M. (2006). Quantitative determination of free-DNA uptake in river bacteria at the single-cell level by in situ rolling-circle amplification. *Appl. Environ. Microbiol.* 72, 6248–6256.

McCarthy, E.L., Egeler, T.J., Bickerstaff, L.E., Pereira da Cunha, M., and Millard, P.J. (2006). Detection and identification of IHN and ISA viruses by isothermal DNA amplification in microcapillary tubes. *Anal. Bioanal. Chem.* 386, 1975–1984.

McCarthy, E.L., Bickerstaff, L.E., da Cunha, M.P., and Millard, P.J. (2007). Nucleic acid sensing by regenerable surface-associated isothermal rolling circle amplification. *Biosens. Bioelectron.* 22, 1236–1244.

Melin, J., Jarvius, J., Göransson, J., and Nilsson, M. (2007). Homogenous amplified single-molecule detection – characterization of key parameters. *Anal. Biochem.* 368, 230–238.

Millard, P.J., Bickerstaff, L.E., LaPatra, S.E., and Kim, C.H. (2006). Detection of infectious haematopoietic necrosis virus and infectious salmon anaemia virus by molecular padlock amplification. *J. Fish Dis.* 29, 201–213.

Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B.P., and Landegren, U. (1994). Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* 265, 2085–2088.

Nilsson, M., Krejci, K., Koch, J., Kwiatkowski, M., Gustavsson, P., and Landegren, U. (1997). Padlock probes reveal single-nucleotide differences, parent of origin and in situ distribution of centromeric sequences in human chromosomes 13 and 21. *Nat. Genet.* 16, 252–255.

Nilsson, M., Barbany, G., Antson, D.O., Gertow, K., and Landegren, U. (2000). Enhanced detection and distinction of RNA by enzymatic probe ligation. *Nat. Biotechnol.* 18, 791–793.

Nilsson, M., Antson, D.O., Barbany, G., and Landegren, U. (2001). RNA-templated DNA ligation for transcript analysis. *Nucleic Acids Res.* 29, 578–581.

Novais, R.C., Borsuk, S., Dellagostin, O.A., and Thorstenson, Y.R. (2008). Molecular inversion probes for sensitive detection of Mycobacterium tuberculosis. *J. Microbiol. Methods* 72, 60–66.

Okou, D.T., Steinberg, K.M., Middle, C., Cutler, D.J., Albert, T.J., and Zwick, M.E. (2007). Microarray-based genomic selection for high-throughput resequencing. *Nat. Methods* 4, 907–909.

Pickering, J., Bamford, A., Godbole, V., Briggs, J., Scozzafava, G., Roe, P., Wheeler, C., Ghouze, F., and Cuss, S. (2002). Integration of DNA ligation and rolling circle amplification for the homogeneous, end-point detection of single nucleotide polymorphisms. *Nucleic Acids Res.* 30, e60.

Porreca, G.J., Zhang, K., Li, J.B., Xie, B., Austin, D., Vassallo, S.L., LeProust, E.M., Peck, B.J., Emig, C.J., Dahl, F., Gao, Y., Church, G.M., and Shendure, J. (2007). Multiplex amplification of large sets of human exons. *Nat. Methods* 4, 931–936.

Saiyed, Z., Telang, S., and Ramchand, C. (2003). Application of magnetic techniques in the field of drug discovery and biomedicine. *Biomagn. Res. Technol.* 1, 2.

Salas, M. (2004). Phi29 DNA polymerase, a potent amplification enzyme. In *DNA Amplification: Current Technologies and Applications* (Demidov, V.V., and Broude, N.E., eds), pp. 21–34. Horizon Bioscience, Wymondham, England.

Salmon Hillbertz, N.H., Isaksson, M., Karlsson, E.K., Hellmen, E., Rosengren Pilgren, G., Savolainen, P., Wade, C.M., von Euler, H., Gustafson, U., Hedhammar, Å., Nilsson, M., Lindblad-Toh, K., Andersson, L., and Andersson, G. (2007). A duplication of FGF3, FGF4, FGF9 and ORAOV1 causes the hair ridge and predisposes to dermoid sinus in Ridgeback dogs. *Nat. Genet.* 39, 1318–1320.

Shaposhnikov, S., Larsson, C., Henriksson, S., Collins, A., and Nilsson, M. (2006). Detection of *Alu* sequences and mtDNA in comets using padlock probes. *Mutagenesis* 21, 243–247.

Smolina, I., Lee, C., and Frank-Kamenetskii, M. (2007). Detection of low-copy-number genomic DNA sequences in individual bacterial cells by using peptide nucleic acid-assisted rolling-circle amplification and fluorescence in situ hybridization. *Appl. Environ. Microbiol.* 73, 2324–2328.

Smolina, I.V., Kuhn, H., Lee, C., and Frank-Kamenetskii, M.D. (2008). Fluorescence-based detection of short DNA sequences under non-denaturing conditions. *Bioorg. Med. Chem.* 16, 84–93.

Smyth, D.J., Cooper, J.D., Bailey, R., Field, S., Burren, O., Smink, L.J., Guja, C., Ionescu-Tirgoviste, C., Widmer, B., Dunger, D.B., Savage, D.A., Walker, N.M., Clayton, D.G., and Todd, J.A. (2006). A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat. Genet.* 38, 617–619.

Stougaard, M., Lohmann, J.S., Zajac, M., Hamilton-Dutoit, S., and Koch, J. (2007). In situ detection of non-polyadenylated RNA molecules using Turtle Probes and target primed rolling circle PRINS. *BMC Biotechnol.* 7, 69.

Strömberg, M., Göransson, J., Gunnarsson, K., Nilsson, M., Svedlindh, P., and Stromme, M. (2008). Sensitive molecular diagnostics using volume-amplified magnetic nanobeads. *Nano Lett.* 8, 816–821.

Thomas, D.C., Nardone, G.A., and Randall, S.K. (1999). Amplification of padlock probes for DNA diagnostics by cascade rolling circle amplification or the polymerase chain reaction. *Arch. Pathol. Lab. Med.* 123, 1170–1176.

Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., and Church, G. (2004). Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* 432, 1050–1054.

Tong, Z., Kong, F., Wang, B., Zeng, X., and Gilbert, G.L. (2007). A practical method for subtyping of Streptococcus agalactiae serotype III, of human origin, using rolling circle amplification. *J. Microbiol. Methods* 70, 39–44.

van Doorn, R., Szemes, M., Bonants, P., Kowalchuk, G.A., Salles, J.F., Ortenberg, E., and Schoen, C.D. (2007). Quantitative multiplex detection of plant pathogens using a novel ligation probe-based system coupled with universal, high-throughput real-time PCR on OpenArrays. *BMC Genomics* 8, 276.

van Eijk, M.J., Broekhof, J.L., van der Poel, H.J., Hogers, R.C., Schneiders, H., Kamerbeek, J., Verstege, E., van Aart, J.W., Geerlings, H., Buntjer, J.B., van Oeveren, A.J., and Vos, P. (2004). SNPWave: a flexible multiplexed SNP genotyping technology. *Nucleic Acids Res.* 32, e47.

Vogelstein, B., and Kinzler, K.W. (1999). Digital PCR. *Proc. Natl. Acad. Sci. USA* 96, 9236–9241.

Wang, Y., Moorhead, M., Karlin-Neumann, G., Falkowski, M., Chen, C., Siddiqui, F., Davis, R.W., Willis, T.D., and Faham, M. (2005). Allele quantification using molecular inversion probes (MIP). *Nucleic Acids Res.* 33, e183.

Zanke, B.W., Greenwood, C.M., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S.M., Prendergast, J., Olschwang, S., Chiang, T., Crowdy, E., Ferretti, V., Laflamme, P., Sundararajan, S., Roumy, S., Olivier, J.F., Robidoux, F., Sladek, R., Montpetit, A., Campbell, P., Bezieau, S., O'Shea, A.M., Zogopoulos, G., Cotterchio, M., Newcomb, P., McLaughlin, J., Younghusband, B., Green, R., Green, J., Porteous, M.E., Campbell, H., Blanche, H., Sahbatou, M., Tubacher, E., Bonaiti-Pellie, C., Buecher, B., Riboli, E., Kury, S., Chanock, S.J., Potter, J., Thomas, G., Gallinger, S., Hudson, T.J., and Dunlop, M.G. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* 39, 989–994.

Zhang, D.Y., Brandwein, M., Hsuih, T.C., and Li, H. (1998). Amplification of target-specific, ligation-dependent circular probe. *Gene* 211, 277–285.

Zheng, S.L., Liu, W., Wiklund, F., Dimitrov, L., Balter, K., Sun, J., Adami, H.O., Johansson, J.E., Sun, J., Chang, B., Loza, M., Turner, A.R., Bleecker, E.R., Meyers, D.A., Carpten, J.D., Duggan, D., Isaacs, W.B., Xu, J., and Gronberg, H. (2006). A comprehensive association study for genes in inflammation pathway provides support for their roles in prostate cancer risk in the CAPS study. *Prostate* 66, 1556–1564.

Zhong, X., Lizardi, P.M., Huang, X., Bray-Ward, P.L., and Ward, D.C. (2001). Visualization of oligonucleotide probes and point mutations in interphase nuclei and DNA fibers using rolling circle DNA amplification. *Proc. Natl. Acad. Sci. USA* 98, 3940–3945.

Chapter 10

Molecular Cytogenetics in Molecular Diagnostics

Holger Tönnies
Institute of Human Genetics, Universitätsklinikum Schleswig-Holstein, Campus Kiel, Christian-Albrechts-Universität zu Kiel, Germany

10.1 INTRODUCTION

The fundamental cellular processes of DNA replication, DNA repair, mitosis, and meiosis ensure the high integrity of the human genome. However, mutations can occur in germline and somatic cells either induced or spontaneously. Depending on their physical size, these mutations manifest as altered sequences on the DNA level (e.g. gene mutations) or gross alterations as numerical and structural chromosome aberrations. Chromosomal aberrations are the major cause of congenital anomalies, mental retardation, and infertility in humans. Furthermore, they account for approximately half of all early spontaneous pregnancy losses (Warburton, 2000). Overall, 1 in 120 liveborn children is a carrier of a chromosomal abnormality. Half of these children are phenotypically conspicuous as a result of a chromosomal imbalance (Gardner and Sutherland, 2004). Four out of 1,000 newborns are carriers of an unbalanced chromosome abnormality, whereas 4.3 out of 1,000 newborns are showing balanced chromosomal rearrangements without a phenotype. Partial losses (monosomy) or gains (trisomy) of euchromatic chromosomal material resulting in different clinical phenotypes have been described for all human chromosomes (Schinzel, 2001; http://www.ecaruca.net; https://decipher.sanger.ac.uk). In genetic pre-, postnatal, and routine tumor genetic diagnostics, conventional cytogenetics using classical karyotyping of chromosomes is, until today, the most widely used technique to characterize numerical and structural intra- and interchromosomal aberrations (see Fig. 10.1).

Chromosome analysis is the gold standard for the diagnosis and prognosis of congenital (inborn) and acquired disorders (e.g. neoplasia). The precise characterization of aberrant cytogenetic findings is imperative for syndromologic assignment, phenotype–karyotype correlations, and genetic counseling in pre- and postnatal diagnostics. In tumor genetics conventional single cell and metaphase cytogenetics is essential for disease monitoring, tumor staging and research purposes to identify chromosomal regions harboring putative tumor suppressor and proto-oncogenes. However, limited chromosome specific resolution obtained by conventional chromosome banding techniques makes the recognition and interpretation of masked or cryptic chromosome aberrations difficult if not impossible to ascertain in several cases.

In the last three decades, molecular cytogenetic techniques based on fluorescence *in situ* hybridization (FISH) applications locating specific, fluorescence-labeled nucleic acid sequences in interphase cells or metaphase chromosomes, have become fast, sensitive, and important complementing tools in genetic diagnostics (reviewed in Tönnies, 2002). The use of diverse, locus-specific FISH probes and multicolor assays enhances the thorough characterization of numerical and complex chromosome aberrations regardless of their complexity, filling in parts the gap between conventional chromosome banding analysis and molecular genetic studies on the DNA level. Until today, multiple different molecular cytogenetic strategies and applications are used to extend the understanding of diseases and cellular pathology and to generate routine diagnostic tests.

There are numerous additional experimental-based techniques established, which are used for comparative genomics and other fundamental research questions. This chapter focuses on diagnostic approaches used in routine laboratories to detect and characterize genetic changes over the base pair level. In comparison to the first two decades of molecular cytogenetics, the development of new FISH approaches implemented in the routine diagnostic setup has been decreased because of the development of new high-resolution microarray-based approaches and high-throughput PCR- or sequencing-based techniques, which have the

potential to be automated, and are fast and efficient in comparison to cell culture-dependent assays (for more details see Chapters 16 and 24).

For the understanding of array-based techniques (Chapter 16) and for professional decision-making in the diagnostic context, it is essential to comprehend the development from conventional single cell chromosome analysis to FISH-based approaches in the early 1980s to modern array-based techniques. Furthermore, due to first array-based studies of the human genome, it became clear that single cell-based approaches, such as interphase FISH (I-FISH) and karyotyping, are, together with quantitative PCR-based approaches, essential for the validation and verification of whole genome screening results to differentiate between pathogenic and benign copy number variations in the human genome. In the routine diagnostic setting, the ''indirect'' whole genome array methods pave the way to investigate the whole genome in a high-resolution manner, whereas focused, mainly single cell (''direct'') methods are cost effective and indispensable to detect specific submicroscopic alterations commonly found in clinically delineated syndromes (e.g. microdeletion syndromes) and neoplasias.

This chapter will summarize modern molecular cytogenetic diagnostics and give a brief overview over the multiple molecular cytogenetic FISH techniques and multicolor assays currently used in routine genetic diagnostics.

10.2 FROM CONVENTIONAL TO MOLECULAR CYTOGENETICS

The first drawings of human chromosomes were published more than 100 years ago (Arnold, 1879; Flemming, 1881; Hansemann 1890; for review see Cremer, 1985). It took several decades to develop stable protocols to prepare human chromosomes and to establish the correct number of human chromosomes. For clinical cytogeneticists the 1960s were the most exciting. With the establishment of hypotonic treatment resulting in successful metaphase chromosome spreading (Hsu and Pomerat, 1953), Tijo and Levan (1956) managed to determine the correct number of chromosomes in human diploid cells to 46. Three years later, Lejeune and coworkers (1959) published their observation, that full trisomy 21 resulting in 47 chromosomes is the main cause for Down syndrome. The next technical improvement in cytology was the addition of colchicine to *in vitro* cultures for the accumulation of dividing cells prior to fixation (Moorhead *et al.*, 1960).

In tumor genetics, the first finding of a chromosomal abnormality was the identification of a minute chromosome, named by its location of discovery ''Philadelphia chromosome'', which was regularly found in peripheral blood cells of patients with chronic myeloid leukemia (CML) (Novell and Hungerford, 1960). In the late 1960s, with the development and establishment of new conventional cytogenetic banding

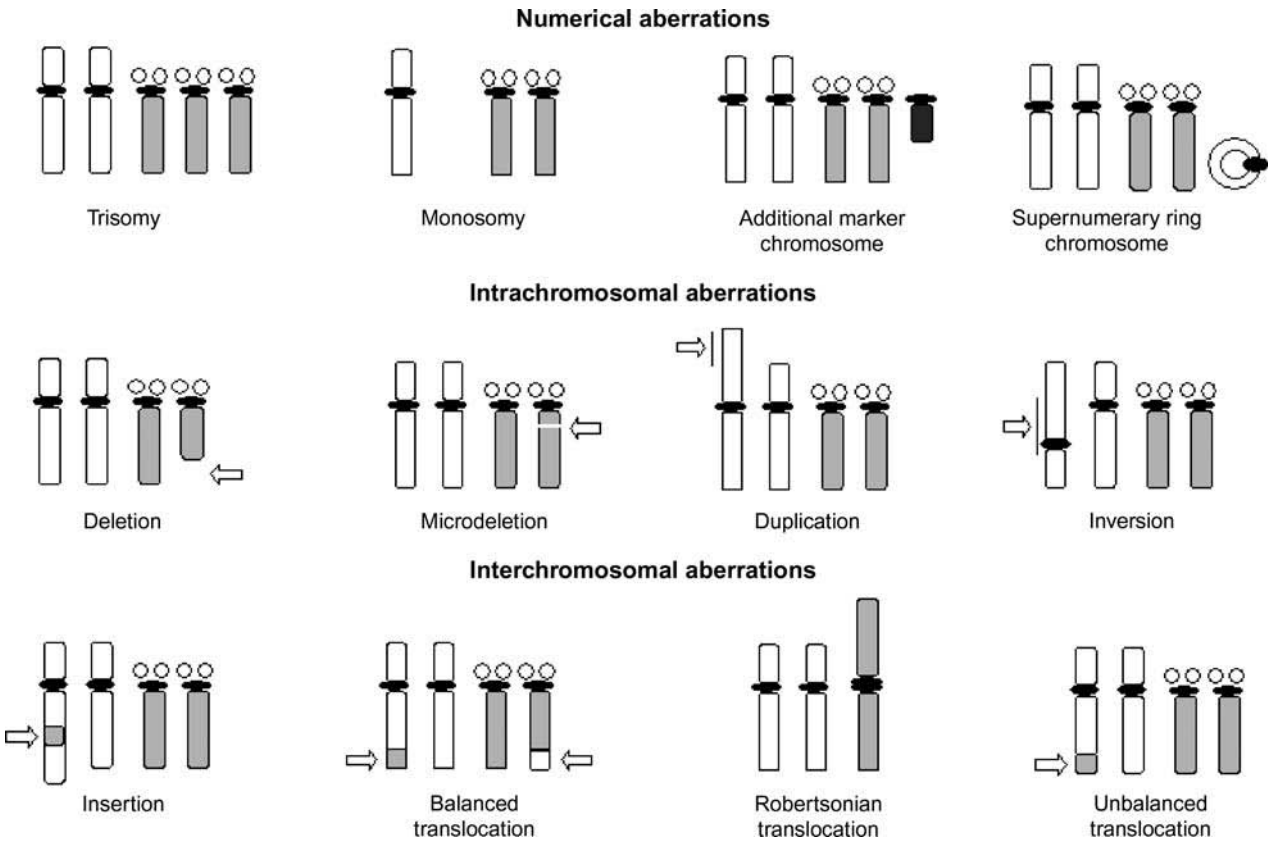


FIGURE 10.1 Schematic illustration of common numerical and structural intra- and interchromosomal aberrations in humans.

techniques (Casperson *et al.*, 1970) to detect numerical and structural chromosome abnormalities in humans, the era of clinical cytogenetics began. Over the last decades multiple additional conventional banding techniques were introduced in routine use (for overview see Verma *et al.*, 1989; Barch *et al.*, 1997) to detect chromosomal aberrations.

Today, the identification and characterization of numerical and structural chromosomal aberrations by karyotyping banded chromosomes is performed routinely in prenatal, postnatal, and tumor cytogenetic diagnostics as in basic genomic research.

Detectable chromosomal aberrations can be divided into numerical and structural aberrations (see Fig. 10.1). In humans, numerical chromosome aberrations as aneusomies of whole chromosomes (e.g. trisomy 21, trisomy 18, or the chromosome constitution 45,X) are frequent, well-characterized chromosome abnormalities and easily detectable by morphologic and numerical chromosome examination. Structural chromosomal abnormalities can be subgrouped into intra-chromosomal (deletions, duplications, inversions) and inter-chromosomal (balanced and unbalanced translocations, insertions) aberrations affecting more than one chromosome (see Fig.10.1; see also ISCN, 2005).

More laborious to identify by conventional cytogenetics alone are the so-called marker chromosomes. Marker chromosomes are structurally abnormal chromosomes of which the origin of the euchromatic content cannot be determined by conventional cytogenetic analysis alone. Often supernumerary, their incidence varies from 0.3 to 3.7/1,000 in newborns and mentally/developmentally delayed patients (Buckton *et al.*, 1985; http://www.med.uni-jena.de/fish/sSMC/00START.htm). Ring chromosomes, some also supernumerary (numerical aberration; e.g. due to mitotic malsegregation), mainly originate from rearranged normal chromosomes due to deletions of the telomeric ends of the chromosome with following ring formation (intrachromosomal aberration; Tönnies *et al.*, 2003a). However, the effectiveness of conventional cytogenetics depends directly on the size of chromosome aberrations appearing in human cells. The limited chromosome-specific banding resolution and assignment obtained by conventional chromosome banding makes the characterization and correct interpretation of complex and subtle chromosome aberrations difficult to ascertain and is therefore by nature often imprecise.

Furthermore, some chromosomal aberrations detected in an affected child are *de novo*, defined by a normal constitutional karyotype of the parents. Therefore, a precise definition of the lost or duplicated chromosomal material is indispensable for any exact genotype–phenotype correlation and provides diagnostic and possibly prognostic information. By conventional cytogenetic methods, rearrangements in the size of 5–10 megabases (Mb) affecting single chromosomes are detectable if high-resolution banding is achieved. However, aberrations smaller than 5 Mb (e.g. microdeletion syndromes; see Table 10.1), complex chromosome aberrations involving three or more chromosomes, and marker chromosomes composed of unknown chromatic material often give unsatisfactory results using conventional cytogenetic techniques only. The dependency on dividing cells to prepare metaphase spreads, which is time consuming, and the occurrence of chromosome aberrations not to be characterized exactly using monochrome banding techniques enhanced the development of alternative, more sensitive approaches based on the hybridization of DNA probes to patient DNA targets (e.g. metaphase chromosomes or interphase nuclei) overcoming these conventional cytogenetics limitations.

TABLE 10.1 Examples of common autosomal microdeletion syndromes in clinical diagnostics.

Chromosomal localization	Name	Involved gene(s)/ hybridization target	OMIM[a] entry
del(1)(p36.3)	Monosomy 1p syndrome		#607872
del(4)(p16.3)	Wolf-Hirschhorn syndrome	Wolf-Hirschhorn critical region	#194190
del(5)(p15.2p15.3)	Cri-du-chat syndrome	Cri-du-chat critical region	#123450
del(7)(q11.23q11.23)	Williams-Beuren syndrome	*ELN*	#194050
del(8)(q24.1q24.1)	Langer-Giedion syndrome	*EXT1, TRPS I, TRPS II*	#150230
del(11)(p13p13)	WAGR syndrome	*WT1, PAX6*	#194072
del(15)(q11q13)pat	Prader-Willi syndrome	*SNRPN*	#176270
del(15)(q11q13)mat	Angelman syndrome	*UBE3A*	#105830
del(16)(p13.3)	Rubinstein-Taybi syndrome	*CBP*	#180849
del(17)(p11.2p11.2)	Smith-Magenis syndrome	Smith-Magenis critical region	#182290
del(17)(p13.3)	Miller-Dieker syndrome	*LIS1*	#247200
del(20)(p11.23p11.23)	Alagille syndrome	*JAG1*	#118450
del(22)(q11.2q11.2)	VCF/DiGeorge syndrome	*VCF*/DiGeorge critical region	#192430 *188400

[a]*OMIM at http://www.ncbi.nlm.nih.gov/omim*

10.3 FLUORESCENCE *IN SITU* HYBRIDIZATION

By *in situ* hybridization (ISH) on metaphase spreads or interphase nuclei, genetic changes can be analyzed at the single cell level allowing the simultaneous assessment of different chromosomes or chromosome regions and the determination of clonal variability or mosaicism (see Fig. 10.2). Classical *in situ* hybridization is based on the binding (hybridization, annealing) of complementary, single-stranded-labeled nucleic acids to the fixed and denatured target DNA of metaphase chromosomes, whole interphase nuclei, or DNA fibers. During hybridization the probe penetrates to the target nucleic acid sequence and anneals to the complementary partner resulting in DNA duplexes of the bound probe and the former single-stranded target. After removing unbound probe material by stringent washing, results can be inspected by microscopy.

First described by Pardue and Gall (1969) using radio-labeled, repetitive DNA probes hybridizing to mice cell preparations, the detection has been carried out by autoradiography enabling the morphological (*in situ*) visualization of the presence of complementary nucleic acid sequences in the target material (John *et al.*, 1969). Hybridization and detection of the first single-copy DNA sequences by autoradiography were described by Harper and Saunders (1981).

Molecular cytogenetics is based on fluorescence *in situ* hybridization skip (FISH). Using non-radioactively labeled probes, Pinkel and coworkers (1986) introduced the basic protocol for FISH, improving the spatial resolution in comparison to radio-labeled probes. FISH is a sensitive and specific method providing precise information about the physical location of DNA sequences in cell nuclei or chromosomes.

10.4 BASIC TECHNICAL ELEMENTS AND MATERIALS

Fluorescence *in situ* hybridization skip, or FISH, is a technique for the high-resolution visualization of labeled nucleic acid probes on target material (see Fig. 10.2). FISH is a stepwise process beginning with the selection and

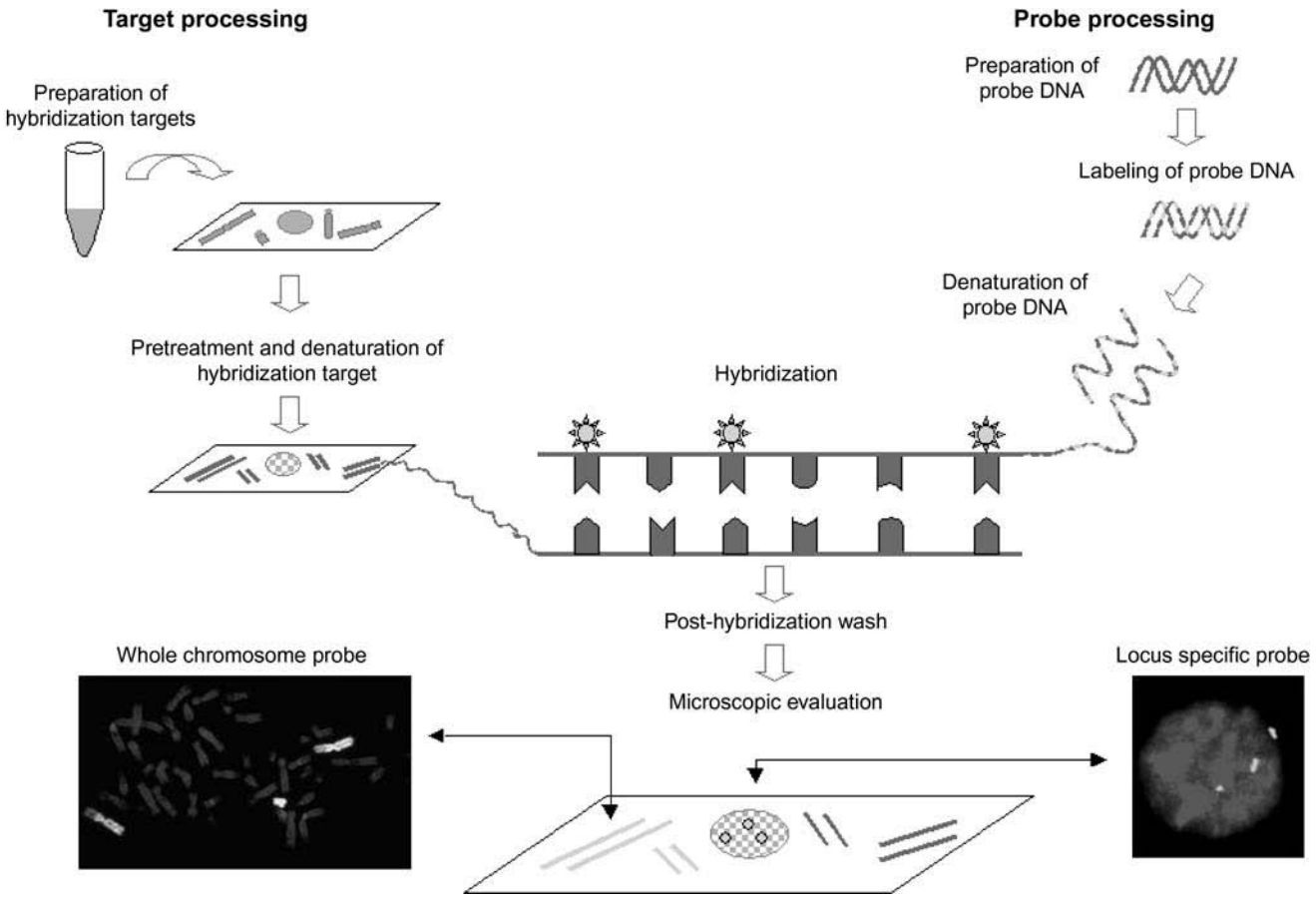


FIGURE 10.2 Flowchart summarizing the standard steps of fluorescence *in situ* hybridization. Target material (cell suspension including metaphase spreads and interphase nuclei) is dropped onto glass slides, pretreated, and denatured to prepare single-stranded DNA. After hybridizing the single-stranded, labeled probe to the target material and post-hybridization wash, results can be analyzed by microscopic inspection. By using a whole chromosome probe for chromosome 3 (left), a partial trisomy is visible. Using a locus-specific probe (e.g. YAC, right), three signals are detectable in the interphase nucleus reflecting a partial trisomy 3.

pretreatment of the target material, followed by the production and labeling of the appropriate probe to be hybridized, the hybridization process itself, and the detection, inspection, and documentation of the results. For the user, a huge number of variable targets, probes, and labeling procedures are available nowadays. The resolution of FISH analyses directly depends on two crucial elements: the target and the probe.

10.4.1 Targets for FISH

A wide variety of cellular materials can be used as targets for investigation by FISH. In routine diagnostics, in particular interphase nuclei and metaphase chromosomes of peripheral blood (routinely T-lymphocytes), primary fibroblasts, bone marrow cells or, in prenatal diagnostics, amniocytes and chorionic villi cells are used after tissue cultivation (see Fig. 10.2; for technical details see Wegner, 1999). In leukemic specimens, where the number of viable, dividing cells is sometimes low, so-called hypermetaphase spreads obtained by long-term exposure of cells to colcemide can be produced (Seong *et al.*, 1995). The yield of metaphase spread-like targets is higher vs. normal chromosome preparation with short-term colcemide treatment. The advantage in comparison to I-FISH is the ability to see a chromosome-like morphology, and to minimize false positive results due to stochastic signal overlap often seen in I-FISH experiments. Furthermore, without cultivation, buccal smear cells, sperm cells, and cells originating from the urinary tract can be used for I-FISH either to get primary informations or to verify former results obtained by peripheral blood cell karyotyping or array-CGH analyses (see also section 10.10). In preimplantation genetic diagnosis (see also Chapter 33), FISH for aneuploidy screening is routinely performed on polar bodies and, if allowed, on blastomeres (Delhanty *et al.*, 1993; Fragouli, 2007).

Mainly used in basic research for, e.g., clone mapping, rather than in diagnostic units, FIBER-FISH using released chromatin fibers from interphase cells (Heng *et al.*, 1992; Wiegant *et al.*, 1992) is utilized to study the structure and organization of mammalian genomes with high resolution. A great advantage of I-FISH on single cells is that it can also be performed on archived or non-dividing material such as fresh frozen and paraffin-embedded tissues slices and touch preparations of pathological specimens. Resolution of FISH experiments depends directly on the target material. Studying metaphase chromosomes, resolution is about 2–5 Mb, whereas in interphase nuclei the resolution is varying between 2 Mb and 50 kb. When performing FIBER-FISH experiments, a resolution of 5–500 kb can be achieved (Speicher and Carter, 2005).

For reproducible hybridization results some targets need different pretreatment steps (e.g. proteolytic steps) for better penetration of the probe. For all target materials, a fixation step is required. In cytogenetics, conventional fixative (methanol:acetic acid = 3:1) is routinely used, but also ethanol fixation (70%) provides good results. For FISH, formerly dividing cells are applied to glass slides for pretreatment by RNase or proteolytic digestion, if necessary. Before hybridization, the double-stranded target DNA has to be denatured either by chemical or heat treatment, melting the DNA double helices to single-stranded DNA (for technical details see Schwarzacher and Heslop-Harrison, 2000).

10.4.2 DNA Probes for FISH

For FISH a variety of DNA probes established by different amplification or cloning techniques are used. Probe DNA can be prepared by locus-specific PCR amplifying single genes, cloning of large human DNA fragments, by chromosome microdissection or flow-sorting of whole chromosomes. A large number of commercial FISH probes is available today (see Fig. 10.3). Region- or band-specific chromosomal DNA can be cloned in vectors of different sizes, such as cosmids, plasmids, P1-derived artificial chromosomes (PACs), bacterial artificial chromosomes (BACs), or yeast artificial chromosomes (YACs). The choice of the vector system depends on the size of the DNA fragment to be cloned. Mapped and sequenced large insert clones are available for almost any chromosomal region and can be

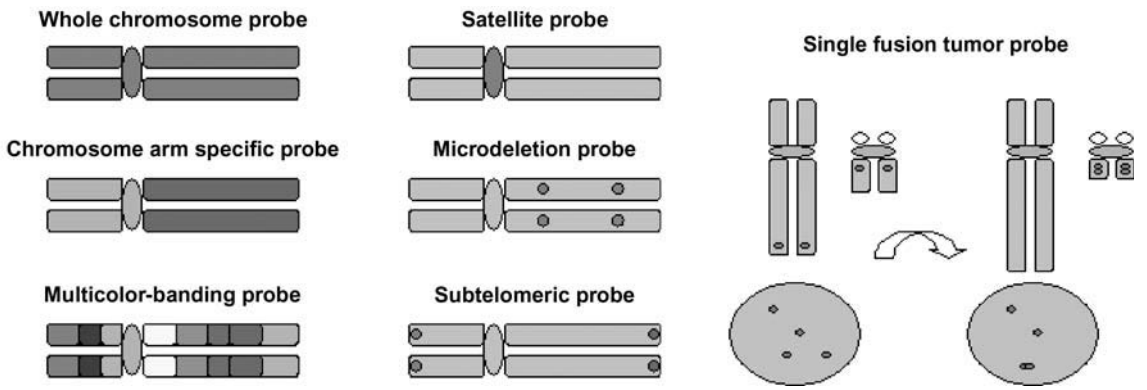


FIGURE 10.3 Schematic illustration of various FISH probes used in routine cytogenetic diagnostics. For explanation, see text.

selected easily by using internet databases (UCSC genome browser, Ensembl Cytoview, NCBI Map-Viewer).

10.4.3 Probe Labeling

Nucleic acid FISH probes are labeled and visualized either directly using fluorochrome-conjugated nucleotides (e.g. fluorescein isothiocyanate; FITC-dUTP) or indirectly using reporter molecules (e.g. biotin-dUTP, digoxygenin (DIG)-dUTP; for details see Schwarzacher and Heslop-Harrison, 2000). Nucleic acid-labeling approaches are based on the use of different enzymatic or chemical methods, and the choice of the labeling technique varies from laboratory to laboratory (for technical details see Rautenstrauß and Liehr, 2002). Most labeling strategies are based on the enzymatic incorporation of label-conjugated nucleotides in a new synthesized DNA strand. Examples for these approaches are the polymerase chain reaction (PCR; amplification and labeling of the probe) using degenerated oligonucleotide primers (DOP; Telenius *et al.*, 1992), the nick translation (no amplification but fragmentation of probe size; Rigby *et al.*, 1977), or random priming using hexamer primers resulting in probe amplification during labeling procedure (Feinberg and Vogelstein, 1983). For small oligonucleotides, terminal fluorochrome-labeling reactions (Bauman *et al.*, 1980) are commonly used.

In the last years, various other labeling approaches such as chemical cross-linking have been established (van Gijlswijk *et al.*, 2001), and numerous kits are commercially available to label probe DNA with high efficiency.

For multicolor assays discriminating numerous targets simultaneously (for details see section 10.6), combinations of five fluorochromes can result in more than 24 colors using a ratio labeling scheme (same fluorochrome combination but different proportions; Dauwerse *et al.*, 1992; Nederlof *et al.*, 1992), combinatorial labeling (unique combination of different fluorochromes) as described by Nederlof and coworkers (1990), or the COBRA (combined binary ratio labeling) strategy introduced by Tanke and coworkers (1999).

A crucial step for all labeling systems is to achieve an optimal fragment size of the labeled FISH probe, in order to assure appropriate penetration and hybridization efficiency. For FISH experiments using locus-specific probes, the fragment size after labeling should not exceed 300 bp. Therefore, some of the labeling approaches also require fragmentation of probe DNA by DNase or ultrasonic treatment before *in situ* hybridization.

10.4.4 Hybridization, Post-Hybridization Wash, Detection, and Documentation

For DNA duplex formation, probe DNA and target DNA has to be single stranded (see Fig. 10.2). For denaturation, the probe DNA dissolved in hybridization buffer (also known as hybridization mix), containing formamide and appropriate salt-concentrations, is dissociated in a water bath at the appropriate temperature (depending on the probe composition). Many probes of higher complexity contain repetitive DNA elements that are scattered throughout the genome (e.g. whole chromosome painting probes). In order to suppress their unspecific cross-hybridization, unlabeled competitor (Cot-1) DNA, a highly repetitive fraction of the human genome, is added to the hybridization mix. The Cot-1 DNA binds to the highly repetitive elements in the probe, suppressing their unspecific cross-hybridization to the target (chromosome *in situ* suppression hybridization; Cremer *et al.*, 1990; Hulten *et al.*, 1991). After appropriate prehybridization or preannealing, the fluid hybridization mix is applied to the target material for an appropriate time (hours to days) depending on the size and complexity of the probe and target. Non- and unspecific bound probe is washed off using stringent washes (low salt concentration) under thermal conditions that do not denature the specific DNA duplexes (washing temperature lower than melting temperature).

For direct-labeled probes, no detection steps are necessary, whereas indirect-labeled probes (e.g. biotin- or digoxigenin-labeled probes) have to be detected using appropriate fluorochrome-conjugated antibodies. After counterstaining the target material using either DAPI (4,6-diamidino-2-phenylindole) or propidium iodide, and application of an antifade solution to prevent bleaching of the fluorescent signals, hybridization results can be visually inspected using an epifluorescence microscope equipped with appropriate single- or multiband pass filter sets. Digital documentation of hybridization results is routinely done using a charge-coupled device (CCD) camera, connected to a computer and appropriate analysis software.

10.5 TYPES OF FISH PROBES AND RECENT FISH APPROACHES FOR METAPHASE AND INTERPHASE FISH

The ability to detect and characterize chromosomal abnormalities in metaphase spreads and interphase cells using FISH has been greatly enhanced by the rapidly increasing availability of numerous chromosome and locus specific probes (see Fig. 10.3 and Table 10.2). The choice of probe and the simultaneous use of multiprobe assays depend on the particular application in question.

10.5.1 Centromeric Satellite Probes

The first routinely used FISH probes were centromere-specific probes detecting highly repetitive centromeric α-satellite DNA sequences (Cremer *et al.*, 1986). In molecular diagnostics, these probes are mainly used for chromosome enumeration and marker chromosome identification (Figs 10.3, 10.4a and b). Especially for interphase cell analysis, which can be performed on cytological preparations as well as in sections

of formalin-fixed and paraffin-embedded (FFPE) tissues, these probes are particularly attractive because of high sensitivity and excellent hybridization efficiencies. The hybridization time of these highly repetitive probes is short (only hours) and provides rapid results without the need of specific cytogenetic expertise for chromosome recognition or cell culture/metaphase preparation. The so-called all-human centromeric probes, a mix of all α-satellite repeats of the human genome hybridizing to all human centromeres simultaneously, are used for the detection of dicentric chromosomes or for the definition of neocentromeres, always lacking α-satellite DNA sequences.

However, by using chromosome-specific centromeric probes for interphase cytogenetics, centromeric polymorphisms can result in split signals and therefore in false positive findings. Additionally, for some chromosomes, sequence homologies of the α-satellite sequences are resulting in cross-hybridizations (e.g. chromosomes 13/21 and 14/22), making these probes less suitable for interphase cytogenetics. Furthermore, the use of centromeric probes for supernumerary marker chromosome detection gives no information about the euchromatic chromosome contents accounting for most phenotypic features.

10.5.2 Whole Chromosome "Painting" Probes

Whole chromosome probes (WCP), also called painting probes, are DNA libraries representing a cocktail of DNA fragments of a single human chromosome (Deaven *et al.*, 1986; Guan *et al.*, 1993; Figs 10.3, 10.4c and d). These DNA probes, obtained by chromosome flow sorting or chromosome microdissection (see Fig. 10.5 and section 10.5.3), allow the labeling of individual chromosomes in metaphase spreads, and subsequently the identification and characterization of both numerical and interchromosomal structural aberrations as translocations and non-homologous insertions. Partial chromosome probes (Fig. 10.3), generated by chromosome microdissection and representing the short or the long arm of chromosomes, are valuable tools for the detection of intrachromosomal pericentric inversions including the centromere.

In routine use, the application of appropriate whole chromosome probes to ascertain chromosomal aberrations needs the prior knowledge (and therefore cytogenetic expertise) of the affected chromosome(s) in question. Otherwise, for example for euchromatin containing marker chromosome characterization, a whole repertoire of different DNA probes has to be hybridized and analyzed to narrow down or to identify the origin of unknown chromosomal material (Blennow *et al.*, 1995).

However, a locus-specific determination of the affected chromosomal region in intrachromosomal rearrangements, such as deletions and duplications, and paracentric inversions, affecting only one chromosome arm, is not possible by using whole or partial chromosome probes. The detection of small interchromosomal translocations involving only very small regions of the chromosome ends often cannot be performed with the adequate diagnostic sensitivity by using whole chromosome probes. In contrast to centromeric probes, chromosome enumeration by whole chromosome probes can be performed only on metaphase spreads, demanding the availability of proliferating material.

10.5.3 Probe Generation by Chromosome Microdissection (Micro-FISH)

The characterization of the composition and chromosomal origin of marker chromosomes or parts thereof can be performed straightforwardly by chromosome

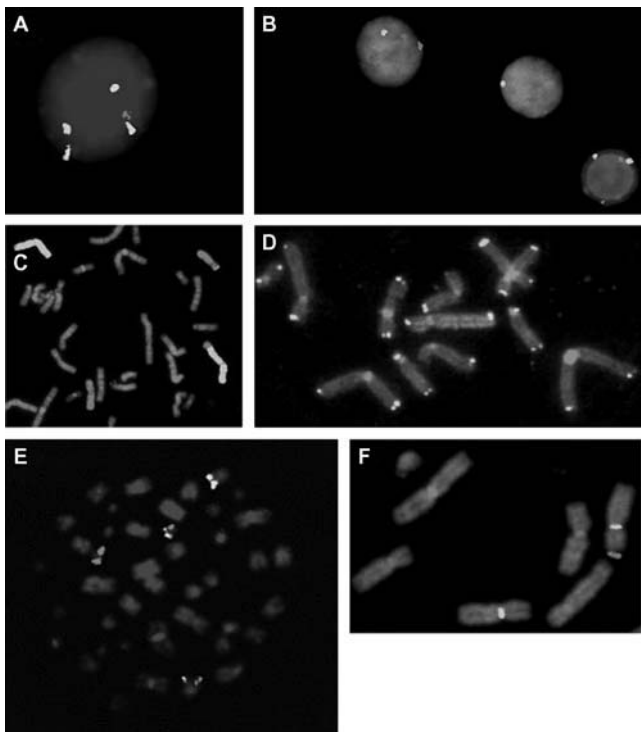


FIGURE 10.4 Example combinations of single probes used in routine diagnostics. **A.** Interphase nuclei of an uncultured amniocytes hybridized with three different alphoid probes (cep-X, red, cep-Y, green, cep-18, blue). A normal male signal constitution has been detected for the chromosomes X and Y. The three blue signals gave evidence for a trisomy 18 in the amniocytes (courtesy of Markus Stumm, Berlin, Germany). **B.** I-FISH using centromeric probes for chromosomes 7 (green) and 9 (red), detecting a monosomy 7 mosaicism. **C.** Whole chromosome probes for chromosome 3 (green) and 11 (red), detecting a reciprocal translocation t(3;11). **D.** Combinatorial hybridization of whole chromosome painting for chromosome 4 (red) and an all telomere repeat probe. **E.** Result of a D-FISH experiment performed on a bone marrow metaphase of a chronic myelogenous leukemia (CML) patient detecting a Philadelphia-positive metaphase spread resulting in mix-color signals (fusion signals) on the derivative chromosomes der(9) and der(22) (courtesy of Ivan Loncarevic, Institute of Human Genetics and Anthropology, Jena, Germany). **F.** Microdeletion of the SHOX gene (red) at the tip of one X-chromosome in a girl. As a control probe, a cep-X (green) has been hybridized. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

TABLE 10.2 FISH probes for the characterization of specific chromosome alterations.

Probes for FISH	Hybridization target	Patient material	Aberration type detectable
Commercial probes			
Centromere specific α-satellite probes	Centromeres	Metaphase spreads, interphase spreads	Numerical aberrations, identification of marker chromosome origin
Whole chromosome probes (WCP) and partial chromosome arm-specific probes (PCP)	Whole chromosomes	Metaphase spreads, interphase spreads	Numerical aberrations, interchromosomal aberrations as balanced translocations or non-homologous insertions
Locus-specific probes (microdeletion and microduplication probes)	Submicroscopic chromosomal loci	Metaphase spreads Interphase cells	Microdeletions and microduplications
Tumor probes (single fusion probes, break apart probes, double fusion probes, oncogene amplification probes)	Submicroscopic fused chromosomal loci (translocation breakpoints), homogeneously staining regions (gene amplifications)	Interphase cells (Metaphase spreads)	Chimeric gene fusions, gene amplifications, inversions, deletions
All telomeric probes (Q-FISH probes)	Telomere repeats	Metaphase spreads	Quantification of telomere size, loss of telomere
Subtelomeric probes	Subtelomeric regions	Metaphase spreads	Cryptic terminal deletions, duplications, translocations
Non-commercial probes			
PCR amplified microdissected or chromosomes or chromosomal bands (Micro-FISH)	Chromosomes or parts thereof	Patient and control metaphase spreads, forward and reverse ''painting''	Marker chromosome characterization
PCR products *in situ* (PRINS) PCR primer for specifc chromosomal loci, genes or alphoid sequences	Telomere repeats, microdeletions, single genes	Metaphase spreads	Telomeric aberrations, microdeletions, single gene deletions

microdissection or micro-FISH (Meltzer *et al.*, 1992). This method is based on the micro-manipulated or laser capture-based chromosome dissection of the marker chromosome followed by PCR-mediated DNA amplification. By reverse (on normal metaphase spread) or forward (on patients' metaphase spreads) hybridization, the complete euchromatic content of the marker chromosome can be explored without the use of commercial probes (see Fig. 10.5). A prerequisite for this assay is proliferating cell material of the patient. In addition to the fact that some chromosome aberrations like inversions and small duplications are not detectable by using dissected material as a probe, precise breakpoint identification is limited due to the hybridization on metaphase chromosomes. Furthermore, there are practical limitations to micro-FISH for diagnostic purposes since specialized equipment, technical skills, and a profound cytogenetic knowledge for the undoubtful recognition of the marker chromosome are necessary.

10.5.4 Region and Locus-Specific Probes

10.5.4.1 Locus-Specific Probes

For the investigation of small, submicroscopic chromosomal loci of the human genome, which are too small to be visualized by conventional cytogenetics, a wide spectrum of so-called locus-specific probes are used routinely in diagnostic FISH laboratories. Most of these probes are vector-cloned probes. For the molecular-cytogenetic verification of syndromes with submicroscopic gain or loss, such as the classical microdeletion and microduplication syndromes (see Table 10.1), single-copy probes detecting the common region of interest are hybridized to patient cells (see Fig. 10.4f). Although metaphase spreads are used routinely as targets for microdeletion syndromes, interphase nuclei are utilized for the detection of microduplications because of the better separation of closely spaced hybridization signals (Shaffer and Lupski, 2000). Today, most commercial probes include a second

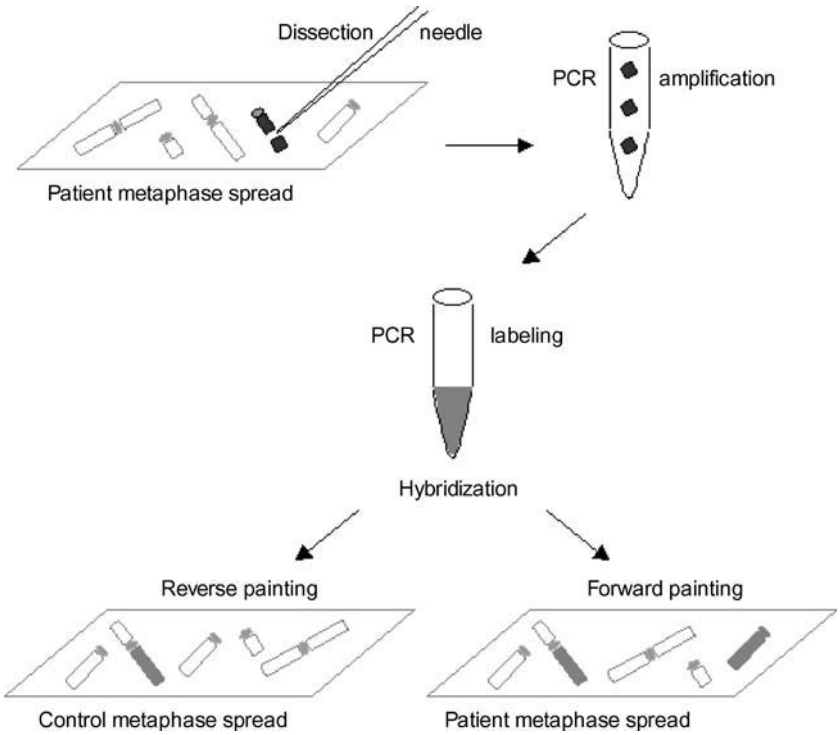


FIGURE 10.5 Flowchart summarizing the practical steps of marker chromosome microdissection (Micro-FISH). Using a glass dissection needle that is controlled by a micromanipulator, some copies of the marker chromosome are dissected and transferred to a microcentrifuge tube. Subsequently, the chromosomal DNA is amplified and labeled by degenerate oligonucleotide primed-PCR (DOP-PCR). The resulting chromosome probe is hybridized to metaphases containing the marker (patient metaphase spread; forward painting) to verify the regional authenticity of the probe. Simultaneously, the chromosomal composition of the marker is determined by hybridization of the probe to normal metaphase chromosomes.

control probe hybridizing on the non-affected chromosome arm for the validation of successful hybridization.

In prenatal diagnostics, differently labeled single probes can be used for rapid testing for chromosomal aneuploidies on uncultivated amniocytes (Kuo *et al.*, 1991; Klinger *et al.*, 1992; Ward *et al.*, 1993; Stumm *et al.*, 2006; Fig. 10.4a). By this approach, a combination of locus-specific and alphoid probes of human chromosomes 21, 18, 13, X, and Y are hybridized in two combinations for the detection of the most common aneuploidies in humans. However, this FISH test is not designed to detect all chromosome aneuploidies and can be utilized only as an adjunctive test to conventional cytogenetics.

10.5.4.2 Tumor Probes

In tumor genetics, most chromosomal rearrangements have major diagnostic and prognostic relevance. A number of single-copy locus-specific probes, detecting characteristic translocations/fusion genes, inversions, deletions, and oncogene amplifications on the single cell level are commercially available (Wang, 2002). These probes allow the investigation of metaphase spreads and interphase cells in one experiment. Furthermore, the detection of minimal residual disease and post-transplantation follow-up can be done with high sensitivity. Some of these locus-specific probes precisely locate translocation and inversion breakpoints, reflecting chimeric gene fusions in neoplastic diseases, especially in leukemias (e.g. *BRC-ABL*). Single-fusion dual-color FISH probes (Figs 10.3 and 10.4e), which were the first available tumor probes, detect specific translocation breakpoints in interphase nuclei by the presence of non-randomly distributed juxtaposed signals. However, in interphase diagnostics, these probes can have high rates of false positive signals. Newer types of translocation probes in cancer genetics are break-apart probes detecting simple splits of normally proximate signals and so-called double fusion probes (D-FISH probes) detecting a fusion signal on both derivative chromosomes (Grand *et al.*, 1998). Oncogene amplification probes, commonly used in solid tumor genetics, visualize gene amplifications in metaphase and interphase cells and clarify the identity of the genes involved.

10.5.4.3 Subtelomeric Probes and Telomere Repeat Probes

The highest gene concentrations in the human genome are in the subtelomeric regions of metaphase chromosomes (Saccone *et al.*, 1992). These chromosomal regions are prone to rearrangements, which could give rise to cryptic aberrations probably accounting for 5% to 10% of unexplained moderate-to-severe mental retardation cases, congenital

anomalies, and spontaneous abortions (Knight *et al.*, 1997, 1999; Anderlid *et al.*, 2002; Yu *et al.*, 2005). Subtelomeric probes were established for all chromosome ends (excluding the short arms of acrocentric chromosomes) to screen metaphase spreads for cryptic translocations or imbalances at the terminal euchromatic parts of the chromosomes.

The non-coding telomere repeat (TTAGGG)n distal to the subtelomeric region at the end of each eukaryotic chromosome protects the chromosome against rearrangements and fusion with other chromosomes (see Fig. 10.4d). Using a repetitive all-telomere probe, ring chromosome formation can be pinpointed, especially in small supernumerary ring chromosomes.

Telomere shortening, resulting from cell divisions over time, can lead to genomic instability and neoplasia (Shay *et al*., 1994). A quantitative FISH assay, using peptide nucleic acid (PNA) telomere oligonucleotide probes, resulting in stronger hybridization signals than standard DNA oligonucleotide probes, has been described by Lansdorp and coworkers (1996). Hybridizing these probes, the fluorescence intensity detected is directly proportional to the amount of telomere repeats.

10.5.5 Special Probe Types and Combinations

In the literature, several additional specific probe designs and combinations have been published over the years, but not all these are used routinely in diagnostics. It would go beyond the scope of this chapter to list all these in detail. All the following FISH-based sensitive approaches are detecting single or few chromosome alterations at a time.

10.5.5.1 Primed in situ Labeling

Primed *in situ* labeling (PRINS) is a complementary FISH approach based on the *in situ* hybridization of short unlabeled DNA primers to metaphase spreads or interphase cells and subsequent *in situ* chain elongation catalyzed by a DNA polymerase (Koch *et al.*, 1989, 1991). The product of this polymerase chain reaction *in situ* is visible due to incorporation of labeled nucleotides. Depending on the primers used, this time- and cost-effective method allows the detection of centromeric alpha satellite DNA (Hindkjaer *et al.*, 1995), telomeric repeats (Krejci and Koch, 1998), microdeletions, and single-copy genes (Cinti *et al.*, 1993; Kadandale *et al.*, 2000; Tharapel *et al.*, 2002). However, PRINS requires the knowledge of the exact target sequence for the primers and high-quality target material.

10.5.5.2 Single-Copy FISH Probes

Rogan and coworkers (2001) introduced the use of a new generation of small single-copy FISH (sc-FISH) probes designed by computational sequence analysis of approximately 100 kb genomic sequences, bridging the gap between molecular genetic data and molecular cytogenetics. These labeled probes, produced by PCR, can be hybridized without preannealing or blocking, as sc-FISH probes lack repetitive DNA sequences. These short probes are produced directly from genomic DNA without recombinant DNA techniques. Applications of these probes include detection of microdeletion syndromes and submicroscopic deletions (Knoll and Rogan, 2003).

10.5.6 MACISH and FICTION: Fluorescence Immunophenotyping and Interphase FISH

MACISH (morphology, antibody, chromosomes, *in situ* hybridization) and FICTION (fluorescence immunophenotyping and interphase cytogenetics) are methods combining immunophenotyping of cells and fluorescence *in situ* hybridization (Knuutila and Teerenhovi, 1989; Weber-Mathiesen *et al.*, 1993). These approaches allow the examination of numerical and structural chromosome abnormalities of immunologically classified cells. They are used especially to detect tumor cells carrying chromosome abnormalities in mixed cell populations contaminated with normal cells as tissue sections or cytological preparations. Recently, Martin-Subero and coworkers (2002) described the use of multicolor-FICTION, allowing the simultaneous detection of the morphological and immunophenotypic characteristics of neoplastic cells together with the most frequent chromosomal aberrations.

10.6 MULTICOLOR FISH SCREENING ASSAYS

For the simultaneous visualization of different chromosome aberrations in one experiment, a variety of multicolor FISH assays have been developed in the last decade (Liehr *et al.*, 2002; Tönnies, 2002). Some of the most important techniques for routine cytogenetics are described here (see Table 10.3).

10.6.1 CenM-FISH and CM-FISH

Recently, two all-human centromere-specific multicolor-FISH approaches for the subsequent determination of the exact origin of structurally abnormal, cytogenetically unidentifiable marker chromosomes have been reported by Henegariu and coworkers (2001a; CM-FISH) and Nietzel and coworkers (2001; cenM-FISH; Fig. 10.6a). These one-step multicolor FISH assays allow the simultaneous characterization of all human centromeres using differently labeled centromeric satellite DNA as probes for all human chromosomes. However, by using centromeric alpha

TABLE 10.3 Molecular cytogenetic multicolor whole metaphase/cell scanning techniques scanning used in human molecular diagnostics.

FISH-technique[a]	Probe setup	Hybridization target (patient material)	Aberration type detectable	Specific advantages of the technique	Diagnostic limitations
CenM-FISH and CM-FISH	All centromere specific α-satellite probes	Centromeres (metaphase spreads)	Numerical aberrations, identification of marker chromosome origin	Fast characterization of supernumerary marker chromosome origin, no need of specific cytogenetic expertise	No information about euchromatic content of the marker
TM-FISH, M-Tel FISH, S-COBRA	All subtelomeric regions	Subtelomeres (metaphase spreads)	Cryptic terminal deletions, duplications and translocations	High sensitivity, simultaneous analysis of all chromosome ends involved in rearrangements	Small analysis spectrum restricted to the chromosome ends, further analysis of affected chromosome(s) necessary
SKY/M-FISH and technical modifications (CCK-FISH, COBRA-FISH, IPM-FISH)	All human WCP-specific probes	Whole chromosomes (metaphase spreads)	Interchromosomal balanced and unbalanced translocations, euchromatic marker chromosome identification	Whole metaphase scanning without prior probe selection, translocation partner detection in one experiment, fast characterization of euchromatic marker chromosome content	Insensitive detection of intrachromosomal aberrations as deletions, duplications and inversions, imprecise breakpoint detection, need of non-overlapping chromosomes
Chromosomal bar codes and Rx-FISH	YAC clones, fragment hybrids, cross-species WCP mix	Whole chromosomes (metaphase spreads)	Inter- and intrachromosomal aberrations as translocations and gross deletions, duplications, inversions	Metaphase-wide detection of gross intra- and interchromosomal chromosome alterations and marker chromosomes	Sub-optimal locus-specific resolution of breakpoints, need of non-overlapping chromosomes
Multicolor-banding, MCB	Microdissection-derived partial chromosome probes	Whole chromosomes (metaphase spreads)	Inter- and intrachromosomal aberrations as translocations, deletions, and inversions	Sensitive whole karyotype detection chromosomal alterations with chromosome band specific resolution	High costs and technically demanding, need of non-overlapping chromosomes
Comparative genomic hybridization, CGH	Patient and control DNA	Whole control chromosomes (metaphase spreads)	Chromosomal inter- and intrachromosomal imbalances, gene amplifications, marker chromosomes	Whole genome scanning technique without need of proliferating patient material, locus-specific detection of gene amplification, band-specific information on imbalance size	Insensitive detection of imbalances in a subpopulation of cells, no detection of balanced aberrations, technically demanding
Matrix-CGH/ Array CGH	Patient and control DNA	Spotted defined DNA probes (cloned sequences or oligonucleotides)	Chromosomal inter- and intrachromosomal imbalances, gene amplifications, marker chromosomes	Whole genome scanning technique without need of proliferating patient material, high-resolution target-specific detection of gene amplification, submicroscopic information on imbalances	Direct dependency on the spotted targets, no detection of balanced translocations, missing of small mosaics

[a]*For abbreviations, see text.*

satellite probes, only the centromeric parts of chromosomes can be explored, whereas other chromosome abnormalities affecting the euchromatin are excluded from investigation by using these assays.

Since most markers originate from the acrocentric chromosomes, Langer and coworkers (2001) developed an acroM-FISH assay using a probe mix, which consists of painting probes and centromere probes for chromosomes 13/21, 14/22, and 15, and a probe specific for rDNA, each labeled with a specific combination of fluorochromes. Using this assay, the origin of approximately 80% of all markers can be identified by one hybridization assay.

10.6.2 Multicolor Assays for Subtelomeric Rearrangements

Special subtelomeric multicolor probe mixes were established to scan metaphase spreads for subtle aberrations at the gene-rich ends of the chromosomes in a scanning assay (see also section 10.5.4.3). Henegariu and coworkers (2001b) reported the analysis of 41 chromosome ends simultaneously using a multicolor hybridization assay including DNA probes located near the end of these chromosomes (0.1–1 Mb from the telomere, TM-FISH). In the same year, Brown and coworkers (2001) presented the so-called M-TEL 12-color FISH assay, which permits the screening of all telomeres in only two hybridizations. As an important parameter for successful hybridizations, all three techniques require well-spread metaphases (proliferating cells) with no cytoplasm.

10.7 MULTICOLOR WHOLE METAPHASE SCANNING TECHNIQUES

In addition to the various locus-specific assays described in this chapter (''direct assays''), which often require the hybridization of a whole repertoire of probes for different chromosomes successively to narrow down the overall composition of a chromosome aberration in single FISH experiments, some unbiased whole metaphase or chromosome scanning approaches (''indirect assays'') are established for the characterization of highly rearranged and unbalanced chromosome aberrations. Two major techniques, multi-fluor-FISH (M-FISH; Speicher *et al.*, 1996; Fig. 10.6b) and spectral karyotyping (SKY; Schröck *et al.*, 1996; Fig. 10.6c) were developed for the detection of non-homologous structural and numerical chromosomal aberrations on the single metaphase level hybridizing of a 24-color whole chromosome probe mix. Each homolog chromosome is displayed in a different color, based on the use of computer-generated false-color chromosome images and karyotyping. Although the principle of hybridizing whole chromosome probes labeled by five different fluorochromes in different combinations (combinatorial labeling) is the same, both approaches underlie a different technical concept: spectral karyotyping combines Fourier spectroscopy, CCD imaging, and optical microscopy to measure simultaneously at all points in the sample emission spectra in the visible and near-infrared spectral range (Schröck *et al.*, 1996). In contrast, M-FISH is based on classical epifluorescence microscopy using fluorochrome-specific optical filter sets for color discrimination (Speicher *et al.*, 1996). The fact that no prior knowledge of the affected chromosome(s) is required (''indirect analysis'') counts as an advantage for both techniques. By both approaches, the detection of chromosome rearrangements such as balanced

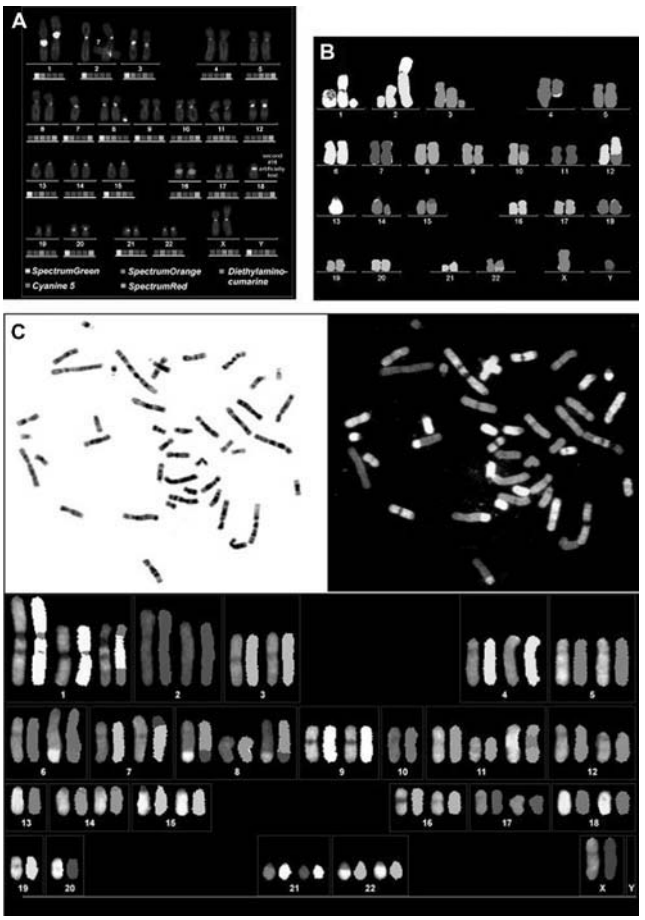


FIGURE 10.6 Different examples of multicolor FISH approaches. **A.** A supernumerary marker chromosome could be characterized as a derivative chromosome 8 using the cen-M-FISH approach (courtesy of Thomas Liehr, Institute of Human Genetics and Anthropology, Jena, Germany). **B.** M-FISH karyotype with complex chromosomal aberrations after radiation of the lymphocyte culture (courtesy of Christian Johannes, Essen, Germany. Hybridization and analysis carried out in collaboration with Metasystems, Germany). **C.** Spectral karyotyping (SKY) analysis of a human multiple myeloma cell-line (NCI-H929, CRL-9068, ATCC, Manassas, VA, USA). Several chromosomal aberrations were identified, e.g. translocations between chromosomes X and 7, 6 and 18, as well as 8 and 20, a dicentric chromosome 1 and a second dicentric chromosome 1 involved in a translocation or insertion with chromosome 10 (courtesy of Isabell Grandy and Evelin Schröck, Institute of Clinical Genetics, Medical Faculty Carl Gustav Carus, Technical University, Dresden, Germany).

or unbalanced translocations and euchromatic marker chromosomes in a single hybridization is possible, even if the chromosome morphology is poor. Therefore, both methods are used routinely in clinical and tumor genetics (McNeil and Ried, 2000; Schröck and Padilla-Nash, 2000; Bayani and Squire, 2002; Teixeira, 2002).

However, by using both techniques only the origin or constitution of an aberrant chromosome can be proved, whereas the exact locus or breakpoint assignment is not possible. Major drawbacks for routine use are the poor sensitivity to detect aberrations like intra-chromosomal deletions, duplications, and inversions. A large number of whole chromosome probes, labeled with different fluorochromes and the special technical equipment required, in particular for SKY analyses, restrict these scanning techniques to some specialized laboratories.

In 1999, Henegariu and coworkers presented a cost-effective alternative to SKY and M-FISH, called color-changing karyotyping (CCK). Compared to both techniques mentioned before, CCK uses only three fluorochromes to discriminate up to 41 DNA probes. The discrimination is done through the difference in signal strength between direct- and indirect-labeled chromosomes. Therefore this approach can be used with a conventional three-filter fluorescent microscope located in every molecular-cytogenetic lab. To achieve a better breakpoint resolution of translocations, Aurich-Costa and coworkers (2001) developed a multicolor karyotyping technique (IPM-FISH) that is based on the use of interspersed polymerase chain reaction generated (IRS-PCR) whole chromosome probes that display, additionally to the color karyotyping, an R-Band-like hybridization pattern.

10.8 MULTICOLOR CHROMOSOME BANDING TECHNIQUES

Until today, cytogenetics using black-and-white bar coding of chromosomes is still the most important and cost-effective standard method to identify chromosome aberrations. However, due to the similar size and shade of some chromosome bands, a detailed identification and chromosomal assignment of altered chromosome regions by conventional cytogenetic banding techniques alone is not always satisfactory. Additionally, a pronounced cytogenetic expertise is needed for aberration detection and characterization, which restricts the number of persons handling these methods. Over time, the idea came up to develop a color bar coding of chromosomes, which allows the detection of inter- and intrachromosomal aberrations as well as an automated karyotyping of the colored chromosomes. The first bar coding attempts, chromosomal bar codes (CBC; Lengauer *et al.*, 1993), were obtained by FISH with pools of Alu-PCR products from YAC clones containing human genomic DNA inserts.

Four years later, Müller and coworkers (1997) described their chromosome bar code for human chromosomes, based on the application of a set of subregional DNA probes (based on human/rodent somatic cell hybrids) that distinguishes each chromosome in a single FISH assay. The result was a multicolor set of 110 distinct signals per haploid chromosome set. Using flow sorted primate chromosomes of two gibbon species, Müller and coworkers (1998) also established the so-called cross-species color segmenting or Rx-FISH. With the exception of six chromosomes, all other human chromosomes can be differentiated by this hybridization assay in at least two and up to six segments. However, allowing the detection of inter- and intrachromosomal aberrations, the resolution of classical banding patterns (approximately 400 to 600 bands/haploid genome) cannot be reached by these approaches (e.g. for breakpoint detection).

High-resolution multicolor banding (MCB) for refined FISH analysis of human chromosomes on the band and subband level was introduced by Chudoba and coworkers (1999) (see Figs 10.7a and b). This technique is based on changing fluorescence intensity ratios of overlapping DNA probes, labeled by five different fluorochromes. After computer-based assignment of distinctive pseudocolors to the overlapping and non-overlapping hybridization signals along the chromosome, this approach allows a higher resolution than the former mentioned, independent of chromosome condensation. By using the MCB technique for single chromosomes, preferentially intrachromosomal aberrations as deletions and peri- or paracentric inversions can be characterized.

Mrasek and coworkers (2001) reported the use of human MCB probes for all chromosomes, permitting the straightforward characterization not only of intrachromosomal aberrations but also of interchromosomal translocations and breakpoint mapping (see Fig. 10.7c). By using 138 microdissection derived MCB probes, a resolution of 450 bands and more can be achieved, enabling a high-resolution FISH banding for the detection of complex chromosome rearrangements (Liehr and Claussen, 2002). Recently, Weise and coworkers (2008) presented the first BAC array-mapped genome-wide human MCB probe set including 169 region-specific microdissection libraries characterized in detail for their size and overlap.

10.9 WHOLE GENOME SCANNING AND COMPARATIVE GENOMIC HYBRIDIZATION

All metaphase or chromosome scanning techniques utilize metaphases of the patient to uncover chromosomal aberrations. In tumor genetics, especially in solid tumor samples, the number of metaphases that can be prepared from these tissues is often small, if available at all. Additionally, the chromosome morphology is often poor so that karyotyping is barely possible.

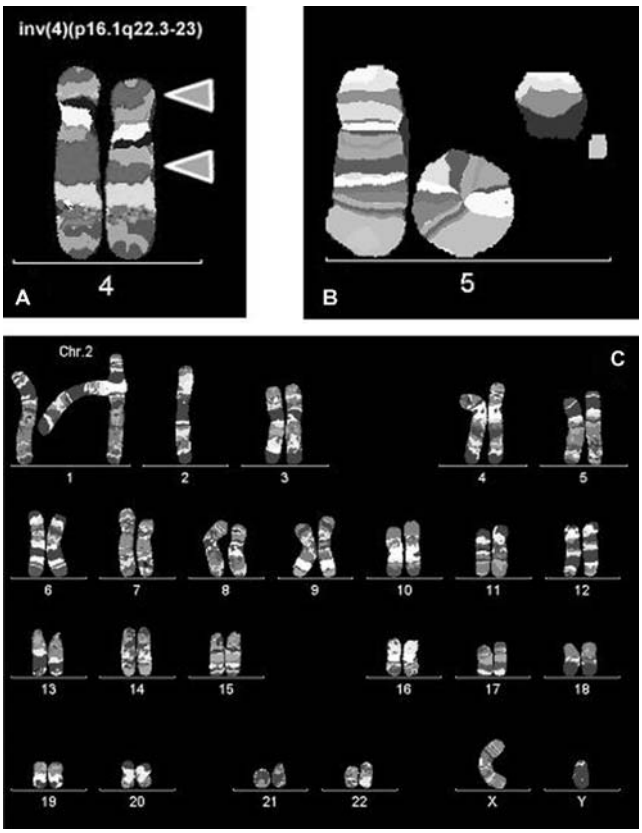


FIGURE 10.7 High-resolution multicolor banding (MCB) for refined FISH analysis of human chromosomes on the band and subband level. **A.** MCB analysis showing a pericentric inversion (including the centromere) of chromosome 4. **B.** MCB analysis of a ring chromosome 5 and a derivative chromosome 5 induced by radiation of lymphocyte cultures. **C.** Example of the use of human MCB probes for all chromosomes simultaneously, permitting the straightforward characterization not only of intrachromosomal aberrations but also of interchromosomal translocations. By using 138 microdissection-derived MCB probes a resolution of 450 bands and more can be achieved enabling a high-resolution FISH banding for the detection of complex chromosome rearrangements (pictures courtesy of Thomas Liehr and Christian Johannes, Institute of Human Genetics and Anthropology, Jena, Germany).

To circumvent these problems, Kallioniemi and coworkers (1992) established a hybridization approach, known as comparative genomic hybridization (CGH), which relies on the use of genomic tumor DNA circumventing difficult cell culture and chromosome preparations of the tumor tissue. CGH is a potent and reliable hybridization approach, linking conventional cytogenetic and molecular genetic techniques. It allows the comprehensive analysis of the entire genome from tissue samples in just one experiment providing global information about chromosomal imbalances and gene amplifications in the test genome (Tönnies *et al.*, 2001a, 2003b; Tönnies and Lage, 2004; Fischer *et al.*, 2004). In brief, CGH is based on the co-hybridization of differentially labeled whole genomic test (tumor) and normal (control or reference) DNA in a ratio of 1:1 to normal control (sex-matched) metaphase spreads (see Fig. 10.8). Fluorochrome-labeled test- and control-DNA probes compete for hybridization on the target chromosomes. Due to deviations in fluorescence ratios of test-DNA to control-DNA copy number changes cannot be detected by visual inspection alone (see Fig. 10.9a). Therefore, software-assisted image capturing, karyotyping, and quantification of fluorescence intensities over the entire length of each chromosome is an integral part of this technique. The endpoint of a CGH analysis is a so-called copy number karyotype displaying all chromosomal imbalances on a conventional cytogenetic resolution level (approximately 5–10 Mb; Fig. 10.9b).

To enhance sensitivity and resolution of CGH, Kirchhoff and coworkers (1998) reported on a high-resolution CGH approach based on a modified CGH software. These authors applied standard reference intervals as detection criteria and detected deletions down to 3 Mb (Kirchhoff *et al.*, 1999). CGH-based investigations provide information not only on the chromosomal assignment of a chromosomal imbalance but also on chromosomal band-specific origin. This approach allows studying DNA from any human source, even though the cells are not viable (e.g. paraffin-embedded tissue sections), or if DNA amount is limited (see also section 10.5.3). Established in tumor genetics, a number of reports describing pre- and postnatal cytogenetic cases are available in the literature (Bryndorf *et al.*, 1995; Levy *et al.*, 1998; Tönnies *et al.*, 2001a, c, d; 2003c, d). Furthermore, even in preimplantation genetic diagnostics, first experimental results have been reported, revealing chromosomal imbalances in blastomeres or polar body cells using DNA of single cells (Malmgren *et al.*, 2002; Wells *et al.*, 2002).

Nevertheless, by CGH the detection of balanced chromosome aberrations, such as reciprocal translocations often found in liquid tumors (leukemias), is not possible. Furthermore, using whole genomic DNA, the information is restricted to the total amount of cells, of which the DNA is extracted. Therefore, if only a small number of cells are affected by a chromosomal imbalance (heterogeneity in tumors, clonal aberrations), it will be missed due to contamination of the DNA with normal cells. However, the analysis of single cells is possible by upstream DNA amplification of single cell DNA as has been shown by Klein and coworkers (1999, 2002).

10.10 MOLECULAR KARYOTYPING – ARRAY CGH

Molecular cytogenetic FISH technologies, as described in this chapter, are based on the combination of molecular genetic and cytogenetic techniques detecting complementary nucleic acid sequences *in situ* on metaphase

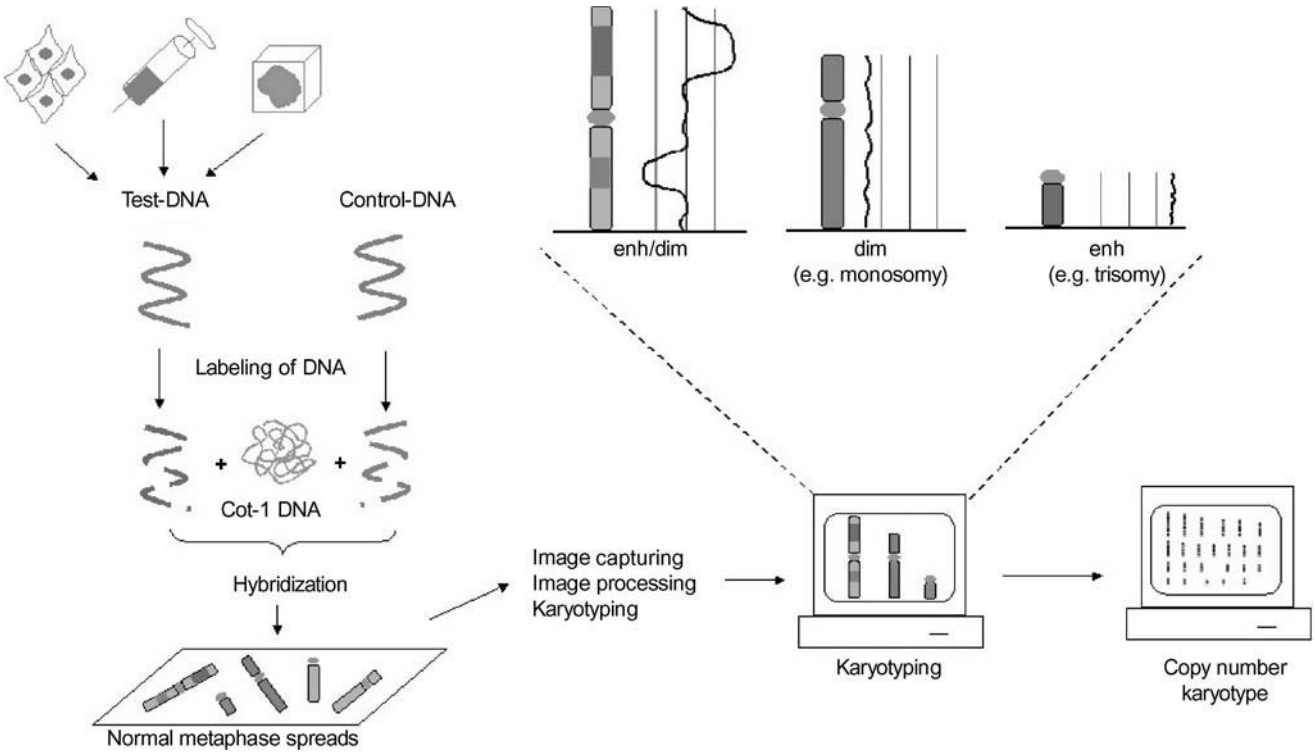


FIGURE 10.8 Simplified flowchart summarizing the steps of CGH technique. Different labeled control and test DNAs extracted from either tissue cultures, whole blood, or paraffin embedded material are hybridized under suppression conditions (Cot-1 DNA) on normal metaphase spreads. After image capturing and karyotyping of the metaphase chromosomes, fluorescence ratio profiles (quotient of test- to control-fluorescence intensity) for each chromosome are calculated. Mean profiles are plotted against the length of each chromosome. The center line (black) in the CGH profiles represents the balanced state of the chromosomal copy number (ratio = 1.0). An upper threshold (green) is used to define a gain of chromosomal material (ENH = enhanced), and a lower threshold (red) is used to interpret a loss of chromosomal material (DIM = diminished). CGH results are displayed in the so-called copy number karyotype. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

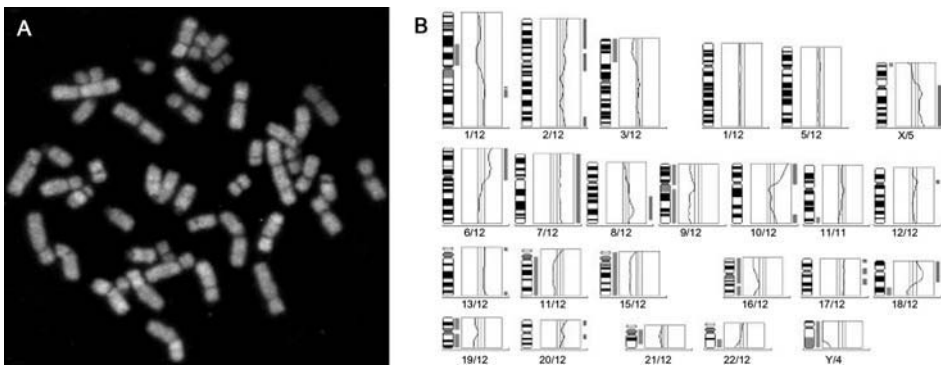


FIGURE 10.9 Comparative genomic hybridization of DNA extracted from ovarian cancer tissue. **A.** Metaphase spread hybridized with normal control DNA (red), tumor DNA (green), and counterstained with DAPI. A high number of fluorescence ratio differences are visible by eye inspection. **B.** Copy-number karyotype of the same CGH analysis. Nearly all chromosomes are showing full (e.g. chromosome 7) or partial (e.g. chromosome 6) imbalances. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

chromosomes or in the cell nuclei. However, the resolution level, as described for CGH, is restricted to approximately 5–10 Mb. To be independent from high-quality metaphases and to enhance the resolution of analyses, Solinas-Toldo and coworkers (1997) established a matrix-based CGH array (Matrix-CGH) that substitutes normal metaphase chromosomes as hybridization target by well-defined genomic DNA fragments arrayed on a solid support, allowing automated analysis of genetic imbalances as small deletions or gene amplifications (Pinkel *et al.*, 1998; Pollack

et al., 1999; Fig. 10.10). In 2001, Snijders and coworkers established a microarray using 2,400 BAC inserts covering nearly the whole human genome for the detection of gains and losses in diploid, polyploid, and heterogeneous backgrounds. Veltman and coworkers (2002) described a DNA microarray spotted with a set of 77 human chromosome specific subtelomeric probes as a target for array-based CGH. The dynamic range of the signal ratios obtained by these non-cytogenetic assays is up to five times higher for matrix CGH in comparison to conventional CGH, allowing for better quantitative assessments of genetic imbalances (Wessendorf *et al.*, 2002). Amplifications and deletions of small genomic regions not detectable by chromosome ana lyses or CGH can be recognized with high resolution directly depending on the number of human DNA fragments spotted on the chip. By the use of 3,500 inserts, a resolution of 1 Mb can easily be obtained (Coe *et al.*, 2007). The robustness and simplicity of these microarrays make them highly suited for routine diagnostic use (Veltman, 2006; Vermeesch *et al.*, 2007).

However, BAC inserts are typically large (80–200 kb) and therefore single-copy number aberrations smaller than 50 kb are not detectable with certainty. Shorter targets in higher numbers can be achieved by using oligonucleotides as targets (25–80 bp). Oligonucleotides synthesized directly onto the matrix offers the possibility to hybridize on 244,000 or 385,000 targets, depending on the company from which the array is ordered. Next generation oligonucleotide arrays have up to 2.1 or 4.2 million targets. Using custom-designed oligonucleotide arrays, an almost nucleotide-level resolution can be achieved by synthesizing overlapping oligonucleotides with single base-pair shifts. Other platforms have been constructed initially for single nucleotide polymorphism (SNP) studies. SNP arrays have been developed with 20 matched and mismatched probe pairs for each SNP allele. Hybridizations of test DNA with reduced complexity due to restriction enzyme digestion and adapter-ligation PCR are performed. SNP arrays allow an overview of chromosomal imbalances as well as loss of heterozygosity (LOH) and uniparental disomies (UPD) to be obtained. The signal intensities of the matched probes are compared to *in silico* data sets. A current 500k GeneChip shows a median spacing of 2.5 kb. Other arrays, using 50 bp oligonucleotides are based on 650,000 different oligonucleotides with a median spacing of 2 kb. However, the targets of SNP arrays are not uniformly distributed across the genome, in comparison to commercial oligonucleotide CGH arrays.

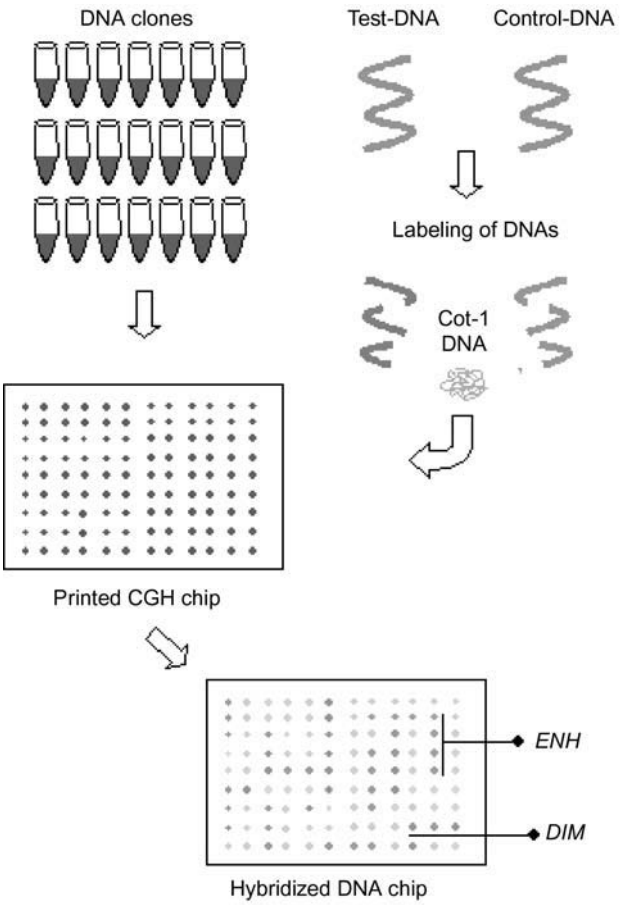


FIGURE 10.10 Schematic illustration of the matrix-comparative genomic hybridization procedure. Cloned DNAs reflecting the human genome or parts thereof are printed on a CGH microarray and serve as a target for labeled control and test DNAs. After hybridization, the resulting relative fluorescence intensities are measured by computer-supported scanning. DNA loss in the test DNA results in red spots, gains in green spots. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

10.11 CONCLUSIONS AND PERSPECTIVES

Conventional and molecular cytogenetic approaches are important to characterize chromosomal aberrations deeply, and can distinguish between a simple gain or loss of genetic material and more complex aberrations as insertions and cryptic abnormalities. The cytogenetic knowledge based on chromosomes and parts thereof has important impact on genetic counseling and further clinical management. Therefore, both conventional and molecular cytogenetics will have important functions in the future of medical diagnostics. Recent advances, such as improved hybridization protocols and strategies, novel commercial and non-commercial probe sets, coupled with hardware development such as image analysis software and specific filter sets for newly established fluorochromes, facilitate the collection of new important data for the understanding and diagnosis of genetic diseases.

Recently, whole genome array techniques have emerged that give the opportunity to investigate the whole genome with high resolution in one experiment. Initially extremely expensive, this technique will eventually oust other molecular cytogenetic approaches due to its cost effectiveness

and wider acceptability. Directly transmitted imbalanced euchromatic variants without phenotypic effects have been detected by conventional cytogenetics and molecular cytogenetic methods for several years (for an overview see Barber, 2005). Using array platforms, a significant fraction of the genome has been identified to vary in copy number between apparently healthy individuals (Sebat *et al.*, 2004; Iafrate *et al.*, 2004; Redon *et al.*, 2006). Current and near future challenges are to interpret these imbalances with caution, using several arguments including parental genotype, verification methods (mostly molecular cytogenetic methods or PCR-based approaches) and last but not least the patients' phenotype. Furthermore, one has to consider that a chromosomal abnormality may also be of clinical significance even though it has been inherited from a healthy parent because more complex mechanisms such as phenotypic variation, incomplete penetrance, imprinting, position effect or even point mutation of a recessive gene may be involved. The amount of data mined today by different high-resolution molecular cytogenetic techniques requires a close interaction between the clinician, the genetic counselor and last, but not least, the cytogeneticist using this palette of techniques. However, molecular cytogenetics has not been able to replace conventional cytogenetics – therefore the new microarray-based techniques are expected not to completely replace molecular cytogenetics in the near future of modern molecular diagnostics.

REFERENCES

Anderlid, B.M., Schoumans, J., Anneren, G., Sahlen, S., Kyllerman, M., Vujic, M., Hagberg, B., Blennow, E., and Nordenskjold, M. (2002). Subtelomeric rearrangements detected in patients with idiopathic mental retardation. *Am. J. Med. Genet.* 107, 275–284.

Arnold, J. (1879). Beobachtungen über Kerntheilungen in den Zellen der Geschwülste. *Virchows Archiv* 78, 279–301.

Aurich-Costa, J., Vannier, A., Gregoire, E., Nowak, F., and Cherif, D. (2001). IPM-FISH, a new M-FISH approach using IRS-PCR painting probes: application to the analysis of seven human prostate cell lines. *Genes Chromosomes Cancer* 30, 143–160.

Barber, J.C. (2005). Directly transmitted unbalanced chromosome abnormalities and euchromatic variants. *J. Med. Genet.* 42, 609–629.

Barch, M.J., Knutsen, T., and Spurbeck, J.L. (1997). *The AGT Cytogenetics Laboratory Manual*. Lippincott-Raven, Philadelphia.

Bauman, J.G., Wiegant, J., Borst, P., and van Duijn, P. (1980). A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome labelled RNA. *Exp. Cell. Res.* 128, 485–490.

Bayani, J.M., and Squire, J.A. (2002). Applications of SKY in cancer cytogenetics. *Cancer Invest* 20, 373–386.

Blennow, E., Nielsen, K.B., Telenius, H., Carter, N.P., Kristoffersson, U., Holmberg, E., Gillberg, C., and Nordenskjold, M. (1995). Fifty probands with extra structurally abnormal chromosomes characterized by fluorescence in situ hybridization. *Am. J. Med. Genet.* 55, 85–94.

Brown, J., Saracoglu, K., Uhrig, S., Speicher, M.R., Eils, R., and Kearney, L. (2001). Subtelomeric chromosome rearrangements are detected using an innovative 12-color FISH assay (M-TEL). *Nat. Med.* 7, 497–501.

Bryndorf, T., Kirchhoff, M., Rose, H., Maahr, J., Gerdes, T., Karhu, R., Kallioniemi, A., Christensen, B., Lundsteen, C., and Philip, J. (1995). Comparative genomic hybridization in clinical cytogenetics. *Am. J. Hum. Genet.* 57, 1211–1220.

Buckton, K.E., Spowart, G., Newton, M.S., and Evans, H.J. (1985). Forty-four probands with an additional marker chromosome. *Hum. Genet.* 69, 353–370.

Casperson, T., Zech, L., and Johhansson, C. (1970). Differential banding of alkylating fluorochromes in human chromosomes. *Exp. Cell. Res.* 60, 315–319.

Chudoba, I., Plesch, A., Lorch, T., Lemke, J., Claussen, U., and Senger, G. (1999). High resolution multicolor-banding: a new technique for refined FISH analysis of human chromosomes. *Cytogenet. Cell Genet.* 84, 156–160.

Cinti, C., Santi, S., and Maraldi, N.M. (1993). Localization of single copy gene by PRINS technique. *Nucleic Acids Res.* 21, 5799–5800.

Coe, B.P., Ylstra, B., Carvalho, B., Meijer, G.A., Macaulay, C., and Lam, W.L. (2007). Resolving the resolution of array CGH. *Genomics* 89, 647–653.

Cremer, T. (1985). *Von der Zellenlehre zur Chromosomentheorie*. Springer Verlag, Berlin, Heidelberg, New York, Tokyo.

Cremer, T., Landegent, J., Bruckner, A., Scholl, H.P., Schardin, M., Hager, H.D., Devilee, P., Pearson, P., and van der Ploeg, M. (1986). Detection of chromosome aberrations in the human interphase nucleus by visualization of specific target DNAs with radioactive and non-radioactive in situ hybridization techniques: diagnosis of trisomy 18 with probe L1.84. *Hum. Genet.* 74, 346–352.

Cremer, T., Popp, S., Emmerich, P., Lichter, P., and Cremer, C. (1990). Rapid metaphase and interphase detection of radiation-induced chromosome aberrations in human lymphocytes by chromosomal suppression in situ hybridization. *Cytometry* 11, 110–118.

Dauwerse, J.G., Wiegant, J., Raap, A.K., Breuning, M.H., and van Ommen, G.J. (1992). Multiple colors by fluorescence in situ hybridization using ratio-labelled DNA probes create a molecular karyotype. *Hum. Mol. Genet.* 1, 593–598.

Deaven, L.L., Van Dilla, M.A., Bartholdi, M.F., Carrano, A.V., Cram, L.S., Fuscoe, J.C., Gray, J.W., Hildebrand, C.E., Moyzis, R.K., and Perlman, J. (1986). Construction of human chromosome-specific DNA libraries from flow-sorted chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* 51, 159–167.

Delhanty, J.D., Griffin, D.K., Handyside, A.H., Harper, J., Atkinson, G.H., Pieters, M.H., and Winston, R.M. (1993). Detection of aneuploidy and chromosomal mosaicism in human embryos during preimplantation sex determination by fluorescent in situ hybridisation, (FISH). *Hum. Mol. Genet.* 2, 1183–1185.

Feinberg, A.P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132, 6–13.

Fischer, T.C., Gellrich, S., Muche, M., Sherev, T., Audring, H., Neitzel, H., Walden, P., Sterry, W., and Tönnies, H. (2004). Genomic aberrations and survival in cutaneous T cell lymphomas. *J. Invest. Dermatol.* 122, 579–586.

Flemming, W. (1881). Beiträge zur Kenntnis der Zelle und ihrer Lebenserscheinungen. 3. Teil. *Archiv für mikroskopische Anatomie* 20, 1–86.

Fragouli, E. (2007). Preimplantation genetic diagnosis: present and future. *J. Assist. Reprod. Genet.* 24, 201–207.

Gardner, R., and Sutherland, G. (2003). *Chromosome Abnormalities and Genetic Counseling*. Oxford University Press.

Grand, F.H., Chase, A., Iqbal, S., Nguyen, D.X., Lewis, J.L., Marley, S.B., Davidson, R.J., Goldman, J.M., and Gordon, M.Y. (1998). A two-color BCR-ABL probe that greatly reduces the false positive and false negative rates for fluorescence in situ hybridization in chronic myeloid leukemia. *Genes Chromosomes Cancer* 23, 109–115.

Guan, X.Y., Trent, J.M., and Meltzer, P.S. (1993). Generation of band-specific painting probes from a single microdissected chromosome. *Hum. Mol. Genet.* 2, 1117–1121.

Hansemann, D. (1890). Über assymmetrische Zellteilung in Epithelkrebsen und deren biologische Bedeutung. *Arch. Pathol. Anat.* 119, 299–326.

Harper, M.E., and Saunders, G.F. (1981). Localization of single copy DNA sequences of G-banded human chromosomes by in situ hybridization. *Chromosoma* 83, 431–439.

Henegariu, O., Heerema, N.A., Bray-Ward, P., and Ward, D.C. (1999). Colour-changing karyotyping: an alternative to MFISH/ SKY. *Nat. Genet.* 23, 263–264.

Henegariu, O., Bray-Ward, P., Artan, S., Vance, G.H., Qumsyieh, M., and Ward, D.C. (2001a). Small marker chromosome identification in metaphase and interphase using centromeric multiplex fish (CM-FISH). *Lab. Invest* 81, 475–481.

Henegariu, O., Artan, S., Greally, J.M., Chen, X.N., Korenberg, J.R., Vance, G.H., Stubbs, L., Bray-Ward, P., and Ward, D.C. (2001b). Cryptic translocation identification in human and mouse using several telomeric multiplex fish (TM-FISH) strategies. *Lab. Invest* 81, 483–491.

Heng, H.H., Squire, J., and Tsui, L.C. (1992). High-resolution mapping of mammalian genes by in situ hybridization to free chromatin. *Hum. Mol. Genet.* 1, 587–591.

Hindkjaer, J., Brandt, C.A., Koch, J., Lund, T.B., Kolvraa, S., and Bolund, L. (1995). Simultaneous detection of centromere-specific probes and chromosome painting libraries by a combination of primed in situ labelling and chromosome painting (PRINS-painting). *Chromosome Res.* 3, 41–44.

Hsu, T.C., and Pomerat, C.M. (1953). Mammalian chromosomes in vitro. II. A method for spreading the chromosomes of cells in tissue culture. *J. Hered.* 44, 23–29.

Hulten, M.A., Gould, C.P., Goldman, A.S., and Waters, J.J. (1991). Chromosome in situ suppression hybridisation in clinical cytogenetics. *J. Med. Genet.* 28, 577–582.

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. *Nat. Genet.* 36, 949–951.

ISCN (2005). *An International System for Human Cytogenetic Nomenclature* (Mitelman, F., ed.). S. Karger, Basel.

John, H., Birnstiel, M., and Jones, K. (1969). RNA-DNA hybrids at the cytological level. *Nature* 223, 582–587.

Kadandale, J.S., Tunca, Y., and Tharapel, A.T. (2000). Chromosomal localization of single copy genes SRY and SOX3 by primed in situ labeling (PRINS). *Microb. Comp. Genomics* 5, 71–74.

Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F.M., and Pinkel, D. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258, 818–820.

Kirchhoff, M., Gerdes, T., Rose, H., Maahr, J., Ottesen, A.M., and Lundsteen, C. (1998). Detection of chromosomal gains and losses in comparative genomic hybridization analysis based on standard reference intervals. *Cytometry* 31, 163–173.

Kirchhoff, M., Gerdes, T., Maahr, J., Rose, H., Bentz, M., Dohner, H., and Lundsteen, C. (1999). Deletions below 10 megabasepairs are detected in comparative genomic hybridization by standard reference intervals. *Genes Chromosomes Cancer* 25, 410–413.

Klein, C.A., Schmidt-Kittler, O., Schardt, J.A., Pantel, K., Speicher, M.R., and Riethmüller, G. (1999). Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc. Natl. Acad. Sci. USA* 96, 4494–4499.

Klein, C.A., Seidl, S., Petat-Dutter, K., Offner, S., Geigl, J.B., Schmidt-Kittler, O., Wendler, N., Passlick, B., Huber, R.M., Schlimok, G., Baeuerle, P.A., and Riethmüller, G. (2002). Combined transcriptome and genome analysis of single micrometastatic cells. *Nat. Biotechnol.* 20, 387–392.

Klinger, K., Landes, G., Shook, D., Harvey, R., Lopez, L., Locke, L., Osathanondh, R., Leverone, B., Houseal, T., Pavelka, K., and Dackowski, W. (1992). Rapid detection of chromosome aneuploidies in uncultured amniocytes by using fluorescence in situ hybridization (FISH). *Am. J. Hum. Genet.* 51, 55–65.

Knight, S.J., Horsley, S.W., Regan, R., Lawrie, N.M., Maher, E.J., Cardy, D.L., Flint, J., and Kearney, L. (1997). Development and clinical application of an innovative fluorescence in situ hybridization technique which detects submicroscopic rearrangements involving telomeres. *Eur. J. Hum. Genet.* 5, 1–8.

Knight, S.J., Regan, R., Nicod, A., Horsley, S.W., Kearney, L., Homfray, T., Winter, R.M., Bolton, P., and Flint, J. (1999). Subtle chromosomal rearrangements in children with unexplained mental retardation. *Lancet* 354, 1677–1681.

Knoll, J.H., and Rogan, P.K. (2003). Sequence-based, in situ detection of chromosomal abnormalities at high resolution. *Am. J. Med. Genet.* 121A, 245–257.

Knuutila, S., and Teerenhovi, L. (1989). Immunophenotyping of aneuploid cells. *Cancer Genet. Cytogenet.* 41, 1–17.

Koch, J.E., Kolvraa, S., Petersen, K.B., Gregersen, N., and Bolund, L. (1989). Oligonucleotide-priming methods for the chromosome-specific labelling of alpha satellite DNA in situ. *Chromosoma* 98, 259–265.

Koch, J., Hindkjaer, J., Mogensen, J., Kolvraa, S., and Bolund, L. (1991). An improved method for chromosome-specific labeling of alpha satellite DNA in situ by using denatured doublestranded DNA probes as primers in a primed in situ labeling (PRINS). procedure. *Genet. Anal. Tech. Appl.* 8, 171–178.

Krejci, K., and Koch, J. (1998). Improved detection and comparative sizing of human chromosomal telomeres in situ. *Chromosoma* 107, 198–203.

Kuo, W.L., Tenjin, H., Segraves, R., Pinkel, D., Golbus, M.S., and Gray, J. (1991). Detection of aneuploidy involving chromosomes 13, 18, or 21, by fluorescence in situ hybridization (FISH) to interphase and metaphase amniocytes. *Am. J. Hum. Genet.* 49, 112–119.

Langer, S., Fauth, C., Rocchi, M., Murken, J., and Speicher, M.R. (2001). AcroM fluorescent in situ hybridization analyses of marker chromosomes. *Hum. Genet.* 109, 152–158.

Lansdorp, P.M., Verwoerd, N.P., van de Rijke, F.M., Dragowska, V., Little, M.-T., Dirks, R.W., Raap, A.K., and Tanke, H.J. (1996). Heterogeneity in telomere length of human chromosomes. *Hum. Mol. Genet.* 5, 685–691.

Lejeune, J., Turpin, R., and Gautier, M. (1959). Chromosomic diagnosis of mongolism. *Arch. Fr. Pediatr.* 16, 962–963.

Lengauer, C., Speicher, M.R., Popp, S., Jauch, A., Taniwaki, M., Nagaraja, R., Riethman, H.C., Donis-Keller, H., D'Urso, M., and Schlessinger, D.

(1993). Chromosomal bar codes produced by multicolor fluorescence in situ hybridization with multiple YAC clones and whole chromosome painting probes. *Hum. Mol. Genet.* 2, 505–512.

Levy, B., Dunn, T.M., Kaffe, S., Kardon, N., and Hirschhorn, K. (1998). Clinical applications of comparative genomic hybridization. *Genet. Med.* 1, 4–12.

Liehr, T., Heller, A., Starke, H., Rubtsov, N., Trifonov, V., Mrasek, K., Weise, A., Kuechler, A., and Claussen, U. (2002). Microdissection based high resolution multicolor banding for all 24 human chromosomes. *Int. J. Mol. Med.* 9, 335–339.

Liehr, T., and Claussen, U. (2002). Multicolor-FISH approaches for the characterization of human chromosomes in clinical genetics and tumor cytogenetics. *Curr. Genomics* 3, 213–235.

Malmgren, H., Sahlen, S., Inzunza, J., Aho, M., Rosenlund, B., Fridstrom, M., Hovatta, O., Ahrlund-Richter, L., Nordenskjold, M., and Blennow, E. (2002). Single cell CGH analysis reveals a high degree of mosaicism in human embryos from patients with balanced structural chromosome aberrations. *Mol. Hum. Reprod.* 8, 502–510.

Martin-Subero, J.I., Chudoba, I., Harder, L., Gesk, S., Grote, W., Novo, F.J., Calasanz, M.J., and Siebert, R. (2002). Multicolor-FICTION: expanding the possibilities of combined morphologic, immunophenotypic, and genetic single cell analyses. *Am. J. Pathol.* 161, 413–420.

McNeil, N., and Ried, T. (2000). Novel molecular cytogenetic techniques for identifying complex chromosomal rearrangements: technology and applications in molecular medicine. *Expert Rev. Mol. Med.* 14, 1–14.

Meltzer, P.S., Guan, X.Y., Burgess, A., and Trent, J.M. (1992). Rapid generation of region specific probes by chromosome microdissection and their application. *Nat. Genet.* 1, 24–28.

Moorhead, P.S., Nowell, P.C., Mellman, W.J., Battips, D.M., and Hungerford, D.A. (1960). Chromosome preparations of leukocytes cultured from human peripheral blood. *Exp. Cell. Res.* 20, 613–616.

Mrasek, K., Heller, A., Rubtsov, N., Trifonov, V., Starke, H., Rocchi, M., Claussen, U., and Liehr, T. (2001). Reconstruction of the female gorilla karyotype using 25-color FISH and multicolor banding (MCB). *Cytogenet. Cell Genet.* 93, 242–248.

Müller, S., Rocchi, M., Ferguson-Smith, M.A., and Wienberg, J. (1997). Toward a multicolour chromosome bar code. *Hum. Genet.* 100, 271–278.

Müller, S., O'Brien, P.C., Ferguson-Smith, M.A., and Wienberg, J. (1998). Cross-species colour segmenting: a novel tool in human karyotype analysis. *Cytometry* 33, 445–452.

Nederlof, P.M., van der Flier, S., Wiegant, J., Raap, A.K., Tanke, H.J., Ploem, J.S., and van der Ploeg, M. (1990). Multiple fluorescence in situ hybridization. *Cytometry* 11, 126–131.

Nederlof, P.M., van der Flier, S., Vrolijk, J., Tanke, H.J., and Raap, A.K. (1992). Fluorescence ratio measurements of doublelabeled probes for multiple in situ hybridization by digital imaging microscopy. *Cytometry* 13, 839–845.

Nietzel, A., Rocchi, M., Starke, H., Heller, A., Fiedler, W., Wlodarska, I., Loncarevic, I.F., Beensen, V., Claussen, U., and Liehr, T. (2001). A new multicolor-FISH approach for the characterization of marker chromosomes: centromere-specific multicolor-FISH (cenM-FISH). *Hum. Genet.* 108, 199–204.

Novell, P.C., and Hungerford, D.A. (1960). A minute chromosome in human chronic granulocyte leukaemia. *Science* 132, 1497.

Pardue, M.L., and Gall, J.G. (1969). Molecular hybridization of radioactive DNA to the DNA of cytological preparations. *Proc. Natl. Acad. Sci. USA* 64, 600–604.

Pinkel, D., Straume, T., and Gray, J.W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. USA* 83, 2934–2938.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W., and Albertson, D.G. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* 20, 207–211.

Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., and Brown, P.O. (1999). Genome-wide analysis of DNA copy number changes using cDNA microarrays. *Nat. Genet.* 23, 41–46.

Rautenstrauß, B.W., and Liehr, T. (2002). *FISH Technology. Springer Lab Manual*. Springer Verlag, Berlin, Heidelberg.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., and Hurles, M.E. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454.

Rigby, P.W., Dieckmann, M., Rhodes, C., and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with I. *J. Mol. Biol.* 113, 237–251.

Rogan, P.K., Cazcarro, P.M., and Knoll, J.H. (2001). Sequencebased design of single-copy genomic DNA probes for fluorescence in situ hybridization. *Genome Res.* 11, 1086–1094.

Saccone, S., De Sario, A., Della Valle, G., and Bernardi, G. (1992). The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. *Proc. Natl. Acad. Sci. USA* 89, 4913–4917.

Schinzel, A. (2001). *Catalogue of Unbalanced Chromosome Aberrations in Man*, 2nd ed. De Gruyter, Berlin, New York.

Schröck, E., du Manoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M.A., Ning, Y., Ledbetter, D.H., Bar-Am, I., Soenksen, D., Garini, Y., and Ried, T. (1996). Multicolor spectral karyotyping of human chromosomes. *Science* 273, 494–497.

Schröck, E., and Padilla-Nash, H. (2000). Spectral karyotyping and multicolor fluorescence in situ hybridization reveal new tumor-specific chromosomal aberrations. *Sem. Hematol.* 37, 334–347.

Schwarzacher, T., and Heslop-Harrison, P. (2000). *Practical in Situ Hybridisation*. BIOS Scientific Publishers Limited, Oxford.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A., and Wigler, M. (2004). Large-scale copy number polymorphism in the human genome. *Science* 305, 525–528.

Seong, D.C., Kantarjian, H.M., Ro, J.Y., Talpaz, M., Xu, J., Robinson, J.R., Deisseroth, A.B., Champlin, R.E., and Siciliano, M.J. (1995). Hypermetaphase fluorescence in situ hybridization for quantitative monitoring of Philadelphia chromosome-positive cells in patients with chronic myelogenous leukemia during treatment. *Blood* 86, 2343–2349.

Shaffer, L.G., and Lupski, J.R. (2000). Molecular mechanisms for constitutional chromosomal rearrangements in humans. *Annu. Rev. Genet.* 34, 297–329.

Shay, J.W., Werbin, H., and Wright, W.E. (1994). Telomere shortening may contribute to aging and cancer: a perspective. *Mol. Cell. Differ.* 2, 1–21.

Snijders, A.M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J.P., Gray, J.W., Jain, A.N., Pinkel, D., and Albertson, D.G. (2001). Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.* 29, 263–264.

Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T., and Lichter, P. (1997). Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 20, 399–407.

Speicher, M.R, and Carter, N.P. (2005). The new cytogenetics: blurring the boundaries with molecular biology. *Nat. Rev. Genet.* 6, 782–792.

Speicher, M.R., Ballard, S.G., and Ward, D. (1996). Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat. Genet.* 12, 368–375.

Stumm, M., Wegner, R.D., Bloechle, M., and Eckel, H. (2006). Interphase M-FISH applications using commercial probes in prenatal and PGD diagnostics. *Cytogenet. Genome Res.* 114, 296–301.

Tanke, H.J., Wiegant, J., van Gijlswijk, R.P., Bezrookove, V., Pattenier, H., Heetebrij, R.J., Talman, E.G., Raap, A.K., and Vrolijk, J. (1999). New strategy for multi-colour fluorescence in situ hybridisation: COBRA: COmbined Binary RAtio labelling. *Eur. J. Hum. Genet.* 7, 2–11.

Teixeira, M.R. (2002). Combined classical and molecular cytogenetic analysis of cancer. *Eur. J. Cancer* 38, 1580–1584.

Telenius, H., Carter, N.P., Nordenskjold, M., Ponder, B.A., and Yunnacliffe, A. (1992). Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13, 718–725.

Tharapel, A.T., Kadandale, J.S., Martens, P.R., Wachtel, S.S., and Wilroy, R.S., Jr. (2002). Prader Willi/Angelman and DiGeorge/velocardiofacial syndrome deletions: diagnosis by primed in situ labeling (PRINS). *Am. J. Med. Genet.* 107, 119–122.

Tijo, J.H., and Levan, A. (1956). The chromosomes of man. *Hereditas* 42, 1–6.

Tönnies, H., Stumm, M., Wegner, R.D., Chudoba, I., Kalscheuer, V., and Neitzel, H. (2001a). Comparative genomic based strategy for the analysis of different chromosome imbalances detected in conventional cytogenetic diagnostics. *Cytogenet. Cell. Genet.* 93, 188–194.

Tönnies, H., Toliat, M.R., Ramel, C., Pape, U.F., Neitzel, H., Berger, W., and Wiedenmann, B. (2001b). Analysis of sporadic neuroendocrine tumors of the enteropancreatic system by comparative genomic hybridization. *Gut* 48, 536–541.

Tönnies, H., Stumm, M., Neumann, L., Volleth, M., Grumpelt, U., Musebeck, J., Annuss, G., and Neitzel, H. (2001c). Two further cases of WHS with unbalanced de novo translocation t(4;8) characterised by CGH and FISH. *J. Med. Genet.* 38, E21.

Tönnies, H., Schulze, I., Hennies, H.C., Neumann, L.M., Keitzer, R., and Neitzel, H. (2001d). De novo terminal deletion of chromosome 15q26.1 characterised by comparative genomic hybridisation and FISH with locus specific probes. *J. Med. Genet.* 38, 617–620.

Tönnies, H. (2002). Modern molecular cytogenetic techniques in genetic diagnostics. *Trends Mol. Med.* 8, 246–250.

Tönnies, H., Neumann, L.M., Gruneberg, B., and Neitzel, H. (2003a). Characterization of a supernumerary ring chromosome 1 mosaicism in two cell systems by molecular cytogenetic techniques and review of the literature. *Am. J. Med. Genet.* 121A, 163–167.

Tönnies, H., Huber, S., Kuhl, J.S., Gerlach, A., Ebell, W., and Neitzel, H. (2003b). Clonal chromosome aberrations in bone marrow cells of Fanconi anemia patients: gains of the chromosomal segment 3q26q29 as an adverse risk factor. *Blood* 101, 3872–3874.

Tönnies, H., Neumann, L.M., Grüneberg, B., and Neitzel, H. (2003c). Characterization of a supernumerary ring chromosome 1 mosaicism in two cell systems by molecular cytogenetic techniques and review of the literature. *Am. J. Med. Genet.* 121, 163–167.

Tönnies, H., Hennies, H.-C., Spohr, H.L., and Neitzel, H. (2003d). Characterization of the first supernumerary tricentric ring chromosome 1 mosaicism by conventional and molecular cytogenetic techniques. *Cytogenet. Genome Res.* 103, 28–33.

Tönnies, H., and Lage, H. (2004). Chromosomal imbalances associated with drug resistance and thermoresistance in human pancreatic carcinoma cells. *Eur. J. Cell Biol.* 83, 591–601.

van Gijlswijk, R.P., Talman, E.G., Janssen, P.J., Snoeijers, S.S., Killian, J., Tanke, H.J., and Heetebrij, R.J. (2001). Universal Linkage System: versatile nucleic acid labeling technique. *Expert Rev. Mol. Diagn.* 1, 81–91.

Veltman, J.A., Schoenmakers, E.F., Eussen, B.H., Janssen, I., Merkx, G., van Cleef, B., van Ravenswaaij, C.M., Brunner, H.G., Smeets, D., and van Kessel, A.G. (2002). High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. *Am. J. Hum. Genet.* 70, 1269–1276.

Veltman, J.A. (2006). Genomic microarrays in clinical diagnosis. *Curr. Opin. Pediatr.* 18, 598–603.

Verma, R.S., and Babu, A. eds. (1989). *Human Chromosomes: Manual of Basic Techniques.* Pergamon Press, New York.

Vermeesch, J.R., Fiegler, H., de Leeuw, N., Szuhai, K., Schoumans, J., Ciccone, R., Speleman, F., Rauch, A., Clayton-Smith, J., Van Ravenswaaij, C., Sanlaville, D., Patsalis, P.C., Firth, H., Devriendt, K., and Zuffardi, O. (2007). Guidelines for molecular karyotyping in constitutional genetic diagnosis. *Eur. J. Hum. Genet.* 15, 1105–1114.

Wang, N. (2002). Methodologies in cancer cytogenetics and molecular cytogenetics. *Am. J. Med. Genet.* 115, 118–124.

Warburton, D. (2000). Cytogenetics of reproductive wastage: from conception to birth. In *Medical Cytogenetics* (Mark, H.F.L., ed.), pp. 213–246. Marcel Dekker, New York, USA.

Ward, B.E., Gersen, S.L., Carelli, M.P., McGuire, N.M., Dackowski, W.R., Weinstein, M., Sandlin, C., Warren, R., and Klinger, K.W. (1993). Rapid prenatal diagnosis of chromosomal aneuploidies by fluorescence in situ hybridization: clinical experience with 4500 specimens. *Am. J. Hum. Genet.* 52, 854–865.

Weber-Matthiesen, K., Pressl, S., Schlegelberger, B., and Grote, W. (1993). Combined immunophenotyping and interphase cytogenetics on cryostat sections by the new FICTION method. *Leukemia* 7, 646–649.

Wegner, R.D. (1999). *Diagnostic Cytogenetics. Springer Lab Manual*. Springer Verlag, Berlin, Heidelberg.

Weise, A., Mrasek, K., Fickelscher, I., Claussen, U., Cheung, S.W., Cai, W.W., Liehr, T., and Kosyakova, N. (2008). Molecular definition of high-resolution multicolor banding probes: first within the human DNA sequence anchored FISH banding probe set. *J. Histochem. Cytochem.* 56, 487–493.

Wells, D., Escudero, T., Levy, B., Hirschhorn, K., Delhanty, J.D., and Munne, S. (2002). First clinical application of comparative genomic hybridization and polar body testing for preimplantation genetic diagnosis of aneuploidy. *Fertil. Steril.* 78, 543–549.

Wessendorf, S., Fritz, B., Wrobel, G., Nessling, M., Lampel, S., Goettel, D., Kuepper, M., Joos, S., Hopman, T., Kokocinski, F., Dohner, H., Bentz, M., Schwaenen, C., and Lichter, P. (2002). Automated screening for genomic imbalances using matrixbased comparative genomic hybridization. *Lab. Invest.* 82, 47–60.

Wiegant, J., Kalle, W., Mullenders, L., Brookes, S., Hoovers, J.M., Dauwerse, J.G., van Ommen, G.J., and Raap, A.K. (1992). High-resolution in situ hybridization using DNA halo preparations. *Hum. Mol. Genet.* 1, 587–591.

Yu, S., Baker, E., Hinton, L., Eyre, H.J., Waters, W., Higgins, S., Sutherland, G.R., and Haan, E. (2005). Frequency of truly cryptic subtelomere abnormalities – a study of 534 patients and literature review. *Clin. Genet.* 68, 436–441.

This page intentionally left blank

Chapter 11

Analysis of Human Splicing Defects Using Hybrid Minigenes

Franco Pagani and Francisco E. Baralle
International Centre for Genetic Engineering and Biotechnology, Padriciano 99, Trieste, Italy

11.1 INTRODUCTION

The availability of fast sequencing protocols is at the base of an ongoing genetic diagnostics revolution. It is now possible to scan at least the coding sequences of any gene for pathological variation. While identifying mutations that produce amino acid changes, stop codons or frame shift of the reading frame is straightforward, the identification of DNA variations that cause aberrant splicing is not as simple. The unexpected complexity of the splicing process, which correctly selects the coding sequences, the exons, from the more abundant non-coding sequences, the introns, has revealed in the past few years the existence of new splicing regulatory elements difficult to identify exclusively by sequence inspection. Variations in these new elements both in coding and non-coding regions may result in an unexpected deleterious effect on the pre-mRNA splicing. The result is that distinguishing between benign and disease-causing sequence substitutions is a challenge for medical geneticists, as well as identifying the nature of the splicing regulatory elements involved. In this chapter, we will describe the hybrid minigenes assay as a tool for the study and characterization of the effect of human DNA variations on the pre-mRNA splicing process. We will also illustrate how the hybrid minigenes in combination with other RNA-based methods such as ribozymes and RNA interference can be used to map splicing regulatory elements and study splicing factors interacting with them.

11.2 BASIC PRE-mRNA SPLICING PROCESS AND ALTERNATIVE SPLICING

Pre-mRNA splicing is a fundamental process in gene expression and in the generation of proteome diversity. In the nucleus, the splicing process acts on the precursor (pre-) mRNA, recognizing the protein coding exonic sequences from the more abundant intronic ones, and joins them together to form the mature mRNA that is then transported to the cytoplasm and translated into a protein. The correct splicing process needs intact *cis* acting elements such as: the 5′ and 3′ splice sites, the polypyrimidine tract and the branch site (Fig. 11.1). The 5′ splice site marks the exon/intron junction at the 5′ end on the intron and its hallmark is the invariant GU dinucleotide. At the other end of the exon, the 3′ splice site region has three conserved elements: the branch site, the polypyrimidine tract, followed

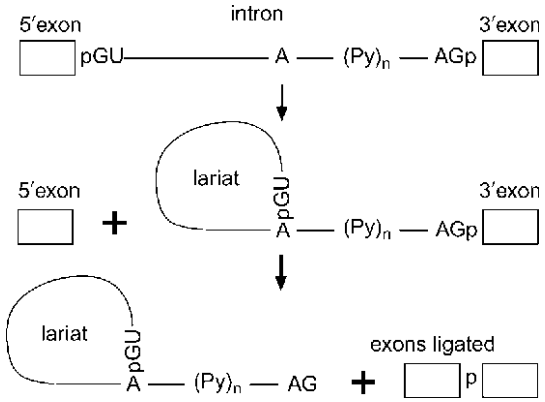


FIGURE 11.1 The splicing process requires two transesterification reactions. Essential splicing signals are the GU/AG dinucleotides at the exon/intron and intron/exon junctions (5′and 3′ splice sites), respectively, the polypyrimidine tract (Py) and the A nucleotide of the branch site. Splicing takes places in two transesterification steps. In the first step the 2′-hydroxyl group of the A residue at the branch site attacks the phosphate at the GU 5′ splice site. This leads to cleavage of the 5′ exon from the intron and the formation of lariat intermediate. In the following step the two exons are ligated by a second transesterification reaction that involves the phosphate at the 3′ end of the intron and the 3′ hydroxyl of the detached exon. This releases the intron, still in the form of a lariat.

by the terminal conserved AG dinucleotide (Fig. 11.1). The splicing cut-and-paste catalytic reaction consists of two transesterification steps (Fig. 11.1)

The multicomponent splicing complex, known as spliceosome, that operates the splicing reaction is composed by five small ribonucleoproteins (snRNP) and more than 100 proteins (Fig. 11.2a). The essential splicing signals on transcribed pre-mRNA are rather short and degenerate sequences at the intron/exon border (Fig. 11.2a). The first event in the splicing reaction is the correct recognition of these classical splicing signals from the abundance and never used pseudo-splice sites. For a detail account of the splicing process see Burge *et al.* (1999).

In a typical gene, the majority of exons are constitutive, which means they are always included in the final mRNAs. However, alternative usage of different splice sites is a common event in human cells. It may occur in different manners, such as usage of alternative splice sites or mutually exclusive exons, exon skipping/inclusion or intron retention (Caceres and Kornblihtt, 2002; Black, 2003). The resulting mRNA variable segments in the mature transcript can insert or remove amino acids, shift the reading frame or introduce a termination codon or even modify regulatory elements for translation or mRNA stability or localization. Alternative splicing results in the production of multiple isoforms from the same transcription unit and is responsible for a substantial part of the complexity of the proteome. It has been estimated that up of 59% of human genes generates multiple mRNAs, an observation that can explain the low unexpected number of genes found in the human genome. A large fraction of alternative spliced transcripts has a restricted developmental and/or cell-type regulation or responds to a variety of external stimuli and this complexity cannot be explained simply by considering the composition of classical splicing signals.

11.3 NOVEL *CIS*-ACTING ELEMENTS INVOLVED IN SPLICING REGULATION

The correct splice site differentiation from the abundance of pseudo-splice sites and the fine-tuning regulation of the alternative splicing process requires auxiliary *cis*-acting elements on the pre-mRNA that assist the spliceosome in the selection of the splice sites. These elements, based on their effect on splicing, have been schematically divided in enhancers and silencers (Cartegni *et al.*, 2002b). Enhancers and silencers act respectively by stimulating or inhibiting the splicing reactions and can be located either directly in the exons (ESE and ESS), and thus overlap with the selection of the amino acid sequence and codon usage, or in introns (ISE and ISS) even at a very long distance from the splice sites (Fig. 11.2a).

Enhancers and silencers are involved both in constitutive and alternative splicing and in the majority of cases they lack a well-defined consensus sequence. Furthermore, these elements are not always univocally defined and their functions may overlap. In fact in some systems it may be more appropriate to talk about composite exonic regulatory elements of splicing (CERES) as it has been described for CFTR exons 9 and 12 (Pagani *et al.*, 2003a, b). Several *trans*-acting splicing factors can interact with enhancers and silencers and accordingly they have been divided into two major groups: members of the serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs) (Fig. 11.2a). In general, but not exclusively,

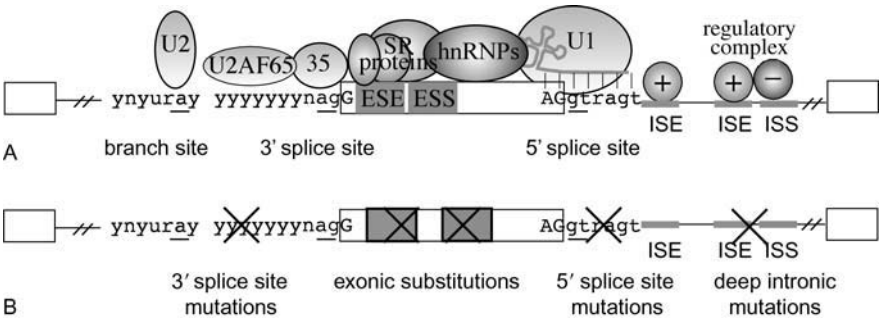


FIGURE 11.2 Regulatory elements in pre-mRNA splicing and mutations that can affect them. **A.** The essential splicing signals that define the exon boundaries are relatively short and poorly conserved sequences. Only the GU and the AG dinucleotides that directly flank the exon (at the 3′ and 5′ ends, respectively) and the branch point adenosine (all in red) are always conserved. In most cases there is also a polypyrimidine tract of variable length (a pyrimidine base, cytosine, or thymine is represented by the consensus symbol "y") upstream of the 3′ splice site. The 5′ splice site binds to the U1 RNA by complementarity. The branch point is typically located 18–40 nucleotides upstream from the polypyrimidine tract. Additional enhancer and silencer elements in the exons (ESE; ESS) and/or introns (ISE; ISS) allow the correct splice sites to be distinguished from the many cryptic splice sites that have identical signal sequences. *Trans*-acting splicing factors can interact with enhancers and silencers and can accordingly be subdivided into two major groups: members of the serine arginine (SR) family of proteins and heterogeneous nuclear ribonuclear particles (hnRNPs). In general, but not exclusively, SR protein binding at ESE facilitates the exon recognition whereas hnRNPs are inhibitory. **B.** Mutations can result in aberrant splicing by affecting different splicing regulatory elements. Exonic variations in ESE or ESS can either change the amino acid or result in synonymous substitutions. Mutations in the intron might occur at the splice sites or deep in the introns. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

SR protein binding at ESE facilitates the exon recognition whereas hnRNPs are inhibitory. These proteins, in common with the majority of RNA-binding proteins, have a modular structure, which consists of one or more RNA-binding domains associated to an auxiliary domain that is often involved in protein–protein interactions.

11.4 HUMAN GENETIC DEFECTS INVOLVING PRE-mRNA SPLICING

Genome sequences from different individuals may reveal sequence variations. A difficult task is to correctly distinguish between benign polymorphisms from disease-causing variations. The identification of disease-causing mutations is based primarily on linkage of the mutation with the disease phenotype and on the effect of the mutations on gene expression. This effect is generally assumed to depend on the location of the nucleotide variants. In several cases, the correct interpretation of the molecular nature of the substitution may not be immediately evident (Fig. 11.3).

The deleterious effect on gene expression of large deletions, nucleotide changes at the consensus GU/AG splice sites, exonic variations that produce stop codons or radical amino acid changes and significant promoter variations are relatively easy to predict. Variations in nucleotides flanking the canonical GU/AG dinucleotides may be more difficult to assess, because the consensus of these regions shows some variability (Zhang, 1998). Exonic sequence variations overlap with coding sequences and for this reason they are frequently considered only for their effect at the protein function level. Missense mutations modify the amino acid composition and synonymous variations are largely ignored as potentially deleterious. In the same way, deep intronic variations are not evaluated for their potential effect on splicing. Recent evidence from many laboratories have now indicated that the primary mechanism of disease in a significant fraction of disease-causing mutations is catastrophic splicing abnormalities that disrupt previously unrecognized splicing regulatory elements (Cartegni *et al.*, 2002b; Faustino and Cooper, 2003; Pagani and Baralle, 2004; Wang and Cooper, 2007). Given the abundance of *cis*- and *trans*-acting factors involved in the splicing process, it was not completely surprising that a great proportion of human defects have something to do with the splicing process. Genetic analysis in *NF1* and *ATM* genes has shown that, in about 50% of the affected patients, mutations that caused splicing alterations were involved (Teraoka *et al.*, 1999; Ars *et al.*, 2000). Of these mutations, 13% and 11%, respectively, would have been erroneously classified as frameshift, missense or nonsense mutations, if the analysis had been limited to genomic sequences. Studies on the mRNA are disclosing that many of them are really exon skipping mutations. Interestingly, most of the splicing mutations identified in these genes did not involve the conserved essential splice sites. So in most studies attempting to identify disease-causing mutations from the DNA sequence alone, SNPs such as these would be overlooked or not correctly classified as producing splicing errors. These sequence variations may affect those auxiliary enhancer and silencer elements which, having a very loose consensus sequence and in several cases overlap with coding regions, are difficult to identify. Several computer-assisted tools have been proposed to evaluate the effect of SNPs on splicing (Table 11.1).

Practically, visual inspection alone can be used in the case of mutations at the invariant canonical splice sites. These cases will also be easily picked up by the programs designed to identify 5′ and 3′ splice sites. These programs provide a score that is variable in each program and if the score is above a threshold the site is considered authentic. In several cases though the scores only partially correlate with the splicing changes (Baralle *et al.*, 2003; Buratti *et al.*, 2007) and the current programs cannot predict if the

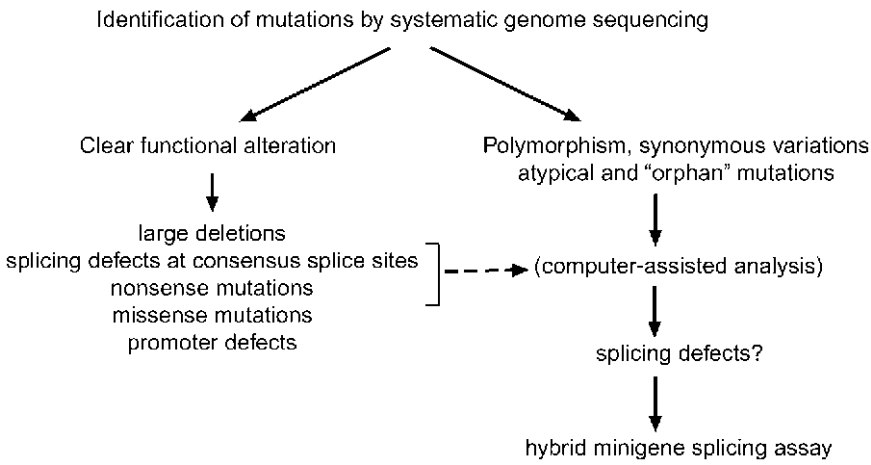


FIGURE 11.3 Functional genomics and pre-mRNA processing alterations.

TABLE 11.1 Useful websites to identify 5′ and 3′ splice sites, exonic regulatory elements, ESE and ESS.

Useful websites to identify 5′ and 3′ splice sites
1. http://www.uni-duesseldorf.de/rna/html/hbond_score.php
2. http://www.fruitfly.org/seq_tools/splice.html
3. http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq_acc.html
4. http://ast.bioinfo.tau.ac.il/SpliceSiteFrame.htm
Useful programs to identify exonic regulatory elements, ESE and ESS
1. http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home
2. http://genes.mit.edu/exonscan
3. http://ast.bioinfo.tau.ac.il
4. http://cubweb.biology.columbia.edu/pesx

mutation will result in exon skipping, activation of cryptic splice sites or intron retention. On the other hand, with the increasing number of splicing mutations reported it is possible to try a ''reverse'' approach which starts from the study of the splicing affecting mutations to explore possible common rules involved. Currently, this line has been investigated for mutations at the relatively ''simple'' splice sites and in particular at the donor site consensus. A database of aberrant splice sites in human disease genes is currently freely available at http://www.dbass.org.uk and contains 625 records of aberrant 3′ and 5′ splice sites generated by 682 different mutations in 235 genes (Buratti *et al.*, 2007). Each data can be searched by phenotype, gene, mutation, location of aberrant splice sites in introns and exons, their distance from authentic counterparts, and by bibliographic references.

Bioinformatic approaches present serious limitations for more degenerated splicing regulatory elements like exonic or intronic regulatory sequences. In general, computer-based approaches identify splicing regulatory elements on a statistical basis providing a large number of consensus sequences. For example, programs that identify putative ESEs are based either on the assumption that ESE sequences are more represented in exons than in introns or on sequences known to be *in vitro* binding targets of positive splicing factors such as SR proteins (Fairbrother *et al.*, 2002; Cartegni *et al.*, 2003; Zhang and Chasin, 2004). As a consequence, an obvious approach is first to check for a given SNP disrupt or create an ESE in which case a potentially pathogenic defect may arise. However, at the moment computer-assisted analysis can suggest only if an SNP has some probability of affecting a splicing regulatory element: clarification of its real pathogenetic role requires transcript analysis in patients' derived cells or in the minigene splicing assay. With the increasing number of disease-causing mutations reported in exonic splicing regulatory elements it would be possible to apply a similar ''reverse'' approach as the one described above for 5′ and 3′ splice sites to improve the ESE recognition programs.

The direct analysis of the processed transcript, mainly in the affected tissue, is the best way to establish with certainty if a particular DNA substitution affects splicing. However, samples obtained for clinical diagnosis are almost always leukocytes from which DNA is prepared; RNA samples from the affected individuals are not always available or not available at all for some tissue-specific expressed genes, such as brain- or heart-specific genes. The development of functional splicing assay is of utmost importance to determine if a particular sequence variation represents a benign polymorphism or a disease-causing mutation and to study the mechanism involved in normal and aberrant splicing.

11.5 GENERAL STRATEGY OF THE HYBRID MINIGENE ASSAY FOR THE IDENTIFICATION OF SPLICING DEFECTS

The basic principle of hybrid minigene splicing assay is shown in Fig. 11.4. Any genomic region of interest (i.e. exon and short intronic flanking regions), which is suspected of causing a splicing defect because it contains an ''orphan'' mutation, can be amplified from normal and affected individuals and cloned into the minigene. The minigene plasmid is then transiently transfected in the appropriate cell line where it will be transcribed by RNA polymerase II and the resulting pre-mRNA processed to obtain a mature mRNA. The mRNA splicing pattern is analyzed mainly by RT-PCR with primers specifically designed to amplify processed transcripts derived from the minigene to distinguish from endogenous transcripts. The use of RT-PCR does not allow an exact quantification of the amount of each processed transcript but of relative proportion of each splicing variant. To better quantify the absolute amounts of splicing variants other methods should be used such as RNase protection assay, real-time PCR (see also Chapter 7) or Northern blot analysis.

The length and exon/intron composition of the gene portion to be cloned in the minigene have to be selected carefully and depend on the location of the presumed defect. Given the number and diversity of regulatory splicing signals, it is worth evaluating the gene structure as completely as possible. However, for practical reasons, in the majority of cases, sequence substitutions in constitutive included exons are analyzed for exon skipping and require at minimum the cloning of the exon itself along with a portion of the flanking introns. The same approach can be used for testing variations near the splice sites which can induce exon skipping or/and activation of cryptic splice sites. However, in the presence of short introns, it is better

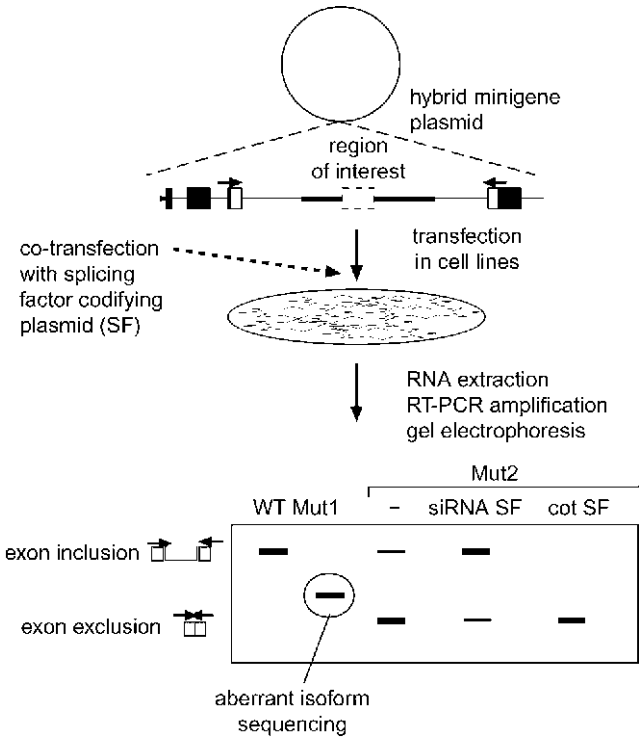


FIGURE 11.4 The hybrid minigene splicing assay. Study of human variations involved in aberrant splicing using the hybrid minigene. A typical hybrid minigene is a plasmid that contains a "simplified" version of the gene that will be evaluated for pre-mRNA splicing. It contains at the 5′ end promoter/enhancer sequences indicated by a long arrow, to allow polymerase II transcription in the transfected cell lines. This is followed by a series of exonic and intronic sequences (indicated as boxes and lines, respectively) that may derive from a reported gene or from the gene context itself. In this case the reporter gene is composed by α-globin (black boxes) and fibronectin exons (dashed boxes). The genomic DNA region of interest that contains a putative "orphan" splicing mutation is introduced in the minigene (thick line represents intronic sequences and white dashed box the presumed exonic sequences). The minigenes (normal and mutated) are transfected in cell lines, followed by RNA extraction, RT-PCR, and gel electrophoresis. In the example, the analysis of the transcript derived from the WT minigene shows the correct inclusion of an exon. Mut1 shows an aberrant spliced form whose identity can be verified by sequence analysis. The Mut2 causes significant exon skipping. In this case, co-transfection of a positive regulatory splicing factor or siRNA-mediated inhibition of a negative splicing factor induces exon inclusion restoring normal splicing.

to include them along with the nearby exons in the minigene as in these cases intron retention may be the observed splicing defect. In some cases, the presence of multiple exons may be necessary (see section 11.6.3, below), and if the introns are too long, they can be internally deleted. In any case, comparison between transcripts derived from normal and mutated minigene is always necessary to identify the disease-causing role of "orphan" splicing variations.

To get more insight on the basic mechanism involved and in particular to identify elements required for regulation by a specific *trans*-acting splicing factor, the hybrid minigenes assay can be performed in cells in which a particular splicing factor is overexpressed or depleted. Overexpression is obtained by transient expression of the protein of interest, which is normally distinguished from the endogenous by epitope tagging. A potential drawback of this method is represented by the fact that the overexpressed protein may exert an unspecific effect on splicing or may be present in the cells at maximum levels. Thus siRNA-mediated silencing of the splicing factor represents a complementary powerful approach to identify elements required for regulation by a specific *trans*-acting splicing factor. In this case, the cells are treated in advance with the siRNAs, typically synthetic double-stranded RNAs, followed by transfection with the minigenes. The efficiency of silencing can be monitored at the level of mRNA by RT-PCR or at the level of the protein. The protein depletion is monitored by Western blotting and analysis of the splicing pattern evaluated by the RNA splicing assay. A detailed technical description of the hybrid minigene splicing assay and siRNA-mediated depletion of splicing factors can be found in Ayala and coworkers (2006) and Goina and coworkers (2008).

An alternative method useful for studying splicing which will not be described in detail here is the *in vitro* splicing assay. In *in vitro* splicing assay, labeled pre-formed RNA molecules transcribed with bacterial polymerases are incubated in the presence of nuclear extracts and the resulting spliced products resolved on polyacrylamide denaturing gel. With this assay the intermediates of the splicing reactions, such as the lariat formation, can be evaluated. However, the most used *in vitro* assays do not take into account the fact that transcription and splicing are intimately connected in the cell. For a description of the mechanisms that couple transcription and pre-mRNA processing see recent reviews (Caceres and Kornblihtt, 2002; Maniatis and Reed, 2002). To overcome this problem some *in vitro* systems have been recently explored. These systems are capable of transcribing and splicing hybrid minigenes in a test tube (Das *et al.*, 2006, 2007). However, also in this *in vitro* splicing assay only relatively short sequences, frequently containing a reduced single intron, can be evaluated. The rest of the chapter will be dedicated to specific examples that will illustrate the analysis of splicing affecting mutations.

11.6. APPLICATIONS OF THE HYBRID MINIGENE ASSAY

11.6.1 3′ Splice Site Variations

Cystic fibrosis is caused by mutations in the CF transmembrane regulator (CFTR) gene and is characterized by pathological features of variable severity at the level of lungs, pancreas, sweat glands, testis, ovaries, and intestine. Some patients have evidence of a clinical disease in only a subgroup of the organ systems. These non-classical CF forms

include late-onset pulmonary disease, male sterility due to congenital bilateral absence of vas deference (CBAVD) and idiopathic pancreatitis. These non-classical forms have been found associated with a peculiar allele at the polymorphic CFTR intron 8/exon 9 junction. At this locus a variable number of dinucleotide UG repeats (from 9 to 13) followed by a U repeat (U5, U7 or U9) can be found in the normal population. The U5 allele is considered a disease mutation with incomplete penetrance as it can be found both in normal individuals and in affected ones. The pathologic effect of the U5 allele has been associated with the alternative splicing of the CFTR exon 9. This exon encodes part of the functionally important first nucleotide-binding domain and its skipping produces a non-functional CFTR protein. The UG13U3 allele has been found in an individual with classical CF (Buratti *et al.*, 2001).

The hybrid minigene experiments shown in Fig. 11.5 indicate the importance of the UGmUn polymorphism in determining the amount of transcripts that contain exon 9 and explore some of the key splicing factors involved. Exon 9 sequences, along with part of flanking introns, were introduced in hybrid minigenes, and different polymorphic variants were studied. Transfection experiments showed that the UG11U7 allele found in normal individuals produced about 85% of normal exon 9 plus mRNAs. The UG11U5 allele, variably associated to non-classical CF, reduces the amount of normal transcript to about 65% while the UG13U3 allele induces significant exon skipping (only 15% include the exon; Fig. 11.5b). The residual amount of normal exon 9 inclusion correlates with the severity of the phenotype. *In vitro* analysis with naked RNAs containing the UG sequence showed that this particular repeat binds efficiently to a novel splicing factor TDP43 (Buratti *et al.*, 2001). To prove the functional role of TDP43 in the CFTR exon 9 regulation, its silencing was performed. The cells were treated with synthetic siRNAs oligonucleotides followed by transfection with the CFTR hybrid minigene. TDP43 protein concentration, monitored by Western blotting, showed a significant reduction after siRNA whereas the internal control tubulin was not affected (Fig. 11.5c). Analysis of the splicing pattern showed that silencing of TDP43 induced a consistent rescue of the splicing pattern (Ayala *et al.*, 2006). In the transfected UG13U3 minigene, the percentage of exon inclusion increased from 15% to 80% (Fig. 11.5d).

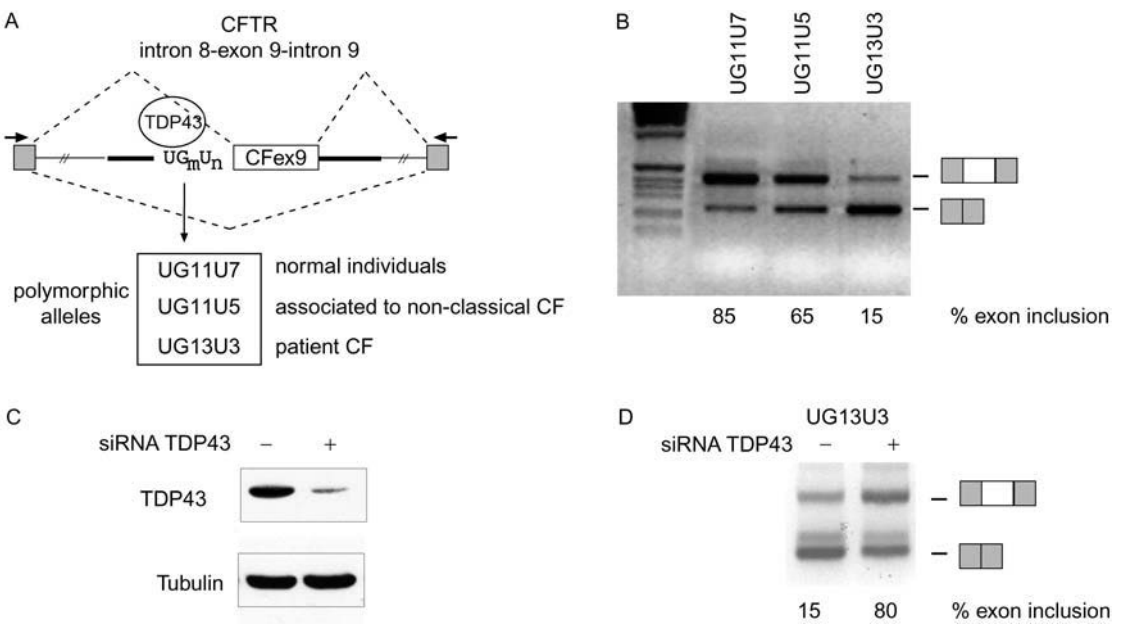


FIGURE 11.5 3′ splice site variations. The example illustrates the use of the minigene approach to study a polymorphic region near the 3′ splice site in CFTR that causes aberrant exon 9 skipping. **A.** The polymorphic UGmUn locus at the end of CFTR intron 9 contains a variable number of dinucleotide UG repeats (from 9 to 13) followed by a U repeat (U3, U5, U7, or U9) variably present in the human population. Variants at the locus (mainly the U5 allele) have been associated to phenotypes of different severity, ranging from classical to non-classical forms of CF, and in some cases found also in normal individuals. The UG13U3 is found in classical form of CF. In the figure, three representative alleles are shown. The three variant minigenes that contain the indicated number of UG and U near the AG 3′ splice site (underlined) were analyzed in splicing assay, and the results are shown in panel b. **B.** Hybrid minigene transient transfection assay showing the effect of the polymorphic variants. The variations induce variable skipping of exon 9 that correlates with the severity of the phenotype. Two processed transcripts are evident in the agarose gel, a normal one that contains the exon 9, and an aberrant one without the exon. Several splicing factors modulate the amount of aberrant exon 9 skipping (not shown), thus possibly influencing the phenotypic expression (Pagani *et al.*, 2000; Buratti *et al.*, 2001). **C.** siRNA-mediated silencing of the splicing factor TDP43, which was found to bind to UG repeats *in vitro*. The Western blotting shows the effect of siRNA against TDP43 on the protein levels in comparison to the control tubulin. **D.** Effect of the siRNA-mediated silencing of the splicing factor TDP43 on the UG13U3 minigene. The siRNA (+) induces rescue of aberrant splicing.

11.6.2 Nucleotide Substitutions in Exonic Regulatory Elements Involving Splicing

11.6.2.1 Missense and Silent Mutations in Composite Exonic Regulatory Element of Splicing

Skipping of exon 12 removes a highly conserved region encoding part of the first nucleotide-binding fold of CFTR rendering the protein non-functional. In this exon, some nucleotide substitutions that cause a change in coding sequence did not show a clear association with loss of protein functionality or disease phenotype. This is the case for two interesting and enigmatic missense mutations, D565G and G576A (the latter having previously considered a neutral polymorphism; Fig. 11.6a). We have analyzed the splicing defects associated with exon 12 D565G and G576A variations.

Using the hybrid minigene, shown in Fig. 11.6b, the natural mutations D565G and G576A induce a variable extent of exon 12 skipping that leads to reduced levels of normal transcripts. Interestingly, other missense mutations, Y577F, that have been reported in a patient with severe CF, did not cause exon skipping but on the contrary increased the amount of transcript with the exon (Fig. 11.6c). This study was reinforced by the opportunity to study the splicing pattern found in patients' derived cells (Fig. 11.6d). This analysis showed a perfect concordance between the pattern observed with the minigene system and the one observed in the cells harboring the natural mutation as can be seen by comparing Figs 11.6c and 11.6d.

To study the nature of the exonic regulatory elements involved, a systematic site-directed mutagenesis near the natural substitutions was performed (Pagani *et al.*, 2003b). These experiments clearly showed an overlapping enhancer and silencer functions (Fig. 11.7a). In fact, nearby mutations or even variations at the same position may result in an enhancing or silencing effect. Due to their peculiar behavior, these elements were named composite exonic regulatory element of splicing (CERES) (Pagani *et al.*, 2003b). Interestingly, some site-directed mutants that affect splicing (Fig. 11.7a underlined) are at the third position of codons and do not modify the amino acid code.

The effect on splicing of synonymous substitutions was systematically studied in the human CFTR exon 12 where the synonymous nucleotides between positions 13 and 52 were individually substituted and the resulting minigenes evaluated for splicing efficiency (Pagani *et al.*, 2005). Compared with the normal human exon that generated about 80% of transcript containing exon 12 (Fig. 11.7c, wt),

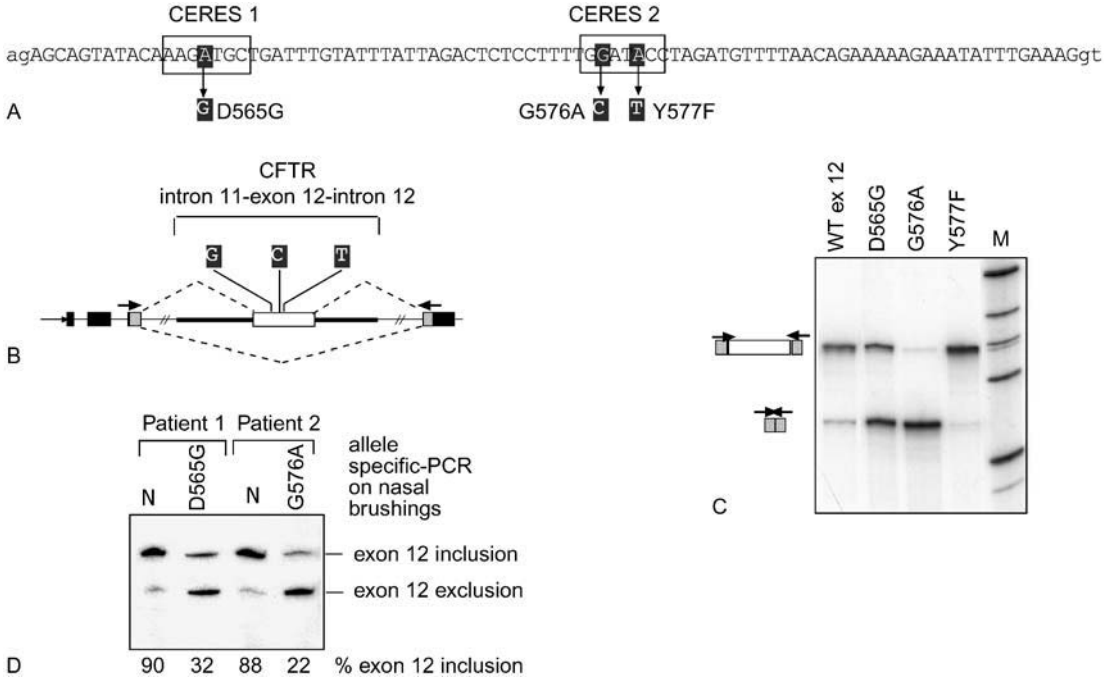


FIGURE 11.6 Missense mutations in composite regulatory elements of splicing (CERES) in CFTR exon 12. **A.** Nucleotide sequence of the CFTR exon 12 showing the position of the two composite exonic regulatory elements of splicing (CERES) (boxed) and natural missense variations. **B.** Schematic representation of the three hybrid minigenes analyzed. These minigenes contain normal exon 12 (WTex12) or the natural mutations. **C.** Hybrid minigenes transient transfection assay showing the effect of natural CFTR exon 12 mutants. The two bands correspond to inclusion (upper) or exclusion (lower) of CFTR exon 12 in final transcripts. WT exon 12 shows incomplete exon inclusion as found *in vivo*. D565G and G576A induce significant exon skipping. On the contrary, Y577F increases the percentage of exon 12 inclusion. **D.** Allele specific RT-PCR on RNA extracted from nasal epithelial cells from patient 1 (heterozygote for D565G) and patient 2 (heterozygote for G576A). The RNAs were amplified with allele-specific primers that can detect the alleles with the missense variations (D565G or G576A) or the other allele (N). Note the perfect correlation between the natural amount of exon inclusion (panel d) and the one obtained with the minigene system (panel c).

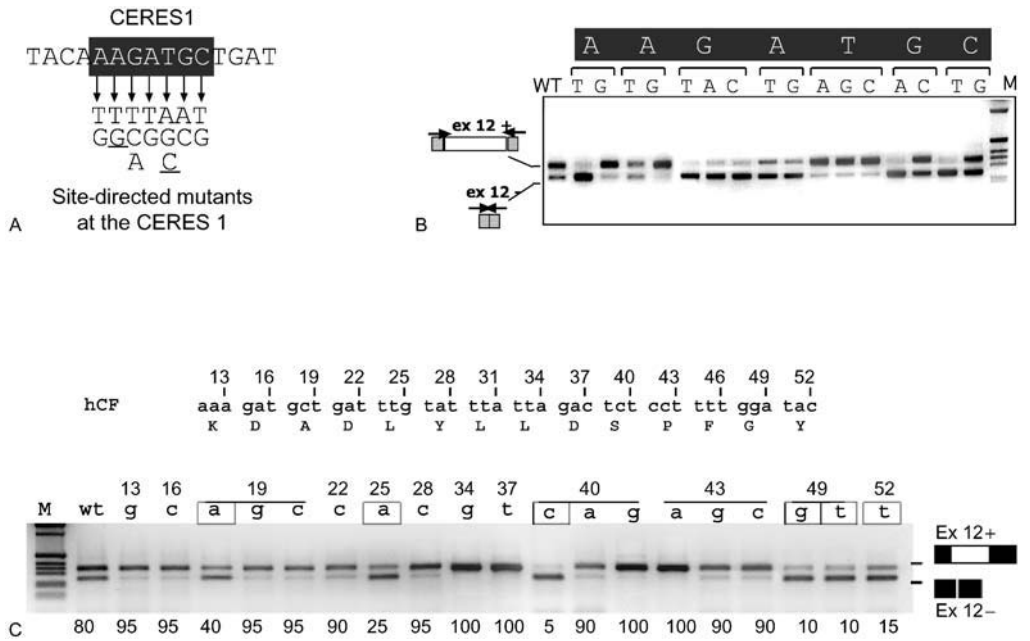


FIGURE 11.7 Exonic variations in composite exonic regulatory element of splicing (CERES) in CFTR exon 12 and role of synonymous mutations. **A.** Nucleotide sequence of the composite exonic regulatory element of splicing 1 (CERES) in CFTR exon 12 and side-directed variations. The site-directed mutants at the third position of the codon usage that do not change the amino acid code are underlined. The different site-directed mutant minigenes were transfected in cells in culture and the percentage of exon 12 inclusion analyzed by RT-PCR. **B.** Hybrid minigenes transient transfection assay showing the effect of natural CFTR exon 12 mutants. Agarose gel electrophoresis of RT-PCR products from splicing assay from the site-directed mutants in the AAGATGC CERES showing the exon 12+ and exon 12− transcripts. **C.** Synonymous mutations in human CFTR exon 12 induce exon skipping. The upper panel shows the nucleotide and amino acid composition (in one letter code) of part of the human CFTR exon 12 (hCF). Third nucleotide positions are numbered according to their location in the exon. The lower panel shows the RT-PCR products from splicing assay from 19 out of 22 possible single synonymous changes between positions 13 and 52 of the human CFTR exon 12. Above each lane is indicated the position and the nucleotide substitution analyzed. The percentages of exon inclusion are reported at the bottom of each lane. The six synonymous substitutions that induce significant exon skipping are boxed.

most mutations produced a small increase in exon inclusion while surprisingly six changes, 19A, 25A, 40C, 49G, 49T and 52T, induced exon skipping (Fig. 11.7c). In two positions, 19 and 40, the effect on splicing was dependent on the nucleotide substitution: 19A and 40C caused an increase in exon skipping whereas 19G, 19C, 40G, and 40A increased the exon inclusion.

These synonymous changes in CFTR exon 12 that cause exon skipping would be labeled as neutral variations, if found in a classical genome scanning analysis. Interestingly, this systematic analysis indicates the high probability of synonymous substitutions in CFTR exon 12 in inducing exon skipping (about one of three). "Neutral" variations frequently considered not to have pathological consequences may on the contrary induce a severe splicing defect. An interesting case of a splicing regulatory synonymous variation with therapeutic implication in the degenerative neurological disorder spinal muscular atrophy is described below.

11.6.2.2 The Importance of the Genomic Context when Assessing Minigene Analysis: The example of NF1 Exon 37

Neurofibromatosis type 1 (NF1) is a common autosomal dominant genetic disorder with a prevalence of ~1 in 3,000 individuals and is one of the most common single gene disorders influencing neurological function in humans. The NF1 gene maps to chromosome 17q11.2 and is thought to be a tumor suppressor gene because loss of heterozygosity is associated with the occurrence of benign and malignant tumors in neural crest-derived tissues as well as myeloid malignancies. It spans a region of about 350 kb of genomic DNA and contains 60 exons. The NF1 gene transcribes several mRNAs in the size range 11–13 kb; these are markedly expressed in neurons, oligodendrocytes, and non-myelinating Schwann cells. The most common transcript codes for a polypeptide of 2,818 amino acids called neurofibromin. About 50% of the mutations identified result in aberrant splicing, and most of these defects did not involve the conserved dinucleotides at the splice sites (Ars *et al.*, 2000). The literature describes a very interesting set of mutations that introduce stop codons in exon 37 and may produce significant alteration of the splicing patterns (Hoffmeyer *et al.*, 1998).

One of these nonsense mutations, a C to A substitution, was studied in detail in different minigene contexts (Fig. 11.8) and gave interesting information on the regulatory exonic sequences involved. In exon 37, substitutions at the same position may code for a stop codon (TAA and TAG) or are synonymous (TAT). The three substitutions induced the same aberrant splicing pattern in the minigenes

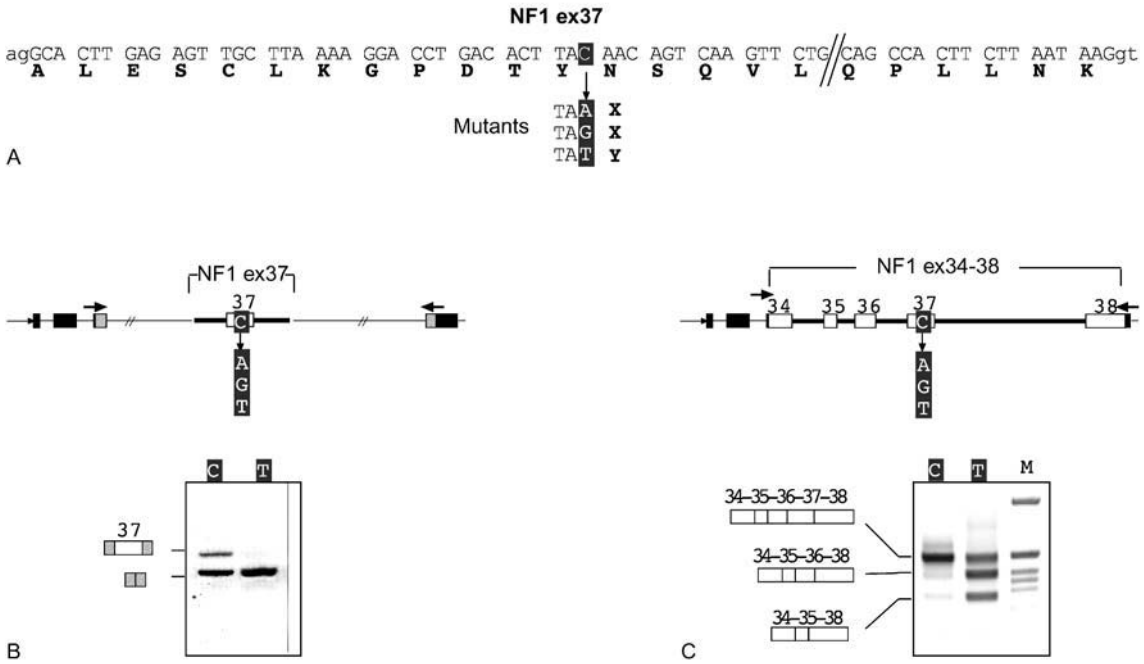


FIGURE 11.8 The genomic context may influence the splice site selection: the example of NF1. **A.** Nucleotide sequence of part of the NF1 exon 37 showing the three different variants at the same position. The natural C to A and C to G mutants are nonsense substitutions. The site-directed C to T substitution is a silent mutation maintaining a tyrosine codon. The normal and mutant exon 37 variants were introduced in two different minigene contexts. In panel b exon 37 is flanked only by nearby intronic sequences; in panel c exon 37 variants are embedded in the entire genomic region from exon 34 to 38. **B.** The upper part shows a schematic view of the hybrid minigene containing only the NF1 exon 37 and part of flanking introns. In the lower panel, the resulting pattern of splicing is evident with the exon 37 inclusion and exclusion bands. It is important to note that in the normal C minigene, only a low amount of the transcripts contains exon 37 and this differs from the natural context where exon 37 is constitutively included. The three variants induce complete exon skipping. **C.** The upper part shows the hybrid minigene that contains the genomic region of NF1 from exon 34 to 38. Lower panel shows the processed transcripts. Contrary to results shown in panel b, the wild-type C construct is normally processed as in the natural context and includes all the five NF1 exons present in the minigene. The three variations induce two aberrant spliced forms: one with skipping of exon 37 alone and one where both exon 36 and 37 are missing. This example shows clearly the importance of the context and the message is that the minigene system can be used only when the wild-type sequence gives the same splicing pattern as the chromosomal gene (see CFTR exon 12 example, Fig. 11.6c and 11.6d).

causing multiple exon skipping, indicating that this aberrant splicing was not necessarily linked to the presence of a stop codon (Fig. 11.8c). In addition, this experiment highlights the importance of the appropriate genomic context in determining the effect of the DNA variations on splicing. If only exon 37 and part of its flanking introns is introduced in the minigene, even the wild-type exon 37 is not fully recognized by the splicing machinery (Fig. 11.8b, lane C). These data indicate that additional genomic sequences extending beyond the flanking introns are necessary for correct processing of this exon. Indeed, the minigene that contains the entire normal genomic sequences from exon 34 to exon 38 showed the complete inclusion of wild-type exon 37 (Fig. 11.8c, lane C). Interestingly, in this minigene, the substitutions induced a complex splicing defect that extended beyond the recognition of exon 37. In fact, exon 37 mutants induced not only its skipping but some transcripts lack both exon 36 and 37 (Fig. 11.8c, lane T).

Thus, to study appropriately the effect of human DNA variations on splicing, the first step is to ensure that there is a good correlation between the wild-type splicing pattern deriving from the chromosomal gene and the minigene construct. If this correlation does not exist the context of the minigene should be appropriately modified.

11.6.2.3 A Silent Mutation in SMN2

Spinal muscular atrophy is a pediatric neurodegenerative disorder caused by homozygous loss of function of the survival motor neuron 1 (*SMN1*) gene. *SMN1* is duplicated in the human genome and its highly homologous copy is called *SMN2* and both genes are transcribed. The *SMN2* gene is present in all the patients but is not able to compensate for the *SMN1* gene defect. *SMN2* differs from *SMN1* by five nucleotides. These variations, either intronic or exonic, are translationally silent. One of these variations, a translationally silent C > T substitution in exon 7, six nucleotides downstream the 3'ss, has been shown (Lorson *et al.*, 1999; Lorson and Androphy, 2000) to be the cause of the inability of *SMN2* to compensate for *SMN1*. Two minigenes were prepared, SMN1 and SMN2, that differ only for the C to T variant. Genomic SMN1 and SMN2 DNAs

including exons 6–8 were cloned into a mammalian expression plasmid, downstream of the constitutively expressing cytomegalovirus promoter. In SMN2, the synonymous substitution causes alternative splicing with skipping of exon 7 in the majority of transcripts (Lorson *et al.*, 1999; Lorson and Androphy, 2000) (Fig. 11.9).

The truncated transcripts skipping exon 7 encode a protein lacking the 16 C-terminal residues, which is unstable and non-functional. Surprisingly, the exact mechanism by which this silent substitution causes exon skipping is not yet clear. Two models have been proposed to explain the effect of this substitution: one suggests that it causes the inactivation of an ESE binding the SR protein SF2/ASF (Cartegni and Krainer, 2002a; Cartegni *et al.*, 2006) and the other that it creates a new ESS binding hnRNPA1 (Kashima and Manley, 2003; Kashima *et al.*, 2007). Some novel insight comes from the analysis of a similar exon skipping mutation in *BRCA1* exon 18 (Kashima *et al.*, 2007; Goina *et al.*, 2008). A natural G to T mutation at position +6 of *BRCA1* exon 18 presents common features with the SMN system as it occurs at the same exonic position of *SMN2*, creates a similar sequence, and induces exon skipping. In this case, exon skipping has been recently associated with binding of additional splicing factors with silencer properties (Goina *et al.*, 2008). The elucidation of the basic mechanisms is important for the spinal muscular atrophy as the splicing reactivation of the dormant *SMN2* paralog might represent a new strategic therapy in affected patients (see comments in Buratti *et al.*, 2003; Khoo *et al.*, 2003).

11.6.3 Variations at the 5′ Splice Site

The effect of variations near the invariant AG/GU dinucleotides at the splice sites are not always easy to evaluate. An exon 3 + 5G > C substitution in neurofibromatosis type 1 gene (NF1) identified during an NF1 genomic sequence clinical screen represents a diagnostic challenge (Baralle *et al.*, 2003). The proximity of the substitution to the 5′ splice site suggests that it may interfere with splicing (Fig. 11.10a). However, there are examples of wild-type 5′ splice sites similar to the exon 3 + 5G > C, in which the corresponding exon is spliced efficiently. For example, in normal *NF-1*, introns 1 and 7 share identical −1 to +5 sequence with the mutated intron 3 and intron 37 carries a T instead of a G at the +5 position (Fig. 11.10a). It is obvious that sequence analysis alone cannot predict if the nucleotide substitution is pathogenic and like in most clinical genetics situations RNA was not available for diagnosis. A region comprising exon 3 and its flanking intronic sequences (Fig. 11.10b) of the potentially abnormal *NF-1* gene was amplified from the available genomic DNA and was inserted into the hybrid minigene constructs. Following transfection and expression of the construct in Hep3B cells, the mRNA produced by the cells was analyzed for splicing pattern by RT-PCR. As shown in Fig. 11.10c, the exon 3 + 5G > C mutation dramatically affected pre-mRNA processing, causing exon 3 to be completely skipped. One of the splicing factors that could be involved in the pathogenesis of this defect is U1 snRNP. This ribonucleoprotein particle contains several proteins and a unique RNA, the U1 snRNA.

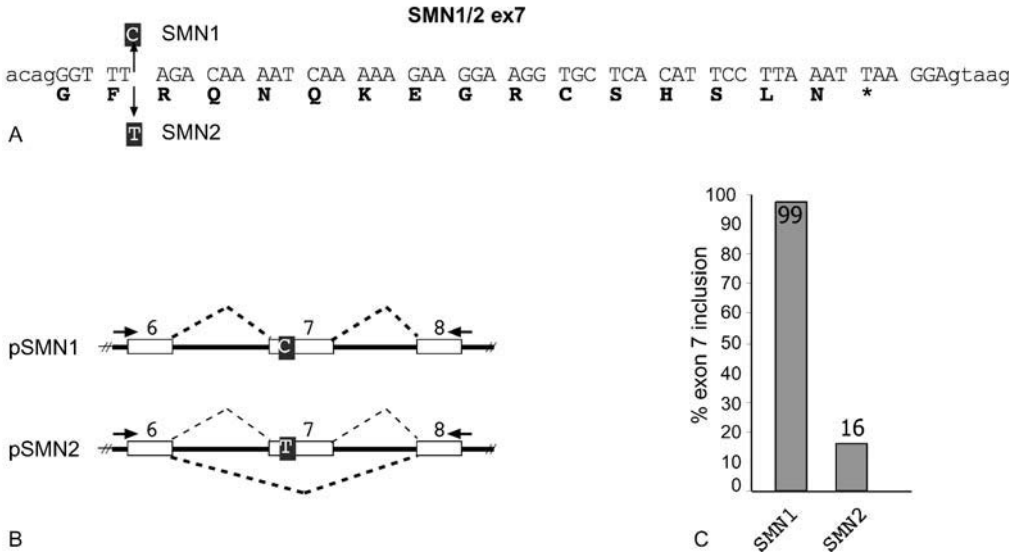


FIGURE 11.9 A silent variation in SMN2 causes exon skipping. **A.** Nucleotide sequence of the exon 7 of SMN genes showing the position of the synonymous C to T substitution. In exon 7, the SMN1 gene contains a C whereas SMN2 contains a T. **B.** Two minigenes were prepared, SMN1 and SMN2, that differ only for the C to T variant. Genomic SMN1 and SMN2 DNAs including exons 6–8 were cloned into a mammalian expression plasmid, downstream of the constitutively expressing cytomegalovirus promoter. The indicated hybrid minigenes were used to demonstrate that the T variant in SMN2 induce exon skipping resulting in a non-functional protein. **C.** Percentage of exon 7 inclusion derived from transient transfection experiments. Data are from Lorson and Androphy (2000).

The U1 snRNA is involved in the recognition of the 5′ splice site by base pair complementarity. The substitution of the guanosine by a cytidine in position +5 of IVS 3 lessens the degree of U1-snRNA base pairing with the 5′ splice site (Fig. 11.10d) although not to an extent that would render the 5′ splice site non-functional as shown by the sequences compared in Fig. 11.10a. To demonstrate that U1 is involved in the splicing defect, a modified version of U1 snRNA was prepared in which the complementary cytidine to guanosine in the U1 snRNA restores normal base pairing with the mutant (Fig. 11.10d). Co-expression of this modified U1 snRNA with the minigene carrying the mutation resulted in rescue of exon 3 splicing (Fig. 11.10c). This experiment proves that the exon 3 +5G > C variation is a disease-causing mutation that induces aberrant skipping of exon 3 by interfering with the recognition of the 5′ splice site that in this context is sensitive to a shortening of the complementary region.

11.6.4 Identification of Deep Intronic Mutations: The Example of the ATM Gene

In the Ataxia telengectasia (ATM) gene a novel type of genetic mutation causes a splicing processing defect affecting an intronic element, which is important for correct intron removal (Pagani *et al.*, 2002).

This last example illustrates the importance of intronic sequences in the splicing process, even if they are located far away from the 5′ and 3′ splice sites. In the affected patient, sequence analysis revealed in genomic DNA a 4 bp deletion (GTAA) in intron 20, about 3 kb from exon 20 and 0.8 kb from exon 21 (Fig. 11.11a). To test if this deletion was a benign polymorphism or a disease causing mutation, two hybrid minigenes were constructed, one containing the normal ATM intron 20 and the other containing the intron deleted (Fig. 11.11b). Transfection experiments showed that normal ATM sequences were properly recognized as an intron by the splicing machinery and accordingly excluded in the mature transcript. On the contrary, the GTAA deletion caused the activation of a 65 nucleotide cryptic exon (Fig. 11.11c). Cryptic or pseudo exons are normally activated by intronic mutations that create or strengthen splice sites or create a new branch site. Interestingly, the GTAA deletion did not involve directly the splice sites of the cryptic exon (Fig. 11.11a). To study the mechanism involved in this new pathogenetic mutation, extensive site-directed mutagenesis was performed leading to the identification of an intronic regulatory sequence that was named intronic splicing processing element (ISPE). The ISPE is complementary to

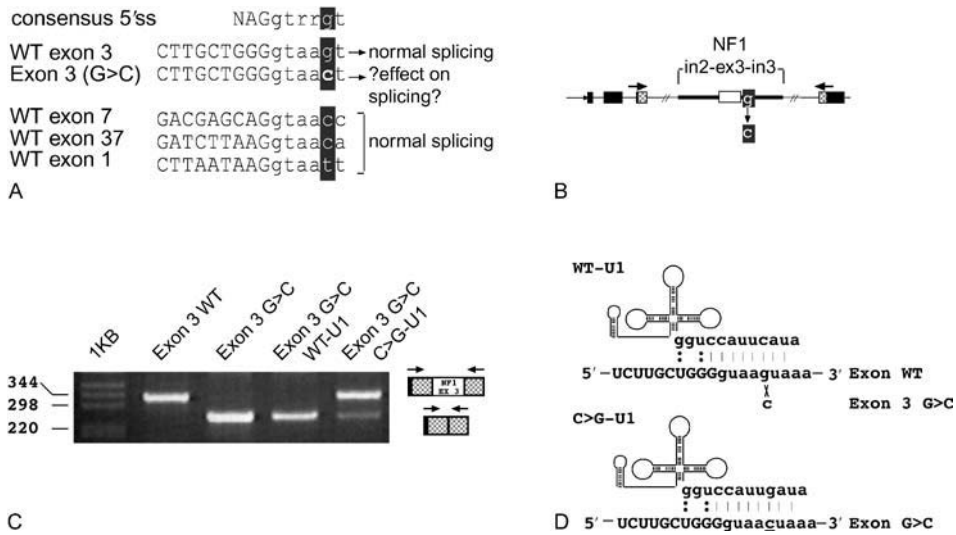


FIGURE 11.10 Identification of 5′ splice site variation in NF1 and correction with U1 snRNA interaction. **A.** Nucleotide sequence of the 5′ splice sites in WT and mutant exon 3 compared to the consensus and with the 5′ splice site of other NF1 exons. The G to C substitution in position +5 in exon 3 deviated from the consensus. However, also the normal 5′ splice site of exons 7, 37 and 1 deviates from the consensus at the same position, but are normally spliced *in vivo*. **B.** Hybrid minigenes containing normal and mutated NF1 exon 37 sequences. **C.** Hybrid minigenes transient transfection assay showing the effect of the exon 3 mutant and the effect of co-transfection of modified U1 snRNAs. The two bands correspond to inclusion (upper) or exclusion (lower) of NF1 exon 3 in the final transcripts. WT exon 3 is completely included in the mRNA while the exon 3 G to C substitution causes complete exon skipping. Co-transfection of WT-U1 has no effect on the splicing pattern of mutant exon 3. On the contrary, co-transfection of C > G-U1 induces inclusion of exon 3, indicating that the aberrant splicing defect is corrected in the presence of complementary U1 snRNA. **D.** Base pairing homology between the 5′ end of normal (WT-U1) and modified (C > G-U1) U1 snRNAs and the exon 3 5′ splice site sequences. The position of the G to C change is indicated. In the upper panel the G to C substitution disrupts the complementarity to normal U1. In the lower panel the modified U1 restores the base pair complementarity with the mutant. In other contexts such as the exons 1, 7 and 37 (panel a) such an extensive complementarity with U1 sn RNA seems not to be necessary, which highlights the difficulties found in predicting the effect of this type of mutation and the importance of functional assays.

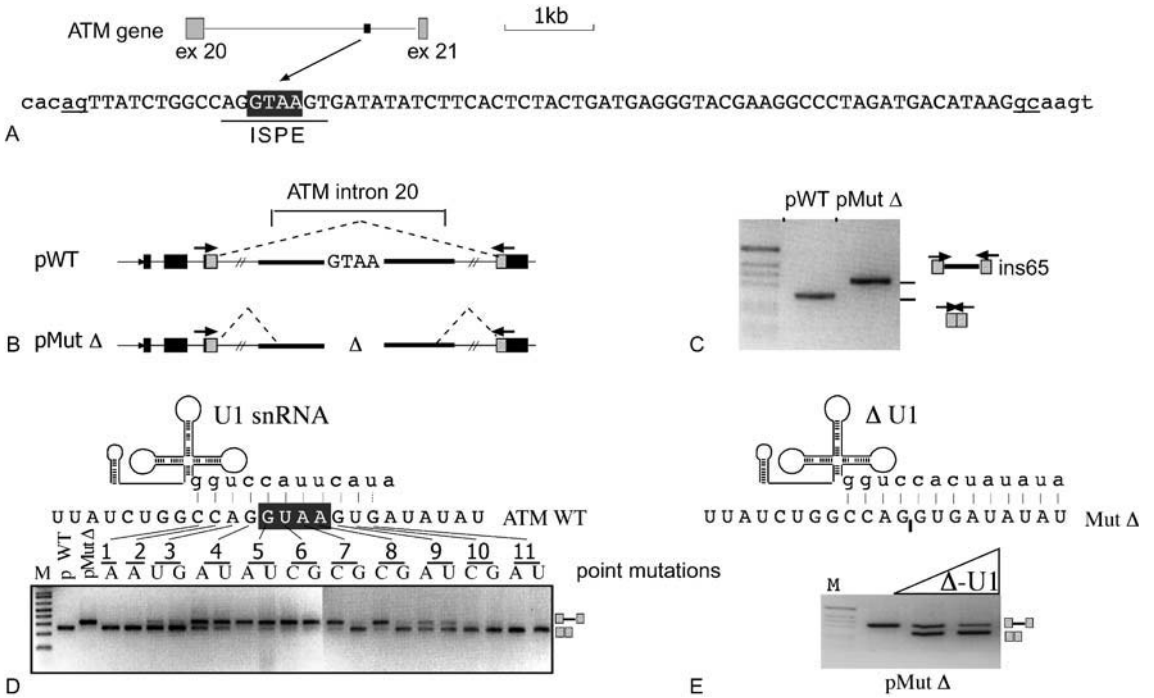


FIGURE 11.11 A splicing processing defect in ATM intron 20 affects a new intronic splicing processing element (ISPE). **A.** Identification of the splicing defect in the affected patient. A deletion was identified by DNA sequence analysis of the ATM gene and of the mRNA containing the cryptic exon shown in panel a. The sequence of the cryptic exon is shown in uppercase, intronic sequences in lowercase and the splice sites used are underlined. The 5′ splice site did not have the almost universal GU donor site but its weak GC variant. The mutated sequences detected in patients' cDNA and genomic DNA result from the deletion of the four intraexonic bases GTAA (boxed). **B.** The 4 bp deletion causes aberrant cryptic exon inclusion in hybrid minigene experiments. The hybrid minigenes contain WT intronic sequences and the ATM variant with the GTAA deletion MutΔ contains the patient's genomic sequences with the 4 nucleotide deletion. The primers used in the RT-PCR assay are shown as superimposed arrows. **C.** Hybrid minigenes were transfected in Hep3B cells and the splicing products analyzed with the specific primers. The agarose gel electrophoresis shows the processed transcripts that correspond to complete intron removal (WT) or inclusion of the 65 bp long cryptic exon (ins65). **D.** Base pairing homology between ATM ISPE and the 5′ end of wild-type U1 snRNA. It is clear that the ATM ISPE base pairing to U1 snRNA regulates intron splicing processivity. We introduced single point mutations, one at time, in the ATM minigenes at the ISPE in the entire U1 snRNA complementarity region between positions 1 and 11 (panel d). Hybrid minigene experiments with the different site-directed mutants showed that mutations between positions 4 and 9 induced cryptic exon inclusion. Interestingly, mutations 7G and 8G did not induce the aberrant splicing products while 7C and 8C did. This is because these variants preserve the complementarity to U1 by G:U "wobble" base pairing. **E.** Restoring U1 snRNAs-ISPE complementarity induces *in vivo* normal intron processivity. We evaluated the effect of mutant U1 complementary to MutΔ variant (Δ-U1) on the cryptic exon inclusion. The upper panel shows the base pairing homology between the 5′ end of the mutant U1 snRNA and the MutΔ minigene. The position of the deletion (Δ) in the ATM mRNAs is indicated. Lower panel shows MutΔ minigene variant co-transfected with increasing amounts of the indicated Δ-U1 snRNA. The aberrant splicing defect is corrected in the presence of complementary U1.

U1 snRNA, which was directly related to the activation of the cryptic exon (Fig. 11.11d). A clear functional interaction between U1 snRNP and the ISPE was obtained by complementation experiments with mutant U1 snRNAs engineered to bind specifically at ΔISPE (Fig. 11.11e). Co-transfection with increasing amounts of Δ-U1 snRNAs with pATMΔ showed the progressive disappearance of the cryptic exon, restoring normal intron processivity (Fig. 11.11e). These results indicate that the strength of the complementarity of the ISPE to U1 snRNA is functionally related to the splicing defect and modulates the efficiency of intron removal. Thus a new type of intronic U1 snRNP binding site (ISPE) performs an essential function for accurate intron removal. The function of the U1 snRNP-ISPE interaction is apparently quite different from the well-established interaction for the 5′ splice site initial recognition. Deletion of this sequence is directly involved in a splicing processing defect in a human disease that again could not have been predicted from simple DNA sequence inspection.

11.6.5 Mapping Intronic Splicing Regulatory Elements Using Ribozymes

Hybrid minigenes are also useful tools to identify in a particular disease gene splicing regulatory elements and to evaluate their contribution in aberrant splicing events. For example, the pathological effect of the polymorphic GUmUn allele described above in the CFTR exon 9 has been shown to be modulated by a downstream intronic splicing silencer, which unexpectedly bind to SR proteins (Pagani *et al.*, 2000). The classical way to identify these

elements is by making deletions or mutagenesis in the minigene and then evaluating the resulting splicing pattern. Comparison between percentages of exon inclusion derived from the normal and the mutant minigenes can identify if the sequence behaves like an enhancer or a silencer (i.e. if there is exon skipping, the target element is an enhancer; if there is exon inclusion, the sequence is a silencer). In general, deletion analysis is performed first, followed by an accurate mapping with point substitutions. In the case of intronic splicing regulatory elements, which can be far away from the exons, they can be identified using a novel approach that, through the insertion into the minigene of a ribozyme, break off the nascent transcript. Due to the exon tethering activity of the polymerase II this cleavage does not preclude splicing of the nascent transcript (Dye *et al.*, 2006). Once transcribed, the ribozyme of about 80 bp undergo rapid autocatalytic cleavage thus interrupting the continuity of the intronic nascent RNA (Fig. 11.12a). In this manner, any intronic sequences present on the nascent RNA cannot exert its regulatory effect resulting in changes of the splicing pattern. This system has been successfully used in the study of alternative splicing of the fibronectin EDA exon (Gromak *et al.*, 2008) (Fig. 11.12).

In the fibronectin EDA minigene, the alternatively spliced EDA exon is skipped to an 80% level. The engineered N117 hammerhead ribozyme (88 bp long) and its catalytically inactive mutant (14A to G) was inserted in two intronic positions at the indicated *Nco*I and *Nde*I sites downstream of the alternatively spliced EDA exon (Fig. 11.12b). The resulting constructs were transiently transfected into Hep3b cells and the EDA alternative splicing pattern was evaluated by RT-PCR analysis. Positioning of the active ribozyme, but not the mutant one at the *Nco*I site, induced a major decrease in the percentage of EDA exon skipping as compared to the parent pEDA (WT) construct. In contrast, the percentage of EDA exon skipping was unaffected in the Nde RZ constructs. The position-specific effect of RZ at the *Nco*I site on EDA skipping is due to the presence of a nearby negative downstream regulatory element (DRE), which cannot interact with the polymerase II tethered exon when the nascent transcript is co-transcriptionally cleaved. This DRE with silencing activity was identified also by classical deletion analysis. The invDRE and ΔDRE minigenes where the potential DRE between the *Nco*I and *Nde*I sites was either inverted or deleted showed a lower percentage of EDA exon skipping, similar to that seen in the Nco RZ construct (Fig. 11.12c).

These experiments suggest that positioning of the ribozyme at the *Nco*I site interferes with pre-mRNA processing by preventing negative interactions between

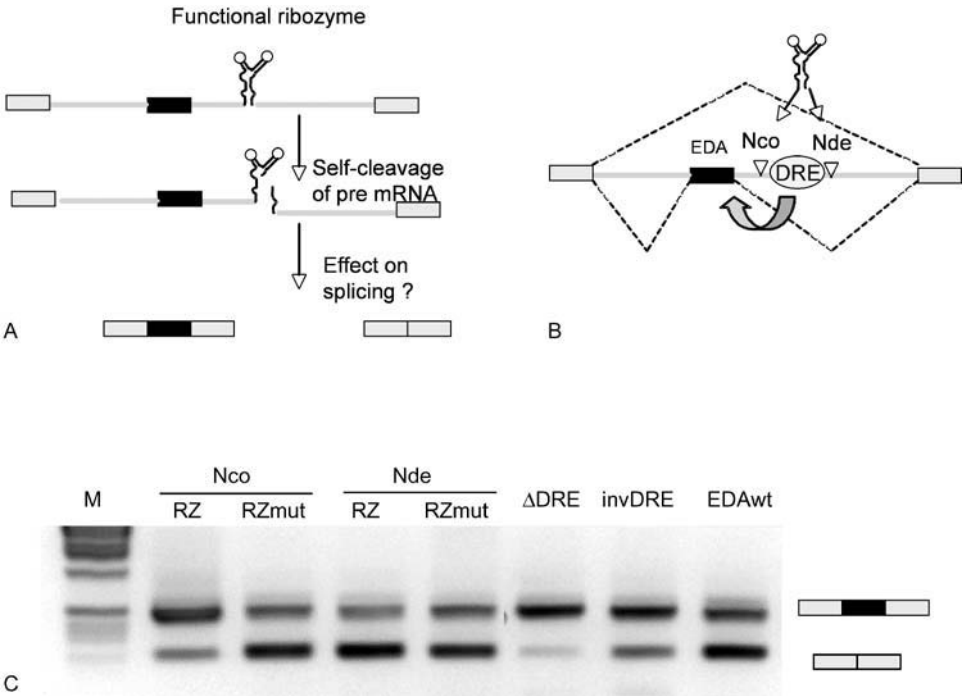


FIGURE 11.12 Identification of an intronic splicing regulatory element using a ribozyme. **A.** The ribozyme inserted in the intron of a minigene induces co-transcriptional cleavage of the nascent transcript, which may influence the splicing pattern according to the presence of a nearby splicing regulatory element. A mutant inactive ribozyme (RZmut) contains a single nucleotide substitution and is not self-cleaved (not shown). **B.** Diagram of the fibronectin EDA minigene. Fibronectin exons are boxed and introns are lines. The unique restriction sites in the EDA downstream intron, used for RZ insertion, are shown. **C.** RT-PCR analysis of total RNA from cells expressing each minigene. Minigenes transfected into Hep3B cells and RT-PCR products resolved on 1.5% agarose gels. RZ and RZmut are the 88 bp active and inactive hammerhead ribozymes. EDA exon inclusion and exclusion forms are indicated. ΔDRE and invDRE minigenes contain the deleted and inverted DRE sequences. Lane M, 1 kb-ladder size marker.

the EDA exon and the downstream DRE on the nascent transcript.

11.7 CONCLUSIONS

The genomic diversity and genomic pathology data currently available has revealed the extent of ignorance of the basic molecular mechanisms underlying the pre-mRNA splicing process (Teraoka *et al.*, 1999; Ars *et al.*, 2000; Cartegni *et al.*, 2002b). Even more worrying is the fact that a lot of clinically relevant mutations may be slipping through the net because their effect on the splicing process is not even considered. An increasing amount of evidence indicates that single nucleotide substitutions in both coding and non-coding sequences might have unexpected deleterious effects on the splicing of the gene transcript (Ars *et al.*, 2000; Cartegni *et al.*, 2002b; Pagani *et al.*, 2002, 2003a, b, 2005; Fernandez-Cadenas *et al.*, 2003; Kashima and Manley, 2003). The hybrid minigenes assay described in this chapter represents a useful tool both for the diagnosis of human splicing defects and for the identification of new basic splicing mechanisms. In fact, in the diagnostic field, the hybrid minigene assay can be used to distinguish between benign polymorphisms from disease-associated splicing mutations. Furthermore, human splicing errors identified by hybrid minigene assays represent a flag put on new splicing regulatory elements; this knowledge gives important insight into the basic molecular mechanism of pre-mRNA splicing.

REFERENCES

Ars, E., Serra, E., Garcia, J., Kruyer, H., Gaona, A., Lazaro, C., and Estivill, X. (2000). Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum. Mol. Genet.* 9, 237–247.

Ayala, Y.M., Pagani, F., and Baralle, F.E. (2006). *TDP43* depletion rescues aberrant CFTR exon 9 skipping. *FEBS Lett.* 580, 1339–1344.

Baralle, M., Baralle, D., De Conti, L., Mattocks, C., Whittaker, J., Knezevich, A., Ffrench-Constant, C., and Baralle, F.E. (2003). Identification of a mutation that perturbs NF1 agene splicing using genomic DNA samples and a minigene assay. *J. Med. Genet.* 40, 220–222.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336.

Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M., and Baralle, F.E. (2001). Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J.* 20, 1774–1784.

Buratti, E., Baralle, F.E., and Pagani, F. (2003). Can a "patch" in a skipped exon make the pre-mRNA splicing machine run better? *Trends Mol. Med.* 9, 229–232.

Buratti, E., Chivers, M., Kralovicova, J., Romano, M., Baralle, M., Krainer, A.R., and Vorechovsky, I. (2007). Aberrant 5′ splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res.* 35, 4250–4263.

Burge, C., Tuschl, T., and Sharp, P. (1999). Splicing of precursors to mRNAs by the spliceosomes. In *The RNA World* (Gesteland, R., Cech, T., and Atkins, J., eds), pp. 525–560. Cold Spring Harbor Press, Cold Spring Harbor, NY.

Caceres, J.F., and Kornblihtt, A.R. (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* 18, 186–193.

Cartegni, L., and Krainer, A.R. (2002a). Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat. Genet.* 30, 377–384.

Cartegni, L., Chew, S.L., and Krainer, A.R. (2002b). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 3(4), 285–298.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. (2003). ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* 31, 3568–3571.

Cartegni, L., Hastings, M.L., Calarco, J.A., de Stanchina, E., and Krainer, A.R. (2006). Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *Am. J. Hum. Genet.* 78, 63–77.

Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M., and Reed, R. (2006). Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev.* 20, 1100–1109.

Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R. (2007). SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* 26, 867–881.

Dye, M.J., Gromak, N., and Proudfoot, N.J. (2006). Exon tethering in transcription by RNA polymerase II. *Mol. Cell* 21, 849–859.

Fairbrother, W.G., Yeh, R.F., Sharp, P.A., and Burge, C.B. (2002). Predictive identification of exonic splicing enhancers in human genes. *Science* 297, 1007–1013.

Faustino, N.A., and Cooper, T.A. (2003). Pre-mRNA splicing and human disease. *Genes Dev.* 17, 419–437.

Fernandez-Cadenas, I., Andreu, A.L., Gamez, J., Gonzalo, R., Martin, M.A., Rubio, J.C. and Arenas, J. (2003). Splicing mosaic of the myophosphorylase gene due to a silent mutation in McArdle disease. *Neurology* 10, 1432–1434.

Goina, E., Skoko, N., and Pagani, F. (2008). Binding of DAZAP1 and hnRNPA1/A2 to an exonic splicing silencer in a natural BRCA1 exon 18 mutant. *Mol. Cell. Biol.* 18, 3850–3860.

Gromak, N., Talotti, G., Proudfoot, N.J., and Pagani, F. (2008). Modulating alternative splicing by cotranscriptional cleavage of nascent intronic RNA. *RNA* 14, 359–366.

Hoffmeyer, S., Nurnberg, P., Ritter, H., Fahsold, R., Leistner, W., Kaufmann, D., and Krone, W. (1998). Nearby stop codons in exons of the neurofibromatosis type 1 gene are disparate splice effectors. *Am. J. Hum. Genet.* 62, 269–277.

Kashima, T., and Manley, J.L. (2003). A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat. Genet.* 34, 460–463.

Kashima, T., Rao, N., David, C.J., and Manley, J.L. (2007). hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. *Hum. Mol. Genet.* 16, 3149–3159.

Khoo, B., Akker, S.A., and Chew, S.L. (2003). Putting some spine into alternative splicing. *Trends Biotechnol.* 21, 328–330.

Lorson, C.L., and Androphy, E.J. (2000). An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. *Hum. Mol. Genet.* 9, 259–265.

Lorson, C.L., Hahnen, E., Androphy, E.J., and Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. USA* 96, 6307–6311.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature* 416, 499–506.

Pagani, F., Buratti, E., Stuani, C., Romano, M., Zuccato, E., Niksic, M., Giglio, L., Faraguna, D., and Baralle, F.E. (2000). Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a non-evolutionary conserved intronic element. *J. Biol. Chem.* 275, 21041–21047.

Pagani, F., Buratti, E., Stuani, C., Bendix, R., Dork, T., and Baralle, F.E. (2002). A new type of mutation causes a splicing defect in ATM. *Nat. Genet.* 30, 426–429.

Pagani, F., Buratti, E., Stuani, C., and Baralle, F.E. (2003a). Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J. Biol. Chem.* 278, 26580–26588.

Pagani, F., Stuani, C., Tzetis, M., Kanavakis, E., Efthymiadou, A., Doudounakis, S., Casals, T., and Baralle, F.E. (2003b). New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. *Hum. Mol. Genet.* 12, 1111–1120.

Pagani, F., and Baralle, F.E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. *Nat. Rev. Genet.* 5, 389–396.

Pagani, F., Raponi, M., and Baralle, F.E. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc. Natl. Acad. Sci. USA* 102, 6368–6372.

Teraoka, S.N., Telatar, M., Becker-Catania, S., Liang, T., Onengut, S., Tolun, A., Chessa, L., Sanal, O., Bernatowska, E., Gatti, R.A., and Concannon, P. (1999). Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. *Am. J. Hum. Genet.* 64, 1617–1631.

Wang, G.S., and Cooper, T.A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* 8, 749–761.

Zhang, M.Q. (1998). Statistical features of human exons and their flanking regions. *Hum. Mol. Genet.* 7, 919–932.

Zhang, X.H., and Chasin, L.A. (2004). Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.* 18, 1241–1250.

This page intentionally left blank

Chapter 12

Detection of Genomic Duplications and Deletions

Graham R. Taylor[1] and Carol A. Delaney[2]

[1]*Regional Genetics Laboratory and CR-UK Mutation Detection Facility, St James's University Hospital, DNA Laboratory, Leeds, Yorkshire LS9 7TF, UK;*

[2]*SE Scotland Cytogenetics Laboratory, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK*

12.1 INTRODUCTION

Gene and chromosome duplications have been implicated as fundamental evolutionary mechanisms (Ohno *et al*.,1968; Li and Gojobori, 1983; Schughart *et al*., 1989). For the purpose of this chapter, duplications and deletions are defined as those too large to be routinely detected by sequencing or simple PCR-based methods (>100 bp), but too small to be detected by conventional cytogenetic metaphase analysis (<106 kb). Unlike other types of rearrangement, for example LINE insertions or balanced translocations, a duplication or deletion will be associated with a direct change in gene dosage. This four-order size range includes small fragments of genes, exons, entire genes, and multiple genes, and the same methods can, of course, be used to measure changes in chromosome count.

Historically, supernumerary chromosomes are the earliest "gene" duplications reported, revealed by cytogenetic techniques (Lejeune, 1960; Patau *et al*., 1961). Soon afterward, evidence for pathogenic α-globin gene deletions was reported (Ose and Bush, 1962), although more than 10 years elapsed before direct confirmation was achieved (Ottolenghi *et al*., 1974). β-Globin gene deletions were reported soon afterward (Kan *et al*., 1975), followed by gene deletions in other hemoglobinopathies and culminating in the use of Southern blotting for the prenatal diagnosis of globin deletions (Orkin *et al*., 1978). Gene dosage changes were next reported in the immunoglobulin genes (Rabbitts *et al*., 1980; van Loghem *et al*., 1980) and the first deletion of a tumor suppressor gene (in retinoblastoma) was reported soon afterward (Junien *et al*., 1982). The identification of repetitive DNA associated with a gene deletion was described in a form of hereditary persistence of fetal hemoglobin (Jagadeeswaran *et al*., 1982). A pathogenic role was attributed to repetitive DNA elements by Hess and coworkers (1983), who suggested that DNA insertion elements may disrupt gene correction processes in the two duplication units containing *HBA2* and *HBA1* genes. Although the widespread use of Southern blotting (Southern, 1975) from the mid-1970s until the late 1980s may have facilitated the detection of deletions and duplications, the application of the polymerase chain reaction (PCR; Mullis *et al*., 1986) may unintentionally have produced an ascertainment bias away from them, the analysis of dystrophin gene deletions in males being a notable exception (Beggs *et al*., 1990). This is because a typical PCR is not designed for quantitative analysis, but for optimal purity and yield. However, the observation of germ-line deletions in a wide range of genetic conditions has required the development of techniques that can detect gene dosage changes in hemizygotes. These techniques include Southern blotting and PCR modifications, as well as newer methods. Techniques that measure gene dosage can also be adapted to quantify the somatic mosaicism and PCR failure (allele dropout, see also Chapter 33).

12.2 MECHANISMS

Deletions and duplications can be mediated by homologous recombination involving recombinogenic elements, for example *PMP22* (Inoue *et al*., 2001); or gene duplications, for example the type IV collagen genes *COL4A5* and *COL4A6*, paired head to head on chromosome Xq22 deletions in X-linked Alport syndrome. *BRCA1* gene deletions involving a head-to-head of a partial pseudogene have been reported (Brown *et al*., 2002), as well as *Alu*-mediated intragenic deletions (Puget *et al*., 1997; Rohlfs *et al*., 2000). A 26 bp core sequence in two out of five α^0-thalassemia deletions has been reported (Harteveld *et al*., 1997), supporting the idea that *Alu* repeats stimulate recombination events not

only by homologous pairing, but also by providing binding sites for recombinogenic proteins. Deletion of subtelomeric repeats has been implicated in facioscapulohumeral muscular dystrophy (FSHD) where a remnant fragment, the result of a deletion of tandemly arrayed 3.3 kb repeat units (D4Z4) on 4q35 can be detected (Lemmers *et al*., 1998). Non-homologous recombination also has been implicated in gene deletions (Hu and Worton, 1992; Suminaga *et al*., 2000).

12.3 PATHOLOGICAL CONSEQUENCES

12.3.1 Supernumerary Chromosomes

In practice, when considering alterations in autosomal chromosome copy number, only supernumerary chromosomes need to be addressed, as autosomal monosomy is not compatible with life (unless the individual has monosomy mosaicism). Monosomy of the X chromosome can occur (Turner syndrome), although it is estimated that only 1 in 300 conceptuses with monosomy X survive (Kajii *et al*., 1980). Triploidy and trisomies of other autosomes are also a common cause of spontaneous abortion. From a clinical perspective, prenatal detection of triploidy and trisomies 13 and 18 are important as fetuses may survive until birth, but most die soon after due to congenital malformations. Trisomies that have less severe consequences are trisomy 21 (Down syndrome), trisomy X (rarely show any physical abnormalities), and Klinefelter syndrome (XXY). Supernumerary chromosomes are readily detectable by conventional cytogenetics (see also Chapter 10); however, molecular techniques may be faster and less expensive.

12.3.2 Microdeletions

Pathogenic partial deletions and duplications have been widely reported for all chromosomes, and cause a range of symptoms. Some deletions are easily detectable by conventional cytogenetics, others are smaller (microdeletions) and can be difficult or impossible to detect by conventional cytogenetics, and other cytogenetic techniques such as FISH or molecular techniques are required. Some of the more common microdeletions with associated syndromes (summarized in Table 12.1) are 22q11.2 (Di George syndrome), 15q11.2/q12 (Prader-Willi and Angelman syndromes), 17p (Miller-Dieker and Smith-Magenis syndromes), 4p16.3 (Wolf-Hirschhorn syndrome), and 5p15 (Cri du chat syndrome).

12.3.3 Subtelomeric Deletions

Several syndromes that are caused by microscopically visible chromosomal deletions and duplications, including the subtelomeric region, have been known to cause idiopathic mental retardation (Flint *et al*., 1995; Knight *et al*., 1999) (e.g. 1p, 4p, 5p, 9p). Despite their clinical relevance, screening for more cryptic alterations in the subtelomeres cannot be readily observed by G-banding analysis since most terminal bands are G-band negative. More recent studies using molecular methods such as detection of loss of hypervariable DNA polymorphisms or microsatellite markers and multiprobe FISH that detect submicroscopic subtelomeric deletions have shown that these can account for up to 5% of cases with mental retardation (Flint *et al*., 1995; Knight *et al*., 1999; Rio *et al*., 2002). Subtelomeric deletion analysis of all chromosomes can be performed rapidly and cost effectively using new molecular techniques such as MAPH (Sismani *et al*., 2001), MLPA (Rooms *et al*., 2006; Palomares *et al*., 2006), and array CGH (de Vries *et al*., 2005).

TABLE 12.1 Proportion (%) of mutations due to deletions or insertions in various genes.

Gene	Disorder	Deletions or duplications (%)	References
	Familial breast cancer	4–27	Puget *et al*. (1999); Gad *et al*. (2002); Hogervorst *et al*. (2003)
VHL	Von Hippel-Lindau disease	20–47	Shuin *et al*. (1995); Vortmeyer *et al*. (2002)
DMD	Duchenne muscular dystrophy	60	Koenig *et al*. (1989)
MSH2/MLH1	Hereditary non-polyposis colorectal cancer	27–54.8	Wijnen *et al*. (1998); Gille *et al*. (2002); Wagner *et al*. (2002)
RB1	Retinoblastoma	14	Richter *et al*. (2003)

12.3.4 Gene Deletions or Duplications

A proportion of some types of cancers are caused by inherited germ-line mutations in tumor suppressor genes, and full or partial gene deletions account for a significant number of these mutations. There are a range of other single gene disorders where deletion or duplication of part or all of the gene accounts for a significant proportion of detected mutations, such as Duchenne muscular dystrophy, spinal muscular atrophy, Charcot-Marie-Tooth disease,

Fanconi anemia, congenital adrenal hyperplasia, and rare metabolic disorders such as non-ketotic hyperglycinemia. Some examples of genes for which an estimate has been obtained for the proportion of germ-line mutations that are deletions or duplications are shown in Table 12.1.

Since the majority of mutations identified in the tumor suppressor genes lead to the production of a truncated product (Couch and Weber, 1996), until recently most studies performed mutation analysis on genomic DNA using PCR-based techniques such as sequencing, heteroduplex analysis (Chapter 4), or the protein truncation test (PTT; Chapter 19). Deletions (Petrij-Bosch *et al*., 1997; Puget *et al*., 1997; Swensen *et al*., 1997) or duplications (Puget *et al*., 1999) within the *BRCA1* gene, for example, would have been missed by conventional PCR-based methods and were detected by either reverse transcriptase PCR (RT-PCR) or Southern blotting. Although both PCR and Southern blotting have been adapted to provide quantitative data, PCR has become the method of choice for genetic testing. Estimates of gene dosage typically have been based on comparisons with a reference standard. Other approaches, including the study of junction fragments or microsatellite inheritance and more recently long-range PCR (Coulter-Mackie *et al*., 1998), FISH (Voskova-Goldman *et al*., 1997) and array-CGH (Bruder *et al*., 2001), have also been employed.

The overall contribution of deletions and duplications recorded in the May 2000 Human Gene Mutation database was 5.5% of reported mutations. As of March 2008 the proportion had increased to 22.7%, comprising 16% deletions and 6.5% duplications (http://www.hgmd.cf.ac.uk; Stenson *et al*., 2003). Given the greater technical difficulties in identifying deletions and duplications, this is still likely to be an underestimate. This emphasizes the importance of including the measurement of gene dosage in any comprehensive mutation scan. Numerous polymorphisms have been described involving deletion or duplication of large chromosomal segments, sometimes involving entire genes; for example, the common deletion polymorphisms of the cytochrome P450 gene *CYP2D6* (Meyer and Zanger, 1997), the theta-class glutathione S-transferase gene *GSTT1* (Wiencke *et al*., 1995), and the mu-type gene *GSTM1* (Brockmoller *et al*., 1992).

Regions close to telomeres are especially prone to interchromosomal rearrangements that can lead to different forms of presence/absence polymorphism, including the multiallelic variation in the structure of the 16p telomere (Wilkie *et al*., 1991), and the deletion polymorphism near the 12q telomere (Baird *et al*., 2000). The assembly of the human genome sequence is complicated by the presence of duplication/deletion polymorphisms, either because only one form of a region of variable structure is recorded in the sequenced chromosome (Siniscalco *et al*., 2000) or because the presence of unsuspected polymorphism for tandem duplication can lead to missassembly (Bailey *et al*., 2001, 2002).

12.4 DIAGNOSTIC TECHNIQUES

Classical cytogenetic genetic techniques have been valuable in identifying supernumerary chromosomes and large deletions. However, the resolution of these methods is limited to several megabases and they would miss many known submicroscopic deletions. Fluoresence *in situ* hybridization (FISH) techniques have greatly expanded the capabilities of cytogenetics (see also Chapter 10); single copy probes allow detection of microdeletions, which would be impossible to detect by conventional cytogenetics; however, due to the specialized equipment and expertise required and the low throughput, detection of submicroscopic deletions has largely become the realm of specialized molecular genetic diagnostics. The detection of submicroscopic duplications by cytogenetics remains a challenge; the use of interphase FISH together with image analysis is being used increasingly.

12.4.1 Southern Blotting

Southern blotting still has a role in a comprehensive diagnostic service, both for gene dosage measurement and for the measurement of allele expansions. Southern blots can detect deletions or duplications by the identification of novel restriction fragments created by the rearrangement or by measurement of band intensity compared to a control fragment.

Fragment size changes are seen because the region of genomic DNA created by a deletion or duplication may have gained or lost restriction enzyme sites. In those cases, a Southern blot of a genomic digest using an appropriate restriction enzyme will detect novel fragments. There is a risk that some altered fragment sizes may simply be due to restriction fragment length polymorphisms, although the risk can be reduced by performing separate digestions with a different enzyme. Direct visualization of fragment sizes has long been used for detecting α-globin duplications and deletions (Orkin *et al*., 1978), and the increased size range available in pulse field gel electrophoresis (PFGE) allows detection of deletions and duplications even in genes as large as the dystrophin gene (Kenwrick *et al*., 1987).

Estimation of gene dosage by measuring the intensity of probe hybridization (usually in comparison with a control) has identified several instances of gene deletions (Bonifas and Epstein, 1990), though few studies report details of the dose–response curve (such as its linear range) and internal controls are often lacking. Dystrophin deletion carrier testing is possible by assessing relative band intensity (van Essen *et al*., 1997). Accurate measurement of band intensity requires a phosphoimager, which gives a linear dose response over a wider range than film emulsion. Figure 12.1 shows the application of Southern blotting to measure the relative gene dosage of *CYP21A* and *CYP21B*, which can be deleted in congenital adrenal hyperplasia. The risk of inaccuracies introduced by uneven transfer of DNA to the membrane or incomplete washing of the probe requires

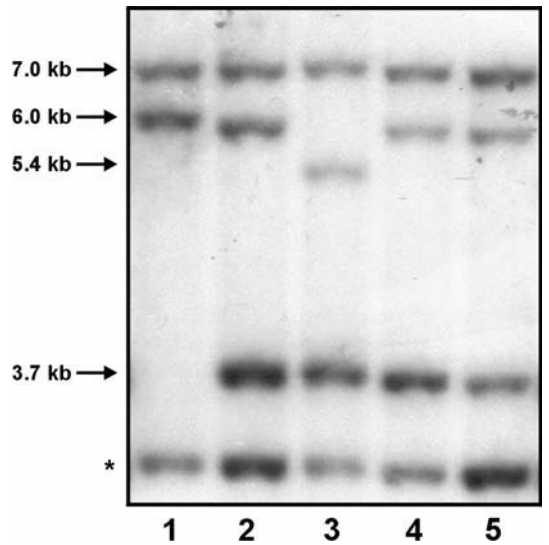


FIGURE 12.1 Gene dosage in 21 hydroxylase by Southern blotting. The Southern blot shows four major bands: C4A (7 kb), C4B (6 or 5.4 kb), 21-hydroxylase (3.7 kb), and 21-hydroxylase pseudogene (indicated by an asterisk). Sample 1 has a homozygous deletion of 21-OH, sample 2 has equal gene dosage at all four loci. Samples 3 and 4 each have deletions of the 21-hydroxylase pseudogene and C4A, and sample 5 has a deletion of 21-hydroxylase and C4B. (Image courtesy of Dr. Kieran Bransfield, Leeds Institute of Molecular Medicine, Leeds.)

that an additional probe for a control locus should be included as a standard.

12.4.2 Microsatellites and SNPs

Microsatellites have been used to detect supernumerary chromosomes (Mansfield, 1993) and large deletions or duplications (Brice *et al*., 1992). This approach is limited by the fact that semi-quantitative PCR methods do not work well using microsatellites, so only cases in which the allele lengths of the microsatellites differ over chromosomal region of interest can be interpreted unambiguously. Peak heights or areas of microsatellites can be hard to interpret because of preferential amplification of the smaller allele and the stuttering effect that produces a series of minor peaks immediately adjacent to the major peak. The difficulties with multiplex quantitative fluorescent PCR (QF-PCR) are reduced by selection of tetranucleotide repeat markers and is finding increasing use in the rapid diagnosis of common aneuploidies (Donaghue *et al*., 2003). Loss of heterozygosity of single nucleotide polymorphisms (SNPs) using high-throughput Affymetrix human SNP arrays has been proposed as a rapid means of identifying allele imbalance caused by genomic deletions in tumor cells (Hoque *et al*., 2003).

12.4.3 Polymerase Chain Reaction

Neubauer and colleagues (1990) described the co-amplification of two amplicons followed by quantitation on the basis of staining intensity as a means of estimating gene dosage. Differential PCR-based methods are semi-quantitative: they determine relative concentrations of two amplicons, but not absolute molar amounts. Taking advantage of the increased sensitivity of fluorescent detection methods Yau and coworkers (1996) described a multiplex fluorescent PCR that was able to detect female carriers of deletions or duplications in the dystrophin gene. Using peak areas, they were able to give statistical estimates of their assay, likely to be essential for diagnostic applications. Similar methods have been used for testing for hereditary motor and sensory neuropathy (HMSN) duplications (Rowland *et al*., 2001) and APC deletions (Flintoff *et al*., 2001). Casilli and coworkers (2002) refined the basic multiplex fluorescent PCR method by selecting shorter fragments, tagging the primers with common tags of 16 nucleotides, and using a modified PCR buffer containing DMSO. The method was termed quantitative multiplex PCR of short fluorescent fragments (QMSFPCR) and enables rapid design of amplicons to fine-map the limits of a deletion or duplication (Casilli *et al*., 2002). Although differential PCR methods offer high orders of multiplicity (e.g. 8–15 amplicons per reaction), end-point multiplex PCR assays rely on the equivalence of amplification of each fragment in the multiplex in the test sample as well as a control sample. This equivalence may be lost if the starting template DNA concentration is too variable, since different fragments within the multiplex may be amplified with different efficiencies. In some situations it may be necessary to choose a more robust, lower order multiplex base on real-time PCR. Real-time PCR (also known as quantitative PCR; qPCR) provides a means for continuous detection and quantification of product throughout the amplification process, and as such can dispense with a gel separation stage and operate in a closed system (see also Chapter 7). The accumulation of PCR product is monitored by staining using interchelating dyes (e.g. *SYBR* Green) or by dual-labeled probes such as TaqMan (Laurendeau *et al*., 1999), molecular beacons (Tyagi and Kramer, 1996), or other fluorescent detection systems. Real-time PCR is becoming widely used as a method for measuring gene dosage (Feldkotter *et al*., 2002; Bertin *et al*., 2003; Covault *et al*., 2003; Gaikovitch *et al*., 2003; Kim *et al*., 2003). Real-time quantitative PCR has been used for screening large gene and chromosomal rearrangements (Ariani *et al*., 2004). Additionally, real-time PCR can increase the sensitivity of mutation detection especially in double-copy genes such as X-linked methyl-CpG-binding protein 2 (*MECP2*) gene, where current mutation scanning techniques such as DGGE, SSCP, DHPLC and direct sequencing are prone to miss gross rearrangements.

12.4.4 MAPH and MLPA

The multiplex amplifiable probe hybridization (MAPH) method for copy-number measurement (Armour *et al*., 2000) combines hybridization as the primary step to detect copy number with end-point multiplex PCR to amplify the

hybridized probes. Sets of short probes corresponding to the segments to be tested, each flanked by the same primer-binding sites, are hybridized with the test genomic DNA immobilized on a solid support. After washing, each specifically bound probe will be present in an amount proportional to its copy number. All probes can then be amplified simultaneously with a single primer pair, and quantified after electrophoretic separation. This study demonstrated the simultaneous assessment of copy number in a set of 40 human loci, including detection of deletions causing Duchenne muscular dystrophy and Prader-Willi/Angelman syndromes. The high order multiplex achieved may be limited only by the need to have probes of varying size for electrophoretic separation. MAPH probes are generated by cloning the target sequences into a plasmid vector, amplifying the cloned sequence using primers directed to the vector with the result that all probes are then flanked with the same sequence. Probes that are intended to be multiplexed must be of sufficient size difference to be resolved by electrophoresis. The membranes are then washed rigorously to remove unbound probe, and the remaining specifically bound probe will be present in an amount proportional to its target copy number. The probes are then stripped from the membrane by boiling, and amplified simultaneously with the universal primer pair. Products are then separated by electrophoresis, and a relative comparison is made between the peak heights. Reduced peak heights compared to internal control probes indicate a reduction in gene copy number (deletion) and an increase in gene copy number (duplication) produces increased peak heights. The assay can be completed in two to three days, requiring one overnight hybridization followed by membrane washing, a PCR step, and a product detection step. The system works well, but the manipulation of small nylon filters presents some difficulties in sample handling and labeling.

A similar technique called multiplex ligatable probe amplification (MLPA) avoids the use of filters by using single-stranded ligatable probes and a thermo-stable DNA ligase to produce the amplifiable target (Schouten *et al*., 2002; see also Chapter 13). In the MLPA technique, genomic DNA is hybridized in solution to probe sets, each of which consists of two halves. One half consists of a target-specific sequence (20–30 nucleotides) flanked by a universal primer sequence, and can be generated synthetically. The other half also has a target-specific sequence at one end (25–43 nucleotides) and a universal primer sequence at the other, but has a variable length stuffer fragment in between (19–370 nucleotides) to generate the size differences necessary in the probes to allow electrophoretic resolution. This larger probe part is generated by cloning the target-specific sequence into M13 derived vectors that already contain the variable length fragments; single-stranded DNA is then purified from the phage particles and made double stranded at two sites by annealing short oligonucleotides in order that the desired probe fragment be liberated by restriction enzyme digestion. The two probe halves are designed such that the target-specific sequences bind adjacently to the target DNA, and can then be joined by use of a ligase. This generates a contiguous probe flanked by universal primer binding sites that can then be amplified by PCR, whereas unbound probe halves cannot be amplified, and hence eliminates the need for removal of excess probe by washing. The amounts of ligated probe produced will be proportional to the target copy number, and after PCR amplification the relative peak heights indicate deletion or duplication of target sequence.

In diagnostic use, applications of these high order multiplex methods include the detection of aneuploidies, unbalanced cryptic translocations, and whole or partial gene duplications or deletions. Commercial kits based on MAPH and MLPA are available (http://www.mrc-holland.com). The high multiplicity of MAPH and MLPA, plus their use of standard genetic laboratory apparatus, make it highly likely that these techniques will become widely used. Both techniques offer a rapid means of scanning up to 40 loci for gene dosage, and are likely to be used widely in research and diagnostic settings. MAPH represents a conceptual breakthrough for the analysis of gene dosage, but the handling of small filter discs is difficult and could pose sample tracking problems in routine medium throughput settings. The liquid-phase solution to sample handling offered by MLPA and the ready availability of robust commercial kits has led to rapid acceptance by diagnostic laboratories worldwide. MLPA uses M13 (single-stranded) probes that are more technically challenging to construct than MAPH probes, which can be made from PCR products.

There is an ever-growing range of MLPA probe sets and kits supplied by MRC-Holland. However, it is also possible to construct additional probes to enable fine mapping and identification of the junction fragments/deletion breakpoints by designing your own synthetic MLPA probes. The advantage of synthetic probes, as compared to the phage M13-derived probes made by MRC-Holland, is that the cloning step is omitted, so they are available sooner, but disadvantages are that fewer probes can be used in a single MLPA reaction, approximately 11 bespoke probes can be added to a probemix. Protocols assisting design of synthetic probes are available from the MRC-Holland website or using AlleleID probe design software (http://www.premierbiosoft.com/datafiles/alleleIDWin.exe). Alternatively, additional loci could be examined by adapting MLPA to an array detection setting, so that fragments were identified by sequence rather than size. This would enable all MLPA probes to be chemically synthesized. Array adaptations of both MAPH and MLPA might enable much higher order multiplexes than are currently possible. However, even 40-plex assays represent a substantial gain in the multiplex order typically available in PCR. MAPH or MLPA could provide competition for the array-based CGH approaches: a 96-well array of 40 probes could interrogate over 3,000 loci, representing a better than 1 cM coverage of the entire human genome or much higher single chromosome resolution. Two color MLPA kits are

available increasing the number of loci examined to some extent by including two probe sets which are amplified by two sets of primers: one labeled with FAM and the other with HEX. An example is the MLPA kit to detect deletions and duplications in the *CBP* (Crebb binding protein) gene. In all, this kit has synthetic probes for 20 out of 32 exons of *CBP*, and also five control probes for unlinked loci. The probes span the *CBP* gene from exons 2 to 32 (the coding region of the gene) and, due to the number of probes, is able to detect deletions and duplications in CBP in cases with Rubinstein-Taybi syndrome previously undetectable by FISH (Roelfsema *et al*., 2005).

12.4.5 Long-range PCR

Long-range PCR (Barnes, 1992) is a modified PCR protocol that includes a proof-reading polymerase and short denaturation times to enable the size of amplicons to be increased from 3–5 kb to beyond 30 kb. This opens the possibility of screening for deletions of this scale by direct PCR, with the advantage of producing the junction fragment for further analysis. Long-range PCR has been used successfully to identify deletions including mitochondrial DNA (Fromenty *et al*., 1996), C4 gene deletions in the MHC complex (Grant *et al*., 2000), the CYP2D6 deletion allele, and LDL deletions (Kim *et al*., 1999). Despite the simplicity in principle of a long-range PCR approach, in practice it is often difficult to design a robust assay. Direct detection of duplications by amplification of entire duplicated regions is likely to prove difficult in diploid genomes since the unduplicated allele will have a significant advantage during the amplification process. However, full characterization of a deletion or duplication requires the sequence of the junction fragments. This can be achieved either by using long-range PCR or by fine mapping, for example using QMSF-PCR followed by conventional PCR and sequencing.

12.4.6 Array-Based CGH

Microarrays have become widely used tools for gene expression studies, and now encouraging developments are taking place in array-based comparative genome hybridization (Array-CGH) as a tool for measuring gene duplications and deletions. Comparative genomic hybridization originally used metaphase chromosomes as targets for differentially labeled probes (e.g. Cy5, Cy3) from control and test samples. Gene dosage changes could be detected by variation of the relative intensity of the two labels (Kallioniemi *et al*., 1992). By replacing the metaphase spread with microarrayed BAC DNA, Pinkel and colleagues (1998) reported the use of array-based CGH analysis to investigate chromosome 20 gene dosage alterations in breast cancer. DNA purified from BAC clones spaced at approximately 3 Mb intervals along the entire chromosome together with some X chromosome controls were arrayed on nitrocellulose membranes as 200–400 micron spots in duplicate. Test DNA from breast cancer cell-lines was labeled by nick translation with fluorescein, control DNA with Texas Red, and the spots were counterstained with DAPI. Scanning used custom-built mercury-arc illumination and CCD detection. Other studies have used different dyes (with Cy3/Cy5 being a popular choice) and omitted counterstaining. Many detection systems use confocal laser scanners, which, although more restricted in fluorochrome options, are brighter and have fewer problems with light scattering.

Growing high numbers (>104) of cloned DNA are expensive and labor intensive. To circumvent large-scale cultures and DNA purifications, array features have been made using degenerate oligonucleotide PCR (DOP-PCR) of BAC templates using 5′ amine-linked primers (Hodgson *et al*., 2001). Fiegler and coworkers (2003) improved the standard DOP-PCR to reduce non-specific host and vector amplicons from BAC clones, improving the signal-to-noise ratio. A ligation-mediated PCR BAC labeling method has also been reported that improved signal-to-noise ratios (Snijders *et al*., 2001). Array CGH is still in development (see also Chapters 10 and 12), but strong commercial and academic interest are likely to result in an increase in research and diagnostic applications in the near future. The supply of robust competitor (Cot1) DNA has been a major problem reported by many array-CGH users (Carter *et al*., 2002). Cot1 DNA is used to block the hybridization of repetitive DNA in BAC-derived amplicons. Buckley and coworkers (2002) constructed a comprehensive microarray representing a human chromosome for analysis of DNA copy-number variation. The chromosome 22 microarray covered 34.7 Mb with an average resolution of 75 kb using a sequence-defined, repeat-free, and non-redundant strategy for array preparation. This enabled an increase in array resolution and eliminated the need for Cot1 DNA. Array targets were made using phi29 DNA polymerase synthesis.

As the technology becomes more widely used, commercial suppliers are starting to emerge; for example, the Cytochip array from BlueGnome (http://www.cytochip.com) is designed for diagnostic use. This is a BAC microarray with an average genome coverage of approximately 565 kb, with median coverage of 250 kb in subtelomeric regions and an average resolution of 100 kb in 90 known clinically relevant regions (microdeletion/duplication syndromes). Higher resolution oligonucleotide-based microarrays for the detection of copy-number changes are becoming increasingly popular allowing genome-wide analysis which can map chromosomal imbalance breakpoints at exon level resolution, including imbalances that are single copy-number genomic alterations (Selzer *et al*., 2005). Commercial suppliers of oligonucleotide-based arrays include Agilent Technologies Inc. (http://www.agilent.com). Several Agilent array CGH platforms exist including arrays with 44,000 60-mer oligonucleotide probes covering both coding and

non-coding human sequences with an average genome resolution of 50 kb through to 244,000 probes with a mediam probe spacing of 9 kb. Other suppliers of high-resolution oligonucleotide-based arrays to detect gene dosage changes include Illumina (http://www.illumina.com) and Affymetrix (http://www.affymetrix.com). Additionally, genome centers, for example, Leiden University in Holland and the Sanger Centre in the UK, are using their resources to manufacture genome- or chromosome-specific arrays.

Since array-CGH enables high-resolution detection of structural variation, this technique is moving from a research tool into diagnostic service provision. Confirmation of the array CGH results are often needed prior to issuing a diagnostic report. Several copy number/dosage techniques confirm array CGH results, including FISH, MLPA, STR analysis, real-time PCR/qPCR, long-range PCR, and, ultimately, DNA sequencing. The choice of confirmation method will depend upon several factors, including the size of the region involved, the precise location of the abnormality, availability of FISH or MLPA probes, the density of STRs, the material available from the proband and the parents (DNA, metaphase spreads, etc.), and access to equipment. Often there is an interplay between several techniques to determine copy-number variation per case (Fig. 12.2).

One of the findings of using higher-resolution array CGH for genomic-wide copy-number changes is the awareness of the significant degree of natural copy-number variation in the genome. At 100 kb array CGH resolution Redon and coworkers (2006) reported an average of 70 polymorphisms per patient and a total of 1,116 unique polymorphisms on over 270 patients. Similarly, de Vries and coworkers (2005) reported 4–9 copy-number variants per patient in a cohort of mental retardation patients, where 96% of these variations were found to be inherited polymorphisms.

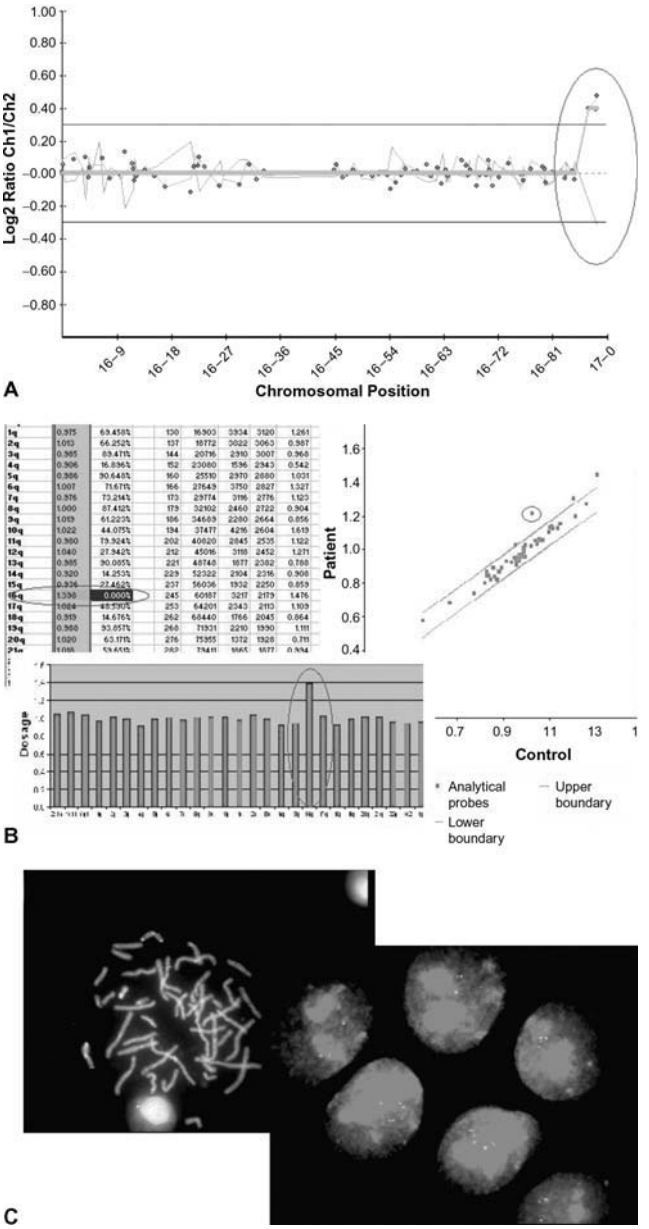


FIGURE 12.2 Detection of a duplication of 16qter in a patient with unexplained learning difficulties using: **A.** A 1 Mb BAC array from BlueGnome. **B.** The MLPA subtelomere kit from MRC-Holland confirms the duplication, and **C.** Interphase FISH using a Vysis subtelomeric probe. Duplication of 16qter is shown by three copies of the red signal as indicated by arrows. *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

12.5 NOMENCLATURE

Recently, the Human Genome Variation Society (http://www.hgvs.org) has taken on the responsibility of establishing a standardized mutation nomenclature that includes deletions and duplications. A nomenclature reference is maintained at this site by Dr. Johan Den Dunnen, Leiden University Medical Centre. The current recommendation for exon or multi-exon deletion nomenclature varies, depending on whether or not the breakpoint has been identified. If the breakpoints are not sequenced (e.g. detected on Southern blot or by MLPA), exonic deletions are described as c.88-?_923+?del, indicating a deletion starting at an unknown position in the intron 5′ of cDNA nucleotide 88 and ending at an unknown position in the intron 3′ of cDNA nucleotide 923. A genomic or cDNA reference (e.g. Genbank, EMBL, DDJB) should be cited, including the version number. If the junction fragments are known, then the form g.390_1458del or (g.390_1458del1069) should be used for a genomic reference sequence or c.13-23_301-143del (c.13-23_301-143del1069) for a cDNA reference sequence.

12.6 GENE DOSAGE APPLICATIONS IN TUMOR PROFILING

Partial chromosomal losses occur relatively frequently in a large number of tumor types, and this is readily detectable by demonstrating loss of heterozygosity (LOH) of polymorphic

microsatellite markers. Such techniques have been widely employed to investigate various aspects of tumor development:

1. Clonality. Loss of chromosomal regions is essentially an irreversible event in tumor cell development, and all subsequent subclones would be expected to demonstrate the same deletions, and may have accumulated more. LOH analysis of multiple tumors from the same patient can be used to determine whether tumors, either synchronous or metachronous, are clonal in origin, and in the latter case may be used to determine whether subsequent tumors are recurrences or metastases of previous tumors or independent primary tumors. For example, in patients with multiple synchronous lung tumors it is important for treatment decisions to discriminate multicentric lung cancers from intrapulmonary metastases, and this can be aided by clonality studies (Shimizu *et al*., 2000).
2. Identification of tumor suppressor genes. It is hypothesized that the reason that tumors frequently undergo LOH, and that LOH of particular regions is commonly associated with specific tumor types, is that the regions lost harbor tumor suppressor genes. Identification of such tumor suppressor genes may provide useful tools for diagnosis, prognosis, and possible therapies. Hence, numerous studies have been undertaken to identify regions of LOH specific to individual tumor types and the genes present in these regions. For example, chromosome 9 is the most frequently deleted chromosome in transitional cell carcinoma of the bladder, and candidate genes have been identified in three of the four regions of minimal deletion (Knowles, 1999). In prostate cancer, gains at Xq and 18q are among the most common chromosomal alterations, and amplification of the AR (Xq12), MYC (8q24), and EIF3S3 (8q23) genes have been found in a large fraction of hormone-refractory prostate cancers (Nupponen and Visakorpi, 2000).
3. Diagnosis. LOH analysis has been employed both in detection of the presence of a tumor, as well as differential diagnosis and typing of tumors. For example, LOH has been reliably demonstrated in DNA extracted from urine sediments in bladder cancer patients, and may have a role in the non-invasive diagnosis of bladder cancer (Linn *et al*., 1997; Berger *et al*., 2002). Differential diagnosis between renal oncocytomas and renal cell carcinomas can be difficult due to morphological similarities, but is important due to their different prognoses. LOH analysis has shown that they can be differentiated on the basis of spectrum of chromosomal loss (Herbers *et al*., 1998). LOH analysis has also been used to characterize the aggressive intraductal carcinoma of the prostate, to differentiate it from the less aggressive high-grade dysplasia (prostatic intraepithelial neoplasia, PIN), and to provide evidence that it does not represent invasion of Gleason grade 3 cancers into the ductal/acinar system (Dawkins *et al*., 2000).
4. Prognosis. In several tumor types, LOH of specific regions have been hypothesized to be key events in tumor evolution and progression. LOH analysis of tumor specimens has been used to determine their value as prognostic markers. It has been found in colorectal cancers that high level LOH correlated with earlier onset and lymphatic invasion, and hence a poorer prognosis, and low level LOH was more common in earlier stage disease and predicted a more favorable outcome (Choi *et al*., 2002). In breast cancers, specific chromosomal regions have been identified for which LOH is a significant predictor of lymph-node metastasis and hence may serve as a negative prognostic indicator (Nagahata *et al*., 2002).

12.7 SUMMARY AND FUTURE DEVELOPMENTS

Gene dosage is a significant contributor to the overall burden of the germ-line and somatic mutations in man. Any comprehensive mutation screen therefore should include measurement of gene dosage. The permeation of genomics into medical practice will increase the demand for mutation screening with applications in diagnosis, predictive testing, and treatment. This, in turn, will encourage the development of more highly automated approaches, the introduction of robust statistical analysis, and quality control of results that may be used in a diagnostic context. Those techniques are most easily adaptable to robust laboratory processes and automated data handling will have a competitive advantage. Ease of handling, hence more robust results, perhaps explains the greater uptake of MLPA compared with MAPH. Both MAPH and MLPA produce data using automated DNA sequencers that are spreadsheet-ready. Statistical parameters may then be easily applied as quality measures (Taylor *et al*., 2003). The same is true of array scanner output and real-time PCR systems, in contrast to classical genetic techniques like Southern blotting or cytogenetics. Whereas MLPA, real-time PCR, QMSF-PCR, and MAPH approaches are focused around relatively low numbers of targets, the array methods offer very high orders of multiplicity, the so-called ''hypothesis-free'' approach. It seems likely that the two approaches will converge as arrays become increasingly targeted toward regions or functional clusters of interest and the PCR-based approaches move toward an array-like format, either in capillary arrays or high density microtiter arrays.

Finally, as clonal sequencing technology matures it may displace the current generation of array technology. This reflects advantages in the clonal/digital nature of the output in contrast to the analog output from arrays, the unbiased readout, not dictated by what has been arrayed, and the ability to read 30–400 base strings of DNA sequence rather than infer sequence results from hybridization to arrayed sequence of 20 to 80 bases. Structural variations, including copy-number variants, can be readily detected using paired

reads, although the sensitivity and specificity of this method has not yet been established. Campbell and coworkers (2008) characterized 306 germ-line structural variants and 103 somatic rearrangements to the base-pair level of resolution.

REFERENCES

Ariani, F., Mari, F., Pescucci, C., Longo, I., Bruttini, M., Meloni, I., Hayek, G., Rocchi, R., and Zappella, M. (2004). Real time quantitative PCR as a routine method for screening large rearrangements in Rett syndrome. Report of one case of MECP2 deletion and one case of MECP2 duplication. *Hum. Mutat.* 24, 172–177.

Armour, J.A., Sismani, C., Patsalis, P.C., and Cross, G. (2000). Measurement of locus copy number by hybridisation with amplifiable probes. *Nucleic Acids Res.* 28, 605–609.

Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., and Eichler, E.E. (2001). Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* 11, 1005–1017.

Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. (2002). Recent segmental duplications in the human genome. *Science* 297, 1003–1007.

Baird, D.M., Coleman, J., Rosser, Z.H., and Royle, N.J. (2000). High levels of sequence polymorphism and linkage disequilibrium at the telomere of 12q: implications for telomere biology and human evolution. *Am. J. Hum. Genet.* 66, 235–250.

Barnes, W.M. (1992). The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* 112, 29–35.

Beggs, A.H., Koenig, M., Boyce, F.M., and Kunkel, L.M. (1990). Detection of 98% of DMD/BMD gene deletions by polymerase chain reaction. *Hum. Genet.* 86, 45–48.

Berger, A.P., Parson, W., Stenzl, A., Steiner, H., Bartsch, G., and Klocker, H. (2002). Microsatellite alterations in human bladder cancer: detection of tumor cells in urine sediment and tumor tissue. *Eur. Urol.* 41, 532–539.

Bertin, R., Acquaviva, C., Mirebeau, D., Guidal-Giroux, C., Vilmer, E., and Cave, H. (2003). CDKN2A, CDKN2B, and MTAP gene dosage permits precise characterization of mono and bi-allelic 9p21 deletions in childhood acute lymphoblastic leukemia. *Genes Chromosomes Cancer* 37, 44–57.

Bonifas, J.M., and Epstein, E.H., Jr. (1990). Detection of carriers for X-linked ichthyosis by Southern blot analysis and identification of one family with a de novo mutation. *J. Invest. Dermatol.* 95, 16–19.

Brice, A., Ravise, N., Stevanin, G., Gugenheim, M., Bouche, P., Penet, C., and Agid, Y. (1992). Duplication within chromosome 17p11.2 in 12 families of French ancestry with Charcot-Marie-Tooth disease type 1a. The French CMT Research Group. *J. Med. Genet.* 29, 807–812.

Brockmoller, J., Gross, D., Kerb, R., Drakoulis, N., and Roots, I. (1992). Correlation between trans-stilbene oxide-glutathione conjugation activity and the deletion mutation in the glutathione S-transferase class mu gene detected by polymerase chain reaction. *Biochem. Pharmacol.* 43, 647–650.

Brown, M.A., Lo, L.J., Catteau, A., Xu, C.F., Lindeman, G.J., Hodgson, S., and Solomon, E. (2002). Germline BRCA1 promoter deletions in UK and Australian familial breast cancer patients: identification of a novel deletion consistent with BRCA1:psiBRCA1 recombination. *Hum. Mutat.* 19, 435–442.

Bruder, C.E., Hirvela, C., Tapia-Paez, I., Fransson, I., Segraves, R., Hamilton, G., Zhang, X.X., Evans, D.G., Wallace, A.J., Baser, M.E., Zucman-Rossi, J., Hergersberg, M., Boltshauser, E., Papi, L., Rouleau, G.A., Poptodorov, G., Jordanova, A., Rask-Andersen, H., Kluwe, L., Mautner, V., Sainio, M., Hung, G., Mathiesen, T., Moller, C., Pulst, S.M., Harder, H., Heiberg, A., Honda, M., Niimura, M., Sahlen, S., Blennow, E., Albertson, D.G., Pinkel, D., and Dumanski, J.P. (2001). High resolution deletion analysis of constitutional DNA from neurofibromatosis type 2 (NF2) patients using microarray-CGH. *Hum. Mol. Genet.* 10, 271–282.

Buckley, P.G., Mantripragada, K.K., Benetkiewicz, M., Tapia- Paez, I., Diaz De Stahl, T., Rosenquist, M., Ali, H., Jarbo, C., De Bustos, C., Hirvela, C., Sinder Wilen, B., Fransson, I., Thyr, C., Johnsson, B.I., Bruder, C.E., Menzel, U., Hergersberg, M., Mandahl, N., Blennow, E., Wedell, A., Beare, D.M., Collins, J.E., Dunham, I., Albertson, D., Pinkel, D., Bastian, B.C., Faruqi, A.F., Lasken, R.S., Ichimura, K., Collins, V.P., and Dumanski, J.P. (2002). A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications. *Hum. Mol. Genet.* 11, 3221–3229.

Carter, N.P., Fiegler, H., and Piper, J. (2002). Comparative analysis of comparative genomic hybridization microarray technologies: report of a workshop sponsored by the Wellcome Trust. *Cytometry* 49, 43–48.

Casilli, F., Di Rocco, Z.C., Gad, S., Tournier, I., Stoppa-Lyonnet, D., Frebourg, T., and Tosi, M. (2002). Rapid detection of novel BRCA1 rearrangements in high-risk breast-ovarian cancer families using multiplex PCR of short fluorescent fragments. *Hum. Mutat.* 20, 218–226.

Campbell, P.J., Stephens, P.J., Pleasance, E.D., O'Meara, S., Li, H., Santarius, T., Stebbings, L.A., Leroy, C., Edkins, S., Hardy, C., Teague, J.W., Menzies, A., Goodhead, I., Turner, D.J., Clee, C.M., Quail, M.A., Cox, A., Brown, C., Durbin, R., Hurles, M.E., Edwards, P.A., Bignell, G.R., Stratton, M.R., and Futreal, P.A. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* 40, 722–729.

Choi, S.W., Lee, K.J., Bae, Y.A., Min, K.O., Kwon, M.S., Kim, K.M., and Rhyu, M.G. (2002). Genetic classification of colorectal cancer based on chromosomal loss and microsatellite instability predicts survival. *Clin. Cancer Res.* 8, 2311–2322.

Couch, F.J., and Weber, B.L. (1996). Mutations and polymorphisms in the familial early-onset breast cancer (BRCA1) gene. Breast Cancer Information Core. *Hum. Mutat.* 8, 8–18.

Coulter-Mackie, M.B., Applegarth, D.A., Toone, J.R., and Gagnier, L. (1998). A protocol for detection of mitochondrial DNA deletions: characterization of a novel deletion. *Clin. Biochem.* 31, 627–632.

Covault, J., Abreu, C., Kranzler, H., and Oncken, C. (2003). Quantitative real-time PCR for gene dosage determinations in microdeletion genotypes. *Biotechniques* 35, 594–596, 598.

Dawkins, H.J., Sellner, L.N., Turbett, G.R., Thompson, C.A., Redmond, S.L., McNeal, J.E., and Cohen, R.J. (2000). Distinction between intraductal carcinoma of the prostate (IDC-P), high-grade dysplasia (PIN), and invasive prostatic adenocarcinoma, using molecular markers of cancer progression. *Prostate* 44, 265–270.

De Vries, B.B., Pfundt, R., Leisink, M., Kooles, D.A., Lisenka, E.L., Vissers, M., Janssen, I.M., van Reijmarsdal, S., Nillesen, W.M., Huys, E.H., de Leeuw, N., Smeets, D., Sistermans, E.A., Feuth, T., van Ravenswaaijarti, C.M., van Kessel, A.G., Schoenmakers, E.F., Brunner, H.G., and Veltman, J.A. (2005). Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.* 77, 606–616.

Donaghue, C., Roberts, A., Mann, K., and Ogilvie, C.M. (2003). Development and targeted application of a rapid QF-PCR test for sex chromosome imbalance. *Prenat. Diagn.* 23, 201–210.

Feldkotter, M., Schwarzer, V., Wirth, R., Wienker, T.F., and Wirth, B. (2002). Quantitative analyses of SMN1 and SMN2 based on real-time lightCycler PCR: fast and highly reliable carrier testing and prediction of severity of spinal muscular atrophy. *Am. J. Hum. Genet.* 70, 358–368.

Fiegler, H., Carr, P., Douglas, E.J., Burford, D.C., Hunt, S., Scott, C.E., Smith, J., Vetrie, D., Gorman, P., Tomlinson, I.P., and Carter, N.P. (2003). DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer* 36, 361–374.

Flint, J., Wilkie, A.O., Buckle, V.J., Winter, R.M., Holland, A.J., and McDermid, H.E. (1995). The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nat. Genet.* 9, 132–140.

Flintoff, K.J., Sheridan, E., Turner, G., Chu, C.E., and Taylor, G.R. (2001). Submicroscopic deletions of the APC gene: a frequent cause of familial adenomatous polyposis that may be overlooked by conventional mutation scanning. *J. Med. Genet.* 38, 129–132.

Fromenty, B., Manfredi, G., Sadlock, J., Zhang, L., King, M.P., and Schon, E.A. (1996). Efficient and specific amplification of identified partial duplications of human mitochondrial DNA by long PCR. *Biochim. Biophys. Acta.* 1308, 222–230.

Gad, S., Caux-Moncoutier, V., Pages-Berhouet, S., Gauthier-Villars, M., Coupier, I., Pujol, P., Frenay, M., Gilbert, B., Maugard, C., Bignon, Y.J., Chevrier, A., Rossi, A., Fricker, J.P., Nguyen, T.D., Demange, L., Aurias, A., Bensimon, A., and Stoppa-Lyonnet, D. (2002). Significant contribution of large BRCA1 gene rearrangements in 120 French breast and ovarian cancer families. *Oncogene* 21, 6841–6847.

Gaikovitch, E.A., Cascorbi, I., Mrozikiewicz, P.M., Brockmoller, J., Frotschl, R., Kopke, K., Gerloff, T., Chernov, J.N., and Roots, I. (2003). Polymorphisms of drug-metabolizing enzymes CYP2C9, CYP2C19, CYP2D6, CYP1A1, NAT2 and of Pglycoprotein in a Russian population. *Eur. J. Clin. Pharmacol.* 59, 303–312.

Gille, J.J., Hogervorst, F.B., Pals, G., Wijnen, J., Van Schooten, R.J., Dommering, C.J., Meijer, G.A., Craanen, M.E., Nederlof, P.M., De Jong, D., McElgunn, C.J., Schouten, J.P., and Menko, F.H. (2002). Genomic deletions of MSH2 and MLH1 in colorectal cancer families detected by a novel mutation detection approach. *Br. J. Cancer* 87, 892–897.

Grant, S.F., Kristjansdottir, H., Steinsson, K., Blondal, T., Yuryev, A., Stefansson, K., and Gulcher, J.R. (2000). Long PCR detection of the C4A null allele in B8-C4AQ0-C4B1-DR3. *J. Immunol. Methods* 244, 41–47.

Harteveld, K.L., Losekoot, M., Fodde, R., Giordano, P.C., and Bernini, L.F. (1997). The involvement of Alu repeats in recombination events at the alpha-globin gene cluster: characterization of two alphazero-thalassaemia deletion breakpoints. *Hum. Genet.* 99, 528–534.

Herbers, J., Schullerus, D., Chudek, J., Bugert, P., Kanamaru, H., Zeisler, J., Ljungberg, B., Akhtar, M., and Kovacs, G. (1998). Lack of genetic changes at specific genomic sites separates renal oncocytomas from renal cell carcinomas. *J. Pathol.* 184, 58–62.

Hess, J.F., Fox, M., Schmid, C., and Shen, C.K. (1983). Molecular evolution of the human adult alpha-globin-like gene region: insertion and deletion of Alu family repeats and non-Alu DNA sequences. *Proc. Natl. Acad. Sci. USA* 80, 5970–5974.

Hodgson, G., Hager, J.H., Volik, S., Hariono, S., Wernick, M., Moore, D., Nowak, N., Albertson, D.G., Pinkel, D., Collins, C., Hanahan, D., and Gray, J.W. (2001). Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. *Nat. Genet.* 29, 459–464.

Hogervorst, F.B., Nederlof, P.M., Gille, J.J., McElgunn, C.J., Grippeling, M., Pruntel, R., Regnerus, R., van Welsem, T., van Spaendonk, R., Menko, F.H., Kluijt, I., Dommering, C., Verhoef, S., Schouten, J.P., van't Veer, L.J., and Pals, G. (2003). Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. *Cancer Res.* 63, 1449–1453.

Hoque, M.O., Lee, J., Begum, S., Yamashita, K., Engles, J.M., Schoenberg, M., Westra, W.H., and Sidransky, D. (2003). High-throughput molecular analysis of urine sediment for the detection of bladder cancer by high-density single-nucleotide polymorphism array. *Cancer Res.* 63, 5723–5726.

Hu, X., and Worton, R.G. (1992). Partial gene duplication as a cause of human disease. *Hum. Mutat.* 1, 3–12.

Inoue, K., Dewar, K., Katsanis, N., Reiter, L.T., Lander, E.S., Devon, K.L., Wyman, D.W., Lupski, J.R., and Birren, B. (2001). The 1.4-Mb CMT1A duplication/HNPP deletion genomic region reveals unique genome architectural features and provides insights into the recent evolution of new genes. *Genome Res.* 11, 1018–1033.

Jagadeeswaran, P., Tuan, D., Forget, B.G., and Weissman, S.M. (1982). A gene deletion ending at the midpoint of a repetitive DNA sequence in one form of hereditary persistence of fetal haemoglobin. *Nature* 296, 469–470.

Junien, C., Despoisse, S., Turleau, C., Nicolas, H., Picard, F., Le Marec, B., Kaplan, J.C., and de Grouchy, J. (1982). Retinoblastoma, deletion 13q14, and esterase D: application of gene dosage effect to prenatal diagnosis. *Cancer Genet. Cytogenet.* 6, 281–287.

Kajii, T., Ferrier, A., Niikawa, N., Takahara, H., Ohama, K., and Avirachan, S. (1980). Anatomic and chromosomal anomalies in 639 spontaneous abortuses. *Hum. Genet.* 55, 87–98.

Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., and Pinkel, D. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258, 818–821.

Kan, Y.W., Holland, J.P., Dozy, A.M., Charache, S., and Kazazian, H.H. (1975). Deletion of the beta-globin structure gene in hereditary persistence of foetal haemoglobin. *Nature* 258, 162–163.

Kenwrick, S., Patterson, M., Speer, A., Fischbeck, K., and Davies, K. (1987). Molecular analysis of the Duchenne muscular dystrophy region using pulsed field gel electrophoresis. *Cell* 48, 351–357.

Kim, S.H., Bae, J.H., Chae, J.J., Kim, U.K., Choe, S.J., Namkoong, Y., Kim, H.S., Park, Y.B., and Lee, C.C. (1999). Long-distance PCR-based screening for large rearrangements of the LDL receptor gene in Korean patients with familial hypercholesterolemia. *Clin. Chem.* 45, 1424–1430.

Kim, S.W., Lee, K.S., Jin, H.S., Lee, T.M., Koo, S.K., Lee, Y.J., and Jung, S.C. (2003). Rapid detection of duplication/deletion of the PMP22 gene in patients with Charcot-Marie-Tooth disease Type 1A and hereditary neuropathy with liability to pressure palsy by real-time quantitative PCR using SYBR Green I dye. *J. Korean Med. Sci.* 18, 727–732.

Knight, S.J., Regan, R., Nicod, A., Horsley, S.W., Kearney, L., Homfray, T., Winter, R.M., Bolton, P., and Flint, J. (1999). Subtle chromosomal rearrangements in children with unexplained mental retardation. *Lancet* 354, 1676–1681.

Knowles, M.A. (1999). Identification of novel bladder tumour suppressor genes. *Electrophoresis* 20, 269–279.

Koenig, M., Beggs, A.H., Moyer, M., Scherpf, S., Heindrich, K., Bettecken, T., Meng, G., Muller, C.R., Lindlof, M., Kaariainen, H., de la Chapelle, A., Kiuru, A., Savontaus, M.-L., Gilgenkrantz, H., Recan, D., Chelly, J., Kaplan, J.-C., Covone, A.E., Archidiacono, N., Romeo, G., Liechti-Gallati, S., Schneider, V., Braga, S., Moser, H., Darras, B.T., Murphy, P., Francke, U., Chen, J.D., Morgan, G., Denton, M., Greenberg, C.R., van Ommen, G.J., and Kunkel, L.M. (1989). The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. *Am. J. Hum. Genet.* 45, 498–506.

Laurendeau, I., Bahuau, M., Vodovar, N., Larramendy, C., Olivi, M., Bieche, I., Vidaud, M., and Vidaud, D. (1999). TaqMan PCR-based

gene dosage assay for predictive testing in individuals from a cancer family with INK4 locus haploinsufficiency. *Clin. Chem.* 45, 982–986.

Lejeune, J. (1960). Mongolism, regressive trisomy. *Ann. Genet.* 2, 1–38.

Lemmers, R.J., van der Maarel, S.M., van Deutekom, J.C., van der Wielen, M.J., Deidda, G., Dauwerse, H.G., Hewitt, J., Hofker, M., Bakker, E., Padberg, G.W., and Frants, R.R. (1998). Inter- and intrachromosomal sub-telomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. *Hum. Mol. Genet.* 7, 1207–1214.

Li, W.H., and Gojobori, T. (1983). Rapid evolution of goat and sheep globin genes following gene duplication. *Mol. Biol. Evol.* 1, 94–108.

Linn, J.F., Lango, M., Halachmi, S., Schoenberg, M.P., and Sidransky, D. (1997). Microsatellite analysis and telomerase activity in archived tissue and urine samples of bladder cancer patients. *Int. J. Cancer* 74, 625–629.

Mansfield, E.S. (1993). Diagnosis of Down syndrome and other aneuploidies using quantitative polymerase chain reaction and small tandem repeat polymorphisms. *Hum. Mol. Genet.* 2, 43–50.

Meyer, U.A., and Zanger, U.M. (1997). Molecular mechanisms of genetic polymorphisms of drug metabolism. *Annu. Rev. Pharmacol. Toxicol.* 37, 269–296.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51, 263–273.

Nagahata, T., Hirano, A., Utada, Y., Tsuchiya, S., Takahashi, K., Tada, T., Makita, M., Kasumi, F., Akiyama, F., Sakamoto, G., Nakamura, Y., and Emi, M. (2002). Correlation of allelic losses and clinicopathological factors in 504 primary breast cancers. *Breast Cancer* 9, 208–215.

Neubauer, A., Neubauer, B., and Liu, E. (1990). Polymerase chain reaction based assay to detect allelic loss in human DNA: loss of beta-interferon gene in chronic myelogenous leukemia. *Nucleic Acids Res.* 18, 993–998.

Nupponen, N.N., and Visakorpi, T. (2000). Molecular cytogenetics of prostate cancer. *Microsc. Res. Tech.* 51, 456–463.

Ohno, S., Wolf, U., and Atkin, N.B. (1968). Evolution from fish to mammals by gene duplication. *Hereditas* 59, 169–187.

Orkin, S.H., Alter, B.P., Altay, C., Mahoney, M.J., Lazarus, H., Hobbins, J.C., and Nathan, D.G. (1978). Application of endonuclease mapping to the analysis and prenatal diagnosis of thalassemias caused by globin-gene deletion. *N. Engl. J. Med.* 299, 166–172.

Ose, T., and Bush, O.B. Jr. (1962). Erythroblastosis fetalis. Report of cases, with one caused by gene deletion. *Bibl. Haematol.* 13, 290–291.

Ottolenghi, S., Lanyon, W.G., Paul, J., Williamson, R., Weatherall, D.J., Clegg, J.B., Pritchard, J., Pootrakul, S., and Boon, W.H. (1974). The severe form of alpha thalassaemia is caused by a haemoglobin gene deletion. *Nature* 251, 389–392.

Palomares, M., Delicado, A., Lapunzina, P., Ajona, D., Aminosa, C., Arcas, J., Martinez Bermejo, A., Fernandez, L., and Lopez-Pajares, I. (2006). MLPA versus multiprobe FISH comparison of 2 methods for the screening of subtelomere rearrangements in 50 patients with idiopathic mental retardation. *Clin. Genet.* 69, 228–233.

Patau, K., Therman, E., Smith, D.W., and Demars, R.I. (1961). Trisomy for chromosome No. 18 in man. *Chromosoma* 12, 280–285.

Petrij-Bosch, A., Peelen, T., van Vliet, M., van Eijk, R., Olmer, R., Drusedau, M., Hogervorst, F.B., Hageman, S., Arts, P.J., Ligtenberg, M.J., Meijers-Heijboer, H., Klijn, J.G., Vasen, H.F., Cornelisse, C.J., van 't Veer, L.J., Bakker, E., van Ommen, G.J., and Devilee, P. (1997). BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. *Nat. Genet.* 17, 341–345.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W., and Albertson, D.G. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* 20, 207–211.

Puget, N., Torchard, D., Serova-Sinilnikova, O.M., Lynch, H.T., Feunteun, J., Lenoir, G.M., and Mazoyer, S. (1997). A 1-kb Alu-mediated germline deletion removing BRCA1 exon 17. *Cancer Res.* 57, 828–831.

Puget, N., Sinilnikova, O.M., Stoppa-Lyonnet, D., Audoynaud, C., Pages, S., Lynch, H.T., Goldgar, D., Lenoir, G.M., and Mazoyer, S. (1999). An Alu-mediated 6-kb duplication in the BRCA1 gene: a new founder mutation? *Am. J. Hum. Genet.* 64, 300–302.

Rabbitts, T.H., Forster, A., Dunnick, W., and Bentley, D.L. (1980). The role of gene deletion in the immunoglobulin heavy chain switch. *Nature* 283, 351–356.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapiro, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalactzopoulos, D., Komua, D., MacDonald, J.R., Maynall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchuda, J., Valsessia, A., Woodwark, C., Yang, F., Zhang, J., Zerial, T., Zhang, J., Armengol, L., Conrad, D.F., Estivil, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., and Hurles, M.E. (2006). Global variation in copy number in the human genome. *Nature* 144, 428–429.

Richter, S., Vandezande, K., Chen, N., Zhang, K., Sutherland, J., Anderson, J., Han, L., Panton, R., Branco, P., and Gallie, B. (2003). Sensitive and efficient detection of RB1 gene mutations enhances care for families with retinoblastoma. *Am. J. Hum. Genet.* 72, 253–269.

Rio, M., Molinari, F., Heuertz, S., Ozilou, C., Gosset, P., Raoul, O., Cormier-Daire, V., Amiel, J., Lyonnet, S., Le Merrer, M., Turleau, C., de Blois, M.C., Prieur, M., Romana, S., Vekemans, M., Munnich, A., and Colleaux, L. (2002). Automated fluorescent genotyping detects 10% of cryptic subtelomeric rearrangements in idiopathic syndromic mental retardation. *J. Med. Genet.* 39, 266–270.

Roelfsema, J.H., White, S.J., Ariyurek, Y., Bartholdi, D., Niedrist, D., Papadia, F., Bacino, C.A., den Dunnen, J.T., van Ommen, G.J., Breuning, M.H., Hennekam, R.C., and Peters, D.J. (2005). Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. *Am. J. Med. Genet.* 76, 572–580.

Rohlfs, E.M., Puget, N., Graham, M.L., Weber, B.L., Garber, J.E., Skrzynia, C., Halperin, J.L., Lenoir, G.M., Silverman, L.M., and Mazoyer, S. (2000). An Alu-mediated 7.1 kb deletion of BRCA1 exons 8 and 9 in breast and ovarian cancer families that results in alternative splicing of exon 10. *Genes Chromosomes Cancer* 28, 300–307.

Rooms, L., Reyniers, E., Wuyts, W., Storm, K., van Luijk, R., Scheers, S., Wauters, J., van den Ende, J., Biervliet, M., Eyskens, F., van Goethem, G., Laridon, A., Ceulemans, B., Courtens, W., and Kooy, R.F. (2006). Multiplex ligation-dependent probe amplification to detect subtelomeric rearrangements in routine diagnosis. *Clin. Genet.* 69, 58–64.

Rowland, J.S., Barton, D.E., and Taylor, G.R. (2001). A comparison of methods for gene dosage analysis in HMSN type 1. *J. Med. Genet.* 38, 90–95.

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., and Pals, G. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* 30, e57.

Schughart, K., Kappen, C., and Ruddle, F.H. (1989). Duplication of large genomic regions during the evolution of vertebrate homeobox genes. *Proc. Natl. Acad. Sci. USA* 86, 7067–7071.

Selzer, R.R., Richmond, T.A., Pofahl, N.J., Green, R.D., Eis, P.S., Nair, P., Brothman, A.R., and Stallings, R.L. (2005). Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer* 44, 305–319.

Shimizu, S., Yatabe, Y., Koshikawa, T., Haruki, N., Hatooka, S., Shinoda, M., Suyama, M., Ogawa, M., Hamajima, N., Ueda, R., Takahashi, T., and Mitsudomi, T. (2000). High frequency of clonally related tumors in cases of multiple synchronous lung cancers as revealed by molecular diagnosis. *Clin. Cancer Res.* 6, 3994–3999.

Shuin, T., Kondo, K., Kaneko, S., Sakai, N., Yao, M., Hosaka, M., Kanno, H., Ito, S., and Yamamoto, I. (1995). Results of mutation analyses of von Hippel-Lindau disease gene in Japanese patients: comparison with results in United States and United Kingdom. *Hinyokika Kiyo* 41, 703–707.

Siniscalco, M., Robledo, R., Orru, S., Contu, L., Yadav, P., Ren, Q., Lai, H., and Roe, B. (2000). A plea to search for deletion polymorphism through genome scans in populations. *Trends Genet.* 16, 435–437.

Sismani, C., Armour, J.A., Flint, J., Girgalli, C., Regan, R., and Patsalis, P.C. (2001). Screening for subtelomeric chromosome abnormalities in children with idiopathic mental retardation using multiprobe telomeric FISH and the new MAPH telomeric assay. *Eur. J. Hum. Genet.* 9, 527–532.

Snijders, A.M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J.P., Gray, J.W., Jain, A.N., Pinkel, D., and Albertson, D.G. (2001). Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.* 29, 263–264.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503–517.

Stenson, P.D., Ball, E.V., Mort, M., Phillips, A.D., Shiel, J.A., Thomas, N.S., Abeysinghe, S., Krawczak, M., and Cooper, D.N. (2003). Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.* 21, 577–581.

Suminaga, R., Takeshima, Y., Yasuda, K., Shiga, N., Nakamura, H., and Matsuo, M. (2000). Non-homologous recombination between Alu and LINE-1 repeats caused a 430-kb deletion in the dystrophin gene: a novel source of genomic instability. *J. Hum. Genet.* 45, 331–336.

Swensen, J., Hoffman, M., Skolnick, M.H., and Neuhausen, S.L. (1997). Identification of a 14 kb deletion involving the promoter region of BRCA1 in a breast cancer family. *Hum. Mol. Genet.* 6, 1513–1517.

Taylor, C.F., Charlton, R.S., Burn, J., Sheridan, E., and Taylor, G.R. (2003). Genomic deletions in MSH2 or MLH1 are a frequent cause of hereditary non-polyposis colorectal cancer: identification of novel and recurrent deletions by MLPA. *Hum. Mutat.* 22, 428–433.

Tyagi, S., and Kramer, F.R. (1996). Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* 14, 303–308.

van Essen, A.J., Kneppers, A.L., van der Hout, A.H., Scheffer, H., Ginjaar, I.B., ten Kate, L.P., van Ommen, G.J., Buys, C.H., and Bakker, E. (1997). The clinical and molecular genetic approach to Duchenne and Becker muscular dystrophy: an updated protocol. *J. Med. Genet.* 34, 805–812.

van Loghem, E., Sukernik, R.I., Osipova, L.P., Zegers, B.J., Matsumoto, H., de Lange, G., and Lefranc, G. (1980). Gene deletion and gene duplication within the cluster of human heavy-chain genes. Selective absence of IgG sub-classes. *J. Immunogenet.* 7, 285–299.

Vortmeyer, A.O., Huang, S.C., Pack, S.D., Koch, C.A., Lubensky, I.A., Oldfield, E.H., and Zhuang, Z. (2002). Somatic point mutation of the wild-type allele detected in tumors of patients with VHL germline deletion. *Oncogene* 21, 1167–1170.

Voskova-Goldman, A., Peier, A., Caskey, C.T., Richards, C.S., and Shaffer, L.G. (1997). DMD-specific FISH probes are diagnostically useful in the detection of female carriers of DMD gene deletions. *Neurology* 48, 1633–1638.

Wagner, A., van der Klift, H., Franken, P., Wijnen, J., Breukel, C., Bezrookove, V., Smits, R., Kinarsky, Y., Barrows, A., Franklin, B., Lynch, J., Lynch, H., and Fodde, R. (2002). A 10-Mb paracentric inversion of chromosome arm 2p inactivates MSH2 and is responsible for hereditary nonpolyposis colorectal cancer in a North-American kindred. *Genes Chromosomes Cancer* 35, 49–57.

Wiencke, J.K., Pemble, S., Ketterer, B., and Kelsey, K.T. (1995). Gene deletion of glutathione S-transferase theta: correlation with induced genetic damage and potential role in endogenous mutagenesis. *Cancer Epidemiol. Biomarkers Prev.* 4, 253–259.

Wijnen, J., van der Klift, H., Vasen, H., Khan, P.M., Menko, F., Tops, C., Meijers Heijboer, H., Lindhout, D., Moller, P., and Fodde, R. (1998). MSH2 genomic deletions are a frequent cause of HNPCC. *Nat. Genet.* 20, 326–328.

Wilkie, A.O., Higgs, D.R., Rack, K.A., Buckle, V.J., Spurr, N.K., Fischel-Ghodsian, N., Ceccherini, I., Brown, W.R., and Harris, P.C. (1991). Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* 64, 595–606.

Yau, S.C., Bobrow, M., Mathew, C.G., and Abbs, S.J. (1996). Accurate diagnosis of carriers of deletions and duplications in Duchenne/Becker muscular dystrophy by fluorescent dosage analysis. *J. Med. Genet.* 33, 550–558.

Chapter 13

Multiplex Ligation-Dependent Probe Amplification (MLPA) and Methylation-Specific (MS)-MLPA: Multiplex Detection of DNA/mRNA Copy Number and Methylation Changes

Eline M. Sepers and Jan P. Schouten
MRC-Holland BV, Amsterdam, The Netherlands

13.1 INTRODUCTION

Multiplex ligation-dependent probe amplification (MLPA) is a high-throughput and sensitive polymerase chain reaction (PCR)-based technique detecting copy-number changes in DNA or RNA. MLPA can quantify up to 50 nucleic acid sequences in one simple reaction, with a resolution down to the single nucleotide level (Schouten *et al.*, 2002). Since its introduction in 2002, over 500 peer-reviewed articles describing its use have been published, including several reviews of the technique itself (Sellner and Taylor, 2004; den Dunnen and White, 2006; Chou *et al.*, 2008). Some MLPA variants, such as methylation-specific MLPA (MS-MLPA) (Nygren *et al.*, 2005), that have been developed more recently are now also in widespread use.

The first and still most common application of MLPA is the detection of small genomic rearrangements which are a frequent cause of disease (see also Chapter 12). Although in the majority of genes, the deletion or duplication of one or more complete exons accounts for less than 10% of all disease-causing mutations, for many others this percentage can be as high as 10–30% (Taylor *et al.*, 2003; Kluwe *et al.*, 2005; Michils *et al.*, 2005; Aldred *et al.*, 2006; Aretz *et al.*, 2007; Kanno *et al.*, 2007; Redeker *et al.*, 2008) or higher still (Beetz *et al.*, 2006; Depienne *et al.*, 2007). In case of the very long dystrophin gene, involved in Duchenne muscular dystrophy, exon deletions and duplications even account for 65–70% of all mutations (Schwartz and Dunø, 2004; Janssen *et al.*, 2005).

Since most methods which are used to detect point mutations, like DNA sequencing and denaturing high-performance liquid chromatography (DHPLC), generally fail to detect copy-number changes, the inclusion of MLPA in clinical test algorithms has increased the detection rate of many genetic disorders. One reason why MLPA has rapidly gained acceptance as the method of choice for the detection of small genomic anomalies is that the method is simple to perform and that all equipment required is already present in most DNA diagnostic laboratories: a thermocycler for performing the MLPA reaction and a capillary sequencer for the detection of MLPA amplification products.

Since MLPA is able to distinguish sequences differing in a single nucleotide, it is widely used for the analysis of complicated genomic areas such as the *CYP21A2*/*CAH* region and the *SMN1*/*SMN2* genes involved in spinal muscular atrophy (SMA), where pseudogenes and gene conversion complicate DNA analysis (Arkblad *et al.*, 2006; Scarciolla *et al.*, 2006; Huang *et al.*, 2007). The relatively low cost and high throughput of MLPA enables large-scale carrier screening for diseases like SMA, in which heterozygous healthy carriers have a relatively high frequency in the general population (Scarciolla *et al.*, 2006). As the multiplex capacity facilitates the detection of copy-number changes of all well-described microdeletion syndromes (Kirchhoff *et al.*, 2007; Ahn *et al.*, 2008; Peppink *et al.*, 2008), all subtelomeric regions (Monfort *et al.*, 2006; Rooms *et al.*, 2006; Bruno *et al.*, 2006; Stegmann *et al.*, 2007; Ahn *et al.*, 2007), or all centromeric regions in

a single reaction, MLPA is now also routinely used in molecular cytogenetics where it has often replaced FISH. For almost all subtelomeric and microdeletion syndrome regions, high-density MLPA probe sets are available that can also reveal many smaller, atypical, deletions (Sutton *et al.*, 2004; Warshawsky *et al.*, 2006; Jalali *et al.*, 2007; van Hagen *et al.*, 2007). Applications of MLPA in other fields including tumor diagnostics (Jeuken *et al.*, 2006; Moerland *et al.*, 2006; Holtkamp *et al.*, 2008; Damato *et al.*, 2008; Bremmer *et al.*, 2008), pharmacogenetics and the detection and subtyping of microorganisms (Bergval *et al.*, 2008; Reijans *et al.*, 2008) are emerging.

As compared to FISH, MLPA has the advantage of being a multiplex technique in which very small (50–70 nt) sequences are detected. Most aberrations disrupting a single gene are too small to be detected by FISH. Moreover, MLPA can be used on purified DNA. Compared to array-based comparative genomic hybridization (CGH; see also Chapters 10 and 12), MLPA is a low cost and technically uncomplicated technique. Although MLPA is not suitable for genome-wide research screening, it is a good alternative to array-based techniques for many routine diagnostic applications. The approximately 300 probe sets now commercially available are dedicated to applications ranging from the relatively common (Duchenne, DiGeorge syndrome, SMA) to the very rare (hereditary pancreatitis, antithrombin deficiency, Birt-Hogg-Dube syndrome).

For routine applications in which traditionally the copy number of only a single gene is measured, such as Her2-neu (*ERBB2*) testing in breast tumors, MLPA has advantages over tests like FISH and real-time PCR (see also Chapter 7), as it can provide, in a single reaction, more information on both the extent of the amplified region(s) and on other rearrangements which may have implications for treatment, such as amplification of the *EGFR* gene or deletion of the *BRCA1* or *BRCA2* genes in the tumor cells (Moerland *et al.*, 2006; Holtkamp *et al.*, 2008; Damato *et al.*, 2008). Although the limited dynamic range of MLPA places some restrictions on its usefulness in detecting very large changes in copy number (>ten-fold), such as may be found in mRNA profiling or the detection of pathogenic microorganisms, MLPA has been described as being cheaper and far more accurate and precise than real-time PCR for applications involving the detection of small copy-number changes of multiple DNA sequences (Damgaard *et al.*, 2005).

Methylation-specific MLPA (MS-MLPA), which combines MLPA with the use of methylation-sensitive restriction endonucleases (Nygren *et al.*, 2005), has proven to be a very useful method for the detection of aberrant methylation patterns in imprinted regions that cause diseases like Prader-Willi/Angelman syndrome (Bittel *et al.*, 2007; Procter *et al.*, 2006; Dikow *et al.*, 2007) and Beckwith-Wiedemann syndrome (Eggermann *et al.*, 2007; Scott *et al*., 2008; Priolo *et al.*, 2008). The MS-MLPA method can also be used for the analysis of aberrant methylation of CpG islands in tumor samples (Jeuken *et al.*, 2007; Hess *et al.*, 2008), using, e.g., DNA derived from formalin-fixed, paraffin-embedded tissues.

This chapter reviews the use of MLPA in the quantification of genomic copy-number changes, detection of specific point mutations, mRNA profiling, and the detection of methylation changes. Although the emphasis is on the detection of MLPA amplification products by capillary electrophoresis, alternative detection methods such as bead and array hybridization are also discussed. Procedures for the design and use of ''home-made'' synthetic MLPA probes for temporary research applications are described.

13.2 PRINCIPLE OF MLPA

In MLPA, a multiplex oligonucleotide ligation assay is combined with a PCR reaction in which all ligation products are amplified with the use of a single PCR primer pair. Each of the up to 50 different ligation products has a unique length and all PCR amplification products of a single MLPA reaction can be separated and quantified by capillary electrophoresis. In contrast to most PCR-based techniques, it is not the actual sample nucleic acids that are amplified in the MLPA PCR reaction, but the probes that are added to the sample. First, sample DNA is denatured and incubated overnight with a mixture of probes. Each probe consists of two oligonucleotides which hybridize to immediately adjacent target sequences (Fig. 13.1, step 1). One of these two oligonucleotides contains the forward PCR primer sequence, the other the reverse PCR primer sequence. Only when both oligonucleotides are hybridized to the template DNA can they be ligated (Fig. 13.1, step 2), permitting subsequent exponential PCR amplification (Fig. 13.1, step 3). Generation of a probe amplification product is thus dependent on the presence in the sample DNA of a small (50–80 nt) sequence that is detected by that particular probe. The large excess of probe oligonucleotides that are not ligated do not have to be removed, as they contain only one of the two PCR primer sequences and can therefore not be amplified exponentially. Furthermore, it is this prerequisite of a ligation reaction in MLPA which provides the possibility of discriminating a single nucleotide difference, since most types of mismatch at the ligation site will effectively prevent ligation.

Most commercially available MLPA probe sets result in fluorescent amplification products with a length between 90 and 500 nt. The output of the capillary electrophoresis (CE) quantification, a file with peak lengths, peak heights, and peak areas (see also Chapter 5), can be analyzed by any spreadsheet software. The peak pattern obtained in MLPA reactions is very reproducible and differences in relative peak height between samples reflect differences in copy number of the probe target sequence (with some exceptions, see section 13.4). The presence of an average of 0, 1, 2, 3, or 4 copies of

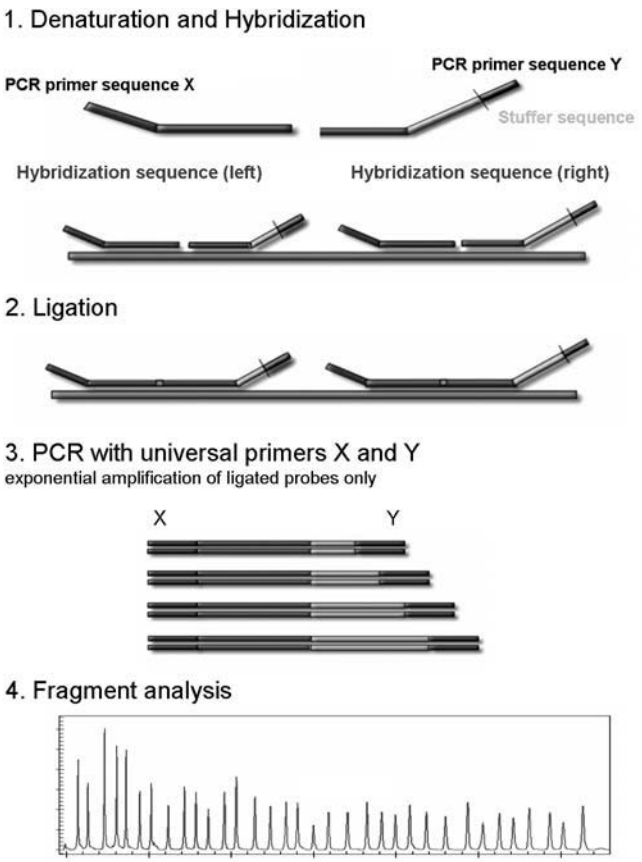


FIGURE 13. 1 Outline of the MLPA technique. 1. The sample DNA is denatured and incubated with a mixture of MLPA probes. Each probe consists of two oligonucleotides: the left probe oligonucleotide (LPO) and the 5′ phosphorylated right probe oligonucleotide (RPO), which hybridize to directly adjacent target sequences. 2. LPOs and RPOs hybridized to their adjacent targets are ligated. 3. Each probe has a ligation product of a unique length, which is amplified exponentially by PCR, using a single primer pair which recognizes the PCR primer sequences included in the LPO and RPO. 4. The resulting peak pattern of the sample is analyzed by comparing it to that of the reference samples.

a given sequence per cell can be readily distinguished. Since only relative copy numbers are determined, it is not possible to distinguish female diploid from triploid or tetraploid cells. Most aberrant copy-number changes causing disease are, however, readily detected by comparing the peak profiles of a patient sample with that of a reference sample (Fig. 13.2).

What makes the MLPA reaction more robust than many other PCR-based multiplex assays? First of all, as previously mentioned, all amplification products are generated with a single PCR primer pair. Second, the complete protocol is simple and the number of steps is limited (Fig. 13.3). Third, each of the three separate steps of the MLPA reaction – probe hybridization, ligation of probe oligonucleotides, and PCR amplification of ligation products – goes to completion. This fact means that small changes in reaction times, reaction temperatures, DNA concentration, and several other parameters have very little influence on the quantitative aspect of the MLPA reaction.

During the MLPA hybridization reaction, approximately one femtomol (600,000,000 copies) of each probe oligonucleotide is present. This is a large excess compared to the amount of sample DNA, since a human DNA sample of, e.g., 60 ng, contains only 20,000 haploid genomes. After incubation for 16 h in a small volume (8 μl) at 60°C, nearly all probe target sequences are covered with probe oligonucleotides. Small changes in volume (6–10 μl), temperature (55–62°C) or duration of the incubation (12–24 h) therefore have very little effect.

For the subsequent ligation step, the 8 μl reaction is diluted to 40 μl by adding a mixture of Mg^{++} ions, the ligase enzyme, and the ligase cofactor NAD, thereby reducing the salt concentration. The ligation reaction is typically performed for 15 minutes. However, the exact duration of the ligation reaction has little influence: for properly designed probes, no change in results can be observed between reactions that are ligated for 5 or 30 minutes, because ligation is already >90% complete within 2 minutes. The ligation is performed at a slightly lower temperature than the hybridization reaction (54°C instead of 60°C), as the effective melting temperature (T_m) of the probe-sample hybrids is lower at the reduced salt concentration present during the ligation reaction. The ligase enzyme used, Ligase-65, is heat inactivated after the ligation reaction. Because the non-ligated probe oligonucleotides do not have to be removed and since the ionic conditions during the ligation reaction resemble those of an ordinary 1 × PCR buffer, the PCR reaction can be initiated immediately by adding the necessary PCR primers, polymerase, and dNTPs. Usually, however, only part of the ligation reaction is used for the subsequent PCR, as this slightly improves the results for some probe mixes.

The PCR reaction in MLPA is more robust than other multiplex PCR reactions, primarily because a single PCR primer pair is used for amplification of all ligated probes. Differences in primer annealing efficiency play no role in MLPA. Another advantage is that the PCR conditions for all MLPA applications are identical. Finally, the amount of each individual amplification product at the end of the PCR is limited, making MLPA less sensitive to differences in PCR conditions. In almost all other PCR types (real-time, multiplex and conventional PCR), only a small amount of the PCR primers is used. PCR amplification comes to a halt as a result of rapid reannealing of complementary amplicon strands, inhibiting further primer binding and elongation. For a standard multiplex PCR, this means that the relative amounts of the amplification products can still change in the last cycles, as different amplicons may reach this point at difference times. Small changes in temperature, PCR conditions, number of cycles or the amount of sample DNA can thus easily affect the final peak pattern. In contrast, in MLPA, it is primer depletion which (after ~30 cycles) results in the termination of the PCR reaction. The majority of the PCR primers are consumed before

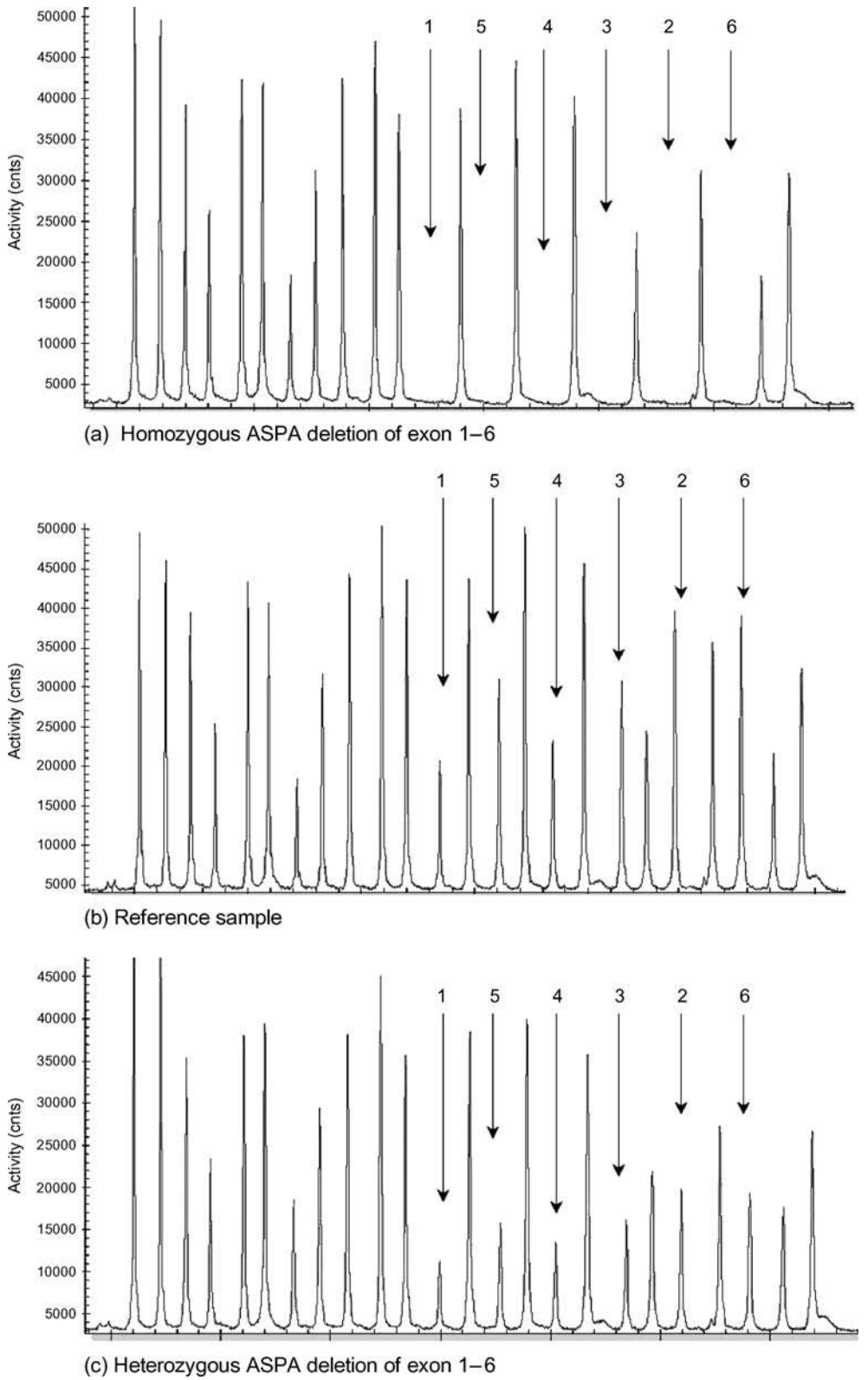


FIGURE 13.2 Detection of copy-number changes by MLPA. MLPA easily allows for the distinction between normal reference samples (b) and samples carrying (a) homozygous or (c) heterozygous deletions. In case of a homozygous loss of the complete *ASPA* gene (a), the six MLPA probes for the *ASPA* gene (arrows) do not generate any signal as there are no target sequences present in the sample. The other probes are on other chromosomes and remain unchanged. Comparison of a heterozygous carrier of a complete *ASPA* gene deletion (c) with a reference sample (b) shows an approximately 50% reduced signal for each of the six *ASPA* probes.

1. DNA denaturation: heat 5 minutes at 98°C
2. Hybridization: add SALSA probemix and MLPA buffer. Incubate 1 minute at 95°C, hybridize for 16 h at 60°C
3. Ligation: add ligase mix and incubate 15 minutes at 54°C Heat inactivate the ligase for 5 minutes at 98°C
4. Add primers, dNTPs and polymerase and start PCR
5. Capillary electrophoresis: export fragment lengths and peak areas to analysis software or spreadsheet. Analyze results

FIGURE 13.3 The MLPA protocol in a nutshell.

any probe amplification product reaches a concentration at which reannealing of complementary strands influences primer binding and elongation. Although the absolute amount of each amplification product increases slightly in the last PCR cycles, the relative amount of each amplification product (i.e. the peak pattern observed) remains unchanged between 30 and 40 PCR cycles. As a consequence, peak patterns obtained on small (20 ng) or large (400 ng) amounts of pure sample DNA are also almost identical and there is thus no need to use equal amounts of DNA in each reaction.

13.3 DETECTION METHODS FOR MLPA AMPLIFICATION PRODUCTS

13.3.1 Capillary Electrophoresis

By far the most common method for analysis of MLPA amplification products is size separation of the different probe amplification products using capillary electrophoresis. In the PCR reaction, one of the two primers is fluorescently labeled and the MLPA PCR products only have to be denatured before being injected into the capillaries. Advantages of capillary electrophoresis are its ease of use and the direct detection and quantification of all amplification products. Main disadvantages are the high cost of the instrument and the limited number of amplification products that can be separated. For proper quantification, probe amplification products should have a length difference of at least 3 nt and preferably 4–6 nt between consecutive fragments.

Chemically synthesized oligonucleotides longer than approximately 60–80 nt are often of low quality. Their use as MLPA probes can result in lower peaks and/or the generation of by-products, i.e. peaks that are one or more nt shorter than the main peak. The reliable amplicon range for chemically synthesized MLPA probe sets is therefore usually between 90 and 150 nt. Using differences of 4 nt between probes, the maximum number of MLPA probes that can be included in a single reaction is limited to approximately 15 synthetic probes. In order to increase the quality of the probes and the number of probes per reaction, the most commonly used MLPA probe sets (produced by the Dutch company MRC-Holland; http://www.mlpa.com) consist of one short (40–55 nt) synthetic oligonucleotide, and one long (70–450 nt) oligonucleotide derived from the single-stranded DNA of an M13 phage-derived clone. For each probe, a separate clone is prepared. This way, good quality probe oligonucleotides with a length of hundreds of nucleotides can be produced, allowing the use of up to 50 MLPA probes in a single reaction with length differences of 5–9 nt. Unfortunately, preparation of these long M13-derived probes is time consuming and expensive, and hence less suitable for more experimental research applications.

13.3.2 Microfluidics

Fragment separation by microfluidics (e.g. Agilent 2100) has not often been described in combination with MLPA, but its rapid progress in resolution opens many possibilities for its use in combination with MLPA. Major advantages of microfluidics include the use of disposable devices and more compact, low cost instruments. In addition, microfluidics avoids the electrokinetic injection procedure of capillary electrophoresis instruments, which results in preferential injection of smaller PCR amplification products. For laboratories that do not perform DNA sequencing, the use of a microfluidics apparatus would therefore provide an excellent alternative for the expensive capillary electrophoresis equipment.

13.3.3 Bead-Arrays and Micro-Arrays

Several articles have described the detection of MLPA amplification products by hybridization to oligonucleotides which are immobilized onto arrays (Berry *et al.*, 2007; Zeng *et al.*, 2008) or on Luminex beads (Lundquist *et al.*, 2005). In contrast to electrophoresis-based detection, such platforms permit the use of synthetic probes with amplification products of identical lengths and could thus considerably increase the number of different probes per reaction. The incorporation of sequence tags in the probes is particularly interesting, as it allows the use of a universal set of oligonucleotides coupled to the beads or arrays (Zeng *et al.*, 2008).

Although the use of array or bead detection is an attractive alternative to electrophoresis, no commercial MLPA tests are yet available for these detection platforms. This is partly due to the higher cost of arrays and beads as well as the lower number of labs that have access to such instruments. Another reason may be the higher standard deviation of individual probes, this is due to the fact that the detection of the probe amplification products requires an extra hybridization step compared to the direct detection by capillary electrophoresis. Furthermore, the reannealing

of the two complementary strands of each probe amplicon will compete with the binding of the labeled strand to the beads or arrays, possibly resulting in a higher variability. The use of exonuclease digestion which exclusively removes the unlabeled PCR strands might solve this, but has not yet been described in conjunction with MLPA.

Increasing the number of probes per reaction is also likely to increase the undesirable occurrence of aspecific amplification products due to interprobe interactions. As said, in order to reach almost complete hybridization of probes to their target sequence, an MLPA reaction contains approximately 1 femtomol, or 6×10^8 copies, of each probe oligonucleotide. Since each probe oligonucleotide already contains a perfect copy of one of the two PCR primer sequences, minor sequence homology between different probes (8–10 nucleotides) can easily result in the formation of aspecific products that are exponentially amplified. Although this occurs only sporadically when less than 50 different probes are present, the number of potential probe interactions increases exponentially with an increasing number of probes. Finally, the linear amplification of one of the probe oligonucleotides which always occurs in an MLPA reaction can also start causing problems with increased numbers of probes, due to its consumption of the unlabeled PCR primer. Since each right probe oligonucleotide (RPO, Fig. 13.1) contains the perfect complement of the reverse (unlabeled) PCR primer, these oligonucleotides will be copied (linear amplification) in each PCR cycle, thereby consuming around 1 fMol of the unlabeled primer per cycle for each of the probes present. The presence of a large number of probes thus results in a rapid depletion of the reverse PCR primer pool, which will interfere with the exponential amplification of ligated probes.

An attractive alternative for applications requiring a large number of probes could be the combination of array-based detection with the closely related multiple amplifiable probe hybridization (MAPH) technique (section 13.8; Armour *et al.*, 2000), in which the non-hybridized probe oligonucleotides are removed from the hybridized probes (Gibbons *et al.*, 2006; Kousoulidou *et al.*, 2008). This circumvents several of the technical bottlenecks mentioned above, but the MAPH technique requires the complicated extra steps of immobilizing sample DNA and the removal of non-hybridized probes before the PCR reaction.

13.4 ANALYSIS OF MLPA RESULTS

The analysis of MLPA results is facilitated by the fact that it is not the sample nucleic acids that are amplified and quantified in an MLPA reaction, but the probes that are added to the samples. An advantage of this approach is that the amplification products always have the same length, regardless of the existence of sequence variants in the probe binding region in the sample DNA. After the different probe amplification products of a single MLPA reaction have been identified and their distinct lengths entered in fragment analysis software (''binning''), all subsequent analysis steps can be automated. Different Excel sheets and dedicated software packages for analysis of MLPA results are available free of charge or at relatively low cost (see www.mlpa.com).

It is not possible to draw any conclusions on the basis of absolute peak areas or heights obtained in an MLPA experiment, as these values depend on many factors, including the number of probes in the MLPA probe mix, bleaching of the fluorescent label, alignment of the laser on each capillary, and the efficiency of the electrokinetic injection in each capillary. The first step in the analysis of MLPA results is therefore always an intra-sample normalization, resulting in relative peak areas or peak heights. The majority of users analyze peak areas, but in some cases, e.g. when the fragment analysis software erroneously separates peaks into a main peak and a shoulder peak, the use of peak heights can have advantages.

Different MLPA applications require slightly different analysis methods. For most applications, intra-sample normalization is performed by dividing the peak area (or height) of each amplicon by either (a) the combined area of only the reference probes (block normalization) or (b) the combined probe area of all peaks in that sample (population normalization). Block normalization is used for instance in MLPA probe sets which target only one or a small number of genes, in which a complete gene deletion (affecting numerous probes) could easily skew up the analysis if all probes were used for normalization. Since reference probes are selected for their location in various genomic regions that are expected to be stable in the application studied, they form a genomic yardstick against which changes of the application-specific genes can be easily detected. In contrast, population analysis is the method of choice for probe sets detecting a wide range of genomic regions. The use of reference probes is often not necessary for these applications. A good example is the MLPA subtelomeric screening for mental retardation. Since these subtelomeric probe sets contain one probe for each of the 46 chromosome arms, the one or two probes that may be aberrant in a single sample will be easily detected when using population normalization. The analysis of MLPA results in tumor diagnostics, where chromosomal aberrations are widespread and unpredictable, is more complicated. The genetic instability of tumor cells means that reference probes, despite their careful selection, can also be subject to copy-number changes. For this reason, the analysis of tumor samples requires more complicated analysis programs (Coffa *et al.*, 2008) which can exclude outlier reference probes from the normalization process by first comparing each target-specific probe to each reference probe separately.

After intra-sample normalization of probe signals, the results obtained on a sample can be compared to the results obtained on one or (preferably) more reference samples.

In most cases, probe ratios below 0.7 or above 1.3 are regarded as indicative of a heterozygous deletion (copy-number change from two to one) or duplication (copy-number change from two to three), respectively. For a correct interpretation of results, probes should be arranged according to chromosomal location as this may reveal more subtle changes, such as those observed in mosaicism. For instance, the SALSA MLPA P095 probemix (MRC-Holland) that is routinely used for the detection of trisomy 13, 18, or 21 in amniotic fluid samples was able to detect a mosaic sample with only 26% trisomic cells for chromosome 18, since a ratio of 1.1 was found for all eight probes present for that chromosome (van Opstal *et al.*, 2008).

It should be kept in mind that next to a true copy-number change, a reduced probe signal can also be due to a mismatch between the target sequence and the probe, for instance due to a (single nucleotide) polymorphism. Sequence changes immediately adjacent to the ligation site can influence probe signals by preventing ligation of the two probe oligonucleotides. Furthermore, sequence alterations at longer distance, even at 15 nt from the ligation site, can result in a reduced probe signal when the mismatch between sample DNA and probe oligonucleotide destabilizes the binding of the probe. Copy-number changes detected by MLPA therefore always require confirmation by other methods, especially when a single probe is affected. Often, DNA sequencing is used to determine whether a mutation or polymorphism is present in the probe's recognition site. Techniques like (long-range) PCR and qPCR are often used to confirm single exon deletions detected by MLPA (Vaughn *et al.*, 2008).

13.5 METHYLATION QUANTIFICATION BY MS-MLPA

A simple modification of the original MLPA technique, methylation-specific MLPA (MS-MLPA, Nygren *et al.*, 2005) is now widely used for the detection of changes in cytosine methylation in specific genomic areas. A major application of MS-MLPA is the detection of aberrant methylation of the imprinting regions involved in Prader-Willi/Angelman syndrome (Bittel *et al.*, 2005; Dikow *et al.*, 2006; Procter *et al.*, 2006) and Beckwith-Wiedemann/RSS syndrome (Eggermann *et al.*, 2007; Scott *et al.*, 2007; Priolo *et al.*, 2008). A second major application lies in the detection of methylation changes of CpG islands that are located near the promoter regions of tumor suppressor genes in DNA derived from tumor cells (Jeuken *et al.*, 2007; Hess *et al.*, 2008).

MS-MLPA probes are identical to standard MLPA probes except for the presence of a recognition site of a restriction endonuclease, such as *HpaII* or *HhaI*. These restriction enzymes are sensitive to cytosine methylation of a CpG dinucleotide in their recognition site: they can only digest double-stranded DNA in which their recognition sequence is unmethylated in both strands. In MS-MLPA, the ligation of MLPA probe oligonucleotides that are hybridized to the sample DNA is combined with digestion by a methylation-sensitive endonuclease. In case the CpG site of the sample DNA is methylated, digestion is effectively prevented and hence a normal MLPA probe signal will be detected. If the CpG site is not methylated, however, the DNA–probe complex will be digested by the endonuclease and no amplification product is formed.

During the development of MS-MLPA, two different approaches were tested. In the first, endonuclease digestion of sample DNA was followed by a normal MLPA reaction. In the second, the ligation of the probes and the digestion of the sample DNA–probe hybrid were performed simultaneously. As the latter method produced good results and was easier to perform, it became the method of choice (Fig. 13.4). By dividing the reaction in two parts after probe hybridization, one part being ligated and digested simultaneously, while the other part is ligated only, just like a normal MLPA, a single MS-MLPA reaction is able to provide information on both copy number and methylation changes. In addition, a very important advantage of digesting the sample DNA–probe hybrids rather than sample DNA itself is that this allows the use of DNA from formaldehyde-fixed, paraffin-embedded (FFPE) tissues. DNA extracted from FFPE tissues is often not completely digested by restriction endonucleases, presumably because the genomic DNA is (partially) denatured during paraffin embedding. Since the probe–DNA hybridization makes the DNA double stranded, this problem is evaded by MS-MLPA.

An important advantage of MS-MLPA as compared to other techniques like methylation-specific PCR is that the difficult-to-standardize bisulfite conversion of sample DNA is circumvented. In addition, both copy numbers and methylation are quantified in a multiplex fashion, as in MS-MLPA too it is possible to quantify up to 50 different sequences per reaction. A disadvantage is the lower sensitivity compared to some other techniques. For example, when a CpG site is methylated in one allele and in only 5–10% of the sample cells, detection by MS-MLPA is difficult due to the low probe signals. A second disadvantage is that probe design requires the presence of a recognition site of a suitable restriction endonuclease (e.g. *HhaI* or *HpaII*) in the probe hybridization sequence. Finally, single-stranded genomic DNA from CpG islands usually has a strong secondary structure, resulting in slower hybridization of probes. As a consequence, a larger percentage of probes fail or larger quantities of probe oligonucleotides have to be added as compared to ordinary MLPA, in which probe sequences typically contain only 40–60% GC nucleotides.

Figure 13.5 shows the use of MS-MLPA in detecting copy-number and methylation changes of the 15q11 region, both of which can result in either Prader-Willi

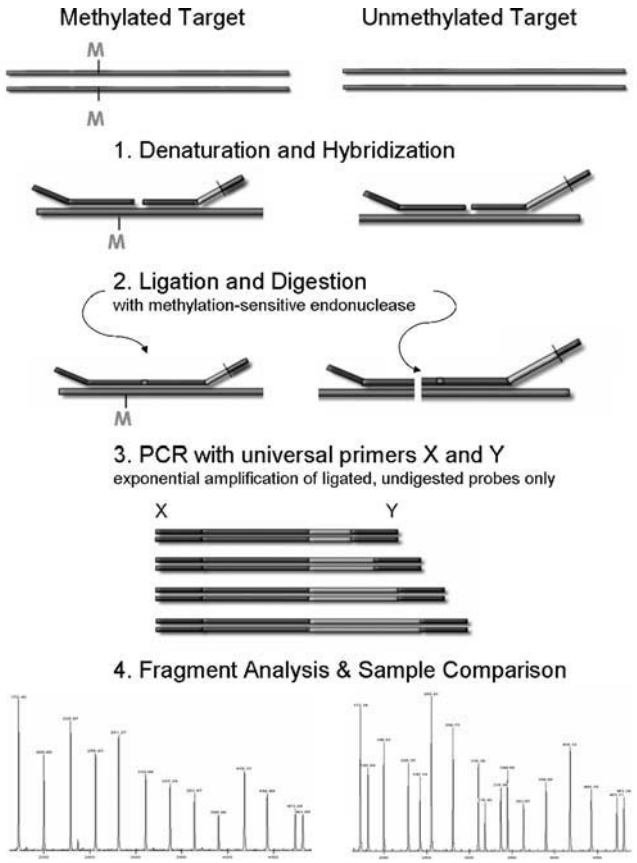


FIGURE 13.4 Outline of the MS-MLPA technique. 1. Sample DNA is denatured. Each MLPA probe consists of two oligonucleotides, hybridizing to directly adjacent target sequences, which can be either methylated or unmethylated. MS-MLPA probes contain a recognition site of a methylation-sensitive restriction endonuclease (e.g. *HhaI* or *HpaII*); reference probes do not. 2. After hybridization, the reaction is divided in two parts. In one tube, LPOs and RPOs that are hybridized to their adjacent targets are ligated. In the other (shown here), ligation and endonuclease digestion take place simultaneously; unmethylated sample DNA–probe hybrids are digested. 3. Each non-digested probe has a ligation product of a unique length, which is amplified exponentially by PCR using a single primer pair recognizing the primer sequences included in each LPO and RPO. 4. For each sample, two peak patterns are produced: one from the ligation-only reaction (for determining copy-number changes (not shown)), and one from the ligation and digestion reaction (for methylation profiling (shown)). The peak patterns are analyzed by comparing them to those obtained on reference samples. Left peak pattern: Ligation and digestion reaction of a healthy control showing only the 13 reference probes not containing a *HhaI* restriction site. The other 26 probes included in this mix all detect a sequence containing a *HhaI* restriction site which is unmethylated in normal blood-derived DNA. The probe–sample hybrids are thus digested and hence the MS-MLPA probes are not visible in this sample. Right peak pattern: Ligation and digestion reaction of a tumor sample. The extra four peaks visible here are due to the fact that some of the sequences detected by the MS-MLPA were methylated in this tumor-derived DNA, saving the probe–DNA hybrid from endonuclease digestion.

(PWS) or Angelman syndrome (AS). This 4 Mb region is subject to imprinting: normal individuals inherit one methylated copy from their mother and one unmethylated copy from their father. Inheritance of two chromosome 15 copies from a single parent (uniparental disomy) results in either Prader-Willi syndrome (two maternal, methylated copies) or Angelman syndrome (two paternal, unmethylated copies). Deletions, too, can be pathogenic: a loss of the paternal copy causes PWS, whereas losing the maternal allele results in AS. Chromosomal deletions can be easily identified by comparing the 15q11 probes with reference probes between different samples (Fig. 13.5: top row), while methylation changes can be detected by comparing the peak profiles of different samples after *HhaI* digestion (Fig. 13.5: bottom row). For instance, when examining the MS-MLPA probe for the *SNRPN* gene (peak 5 in all undigested samples), this peak is halved after digestion with the *HhaI* enzyme in a healthy reference sample. Conversely, the probe signal is not reduced by *HhaI* digestion in PWS patients (both PWS deletion and PWS disomy): the peak signal is the same in undigested and digested samples because all copies present are methylated. In contrast, since in AS all copies (both in the case of AS disomy and AS deletion) are unmethylated and hence digested, no signal is generated by this probe when testing Angelman patients. The complete ME028 probe set for PWS/AS contains more than 25 probes in this 15q11 region (several of which indicating methylation changes), making it also suitable for the detection of atypical smaller deletions. In addition, the set contains several digestion control probes (confirming complete endonuclease digestion) and several reference probes on other chromosomes.

13.6 RT-MLPA FOR mRNA PROFILING

Reverse transcriptase MLPA (RT-MLPA) (Eldering *et al.*, 2003) is another adaptation of the MLPA technique which can be used for mRNA profiling as an alternative to real-time PCR and micro-arrays. There are two main differences between the MLPA protocol for RNA and the original MLPA method developed for DNA. First, RT-MLPA starts with the conversion of mRNA into cDNA. To this end, a reverse transcriptase enzyme and a special RT primer mix are added to the RNA sample. This primer mix contains one RT primer for each of the MLPA probes. The RT primers are located just downstream from the probe's target sequences. The probes that are used in the subsequent MLPA reaction are thus complementary to the cDNA, not to the mRNA. The conversion into cDNA is essential as NAD-dependent ligases, such as the Ligase-65 enzyme used in MLPA, cannot ligate DNA probe oligonucleotides that are hybridized to RNA.

The second difference is that, whenever possible, RT-MLPA probes are designed with an exon boundary within their target sequence. For instance, one part of an RT-MLPA probe may hybridize to the last 25 nucleotides of exon 1, while the other part hybridizes to the first 35 nucleotides of exon 2. Such an ''intron spanning'' design prevents the probe from generating a signal on contaminating (genomic) DNA that is often present in RNA samples.

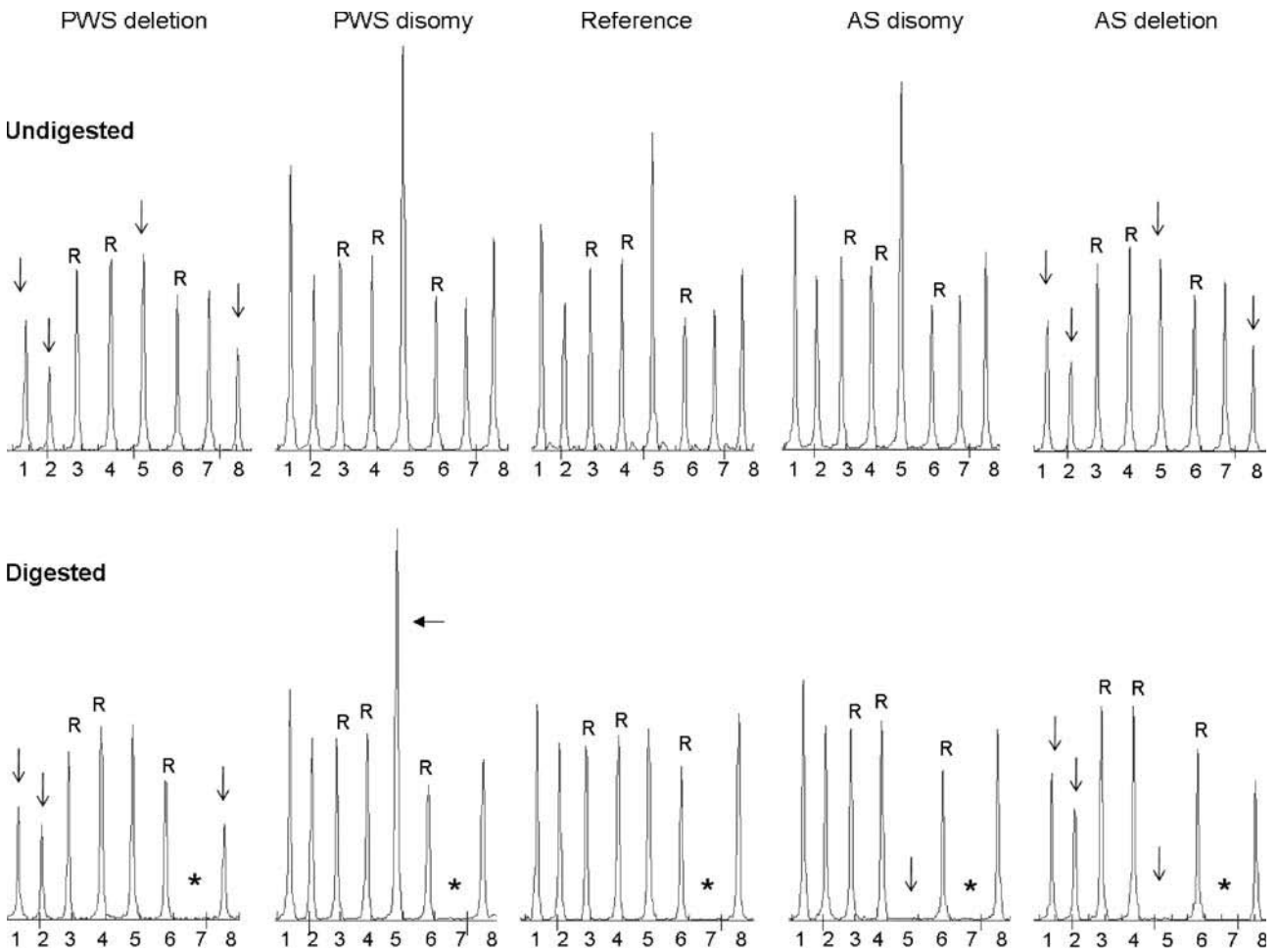


FIGURE 13.5 Copy-number and methylation changes in 15q11 detected by MS-MLPA in Prader-Willi and Angelman syndrome patients. Partial peak patterns (eight out of 48 probes shown) of the ME028 probemix (MRC-Holland) used on one reference and various patient samples. **Top:** Detection of copy-number changes (ligation only). **Bottom:** Detection of methylation changes (ligation and digestion with *HhaI* restriction endonuclease). **R:** Reference probe on other chromosome not affected by *HhaI* digestion (probes 3, 4, 6). **Probe 1:** Detects *SNRPN* gene, no *HhaI* site; **2:** Detects *UBE3A* gene, no *HhaI* site; **5:** Detects *SNRPN* gene, contains *HhaI* site; **7:** Detects 15q26, contains *HhaI* site, sequence completely unmethylated in normal blood-derived DNA; **8:** Detects *UBE3A* gene, no *HhaI* site. **Arrow:** Probes showing aberrations as compared to the reference sample *SNRPN* and *UBE3A* genes are located in the 15q11 PWS/AS region. Probes 5 and 7 are MS-MLPA probes containing an *HhaI* site: when the target sequence is unmethylated and *HhaI* digestion is performed (bottom row), they are digested: no probe signal. Probe 5 detects methylation changes of a sequence on the 15q11 region; a normal reference thus has one copy remaining upon digestion. Probe 7 is a digestion control detecting a 15q26sequence that is unmethylated in blood-derived DNA; this probe controls for complete endonuclease digestion.

When comparing RT-MLPA to mRNA profiling by micro-arrays, RT-MLPA assays are more labor intensive to develop, have a much lower multiplexing capacity, and a smaller dynamic range. Advantages of RT-MLPA over micro-arrays are its ease of use, the much lower costs per sample and the lower standard deviation of individual probes. Furthermore, in RT-MLPA there is no need to perform a sample labeling reaction; it offers higher throughput capabilities and the ability to distinguish sequences differing in a single nucleotide. In a clinical setting, RT-MLPA has been demonstrated to be a good alternative to micro-arrays, as it can quantify up to 50 different transcripts in a single-tube assay, while showing excellent conformity with real-time PCR and micro-array results (Hess *et al.*, 2004, 2007; Lok *et al.*, 2006; Cillessen *et al.*, 2007).

The variability in RNA expression levels combined with the rather low dynamic range of RT-MLPA means that a probe set which performs well on RNA samples derived from a given tissue may require adjustments when used on another tissue. The level of RNA expression can vary so greatly that peak signals of highly expressed genes may get off scale. The inclusion of so-called competitor oligonucleotides in the probe mix allows for tissue-specific fine-tuning of RT-MLPA assays. These competitors compete with a given probe because they are designed to recognize the same target sequence as one of the two probe oligonucleotides. Although the competitor will bind to the target sequence and may even be ligated to the second probe oligonucleotide, the resulting product will never be amplified because the competitor, unlike the real probe, lacks

the complete PCR primer sequence. As a consequence, the competition between competitor and probe for the limited target sites reduces the signal of the probe's peak signal. The simple inclusion of one or more of these competitors allows ''dampening'' of specific peak signals, thereby making it possible to quantify them again.

An important advantage of RT-MLPA is the possibility of using highly degraded RNA, such as RNA derived from FFPE tissues. At MRC-Holland, encouraging results have been obtained on RNA samples from slides that had been stored at room temperature for many years. Small stretches of cDNA are sufficient to generate RT-MLPA probe signals, since the probe-specific RT primers which are supplied with an RT-MLPA probe set can partly overlap with their corresponding probe and have to be elongated by only 50 nucleotides. Also, the use of one dedicated RT primer for each probe equalizes the efficiency of the mRNA-to-cDNA conversion, making it less dependent on the GC content and the degree of RNA breakdown as compared to the use of hexamers or oligo-dT.

13.7 DESIGN OF MLPA PROBE SETS

The majority of MLPA reactions performed worldwide use MLPA probe mixes and reagents developed by MRC-Holland (http://www.mlpa.com). Most MLPA probes manufactured by this company consist of one synthetic and one M13 phage-derived probe oligo. The advantage of M13-derived oligonucleotides over their synthetic counterparts is that the former can be much longer as their production process is more reliable. In every cycle of chemical oligonucleotide production, one new dNTP is added to the growing oligo. Approximately 1% of these additions fail, meaning the quality of synthetic oligos decreases considerably with increased length. By making one of the probe oligos phage M13 derived, it is possible to use a much larger range in probe length. A typical MRC-Holland probe set contains up to 50 probes with an amplicon range between 120 and 500 nucleotides.

Preparation of these M13-derived oligonucleotides is costly and time consuming and therefore less suitable for temporary research applications. For each probe, one M13 clone is prepared containing a target-specific hybridization sequence of 25–50 nt, a stuffer sequence of variable length and the reverse PCR primer sequence. Single-stranded DNA of each clone is produced and the probe part is subsequently cut out from the bulk of the M13 DNA by restriction endonuclease digestion. For research applications, many laboratories develop their own fully synthetic probe sets (Stern *et al.*, 2004; Harteveld *et al.*, 2005; Ehlert *et al.*, 2007) or add a small number of synthetic probes to an MRC-Holland probe mix (Scott *et al.*, 2008). Because of the low quality of long synthetic probes, the amplification products of synthetic probe sets generally range between 90 and 150 nucleotides, with a typical length difference of 3 or 4 nucleotides between probes. In order to increase the number of probes per MLPA reaction, the use of multiple probe sets, each amplified by primers labeled with a different fluorescent dye, has been described (White *et al.*, 2004; Harteveld *et al.*, 2005). However, since such probe sets are more difficult to develop, most laboratories use home-made probe sets consisting of less than 20 probes for their research applications. Guidelines for the design of synthetic MLPA probes are described in detail on http://www.mlpa.com. The most important rules are:

1. Each probe consists of two oligonucleotides: one containing the sequence of the fluorescent forward PCR primer at its 5′ end and the hybridizing sequence at its 3′ end (left probe oligonucleotide or LPO), the other having the hybridizing sequence at its (phosphorylated) 5′ end and the reverse-complement sequence of the unlabeled reverse PCR primer at its 3′ end (right probe oligonucleotide or RPO).
2. The hybridizing sequences of the LPO and RPO should be directly adjacent, allowing the two oligonucleotides to be ligated when both are hybridized to their target sequence.
3. Different probes should not overlap as they will compete for the same target sequence. In addition, non-specific amplification products can be generated even with relatively small homologous regions (ten nucleotides) if probes are not properly designed.
4. Both LPO and RPO should have a hybridizing sequence that allows a stable binding to the target sequence at the temperature and salt concentrations used during hybridization and ligation. In practice, a minimum melting temperature of 68 °C at 100 mM salt is used, as calculated using RAW software (available on http://www.mlpa.com). RAW uses an average of several published formulas to calculate the T_m. There is no upper limit for the T_m and it is thus possible to use long hybridizing sequences. The use of a non-hybridizing stuffer sequence, derived from, e.g. T7 or λ-phage sequences, may have advantages, however; for instance when the target sequence is very GC rich.
5. When electrophoresis is used for separation and quantification of probe amplification products, each probe should generate an amplification product of unique length.
6. The nucleotide at the 3′ end of the LPO is the most discriminative site for distinguishing related sequences such as pseudogenes. A mismatch between probe and sample DNA at this position is usually sufficient to prevent the generation of a probe signal. However, as thymine and guanine are also capable of forming hydrogen bonds, probes that only have a G/T mismatch when bound to the sample DNA will still generate a ligation product, even when this mismatch is located at the 3′ end of the LPO. In that case, a probe can usually be designed detecting the

sequence on the complementary strand. When properly designed, the probe signal on the sequence with a mismatch at the 3′ end of the LPO will be less than 2% of that of the sequence without mismatch under the conditions recommended by MRC-Holland.

7. The first nucleotide after the PCR primer sequence has a considerable influence on the height of the probe signal obtained. When the first nucleotide with which the forward (fluorescent) PCR primer is elongated is an adenine, a much lower signal is obtained as compared to a cytosine. The probe signal increases in the order A, T, G, C.
8. For MS-MLPA probes, the sequence detected by the probe should contain the recognition sequence of a methylation-sensitive restriction endonuclease such as *HhaI* or *HpaII*. This recognition sequence should not be located at the very start or end of the hybridizing part of the probe: on either side of the recognition sequence, at least 5 nt should be complementary to the sample DNA.

Various control fragments can be included in the MLPA probe set to check the quality of the reaction just performed. First, all MRC-Holland probe sets contain four small fragments (the so-called Q-fragments) whose signal is inversely related to the amount of sample DNA used. These control fragments are used as indicators of insufficient amounts of DNA. As these Q-fragments already contain both PCR primer sequences in a single DNA fragment, they do not require ligation and are thus also amplified in reactions in which no sample DNA was present. As little as approximately 1,000 molecules of each Q-fragment are present in an MLPA reaction. Peak signals of these Q-fragments are relatively high when very low amounts of sample DNA are used (less than 20 ng DNA, corresponding to 6,000 haploid genomes or 3,000 cells). In contrast, when working with larger amounts of DNA, the amplification of the more numerous target-specific, ligation-dependent MLPA probes will outcompete that of the Q-fragments and hence their peak signals will dwarf those of the latter. In the automatic analysis of MLPA results by the Coffalyser software, reactions with high Q-fragment signals are flagged as less reliable because of the low sample DNA input.

In addition, MRC-Holland probe mixes contain two small control fragments, termed D-fragments, as indicators for poor DNA denaturation, something which can also result in unreliable results. As these D-fragments are located in strong CpG islands, a low signal as compared to a third control probe, detecting a more AT-rich region, is an indication of incomplete DNA denaturation. Third, MRC-Holland probe mixes contain one chromosome X- and one chromosome Y-specific probe.

13.8 RELATED TECHNIQUES

Several techniques resembling MLPA have been described. The MAPH technique (Armour *et al.*, 2000; see also Chapter 12), developed prior to MLPA, is very similar to MLPA as it also ends with a multiplex PCR in which all probes are amplified with a single PCR primer pair. However, MAPH does not use the ligation reaction that allows MLPA to distinguish even closely related sequences. In MAPH, complete probes containing both PCR primer sequences are hybridized to the sample DNA. As there is a large excess of probes compared to the number of target sequences in the sample, all non-hybridized probe oligonucleotides have to be removed as they are inherently amplifiable. This is accomplished by using immobilized sample DNA. The extra steps involved in MAPH for immobilization of sample DNA and removal of non-hybridized probes have precluded its widespread use. For reasons mentioned in section 13.3.3, MAPH might, however, have advantages over MLPA when combined with probes sets containing a very large number of probes and an array-based detection of the amplification products (Kousoulidou *et al.*, 2008). Extension-MLPA, or e-MLPA, is a recently described variation of MLPA (Leonhard *et al.*, 2008) in which the two probe oligonucleotides are not binding to adjacent, but to nearby sequences of the sample DNA. One of the probe oligonucleotides is elongated by polymerase activity until it reaches the second probe oligonucleotide and can be ligated. Application of e-MLPA could include the detection of variable number of tandem repeats. Interpretation of e-MLPA results is more complicated than that of MLPA, as the length of amplification product of each probe is not fixed.

MLGA is a recently described method (Isaksson *et al.*, 2007) in which selectively circularized genomic fragments are amplified. In contrast to MLPA, the genomic DNA itself is amplified rather than the probe oligonucleotides. Small deletions or insertions (indels) will therefore result in differently sized amplification products, complicating analysis of results. Although MLGA reactions can be performed in a shorter time than MLPA, a larger number of different enzymes are used and the hands-on time is longer than for MLPA.

No commercial assays based on MAPH, e-MLPA, or MLGA are available at this moment.

13.9 PITFALLS OF MLPA REACTIONS

Variability of results with particular MLPA probes can have many different causes, including improper probe design, poor sample quality, insufficient amounts of sample DNA, differential polymerase sensitivity of probes, and secondary structure of the genomic DNA to which the probe binds.

13.9.1 Probe Design

Improper probe design can include the use of probe oligonucleotides that cannot form a (sufficiently) stable binding with their target, for instance due to a polymorphism within

the detected sequence or because of having an insufficiently long hybridizing sequence. When designing probes, it should be kept in mind that a polymorphism or mutation in the middle of the hybridizing sequence can have a stronger effect on probe stability and signal than one located just 3 or 4 nt from the ligation site. Moreover, sequence variations within two nucleotides of the probe ligation site can result in a reduced probe signal by hampering the ligation of the two probe oligonucleotides.

Once bound, the chance of a probe leaving its target during the remaining part of the 16 hours of hybridization reaction should be very low. In case the binding of one of the probe oligonucleotides to the DNA template is not sufficiently stable, an equilibrium will be reached between probe binding and probe "denaturation" which is extremely sensitive to incubation temperature and probe and salt concentrations. In that case, small differences in evaporation between different samples can also have a strong effect on the MLPA results.

When designing probes, it should be kept in mind that the secondary structure of the (single-stranded) sample DNA can affect the performance of certain probes. Dependent on the genomic DNA sequence, a strong secondary structure can be formed immediately after reducing the temperature following the DNA denaturation step, which may prevent probes from binding to their target. This secondary structure can depend on the degree of sample DNA fragmentation; results obtained with some probes might therefore depend on sample DNA integrity. Unfortunately, the secondary structure of sample DNA is very difficult to predict. The MLPA reaction starts with a 5 minute incubation at 98 °C which is intended to both denature and partially fragment the sample DNA. In MS-MLPA reactions, in which many probes are used that detect sequences in CpG islands, this 98 °C incubation is increased to 10 minutes in order to obtain even shorter DNA fragments.

13.9.2 Sample Quantity

Insufficient amounts of sample DNA result in an increased variability of probe signals. At present, a minimum of 20 ng human DNA (~3,000 cells) is required to obtain good results with MLPA. Using an optical density measurement to determine DNA concentration has its shortcomings, as the presence of RNA or other impurities can easily lead to overestimation of the sample DNA concentration. However, the use of insufficient amounts of sample DNA is easily detected by the aforementioned Q-control fragments included in each MRC-Holland probe set (1.7). Very high sample amounts (>500 ng) can also result in more variable results, perhaps due to increased viscosity leading to slower probe hybridization. For very small sample quantities, the combination of MLPA with whole genome amplification methods has been described (Chou *et al.*, 2008) but this is less reliable due to the uneven amplification of the genomic DNA (Pugh *et al.*, 2008). Recently, new PCR primers and PCR conditions have been developed at MRC-Holland that provide good MLPA results on samples containing only 5 ng of sample DNA without the introduction of extra steps in the MLPA protocol. This may allow the use of DNA extracted from a small punch of a Guthrie card and will increase the possibilities of using DNA purified from amniotic fluid samples for MLPA reactions.

13.9.3 Incomplete Sample DNA Denaturation

Long genomic DNA strands are more difficult to denature than shorter fragments. However, as the 5–10 minute 98 °C heat treatment which initiates the MLPA reaction both denatures and fragments the sample DNA, this usually does not pose many problems. Apart from affecting the PCR reaction, the presence of certain ions such as Mg^{++} or Fe^{++} in the sample can in some cases prevent complete DNA denaturation, in particular of regions with a high GC content, such as CpG islands. A concentration of as little as 0.1 mM $MgCl_2$ in the sample can already increase the denaturation temperature (T_m) by 12 °C (Eichhorn, 1962). Detecting incomplete sample DNA denaturation is important, as it can affect the signals of all MLPA probes for sequences within or nearby a CpG island. For DNA samples that have only been partially denatured, the possibility of probes binding to their target sequence and generating a probe signal will depend on the level of sample DNA degradation and the distance between the probe recognition sequence and the nearest CpG island. Incomplete DNA denaturation can lead to false positive results which – erroneously – seem reliable as more than one probe is affected. Fortunately, a poor denaturation of the sample DNA is usually detected when the aforementioned denaturation control fragments are present (section 13.7).

In this respect, it is important to recall that long genomic DNA fragments containing CpG islands are not denatured at 95 °C in 1x PCR buffer. Performing a conventional PCR on such incompletely denatured sample DNA will result in weak bands, as there is always a subset of degraded sample DNA that is denatured and can be amplified. Real-time PCR reactions in which the genomic sample DNA is denatured after mixing with PCR buffer can result in unreliable results for CG-rich regions, such as CpG islands. As denatured human genomic DNA requires many months to renature, it is wise to denature the sample DNA by heating to 90–95°C in TE, or by alkali treatment, before addition of PCR buffer and polymerase.

13.9.4 Sample Quality

As compared to many other techniques, MLPA is less influenced by DNA and mRNA fragmentation, with the exception of probes detecting sequences prone to secondary

structure formation. This is an advantage when degraded nucleic acids that have been derived from, e.g., paraffin embedded tissues, feces, apoptotic cells, or processed foods are analyzed (Ehlert *et al.*, 2007). MLPA is, however, more sensitive to the presence of certain impurities than conventional PCR. Ionic impurities, such as iron derived from blood cells, can have a particularly strong effect on the relative peak patterns obtained in the MLPA PCR reaction, as amplification of some probes is more sensitive to such impurities than others.

Using pure DNA samples, it can be shown that some MLPA probes are very sensitive to the amount of polymerase enzyme used in the PCR reaction. Approximately 5% of the probes have a much lower relative probe signal when two-fold higher or lower amounts of polymerase are used in the PCR reaction, as a result of certain sequences in the amplicon. The effect of particular impurities on MLPA results might be a result of faster inactivation of the polymerase in the PCR reaction. It is therefore important to test probes and select those that are not strongly influenced by the amount of polymerase activity in the PCR.

Many types of impurities do not affect MLPA results. Crude cell lysates, for instance from amniotic fluid cells, can be used with excellent results, provided that the cells are not contaminated by blood cells. Proteinase K digests from paraffin-embedded tissues also provide good results, especially when low molecular weight impurities are first removed from the tissue by extensive washing.

Differences in purity between samples can sometimes make it difficult to compare results on samples obtained from different laboratories or purified by different methods. In case of problems, an extra purification step usually helps. In general, it is recommended to use reference samples that have been purified by the same method and that are derived from the same tissue as the patient samples. Most published studies have used DNA that was purified from blood or from paraffin-embedded tissues (van Dijk *et al.*, 2005; Takata, 2008), although the use of DNA derived from, e.g., buccal smears (Peppink *et al.*, 2008) in MLPA reactions has also been described.

13.9.5 Interpretation Problems

As with all genetic testing methods, it should be kept in mind that the ''complete'' sequence of only a handful of individual human genomes is known at this moment and that gaps in the sequences still exist. Moreover, individual variation in copy number of certain genomic regions is larger than previously expected, and many deletions and duplications of genomic sequences, even up to several Mb in length, have been detected in healthy individuals. Copy-number variations (CNVs) and low-frequency single nucleotide polymorphisms (SNPs) are constantly being discovered. Not only does this complicate the design of MLPA probes, it also makes it difficult at times to draw a firm conclusion about the significance of certain MLPA results. For example, a complete deletion of exon 16 of the *DMD* gene has been reported in a healthy adult male (Schwartz *et al.*, 2007). As the gene sequence remained in-frame, this deletion was expected not to lead to Duchenne muscular dystrophy but to the milder Becker muscular dystrophy. Surprisingly, however, the deletion of the amino acids encoded by exon 16 in this long protein apparently had no phenotypic effect.

Since the link between genotype and phenotype is not always clear, it is often necessary also to analyze parental samples. A good example concerns the use of MLPA for finding the cause of mental retardation, in which all subtelomeres are screened using a single probe set containing one probe per chromosomal arm. Any unusual result found with these broad first line screening kits can be further investigated with specialized MLPA probe sets containing more probes per subtelomere, enabling users to confirm the results and to determine the extent of the deletion/duplication. When the same deletion or duplication is also found in a healthy parent, it is unlikely to be the actual cause of the mental retardation. In contrast, when the mutation is *de novo*, the observation that it is indeed causative is more plausible. It will require many years of diagnostic testing in various ethnic populations before the effect of each variation in the human genome is known or can be correctly predicted.

13.10 SUMMARY: ADVANTAGES AND LIMITATIONS OF MLPA

Major advantages of MLPA for copy-number detection are:

1. Specificity of probes is very high as demonstrated by the complete absence of probe signals of chromosome Y-specific probes on female DNA, and the finding of the expected ratio of 2:1 for X-specific probes between male and female samples. Well-designed probes have single nucleotide specificity.
2. The sequence detected by each MLPA probe is only 50–70 nt, allowing the detection of small deletions that cannot be detected by FISH or BAC-arrays.
3. MLPA can be performed on purified DNA from a variety of sources, including degraded DNA extracted from formaldehyde-fixed, paraffin-embedded tissues. No intact cells are required.
4. MLPA is a relatively low-cost and technically uncomplicated method that requires only limited amounts of sample. Detection of copy-number changes and known point mutations in a single probe set is possible.
5. Identical protocol and PCR conditions for hundreds of applications. Results are obtained within 24 hours and are easy to analyze, as the length of amplification product of each probe is fixed. The instruments required,

thermocycler and capillary electrophoresis equipment, are available in most laboratories involved in mutation screening.

6. MLPA allows the inclusion of control fragments for sufficient DNA quantities, proper DNA denaturation, presence of DNA contamination in RNA samples, etc.

Major advantages of MLPA for methylation detection are:

1. No bisulfite conversion is required.
2. Up to 50 probes can be multiplexed in one reaction.
3. The method is semi-quantitative.

Some limitations of MLPA are:

1. Similarly to real-time PCR and array-based techniques, MLPA cannot detect inversions and translocations that do not result in relative copy-number changes, unless specific probes are available detecting the exact breakpoint. Also, most point mutations will not be detected by MLPA.
2. Similarly to array-CGH, MLPA cannot distinguish female diploid, triploid, or tetraploid cells.
3. Reduced probe signals may not only reflect copy-number changes, but can also be due to SNPs or point mutations in the sequence detected by the probe that influence the ligation or destabilize the binding of the probe oligonucleotides to the sample DNA. Results obtained with a single probe should always be confirmed by other techniques.
4. The limited dynamic range of MLPA can complicate the development of mRNA-specific probe mixes.
5. The number of probes in a single assay is currently limited to approximately 50, making MLPA less suitable for genome-wide screening for copy-number changes.
6. In contrast to BAC-arrays, whole genome amplification methods are difficult to combine with MLPA.
7. Development of MLPA probe mixes is time consuming.

REFERENCES

Ahn, J.W., Ogilvie, C.M., Welch, A., Thomas, H., Madula, R., Hills, A., Donaghue, C., and Mann, K. (2007). Detection of subtelomere imbalance using MLPA: validation, development of an analysis protocol, and application in a diagnostic centre. *BMC Med. Genet.* 8, 9.

Ahn, J.W., Mann, K., Docherty, Z., and Ogilvie, C.M. (2008). Submicroscopic chromosome imbalance in patients with developmental delay and/or dysmorphism referred specifically for Fragile X testing and karyotype analysis. *Mol. Cytogenet.* 1, 2.

Aldred, M.A., Vijayakrishnan, J., James, V., Soubrier, F., Gomez-Sanchez, M.A., Martensson, G., Galie, N., Manes, A., Corris, P., Simonneau, G., Humbert, M., Morrell, N.W., and Trembath, R.C. (2006). BMPR2 gene rearrangements account for a significant proportion of mutations in familial and idiopathic pulmonary arterial hypertension. *Hum. Mutat.* 27, 212–213.

Aretz, S., Stienen, D., Uhlhaas, S., Stolte, M., Entius, M.M., Loff, S., Back, W., Kaufmann, A., Keller, K.M., Blaas, S.H., Siebert, R., Vogt, S., Spranger, S., Holinski-Feder, E., Sunde, L., Propping, P., and Friedl, W. (2007). High proportion of large genomic deletions and a genotype phenotype update in 80 unrelated families with juvenile polyposis syndrome. *J. Med. Genet.* 44, 702–709.

Arkblad, E.L., Darin, N., Berg, K., Kimber, E., Brandberg, G., Lindberg, C., Holmberg, E., Tulinius, M., and Nordling, M. (2006). Multiplex ligation-dependent probe amplification improves diagnostics in spinal muscular atrophy. *Neuromuscul. Disord.* 12, 830–838.

Armour, J.A., Sismani, C., Patsalis, P.C., and Cross, G. (2000). Measurement of locus copy number by hybridisation with amplifiable probes. *Nucleic Acids Res.* 28, 605–609.

Beetz, C., Nygren, A.O., Schickel, J., Auer-Grumbach, M., Bürk, K., Heide, G., Kassubek, J., Klimpe, S., Klopstock, T., Kreuz, F., Otto, S., Schüle, R., Schöls, L., Sperfeld, A.D., Witte, O.W., and Deufel, T. (2006). High frequency of partial SPAST deletions in autosomal dominant hereditary spastic paraplegia. *Neurology* 67, 1926–1930.

Bergval, I.L., Vijzelaar, R.N., Dalla Costa, E.R., Schuitema, A.R., Oskam, L., Kritski, A.L., Klatser, P.R., and Anthony, R.M. (2008). Development of a multiplex assay for the rapid characterisation of Mycobacterium tuberculosis. *J. Clin. Microbiol.* 46, 689–699.

Berry, I.R., Delaney, C.A., and Taylor, G.R. (2007). Detecting ligated fragments on oligonucleotide microarrays: optimizing chip design, array multiplex ligation-dependent probe amplification modification, and hybridization parameters. *Methods Mol. Biol.* 381, 247–265.

Bremmer, J.F., Braakhuis, B.J., Brink, A., Broeckaert, M.A., Beliën, J.A., Meijer, G.A., Kuik, D.J., Leemans, R.C., Bloemena, E., van der Waal, I., and Brakenhoff, R.H. (2008). Comparative evaluation of genetic assays to identify oral pre-cancerous fields. *J. Oral Pathol. Med.* 37, 599–606.

Bittel, D.C., Kibiryeva, N., and Butler, M.G. (2007). Methylation-specific multiplex ligation-dependent probe amplification analysis of subjects with chromosome 15 abnormalities. *Genet. Test.* 11, 467–475.

Bruno, D.L., Burgess, T., Ren, H., Nouri, S., Pertile, M.D., Francis, D.I., Norris, F., Kenney, B.K., Schouten, J., Choo, K.H., and Slater, H.R. (2006). High-throughput analysis of chromosome abnormality in spontaneous miscarriage using an MLPA subtelomere assay with an ancillary FISH test for polyploidy. *Am. J. Med. Genet.* 140, 2786–2793.

Cillessen, S.A., Hess, C.J., Hooijberg, E., Castricum, K.C., Kortman, P., Denkers, F., Vos, W., van de Wiel, M.A., Schuurhuis, G.J., Ossenkoppele, G.J., Meijer, C.J., and Oudejans, J.J. (2007). Inhibition of the intrinsic apoptosis pathway downstream of caspase-9 activation causes chemotherapy resistance in diffuse large B-cell lymphoma. *Clin. Cancer Res.* 13, 7012–7021.

Chou, L.S., Lyon, E., and Mao, R. (2008). Molecular diagnosis utility of multiplex ligation-dependent probe amplification. *Exp. Opin. Med. Diagn.* 2, 373–385.

Coffa, J., van de Wiel, M.A., Diosdado, B., Carvalho, B., Schouten, J., and Meijer, G.A. (2008). MLPAnalyzer: data analysis tool for reliable automated normalization of MLPA fragment data. *Cell. Oncol.* 30, 323–335.

Damato, E.M., Damato, B., Sibbring, J.S., and Coupland, S.E. (2008). Ciliary body melanoma with partial deletion of chromosome 3 detected with multiplex ligation-dependent probe amplification. *Graefes Arch. Exp. Ophthalmol.* 246, 1637–1640.

Damgaard, D., Nissen, P.H., Jensen, L.G., Nielsen, G.G., Stenderup, A., Larsen, M.L., and Faergeman, O. (2005). Detection of large deletions in the LDL receptor gene with quantitative PCR methods. *BMC Med. Genet.* 6, 15.

Depienne, C., Fedirko, E., Forlani, S., Cazeneuve, C., Ribaï, P., Feki, I., Tallaksen, C., Nguyen, K., Stankoff, B., Ruberg, M., Stevanin, G., Durr, A., and Brice, A. (2007). Exon deletions of SPG4 are a frequent cause of hereditary spastic paraplegia. *J. Med. Genet.* 44, 281–284.

Dikow, N., Nygren, A.O., Schouten, J.P., Hartmann, C., Krämer, N., Janssen, B., and Zschocke, J. (2007). Quantification of the methylation status of the PWS/AS imprinted region: comparison of two approaches based on bisulfite sequencing and methylation-sensitive MLPA. *Mol. Cell. Probes* 21, 208–215.

den Dunnen, J.T., and White, S.J. (2006). MLPA and MAPH: sensitive detection of deletions and duplications. *Curr. Protoc. Hum. Genet.* 12. Chapt. 7: Unit 7.14.

Eggermann, T., Schönherr, N., Eggermann, K., Buiting, K., Ranke, M.B., Wollmann, H.A., and Binder, G. (2007). Use of multiplex ligation-dependent probe amplification increases the detection rate for 11p15 epigenetic alterations in Silver-Russell syndrome. *Clin. Genet.* 73, 79–84.

Ehlert, A., Moreano, F., Busch, U., and Engel, K.H. (2007). Development of a modular system for detection of genetically modified organisms in food based on ligation-dependent probe amplification. *Eur. Food Res. Technol.* 227, 805–812.

Eichhorn, G.L. (1962). Metal ions as stabilizers or destabilizers of the deoxyribonucleic acid structure. *Nature* 194, 474–475.

Eldering, E., Spek, C.A., Aberson, H.L., Grummels, A., Derks, I.A., de Vos, A.F., McElgunn, C.J., and Schouten, J.P. (2003). Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signalling pathways. *Nucleic Acids Res.* 31, e153.

Gibbons, B., Datta, P., Wu, Y., Chan, A., and Armour, A.J. (2006). Microarray MAPH: accurate array-based detection of relative copy number in genomic DNA. *BMC Genomics* 7, 163.

Harteveld, C.L., Voskamp, A., Phylipsen, M., Akkermans, N., den Dunnen, J.T., White, S.J., and Giordano, P.C. (2005). Nine unknown rearrangements in 16p13.3 and 11p15.4 causing alpha- and beta-thalassaemia characterised by high resolution multiplex ligation-dependent probe amplification. *J. Med. Genet.* 42, 922–931.

Hess, C.J., Denkers, F., Ossenkoppele, G.J., Waisfisz, Q., McElgunn, C.J., Eldering, E., Schouten, J.P., and Schuurhuis, G.J. (2004). Gene expression profiling of minimal residual disease in acute myeloid leukaemia by novel multiplex-PCR-based method. *Leukemia* 12, 1981–1988.

Hess, C.J., Berkhof, J., Denkers, F., Ossenkoppele, G.J., Schouten, J.P., Oudejans, J.J., Waisfisz, Q., and Schuurhuis, G.J. (2007). Activated intrinsic apoptosis pathway is a key related prognostic parameter in acute myeloid leukemia. *J. Clin. Oncol.* 25, 1209–1215.

Hess, C.J., Errami, A., Berkhof, J., Denkers, F., Ossenkoppele, G.J., Nygren, A.O., Schuurhuis, G.J., and Waisfisz, Q. (2008). Concurrent methylation of promoters from tumor associated genes predicts outcome in acute myeloid leukemia. *Leuk. Lymphoma* 49, 1132–1141.

Holtkamp, N., Malzer, E., Zietsch, J., Okuducu, A.F., Mucha, J., Mawrin, C., Mautner, V., Schildhaus, H.U., and von Deimling, A. (2008). EGFR and erbB2 in malignant peripheral nerve sheath tumors and implications for targeted therapy. *Neuro. Oncol.* 10, 946–957.

Huang, C.H., Chang, Y.Y., Chen, C.H., Kuo, Y.S., Hwu, W.L., Gerdes, T., and Ko, T.M. (2007). Copy number analysis of survival motor neuron genes by multiplex ligation-dependent probe amplification. *Genet. Med.* 9, 241–248.

Isaksson, M., Stenberg, J., Dahl, F., Thuresson, A.C., Bondeson, M.L., and Nilsson, M. (2007). MLGA – a rapid and cost-efficient assay for gene copy number analysis. *Nucleic Acids Res.* 35, e115.

Jalali, G.R., Vorstman, J.A., Errami, A., Vijzelaar, R., Biegel, J., Shaikh, T., and Emanuel, B.S. (2007). Detailed analysis of 22q11.2 with a high density MLPA probe set. *Hum. Mutat.* 28, 433–440.

Janssen, B., Hartmann, C., Scholz, V., Jauch, A., and Zschocke, J. (2005). MLPA analysis for the detection of deletions, duplications and complex rearrangements in the dystrophin gene: potential and pitfalls. *Neurogenetics* 6, 29–35.

Jeuken, J., Cornelissen, S., Boots-Sprenger, S., Gijsen, S., and Wesseling, P. (2006). Multiplex ligation-dependent probe amplification: a diagnostic tool for simultaneous identification of different genetic markers in glial tumors. *J. Mol. Diagn.* 8, 433–443.

Jeuken, J.W., Cornelissen, S.J., Vriezen, M., Dekkers, M.M., Errami, A., Sijben, A., Boots-Sprenger, S.H., and Wesseling, P. (2007). MS-MLPA: an attractive alternative laboratory assay for robust, reliable, and semiquantitative detection of MGMT promoter hypermethylation in gliomas. *Lab. Invest* 87, 1055–1065.

Kanno, J., Hutchin, T., Kamada, F., Narisawa, A., Aoki, Y., Matsubara, Y., and Kure, S. (2007). Genomic deletion within GLDC is a major cause of non-ketotic hyperglycinaemia. *J. Med. Genet.* 44, e69.

Kirchhoff, M., Bisgaard, A.M., Bryndorf, T., and Gerdes, T. (2007). MLPA analysis for a panel of syndromes with mental retardation reveals imbalances in 5.8% of patients with mental retardation and dysmorphic features, including duplications of the Sotos syndrome and Williams-Beuren syndrome regions. *Eur. J. Med. Genet.* 50, 33–42.

Kluwe, L., Nygren, A.O., Errami, A., Heinrich, B., Matthies, C., Tatagiba, M., and Mautner, V. (2005). Screening for large mutations of the NF2 gene. *Genes Chromosomes Cancer* 42, 384–391.

Kousoulidou, L., Männik, K., Sismani, C., Zilina, O., Parkel, S., Puusepp, H., Tõnisson, N., Palta, P., Remm, M., Kurg, A., and Patsalis, P.C. (2008). Array-MAPH: a methodology for the detection of locus copy-number changes in complex genomes. *Nat. Protoc.* 3, 849–865.

Leonhard, W.N., Roelfsema, J.H., Lantinga-van Leeuwen, I.S., Breuning, M.H., and Peters, D.J. (2008). Quantification of Cre-mediated recombination by a novel strategy reveals a stable extra-chromosomal deletion-circle in mice. *BMC Biotechnol.* 8, 18.

Lok, C.A., Böing, A.N., Reitsma, P.H., van der Post, J.A., van Bavel, E., Boer, K., Sturk, A., and Nieuwland, R. (2006). Expression of inflammation-related genes in endothelial cells is not directly affected by microparticles from preeclamptic patients. *J. Lab. Clin. Med.* 147, 310–320.

Lundquist, P., Flynn, G., Thorland, E., and Dawson, D. (2005). Detection of chromosome 18q11–12 microdeletions in two Niemann-Pick type C families using luminex-based MLPA and FISH. *J. Mol. Diagn.* 7, 657–658.

Michils, G., Tejpar, S., Thoelen, R., van Cutsem, E., Vermeesch, J.R., Fryns, J.P., Legius, E., and Matthijs, G. (2005). Large deletions of the APC gene in 15% of mutation-negative patients with classical polyposis (FAP): a Belgian study. *Hum. Mutat.* 25, 34–125.

Moerland, E., van Hezik, R.L., van der Aa, T.C., van Beek, M.W., and van den Brule, A.J. (2006). Detection of HER2 amplification in breast carcinomas: comparison of Multiplex Ligation-dependent Probe Amplification (MLPA) and Fluorescence In Situ Hybridization (FISH) combined with automated spot counting. *Cell. Oncol.* 28, 151–159.

Monfort, S., Orellana, C., Oltra, S., Roselló, M., Guitart, M., and Martínez, F. (2006). Evaluation of MLPA for the detection of cryptic subtelomeric rearrangements. *J. Lab. Clin. Med.* 147, 295–300.

Nygren, A.O., Ameziane, N., Duarte, H.M., Vijzelaar, R.N., Waisfisz, Q., Hess, C.J., Schouten, J.P., and Errami, A. (2005). Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. *Nucleic Acids Res.* 33, e128.

Peppink, D., Douma-Kloppenburg, D.D., de Rooij-Askes, E.S., van Zoest, I.M., Evenhuis, H.M., Gille, J.J., and van Hagen, J.M. (2008). Feasibility and outcomes of multiplex ligation-dependent probe amplification on buccal smears as a screening method for microdeletions and duplications among 300 adults with an intellectual disability of unknown aetiology. *J. Intellect. Disabil. Res.* 52, 59–67.

Priolo, M., Sparago, A., Mammì, C., Cerrato, F., Laganà, C., and Riccio, A. (2008). MS-MLPA is a specific and sensitive technique for detecting all chromosome 11p15.5 imprinting defects of BWS and SRS in a single-tube experiment. *Eur. J. Hum. Genet.* 16, 565–571.

Procter, M., Chou, L.S., Tang, W., Jama, M., and Mao, R. (2006). Molecular diagnosis of Prader-Willi and Angelman syndromes by methylation-specific melting analysis and methylation-specific multiplex ligation-dependent probe amplification. *Clin. Chem.* 52, 1276–1283.

Pugh, T.J., Delaney, A.D., Farnoud, N., Flibotte, S., Griffith, M., Li, H.I., Qian, H., Farinha, P., Gascoyne, R.D., and Marra, M.A. (2008). Impact of whole genome amplification on analysis of copy number variants. *Nucleic Acid Res.* 36, e80.

Redeker, E.J., de Visser, A.S., Bergen, A.A., and Mannens, M.M. (2008). Multiplex ligation-dependent probe amplification (MLPA) enhances the molecular diagnosis of aniridia and related disorders. *Mol. Vis.* 14, 836–840.

Reijans, M., Dingemans, G., Klaassen, C.H., Meis, J.F., Keijdener, J., Mulders, B., Eadie, K., van Leeuwen, W., van Belkum, A., Horrevorts, A.M., and Simons, G. (2008). RespiFinder: a new multiparameter test to differentially identify fifteen respiratory viruses. *J. Clin. Microbiol.* 46, 1232–1240.

Rooms, L., Reyniers, E., Wuyts, W., Storm, K., van Luijk, R., Scheers, S., Wauters, J., van den Ende, J., Biervliet, M., Eyskens, F., van Goethem, G., Laridon, A., Ceulemans, B., Courtens, W., and Kooy, R.F. (2006). Multiplex ligation-dependent probe amplification to detect subtelomeric rearrangements in routine diagnostics. *Clin. Genet.* 69, 58–64.

Scarciolla, O., Stuppia, L., De Angelis, M.V., Murru, S., Palka, C., Giuliani, R., Pace, M., Di Muzio, A., Torrente, I., Morella, A., Grammatico, P., Giacanelli, M., Rosatelli, M.C., Uncini, A., and Dallapiccola, B. (2006). Spinal muscular atrophy genotyping by gene dosage using multiple ligation-dependent probe amplification. *Neurogenetics* 4, 269–276.

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., and Pals, G. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* 30, e57.

Schwartz, M., and Dunø, M. (2004). Improved molecular diagnosis of dystrophin gene mutations using the multiplex ligation-dependent probe amplification method. *Genet. Test.* 8, 361–367.

Schwartz, M., Dunø, M., Palle, A.L., Krag, T., and Vissing, J. (2007). Deletion of exon 16 of the dystrophin gene is not associated with disease. *Hum. Mutat.* 28, 205.

Scott, R.H., Douglas, J., Baskcomb, L., Nygren, A.O., Birch, J.M., Cole, T.R., Cormier-Daire, V., Eastwood, D.M., Garcia-Minaur, S., Lupunzina, P., Tatton-Brown, K., Bliek, J., Maher, E.R., and Rahman, N. (2008). Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) robustly detects and distinguishes 11p15 abnormalities associated with overgrowth and growth retardation. *J. Med. Genet.* 45, 106–113.

Sellner, L.N., and Taylor, G.R. (2004). MLPA and MAPH: new techniques for detection of gene deletions. *Hum. Mutat.* 24, 413–419.

Stegmann, A.P., Jonker, L.M., and Engelen, J.J. (2007). Prospective screening of patients with unexplained mental retardation using subtelomeric MLPA strongly increases the detection rate of cryptic unbalanced chromosomal rearrangements. *Eur. J. Med. Genet.* 51, 93–105.

Stern, R.F., Roberts, R.G., Mann, K., Yau, S.C., Berg, J., and Ogilvie, C.M. (2004). Multiplex ligation-dependent probe amplification using a completely synthetic probe set. *Biotechniques* 37, 399–405.

Sutton, I.J., Mocroft, A.P., Lindley, V.H., Barber, R.M., Bryon, R.J., Winer, J.B., and MacDonald, F. (2004). Application of multiplex ligation-dependent probe analysis to define a small deletion encompassing PMP22 exons 4 and 5 in hereditary neuropathy with liability to pressure palsies. *Neuromuscul. Disord.* 12, 804–809.

Takata, M. (2008). Detection of copy number changes at multiple loci in DNA prepared from formalin-fixed, paraffin-embedded tissue by multiplex ligation-dependent probe amplification. *Methods Mol. Biol.* 439, 101–108.

Taylor, C.F., Charlton, R.S., Burn, J., Sheridan, E., and Taylor, G.R. (2003). Genomic deletions in MSH2 or MLH1 are a frequent cause of hereditary non-polyposis colorectal cancer: identification of novel and recurrent deletions by MLPA. *Hum. Mutat.* 23, 428–433.

van Dijk, M.C., Rombout, P.D., Boots-Sprenger, S.H., Straatman, H., Bernsen, M.R., Ruiter, D.J., and Jeuken, J.W. (2005). Multiplex ligation-dependent probe amplification for the detection of chromosomal gains and losses in formalin-fixed tissue. *Diagn. Mol. Pathol.* 14, 9–16.

van Hagen, J.M., Eussen, H.J., van Schooten, R., van der Geest, J.N., Lagers-van Haselen, G.C., Wouters, C.H., De Zeeuw, C.I., and Gille, J.J. (2007). Comparing two diagnostic laboratory tests for Williams syndrome: fluorescent in situ hybridization versus multiplex ligation-dependent probe amplification. *Genet. Test.* 11, 321–327.

Van Opstal, D., Boter, M., de Jong, D., van den Berg, C., Bruggenwirth, H.T., Wildschut, H.I.J., de Klein, A., and Galjaard, R.J.H. (2009). Rapid aneuploidy detection with multiplex ligation-dependent probe amplification: a prospective study of 4000 amniotic fluid samples. *Eur. J. Hum. Genet.* 17, 112–121.

Vaughn, C.P., Lyon, E., and Samowitz, W.S. (2008). Confirmation of single exon deletions in MLH1 and MSH2 using quantitative polymerase chain reaction. *J. Mol. Diagn.* 10, 355–360.

Warshawsky, I., Chernova, O.B., Hübner, C.A., Stindl, R., Henneke, M., Gal, A., and Natowicz, M.R. (2006). Multiplex ligation-dependent probe amplification for rapid detection of proteolipid protein 1 gene duplications and deletions in affected males and carrier females with Pelizaeus-Merzbacher disease. *Clin. Chem.* 52, 1267–1275.

White, S.J., Vink, G.R., Kriek, M., Wuyts, W., Schouten, J., Bakker, B., Breuning, M.H., and den Dunnen, J.T. (2004). Two-color multiplex ligation-dependent probe amplification: detecting genomic rearrangements in hereditary multiple exostoses. *Hum. Mutat.* 24, 86–92.

Zeng, F., Ren, Z.R., Huang, S.Z., Kalf, M., Mommersteeg, M., Smit, M., White, S., Jin, C.L., Xu, M., Zhou, D.W., Yan, J.B., Chen, M.J., van Beuningen, R., Huang, S.Z., den Dunnen, J., Zeng, Y.T., and Wu, Y. (2008). Array-MLPA: comprehensive detection of deletions and duplications and its application to DMD patients. *Hum. Mutat.* 29, 190–197.

Chapter 14

Molecular Techniques for DNA Methylation Studies

Jörg Tost[1] and Ivo G. Gut[2]

[1] *Laboratory for Epigenetics, Centre National de Génotypage, CEA – Institut de Génomique, Evry, France;*

[2] *Department for Translational Research, Centre National de Génotypage, CEA – Institut de Génomique, Evry, France*

14.1 INTRODUCTION

All cells of a multicellular organism carry the same genetic code in their DNA sequence, while cells display broad morphological and functional diversity. The heterogeneity of cells is caused by differential expression of genes. Epigenetics can be defined as the study of mitotically (and in some cases meiotically) heritable changes of a phenotype, such as the gene expression of specific cell types that are not due to changes in the genetic code (Waddington, 1942). Epigenetic regulation mediates the adaption to an environment ultimately contributing to the phenotype.

Epigenetic phenomena are mediated by a variety of molecular mechanisms including post-transcriptional histone modifications, histone variants, ATP-dependent chromatin remodeling complexes, polycomb/trithorax protein complexes, small and other non-coding RNAs including siRNA and miRNAs, and DNA methylation (Tost, 2008). These diverse molecular mechanisms have all been found to be closely intertwined and stabilize each other to ensure the faithful propagation of an epigenetic state over time and especially through cell division. As the description of all methods used for the analysis of the different kinds of epigenetic modifications is beyond the scope of this chapter, the authors will concentrate on the DNA methylation as the best studied epigenetic biomarker. As such, DNA methylation has great potential as a diagnostic and predictive tool as well as therapeutic target. It should nonetheless be mentioned that the expression of small RNAs has recently been shown to be altered in many diseases (Boyd, 2008) and miRNA profiles allow not only the detection early-stage cancer but also the determination of its tissue of origin using blood-based miRNA profiling (Rosenfeld *et al.*, 2008). Further miRNA expression signatures have been identified that are associated with relapse and progression of human cancers (Yu *et al.*, 2008).

While this chapter focuses on DNA methylation as a covalent DNA modification it should be kept in mind that transcription does not occur on unpacked DNA but in the context of chromatin which critically influences the accessibility of the DNA to transcription factors and the DNA polymerase complexes. Chromatin modulations play a central role in shaping the epigenome and delineate a functional chromatin topology which serves as the platform forming regulatory circuits in all cells. Open (euchromatin) and closed (heterochromatin) chromatin states are controlled by histone modifications, histone composition, and the ATP-dependent chromatin remodeling machinery in close crosstalk with the binding of a plethora of non-histone proteins (Fig. 14.1). Positioning of the nucleosome at specific regulatory elements, such as a promoter as well as the degree of compaction (i.e. nucleosome spacing) contributes to the transcriptional potential of the corresponding gene. DNA methylation is highly related to these certain chromatin modifications and enzymes that modify DNA and histones have been shown to directly interact and constitute links between local DNA methylation and regional chromatin structure (Geiman and Robertson, 2002).

14.1.1 The Biology of DNA Methylation

DNA methylation is the only genetically programed DNA modification in mammals. This post-replication modification is almost exclusively found on the 5 position of the pyrimidine ring of cytosines in the context of the dinucleotide sequence CpG (Fig. 14.2; Bird, 2002). 5-Methylcytosine accounts for ~1% of all bases, varying

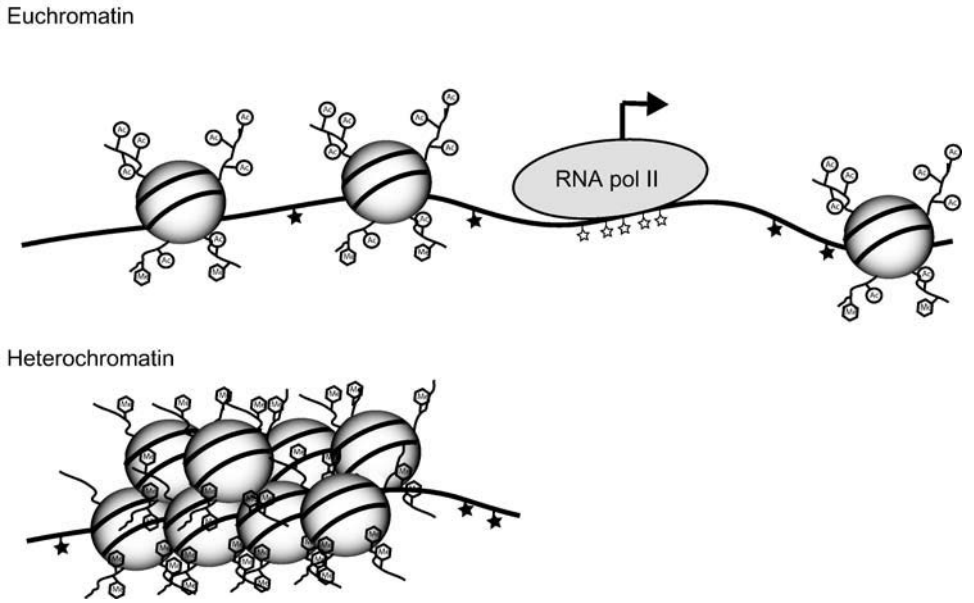


FIGURE 14.1 Simplified representation of euchromatin and heterochromatin. In euchromatin, RNA polymerase II can access the promoter CpG island (unfilled stars), which corresponds to a nucleosome-free region and the region is permissive to transcription. Histones are hyperacetylated (circle with Ac) including lysine 9 of histone H3 (H3K9) and lysine 4 of H3 (H3K4) is methylated (polygon with Me). In contrast, the DNA and nucleosome structure is compacted in heterochromatin. H3K4 is demethylated; H3K9, H3K27, and H4K20 are methylated. DNA is hypermethylated and the region is transcriptionally silent. For reasons of clarity, only selected histone modifications associated with the chromatin state are shown.

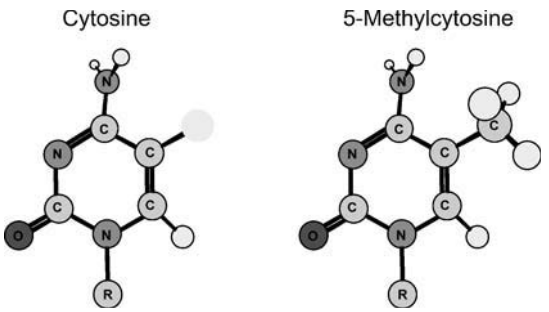


FIGURE 14.2 Chemical structure of cytosine and 5-methylcytosine. Cytosine is incorporated into the DNA using deoxycytidinetriphosphate as building block and methylated after its incorporation by DNA methyltransferases.

slightly in different tissue types and the majority (75%) of CpG dinucleotides throughout mammalian genomes are methylated. Other types of methylation, such as methylation of cytosines in the context of CpNpG or CpA sequences, have been detected in mouse embryonic stem cells and plants, but are generally rare in somatic mammalian/human tissues. CpGs are underrepresented in the genome, probably because they are mutational hotspots (deamination of methylated CpGs to TpGs). Despite this general trend, CpG-rich clusters of approximately 1–4 kb in length – so-called CpG islands – are found in the promoter region and first exons of many genes. CpGs in these islands are mostly not methylated which corresponds to the maintenance of an open chromatin structure and a potentially active state of transcription (Antequera, 2003). There are around 30,000 CpG islands in the human genome. About three quarters of transcription start sites and 88% of active promoters are associated with CpG-rich sequences and might be regulated by DNA methylation. Cytosine methylation of CpG dinucleotides is found in close proximity to critically important *cis*-elements within promoters and is often associated with a repressed chromatin state and inhibition of transcription.

Notably an unmethylated state of a CpG island does not necessarily correlate with the transcriptional activity of the gene, but rather with its potential for activation. On the other hand, the simple presence of methylation does not necessarily induce silencing of nearby genes. Only when a specific core region of the promoter becomes hypermethylated is the expression of the associated gene modified (Ushijima, 2005). DNA methylation often changes after modification of the chromatin structure and is used as a molecular mechanism to permanently and thus heritably lock the gene in its inactive state (Bird, 2002). Methylation can interfere with transcription in several ways. It can inhibit the binding of transcriptional activators with their cognate DNA recognition sequence such as Sp1 and Myc through sterical hindrance. Methyl-CpG-binding domain (MBD) proteins and the DNA methyltransferases (DNMTs) themselves bind to methylated DNA and thereby prevent binding of potentially activating transcription factors. These two protein families also recruit additional proteins with repressive function such as histone deacetylases (HDACs) and chromatin remodeling complexes to the methylated DNA to establish a repressive chromatin configuration creating a self-enforcing mechanism to silence genes.

14.1.2 Dynamics of DNA Methylation

During development and differentiation the mammalian organism creates a number of cell-type specific differentially marked epigenomes, whose identity is *inter alia* defined by their respective DNA methylation patterns. Consequently, one genome contains approximately 180 different epigenomes. Cytosine methylation is essential for mammalian embryogenesis, which is characterized by two waves of genome-wide epigenetic reprograming, in the zygote and in the primordial germ cells (Reik and Walter, 2001). Incomplete reprograming of these epigenetic patterns is one of the reasons for the low success rate of somatic cell nuclear transfer, i.e. the fusion of a somatic cell with an enucleated oocyte (''cloning'') (Niemann *et al.*, 2008). Modifications in the environment during early development can lead to permanent changes in the patterns of epigenetic modifications. This modulation of epigenetic patterns *in utero* has given rise to the developmental origin of disease hypothesis which postulates that the *in utero* environment can cause permanent changes to metabolic processes that directly affect postnatal phenotype, confers susceptibility to multifactorial disease at adult age, and may also be transmitted to subsequent generations (Tang and Ho, 2007). Epigenetics also holds the promise to explain at least a part of the influences the environment has on a phenotype as epigenetic changes are an integral part of aging and cellular senescence (Richardson, 2003). Studies in monozygotic twins demonstrated that epigenetic differences in genetically identical humans (monozygotic twins) accumulate with age and different environments create different patterns of epigenetic modifications (Fraga *et al.*, 2005). Chemical and environmental toxins have shown to induce changes to DNA methylation patterns without altering the genetic sequence and leading to epimutations associated phenotypes (Feil, 2006; Bollati *et al.*, 2007). DNA methylation has multiple essential roles in mammals. It is critical for imprinting, i.e. the asymmetric expression of either the maternal or paternal allele in a parent-of-origin-specific manner in somatic cells of the offspring (Reik and Walter, 2001) is implicated in X chromosome inactivation, the random silencing of one of the two X chromosomes in embryonic tissues of female mammals to achieve dosage compensation (Heard and Disteche, 2006). It is required for the maintenance of genome integrity (Eden *et al.*, 2003) and plays a key role in transcriptional silencing of repetitive DNA sequences and endogenous transposons (Yoder *et al.*, 1997).

14.1.3 DNA Methylation and Disease

DNA methylation and chromatin structure are strikingly altered in many pathological situations, particularly cancer. Although a number of genetic variations associated with disease susceptibility have recently been identified by genome-wide studies (Easton *et al.*, 2007; Hung *et al.*, 2008), they confer only small increases in risk and do not report on the onset of disease. Environmental factors undoubtedly play a large role in the actual occurrence of disease. The epigenome constitutes a memory of an organism to all the stimuli and influences it has ever been exposed to. Aberrant methylation patterns have been reported in various neurodevelopmental disorders and imprinting anomalies lead to disorders such as Prader-Willi, Angelman and Beckwith-Wiedemann syndromes (Robertson, 2005). DNA methylation patterns are globally disturbed in autoimmune diseases such as the lupus erythematosus (Balada *et al.*, 2007) or rheumatoid arthritis (Neidhart *et al.*, 2000) and it is very probable that epigenetic changes contribute to the susceptibility and development of many complex or multifactorial diseases (Junien and Nathanielsz, 2007; van Vliet *et al.*, 2007). Epigenetic mechanisms are consistent with various non-mendelian features of multifactorial diseases such as the relatively high degree of discordance in monozygotic twins. To further underline the scope of epigenetic alterations in disease, it is interesting to point out that monogenetic diseases such as α-thalassemia that have previously been attributed solely to genetic alterations can also be caused by epigenetic alterations at the same locus (Tufarelli *et al.*, 2003).

The high frequency of DNA methylation changes in cancer and the potential diagnostic use of DNA methylation for diagnostics and prognostics have fueled much of the rapid catch-up of technologies for epigenetic analyses that we have seen in the past few years and we will devote most of the rest of the chapter to describe applications of DNA methylation techniques in oncology.

14.1.3.1 DNA Methylation Changes in Cancer

Cancer is probably the best studied disease with a strong epigenetic component (Laird, 2005; Jones and Baylin, 2007). In tumors, a global loss of DNA methylation (hypomethylation) of the genome is observed (Feinberg and Vogelstein, 1983) and has been suggested to initiate and propagate oncogenesis by inducing chromosome instabilities and transcriptional activation of oncogenes and prometastatic genes (Ehrlich, 2002). The overall decrease in DNA methylation is accompanied by a region- and gene-specific increase of methylation (hypermethylation) of multiple CpG islands (Fig. 14.3; Laird, 2005; Jones and Baylin, 2007). Hypermethylation of CpG islands in the promoter region of a tumor suppressor or otherwise cancer-related gene is often associated with transcriptional silencing of the associated gene. The number of gene-associated promoters that are known to become hypermethylated during carcinogenesis is rapidly growing. Genes of numerous pathways involved in signal transduction (*APC*), DNA repair (*MGMT*, *MLH1*, *BRCA1*), detoxification (*GSTP1*), cell cycle regulation (*p15, p16, RB*), differentiation (*MYOD1*), angiogenesis (*THBS1*, *VHL*), and apoptosis (*Caspases,*

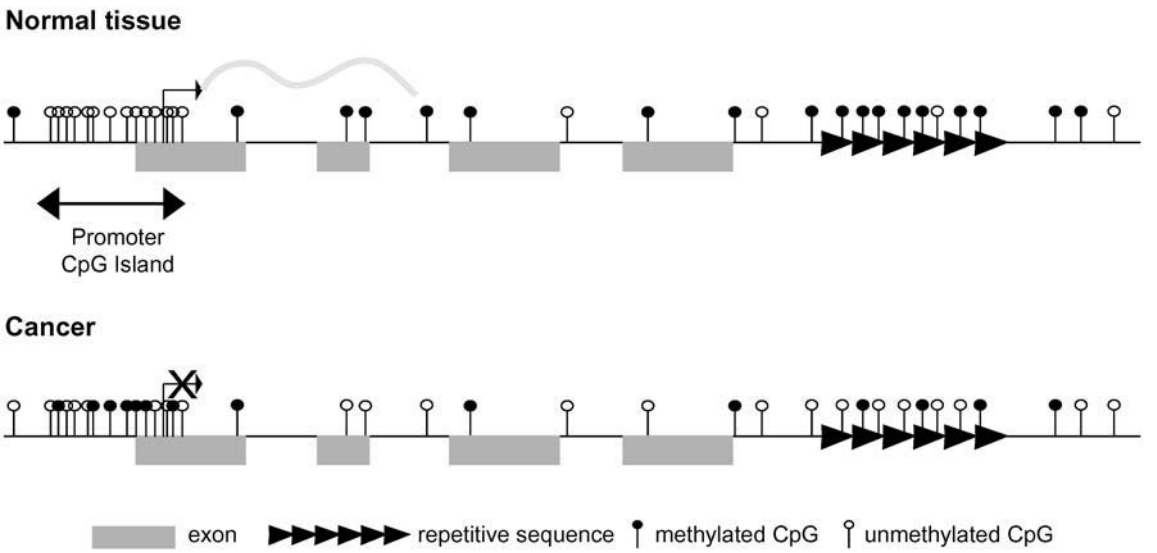


FIGURE 14.3 Distribution of DNA methylation in normal tissue and cancer. In the normal tissue, most promoter CpG islands are free of DNA methylation (indicated by white circles) even if the gene is not expressed. Repetitive elements as well as interspersed CpG dinucleotides are mostly methylated (indicated by black circles). In tumors, a global loss of DNA methylation (hypomethylation of the cancer genome) is observed while some promoter CpG islands become methylated in a tumor-type specific manner. Methylation patterns are dynamic and also change to a lesser extent during age and in response to environmental factors.

p14, DAPK) are often inappropriately inactivated by DNA methylation. It should be noted that so far no single gene has been identified that is always methylated in a certain type of cancer. Both hypo- and hypermethylation are found in the same tumor, but the underlying mechanisms for both phenomena have so far not been elucidated. A new dimension has recently been added to epigenetic cancer research with the demonstration of long-range gene silencing by epigenetic modifications (Frigola *et al.*, 2006). Long-range epigenetic silencing seems to be a prevalent phenomenon during carcinogenesis as a recent survey identified 28 regions of copy-number independent transcriptional deregulation in bladder cancer that are potentially regulated through epigenetic mechanisms (Stransky *et al.*, 2006). While the contribution of genetic factors to carcinogenesis such as the high-penetrance germ-line mutations in genes such as *BRCA1* and *p53* in familial cancers has long been recognized, it has become evident that epigenetic changes leading to transcriptional silencing of tumor suppressor genes constitute an at least equally contributing mechanism. For example, microarray expression profiles of breast tumors with *BRCA1* mutations are very similar to those of sporadic breast cancer cases with *BRCA1* promoter hypermethylation demonstrating that disruption of *BRCA1* function by either genetic or epigenetic pathways leads to the same perturbations (Hedenfalk *et al.*, 2001). It has been extrapolated that aberrant promoter methylation is initiated at ~1% of all CpG islands and as much as 10% become methylated during the multistep process of tumorigenesis (Costello *et al.*, 2000).

The methylation status of CpG islands can be used to characterize and classify cancers. While, for example, head and neck, breast or testicular tumors show overall low levels of methylation some other tumor types such as colon tumors, acute myeloid leukemias or gliomas are characterized by high levels of methylation, although some heterogeneity is observed in almost all tumor types. Methylation patterns can be shared by different types of tumors as well as being tumor type specific and methylation profiling can therefore identify distinct subtypes of human cancers (Costello *et al.*, 2000).

14.1.4 DNA Methylation as a Biomarker

Research has so far mainly focused on the hypermethylation of promoter associated CpG islands where hypermethylation is inversely correlated to their transcriptional activity. Correlation between DNA methylation and gene inactivation is a prerequisite for the identification and validation of novel functionally important genes, namely tumor suppressor genes. However, a large number of promoters become hypermethylated during carcinogenesis where there is no evidence that the corresponding gene acts as a tumor suppressor. In this case DNA methylation might still be a useful biomarker for tumor diagnosis or risk assessment if the methylation pattern is specific for a certain tumor type and/or correlates with clinically important parameters. A good example is the classic panel for the detection of the CpG island methylator phenotype defining a subtype of colorectal cancers with a distinct phenotype that comprises three MINT (methylated IN tumors) fragments (Issa, 2004). These fragments have been identified through differential methylation screening processes, but have only been later mapped to specific genomic loci. As described

above, the analysis of DNA methylation patterns is complicated by the fact that some changes are due to exposure to environmental influences as well as accumulation of DNA methylation at some promoters during aging (Feil, 2006). To be useful as biomarker, age-associated DNA methylation changes have therefore to be distinguished from cancer predisposing alterations.

Biomarkers capable of distinguishing diseased or malignant cells from normal ones must be specific, sensitive, and detectable in specimens obtained through minimally invasive procedures to be clinically applicable. Many biomarkers on the protein, RNA or DNA level fulfilling these criteria have been discovered. DNA-based molecular biomarkers are relatively easily transferred from a research laboratory setting into routine diagnostics due to the amplifiable and stable nature of DNA. Methyl groups on cytosines are part of the covalent structure of the DNA. Once methylation is acquired, it is in most cases chemically and biologically stable over time while expression of mRNA and/or proteins can be modified by non-disease-related environmental conditions and vary over the cell cycle. Most DNA methylation analysis methods determine the ratio of methylated and unmethylated CpGs and are independent of the total amount of starting material. It provides a binary and positive signal that can be detected independently of expression levels and is therefore easier to detect than negative signals like loss of heterozygosity. DNA methylation can be analyzed by an increasing number of methods that are amenable to high throughput, and quantitative assays eliminate the need for normalization. They are applicable to formalin-fixed paraffin-embedded clinical specimens and other archived material. One of the most important criteria for a clinically useful biomarker enabling screening of individuals at potential risk and monitoring of therapy response or disease recurrence is the analysis of the reliable biomarker in surrogate tissues such as blood or body fluids that can be obtained through minimal invasive procedures. The sensitive and specific detection of tumor-specific DNA methylation patterns at distal sites makes DNA methylation a biomarker of choice for the clinical management of cancer patients.

14.2 CLINICAL APPLICATIONS OF DNA METHYLATION ANALYSIS

Survival rates of cancer are greatly improved if the disease is detected at an early stage with locally limited disease and most common cancers can be cured at this stage. However, most traditional detection methods such as cytological screening do not provide the required sensitivity and some methods for early detection such as spiral computed tomography recommended for lung and other cancers are too expensive to be used on a population-wide scale. Tremendous effort has therefore been put into research to identify suitable molecular biomarkers for the detection of pre-neoplastic lesions. Epigenetic changes occur at higher frequency compared to genetic changes and maybe especially important in early-stage human neoplasia. DNA methylation changes often precede malignancy as extensive CpG island hypermethylation can be detected in benign polyps of the colon, in low- as well as in high-grade tumors (Goelz *et al.*, 1985; Costello *et al.*, 2000). It has therefore been suggested that epigenetic lesions in normal tissue set the stage for neoplasia. However, access to the primary tissue may be complicated, inconvenient, and does certainly not meet the needs for population-based screening for risk assessment, which requires the biological material to be obtained by cost-effective and minimally invasive procedures. Recent reports of DNA methylation analyses carried out on tumor biopsies and matched DNA samples extracted from body fluids such as serum, plasma, urine, and sputum have demonstrated a high level of concordance. The most effective way to detect the aberrant methylation is to analyze fluids that have been in physical contact with the site of the respective cancer. Tumors release a substantial amount of genomic DNA into the systemic circulation, probably through cellular necrosis and apoptosis although the exact source and mechanisms behind this phenomenon remain elusive (Jahr *et al.*, 2001). The amount of freely circulating DNA in plasma ranges from 10–30 ng/ml in healthy control subjects up to 500 ng/ml in some cancer patients with metastasized cancers. This freely circulating DNA contains the same genetic and epigenetic alterations that are specific to the primary tumor (Silva *et al.*, 1999). If the analyzed gene-specific methylation patterns are specific for a pathological state, methylation analysis of DNA recovered from plasma and serum can be used as a biomarker for molecular diagnosis and prognosis in various types of malignancies. Aberrant promoter hypermethylation was first demonstrated in the serum of non-small cell lung cancer patients by the analysis of *p16*, *DAPK*, *GSTP1*, and *MGMT* (Esteller *et al.*, 1999). In a recent study, methylation of at least one promoter was found in 51% of sera from stage I lung cancers while serum protein-based markers detected the presence of cancer in only 11% of cases (Fujiwara *et al.*, 2005). Examples include methylation of *p16* in liver (Wong *et al.*, 1999) and colorectal cancer patients (Zou *et al.*, 2002), *GSTP1* in patients with prostate cancer (Goessl *et al.*, 2000), *APC*, *RASSF1A*, and *DAPK* in breast cancer (Dulaimi *et al.*, 2004), or *XIST* methylation in testicular cancer patients (Kawakami *et al.*, 2004).

Measurement of serum prostate-specific antigen (PSA) is a routine test for the early detection screening of prostate cancer. Although one of the best conventional tumor markers, detection of increased PSA levels does not have sufficient sensitivity and specificity to ensure reliable diagnosis making secondary examinations mandatory and missing part of the cancer cases (Gray, 2005). CpG hypermethylation of the promoter region of the glutathione-*S*-transferase 1

(*GSTP1*) is a well-established biomarker for hormone-dependent cancers, especially prostate cancer. It is detectable in 90% of prostate cancers but not in normal prostate cells or benign hyperplasia. It is already detectable in precursor lesions and has been detected in urine and ejaculate of men with prostate cancer and carries great potential as a clinical diagnostic test (Goessl *et al.*, 2000; Cairns *et al.*, 2001; Nakayama *et al.*, 2004).

In tamoxifen-treated breast cancers, low methylation status of the estrogen receptor α (*ESR1*) correlated with reduced survival while increased methylation at the *ARH1* promoter was predictive for prolonged survival exclusively in non-tamoxifen treated breast cancers (Widschwendter *et al.*, 2004). *CYP1B1* methylation correlated with survival albeit less significantly in both classes of breast cancers. The presence of methylation was found to be a negative indicator in tamoxifen treated breast cancers but a protective effect was found in cancers not treated with tamoxifen.

Genome-wide methylation profiles obtained by differential methylation hybridization were used to define tumor subgroups of ovarian cancers that correlate with the time to disease recurrence after chemotherapy in patients and a selection of loci might be useful for prediction chemotherapy outcome (Wei *et al.*, 2002). 112 methylated loci were identified in ovarian cancer patients that enabled the prediction of progression free survival with an accuracy of 95% (Wei *et al.*, 2006).

14.3 METHODS FOR DNA METHYLATION ANALYSIS

For the analysis of DNA, methylation-sensitive and quantitative methods are required to detect even subtle changes in the degree of methylation as biological samples often represent a heterogeneous mixture of different cells, especially tumor and non-tumor cells from tissue biopsies. Realizing the importance of epigenetic changes in development and disease, a variety of techniques for the study of DNA methylation has been developed in the past few years (Fraga and Esteller, 2002; Laird, 2003; Brena *et al.*, 2006b; Schones and Zhao, 2008). No single method has emerged as the ''gold'' standard technique unifying quantitative accuracy and high sensitivity for whole genome analysis and precise investigation of individual CpG positions. So the choice of the method mainly depends on the desired application. Essentially, methods can be classed into two categories, genome-wide and targeted. Several generations of genome-wide methods have been developed adapting to different analytical supports with increasing levels of resolution while targeted approaches vary in function of their application.

14.3.1 The Methylation Content of a Sample

Methods for the analysis of global DNA methylation levels in a sample determine the overall 5-methylcytosine content or changes affecting the entire epigenome, respectively (Table 14.1). They do, however, not give any information about the location or repartition of DNA methylation on the genome. 5-Methylcytosine can be differentiated from its unmethylated counterpart by the different mass or polarity of the two bases (cytosine and 5-methylcytosine) which can be used for chromatographic or mass spectrometric separation (Wiebers, 1976; Eick *et al.*, 1983). The 5-methylcytosine content is measured after hydrolysis to mononucleosides and comparison to an internal standard enables quantification. Bacterial methyltransferases such as *Sss*I transfer a methyl group from the universal methyl donor *S*-adenosyl-L-methionine to unmethylated CpG positions. The methyl acceptor assay makes use of these enzymes and analyzes the amount of incorporated radiolabeled methyl groups into a sample (Bestor and Ingram, 1983). The measured amount of radioactive label correlates thus inversely with the degree of its methylation prior to labeling. Similarly, the cytosine extension assay combines methylation-sensitive restriction digestion and single nucleotide extension with radio- or fluorescently labeled dCTP complementary to the guanine 5′ overhang created by the digestion (Pogribny *et al.*, 1999; Bönsch *et al.*, 2004). The pyrosequencing-based luminometric methylation assay (LUMA) is based on the differential digestion of a sample with a methylation-sensitive endonuclease or its methylation-insensitive isoschizomer and the successive dispensation of four nucleotides complementary to the overhang created by the endonucleases (Karimi *et al.*, 2006). Methods analyzing the total amount of 5-methylcytosine in a sample are used to analyze and follow global DNA methylation changes induced by demethylating pharmaceutical agents in patients with hematological malignancies at various time points of treatment (Mund *et al.*, 2005; Liu *et al.*, 2007) or to investigate the efficacy of novel demethylating agents (Balch *et al.*, 2005).

In situ hybridization methods with antibodies directed against 5-methylcytosine allow the measurement of the methylation content and its potentially cell-type specific distribution (Miller *et al.*, 1974; Rougier *et al.*, 1998). As only clustered methylated CpGs prevalent in, for example, repeat elements can be recognized at the chromosomal level, methylation patterns at relatively small loci such as CpG islands contribute little to the overall staining profile. In a recent advancement an ultra-sensitive and rapid fluorescence scanning system with submicrometer resolution achieved the detection of methyl groups at specific promoters isolated from genomic DNA by restriction digestion and hybridization to capture oligonucleotides immobilized on a glass slide (Pröll *et al.*, 2006).

14.3.2 Genome-Wide Analysis of DNA Methylation Patterns

Microarray-based technologies (Table 14.2) are valuable methods for the identification of new epigenetic biomarkers

TABLE 14.1 Methods for the analysis of global DNA methylation levels.

Analysis technique	Principle/Readout format	Remarks (advantages/inconveniences)	Reference
Digestion/hydrolysis to mononucleosides	HPLC	Instrumentation widely available, large amounts of DNA are required (5–100 μg)	Kuo *et al.* (1980); Ramsahoye (2002)
		FFPE samples not suitable	
	(Fluorescent) Capillary electrophoresis	Increased sensitivity (100 ng-1 μg DNA required)	Fraga *et al.* (2002); Stach *et al.* (2003) Wiebers (1976)
	Mass spectrometry	Very sensitive (40 fmol), but requires expensive instrumentation	Friso *et al.* (2002); Liu *et al.* (2007a)
	2D Thin layer chromatography	Inexpensive, no special instrumentation required, but less accurate quantification	Cedar *et al.* (1979); Bestor *et al.* (1984)
Chemical derivatization	Modification of cytosines with chloracetaldehyde	Not restricted to CpG target sequences, independent of enzymatic performance, but multiple reaction steps and toxic chemicals	Oakeley *et al.* (1999)
Methyltransferase assay	Measurement of incorporated radioactive methyl groups	Rapid, inexpensive but requires radioactivity and prone to large inter-experimental variations	Bestor and Ingram (1983); Wu *et al.* (1993)
	Enzymatic regional methylation assay (ERMA), incorporation of radioactive methylgroups in PCR products amplified from bisulfite-treated DNA	Methylation content for a special region of interest, many non-radioactive methods are available	Galm *et al.* (2002)
Solid-phase primer extension	Methylation-sensitive restriction digest and incorporation of radioactive-labeled nucleotides on a solid surface	Multi-step procedure, radioactive	Heiskanen *et al.* (1994)
Cytosine extension assay	Methylation-sensitive digestion and labeled single nucleotide incorporation	Rapid, choice of enzyme permits targeting towards CpG islands	Pogribny *et al.* (1999); Bönsch *et al.* (2004)
Luminometric methylation assay (LUMA)	Luminometric monitoring of the incorporation of nucleotides into overhangs created methylation-sensitive restriction enzyme	Rapid, non-radioactive, internal calibration, only CpGs within the recognition sites are taken into account	Karimi *et al.* (2006)
Immunological detection	Antibody staining with antibodies specific for 5-methylcytosine	Detection of cell–cell specific differences in DNA methylation, low sensitivity	Miller *et al.* (1974); Rougier *et al.* (1998)
	Competitive solid-phase enzyme-linked immunoassay	Increased sensitivity, no information on the distribution	Reynaud *et al.* (1992)
Self-primed *in-situ* labeling (SPRINS)	Methylation-sensitive restriction digest and *in-situ* extension of created overhangs with modified nucleotides that are detected with fluorescent antibodies	Cell-to-cell specific methylation differences, complicated procedure, only subset of CpGs is interrogated	Andersen *et al.* (1998)

that do not require *a priori* knowledge of target DNA sequences. Their power to identify methylation markers for (early) diagnosis and classification of tumors has repeatedly been proven (Adorjan *et al.*, 2002; Cottrell and Laird, 2003; Laird, 2003; Ushijima, 2005). Methylation-based signatures are useful to predict disease progression and risk of relapse (Wei *et al.*, 2006) as well as response to specific cytotoxic drugs (Maier *et al.*, 2005). The aim of the study

TABLE 14.2 Methods for the genome-wide identification of differentially methylated genes.

Method	Principle of differentiation	Method	Enrichment of methylated/ unmethylated fragments	Detection platform	Comment	Reference
Methylation reversal	Reactivation of epigenetically silenced genes by treatment with methylation inhibitors	RNA expression analysis before and after treatment with demethylating agents	–	cDNA/expression microarrays	Identifies epigenetic modifications associated with transcriptional changes, potential confounding effects of demethylating drugs through methylation-independent transcription changes	Suzuki *et al.* (2002)
Methylation-sensitive arbitrarily primed PCR (MS-AP-PCR)	Methylation-sensitive restriction endonucleases (multiple possibilities)	Digestion with methylation-sensitive enzyme and its insensitive isoschizomer, random amplification, high resolution gel electrophoresis	Methylated CpG-rich regions	Polyacrylamide gel	Small amount of sample required, identifies DNA methylation changes independent of their location, limited resolution of the gel, band excision, and cloning required for sequence identification, interference from repetitive elements	Gonzalgo and Jones (1997)
	Same as MS-AP-PCR	Same as MS-AP-PCR, but additional *McrBC* digestion	Methylated CpG-rich regions	Polyacrylamide gel	More complex procedure compared to MS-AP-PCR, but less interference from repetitive elements	Tryndyak *et al.* (2006)
Methylation-sensitive restriction fingerprinting (MSRF)	Same as MS-AP-PCR	Same as MS-AP-PCR	Methylated CpG-rich regions	Polyacrylamide gel	Same as MS-AP-PCR	Huang *et al.* (1997)
Methylation-sensitive representational difference analysis (MS-RDA)	Methylation-sensitive restriction enzymes	Methylation-sensitive digestion of the sample, two–three rounds of adaptor ligation and ligation-mediated PCR, enrichment of unique sequences competitive hybridization to a sample treated with the isoschizomer	Differentially methylated regions	Enriched sequences are identified by cloning	Does not require prior knowledge on the location of methylation changes, technically complex, labor intensive	Ushijima *et al.* (1997)
Methylated CpG island amplification representational difference analysis (MCA-RDA)	Methylation-sensitive restriction enzymes	Methylation-sensitive digest leaving blunt ends, methylation-insensitive isoschizomer digest leaving an overhang, adaptor ligation and ligation-mediated amplification, RDA (see above)	Methylated regions	Identification by cloning	Technical complex, AIMS uses the same principle	Toyota *et al.* (1999)

Amplification of intermethylated sites (AIMS)	Methylation-sensitive restriction enzymes	Like MCA, but amplified digestion products are resolved on a polyacrylamide gel, bands are excised for identification by cloning	Methylated regions	Identification by polyacrylamide gel electrophoresis	Identifies DNA methylation changes independent of their location, limited resolution of the gel, band excision, and cloning required for sequence identification, interference from repetitive elements	Frigola *et al.* (2002)
Differential methylation hybridization (DMH)	Methylation-sensitive restriction endonucleases (multiple possibilities)	Adaptor ligation, digestion, ligation mediated amplification, hybridization	Methylated regions	CpG island library and other microarrays	Widely used, simple, analysis restricted to sequences with restriction enzyme recognition sets, sensitive to genetic polymorphisms (SNPs)	Huang *et al.* (1999); Shi *et al.* (2003a)
Methylation-sensitive amplified fragment length polymorphism (MS-AFLP)	Methylation-sensitive restriction endonucleases (multiple possibilities)	Digestion, adaptor ligation, and ligation-mediated amplification, hybridization	Unmethylated regions	Microarrays	Restricted to sequences with restriction enzyme recognition sets, sensitive to genetic polymorphisms (SNPs)	Yamamoto and Yamamoto (2004)
BAC microarrays	Methylation-sensitive restriction endonucleases (multiple possibilities)	Digestion, biotinylation, comparative hybridization	Unmethylated CpG-rich regions	Universal microarrays	PCR free, large amounts of DNA required, array suitable for a large number of different enzymes, improved resolution can be obtained with whole genome oligonucleotide tiling arrays	Ching *et al.* (2005)
Microarray-based integrated analysis of methylation by isochizomers (MIAMI)	Methylation-sensitive digestion and isoschizomer (*Hpa*I/*Msp*I)	Digestion, ligation-mediated amplification, co-hybridization	Differential display of unmethylated and methylated regions	Oligonucleotide microarrays	Simple, well suited for the identification of hypomethylated regions, less sensitive to genetic polymorphisms, requires enzyme recognition sites	Hatada *et al.* (2006)
*Hpa*II tiny fragment enrichment by ligation-mediated PCR (HELP) or	Methylation-sensitive digestion and isoschizomer (*Hpa*I/*Msp*I)	Digestion, ligation-mediated amplification, co-hybridization	Differential display of unmethylated and methylated regions	Oligonucleotide microarrays	Simple, well suited for the identification of hypomethylated regions, less sensitive to genetic polymorphisms, requires enzyme recognition sites	Khulan *et al.* (2006)
McrBC digestion	Methylation-dependent restriction endonucleases	Digestion, ligation-mediated amplification, hybridization	Unmethylated regions	Microarrays	Widely used, higher flexibility and coverage, not sequence specific, avoids interference from highly methylated repetitive elements	Lippman *et al.* (2005); Schumacher *et al.* (2006)

(*Continued*)

TABLE 14.2 (Continued)

Method	Principle of differentiation	Method	Enrichment of methylated/ unmethylated fragments	Detection platform	Comment	Reference
Restriction landmark genomic scanning (RLGS)	Methylation-sensitive restriction endonucleases (multiple possibilities)	Digestion, 2-dimensional electrophoresis	Unmethylated CpG-rich regions	2D gel	Very accurate and reproducible results, PCR free, computational spot identification possible, labor intensive, limited number of fragments can be resolved, large amounts of DNA required	Costello *et al.* (2002); Rush and Plass (2002)
Methylation single nucleotide polymorphism (MSNP)	Methylation-sensitive digestion	SNP genotyping of DNA before and after methylation-sensitive digestion	Methylated regions	SNP genotyping arrays	Identification of allele-specific DNA methylation, integration of a methylation-sensitive step in the optimized genotyping protocol, only large methylation differences can be resolved, arrays with 10k–1M probes readily available, copy-number changes, loss of heterozygosity is detected simultaneously	Yuan *et al.* (2006); Kerkel *et al.* (2008)
Methylated DNA immunoprecipitation (MeDIP)	5-Methylcytosine antibody	Sonication, immunoprecipitation, competitive hybridization against input DNA (after whole-genome amplification)	Methylated regions	BAC microarrays, oligonucleotide tiling arrays, next generation sequencing	Genome-wide analysis, not dependent on the presence of restriction enzymes, insensitive to genetic polymorphisms, bias towards CpG-rich regions (computational correction required), sensitivity is still discussed	Weber *et al.* (2005); Down *et al.* (2008); Rakyan *et al.* (2008)
Methylcytosine DNA-binding domain (MBD) proteins	Chromatin immunoprecipitation assay	Crosslink DNA to proteins, sonication, immunoprecipitation of crosslinked DNA with antibody to MBDs	Methylated regions	CpG island microarray	Sequence specificity of MBDs, potentially applicable to second generation sequencing	Ballestar *et al.* (2003)
Methylated CpG island recovery assay (MIRA)	DNA-binding domain (MBD2/MDB3L1) protein affinity selection	Sonication, immunopurification, competitive hybridization against input DNA (after whole-genome amplification)	Methylated regions	BAC microarrays, oligonucleotide tiling arrays, potentially next generation sequencing	Genome-wide analysis, not dependent on the presence of restriction enzymes, insensitive to genetic polymorphisms, might have a higher sensitivity than MeDIP	Rauch and Pfeifer (2005); Rauch *et al.* (2006)

will determine which method will be best suited. While some technologies identify specifically hypermethylated genes or CpG islands associated with specific promoters, other techniques will analyze regions of differential methylation between two samples that are not necessarily associated with genes and others again correspond to regions where pharmacological agents with demethylating activity induce changes on the transcriptional level. Independent of the method for sample preparation, technologies use more or less high-density nucleic acid microarrays, gels, or more recently second generation sequencing devices as detection platform. While microarrays with a selected probe library restrict the analysis to the loci present on the array, they directly identify the targets of interest, while gel-based approaches do detect specific loci but only after follow-up cloning and sequencing to identify sequences differentially methylated between two samples. It should be noted that, for example, restriction landmark genome scanning (RLGS) patterns (see below) are highly reproducible and cloning might only be necessary once to confirm the spot identity. Interference of normally highly methylated repetitive elements in the analysis might be another important point to consider. Microarrays can be roughly divided into arrays with immobilized bacterial artificial chromosomes (BAC) clones or PCR products which have been used extensively in the last decade and oligonucleotide arrays that depending on the platform are nowadays synthesized *in situ* by different procedures and contain small oligonucleotides with a length between 20 and 80 bp. The latter of course have a significantly improved resolution compared to the former. With the completion of the human genome sequence and the availability of a rapidly growing number of genome sequences for model organisms and the ever more advanced technologies for microarray fabrication the high-resolution interrogation of entire genomes or selected parts such as individual chromosomes or all promoters has become reality for different genomic applications including the analysis of genome-wide DNA methylation patterns (Mockler *et al.*, 2005). Tiling arrays containing several million features represent contiguous stretches of sequence and include annotated as well as regions of so far unknown function. They therefore enable unbiased and hypothesis-free approaches for the identification of novel transcriptional or regulatory elements as well as sequences subject to epigenetic modification during development or pathogenesis (Mockler *et al.*, 2005). Most methods for the genome-wide discovery of genes subject to DNA methylation changes rely on enrichment of sequences by either methylation-sensitive restriction enzymes or immunoprecipitation and in both cases subsequent PCR amplification. It should be noted that the antibody-based approach leads to a positive display of DNA methylation changes while the use of methylation-sensitive enzymes results, depending on the method, in a positive (DMH, AP-PCR) or negative (RLGS) genome-wide picture.

14.3.2.1 Methylation Reversal Approach

A method that sets itself a little bit apart is the methylation reversal approach, which identifies methylation patterns not by analyzing changes in the DNA methylation patterns itself, but by measuring mRNA expression after treating cells with a DNA methylation inhibitor (Suzuki *et al.*, 2002). Pharmacological agents such as the nucleoside analog 5-aza-2′-deoxcytidine (5-aza-dC) inhibit DNA methylation by forming a covalent complex between the methyltransferase and DNA which leads to a progressive global demethylation (Yoo and Jones, 2006). This approach is therefore restricted to cell lines that can be cultured. Cells can be treated with 5-aza-dC alone or in combination with a histone deacetylase inhibitor such as trichostatin A (TSA) that lead synergistically to gene activation of epigenetically silenced genes (Suzuki *et al.*, 2002). This approach thereby uses well-established expression microarrays as detection platform which are available from a number of commercial or academic sources. The main advantage of this approach is that the detection of hypermethylated sites is linked to their transcriptional status. However, modified expression patterns due to changes of methylation in *cis* such as demethylation of the promoter, enhancer, or other regulatory element of a target gene, or changes due to the activation of a gene that is in the same pathway as the target gene will be indistinguishable from each other. A further complication are the multiple effects of drugs such as 5-aza-dC on additional numerous metabolic pathways (Stresemann *et al.*, 2006), which might lead to methylation-independent changes in gene expression.

14.3.2.2 Methylation-Sensitive and -Dependent Restriction Enzyme-Based Methods

Methylation-sensitive restriction endonucleases, i.e. enzymes that are blocked by methylated cytosines in their recognition sequence (Bird and Southern, 1978), are widely used for the analysis of methylation patterns in combination with their methylation-insensitive isoschizomers (Table 14.2). One of the most commonly used pairs of enzymes is *Hpa*II/*Msp*I; both recognize and cleave the four base palindrome C|CGG in double-stranded DNA, but while *Msp*I cleaves the DNA independently of its methylation status, *Hpa*II is unable to cleave when the second cytosine is methylated (C|^{me}CGG). Another frequently used enzymatic combination is the methylation-sensitive enzyme *Sma*I (CC^{me}C|GGG) leaving blunt ends and *Xma*I (C|C^{me}CGGG) which is less sensitive to methylation and creates an overhang that can be used for the ligation of adaptors for subsequent PCR amplification. Although methods based on methylation-sensitive restriction enzymes are simple and cost effective as they do not require any special instrumentation, they are hampered by the limitation to specific restriction sites as only CpG sites found within these

sequences can be analyzed. For example, only ~4% of CpG sites in non-repetitive sequences are in recognition sites for the frequently used restriction enzyme *Hpa*II and only 0.03% can be cleaved by *Not*I (Fazzari and Greally, 2004). In addition, methods using these enzymes might be prone to false positive results due to incomplete cleavage and some sequences are intrinsically resistant to digestion if not appropriately controlled. For example, non-CpG methylation on cytosines or DNA adducts in the vicinity of the cleavage site might influence the restriction capacity of an enzyme. Digestions are therefore difficult to perform on material extracted from formaldehyde-fixed paraffin-embedded (FFPE) material. For the analysis of gene-specific methylation patterns or individual CpG positions, methods using methylation-sensitive endonucleases have largely been replaced by PCR-based methods following treatment of genomic DNA with sodium bisulfite, while many genome-wide discovery methods still rely on these enzymes for the isolation of either the methylated or the unmethylated fraction of the genome.

Differential methylation hybridization (DMH; Huang *et al.*, 1999) is a technically simple method based on comparative genomic hybridization (CGH, see also Chapter 12) to identify differentially methylated sequences. For DMH, two samples such as a paired tumor and normal sample or two differentially treated fractions of the same sample are hybridized to an array-based probe library and relative (fluorescent) intensities are measured. Genomic DNA is fragmented by digestion with a frequently cutting restriction enzyme that preferentially cuts outside CpG islands. Linkers for PCR amplification are subsequently ligated to the digestion products. The sample is split into two parts and one half digested with a methylation-sensitive restriction enzyme. In this sample only methylated fragments that are resistant to the digestion are amplified in the subsequent PCR, while in the reference sample all fragments are amplified. Fractions are labeled with two different fluorescent dyes and hybridized to microarrays with an immobilized CpG island library (Shi *et al.*, 2002). Immobilized CpG island libraries also permit the simultaneous analysis of gene expression using the subset of the CpG island library overlapping with the first coding exons of genes (Shi *et al.*, 2002) and histone modifications (Shi *et al.*, 2003b). Similar procedures focusing on unmethylated CpG-rich regions have been devised (Yamamoto and Yamamoto, 2004; Ching *et al.*, 2005).

Methylation-sensitive representational difference analysis (MS-RDA; Ushijima *et al.*, 1997) identifies differential methylation in two samples by a methylation-sensitive digestion. Adaptors are then ligated to the restriction products that are subsequently used for PCR amplification. Fragments that are only present in one library due to methylation at the restriction site can be isolated by representational difference analysis (RDA; Lisitsyn *et al.*, 1993). Amplification products are competitively hybridized together and sequences present at higher frequency in one of the samples are preferentially amplified by sequential cycles of PCR. Fragments are subsequently cloned for identification. Although technically complex, labor intensive and prone to several biases as it relies on repeated cycles of hybridization and PCR kinetics, the method permits the identification of differentially methylated regions without prior knowledge about its location and is used for the identification of novel tumor-related genes (see Noda and coworkers (2007) for a recent example).

Methylated CpG island amplification (MCA-RDA) selects methylated fragments like DMH (Toyota *et al.*, 1999). This is achieved by a first digestion with the methylation-sensitive restriction endonucleases *Sma*I, leaving blunt ends. The remaining and therefore methylated *Sma*I sites are cleaved with the methylation-insensitive isoschizomer *Xma*I, leaving a four base overhang that is used to ligate adaptors for subsequent PCR amplification. DNA fragments flanked by the two ligated adaptors are therefore amplified. Differences between two samples are identified through subtractive hybridization as described above. However, the subclone libraries contain a large number of repetitive fragments and only a small number of sequences with functional significance can be identified.

Amplification of inter-methylated sites (AIMS) is an improved version of MCA-RDA providing a fingerprint of sequences that are positioned between two methylated sites displayed on a high-resolution polyacrylamide gel (Frigola *et al.*, 2002).

Information complementary to those obtained by methylation-sensitive restriction digest can be obtained by the methylation-dependent restriction enzyme *McrBC*, which cleaves between two non-palindromic G/A^{me}C sites 40–3,000 bp apart from each other but cleavage is best if the two sites are separated by 55–103 bp (Stewart *et al.*, 2000). In a recent DMH approach, a cocktail of methylation-sensitive enzymes was used to enrich the methylated fraction of the genome potentially interrogating 32% of all CpG dinucleotides and in parallel enriching the unmethylated fraction by digestion with *McrBC* (Schumacher *et al.*, 2006). In parallel amplified genomic fragments that had not undergone any methylation-sensitive or -dependent digest were hybridized to microarrays to identify genetic variation such as polymorphisms that, if occurring at CpGs or in the restriction enzyme recognition sites, might influence hybridization efficiency independent of the methylation status (Schumacher *et al.*, 2006). *McrBC* can also be used on size-selected (1.5–4.0 kb) DNA to fractionate unmethylated (i.e. gel-purified high molecular weight) DNA after digestion, which is then comparatively hybridized with DNA similarly processed but not cut with the enzyme, on high density arrays (Ordway *et al.*, 2006; Irizarry *et al.*, 2008). Similarly, DNA fragments obtained after restriction endonucleases and McrBC digestion can be cloned and analyzed by high-throughput sequencing (Rollins *et al.*, 2006).

Instead of using the input DNA as reference sample, DMH has also been performed by comparing fluorescence intensities of a sample amplified from a methylation-sensitive digest and the same sample digested with its methylation-insensitive isoschizomer. These approaches have been termed MIAMI (microarray-based integrated analysis of methylation by isoschizomers; Hatada *et al.*, 2006) or HELP (*Hpa*II tiny fragment enrichment by ligation-mediated PCR; Khulan *et al.*, 2006). This approach has the advantage of being less susceptible to false positive results due to genomic alterations as the reference sample interrogates exactly the same sequences as the methylation-sensitive digest, but has a limited genomic coverage (Irizarry *et al.*, 2008).

An alternative approach – termed methylation single nucleotide polymorphism (MSNP) – analyzes the methylated fraction of the genome enriched by methylation-sensitive digestion on commercial SNP genotyping microarrays (Yuan *et al.*, 2006; Kerkel *et al.*, 2008). This approach permits the simultaneous genetic (DNA copy number/loss of heterozygosity) and epigenetic analysis of a genome with dense coverage. The advantage of this method is the integration of a methylation-dependent step into a highly standardized protocol which has been thoroughly validated, and random variations in signal intensities due to differential hybridization efficiencies have been minimized. It should be noted that depending on the methylation-sensitive restriction enzyme used only ~40% of SNPs are situated in amplicons containing restriction sites for the respective enzyme and are therefore informative (Yuan *et al.*, 2006). On the other hand, this still provides a large amount of data and as stated above the non-informative sites can still be used for complementary genetic analysis and for normalization of signal.

A more rapid but low-resolution picture of DNA methylation changes can be obtained by methylation-sensitive arbitrarily primed PCR (MS-AP-PCR) or methylation-sensitive restriction fingerprinting (MSRF; Gonzalgo *et al.*, 1997; Huang *et al.*, 1997). Genomic DNA is digested with a methylation-sensitive restriction enzyme or its methylation-insensitive isoschizomer followed by PCR amplification with random CpG-rich primers. Amplification products are separated by size on standard sequencing polyacrylamide gels under denaturing conditions. The absence or presence of bands between samples leads to a rapid discovery of differentially methylated DNA fragments. MS-AP-PCR suffers from the low resolution of gels which limits the number of fragments that can be analyzed and most of the fragments do not correspond to promoter regions of genes.

Restriction landmark genomic scanning (RLGS) is based on two-dimensional electrophoresis of fragments generated by methylation-sensitive digestion (Costello *et al.*, 2002). It permits the genome-wide quantitative assessment of epigenetic alterations between samples by a first digestion with a rare-cutting methylation-sensitive enzyme that cuts preferably in large CpG-rich sequences that are frequently found in CpG islands such as *Not*I. The created endonuclease sites are subsequently radiolabeled with ^{32}P-dCTP and ^{32}P-dGTP. Methylated CpGs are not cleaved and therefore not labeled which avoids interference with normally highly methylated repetitive elements in the subsequent readout. Fragments are further digested with frequently cutting restriction endonucleases and separated by two-dimensional electrophoresis. Differential methylation is analyzed by comparing spot intensities on the two-dimensional gel between samples or to an *in silico* generated pattern that also facilitates spot identification. RLGS patterns are characterized by a high reproducibility of fragment position and intensity. RLGS detects a large number of differentially methylated CpG islands (up to 1,800). RLGS avoids some potential pitfalls compared to some other discovery methods as it does not rely on hybridization kinetics and PCR amplification. This technology has been widely used to identify DNA methylation changes in cancer (Smiraglia and Plass, 2002) and to establish methylation patterns that are specific for a tumor type or shared between different tumor types (Costello *et al.*, 2000). Although RLGS is a method delivering high-quality results, its routine use is hampered by the requirement for a large amount of high-quality molecular weight DNA and the rather complex multistep procedure.

Several of the restriction enzyme-based strategies are currently being transferred to second generation sequencing instruments (FLX, 454 Corporation/Roche, Genome Analyzer Solexa/Illumina, SOLID Applied Biosystems; see also Chapter 24), which will permit a more quantitative analysis of the isolated (methylated or unmethylated) fraction of the genome. Second generation sequencing does not restrict the analysis to a subset of loci of the genome as microarrays do and the sequence identity is resolved in contrast to gel-based readouts. Although many questions concerning coverage of the genome, cost, and specificity remain unanswered at the moment, methods will be devised in the very near future.

14.3.2.3 Methylated DNA Immunoprecipitation (MeDIP) and Methyl Binding Protein Affinity Chromatography

The above described approaches rely on methylation-sensitive restriction endonucleases and analysis is restricted to sequences that contain recognition sequences for the enzymes in their proximity. An alternative approach isolates the methylated fraction of a genome by immunoprecipitation or affinity purification of methylated DNA with MBD proteins. The MBD2b protein has the highest affinity for methylated DNA of the different MBD proteins (Fraga *et al.*, 2003). In the recently devised methylated-CpG island recovery assay (MIRA) a glutathione S-transferase (GST)-tagged full length MBD2b has therefore been used to bind sonicated methylated DNA fragments and the affinity to

methylated CpG dinucleotides is further enhanced in a dose-dependent manner by the addition of the MBD3-like-1 protein (Rauch and Pfeifer, 2005). The combined effect significantly improves the sensitivity of the assay and a single methylated CpG dinucleotide enabled capture of the corresponding DNA molecule. Ligation of oligonucleotide linkers to enzymatically digested DNA prior to affinity chromatography permits efficient amplification of eluted fractions and subsequent analysis of differentially labeled input DNA and MIRA-enriched amplification products by DMH on a CpG island microarray (Rauch *et al.*, 2006) or potentially by second generation sequencing. A potential complication is the target specificity of MBDs and MBD columns enrich significantly but do not fully purify methylated sequences (Selker *et al.*, 2003).

An alternative approach – methylated DNA immunoprecipitation (MeDIP) – follows the protocol for chromatin immunoprecipitation analysis (ChIP), enriching methylated sequences independent of their surrounding sequence. Methylated DNA fragments are precipitated with a bead-immobilized antibody specific for 5-methylcytosine and analyzed locus-specifically by PCR amplification (Luo and Preuss, 2003) or genome-wide on various microarrays (Weber *et al.*, 2005). Precipitated DNA is amplified by linker mediated universal amplification, labeled and hybridized to the microarray. Hybridization can be carried out competitively, labeling input and enriched fraction with two different dyes (Weber *et al.*, 2005) or only the enriched fraction is labeled and put onto a single color array (Zhang *et al.*, 2006; Hayashi *et al.*, 2007). Resolution of the arrays depends mainly on the length of the immobilized probe. Tiling arrays are high-density oligonucleotide arrays potentially covering large genomic regions at high resolution (potentially at single cytosine residues). This approach created the first complete DNA methylation map of an organism resolving the ''methylome'' of *Arabidopsis thaliana* at a 35 bp resolution (Zhang *et al.*, 2006). Candidate methylated sites (CMS) identified by the MeDIP/tiling array approach were validated in a subset of genes positively confirming the hypermethylated status in 24/25 regions (Hayashi *et al.*, 2007) and the location of a large proportion of CMS correlated well with the location of core promoters determined by reporter assays to be situated most often at ~500-1 bp upstream of the transcription start site (Cooper *et al.*, 2006). The combination of MeDIP with tiling arrays was also used to define the methylome of human promoters (Shen *et al.*, 2007; Weber *et al.*, 2007) as well as 13 human somatic tissues identifying tissue-specifically methylated genes (Rakyan *et al.*, 2008). MeDIP has created a lot of interest in recent years; however, there might be bias in the enrichment of sequences preferentially precipitating sequences with a high density of methylated CpGs such as repetitive elements or methylated CpG islands (Irizarry *et al.*, 2008). Computational algorithms are required to account for this biased amplification and to convert the fluorescent signal intensities into methylation percentage (Down *et al.*, 2008; Pelizzola *et al.*, 2008).

The recent appearance of second generation DNA sequencers (FLX, 454 Corporation/Roche, Genome Analyzer Solexa/Illumina, SOLID Applied Biosystems) that allow sequencing many fragments in parallel very efficiently has led to the transfer of many array-based readouts to direct sequencing. MeDIP-seq is an approach that combines methylation-specific immunoprecipitation with sequencing of the pull-down product (Down *et al.*, 2008). The resulting short reads are mapped back onto the genome sequence. Stacked-up sequences indicate regions of the genome where the cytosines are methylated. A key problem of this is that repetitive elements are methylated and so result in a lot of sequence reads attributable to repeats. While the algorithms for sequence analysis largely eliminate sequences with multiple hits in the genome, sequencing space is lost on the sequencer.

Microarray-based discovery methods are well suited for the mapping of large-scale methylation variation in the whole genome, but do not provide quantitative high-resolution information of individual CpG sites. Genome-wide approaches yield a large number of candidate genes or regions potentially involved in the biological or medical question under investigation. Identified potential biomarkers have therefore to be validated using more quantitative methods in a larger sample set (potentially hundreds of samples).

14.3.3 Analysis of Candidate Gene Regions

The following methods permit analyzing the overall density of DNA methylation in an amplified target region after bisulfite treatment. Sodium bisulfite conversion of genomic DNA has revolutionized the field of DNA methylation analysis as it permits the use of well-established DNA amplification procedures from minute amounts of samples such as clinical specimens (Fig. 14.4; Frommer *et al.*, 1992). Bisulfite treatment of genomic DNA samples results in the hydrolytic deamination of non-methylated cytosines to uracils, while methylated cytosines are resistant to conversion (Shapiro *et al.*, 1974; Wang *et al.*, 1980). Following PCR amplification, in which uracils are replaced by their DNA analog thymine, the methylation status at a given position is manifested in the ratio C (former methylated cytosine) to T (former non-methylated cytosine) and can be analyzed as a virtual C/T polymorphism in the bisulfite-treated DNA. This chemical conversion of DNA is required to ''freeze'' the methylation status of a sample as DNA polymerases used for PCR amplification do not distinguish between methylated and non-methylated cytosines and therefore the methylation information is not retained after amplification. Sodium bisulfite treatment therefore enables the analysis of any CpG site of interest with quantitative resolution at the nucleotide level displaying CpG methylation as a positive signal.

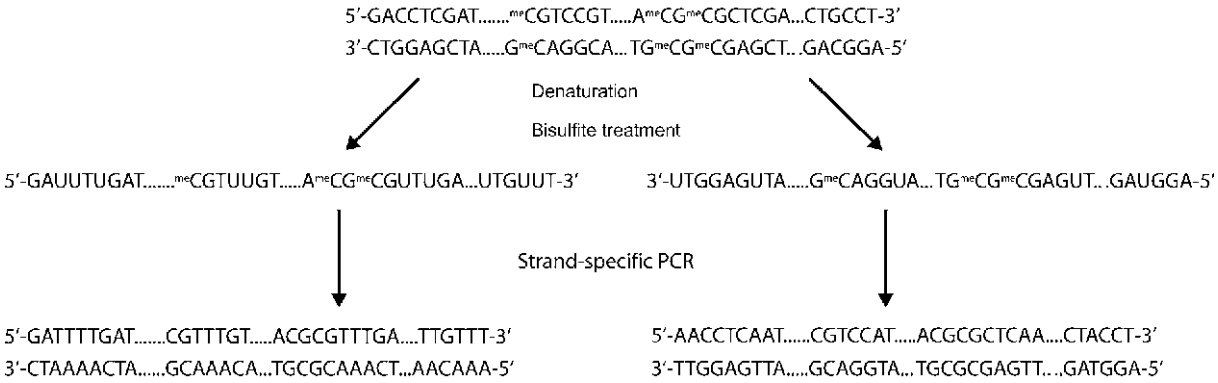


FIGURE 14.4 Principle of the treatment of genomic DNA with sodium bisulfite. Unmethylated cytosines are hydrolytically deaminated to uracils which in turn are replaced by the DNA analog thymine during subsequent PCR amplification. Methylated cytosines are resistant to the treatment under the carefully controlled reaction conditions and remain cytosines. Therefore any cytosine still present in the bisulfite-treated DNA corresponds to a methylated cytosine prior to conversion. The Watson and Crick strands that are complementary in native genomic DNA are no longer complementary after bisulfite treatment and both strands can be independently amplified by PCR. This process doubles the genome size while simultaneously reducing its complexity to a three letter code which makes primer design more challenging on bisulfite-treated DNA than on untreated genomic DNA.

Methods for the analysis of the global methylation levels of a target region do not yield precise information on the methylation status of individual CpGs and the distribution of allelic methylation patterns is only more or less resolved depending on the technique. These methods are valuable screening tools that might help select informative samples for more detailed analysis using techniques with increased resolving power such as sequencing, mass spectrometry-based sequencing and pyrosequencing. Analysis of DNA methylation levels in candidate regions can be semi-quantitatively analyzed by techniques like denaturing high-performance liquid chromatography (dHPLC; Deng *et al.*, 2002) or methylation-sensitive single-strand conformation analysis (Burri and Chaubert, 1999) making all use of the differential retention or migration behavior due to the sequence difference between former methylated and unmethylated molecules induced by bisulfite treatment. However, heterogeneous methylation levels might result in patterns which are difficult to interpret. Melting curve analysis (Fig. 14.5; see also Chapter 15) monitors the melting temperature of a PCR product amplified from bisulfite-treated DNA in real time using a thermocycler coupled to a fluorometer (Worm *et al.*, 2001; Akey *et al.*, 2002). The melting profiles of PCR products originating from methylated and unmethylated variants of the same template are significantly different due to their different GC content. Therefore the methylation status of an unknown sample can be determined by comparing the melting profile of the sample to calibration standards. A gradual increase of the temperature leads to a step-wise dissociation of the double strand in domains of the PCR product in function of their GC content differing between methylated and unmethylated molecules after bisulfite treatment. The technology has previously been hampered by the toxicity of the intercalating agent *SYBR Green I* to DNA polymerases which prohibited working at the required saturating concentrations. Recent advances in fluorescence detection technology, new algorithms for data calculation, and the use of novel dyes have allowed the development of this high-resolution melting analysis (HRM; Wojdacz and Dobrovic, 2007). This method is a promising advance as it is extremely simple and relatively inexpensive allowing the rapid scanning of a large number of genes for the presence of differential DNA methylation. First diagnostic applications for the detection of aberrant methylation profiles in imprinting disorders (White *et al.*, 2007; Alders *et al.*, 2008; Wojdacz *et al.*, 2008) and cancer (Balic *et al.*, 2009) have been proposed.

14.3.3.1 Methylation-Specific Oligonucleotide (MSO)

The methylation status of multiple regions can also be quantitatively and simultaneously interrogated by hybridization of fluorescently labeled DNA amplified from bisulfite-treated genomic DNA to microarrays. These contain paired oligonucleotide probes corresponding to either the methylated or the unmethylated allele of a specific target sequence after bisulfite treatment whereby each probe might cover one or several CpG positions (Adorjan *et al.*, 2002; Gitan *et al.*, 2002). MSO enables rapid screening of many CpG islands and/or promoters in parallel and is highly flexible as any sequence can be spotted onto the array independent of the presence of restriction sites. In an improved method bisulfite-treated DNA is amplified with random fluorescently labeled primers to interrogate simultaneously CpG positions in hundreds of genes (Yu *et al.*, 2005). Standard curves are required to account for differential hybridization kinetics of methylated and unmethylated sequences. However, potential cross-hybridization due to the decreased sequence complexity of bisulfite-treated

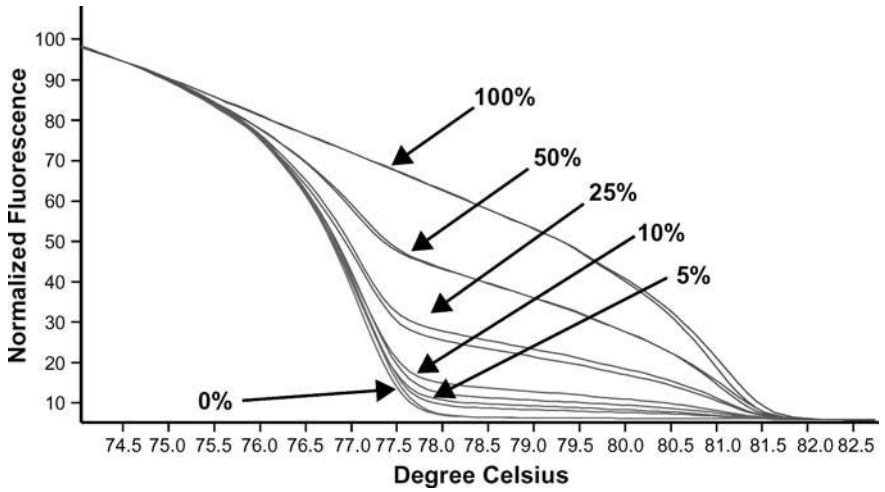


FIGURE 14.5 High-resolution melting analysis of different mixtures of methylated and unmethylated DNA (0, 5, 10, 25, 50, and 100%) in the promoter region of *RARB* gene.

DNA requires substantial optimization and validation for setting up a new array (Shi *et al.*, 2003a). The establishment of novel multiplexed high-throughput sequencing applications (see below) has made this technology somewhat redundant.

14.3.4 Sequencing Technologies

14.3.4.1 Sanger Sequencing of Bisulfite-Treated DNA

Direct fluorescent Sanger sequencing of PCR products amplified from bisulfite-treated DNA can be used for detailed analysis of DNA methylation in long CpG stretches from small amounts of starting material (Clark *et al.*, 1994). However, direct sequencing has proven to be technically demanding in regions with heterogeneous or low (<15%) methylation levels. Further, quantitation of the readout is imprecise and difficult due to the low sequence diversity of bisulfite-converted DNA requiring sophisticated computational algorithms such as ESME (Lewin *et al.*, 2004). Therefore in most cases PCR products are cloned and multiple colonies are subsequently sequenced. Cloning and sequencing of multiple colonies (Frommer *et al.*, 1992) is currently regarded as the reference method for the analysis of DNA methylation due to the richness of information provided and most novel methods are validated by comparison to the results obtained by sequencing and cloning. It provides detailed information on the methylation status of CpG positions in relatively long sequence stretches of target sequence and has been widely used in the past. Prior to the advent of second generation sequencing, cloning and subsequent Sanger sequencing has been the only technology capable of addressing adequately the problem of allele-specific and/or mosaic methylation patterns. However, the quantitative resolution is limited by the number of clones analyzed. As the procedure gets expensive, time consuming and cumbersome when applied in high throughput, only a small number of alleles in the order of about ten to 20 colonies are usually analyzed from which it is difficult to infer statistically meaningful results. Despite its status as a reference method, it should also be stressed that this procedure is very much prone to a variety of biases (PCR and cloning procedure) which distort the quantitative result. A variety of critical parameters as well as potential sources of artifacts and their remediation have been investigated in detail to avoid these pitfalls (Grunau *et al.*, 2001; Warnecke *et al.*, 2002).

14.3.4.2 Second Generation Sequencing of Bisulfite-Treated DNA

Much greater coverage and quantitative resolution can be obtained by second generation sequencing. The 454/Roche sequencing technology was applied to the analysis of target regions in great depth with several hundreds of clonal molecules that were analyzed for five samples and 25 target regions (Taylor *et al.*, 2007). Disease (patient)-specific nucleotide tails were added to the gene-specific amplification primers to allow multiplexing and to suit the large sequencing capacities of the instrument. Target regions were separately amplified and equal amounts of each PCR product were pooled prior to sequencing. Similarly, four genomic regions were simultaneously analyzed in 50 breast cancer patients (Korshunova *et al.*, 2008). Individual sequences could be attributed to the patients through a similar ''barcoding'' strategy. Both analyses required the individual amplification of regions of interest, making the analysis of a large number of target regions impractical. An alternative approach would be to use capture arrays with immobilized sequences complementary to genes/regions of interest which has been demonstrated for the resequencing

of genomic regions (Albert *et al.*, 2007). However, the low amount of material that can be recovered by this method and the therefore required large number of amplification cycles might bias the quantitative output.

For organisms with a limited genome size such as the plant *Arabidopsis thaliana* whole genome bisulfite sequencing (BS-seq) at single nucleotide resolution has been demonstrated (Cokus *et al.*, 2008; Lister *et al.*, 2008). A strategy to apply BS-seq to genome-wide studies in more complex organisms such as mice and humans consists of reducing the complexity of the genome by the creation of reduced representation libraries (Meissner *et al.*, 2008). The methylation-insensitive restriction endonuclease *Msp*I followed by size selection was used to isolate small fragments originating from CpG-rich regions thereby enriching significantly for CpGs in CpG islands. This fraction of the genome was then comprehensively sequenced. However, the isolated fraction covered only 1% of the genome. Whole genome BS-seq would be an ideal method as it would combine single nucleotide resolution with deep coverage of coordinate methylation patterns *in cis*. However, for the moment, BS-seq is unlikely to be widely adopted as it is quite cost prohibitive and a daunting task as the two strands of the DNA are no longer complementary after bisulfite treatment effectively doubling the size of the genome while simultaneously reducing its complexity to a three letter code.

14.3.4.3 Pyrosequencing

One of the methods that has received much attention for the simultaneous analysis and quantification of the methylation degree of several CpG positions in close proximity is Pyrosequencing™ (Uhlmann *et al.*, 2002; Colella *et al.*, 2003; Tost *et al.*, 2003a; Dupont *et al.*, 2004; see also Chapter 8). The pyrosequencing technology, in contrast to conventional Sanger sequencing, is based on sequencing-by-synthesis. A single nucleotide is added to a primer hybridized to a template strand and its incorporation is monitored by the luminometric detection of pyrophosphate that is released upon the nucleotide incorporation and converted into a light signal by a cascade consisting of three enzymes (Ronaghi *et al.*, 1998). Remaining nucleotides are degraded by the fourth enzyme – an apyrase – enabling cyclic dispensation of nucleotides in a pre-defined order. One of the major strengths of the technology is the quantitative nature of the results (see Chapter 8). The bioluminometric response is linear ($R^2 > 0.99$) for the sequential addition of up to five identical nucleotides (C, G, T) or three α-S-dATP. Pyrosequencing is therefore ideally suited for DNA methylation analysis after bisulfite treatment as it combines the ability of direct quantitative sequencing, reproducibility, speed, and ease of use (Tost and Gut, 2007). Information about an entire amplified region can be obtained by repeated stripping of the *de novo* synthesized strand from the template and hybridization of new sequencing primers (serial pyrosequencing; Tost and coworkers (2006)). Pyrosequencing provides quantitative information on the methylation status of all CpG positions in a sequence of interest and is therefore in contrast to MSP methods also ideally suited for the analysis of regions displaying heterogeneous methylation levels. Another advantage of pyrosequencing in comparison to fluorescent (real-time) MSP methods is the detection of the sequence surrounding the polymorphic CpG positions instead of a fluorescent signal which does not contain direct information about the actual amplified target. It thereby also permits the detection of SNPs present in the bisulfite-treated sequence. Besides several studies on genes aberrantly silenced by promoter hypermethylation in cancer (Fig. 14.6), pyrosequencing has been used for monitoring chemically induced demethylation in leukemia patients (Yang *et al.*, 2006) as well as for a diagnostic test for aberrant methylation in the imprinting disorders Prader-Willi and Angelman syndrome (White *et al.*, 2006).

14.3.4.4 MALDI Mass Spectrometry-Based Sequencing

A relatively new method for sequencing of PCR products amplified from bisulfite-treated DNA uses matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-ToF-MS) for which fragile biomolecules are mixed with a matrix of small organic molecules absorbing most of the energy of the laser which is used to volatilize the analyte and transfer part of it onto the analyte. MALDI has been proven very useful for complex analysis of nucleic acids such as HLA-typing, haplotyping, gene expression, DNA methylation analysis, and tissue imaging (Tost and Gut, 2006). As the direct analysis of DNA in a MALDI mass spectrometer is associated with some problems (Tost and Gut, 2002), bisulfite-treated DNA is transcribed into RNA which is subsequently base-specifically cleaved using a variety of RNAses (Schatz *et al.*, 2004; Ehrich *et al.*, 2005). In Sequenom's MassCLEAVE protocol, a modified T7 polymerase is used that does not discriminate between ribonucleotides and deoxyribonucleotides (Ehrich *et al.*, 2005). Transcription is carried out with three ribonucleotides and either CTP or UTP is replaced by the respective non-cleavable deoxynucleotide. RNAse A is used for the subsequent C or U specific cleavage. CpGs that are differentially methylated between two samples in the amplified region can be identified by shifts in the peak patterns due to the mass difference in a CpG dinucleotide containing cleavage product or the presence or absence of a peak corresponding to a certain fragment, respectively. The use of universal reverse transcription primers makes these procedures suitable for high-throughput applications with high quantitative resolution (5–10%) as recently shown by the analysis of more than 400 loci in 59 cancer cell lines

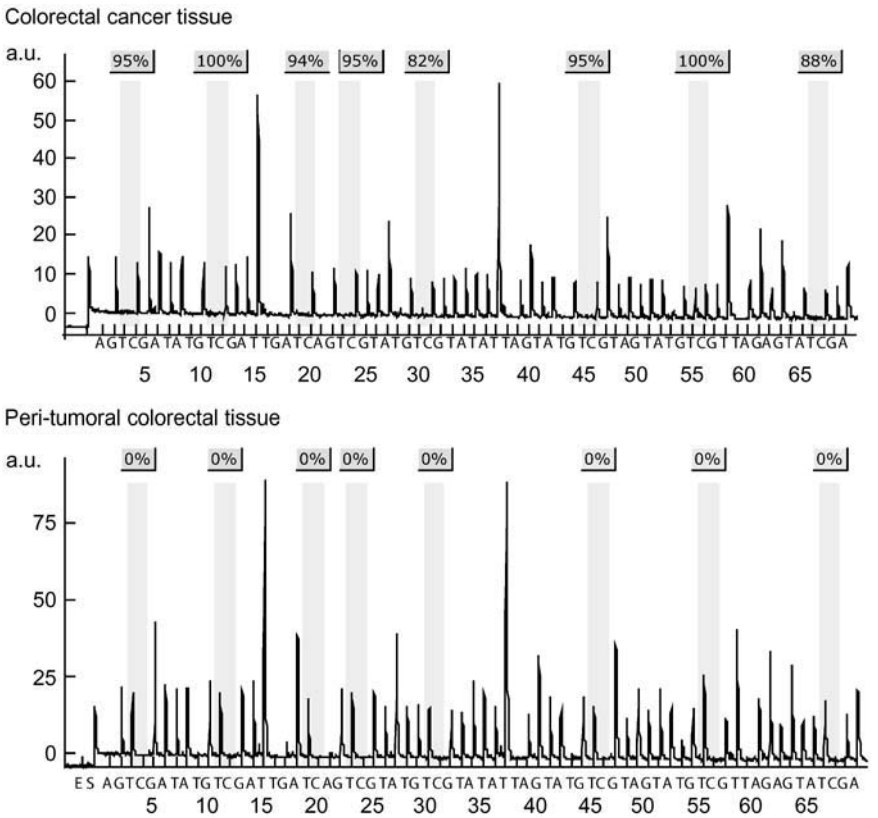


FIGURE 14.6 Pyrograms of the promoter region of the mismatch repair gene *MLH1*, which is commonly inactivated in microsatellite instable colorectal cancer by promoter hypermethylation. Top: DNA extracted from the tumor. Bottom: DNA extracted from distant peri-tumoral tissue of the same patient.

(Ehrich *et al.*, 2008). This approach can also be used to determine allele-specific DNA methylation patterns if SNPs are present within the analyzed fragment (Coolen *et al.*, 2007). Potential complications might, however, arise if several fragments of the same mass are created by the cleavage and peaks can therefore not be unambiguously assigned.

14.3.5 Analysis of Individual CpG Positions

14.3.5.1 Restriction Enzyme-Based Analysis of Single CpG Positions

Southern blotting is a simple and inexpensive procedure to analyze the methylation status at specific CpGs by digesting genomic DNA with methylation-sensitive restriction enzymes and subsequent Southern analysis (Bird and Southern, 1978). This approach enabled some of the most intriguing findings in the beginning of DNA methylation research such as the genome-wide hypomethylation of human cancers (Feinberg and Vogelstein, 1983) and provides authentic information on the methylation status of a CpG in an enzymatic recognition sequence as it does not rely on PCR amplification. However, the trade-off is that a large amount of DNA (several micrograms) is required and it is not well suited for heterogeneous population of methylated molecules that might occur, for example, in cancerous tissues. Exponential amplification by PCR amplification following methylation-sensitive restriction digestion is an alternative requiring substantially less DNA and no prior bisulfite conversion treatment making them well suited as rapid screening tools for differential methylation (Singer-Sam *et al.*, 1990). Multiple targets can be simultaneously analyzed by locus-specific multiplex PCR following methylation-sensitive restriction digest of genomic DNA (Melnikov *et al.*, 2005). Additional information on the methylation status of a target region can be achieved by digesting the DNA with either methylation-sensitive restriction enzymes or methylation-dependent enzymes such as the above described *McrBC* which distinguishes complete methylation, partial methylation, or absence of methylation in the sequence (Yamada *et al.*, 2004). Quantification can be improved by using ligation-mediated PCR to amplify the digested DNA and subsequent radioactive labeling (McGrew and Rosenthal, 1993) or – more easily – monitoring the increase in fluorescence by quantitative real-time PCR with the intercalating dye *SYBR Green* (Bastian *et al.*, 2005; Oakes *et al.*, 2006).

A method that has recently attracted a lot of interest for diagnostic applications is methylation-sensitive multiplex ligation-dependent amplification (MS-MLPA; Nygren *et al.*,

2005; see also Chapter 13). While initially devised for the analysis of copy-number alterations this technique has been useful for the parallel analysis of up to 40 loci permitting a comprehensive analysis for all possible variations of DNA methylation aberrations in imprinting disorders (Dikow *et al.*, 2007; Priolo *et al.*, 2008), the combined analysis of genetic and epigenetic alterations in imprinting disorders (Scott *et al.*, 2008), as well as tumor analysis (Berkhout *et al.*, 2007). Two oligonucleotides with universal primer binding sites are annealed to a target region and are ligated in case of complete target complementarity. A methylation-sensitive enzyme is added to the ligation reaction digesting unmethylated templates and reducing the amount of ligated product. A semi-quantitative readout is then performed using capillary electrophoresis.

Another method for rapid screening of samples relies on the creation of new restriction enzyme recognition sites through bisulfite treatment. Bisulfite-converted DNA samples are PCR amplified following restriction digestion, and methylation is analyzed through the presence or absence of bands on an agarose gel (Sadri and Hornsby, 1996). The quantitative accuracy of the assay can be improved if the analysis is carried out by polyacrylamide gel electrophoresis with subsequent electroblotting, hybridization with labeled oligonucleotides, and quantitation using a phosphoimager (COmbined Bisulfite Restriction Analysis; COBRA; Xiong and Laird, 1997). Though simple and inexpensive, COBRA becomes very labor intensive if quantitation is required. The latter point has been greatly improved by the implementation of the Agilent Bioanalyzer as detection platform (Brena *et al.*, 2006a). The Bioanalyzer separates and quantifies DNA fragments via electrophoresis in microfluidic chips for direct quantitative visualization of the restriction products created by COBRA.

14.3.5.2 Methylation-Sensitive Single Nucleotide Primer Extension (MS-SNuPE) Assays

The methylation degree at given CpG positions is manifested in the ratio C (former methylated cytosine) to T (former non-methylated cytosine) after bisulfite treatment and can be analyzed as a virtual C/T polymorphism after a strand-specific PCR. Quantitative and accurate information about individual methylation variable positions can therefore be obtained by methylation-sensitive single nucleotide primer extension (Ms-SNuPE). A PCR template is generated irrespective of the methylation status of the target region; internal primers are hybridized to the target sequence and terminate immediately 5′ of the CpG position to be assayed. Extension is carried out with (labeled) terminating dideoxynucleotides. The analysis of individual CpG positions requires prior knowledge about which CpG positions are of interest as, for example, CpG islands can contain hundreds of CpGs. However, once these positions are known, methods based on primer extension are very well suited to high-throughput analysis of a few CpGs in large sample cohorts enabling epigenetic analyses on an epidemiological scale. The close proximity of potentially polymorphic positions in a CpG-rich region such as CpG islands complicates the design of SNuPE assays irrespective of the used detection platform as this might cause preferential annealing of the primers to a subpopulation of methylated molecules displaying a specific methylation pattern. A variety of detection platforms have been combined with MS-SNuPE (Table 14.3). Mass spectrometry-based assays (Tost *et al.*, 2003b; Ragoussis *et al.*, 2006) permit higher levels of multiplexing and were applied in a multiplexed form as reference method for verification and quantitative fine typing of the results obtained by direct bisulfite sequencing for the Human Epigenome Project (Rakyan *et al.*, 2004). However, they require more sophisticated instrumentation in contrast to capillary electrophoresis-based methods such as MS-SNaPshot (Uhlmann *et al.*, 2002; Kaminsky *et al.*, 2005).

Illumina Inc. (http://www.illumina.com) has presented a modified version of the Golden Gate and Infinium genotyping assay for the analysis of DNA methylation. The Golden Gate assay analyzes 1505 CpGs in 807 genes using 250 ng of bisulfite-converted DNA as input (Bibikova *et al.*, 2006b). Allele discrimination is achieved prior to amplification: two allele-specific oligonucleotides that have their 3′ terminus complementary to the methylated or unmethylated alleles of a target CpG and that are extended only if the primer perfectly matches the target and two locus-specific oligonucleotides that are also complementary to an either unmethylated or methylated allele (Bibikova *et al.*, 2006b). Extended allele-specific oligonucleotides are then ligated to the locus-specific oligonucleotides to create templates for subsequent amplification with fluorescently labeled primers specific for the respective methylation state. Allele discrimination is thus achieved at two different levels increasing the specificity of the reaction and minimizing false positive results. The locus-specific oligonucleotide also carries a sequence that is complementary to tag sequences immobilized on the bead arrays which is then used for readout of the quantitative methylation degree at selected CpG positions with a resolution of about 20%. Results are normally confirmed by high-resolution quantitative methods such as pyrosequencing (Ladd-Acosta *et al.*, 2007). The Golden Gate assay has been used to study the DNA methylation profile of embryonic stem cells (Bibikova *et al.*, 2006a), brain tissue (Ladd-Acosta *et al.*, 2007) as well as to identify DNA methylation changes specific to Wilms tumor (Bjornsson *et al.*, 2007). This approach will probably be very useful to scan a larger number of genes or CpG positions to identify targets of epigenetic deregulation in disease, but as only one to three CpGs are analyzed per gene the chances of missing changes that are restricted to a few CpG sites is quite high.

TABLE 14.3 Single nucleotide primer extension methods for the analysis of individual CpG positions after bisulfite treatment.

Method	Label of ddNTPs	Separation	Detection platform	Multiplexing level	Comment	Reference
MS-SNuPE	Radioactive	Polyacrylamide gel electrophoresis	Phosphoimager	3–5 (primers differing in length)	Accurate, but quite labor intensive. Two reactions required for quantification	Gonzalgo and Jones (1997); Gonzalgo and Jones (2002)
SNuPE IP RP HPLC	None	Mass and hydrophobicity	Ion-pair reverse phase HPLC	3 (non-complementary tails at the 5′end)	Medium throughput	El-Maarri *et al.* (2002)
MethylQuant	None	Allele-specific primer extension with locked nucleic acid-modified primers	Real-time thermocycler	No	Easy to implement, readily available instrumentation, high throughput, but two separate reactions for the methylated and unmethylated allele	Thomassin *et al.* (2004)
GOOD assay	None	Mass	MALDI mass spectrometry	3–5 (different masses)	High throughput, simple mass spectrometric signatures	Tost *et al.* (2003b)
SNaPshot	Fluorescence	Capillary electrophoresis	Capillary electrophoresis	3 (non-complementary tails at the 5′end)	Medium throughput, instrumentation available in many laboratories	Uhlmann *et al.* (2002); Kaminsky *et al.* (2005)
Iplex	None	Mass	MALDI mass spectrometry	Up to 27	High throughput	Ragoussis *et al.* (2006)

Further, the assay is based on the hypothesis that closely neighbored CpG positions display a high degree of co-methylation which might be a simplified assumption. The Infinium assay analyzes for 27,578 CpG sites in more than 14,000 genes. Twelve samples are processed in parallel. Similar to the respective genotyping assay, the 3′ base is specific for the methylation state which is extended only in the case of complete complementary with one of the four fluorescently labeled ddNTPs. A different bead ''barcode'' distinguishes between probes specific for the methylated and unmethylated state of a given CpG. This assay allows comprehensive analysis of the DNA methylation state of a sample in high-throughput studies, but again only 1–3 CpGs are analyzed per gene.

14.3.6 Methylation-Specific PCR and Real-Time PCR Methods

DNA methylation has received a lot of attention in recent years because of its potential as a stable and amplifiable biomarker for early diagnosis, prognosis, or response to treatment in various cancers (Laird, 2003). Biomarkers capable of distinguishing cancerous cells from normal ones must be specific, sensitive, and detectable in specimens obtained through minimally invasive procedures to be clinically applicable. Methylation-specific PCR (MSP), (Herman *et al.*, 1996) and methylation-specific real-time PCR-based methods such as MethyLight (Eads *et al.*, 2000a), HeavyMethyl (Cottrell *et al.*, 2004), or QAMA (Zeschnigk *et al.*, 2004) have proven very well suited for the detection of cancer-specific methylation patterns in primary tumor tissues but also – and more importantly – for the detection of very low levels of methylation in circulating DNA (Fig. 14.7). Tumor-derived methylated DNA molecules can be found in various body fluids such as urine, sputum, or serum/plasma of cancer patients. Detection and monitoring of these methylated molecules in the presence of an excess of normal (and usually unmethylated) DNA can be achieved with various approaches described below. As the amplicons of the different MSP variants are small (~100 bp), these methods usually work well with DNA of lower quality such as DNA extracted from FFPE samples

(Herman *et al.*, 1996). No special equipment is required for conventional methylation-specific PCR (MSP) and real-time PCR machines are available at most research institutions for the quantitative analysis of gene expression. The similarity of the approaches described below to real-time expression analysis also facilitates implementation of the technology, execution of the experiments, and interpretation of the results for laboratories not yet very familiar with DNA methylation analysis. The design of assays and optimization of amplification are probably the most important part to ensure specific amplification of the desired locus. Sensitivity and specificity vary largely between assays depending on primers (and probes in case of techniques like MethyLight) and conditions.

14.3.6.1 Methylation-Specific PCR (MSP)

MSP allows the amplification of virtually any CpG sites after bisulfite treatment with three pairs of primers for amplification, complementary to the former methylated, the former unmethylated sequences or to genomic, unconverted DNA, respectively (Fig. 14.7; Herman *et al.*, 1996). The latter serves as control for complete bisulfite conversion. Primers need to hybridize to sequences with at least two methylation variable positions (CpGs) to obtain the necessary specificity for selective amplification. The presence or absence of an amplification product analyzed on a conventional agarose gel reveals the methylation status of the CpGs underlying the amplification primers. It is the most widely used technology for DNA methylation analysis as it does not require any expensive instrumentation and a large number of samples can be rapidly assessed. The main advantage of MSP is the high sensitivity, which enables the detection of one allele in the presence of a 1,000-fold excess of the other (Herman *et al.*, 1996). It permits the detection of methylation at high throughput but this is limited to the primer binding sites. The detection limit can be further decreased by fluorescent labeling of one of the MSP primers and analysis of the amplification product on a sequencer (Fig. 14.7; Goessl *et al.*, 2000). MSP is well suited for the rapid and sensitive detection of methylation patterns at specific sequences, but does not provide resolution at the individual nucleotide level and heterogeneous methylation patterns at the primer binding sites can induce failure of amplification. Further, the biased amplification leads to a more qualitative than quantitative result making it difficult to distinguish different degrees of methylation at the target sites. An additional major drawback of the MSP approach is the gel-based detection of the amplification products. While acceptable for use in a research laboratory, diagnostics in a clinical setting require homogeneous high-throughput assays such as the fluorescent real-time MSP approaches described below (Cottrell and Laird, 2003).

Most real-time PCR-based methods use the same principle as the TaqMan® assay (Holland *et al.*, 1991; see also Chapter 7). In addition to the two amplification primers, a third oligonucleotide called probe, which is dually labeled with a fluorescent reporter (e.g. FAM) and a quencher dye (e.g. TAMRA), hybridizes to a target sequence in the amplified region. The technology makes use of fluorescence resonance energy transfer (FRET): if the probe is fully complementary to the amplified sequence, the probe is cleaved by the $5' > 3'$ exonucleolytic activity of Taq polymerase during the extension phase of the amplification

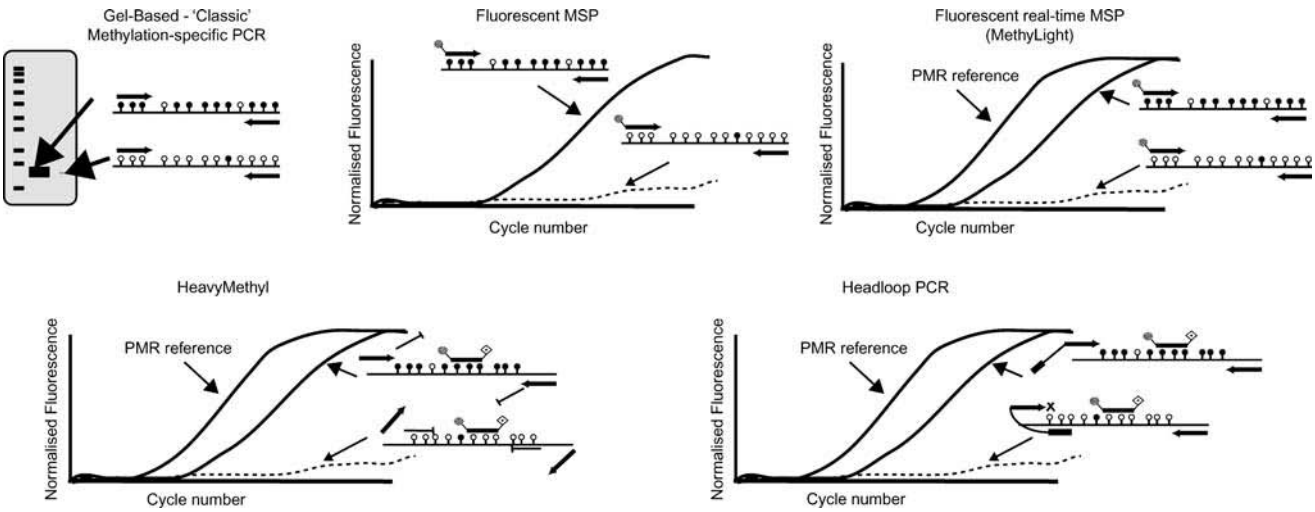


FIGURE 14.7 Real-time PCR methods for the sensitive detection of DNA methylation in, e.g., body fluids. For simplification only amplification with a primer complementary to a completely methylated allele is shown. PMR: percentage of methylated allele, a completely methylated DNA standard that is used for the calculation of the percentage of methylation contained within a sample. All methods use sodium bisulfite treatment prior to PCR amplification. CpGs are depicted as lollipops; former unmethylated CpGs are shown as empty lollipops while filled ones correspond to former methylated CpGs. Although methylation is retained as a sequence difference after bisulfite treatment, lollipops are shown for easier differentiation of the alleles. Details of the different techniques are given in the text.

resulting in spatial separation of the two dyes and a fluorescent signal proportional to the amount of PCR product generated. Real-time quantitative PCR enables thereby the non-isotopic, rapid, and accurate quantitative analysis over a large dynamic range of several CpG positions in a small target region. The addition of a third probe that has to anneal correctly to the synthesized template improves sensitivity as well as specificity compared to conventional MSP with the drawback that expensive hybridization probes are required. The simple one-step procedure makes real-time methylation-specific PCRs rapid high-throughput assays for quantitative DNA methylation analysis that are robust and quite resistant to carryover contamination. These approaches do not only provide information if molecules with a certain methylation pattern are present in the sample, like conventional MSP, but also report on how many of them there are. Discrimination between methylated and unmethylated alleles can be achieved at different levels of the primers and/or the hybridization probe (Eads *et al.*, 2000a). Although in principle primers and probes could be designed for different combinations of methylated and unmethylated alleles, the most widely used approaches such as MethyLight use primers and probes that are specific for the same methylation patterns mostly completely methylated molecules. Heterogeneous methylation patterns that display large variations between consecutive CpGs complementary to primers or probe will therefore probably lead to failure of the assay or biased quantitative results.

14.3.6.2 MethyLight

MethyLight can detect a single hypermethylated allele against a background of 10,000 unmethylated alleles (Fig. 14.7; Lo *et al.*, 1999; Eads *et al.*, 2000a). Absolute quantification of the number of molecules corresponding to the investigated pattern of methylation is calculated by the ratio between the gene of interest and a reference gene, for example the *β-actin* (*ACTB*) gene. The percentage of fully methylated molecules (Fig. 14.7; PMR, percentage of fully methylated reference) is calculated by dividing the gene-to-reference ratio by the gene-to-completely methylated reference ratio (which is obtained by *Sss*I treatment of the normally unmethylated reference) and multiplying by 100 (Eads *et al.*, 2000b). In a recent evaluation, MethyLight displayed a high level of precision and reproducibility with an average variation of ~0.8 PMR (0.8%), with slightly larger variations induced by different bisulfite treatments (Ogino *et al.*, 2006). The addition of an additional probe marked with a different fluorescent dye allows for the simultaneous detection of unconverted sequences that might co-amplify with the bisulfite-converted molecules avoiding potential false-positive results (ConLight (Rand *et al.*, 2002)).

In a recent variation, bisulfite-treated DNA was amplified independently of the methylation status in a first round of amplification (Fackler *et al.*, 2004). This reaction can be performed in multiplex amplifying several target regions in a single reaction. In a nested PCR approach, specifically methylated or unmethylated molecules are amplified in two separate reactions using the above-described fluorescent real-time approach. This is in contrast to MethyLight where normally only the methylated molecules are interrogated (Weisenberger *et al.*, 2006). Quantification of the methylation percentage is performed by dividing the intensity of fluorescence corresponding to the methylated primer set to the added intensities of both sets yielding comparable results to the reference gene approach. However, multiplex reactions might be difficult to set up on bisulfite-converted DNA with its low sequence diversity and the equal amplification of methylated and unmethylated molecules (absence of a PCR bias) is a prerequisite for accurate quantification. Furthermore, partially methylated molecules are not taken into account thereby potentially overestimating the methylation content.

14.3.6.3 Quantitative Analysis of Methylated Alleles (QAMA)

Quantitative analysis of methylated alleles QAMA is a variation of MethyLight which uses a TaqMan probe conjugated to a minor groove binder for discrimination at single base level by forming hyperstabilized duplexes with complementary DNAs (Zeschnigk *et al.*, 2004). Methylated and unmethylated alleles are simultaneously quantified using two probes modified with a quencher, a minor groove binder, and one of the two fluorophores VIC or FAM. Thus, amplification of the bisulfite-treated DNA can be carried out with primers amplifying simultaneously formerly methylated and unmethylated alleles and differentiation of the methylation status of alleles is achieved only at the probe level.

14.3.6.4 HeavyMethyl and Headloop PCR

HeavyMethyl further increases sensitivity and specificity of real-time PCR-based assays for the analysis of DNA methylation using methylation-dependent blocking oligonucleotides (Fig. 14.7; Cottrell *et al.*, 2004). A similar approach had previously been successfully used to detect genetic mutations diluted in an excess of normal DNA (Sun *et al.*, 2002). In contrast to MethyLight amplification, primers are not specific for a certain methylation pattern, but positioned in sequence stretches containing no CpG positions. Only the fluorescent probe is specific usually to a consistently hypermethylated sequence. The increased specificity and sensitivity is achieved through a second pair of non-extendable (3′ phosphorylated) oligonucleotides that hybridize specifically to a methylation pattern opposite the investigated one, usually the unmethylated sequence. The annealing sites of these oligonucleotides

overlap with the target sequences for PCR amplification and thereby efficiently block any amplification of the bisulfite sequence corresponding to the undesired methylation pattern. HeavyMethyl was able to specifically detect 25 pg of *in vitro* methylated DNA in the background of 400 ng of unmethylated DNA (relative sensitivity up to 1:8,000). As for MethyLight, a PMR standard is used to identify samples with negligible amounts of methylation. The use of four to five different oligonucleotides contributes significantly to the cost of the assay and design might be more complex compared to conventional MSP or MethyLight. However, HeavyMethyl shows the necessary sensitivity and specificity necessary for clinical applications and might be a useful alternative for situations where other real-time base methylation assays reach their limits.

An alternative but similar approach uses amplification primers specific for a target after bisulfite treatment that carry a 5′-tail sequence complementary to a sequence that is present in the amplicon corresponding to a specific methylation pattern (Fig. 14.7; Rand *et al.*, 2005). After incorporation of the primer in the synthesized PCR product, the tail folds back onto the template creating a secondary structure refractory to amplification. For example, if the tail is complementary to an unmethylated sequence, only methylated molecules are amplified. The amplification is monitored in real time with *SYBR Green* or by TaqMan probes. Sensitivity of the approach is similar to the others described above detecting a methylated allele in the presence of a 4,000-fold excess of unmethylated ones. This approach is therefore well suited for the selective amplification of tumor-derived methylated molecules in body fluids.

14.4 CONCLUSIONS

DNA methylation analysis will undoubtedly play a key role in the diagnosis, prognostic assessment, and treatment of various diseases. In the past few years a multitude of epigenetic markers has been discovered. However, with an estimation of about 10% of all genes epigenetically deregulated in a specific type of cancer, we only have scratched the tip of the iceberg. A large increase in the number of identified epigenetic changes will be seen over the next few years together with first genome-wide mammalian epigenome maps. Technology has been developed at a breathtaking speed with many companies now providing easy-to-implement assays and reproducible protocols for bisulfite treatment. With the technology available, research is shifting from gene-centered study to genome-wide analyses.

One of the most important tasks now is to validate the already known biomarkers in larger cohorts as most studies so far have analyzed tens of samples instead of the hundreds or thousands required to definitely confirm the utility of a biomarker. The discovery of viable markers will crucially depend on the access to correctly characterized and classified biological material. Another task is to translate the knowledge on epigenetic changes and their correlation with pathophysiological parameters into clinical practice. The sensitive and specific detection of tumor-specific DNA methylation patterns at distal sites makes DNA methylation a biomarker of choice for the clinical management of cancer patients and predictive epigenetic biomarkers will allow a personalized treatment of disease based on the individual methylation profile.

REFERENCES

Adorjan, P., Distler, J., Lipscher, E., Model, F., Muller, J., Pelet, C., Braun, A., Florl, A.R., Gutig, D., Grabs, G., Howe, A., Kursar, M., Lesche, R., Leu, E., Lewin, A., Maier, S., Muller, V., Otto, T., Scholz, C., Schulz, W.A., Seifert, H.H., Schwope, I., Ziebarth, H., Berlin, K., Piepenbrock, C., and Olek, A. (2002). Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res.* 30, e21.

Akey, D.T., Akey, J.M., Zhang, K., and Jin, L. (2002). Assaying DNA methylation based on high-throughput melting curve approaches. *Genomics* 80, 376–384.

Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X., Richmond, T.A., Middle, C.M., Rodesch, M.J., Packard, C.J., Weinstock, G.M., and Gibbs, R.A. (2007). Direct selection of human genomic loci by microarray hybridization. *Nat. Methods* 4, 903–905.

Alders, M., Bliek, J., Vd Lip, K., Vd Bogaard, R., and Mannens, M. (2009). Determination of KCNQ1OT1 and H19 methylation levels in BWS and SRS patients using methylation-sensitive high-resolution melting analysis. *Eur. J. Hum. Genet.* 17, 467–473.

Andersen, C.L., Koch, J., and Kjeldsen, E. (1998). CpG islands detected by self-primed in situ labeling (SPRINS). *Chromosoma* 107, 260–266.

Antequera, F. (2003). Structure, function and evolution of CpG island promoters. *Cell. Mol. Life Sci.* 60, 1647–1658.

Balada, E., Ordi-Ros, J., and Vilardell-Tarres, M. (2007). DNA methylation and systemic lupus erythematosus. *Ann. NY Acad. Sci.* 1108, 127–136.

Balch, C., Yan, P., Craft, T., Young, S., Skalnik, D.G., Huang, T.H., and Nephew, K.P. (2005). Antimitogenic and chemosensitizing effects of the methylation inhibitor zebularine in ovarian cancer. *Mol. Cancer Ther.* 4, 1505–1514.

Balic, M., Pichler, M., Strutz, J., Heitzer, E., Ausch, C., Samonigg, H., Cote, R.J., and Dandachi, N. (2009). High quality assessment of DNA methylation in archival tissues from colorectal cancer patients using quantitative high-resolution melting analysis. *J. Mol. Diagn.* 11, 102–108.

Ballestar, E., Paz, M.F., Valle, L., Wei, S., Fraga, M.F., Espada, J., Cigudosa, J.C., Huang, T.H., and Esteller, M. (2003). Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. *EMBO J.* 22, 6335–6345.

Bastian, P.J., Palapattu, G.S., Lin, X., Yegnasubramanian, S., Mangold, L.A., Trock, B., Eisenberger, M.A., Partin, A.W., and Nelson, W.G.

(2005). Preoperative serum DNA GSTP1 CpG island hypermethylation and the risk of early prostate-specific antigen recurrence following radical prostatectomy. *Clin. Cancer Res.* 11, 4037–4043.

Berkhout, M., Nagtegaal, I.D., Cornelissen, S.J., Dekkers, M.M., van de Molengraft, F.J., Peters, W.H., Nagengast, F.M., van Krieken, J.H., and Jeuken, J.W. (2007). Chromosomal and methylation alterations in sporadic and familial adenomatous polyposis-related duodenal carcinomas. *Mod. Pathol.* 20, 1253–1262.

Bestor, T.H., and Ingram, V.M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc. Natl. Acad. Sci. USA* 80, 5559–5563.

Bestor, T.H., Hellewell, S.B., and Ingram, V.M. (1984). Differentiation of two mouse cell lines is associated with hypomethylation of their genomes. *Mol. Cell. Biol.* 4, 1800–1806.

Bibikova, M., Chudin, E., Wu, B., Zhou, L., Garcia, E.W., Liu, Y., Shin, S., Plaia, T.W., Auerbach, J.M., Arking, D.E., Gonzalez, R., Crook, J., Davidson, B., Schulz, T.C., Robins, A., Khanna, A., Sartipy, P., Hyllner, J., Vanguri, P., Savant-Bhonsale, S., Smith, A.K., Chakravarti, A., Maitra, A., Rao, M., Barker, D.L., Loring, J.F., and Fan, J.B. (2006a). Human embryonic stem cells have a unique epigenetic signature. *Genome Res.* 16, 1075–1083.

Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E.W., Wu, B., Doucet, D., Thomas, N.J., Wang, Y., Vollmer, E., Goldmann, T., Seifart, C., Jiang, W., Barker, D.L., Chee, M.S., Floros, J., and Fan, J.B. (2006b). High-throughput DNA methylation profiling using universal bead arrays. *Genome Res.* 16, 383–393.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21.

Bird, A.P., and Southern, E.M. (1978). Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. *J. Mol. Biol.* 118, 27–47.

Bjornsson, H.T., Brown, L.J., Fallin, M.D., Rongione, M.A., Bibikova, M., Wickham, E., Fan, J.B., and Feinberg, A.P. (2007). Epigenetic specificity of loss of imprinting of the IGF2 gene in Wilms tumors. *J. Natl. Cancer Inst.* 99, 1270–1273.

Bollati, V., Baccarelli, A., Hou, L., Bonzini, M., Fustinoni, S., Cavallo, D., Byun, H.M., Jiang, J., Marinelli, B., Pesatori, A.C., Bertazzi, P.A., and Yang, A.S. (2007). Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer Res.* 67, 876–880.

Bönsch, D., Lenz, B., Reulbach, U., Kornhuber, J., and Bleich, S. (2004). Homocysteine associated genomic DNA hypermethylation in patients with chronic alcoholism. *J. Neural Transm.* 111, 1611–1616.

Boyd, S.D. (2008). Everything you wanted to know about small RNA but were afraid to ask. *Lab. Invest* 88, 569–578.

Brena, R.M., Auer, H., Kornacker, K., Hackanson, B., Raval, A., Byrd, J.C., and Plass, C. (2006a). Accurate quantification of DNA methylation using combined bisulfite restriction analysis coupled with the Agilent 2100 Bioanalyzer platform. *Nucleic Acids Res.* 34, e17.

Brena, R.M., Huang, T.H., and Plass, C. (2006b). Quantitative assessment of DNA methylation: potential applications for disease diagnosis, classification, and prognosis in clinical settings. *J. Mol. Med.* 84, 365–377.

Burri, N., and Chaubert, P. (1999). Complex methylation patterns analyzed by single-strand conformation polymorphism. *Biotechniques* 26, 232–234.

Cairns, P., Esteller, M., Herman, J.G., Schoenberg, M., Jeronimo, C., Sanchez-Cespedes, M., Chow, N.H., Grasso, M., Wu, L., Westra, W.B., and Sidransky, D. (2001). Molecular detection of prostate cancer in urine by GSTP1 hypermethylation. *Clin. Cancer Res.* 7, 2727–2730.

Cedar, H., Solage, A., Glaser, G., and Razin, A. (1979). Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. *Nucleic Acids Res.* 6, 2125–2132.

Ching, T.T., Maunakea, A.K., Jun, P., Hong, C., Zardo, G., Pinkel, D., Albertson, D.G., Fridlyand, J., Mao, J.H., Shchors, K., Weiss, W.A., and Costello, J.F. (2005). Epigenome analyses using BAC microarrays identify evolutionary conservation of tissue-specific methylation of SHANK3. *Nat. Genet.* 37, 645–651.

Clark, S.J., Harrison, J., Paul, C.L., and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 22, 2990–2997.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215–219.

Colella, S., Shen, L., Baggerly, K.A., Issa, J.P., and Krahe, R. (2003). Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *Biotechniques* 35, 146–150.

Coolen, M.W., Statham, A.L., Gardiner-Garden, M., and Clark, S.J. (2007). Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. *Nucleic Acids Res.* 35, e119.

Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L., and Myers, R.M. (2006). Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* 16, 1–10.

Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C., Schuller, D.E., Yu, L., Bloomfield, C.D., Caligiuri, M.A., Yates, A., Nishikawa, R., Su Huang, H., Petrelli, N.J., Zhang, X., O'Dorisio, M.S., Held, W.A., Cavenee, W.K., and Plass, C. (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat. Genet.* 24, 132–138.

Costello, J.F., Plass, C., and Cavenee, W.K. (2002). Restriction landmark genome scanning. *Methods Mol. Biol.* 200, 53–70.

Cottrell, S.E., and Laird, P.W. (2003). Sensitive detection of DNA methylation. *Ann. NY Acad. Sci.* 983, 120–130.

Cottrell, S.E., Distler, J., Goodman, N.S., Mooney, S.H., Kluth, A., Olek, A., Schwope, I., Tetzner, R., Ziebarth, H., and Berlin, K. (2004). A real-time PCR assay for DNA-methylation using methylation-specific blockers. *Nucleic Acids Res.* 32, e10.

Deng, D., Deng, G., Smith, M.F., Zhou, J., Xin, H., Powell, S.M., and Lu, Y. (2002). Simultaneous detection of CpG methylation and single nucleotide polymorphism by denaturing high performance liquid chromatography. *Nucleic Acids Res.* 30, e13.

Dikow, N., Nygren, A.O., Schouten, J.P., Hartmann, C., Kramer, N., Janssen, B., and Zschocke, J. (2007). Quantification of the methylation status of the PWS/AS imprinted region: comparison of two approaches based on bisulfite sequencing and methylation-sensitive MLPA. *Mol. Cell Probes* 21, 208–215.

Down, T.A., Rakyan, V.K., Turner, D.J., Flicek, P., Li, H., Kulesha, E., Graf, S., Johnson, N., Herrero, J., Tomazou, E.M., Thorne, N.P., Backdahl, L., Herberth, M., Howe, K.L., Jackson, D.K., Miretti, M.M., Marioni, J.C., Birney, E., Hubbard, T.J., Durbin, R., Tavare, S., and Beck, S. (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.* 26, 779–785.

Dulaimi, E., Hillinck, J., Ibanez de Caceres, I., Al-Saleem, T., and Cairns, P. (2004). Tumor suppressor gene promoter hypermethylation in serum of breast cancer patients. *Clin. Cancer Res.* 10, 6189–6193.

Dupont, J.M., Tost, J., Jammes, H., and Gut, I.G. (2004). De novo quantitative bisulfite sequencing using the pyrosequencing technology. *Anal. Biochem.* 333, 119–127.

Eads, C.A., Danenberg, K.D., Kawakami, K., Saltz, L.B., Blake, C., Shibata, D., Danenberg, P.V., and Laird, P.W. (2000a). MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res.* 28, e32.

Eads, C.A., Lord, R.V., Kurumboor, S.K., Wickramasinghe, K., Skinner, M.L., Long, T.I., Peters, J.H., DeMeester, T.R., Danenberg, K.D., Danenberg, P.V., Laird, P.W., and Skinner, K.A. (2000b). Fields of aberrant CpG island hypermethylation in Barrett's esophagus and associated adenocarcinoma. *Cancer Res.* 60, 5021–5026.

Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C.S., Bowman, R., Meyer, K.B., Haiman, C.A., Kolonel, L.K., Henderson, B.E., Le Marchand, L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C.Y., Wu, P.E., Wang, H.C., Eccles, D., Evans, D.G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M.R., Rahman, N., Chenevix-Trench, G., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska, B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K.Y., Noh, D.Y., Ahn, S.H., Hunter, D.J., Hankinson, S.E., Cox, D.G., Hall, P., Wedren, S., Liu, J., Low, Y.L., Bogdanova, N., Schurmann, P., Dork, T., Tollenaar, R.A., Jacobi, C.E., Devilee, P., Klijn, J.G., Sigurdson, A.J., Doody, M.M., Alexander, B.H., Zhang, J., Cox, A., Brock, I.W., MacPherson, G., Reed, M.W., Couch, F.J., Goode, E.L., Olson, J.E., Meijers-Heijboer, H., van den Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R.L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J.L., McCredie, M., Southey, M., Giles, G.G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y.D., Spurdle, A.B., Beesley, J., Chen, X., Mannermaa, A., Kosma, V.M., Kataja, V., Hartikainen, J., Day, N.E., Cox, D.R., and Ponder, B.A. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447, 1087–1093.

Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* 300, 455.

Ehrich, M., Nelson, M.R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G., Cantor, C.R., Field, J.K., and van den Boom, D. (2005). Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci. USA* 102, 15785–15790.

Ehrich, M., Turner, J., Gibbs, P., Lipton, L., Giovanneti, M., Cantor, C., and van den Boom, D. (2008). Cytosine methylation profiling of cancer cell lines. *Proc. Natl. Acad. Sci. USA* 105, 4844–4849.

Ehrlich, M. (2002). DNA methylation in cancer: too much, but also too little. *Oncogene* 21, 5400–5413.

Eick, D., Fritz, H.J., and Doerfler, W. (1983). Quantitative determination of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography. *Anal. Biochem.* 135, 165–171.

El-Maarri, O., Herbiniaux, U., Walter, J., and Oldenburg, J. (2002). A rapid, quantitative, non-radioactive bisulfite-SNuPE- IP RP HPLC assay for methylation analysis at specific CpG sites. *Nucleic Acids Res.* 30, e25.

Esteller, M., Sanchez-Cespedes, M., Rosell, R., Sidransky, D., Baylin, S.B., and Herman, J.G. (1999). Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. *Cancer Res.* 59, 67–70.

Fackler, M.J., McVeigh, M., Mehrotra, J., Blum, M.A., Lange, J., Lapides, A., Garrett, E., Argani, P., and Sukumar, S. (2004). Quantitative multiplex methylation-specific PCR assay for the detection of promoter hypermethylation in multiple genes in breast cancer. *Cancer Res.* 64, 4442–4452.

Fazzari, M.J., and Greally, J.M. (2004). Epigenomics: beyond CpG islands. *Nat. Rev. Genet.* 5, 446–455.

Feil, R. (2006). Environmental and nutritional effects on the epigenetic regulation of genes. *Mutat. Res.* 600, 46–57.

Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301, 89–92.

Fraga, M.F., and Esteller, M. (2002). DNA methylation: a profile of methods and applications. *Biotechniques* 33, 632–649.

Fraga, M.F., Uriol, E., Borja Diego, L., Berdasco, M., Esteller, M., Canal, M.J., and Rodriguez, R. (2002). High-performance capillary electrophoretic method for the quantification of 5-methyl 2′-deoxycytidine in genomic DNA: application to plant, animal and human cancer tissues. *Electrophoresis* 23, 1677–1681.

Fraga, M.F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P.A., and Esteller, M. (2003). The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res.* 31, 1765–1774.

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suner, D., Cigudosa, J.C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T.D., Wu, Y.Z., Plass, C., and Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl. Acad. Sci. USA* 102, 10604–10609.

Frigola, J., Ribas, M., Risques, R.A., and Peinado, M.A. (2002). Methylome profiling of cancer cells by amplification of inter-methylated sites (AIMS). *Nucleic Acids Res.* 30, e28.

Frigola, J., Song, J., Stirzaker, C., Hinshelwood, R.A., Peinado, M.A., and Clark, S.J. (2006). Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nat. Genet.* 38, 540–549.

Friso, S., Choi, S.W., Dolnikowski, G.G., and Selhub, J. (2002). A method to assess genomic DNA methylation using high-performance liquid chromatography/electrospray ionization mass spectrometry. *Anal. Chem.* 74, 4526–4531.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* 89, 1827–1831.

Fujiwara, K., Fujimoto, N., Tabata, M., Nishii, K., Matsuo, K., Hotta, K., Kozuki, T., Aoe, M., Kiura, K., Ueoka, H., and Tanimoto, M. (2005). Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer. *Clin. Cancer Res.* 11, 1219–1225.

Galm, O., Rountree, M.R., Bachman, K.E., Jair, K.W., Baylin, S.B., and Herman, J.G. (2002). Enzymatic regional methylation assay: a novel method to quantify regional CpG methylation density. *Genome Res.* 12, 153–157.

Geiman, T.M., and Robertson, K.D. (2002). Chromatin remodeling, histone modifications, and DNA methylation – how does it all fit together?. *J. Cell. Biochem.* 87, 117–125.

Gitan, R.S., Shi, H., Chen, C.M., Yan, P.S., and Huang, T.H. (2002). Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res.* 12, 158–164.

Goelz, S.E., Vogelstein, B., Hamilton, S.R., and Feinberg, A.P. (1985). Hypomethylation of DNA from benign and malignant human colon neoplasms. *Science* 228, 187–190.

Goessl, C., Krause, H., Muller, M., Heicappell, R., Schrader, M., Sachsinger, J., and Miller, K. (2000). Fluorescent methylation-specific

polymerase chain reaction for DNA-based detection of prostate cancer in bodily fluids. *Cancer Res.* 60, 5941–5945.

Gonzalgo, M.L., and Jones, P.A. (1997). Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.* 25, 2529–2531.

Gonzalgo, M.L., Liang, G., Spruck, C.H., 3rd, Zingg, J.M., Rideout, W.M., 3rd, and Jones, P.A. (1997). Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR. *Cancer Res.* 57, 594–599.

Gonzalgo, M.L., and Jones, P.A. (2002). Quantitative methylation analysis using methylation-sensitive single-nucleotide primer extension (Ms-SNuPE). *Methods* 27, 128–133.

Gray, M.A. (2005). Clinical use of serum prostate-specific antigen: a review. *Clin. Lab.* 51, 127–133.

Grunau, C., Clark, S.J., and Rosenthal, A. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.* 29, e65.

Hatada, I., Fukasawa, M., Kimura, M., Morita, S., Yamada, K., Yoshikawa, T., Yamanaka, S., Endo, C., Sakurada, A., Sato, M., Kondo, T., Horii, A., Ushijima, T., and Sasaki, H. (2006). Genome-wide profiling of promoter methylation in human. *Oncogene* 25, 3059–3064.

Hayashi, H., Nagae, G., Tsutsumi, S., Kaneshiro, K., Kozaki, T., Kaneda, A., Sugisaki, H., and Aburatani, H. (2007). High-resolution mapping of DNA methylation in human genome using oligonucleotide tiling array. *Hum. Genet.* 120, 701–711.

Heard, E., and Disteche, C.M. (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.* 20, 1848–1867.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O.P., Wilfond, B., Borg, A., Trent, J., Raffeld, M., Yakhini, Z., Ben-Dor, A., Dougherty, E., Kononen, J., Bubendorf, L., Fehrle, W., Pittaluga, S., Gruvberger, S., Loman, N., Johannsson, O., Olsson, H., and Sauter, G. (2001). Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* 344, 539–548.

Heiskanen, M., Syvanen, A.C., Siitari, H., Laine, S., and Palotie, A. (1994). A novel method to quantitate methylation of specific genomic regions. *PCR Methods Appl.* 4, 26–30.

Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D., and Baylin, S.B. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. USA* 93, 9821–9826.

Holland, P.M., Abramson, R.D., Watson, R., and Gelfand, D.H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of Thermus aquaticus DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88, 7276–7280.

Huang, T.H., Laux, D.E., Hamlin, B.C., Tran, P., Tran, H., and Lubahn, D.B. (1997). Identification of DNA methylation markers for human breast carcinomas using the methylation-sensitive restriction fingerprinting technique. *Cancer Res.* 57, 1030–1034.

Huang, T.H., Perry, M.R., and Laux, D.E. (1999). Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet.* 8, 459–470.

Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., Bencko, V., Foretova, L., Janout, V., Chen, C., Goodman, G., Field, J.K., Liloglou, T., Xinarianos, G., Cassidy, A., McLaughlin, J., Liu, G., Narod, S., Krokan, H.E., Skorpen, F., Elvestad, M.B., Hveem, K., Vatten, L., Linseisen, J., Clavel-Chapelon, F., Vineis, P., Bueno-de-Mesquita, H.B., Lund, E., Martinez, C., Bingham, S., Rasmuson, T., Hainaut, P., Riboli, E., Ahrens, W., Benhamou, S., Lagiou, P., Trichopoulos, D., Holcatova, I., Merletti, F., Kjaerheim, K., Agudo, A., Macfarlane, G., Talamini, R., Simonato, L., Lowry, R., Conway, D.I., Znaor, A., Healy, C., Zelenika, D., Boland, A., Delepine, M., Foglio, M., Lechner, D., Matsuda, F., Blanche, H., Gut, I., Heath, S., Lathrop, M., and Brennan, P. (2008). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 452, 633–637.

Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddeloh, J.A., Wen, B., and Feinberg, A.P. (2008). Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res.* 18, 780–790.

Issa, J.P. (2004). CpG island methylator phenotype in cancer. *Nat. Rev. Cancer* 4, 988–993.

Jahr, S., Hentze, H., Englisch, S., Hardt, D., Fackelmayer, F.O., Hesch, R.D., and Knippers, R. (2001). DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* 61, 1659–1665.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. *Cell* 128, 683–692.

Junien, C., and Nathanielsz, P. (2007). Report on the IASO Stock Conference 2006: early and lifelong environmental epigenomic programming of metabolic syndrome, obesity and type II diabetes. *Obes. Rev.* 8, 487–502.

Kaminsky, Z.A., Assadzadeh, A., Flanagan, J., and Petronis, A. (2005). Single nucleotide extension technology for quantitative site-specific evaluation of metC/C in GC-rich regions. *Nucleic Acids Res.* 33, e95.

Karimi, M., Johansson, S., Stach, D., Corcoran, M., Grander, D., Schalling, M., Bakalkin, G., Lyko, F., Larsson, C., and Ekstrom, T.J. (2006). LUMA (LUminometric Methylation Assay) – a high throughput method to the analysis of genomic DNA methylation. *Exp. Cell Res.* 312, 1989–1995.

Kawakami, T., Okamoto, K., Ogawa, O., and Okada, Y. (2004). XIST unmethylated DNA fragments in male-derived plasma as a tumour marker for testicular cancer. *Lancet* 363, 40–42.

Kerkel, K., Spadola, A., Yuan, E., Kosek, J., Jiang, L., Hod, E., Li, K., Murty, V.V., Schupf, N., Vilain, E., Morris, M., Haghighi, F., and Tycko, B. (2008). Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat. Genet.* 40, 904–908.

Khulan, B., Thompson, R.F., Ye, K., Fazzari, M.J., Suzuki, M., Stasiek, E., Figueroa, M.E., Glass, J.L., Chen, Q., Montagna, C., Hatchwell, E., Selzer, R.R., Richmond, T.A., Green, R.D., Melnick, A., and Greally, J.M. (2006). Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res.* 16, 1046–1055.

Korshunova, Y., Maloney, R.K., Lakey, N., Citek, R.W., Bacher, B., Budiman, A., Ordway, J.M., McCombie, W.R., Leon, J., Jeddeloh, J.A., and McPherson, J.D. (2008). Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA. *Genome Res.* 18, 19–29.

Kuo, K.C., McCune, R.A., Gehrke, C.W., Midgett, R., and Ehrlich, M. (1980). Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. *Nucleic Acids Res.* 8, 4763–4776.

Ladd-Acosta, C., Pevsner, J., Sabunciyan, S., Yolken, R.H., Webster, M.J., Dinkins, T., Callinan, P.A., Fan, J.B., Potash, J.B., and Feinberg, A.P. (2007). DNA methylation signatures within the human brain. *Am. J. Hum. Genet.* 81, 1304–1315.

Laird, P.W. (2003). Early detection: the power and the promise of DNA methylation markers. *Nat. Rev. Cancer* 3, 253–266.

Laird, P.W. (2005). Cancer epigenetics. *Hum. Mol. Genet.* 14(Spec No. 1), R65–R76.

Lewin, J., Schmitt, A.O., Adorjan, P., Hildmann, T., and Piepenbrock, C. (2004). Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplificates. *Bioinformatics* 20, 3005–3012.

Lippman, Z., Gendrel, A.V., Colot, V., and Martienssen, R. (2005). Profiling DNA methylation patterns using genomic tiling microarrays. *Nat. Methods* 2, 219–224.

Lisitsyn, N., Lisitsyn, N., and Wigler, M. (1993). Cloning the differences between two complex genomes. *Science* 259, 946–951.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133, 523–536.

Liu, Z., Liu, S., Xie, Z., Blum, W., Perrotti, D., Paschka, P., Klisovic, R., Byrd, J., Chan, K.K., and Marcucci, G. (2007). Characterization of in vitro and in vivo hypomethylating effects of decitabine in acute myeloid leukemia by a rapid, specific and sensitive LC-MS/MS method. *Nucleic Acids Res.* 35, e31.

Lo, Y.M., Wong, I.H., Zhang, J., Tein, M.S., Ng, M.H., and Hjelm, N.M. (1999). Quantitative analysis of aberrant p16 methylation using real-time quantitative methylation-specific polymerase chain reaction. *Cancer Res.* 59, 3899–3903.

Luo, S., and Preuss, D. (2003). Strand-biased DNA methylation associated with centromeric regions in Arabidopsis. *Proc. Natl. Acad. Sci. USA* 100, 11133–11138.

Maier, S., Dahlstroem, C., Haefliger, C., Plum, A., and Piepenbrock, C. (2005). Identifying DNA methylation biomarkers of cancer drug response. *Am. J. Pharmacogenomics* 5, 223–232.

McGrew, M.J., and Rosenthal, N. (1993). Quantitation of genomic methylation using ligation-mediated PCR. *Biotechniques* 15, 722–729.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., Gnirke, A., Jaenisch, R., and Lander, E.S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766–770.

Melnikov, A.A., Gartenhaus, R.B., Levenson, A.S., Motchoulskaia, N.A., and Levenson Chernokhvostov, V.V. (2005). MSRE-PCR for analysis of gene-specific DNA methylation. *Nucleic Acids Res.* 33, e93.

Miller, O.J., Schnedl, W., Allen, J., and Erlanger, B.F. (1974). 5-Methylcytosine localised in mammalian constitutive heterochromatin. *Nature* 251, 636–637.

Mockler, T.C., Chan, S., Sundaresan, A., Chen, H., Jacobsen, S.E., and Ecker, J.R. (2005). Applications of DNA tiling arrays for whole-genome analysis. *Genomics* 85, 1–15.

Mund, C., Hackanson, B., Stresemann, C., Lubbert, M., and Lyko, F. (2005). Characterization of DNA demethylation effects induced by 5-Aza-2′-deoxycytidine in patients with myelodysplastic syndrome. *Cancer Res.* 65, 7086–7090.

Nakayama, M., Gonzalgo, M.L., Yegnasubramanian, S., Lin, X., De Marzo, A.M., and Nelson, W.G. (2004). GSTP1 CpG island hypermethylation as a molecular biomarker for prostate cancer. *J. Cell. Biochem.* 91, 540–552.

Neidhart, M., Rethage, J., Kuchen, S., Kunzler, P., Crowl, R.M., Billingham, M.E., Gay, R.E., and Gay, S. (2000). Retrotransposable L1 elements expressed in rheumatoid arthritis synovial tissue: association with genomic DNA hypomethylation and influence on gene expression. *Arthritis Rheum.* 43, 2634–2647.

Niemann, H., Tian, X.C., King, W.A., and Lee, R.S. (2008). Epigenetic reprogramming in embryonic and foetal development upon somatic cell nuclear transfer cloning: focus on mammalian embryogenomics. *Reproduction* 135, 151–163.

Noda, H., Miyaji, Y., Nakanishi, A., Konishi, F., and Miki, Y. (2007). Frequent reduced expression of alpha-1B-adrenergic receptor caused by aberrant promoter methylation in gastric cancers. *Br. J. Cancer* 96, 383–390.

Nygren, A.O., Ameziane, N., Duarte, H.M., Vijzelaar, R.N., Waisfisz, Q., Hess, C.J., Schouten, J.P., and Errami, A. (2005). Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. *Nucleic Acids Res.* 33, e128.

Oakeley, E.J., Schmitt, F., and Jost, J.P. (1999). Quantification of 5-methylcytosine in DNA by the chloroacetaldehyde reaction. *Biotechniques* 27, 744–746, 748–750, 752.

Oakes, C.C., La Salle, S., Robaire, B., and Trasler, J.M. (2006). Evaluation of a quantitative DNA methylation analysis technique using methylation-sensitive/dependent restriction enzymes and real-time PCR. *Epigenetics* 1, 146–152.

Ogino, S., Kawasaki, T., Brahmandam, M., Cantor, M., Kirkner, G.J., Spiegelman, D., Makrigiorgos, G.M., Weisenberger, D.J., Laird, P.W., Loda, M., and Fuchs, C.S. (2006). Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. *J. Mol. Diagn.* 8, 209–217.

Ordway, J.M., Bedell, J.A., Citek, R.W., Nunberg, A., Garrido, A., Kendall, R., Stevens, J.R., Cao, D., Doerge, R.W., Korshunova, Y., Holemon, H., McPherson, J.D., Lakey, N., Leon, J., Martienssen, R. A., and Jeddeloh, J.A. (2006). Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets. *Carcinogenesis* 27, 2409–2423.

Pelizzola, M., Koga, Y., Urban, A.E., Krauthammer, M., Weissman, S., Halaban, R., and Molinaro, A.M. (2008). MEDME: an experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment. *Genome Res.* 18, 1652–1659.

Pogribny, I., Yi, P., and James, S.J. (1999). A sensitive new method for rapid detection of abnormal methylation patterns in global DNA and within CpG islands. *Biochem. Biophys. Res. Commun.* 262, 624–628.

Priolo, M., Sparago, A., Mammi, C., Cerrato, F., Lagana, C., and Riccio, A. (2008). MS-MLPA is a specific and sensitive technique for detecting all chromosome 11p15.5 imprinting defects of BWS and SRS in a single-tube experiment. *Eur. J. Hum. Genet.* 16, 565–571.

Pröll, J., Födermayr, M., Wechselberger, C., Pammer, P., Sonnleitner, M., Zach, O., and Lutz, D. (2006). Ultra-sensitive immunodetection of 5′methyl cytosine for DNA methylation analysis on oligonucleotide microarrays. *DNA Res.* 13, 37–42.

Ragoussis, J., Elvidge, G.P., Kaur, K., and Colella, S. (2006). Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. *PLoS Genet.* 2, e100.

Rakyan, V.K., Hildmann, T., Novik, K.L., Lewin, J., Tost, J., Cox, A.V., Andrews, T.D., Howe, K.L., Otto, T., Olek, A., Fischer, J., Gut, I.G., Berlin, K., and Beck, S. (2004). DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.* 2, e405.

Rakyan, V.K., Down, T.A., Thorne, N.P., Flicek, P., Kulesha, E., Graf, S., Tomazou, E.M., Backdahl, L., Johnson, N., Herberth, M., Howe, K.L., Jackson, D.K., Miretti, M.M., Fiegler, H., Marioni, J.C., Birney, E., Hubbard, T.J., Carter, N.P., Tavare, S., and Beck, S. (2008). An integrated resource for genome-wide identification and analysis of

human tissue-specific differentially methylated regions (tDMRs). *Genome Res.* 18, 1518–1529.

Ramsahoye, B.H. (2002). Measurement of genome-wide DNA cytosine-5 methylation by reversed-phase high-pressure liquid chromatography. *Methods Mol. Biol.* 200, 17–27.

Rand, K., Qu, W., Ho, T., Clark, S.J., and Molloy, P. (2002). Conversion-specific detection of DNA methylation using real-time polymerase chain reaction (ConLight-MSP) to avoid false positives. *Methods* 27, 114–120.

Rand, K.N., Ho, T., Qu, W., Mitchell, S.M., White, R., Clark, S.J., and Molloy, P.L. (2005). Headloop suppression PCR and its application to selective amplification of methylated DNA sequences. *Nucleic Acids Res.* 33, e127.

Rauch, T., and Pfeifer, G.P. (2005). Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Lab. Invest* 85, 1172–1180.

Rauch, T., Li, H., Wu, X., and Pfeifer, G.P. (2006). MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Res.* 66, 7939–7947.

Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* 2, 21–32.

Reynaud, C., Bruno, C., Boullanger, P., Grange, J., Barbesti, S., and Niveleau, A. (1992). Monitoring of urinary excretion of modified nucleosides in cancer patients using a set of six monoclonal antibodies. *Cancer Lett.* 61, 255–262.

Richardson, B. (2003). Impact of aging on DNA methylation. *Ageing Res. Rev.* 2, 245–261.

Robertson, K.D. (2005). DNA methylation and human disease. *Nat. Rev. Genet.* 6, 597–610.

Rollins, R.A., Haghighi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J., and Bestor, T.H. (2006). Large-scale structure of genomic methylation patterns. *Genome Res.* 16, 157–163.

Ronaghi, M., Uhlen, M., and Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* 281, 363, 365.

Rosenfeld, N., Aharonov, R., Meiri, E., Rosenwald, S., Spector, Y., Zepeniuk, M., Benjamin, H., Shabes, N., Tabak, S., Levy, A., Lebanony, D., Goren, Y., Silberschein, E., Targan, N., Ben-Ari, A., Gilad, S., Sion-Vardy, N., Tobar, A., Feinmesser, M., Kharenko, O., Nativ, O., Nass, D., Perelman, M., Yosepovich, A., Shalmon, B., Polak-Charcon, S., Fridman, E., Avniel, A., Bentwich, I., Bentwich, Z., Cohen, D., Chajut, A., and Barshack, I. (2008). MicroRNAs accurately identify cancer tissue origin. *Nat. Biotechnol.* 26, 462–469.

Rougier, N., Bourc'his, D., Gomes, D.M., Niveleau, A., Plachot, M., Paldi, A., and Viegas-Pequignot, E. (1998). Chromosome methylation patterns during mammalian preimplantation development. *Genes Dev.* 12, 2108–2113.

Rush, L.J., and Plass, C. (2002). Restriction landmark genomic scanning for DNA methylation in cancer: past, present, and future applications. *Anal. Biochem.* 307, 191–201.

Sadri, R., and Hornsby, P.J. (1996). Rapid analysis of DNA methylation using new restriction enzyme sites created by bisulfite modification. *Nucleic Acids Res.* 24, 5058–5059.

Schatz, P., Dietrich, D., and Schuster, M. (2004). Rapid analysis of CpG methylation patterns using RNase T1 cleavage and MALDI-TOF. *Nucleic Acids Res.* 32, e167.

Schones, D.E., and Zhao, K. (2008). Genome-wide approaches to studying chromatin modifications. *Nat. Rev. Genet.* 9, 179–191.

Schumacher, A., Kapranov, P., Kaminsky, Z., Flanagan, J., Assadzadeh, A., Yau, P., Virtanen, C., Winegarden, N., Cheng, J., Gingeras, T., and Petronis, A. (2006). Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res.* 34, 528–542.

Scott, R.H., Douglas, J., Baskcomb, L., Nygren, A.O., Birch, J.M., Cole, T.R., Cormier-Daire, V., Eastwood, D.M., Garcia-Minaur, S., Lupunzina, P., Tatton-Brown, K., Bliek, J., Maher, E.R., and Rahman, N. (2008). Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) robustly detects and distinguishes 11p15 abnormalities associated with overgrowth and growth retardation. *J. Med. Genet.* 45, 106–113.

Selker, E.U., Tountas, N.A., Cross, S.H., Margolin, B.S., Murphy, J.G., Bird, A.P., and Freitag, M. (2003). The methylated component of the Neurospora crassa genome. *Nature* 422, 893–897.

Shapiro, R., DiFate, V., and Welcher, M. (1974). Deamination of cytosine derivatives by bisulfite. Mechanism of the reaction. *J. Am. Chem. Soc.* 96, 906–912.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A., and Issa, J.P. (2007). Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet.* 3, 2023–2036.

Shi, H., Yan, P.S., Chen, C.M., Rahmatpanah, F., Lofton-Day, C., Caldwell, C.W., and Huang, T.H. (2002). Expressed CpG island sequence tag microarray for dual screening of DNA hypermethylation and gene silencing in cancer cells. *Cancer Res.* 62, 3214–3220.

Shi, H., Maier, S., Nimmrich, I., Yan, P.S., Caldwell, C.W., Olek, A., and Huang, T.H. (2003a). Oligonucleotide-based microarray for DNA methylation analysis: principles and applications. *J. Cell. Biochem.* 88, 138–143.

Shi, H., Wei, S.H., Leu, Y.W., Rahmatpanah, F., Liu, J.C., Yan, P.S., Nephew, K.P., and Huang, T.H. (2003b). Triple analysis of the cancer epigenome: an integrated microarray system for assessing gene expression, DNA methylation, and histone acetylation. *Cancer Res.* 63, 2164–2171.

Silva, J.M., Dominguez, G., Garcia, J.M., Gonzalez, R., Villanueva, M.J., Navarro, F., Provencio, M., San Martin, S., Espana, P., and Bonilla, F. (1999). Presence of tumor DNA in plasma of breast cancer patients: clinicopathological correlations. *Cancer Res.* 59, 3251–3256.

Singer-Sam, J., LeBon, J.M., Tanguay, R.L., and Riggs, A.D. (1990). A quantitative HpaII-PCR assay to measure methylation of DNA from a small number of cells. *Nucleic Acids Res.* 18, 687.

Smiraglia, D.J., and Plass, C. (2002). The study of aberrant methylation in cancer via restriction landmark genomic scanning. *Oncogene* 21, 5414–5426.

Stach, D., Schmitz, O.J., Stilgenbauer, S., Benner, A., Dohner, H., Wiessler, M., and Lyko, F. (2003). Capillary electrophoretic analysis of genomic DNA methylation levels. *Nucleic Acids Res.* 31, E2.

Stewart, F.J., Panne, D., Bickle, T.A., and Raleigh, E.A. (2000). Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol. Biol.* 298, 611–622.

Stransky, N., Vallot, C., Reyal, F., Bernard-Pierrot, I., de Medina, S.G., Segraves, R., de Rycke, Y., Elvin, P., Cassidy, A., Spraggon, C., Graham, A., Southgate, J., Asselain, B., Allory, Y., Abbou, C.C., Albertson, D.G., Thiery, J.P., Chopin, D.K., Pinkel, D., and Radvanyi, F. (2006). Regional copy number-independent deregulation of transcription in cancer. *Nat. Genet.* 38, 1386–1396.

Stresemann, C., Brueckner, B., Musch, T., Stopper, H., and Lyko, F. (2006). Functional diversity of DNA methyltransferase inhibitors in human cancer cell lines. *Cancer Res.* 66, 2794–2800.

Sun, X., Hung, K., Wu, L., Sidransky, D., and Guo, B. (2002). Detection of tumor mutations in the presence of excess amounts of normal DNA. *Nat. Biotechnol.* 20, 186–189.

Suzuki, H., Gabrielson, E., Chen, W., Anbazhagan, R., van Engeland, M., Weijenberg, M.P., Herman, J.G., and Baylin, S.B. (2002). A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. *Nat. Genet.* 31, 141–149.

Tang, W.Y., and Ho, S.M. (2007). Epigenetic reprogramming and imprinting in origins of disease. *Rev. Endocr. Metab. Disord.* 8, 173–182.

Taylor, K.H., Kramer, R.S., Davis, J.W., Guo, J., Duff, D.J., Xu, D., Caldwell, C.W., and Shi, H. (2007). Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. *Cancer Res.* 67, 8511–8518.

Thomassin, H., Kress, C., and Grange, T. (2004). MethylQuant: a sensitive method for quantifying methylation of specific cytosines within the genome. *Nucleic Acids Res.* 32, e168.

Tost, J., and Gut, I.G. (2002). Genotyping single nucleotide polymorphisms by mass spectrometry. *Mass Spectrom. Rev.* 21, 388–418.

Tost, J., Dunker, J., and Gut, I.G. (2003a). Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques* 35, 152–156.

Tost, J., Schatz, P., Schuster, M., Berlin, K., and Gut, I.G. (2003b). Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. *Nucleic Acids Res.* 31, e50.

Tost, J., El Abdalaoui, H., and Gut, I.G. (2006). Serial pyrosequencing for quantitative DNA methylation analysis. *Biotechniques* 40, 721–726.

Tost, J., and Gut, I.G. (2006). DNA analysis by mass spectrometry – past, present and future. *J. Mass Spectrom.* 41, 981–995.

Tost, J., and Gut, I.G. (2007). DNA methylation analysis by pyrosequencing. *Nat. Protoc.* 2, 2265–2275.

Tost, J. (2008). *Epigenetics*. Horizon Scientific Press, Norwich.

Toyota, M., Ho, C., Ahuja, N., Jair, K.W., Li, Q., Ohe-Toyota, M., Baylin, S.B., and Issa, J.P. (1999). Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res.* 59, 2307–2312.

Tryndyak, V., Kovalchuk, O., and Pogribny, I.P. (2006). Identification of differentially methylated sites within unmethylated DNA domains in normal and cancer cells. *Anal. Biochem.* 356, 202–207.

Tufarelli, C., Stanley, J.A., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W. G., and Higgs, D.R. (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.* 34, 157–165.

Uhlmann, K., Brinckmann, A., Toliat, M.R., Ritter, H., and Nürnberg, P. (2002). Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* 23, 4072–4079.

Ushijima, T., Morimura, K., Hosoya, Y., Okonogi, H., Tatematsu, M., Sugimura, T., and Nagao, M. (1997). Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors. *Proc. Natl. Acad. Sci. USA* 94, 2284–2289.

Ushijima, T. (2005). Detection and interpretation of altered methylation patterns in cancer cells. *Nat. Rev. Cancer* 5, 223–231.

van Vliet, J., Oates, N.A., and Whitelaw, E. (2007). Epigenetic mechanisms in the context of complex diseases. *Cell. Mol. Life Sci.* 64, 1531–1538.

Waddington, C.H. (1942). The epigenotype. *Endeavour* 1, 18–20.

Wang, R.Y., Gehrke, C.W., and Ehrlich, M. (1980). Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res.* 8, 4777–4790.

Warnecke, P.M., Stirzaker, C., Song, J., Grunau, C., Melki, J.R., and Clark, S.J. (2002). Identification and resolution of artifacts in bisulfite sequencing. *Methods* 27, 101–107.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* 37, 853–862.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* 39, 457–466.

Wei, S.H., Chen, C.M., Strathdee, G., Harnsomburana, J., Shyu, C.R., Rahmatpanah, F., Shi, H., Ng, S.W., Yan, P.S., Nephew, K.P., Brown, R., and Huang, T.H. (2002). Methylation microarray analysis of late-stage ovarian carcinomas distinguishes progression-free survival in patients and identifies candidate epigenetic markers. *Clin. Cancer Res.* 8, 2246–2252.

Wei, S.H., Balch, C., Paik, H.H., Kim, Y.S., Baldwin, R.L., Liyanarachchi, S., Li, L., Wang, Z., Wan, J.C., Davuluri, R.V., Karlan, B.Y., Gifford, G., Brown, R., Kim, S., Huang, T.H., and Nephew, K.P. (2006). Prognostic DNA methylation biomarkers in ovarian cancer. *Clin. Cancer Res.* 12, 2788–2794.

Weisenberger, D.J., Siegmund, K.D., Campan, M., Young, J., Long, T.I., Faasse, M.A., Kang, G.H., Widschwendter, M., Weener, D., Buchanan, D., Koh, H., Simms, L., Barker, M., Leggett, B., Levine, J., Kim, M., French, A.J., Thibodeau, S.N., Jass, J., Haile, R., and Laird, P.W. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat. Genet.* 38, 787–793.

White, H.E., Durston, V.J., Harvey, J.F., and Cross, N.C. (2006). Quantitative analysis of SNRPN gene methylation by pyrosequencing as a diagnostic test for Prader-Willi syndrome and Angelman syndrome. *Clin. Chem.* 52, 1005–1013.

White, H.E., Hall, V.J., and Cross, N.C. (2007). Methylation-sensitive high-resolution melting-curve analysis of the SNRPN gene as a diagnostic screen for Prader-Willi and Angelman syndromes. *Clin. Chem.* 53, 1960–1962.

Widschwendter, M., Siegmund, K.D., Muller, H.M., Fiegl, H., Marth, C., Muller-Holzner, E., Jones, P.A., and Laird, P.W. (2004). Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. *Cancer Res.* 64, 3807–3813.

Wiebers, J.L. (1976). Detection and identification of minor nucleotides in intact deoxyribonucleic acids by mass spectrometry. *Nucleic Acids Res.* 3, 2959–2970.

Wojdacz, T.K., and Dobrovic, A. (2007). Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Res.* 35, e41.

Wojdacz, T.K., Dobrovic, A., and Algar, E.M. (2008). Rapid detection of methylation change at H19 in human imprinting disorders using methylation-sensitive high-resolution melting. *Hum. Mutat.* 29, 1255–1260.

Wong, I.H., Lo, Y.M., Zhang, J., Liew, C.T., Ng, M.H., Wong, N., Lai, P.B., Lau, W.Y., Hjelm, N.M., and Johnson, P.J. (1999). Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. *Cancer Res.* 59, 71–73.

Worm, J., Aggerholm, A., and Guldberg, P. (2001). In-tube DNA methylation profiling by fluorescence melting curve analysis. *Clin. Chem.* 47, 1183–1189.

Wu, J., Issa, J.P., Herman, J., Bassett, D.E., Jr., Nelkin, B.D., and Baylin, S.B. (1993). Expression of an exogenous eukaryotic DNA methyltransferase gene induces transformation of NIH 3T3 cells. *Proc. Natl. Acad. Sci. USA* 90, 8891–8895.

Xiong, Z., and Laird, P.W. (1997). COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* 25, 2532–2534.

Yamada, Y., Watanabe, H., Miura, F., Soejima, H., Uchiyama, M., Iwasaka, T., Mukai, T., Sakaki, Y., and Ito, T. (2004). A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. *Genome Res.* 14, 247–266.

Yamamoto, F., and Yamamoto, M. (2004). A DNA microarray-based methylation-sensitive (MS)-AFLP hybridization method for genetic and epigenetic analyses. *Mol. Genet. Genomics* 271, 678–686.

Yang, A.S., Doshi, K.D., Choi, S.W., Mason, J.B., Mannari, R.K., Gharybian, V., Luna, R., Rashid, A., Shen, L., Estecio, M.R., Kantarjian, H.M., Garcia-Manero, G., and Issa, J.P. (2006). DNA methylation changes after 5-aza-2′-deoxycytidine therapy in patients with leukemia. *Cancer Res.* 66, 5495–5503.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* 13, 335–340.

Yoo, C.B., and Jones, P.A. (2006). Epigenetic therapy of cancer: past, present and future. *Nat. Rev. Drug Discov.* 5, 37–50.

Yu, S.L., Chen, H.Y., Chang, G.C., Chen, C.Y., Chen, H.W., Singh, S., Cheng, C.L., Yu, C.J., Lee, Y.C., Chen, H.S., Su, T.J., Chiang, C.C., Li, H.N., Hong, Q.S., Su, H.Y., Chen, C.C., Chen, W.J., Liu, C.C., Chan, W.K., Chen, W.J., Li, K.C., Chen, J.J., and Yang, P.C. (2008). MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell* 13, 48–57.

Yu, Y.P., Paranjpe, S., Nelson, J., Finkelstein, S., Ren, B., Kokkinakis, D., Michalopoulos, G., and Luo, J.H. (2005). High throughput screening of methylation status of genes in prostate cancer using an oligonucleotide methylation array. *Carcinogenesis* 26, 471–479.

Yuan, E., Haghighi, F., White, S., Costa, R., McMinn, J., Chun, K., Minden, M., and Tycko, B. (2006). A single nucleotide polymorphism chip-based method for combined genetic and epigenetic profiling: validation in decitabine therapy and tumor/normal comparisons. *Cancer Res.* 66, 3443–3451.

Zeschnigk, M., Bohringer, S., Price, E.A., Onadim, Z., Masshofer, L., and Lohmann, D.R. (2004). A novel real-time PCR assay for quantitative analysis of methylated alleles (QAMA): analysis of the retinoblastoma locus. *Nucleic Acids Res.* 32, e125.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., and Ecker, J.R. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126, 1189–1201.

Zou, H.Z., Yu, B.M., Wang, Z.W., Sun, J.Y., Cang, H., Gao, F., Li, D.H., Zhao, R., Feng, G.G., and Yi, J. (2002). Detection of aberrant p16 methylation in the serum of colorectal cancer patients. *Clin. Cancer Res.* 8, 188–191.

Chapter 15

High-Resolution Melting Curve Analysis for Molecular Diagnostics

Jared S. Farrar[1], Gudrun H. Reed[1] and Carl T. Wittwer[1,2]
[1]*Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, UT;*
[2]*ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT*

15.1 INTRODUCTION TO MELTING ANALYSIS

DNA melting or denaturation occurs when double-stranded DNA (dsDNA) separates into random coils of single-stranded DNA (ssDNA). The melting of DNA can be followed with dyes, which fluoresce in the presence of dsDNA. When a PCR product is melted in the presence of a dsDNA dye, the fluorescence is monitored continuously and plotted against the temperature. As the temperature increases, there is a characteristic drop in fluorescence at high temperature that follows the denaturation of the PCR product. The melting profile of a PCR product depends on its GC content, length, and sequence. Short PCR products usually melt in a single transition. The melting temperature is the temperature at which 50% of the DNA is double stranded and the remainder is single stranded. Longer PCR products often melt in multiple transitions, corresponding to melting domains of different stability. It has become common practice to melt PCR products after amplification to assess their identity and purity.

Traditional genotyping by melting analysis relies on the use of labeled probes. Variants under the probe are detected, but not those outside of the probe region. Conventional variant scanning identifies variants anywhere within a PCR product, but requires separation of the mixture through a gel or other matrix. In contrast, high-resolution melting is a closed-tube, homogeneous technique that does not require labeled probes or sample processing after PCR. Both genotyping and scanning are performed with only a dsDNA dye and two primers. For fine discrimination of multiple variants within a particular region, an unlabeled probe or snapback primer tail can be included. High-resolution melting can rapidly establish sequence identity when specific genotyping is not required, as in HLA matching or the assessment of repeated DNA segments. Melting is non-destructive, allowing subsequent analysis of the PCR product when necessary. The simplicity and speed of high-resolution melting analysis is attractive and is reflected by its increasing use in molecular diagnostics. The method is enabled by high-resolution melting instruments and dsDNA dyes that detect heteroduplexes. High-resolution melting analysis has been recently reviewed (Reed *et al*., 2007; Erali *et al*., 2008b).

15.1.1 History

Melting curve analysis was introduced as an integral part of real-time PCR (see also Chapter 7) with the LightCycler® in 1997 (Wittwer *et al*., 1997b). Instead of looking at product fluorescence once every cycle, continuous monitoring during PCR was performed (Wittwer *et al*., 1997a). The dsDNA dye SYBR® Green I allowed both template quantification and melting curve analysis to distinguish between PCR products (Ririe *et al*., 1997). In many cases the need for further analysis by electrophoresis was eliminated.

High-resolution amplicon melting was first reported using a 5′-labeled primer to provide the fluorescence signal. Both heterozygous and homozygous single base variants were identified (Gundry *et al*., 2003). However, a fluorescently labeled primer was needed and only variants residing in the melting domain of the labeled primer were detected. The introduction of dsDNA dyes that efficiently detect heteroduplexes eliminated the need for labeled primers (Wittwer *et al*., 2003). For the first time, heterozygous variants anywhere between the primers could be easily identified without any processing or separation steps after PCR.

The first report of genotyping by fluorescent melting curve analysis used hybridization probes and fluorescence

resonance energy transfer (Lay and Wittwer, 1997). A fragment of *F5* containing the variant g.1691A > G (Leiden) was amplified by asymmetric PCR. The reaction included a Cy5-labeled primer and a 3′-fluorescein-labeled probe covering the variant site. When the fluorescein probe was hybridized to the extension product of the Cy5 primer, energy transfer enhanced the Cy5 fluorescence. Melting of the duplexes provides the genotype because the different alleles resulted in different probe melting temperatures. Combined with rapid cycle PCR, genotyping required only 30 minutes. Subsequently, genotyping with two adjacent hybridization probes (HybProbes®), each labeled with a donor or acceptor fluorophore, was first demonstrated using *HFE* variants (Bernard *et al*., 1998). The method was later simplified by using one fluorescein-labeled probe (SimpleProbe®) instead of two probes (Crockett and Wittwer, 2001).

Genotyping by melting without labeled probes was first achieved using SYBR® Green I and allele-specific PCR using three primers, one with a GC tail to discriminate alleles (Germer and Higuchi, 1999). Later on, the need for allele-specific PCR and GC tailing was eliminated by small amplicon melting, a method enabled by high-resolution melting. Only two standard primers and a dsDNA dye that detects heteroduplexes were required (Liew *et al*., 2004). For probe-based genotyping, the same dsDNA dye can detect the melting of unlabeled probes (Zhou *et al*., 2004a). Unlabeled probe genotyping uses three standard oligonucleotides (the probe is blocked at the 3′-end to prevent extension) and can easily identify homozygous variants that may be difficult to genotype by amplicon melting alone.

Alternatively, snapback primer genotyping uses only two primers, one with a 5′-tail that serves as the probe element, forming an intramolecular hairpin (Zhou *et al*., 2008). Asymmetric PCR is usually required for sufficient single-stranded product to hybridize to unlabeled probes or snapback primers. Two melting regions are observed, one for the probed region and another for the amplicon, allowing simultaneous genotyping and variant scanning in the same assay (Zhou *et al*., 2005).

15.1.2 Components of the Technology

High-resolution melting of DNA for genotyping, scanning, and sequence matching was made possible by three developments. First, dsDNA dyes that detect heteroduplexes were identified from existing commercial dyes or by new design and synthesis. Second, dedicated high-resolution melting instruments became available and existing real-time PCR instruments were modified to increase their melting resolution. Finally, software for data normalization, curve shape comparison, and genotype clustering were developed. Used in concert, these tools increase melting curve quality and allow the detection of small sequence differences in PCR products.

15.1.2.1 dsDNA Dyes for Melting Analysis

dsDNA binding dyes monitor the melting of entire PCR products. Traditionally, the dye SYBR® Green I was used. SYBR® Green I is a sensitive and convenient dye for quantitative PCR and product melting analysis (Wittwer *et al*., 1997b). Early reports of closed-tube SYBR® Green I genotyping (Marziliano *et al*., 2000; Pirulli *et al*., 2000) were later questioned (von Ahsen *et al*., 2001). This was followed by a study demonstrating single base genotyping in products up to 167 bp (Lipsky *et al*., 2001). However, the protocol required purifying the samples after PCR, followed by the addition of high concentrations of dye. Unfortunately, SYBR® Green I inhibits PCR at these high saturating concentrations (Wittwer *et al*., 1997a). At dye concentrations that are compatible with PCR, lower melting products were difficult to observe because of dye redistribution during melting and/or G/C base-pair specificity (Giglio *et al*., 2003; Gundry *et al*., 2003; Wittwer *et al*., 2003). In the case of heterozygote amplification, strand re-association from heteroduplexes to homoduplexes may also occur during melting, thereby masking the heteroduplexes. Most dsDNA dyes, including SYBR® Green I, ethidium bromide, SYBR® Gold, Pico Green, TOTO®-1 and YOYO®-1, do not detect heteroduplexes well (Wittwer *et al*., 2003). Nevertheless, SYBR® Green I has been used to detect heterozygotes in a closed tube system (Dufresne *et al*., 2006). Furthermore, SYBR® Green I genotyping may be successful when the melting differences between genotypes is large, as after gap PCR (Pornprasert *et al*., 2008), repeat typing (Price *et al*., 2007), or methylation analysis (Worm *et al*., 2001).

Different dsDNA dyes detect heteroduplexes to a greater or lesser extent. In a systematic study aimed at developing the best dye for heterozygote detection, over 30 existing commercial dyes were compared to over 50 newly synthesized dyes (Wittwer *et al*., 2008). The best dyes were identified and trademarked "LCGreen®" (Idaho Technology) in honor of the LightCycler® system. Figure 15.1 compares LCGreen®+ with SYBR® Green I and to two other commercial dyes that have been used to detect heterozygotes. The apparent heteroduplex percentages observed were 23.5% for LCGreen®+, 17.6% for SYTO® 9 (Invitrogen), and 15.2% for EvaGreen® (Biotium). The percentage of observed heteroduplexes directly correlates with the ease of heterozygote detection and scanning sensitivity. Unlike SYBR® Green I, the other dyes do not inhibit PCR at saturating concentrations and are referred to as "saturating" dsDNA dyes.

LCGreen® dyes are asymmetric cyanines (Dujols *et al*., 2006) specifically developed for heteroduplex detection. SYTO® 9 is marketed as a nucleic acid stain for both gram-positive and gram-negative bacteria. EvaGreen® was developed

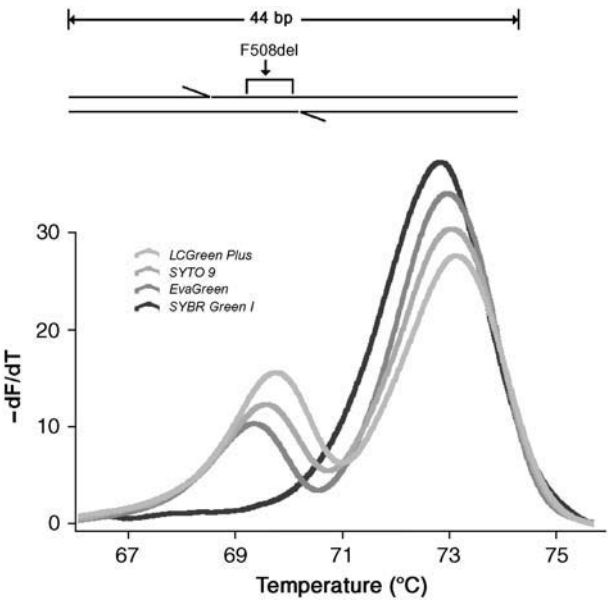


FIGURE 15.1 Heterozygote detection with different dyes. The ability to detect heterozygotes was compared among three saturating dyes (LCGreen®+, SYTO® 9, EvaGreen®), and the non-saturating dye, SYBR® Green I. A short amplicon surrounding the p.F508del site of *CFTR* was amplified, using heterozygous p.F508del DNA as template (Wittwer *et al*., 2003). High-resolution melting analysis (HR-1™, Idaho Technology) was performed, and the data plotted as a negative derivative after normalization and curve overlay. In each case, optimal dye concentrations were used that minimized PCR inhibition and maximized heteroduplex detection. Although heteroduplexes are not observed with SYBR® Green I, increasing heteroduplex signals are observed for EvaGreen®, SYTO® 9, and LCGreen®+.

for DNA quantification and has been used in capillary gel electrophoresis (Sang and Ren, 2006) and quantitative PCR (Mao *et al*., 2007). dsDNA dyes generally increase DNA melting temperatures and may require adjustment of cycling parameters. If the melting temperature of the amplicon is already high, an agent such as DMSO, glycerol, or betaine may be required to prevent apparent ''inhibition''. These dyes are stable to PCR conditions and can monitor PCR in real-time using the SYBR® Green I/fluorescein channel of real-time PCR instruments. Potential advantages of SYTO® 9 (Paul *et al*., 2005) and EvaGreen® (Mao *et al*., 2007) over SYBR® Green I for real-time PCR and standard melting analysis have been detailed. For example, SYTO® 9 detects multiple PCR products and melting domains better than SYBR® Green I and can effectively identify different strains of *Naegleria* (Robinson *et al*., 2006) and *Mycoplasma* (Jeffery *et al*., 2007). Furthermore, both SYTO® 9 and EvaGreen® have been used in methylation analysis where heteroduplex detection is not required or even desired (Wojdacz and Dobrovic, 2007; Snell *et al*., 2008; Kristensen *et al*., 2008; Wojdacz *et al*., 2008; White *et al*., 2007). The LCGreen® dyes were the first commercial dyes for heteroduplex detection and remain the best dyes for genotyping and scanning applications where heteroduplex detection is critical (Fig. 15.1). Nevertheless, SYTO® 9 has been used for heteroduplex scanning (Krypuy *et al*., 2006, 2007; Takano *et al*., 2008; Do *et al*., 2008) and single base genotyping of diploid DNA (Kristensen and Dobrovic, 2008).

15.1.2.2 Instrumentation

High-resolution melting requires discriminating small melting curve differences. As PCR products become larger, these differences become smaller (Gundry *et al*., 2003; Reed and Wittwer, 2004) and instrument precision becomes more critical. Standard real-time instruments are not designed for high-resolution melting. Figure 15.2 compares different high-resolution melting and real-time PCR platforms using identical analysis software (Herrmann *et al*., 2006, 2007a, b). By measuring multiple identical samples, the instrument contributions to variation were determined. The most important technical metric is variation in curve shape, used in variant scanning and heterozygote identification. This is the variation remaining after the curves are overlaid so that curve shapes can be compared. The instruments in Fig. 15.2 are arranged in order of scanning sensitivity (lower standard deviations are better). As expected, instruments designed for high-resolution melting are better scanning platforms than other instruments. Another technical metric is variation in absolute temperature, important for distinguishing different homozygotes. Air-controlled and single sample systems perform better than plate-based systems by this metric.

In addition to technical performance of melting instruments, there are many other factors to consider. For example, although the HR-1™ (Idaho Technology) has the best technical specifications and the fastest ramping rate (0.3°C/s) for the shortest turnaround time (1–2 min), it is not acceptable for high-throughput users. Sample tracking errors have occurred when many samples are manually processed on the HR-1™ (Vandersteen *et al*., 2007). The LightScanner® (Idaho Technology) has excellent scanning sensitivity and the highest throughput because multiple thermocyclers can feed into one 96- or 384-well LightScanner® with an analysis cycle of only 5–10 min. However, genotyping accuracy is limited unless internal temperature controls are included (Gundry *et al*., 2008). Furthermore, the HR-1™ and the LightScanner® do not perform PCR. The LightCycler® 480 (Roche) is a 96/384 microtiter plate system with integrated real-time PCR and excellent scanning sensitivity. However, genotyping accuracy is limited and the melting cycle is slower. The real-time Rotor-Gene® 6000 HRM series (Corbett Life Science) has excellent genotyping accuracy. However, the scanning sensitivity is less, and the melting cycle is also slow. The ''best'' instrument depends on the primary application and the relative needs for throughput, turnaround time, sample format, economy, and PCR integration. Competition should improve melting hardware in the future as genotyping

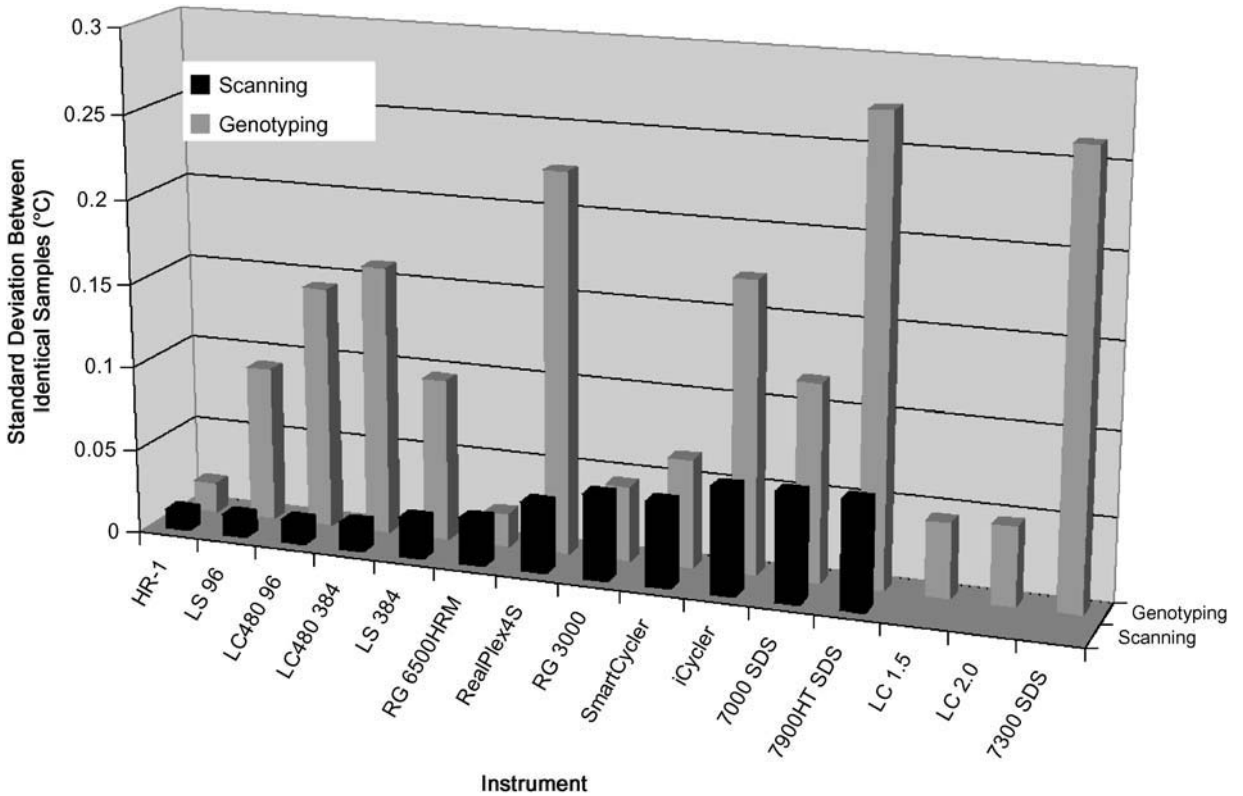


FIGURE 15.2 Instrument variation for scanning and genotyping. Fifteen different instruments from seven different vendors are ordered according to scanning variation (black bars), with genotyping variation (gray bars) also indicated. Scanning variation (curve shape) is nearly always less than genotyping variation (absolute temperature). Some instruments (e.g. LightCycler® 1.5, LightCycler® 2.0, 7300 SDS) could not be evaluated for scanning variation because of low data density. Data plotted here were taken from Herrmann and coworkers (2007b). Instruments and vendors are HR-1™ (Idaho Technology), LS96 (LightScanner® 96-well version, Idaho Technology), LC480 96 (96-well block on the LightCycler® 480, Roche Applied Science), LC480 384 (384-well block on the LightCycler® 480, Roche Applied Science), LS384 (LightScanner® 384-well version, Idaho Technology), RG 6500HRM (Rotor-Gene® 6500 high-resolution melting version, Corbett Life Science), RealPlex4S (Mastercycler® Realplex4S, Eppendorf), RG3000 (Rotor-Gene® 3000, Corbett Life Science), Smartcycler® (Cepheid), iCycler® (BioRad), 7000, 7300, and 7900HT SDS (ABI), LightCycler® (LC) versions 1.5 and 2.0 (Roche Applied Science).

accuracy and scanning sensitivity depend solely on instrument precision to resolve small differences between melting curves.

15.1.2.3 Data Analysis

Melting data is usually processed before final presentation. However, viewing the original data is important for quality control and to establish the data density (points per °C). The actual data points (not smoothed curves) should be displayed after analog to digital conversion but before software manipulation. The absolute sample fluorescence depends on the sample volume, instrument optics, and PCR amplification. Samples that amplify poorly (including any negative controls) will have low fluorescence and should not be analyzed further. If the melting transition is unexpectedly meager or multiple transitions are observed when only one is expected, unintended (''non-specific'') PCR products are likely. It is best first to optimize the PCR by annealing temperature gradients and gels before melting analysis is attempted. Approximate melting temperatures and profiles can be predicted (Rasmussen *et al*., 2007) and if the experimental curves are very different from these predictions, the PCR is again suspect. Different samples will usually vary in absolute fluorescence as a result of optical and/or volume differences (Fig. 15.3a). Given a pure PCR product, a few simple steps can turn data that appear highly variable (Fig. 15.3a) into discrete clusters that correlate with genotype (Fig. 15.3c–e).

15.1.2.3.1 Background Removal and Normalization

High-resolution melting curves are easier to compare after the fluorescence background is removed and they are normalized. Even though the sharpest drop in fluorescence occurs as the amplicon melts, fluorescence also decreases with increasing temperature outside of the melting transition. This decrease is nearly linear on each side of a PCR product (Gundry *et al*., 2003), but is better approximated by an exponential when both low (probe) and high (amplicon) temperature transitions are present (Erali *et al*., 2008a).

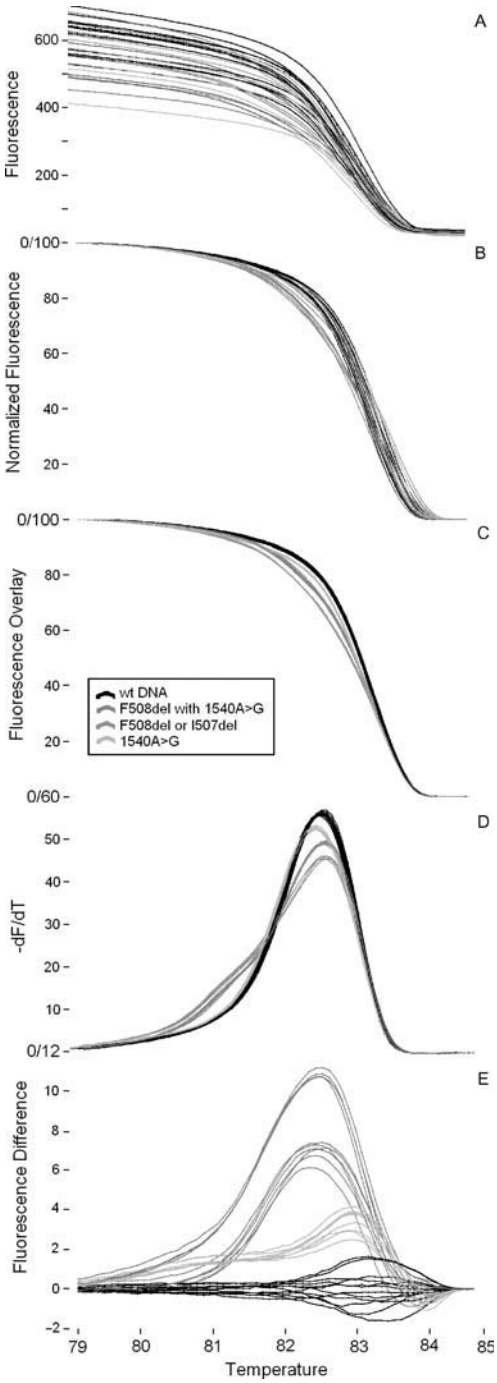


FIGURE 15.3 High-resolution melting curve analysis. PCR amplicon melting curves reveal duplex melting as a sharp drop in fluorescence with increasing temperature. The data displayed in all panels include four melting curve clusters of *CFTR*, exon 10 (Montgomery *et al.* 2007a). **A.** The original data, showing variation in fluorescence among samples at low temperature due to optical or sample differences, does not separate into 4 clusters. **B.** Normalized data, correcting for the temperature effect on fluorescence, still does not completely separate the four clusters. **C.** Curve overlay of normalized data compares the shape of the melting curves. Although the differences between clusters are small, they are now discernible. **D.** The negative derivative of overlaid data further visualizes the differences between clusters. **E.** The preferred data display for high-resolution amplicon melting analysis is shown here as a difference plot, where all clusters are easy to discern.

If the background is properly removed, the fluorescence both before and after duplex melting will be horizontal. Normalization plots all curves between 100% (completely hybridized) and 0% (completely single stranded). Although differences in fluorescence magnitude are lost with normalization, the melting transitions are easier to compare. Melting curves after background removal and normalization are shown in Fig. 15.3b.

15.1.2.3.2 Curve Overlay

Overlay of melting curves provides the best sensitivity for heterozygote detection. This is true, irrespectively of the instrument because all instruments show some temperature variation between samples and all samples are not identical because of pipetting and evaporation variation. These differences can be decreased by overlaying all curves so that they are superimposed over the high-temperature region (usually between 2 and 5% normalized fluorescence). This curve overlay allows for easy identification of heteroduplexes by their melting curve shapes. However, temperature differences between samples are scrambled, so that variant homozygotes may be more difficult to identify after curve overlay. In general, when it is important to identify homozygous variants, data should be examined both with and without curve overlay. Overlaid melting curves after background subtraction and normalization are shown in Fig. 15.3c.

15.1.2.3.3 Derivative Plots

Negative first derivative plots are used extensively in labeled probe genotyping (Lay and Wittwer, 1997; Bernard *et al.*, 1998) and low resolution amplicon analysis (Ririe *et al.*, 1997). The negative first derivative is usually determined by Solvitsky-Golay polynomial estimation (Wittwer and Kusukawa, 2004). When melting curves are converted to derivative plots, curve peaks approximate melting temperatures. Unlabeled probe (Zhou *et al.*, 2004a) and snapback primer genotyping (Zhou *et al.*, 2008) typically use derivative plots for analysis. Although derivative plots can be used for high-resolution amplicon melting (Fig. 15.3d), high-resolution data are best presented without the smoothing process inherent in taking a derivative.

15.1.2.3.4 Difference Plots

High-resolution scanning data are best presented on difference plots (Fig. 15.3e). Different genotypes are easiest to identify on these magnified graphs (Wittwer *et al.*, 2003). They are created by plotting the difference between a reference curve and all other melting curves. The reference curve is usually the average of all wild-type curves so that this group clusters around the horizontal axis. Different genotypes trace different paths for easy visual discrimination. Automatic clustering of genotypes can be performed, for example, by unbiased hierarchal clustering (Zhou *et al.*, 2005; Vandersteen *et al.*, 2007).

15.2 GENOTYPING BY HIGH-RESOLUTION MELTING

High-resolution melting is a simple, yet powerful, closed-tube, homogeneous technique for genotyping. Melting methods that use saturating dsDNA dyes include amplicon melting, unlabeled probe, and snapback primer genotyping.

15.2.1 Amplicon Melting

Amplicon melting is commonly used in three ways: (1) specific genotyping by small amplicon melting, (2) sequence matching when identity is more important than specific genotyping, e.g. in HLA matching and identity assessment using repeats, and (3) methylation analysis.

15.2.1.1 Small Amplicon Melting

Direct amplicon melting was the first genotyping method introduced with dsDNA saturation dyes (Wittwer *et al*., 2003). Although long PCR products (>500 bp) can be genotyped by high-resolution amplicon melting, melting curve differences between genotypes are easier to distinguish in small amplicons (Liew *et al*., 2004). Very rapid PCR is possible with small amplicons by reducing the denaturation temperature and eliminating all temperature holds. PCR and genotyping requires only 10–20 minutes on real-time instruments like the carousel LightCycler® (Roche) or inexpensive platforms, such as the RapidCycler® 2 (Idaho Technology) matched with the HR-1™. No probes are required as the dye labels the entire PCR product. Genotyping of a single base variant is illustrated in Fig. 15.4, showing normalized and background subtracted melting curves. Heterozygotes are easily identified by a change in curve shape. Homozygotes are typically identified by a shift in melting temperature, so curve overlay is not performed.

Biallelic single base variants in diploid DNA (C/T, G/A, C/A, G/T, C/G, and T/A) can be grouped into four classes based on the homoduplexes and heteroduplexes formed after PCR (Liew *et al*., 2004). In all cases, heterozygotes are easy to identify by curve shape. Homozygotes in either Class 1 or Class 2 are easy to distinguish by melting temperature because one homozygote contains an A:T and the other a G:C pair. In short amplicons, the differences in melting temperature are between 0.8 and 1.4°C for these two classes and make up more than 84% of human single base variants. Homozygotes in either Class 3 or Class 4 are more difficult to distinguish because the base pair (A:T or G:C) stays the same (e.g. A > T, C > G, etc.) with the bases merely switching strands. Differences in melting temperature do occur, however, because of nearest-neighbor interactions with adjacent bases, but the differences are generally <0.40°C. Class 3 and 4 variants make up approximately 16% of human single base variants. Within these

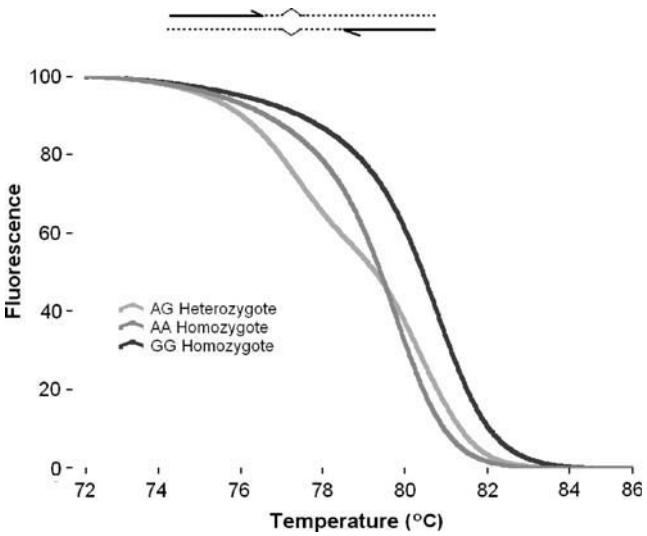


FIGURE 15.4 Genotyping by small amplicon melting (Liew *et al*., 2004). Typical normalized melting curves representing all genotypes of an A > G variant are shown (456p). The homozygotes are differentiated by absolute temperature and the heterozygote is discerned by the change in curve shape. Note that no curve overlay is performed for small amplicon genotyping.

two classes, nearest-neighbor symmetry occurs in one case out of four (Liew *et al*., 2004; Palais *et al*., 2005), predicting identical homozygous melting temperatures in 4% of human single base variants. Homozygous small insertions/deletions may also be difficult or impossible to distinguish by melting temperature alone, for example p.F508del in cystic fibrosis (Montgomery *et al*., 2007a). Similarly, in haploid organisms (most microbes) and for hemizygous genes, heterozygotes do not occur and different sequences (strains, genotypes) may have very similar or identical melting temperatures. As melting temperature differences become smaller, instrument precision and solution consistency become more critical for sequence resolution.

There are several ways to better discriminate genotypes with very similar melting temperatures. Different alleles can be linked by allele-specific amplification to primer tails that modify amplicon melting temperature. The tails can be GC or AT rich or may incorporate locked nucleic acids (Germer and Higuchi, 1999; Erali *et al*., 2006; Seipp *et al*., 2008). Internal temperature standards, which increase absolute temperature precision, are particularly useful on plate-based systems (Liew *et al*., 2007; Seipp *et al*., 2007; Nellaker *et al*., 2007). For example, most nearest-neighbor symmetric single base variants (predicted to have identical melting temperatures) can be distinguished by including internal temperature standards that control for both instrument and solution variance (Gundry *et al*., 2008). Finally, unknown samples may be mixed with a known genotype. Typically, the known genotype is wild type. If the unknown is wild type, the melting curve of the mixture will be the same. If the unknown is a homozygous variant, the mixture will be

heterozygous and heteroduplexes will alter the curve shape. Mixing can be performed either before or after PCR. If mixing is done after PCR, only the homozygous samples need to be mixed, but the benefits of a closed-tube system are lost. If mixing is done before PCR, quantitative heteroduplex analysis can be performed with optimal mixing ratios to easily distinguish all three genotypes (Palais *et al*., 2005).

15.2.1.2 Sequence Matching

In some cases, complete genotyping is not as important as knowing whether the same sequences are present. Interesting examples are HLA analysis for transplantation compatibility, repeat typing for identity matching, and genetic mapping.

15.2.1.2.1 HLA Matching

Conventional HLA testing requires specific typing of four to six highly polymorphic loci, a rather tedious task. In contrast, high-resolution melting can easily establish HLA sequence identity and transplant compatibility, at least for living related transplants. For example, high-resolution melting was applied to the highly polymorphic HLA-A locus (Zhou *et al*., 2004b). All seven cases of shared alleles among two individuals were assessed and a 17-member family tested. Results from melting curve analysis were concordant with sequencing. HLA identity was suggested when two individuals had the same melting curve. Identity was confirmed by comparing the melting curve of a 1:1 mixture. This technique is being extended to other HLA loci for more complete HLA matching. Analysis of unrelated donor/recipient pairs for allogeneic hematopoietic stem-cell transplantation is a more difficult but intriguing possibility.

15.2.1.2.2 Repeated Sequences

Repeat sequence polymorphisms are usually typed by size separation on gels. A few studies have used high-resolution melting analysis to differentiate these polymorphisms with some success. Internal duplications in the human *FLT3* gene between 6 and 102 bp were easily detected with 100% concordance to sequencing (Vaughn and Elenitoba-Johnson, 2004). Differentiation of clustered short sequence repeats of *Campylobacter jejuni* (Price *et al*., 2007), the complex repeat structure of SpA in *Staphylococcus aureus* (Stephens *et al*., 2008), and multilocus short repeats in *Bacillus anthracis* (Fortini *et al*., 2007) have been reported. High-resolution melting of dinucleotide repeats has also been used to identify varieties of grapes and olives (Mackay *et al*., 2008).

15.2.1.2.3 Genetic Mapping

Single base variants can be used in genetic mapping when the sequence and position of the variant is known. However, high-resolution melting can provide fine mapping in the absence of specific genotyping. Starting with expressed sequence tags of unknown variation, different melting curves were directly correlated to phenotype without knowledge of the sequence change (Lehmensiek *et al*., 2008). Such mapping can be performed without probes in a closed system.

15.2.1.3 Methylation Analysis

The methylation of cytosine in DNA is important in the regulation of gene activity and expression (Jones and Baylin, 2002). Methylation of cytosine (C) to 5-methylcytosine (m^5C) typically occurs within CpG islands near the promoters of protein encoding genes (Brinson *et al*., 1997; Kristensen *et al*., 2008). Methylation information is usually lost during PCR, so either methylation-specific restriction endonuclease or bisulfite treatment is used before amplification. Bisulfite treatment converts Cs to Us, while m^5Cs are not affected, allowing methylation-specific PCR. Alternatively, bisulfite sequencing reveals the position of all m^5C bases.

Bisulfite treatment and PCR convert unmethylated C:G pairs to A:T pairs so that the amplicon melting temperature is directly related to the degree of methylation. In contrast to methods that require heteroduplex detection, saturation dyes are not required; methylation analysis by melting was introduced with SYBR® Green I (Worm *et al*., 2001). High-resolution melting appears to increase the sensitivity and precision of analysis (Wojdacz and Dobrovic, 2007; Dahl and Guldberg, 2007). Variations include melting analysis after real-time methylation-specific PCR (Kristensen *et al*., 2008) and digital methylation-sensitive high-resolution melting (Snell *et al*., 2008).

Methylation analysis by high-resolution melting has been used to diagnose imprinting disorders, including the Angelman and Prader-Willi syndromes (White *et al*., 2007) and the Beckwith Wiedemann and Russell Silver syndromes (Wojdacz *et al*., 2008). Additional studies on promoter methylation include *BRCA1* (Snell *et al*., 2008), *CDH1, DAPK1, CDKN2A,* and *RARB* (Kristensen *et al*., 2008), and *RASEF* (Maat *et al*., 2008). *MGMT* and *BNIP3* methylation down to 0.1–1% could be detected in colorectal cancer samples (Wojdacz and Dobrovic, 2007).

15.2.2 Unlabeled Probe Genotyping

Many sequence variants can be distinguished by amplicon melting alone. However, when greater detail is needed or definitive genotyping is required, unlabeled probes can be used (Zhou *et al*., 2004a). The same saturating dye used for amplicon melting can be used for unlabeled probe genotyping. Although high-resolution melting is not an absolute requirement for unlabeled probe genotyping, more genotypes can be distinguished on high-resolution instruments. Asymmetric PCR is performed for 40–50 cycles to produce excess strand complementary to the probe. A 1:5 to 1:10 primer ratio usually produces enough double-stranded product for amplicon melting as well as enough single-stranded

product for probe melting (Dujols *et al*., 2006). The 3′-end of the unlabeled probe is blocked in order to prevent extension during amplification. This is typically accomplished by 3′-phosphorylation. However, Dames and coworkers (2007a) reported that incomplete 3′-phosphorylation may result in probe extension and aberrant melting profiles. Probe blocking can be improved with amino-modified C6, inverted dT, or a C3 spacer. Unlabeled probe genotyping has recently been reviewed (Erali *et al*., 2008a).

Unlabeled probes between 20 and 35 bases with melting temperatures of 55–70°C are generally recommended (Zhou *et al*., 2004a). If the polymerase lacks 5′ to 3′ exonuclease activity, the probe should melt from all alleles before the PCR extension temperature to prevent possible allele bias and/or PCR inhibition. Higher probe melting temperatures can be used with exonuclease positive polymerases, although it is convenient to keep their melting temperatures below those of primer dimers or other alternative amplification products. Estimated probe melting temperatures are typically 1–4°C lower than observed melting temperatures due to dye stabilization of the hybrid (Zhou *et al*., 2005).

After exponential background subtraction (Erali *et al*., 2008a), unlabeled probe melting data is usually shown on derivative plots in order to easily visualize the melting transitions. Typical single base genotyping with an unlabeled probe is shown in Fig. 15.5. Two melting regions are apparent. At lower temperatures, probe melting occurs for specific genotyping. At higher temperatures, amplicon melting occurs. With high-resolution analysis, amplicon melting can be used to scan for variants anywhere between the primers. Genotyping and scanning can even be performed simultaneously from the same melting curve (Zhou *et al*., 2005; Montgomery *et al*., 2007b).

Probes can be designed to match either the wild-type or variant sequence. For example, common variants revealed by scanning can be definitively identified by matching the probe to the variant (Vandersteen *et al*., 2007). Probes can also be designed to mask benign sequence variations near targeted disease-causing variants by incorporating deletions, unmatched nucleotides or universal bases (Margraf *et al*., 2006b, 2007). Multiple alleles can be genotyped using unlabeled probes. For example, Zhou and coworkers (2005) demonstrated that five variants within exon 10 of the *CFTR* gene could be genotyped with two unlabeled probes in the same reaction. One of the unlabeled probes genotyped two single base variants and two deletions.

15.2.3 Snapback Primer Genotyping

Snapback primers are unlabeled probes attached to the 5′-end of a primer. After PCR, the result is a self-probing amplicon where one PCR strand forms a hairpin (Zhou *et al*., 2008). The stability of the hairpin duplex can be monitored by melting and is sensitive to sequence. Unlike Scorpion® primers, no covalent modifications are necessary.

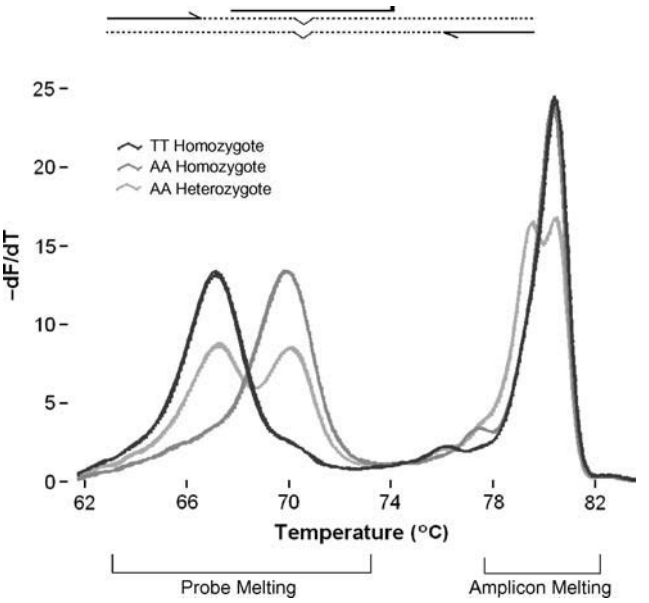


FIGURE 15.5 Genotyping with unlabeled probes (Zhou *et al*. 2004a). A 3′-blocked unlabeled probe covers the variable region, in this case an A > T variant. After asymmetric PCR, 2 melting regions are apparent. At low temperature, the probe melts from single-stranded amplicon, definitively determining genotype. At high temperature the double-stranded amplicon melts, easily identifying the heterozygote but not separating the homozygotes. The small peaks around 76–78°C are extension products of the probes, sometimes seen with incomplete blockage of the probe's 3′-end. The data are displayed as a derivative plot after exponential background removal and normalization.

Asymmetric PCR is usually performed in order to create both intramolecular snapback hairpins as well as intermolecular duplexes of full-length amplicons. The melting of both the amplicon and snapback duplexes allows for investigation of sequence variations within the amplicon as well as variants within the snapback hairpin stem. Similar to unlabeled probe analysis, the melting transitions of both the snapback and full-length amplicon duplexes are usually plotted on derivative plots (Fig. 15.6).

Advantages of snapback primers over unlabeled probes include (1) only two oligonucleotides are needed instead of three, (2) no 3′-blocking is necessary, and (3) short probe lengths (similar to locked nucleic acids or minor groove binders) can be used because of intramolecular stabilization. The hairpin melting temperature is linearly related to the stem length (6–28 bps) and inversely related to the log of the loop size (17–150 bases). A 2 bp mismatch at the 5′-end of the snapback primer is usually included to limit 3′-extension of the minor snapback product.

15.2.4 Applications in Molecular Diagnostics

Genotyping by high-resolution amplicon melting was first reported for common disease variants of *HBB*, *F2*, *F5*, *MTHFR*, *HFE*, and *CFTR* (Gundry *et al*., 2003; Wittwer *et al*., 2003; Liew *et al*., 2004). Subsequent studies

included *CYP2C9* (Hill *et al*., 2006), *LCT* (Liew *et al*., 2007), *LRRK2* (Tedde *et al*., 2007), *FGFR3* (Hung *et al*., 2008), *MTR* and *DNMT3b* (Kristensen and Dobrovic, 2008) and *IL10* (Tedde *et al*., 2008). Amplicon genotyping can be multiplexed if different targets are separated in melting temperature. Examples include duplex assays for human platelet antigen (Liew *et al*., 2006) and *MTHFR* variants (Seipp *et al*., 2007), and triplex assays for *PAH* (Dobrowolski *et al*., 2007a) and *OTC* polymorphisms (Dobrowolski *et al*., 2007b). Furthermore, four coagulation variants (*F2*, *F5*, and 2 *MTHFR*) have been genotyped in one quadruplex reaction (Seipp *et al*., 2008). Clinical applications of high-resolution amplicon melting in microbiology include mycobacteria speciation using *hsp65* (Odell *et al*., 2005), bacterial speciation using the 16S rRNA gene (Cheng *et al*., 2006), *Mycoplasma synoviae* strain typing (Jeffery *et al*., 2007), identifying *gyrA* variants that cause quinolone resistance in *Salmonella* (Slinger *et al*., 2007), subtyping influenza A (Lin *et al*., 2008a), and varicella zoster strain typing (Toi and Dwyer, 2008).

Unlabeled probes have been used to genotype *F5* and *CFTR* (Zhou *et al*., 2004a, 2005), *LCT* and human platelet antigens (Liew *et al*., 2007), and common polymorphisms of hereditary hemorrhagic telangiectasia (Vandersteen *et al*., 2007). Mutations were identified in the *RET* proto-oncogene with unlabeled probes that mask common polymorphisms (Margraf *et al*., 2007). Unlabeled probes have also been used to detect and differentiate HSV-1 and HSV-2 (Dames *et al*., 2007b).

Snapback primers have been used to genotype *F5* g.1691G > A and exon 10 *CFTR* variants (Zhou *et al*., 2008). In the latter, seven different genotypes at two different loci were typed using two snapback primers and symmetric PCR. However, a ten-fold dilution after PCR was required in order to favor intramolecular snapback formation, resulting in an open-tube assay.

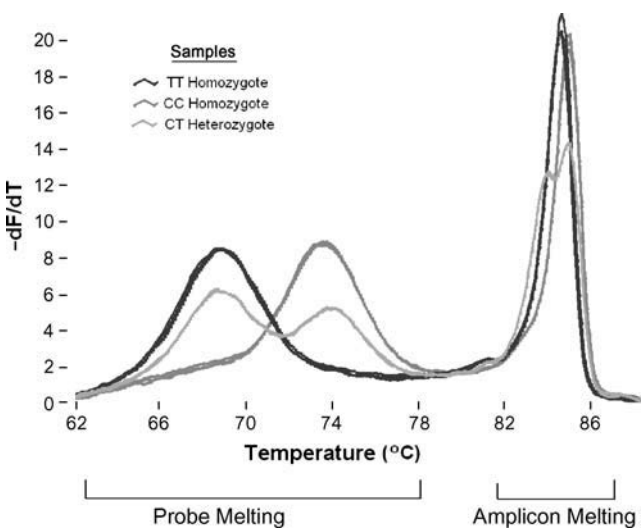


FIGURE 15.6 Genotyping with snapback primers (Zhou *et al*., 2008). A snapback primer includes a 5′-tail complementary to its extension product. After asymmetric PCR, regions of single-stranded hairpin (probe) melting and amplicon melting are revealed, similar to unlabeled probe genotyping (see also Fig. 15.5). High-resolution melting of all genotypes of a C > T variant are shown. The data are displayed as a derivative plot after exponential background removal and normalization.

15.2.4.1 Factor V Leiden Genotyping

Factor V Leiden is a single base variant in the *F5* gene (g.1691G > A) that increases the risk of deep venous thrombosis and pulmonary embolism. Genotyping by melting is typically performed with labeled hybridization probes. However, high-resolution melting enables genotyping with small amplicons, and unlabeled probes or snapback primers can also be used. Small amplicon genotyping was first shown by Liew and coworkers (2004) with wild-type and homozygous mutant curves separated by temperature and heterozygotes identified by curve shape. Graham and coworkers (2005) demonstrated that unexpected heterozygotes near g.1691G > A could also be distinguished (g.1690delC, g.1690C > T, as well as the compound heterozygote g.1696A > G/g.1690G > A). Simultaneous amplicon and unlabeled probe analysis of factor V Leiden was shown by Zhou and coworkers (2005) in a 384-well format. Both probe and amplicon melting provide independent assessments of genotype for increased confidence. Factor V Leiden genotyping has also been demonstrated with snapback primers (Zhou *et al*., 2008).

15.3 VARIANT (HETERODUPLEX) SCANNING BY HIGH-RESOLUTION MELTING

Variant scanning by high-resolution melting depends on detecting heteroduplexes after amplification. After PCR of a heterozygous sequence, four unique duplexes are formed, two homoduplexes (where both strands are completely complementary) and two heteroduplexes (with at least one mismatch between strands). The two homoduplexes are identical to the starting DNA and have melting temperatures that are usually very close to each other. The remaining two duplexes are heteroduplexes with mismatched bases and lower melting temperatures. The observed melting curve is a composite of the melting profiles of these four duplexes. Other scanning techniques also depend on detection of heteroduplexes, but high-resolution melting is unique because no physical separation of heteroduplexes is necessary.

Variant scanning by high-resolution melting is a closed-tube technique that requires no downstream processing to detect heterozygotes. During initial PCR setup, a saturating dsDNA dye that detects heteroduplexes is included. As PCR progresses, product accumulation is optionally monitored each cycle in real time. After PCR, the composite melting

curve of two homoduplexes and two heteroduplexes is measured by high-resolution melting. Heteroduplexes are detected by a change in melting curve shape after normalization (Fig. 15.3b) and curve overlay (Fig. 15.3c). Although these differences are present on derivative plots (Fig. 15.3d), they are easier to identify on difference plots (Fig. 15.3e). Heteroduplex detection by high-resolution melting is favored by rapid cooling before melting, rapid heating during melting, and low Mg^{2+} concentrations (Gundry *et al*., 2003). Depending on the instrument and protocol used, high-resolution melting can be performed in as little as 1 min or as long as 90 min or greater (Wittwer *et al*., 2003; Herrmann *et al*., 2007b).

The sensitivity and specificity of high-resolution melting for detecting heterozygous single base changes was comprehensively evaluated using engineered plasmids (Reed and Wittwer, 2004). The effects of PCR product size, type of base change, GC content, and the position of the base change in relation to the PCR product were considered. The PCR products ranged in size from 50 to 1,000 bp with GC contents of 40–60%. Sensitivity and specificity for products less than 400 bp was 100% ($n = 576$). For products between 400 and 1,000 bp the sensitivity was 96.1% and the specificity 99.4% ($n = 672$). The location or identity of the single base change did not affect sensitivity.

In 19 genetic studies using DNA extracted from human blood (Table 15.1), the weighted average sensitivity for heterozygote detection using high-resolution melting was 99.3% ($n = 839$) and the specificity 98.8% ($n = 2{,}659$). The actual sensitivity may even be higher, considering that of the five false negatives reported, two were clerical errors (Vandersteen *et al*., 2007), two used an arbitrary cutoff of $<5\%$ on difference plots (Laurie *et al*., 2007), and one was based on disease phenotype rather than genotype (Dobrowolski *et al*., 2007a). On the other hand, reporting bias may falsely elevate the sensitivity and specificity reported in these early studies. Furthermore, Tables 15.1 and 15.2 consider the sensitivity and specificity of heterozygote detection per amplicon. Correctly identifying causative mutations in an entire gene or set of genes responsible for a disease phenotype is a more difficult task, and the sensitivity and specificity will be lower. The amplicon lengths were usually between 100 and 500 bp. The LightScanner® instrument was most commonly used, followed by the HR-1™, the LightCycler® 480, and the Rotor-Gene® 6000, while the most common dyes used were LCGreen®+, LCGreen® I, ResoLight®, and SYTO® 9.

High-resolution melting is unusual among scanning techniques in that homozygous variants can also be identified (Dobrowolski *et al*., 2005). However, some homozygous variants are not detected, such as homozygous p.F508del in *CFTR* (Chou *et al*., 2005; Montgomery *et al*., 2007a). From Table 15.1, the average homozygote detection sensitivity was 75% ($n = 60$), using the five studies reporting these data. To detect all homozygous and hemizygous variants, mixing with a known genotype is necessary, either before or after PCR (Liew *et al*., 2004; Palais *et al*., 2005).

DNA is usually extracted from whole blood for genetic scanning studies. However, dried blood spots (Dobrowolski *et al*., 2007a) and whole genome amplified DNA (Margraf *et al*., 2006a) have also been used. The sensitivity and specificity of scanning is often decreased when tumor tissue is analyzed because of normal cell admixture and/or sample fixation/staining before DNA extraction. Table 15.2 summarizes 18 clinical oncology studies using tumor DNA (mostly from formalin fixed, paraffin-embedded tissue). The overall variant sensitivity was 96.9% ($n = 428$) and specificity 97.1% ($n = 3{,}080$). The amplicon lengths were generally shorter than in genetic studies, usually varying between 60 and 250 bp. The HR-1™ was the most commonly cited instrument, followed by the Rotor-Gene® 6000, the LightScanner®, and the LightCycler® 480. LCGreen® I was the most common dye, followed by SYTO® 9 and LCGreen®+. The minimum percentage of variant that could be detected depended on the target, but ranged from 0.1 to 12%.

Common heterozygous variants of no clinical significance are a problem for any scanning method. High-resolution melting can identify these variants and eliminate them from consideration by three different methods. First, the melting curves of most heterozygous variants are different from each other (Graham *et al*., 2005). If common variants are identified by screening a normal population, then only those curves different from the normal variants need to be considered. The risk of two heterozygous variants having the same melting curve is about 7% (Montgomery *et al*., 2007a), so this method is not foolproof. Confirmation of common heterozygotes can be obtained by either unlabeled probes (Vandersteen *et al*., 2007) or by small amplicon melting (Dobrowolski *et al*., 2007b).

15.3.1 Applications of Variant Heteroduplex Scanning in Molecular Diagnostics

Heterozygote scanning by high-resolution melting has been used on numerous genetic targets (Table 15.1) and cancer tissues (Table 15.2). Some disease genes, such as *BRCA1/2*, *CFTR*, *F8*, *c-kit*, *EGFR*, and *TP* have been studied by more than one group. The method is favored by most as an inexpensive, rapid, closed-tube alternative to more complex scanning approaches. It is more sensitive than single-stranded conformational polymorphism or heteroduplex analysis (Highsmith *et al*., 1999a, b). Comparison against denaturing high-performance liquid chromatography (DHPLC) also favors melting analysis (Chou *et al*., 2005), although some heteroduplexes detected by dHPLC may not be detected by melting (Laurie *et al*., 2007).

TABLE 15.1 Heterozygote scanning for mutations leading to human genetic disorders.

Gene	Amplicons	Samples[a]	Length (bp)	Dye[b]	Instrument[c]	Detection sensitivity (%)		Specificity (%)	References
						Heterozygote (n)	Homozygote (n)	Heterozygote (n)	
ACVRL1	12	22	184–302	LCGreen® I	HR-1	96.4 (28)			Vandersteen *et al.* (2007)
APOB	2	57	156, 365	EvaGreen®	HR-1	100 (22)		100 (64)	Liyanage *et al.* (2008)
BRCA1/2	3	29	159–534	ResoLight®	LC480	100 (29)		100 (26)	Takano *et al.* (2008)
BRCA1/2	112	234	136–435	0.5X LCGreen®+	**LS/LC480**	**100/93.4** (212)	"most"	**98.7/98.6** (2464)	De Leeneer *et al.* (2008)
BRCA1/2	46	107	95–280	ResoLight®	LC480	100 (87)			de Juan *et al.* (2008)
CFTR	37	30	146–322	LCGreen®+	LS	100 (40)	75 (8)		Montgomery *et al.* (2007a)
CFTR	9	26	175–458	LCGreen® I	HR-1	100 (20)	50 (2)		Chou *et al.* (2005)
ENG	17	22	138–356	LCGreen® I	HR-1	96.6 (29)			Vandersteen *et al.* (2007)
EXT1/2	9	51	164–276	LCGreen®+	LS	100 (27)			Lonie *et al.* (2006)
F8	14	20	168–436	LCGreen®+	**LS/LC480**	**90/90** (20)			Laurie *et al.* (2007)
F8	52	384	155–250	ResoLight®	LC480	100 (15)	89 (28)	"high"	Lin *et al.* (2008b)
GJB1	4	32	185–458	0.6X LCGreen®+	LS	100 (23)		100 (105)	Kennerson *et al.* (2007)
IGF1	24	95	206–483	LCGreen®+	LS				Palles *et al.* (2008)
MCAD	13	18	129–223	LCGreen® I	HR-1	100 (33)			McKinney *et al.* (2004)
NF2	15	188	177–284	50μM SYTO® 9	RG6000	100 (16)			Sestini *et al.* (2008)
OTC	10	23	146–266	LCGreen®+	LS	100 (23)			Dobrowolski *et al.* (2007b)
PAH	13	95	154–283	LCGreen®+	LS	98.8 (164)	83.3 (12)		Dobrowolski *et al.* (2007a)
RET	6	80	114–235	LCGreen®+	HR-1	100 (30)			Margraf *et al.* (2006a)
SLC22A5	13	21	119–312	LCGreen® I	HR-1	100 (21)	30 (10)		Dobrowolski *et al.* (2005)

[a]*All samples were DNA isolated from fresh whole blood except for PAH (dried blood spots), RET (whole genome amplified), and SLC22A5 (fibroblasts and blood).*

[b]*All dye concentrations were 1X unless shown otherwise.*

[c]*When two instruments were compared in the same study, the entries are shown in bold type. Instrument abbreviations and manufacturers are HR-1™ (Idaho Technology), LC480 (Roche Applied Science), LS (LightScanner®, Idaho Technology), and RG6000 (Rotor-Gene® 6000 series, Corbett Life Science).*

TABLE 15.2 Variant scanning of human tumors[a].

Gene	Sample type[b]	Amplicons	Samples	Length (bp)	Dye[c]	Instrument[d]	Variant detection (%)		Minimum variant (%)	References
							Sensitivity *(n)*	Specificity *(n)*		
ABL1	Blood (cDNA)	4	101	221–241	LCGreen® I	RG6000	98.0 (51)	100 (31)	5	Polakova *et al.* (2008)
BRAF	Melanoma (FFPE)	2	90	190, 250	LCGreen® I	HR-1	96.5 (43)	97.9 (94)		Willmore-Payne *et al.* (2005)
c-kit	Melanoma (FFPE)	4	29	170–235	LCGreen® I	HR-1	100 (2)	100 (114)		Willmore-Payne *et al.* (2005)
c-kit	GIST (FFPE)	4	96	170–235	LCGreen® I	HR-1				Holden *et al.* (2007)
c-kit	Seminoma (FFPE)	2	22	170, 219	LCGreen® I	HR-1	100 (5)	100 (17)		Willmore-Payne *et al.* (2006a)
c-kit	GIST (FFPE)	4	29	170–235	LCGreen® I	HR-1	100 (18)	100 (56)		Willmore *et al.* (2004)
KRAS	Lung (FFPE)	1	200	92	5 μM SYTO9®	RG6000	100 (25)	100 (25)		Do *et al.* (2008)
KRAS	Lung (FFPE)	2	30	**92/189**	5 μM SYTO9®	RG6000	100/100 (9)	100/100 (21)	**5–6/10–12**	Krypuy *et al.* (2006)
EGFR	Lung (FFPE)	5	200	121–250	5 μM SYTO9®	RG6000	100 (73)	92.4 (927)		Do *et al.* (2008)
EGFR	Lung FNA (FFPE)	4	11	186–248	LCGreen® I	HR-1				Smith *et al.* (2008)
EGFR	Lung **(FFPE/ MFPE/slide)**	2	212	61, 70	LCGreen® I	HR-1	92/97/88 (**37/37/16**)	100/100/100 (**89/95/40**)		Takano *et al.* (2007)
EGFR	Lung (Slide)	2	36	61, 70	LCGreen® I	HR-1	90 (19)	100 (15)	0.1, 10	Nomoto *et al.*, 2006
EGFR1/2	SCC (FFPE)	6	24	186–252	LCGreen® I	HR-1	100 (13)	100 (131)		Willmore-Payne *et al.* (2006b)
EGFR1/2	Lung (FFPE)	6	39	186–252	LCGreen® I	HR-1	96.3 (27)	99.8 (207)		Willmore-Payne *et al.* (2006c)
PDGFRA	GIST (FFPE)	2	96	200, 235	LCGreen® I	HR-1				Holden *et al.* (2007)
RASEF	Melanoma (F)	16	46		LCGreen®	LS				Maat *et al.* (2008)
TP53	Breast **(F/ WGA/FFPE)**	21	74	80–150	LCGreen®+	LC480-384	**100/86/86 (12/7/7)**	**99.2/95.4/100 (471/287/287)**	0.5	Bastien *et al.* (2008)
TP53	**Ovary/Breast** (F)	5	40	136–245	5 μM SYTO9®	RG6000	**100/100 (20/7)**	**98.8/100 (80/93)**	5	Krypuy *et al.* (2007)

[a]When more than one sample type or amplicon length were compared in the same study, the entries are shown in bold type.

[b]Abbreviations include FFPE (formalin fixed, paraffin embedded), GIST (gastrointestinal stromal tumor), MFPE (methanol fixed, paraffin embedded), F (frozen), WGA (whole genome amplified from DNA isolated from frozen tissue).

[c]All dye concentrations were 1X unless shown otherwise.

[d]Instrument abbreviations and manufacturers were RG6000 (Rotor-Gene® 6000 series, Corbett Life Science), HR-1™ (Idaho Technology), LS (LightScanner®, Idaho Technology), and LC480-384 (LightCycler®, Roche Applied Science).

In addition to heterozygote scanning of diploid organisms, haploid microbes can be studied by mixing with a known sample. For example, the species of bacteria can be determined using the 16S ribosomal gene (Cheng *et al*., 2006), the antibiotic resistance of mycobacteria can be assessed using *rpoB* (Hoek *et al*., 2008) and influenza A can be subtyped using the *M* gene (Lin *et al*., 2008a). Another interesting use of mixing is to locate RNA editing sites by mixing cDNA with DNA, as demonstrated in *Arabidopsis* (Chateigner-Boutin and Small, 2007).

15.3.1.1 Cystic Fibrosis Variant Scanning

Cystic fibrosis is an autosomal recessive disorder which is caused by mutations in *CFTR* affecting the exocrine glands of the lungs, pancreas, intestines, and liver (Riordan *et al*., 1989). Over 1,600 variants of the *CFTR* gene have been described, each with varying frequencies and distribution across populations (Davies *et al*., 2007; van Baal *et al*., 2007). Different variants in the gene have varying effects on *CFTR* function and result in different phenotypes of the disease.

Complete analysis of all 27 exons of *CFTR*, including all American College of Medical Genetics recommended variants and eight additional disease causing variants, was demonstrated by Montgomery and coworkers (2007a). Initially, common variants were identified in 96 Caucasian blood donors and correlated with their melting patterns. Then, 30 blinded samples enriched for disease variants were used to assess the sensitivity of the technique. Scanning detected all 40 disease-associated heterozygotes for a sensitivity of 100%. However, 47 benign heterozygotes were also detected. By considering the melting patterns of the six most common heterozygotes, 45 of these 47 were matched to benign variants and did not require sequencing or genotyping. Most (6 out of 8) homozygous variants were also distinguished from wild type without mixing or genotyping.

15.4 CONCLUSIONS

High-resolution melting is a new technique for genotyping, variant scanning, and sequence matching. New dyes and high-resolution instruments allow a great deal of information to be extracted from the simple melting of DNA. Dyes that detect heteroduplexes are critical for scanning and genotyping applications, but not for methylation analysis or repeat typing. Scanning and small amplicon melting methods depend strongly on instrument resolution, while genotyping with unlabeled probes or snapback primers can be performed adequately on standard instrumentation. These methods are fast, affordable, and simple. As they gain wider acceptance, high-resolution melting is destined to become more common in research and clinical diagnostic laboratories.

REFERENCES

Bastien, R., Lewis, T.B., Hawkes, J.E., Quackenbush, J.F., Robbins, T.C., Palazzo, J., Perou, C.M., and Bernard, P.S. (2008). High-throughput amplicon scanning of the TP53 gene in breast cancer using high-resolution fluorescent melting curve analyses and automatic mutation calling. *Hum. Mutat.* 29, 757–764.

Bernard, P.S., Ajioka, R.S., Kushner, J.P., and Wittwer, C.T. (1998). Homogeneous multiplex genotyping of hemochromatosis mutations with fluorescent hybridization probes. *Am. J. Pathol.* 153, 1055–1061.

Brinson, E.C., Adriano, T., Bloch, W., Brown, C.L., Chang, C.C., Chen, J., Eggerding, F.A., Grossman, P.D., Iovannisci, D.M., Madonik, A.M., Sherman, D.G., Tam, R.W., Winn-Deen, E.S., Woo, S.L., Fung, S., and Iovannisci, D.A. (1997). Introduction to PCR/OLA/SCS, a multiplex DNA test, and its application to cystic fibrosis. *Genet. Test.* 1, 61–68.

Chateigner-Boutin, A.L., and Small, I. (2007). A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res.* 35, e114.

Cheng, J.C., Huang, C.L., Lin, C.C., Chen, C.C., Chang, Y.C., Chang, S.S., and Tseng, C.P. (2006). Rapid detection and identification of clinically important bacteria by high-resolution melting analysis after broad-range ribosomal RNA real-time PCR. *Clin. Chem.* 52, 1997–2004.

Chou, L.S., Lyon, E., and Wittwer, C.T. (2005). A comparison of high-resolution melting analysis with denaturing high-performance liquid chromatography for mutation scanning: cystic fibrosis transmembrane conductance regulator gene as a model. *Am. J. Clin. Pathol.* 124, 330–338.

Crockett, A.O., and Wittwer, C.T. (2001). Fluorescein-labeled oligonucleotides for real-time PCR: using the inherent quenching of deoxyguanosine nucleotides. *Anal. Biochem.* 290, 89–97.

Dahl, C., and Guldberg, P. (2007). High-resolution melting for accurate assessment of DNA methylation. *Clin. Chem.* 53, 1877–1878.

Dames, S., Margraf, R.L., Pattison, D.C., Wittwer, C.T., and Voelkerding, K.V. (2007a). Characterization of aberrant melting peaks in unlabeled probe assays. *J. Mol. Diagn.* 9, 290–296.

Dames, S., Pattison, D.C., Bromley, L.K., Wittwer, C.T., and Voelkerding, K.V. (2007b). Unlabeled probes for the detection and typing of herpes simplex virus. *Clin. Chem.* 53, 1847–1854.

Davies, J.C., Alton, E.W.F.W., and Bush, A. (2007). Cystic fibrosis. *BMJ* 335, 1255–1259.

de Juan, I., Esteban, E., Palanca, S., Barragan, E., and Bolufer, P. (2008). High-resolution melting analysis for rapid screening of BRCA1 and BRCA2 Spanish mutations. *Breast Cancer Res. Treat*. in press.

De Leeneer, K., Coene, I., Poppe, B., De Paepe, A., and Claes, K. (2008). Rapid and sensitive detection of BRCA1/2 mutations in a diagnostic setting: comparison of two high-resolution melting platforms. *Clin. Chem.* 54, 982–989.

Do, H., Krypuy, M., Mitchell, P., Fox, S., and Dobrovic, A. (2008). High resolution melting analysis for rapid and sensitive EGFR and KRAS mutation detection in formalin fixed paraffin embedded biopsies. *BMC Cancer* 8, 142.

Dobrowolski, S.F., McKinney, J.T., Amat di San Filippo, C., Giak Sim, K., Wilcken, B., and Longo, N. (2005). Validation of dye-binding/high-resolution thermal denaturation for the identification of mutations in the SLC22A5 gene. *Hum. Mutat.* 25, 306–313.

Dobrowolski, S.F., Ellingson, C., Coyne, T., Grey, J., Martin, R., Naylor, E.W., Koch, R., and Levy, H.L. (2007a). Mutations in the phenylalanine hydroxylase gene identified in 95 patients with phenylketonuria

using novel systems of mutation scanning and specific genotyping based upon thermal melt profiles. *Mol. Genet. Metab.* 91, 218–227.

Dobrowolski, S.F., Ellingson, C.E., Caldovic, L., and Tuchman, M. (2007b). Streamlined assessment of gene variants by high resolution melt profiling utilizing the ornithine transcarbamylase gene as a model system. *Hum. Mutat.* 28, 1133–1140.

Dufresne, S.D., Belloni, D.R., Wells, W.A., and Tsongalis, G.J. (2006). BRCA1 and BRCA2 mutation screening using SmartCycler II high-resolution melt curve analysis. *Arch. Pathol. Lab. Med.* 130, 185–187.

Dujols, V.E., Kusukawa, N., McKinney, J.T., Dobrowolski, S.F., and Wittwer, C.T. (2006). High-resolution melting analysis for scanning and genotyping. In *Real-Time PCR* (Dorak, M.T., ed.), pp. 157–171. Garland Science, New York.

Erali, M., Pounder, J.I., Woods, G.L., Petti, C.A., and Wittwer, C.T. (2006). Multiplex single-color PCR with amplicon melting analysis for identification of Aspergillus species. *Clin. Chem.* 52, 1443–1445.

Erali, M., Palais, B., and Wittwer, C.T. (2008a). SNP genotyping by unlabeled probe melting analysis. In *Molecular Beacons – Signaling Nucleic Acid Probes* (Seitz, O., and Marx, A., eds), pp. 199–206. Humana Press, Totowa.

Erali, M., Voelkerding, K.V., and Wittwer, C.T. (2008b). High resolution melting applications for clinical laboratory medicine. *Exp. Mol. Pathol.* 85, 50–58.

Fortini, D., Ciammaruconi, A., De Santis, R., Fasanella, A., Battisti, A., D'Amelio, R., Lista, F., Cassone, A., and Carattoli, A. (2007). Optimization of high-resolution melting analysis for low-cost and rapid screening of allelic variants of Bacillus anthracis by multiple-locus variable-number tandem repeat analysis. *Clin. Chem.* 53, 1377–1380.

Germer, S., and Higuchi, R. (1999). Single-tube genotyping without oligonucleotide probes. *Genome Res.* 9, 72–78.

Giglio, S., Monis, P.T., and Saint, C.P. (2003). Demonstration of preferential binding of SYBR Green I to specific DNA fragments in real-time multiplex PCR. *Nucleic Acids Res.* 31, e136.

Graham, R., Liew, M., Meadows, C., Lyon, E., and Wittwer, C.T. (2005). Distinguishing different DNA heterozygotes by high-resolution melting. *Clin. Chem.* 51, 1295–1298.

Gundry, C.N., Vandersteen, J.G., Reed, G.H., Pryor, R.J., Chen, J., and Wittwer, C.T. (2003). Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. *Clin. Chem.* 49, 396–406.

Gundry, C.N., Dobrowolski, S.F., Martin, Y.R., Robbins, T.C., Nay, L.M., Boyd, N., Coyne, T., Wall, M.D., Wittwer, C.T., and Teng, D.H. (2008). Base-pair neutral homozygotes can be discriminated by calibrated high-resolution melting of small amplicons. *Nucleic Acids Res.* 36, 3401–3408.

Herrmann, M.G., Durtschi, J.D., Bromley, L.K., Wittwer, C.T., and Voelkerding, K.V. (2006). Amplicon DNA melting analysis for mutation scanning and genotyping: cross-platform comparison of instruments and dyes. *Clin. Chem.* 52, 494–503.

Herrmann, M.G., Durtschi, J.D., Bromley, L.K., Wittwer, C.T., and Voelkerding, K.V. (2007a). Instrument comparison for heterozygote scanning of single and double heterozygotes: a correction and extension of Herrmann *et al*., 2006. *Clin. Chem.* 52, 494–503. *Clin. Chem.* 53, 150–152.

Herrmann, M.G., Durtschi, J.D., Wittwer, C.T., and Voelkerding, K.V. (2007b). Expanded instrument comparison of amplicon DNA melting analysis for mutation scanning and genotyping. *Clin. Chem.* 53, 1544–1548.

Highsmith, W.E., Jr., Jin, Q., Nataraj, A.J., O'Connor, J.M., Burland, V.D., Baubonis, W.R., Curtis, F.P., Kusukawa, N., and Garner, M.M. (1999a). Use of a DNA toolbox for the characterization of mutation scanning methods. I: Construction of the toolbox and evaluation of heteroduplex analysis. *Electrophoresis* 20, 1186–1194.

Highsmith, W.E., Jr., Nataraj, A.J., Jin, Q., O'Connor, J.M., El-Nabi, S.H., Kusukawa, N., and Garner, M.M. (1999b). Use of DNA toolbox for the characterization of mutation scanning methods. II: Evaluation of single-strand conformation polymorphism analysis. *Electrophoresis* 20, 1195–1203.

Hill, C.E., Duncan, A., Wirth, D., and Nolte, F.S. (2006). Detection and identification of cytochrome P-450 2C9 alleles *1, *2, and *3 by high-resolution melting curve analysis of PCR amplicons. *Am. J. Clin. Pathol.* 125, 584–591.

Hoek, K.G.P., Pittius, N.C.G.v., Moolman-Smook, H., Carelse-Tofa, K., Jordaan, A., Spuy, G.D.v.d., Streicher, E., Victor, T.C., Helden, P.D.v., and Warren, R.M. (2008). Fluorometric assay for testing rifampin susceptibility of mycobacterium tuberculosis complex. *J. Clin. Microbiol.* 46, 1369–1373.

Holden, J.A., Willmore-Payne, C., Coppola, D., Garrett, C.R., and Layfield, L.J. (2007). High-resolution melting amplicon analysis as a method to detect c-kit and platelet-derived growth factor receptor alpha activating mutations in gastrointestinal stromal tumors. *Am. J. Clin. Pathol.* 128, 230–238.

Hung, C.C., Lee, C.N., Chang, C.H., Jong, Y.J., Chen, C.P., Hsieh, W.S., Su, Y.N., and Lin, W.L. (2008). Genotyping of the G1138A mutation of the FGFR3 gene in patients with achondroplasia using high-resolution melting analysis. *Clin. Biochem.* 41, 162–166.

Jeffery, N., Gasser, R.B., Steer, P.A., and Noormohammadi, A.H. (2007). Classification of Mycoplasma synoviae strains using single-strand conformation polymorphism and high-resolution melting-curve analysis of the vlhA gene single-copy region. *Microbiology* 153, 2679–2688.

Jones, P.A., and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* 3, 415–428.

Kennerson, M.L., Warburton, T., Nelis, E., Brewer, M., Polly, P., De Jonghe, P., Timmerman, V., and Nicholson, G.A. (2007). Mutation scanning the GJB1 gene with high-resolution melting analysis: implications for mutation scanning of genes for Charcot-Marie-Tooth disease. *Clin. Chem.* 53, 349–352.

Kristensen, L.S., and Dobrovic, A. (2008). Direct genotyping of single nucleotide polymorphisms in methyl metabolism genes using probe-free high-resolution melting analysis. *Cancer Epidemiol. Biomarkers Prev.* 17, 1240–1247.

Kristensen, L.S., Mikeska, T., Krypuy, M., and Dobrovic, A. (2008). Sensitive melting analysis after real time-methylation specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. *Nucleic Acids Res.* 36, e42.

Krypuy, M., Newnham, G.M., Thomas, D.M., Conron, M., and Dobrovic, A. (2006). High resolution melting analysis for the rapid and sensitive detection of mutations in clinical samples: KRAS codon 12 and 13 mutations in non-small cell lung cancer. *BMC Cancer* 6, 295.

Krypuy, M., Ahmed, A., Etemadmoghadam, D., Hyland, S., deFazio, A., Fox, S., Brenton, J., Bowtell, D., and Dobrovic, A.Australian Ovarian Cancer Study (2007). High resolution melting for mutation scanning of TP53 exons 5–8. *BMC Cancer* 7, 168.

Laurie, A.D., Smith, M.P., and George, P.M. (2007). Detection of factor VIII gene mutations by high-resolution melting analysis. *Clin. Chem.* 53, 2211–2214.

Lay, M.J., and Wittwer, C.T. (1997). Real-time fluorescence genotyping of factor V Leiden during rapid-cycle PCR. *Clin. Chem.* 43, 2262–2267.

Lehmensiek, A., Sutherland, M.W., and McNamara, R.B. (2008). The use of high resolution melting (HRM) to map single nucleotide

polymorphism markers linked to a covered smut resistance gene in barley. *Theor. Appl. Genet.* 117, 721–728.

Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E., and Wittwer, C.T. (2004). Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clin. Chem.* 50, 1156–1164.

Liew, M., Nelson, L., Margraf, R., Mitchell, S., Erali, M., Mao, R., Lyon, E., and Wittwer, C.T. (2006). Genotyping of human platelet antigens 1 to 6 and 15 by high-resolution amplicon melting and conventional hybridization probes. *J. Mol. Diagn.* 8, 97–104.

Liew, M., Seipp, M., Durtschi, J., Margraf, R.L., Dames, S., Erali, M., Voelkerding, K., and Wittwer, C.T. (2007). Closed-tube SNP genotyping without labeled probes/a comparison between unlabeled probe and amplicon melting. *Am. J. Clin. Pathol.* 127, 1–8.

Lin, J.-H., Tseng, C.-P., Chen, Y.-J., Lin, C.-Y., Chang, S.-S., Wu, H.-S., and Cheng, J.-C. (2008a). Rapid differentiation of influenza a virus subtypes and genetic screening for virus variants by high-resolution melting analysis. *J. Clin. Microbiol.* 46, 1090–1097.

Lin, S.Y., Su, Y.N., Hung, C.C., Tsay, W., Chiou, S.S., Chang, C.T., Ho, H.N., and Lee, C.N. (2008b). Mutation spectrum of 122 hemophilia A families from Taiwanese population by LD-PCR, DHPLC, multiplex PCR and evaluating the clinical application of HRM. *BMC Med. Genet.* 9, 53.

Lipsky, R.H., Mazzanti, C.M., Rudolph, J.G., Xu, K., Vyas, G., Bozak, D., Radel, M.Q., and Goldman, D. (2001). DNA melting analysis for detection of single nucleotide polymorphisms. *Clin. Chem.* 47, 635–644.

Liyanage, K.E., Hooper, A.J., Defesche, J.C., Burnett, J.R., and van Bockxmeer, F.M. (2008). High-resolution melting analysis for detection of familial ligand-defective apolipoprotein B-100 mutations. *Ann. Clin. Biochem.* 45, 170–176.

Lonie, L., Porter, D.E., Fraser, M., Cole, T., Wise, C., Yates, L., Wakeling, E., Blair, E., Morava, E., Monaco, A.P., and Ragoussis, J. (2006). Determination of the mutation spectrum of the EXT1/EXT2 genes in British Caucasian patients with multiple osteochondromas, and exclusion of six candidate genes in EXT negative cases. *Hum. Mutat.* 27, 1160.

Maat, W., Beiboer, S.H., Jager, M.J., Luyten, G.P., Gruis, N.A., and van der Velden, P.A. (2008). Epigenetic regulation identifies RASEF as a tumor-suppressor gene in uveal melanoma. *Invest. Ophthalmol. Vis. Sci.* 49, 1291–1298.

Mackay, J.F., Wright, C.D., and Bonfiglioli, R.G. (2008). A new approach to varietal identification in plants by microsatellite high resolution melting analysis: application to the verification of grapevine and olive cultivars. *Plant Methods* 4, 8.

Mao, F., Leung, W.-Y., and Xin, X. (2007). Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnol.* 7, 76.

Margraf, R.L., Mao, R., Highsmith, W.E., Holtegaard, L.M., and Wittwer, C.T. (2006a). Mutation scanning of the RET protooncogene using high-resolution melting analysis. *Clin. Chem.* 52, 138–141.

Margraf, R.L., Mao, R., and Wittwer, C.T. (2006b). Masking selected sequence variation by incorporating mismatches into melting analysis probes. *Hum. Mutat.* 27, 269–278.

Margraf, R.L., Mao, R., Highsmith, W.E., Holtegaard, L.M., and Wittwer, C.T. (2007). RET Proto-oncogene genotyping using unlabeled probes, the masking technique, and amplicon high-resolution melting analysis. *J. Mol. Diagn.* 9, 184–196.

Marziliano, N., Bevilacqua, E., Pirulli, D., Span, A., Amoroso, A., and Crovella, S. (2000). Single tube melting temperature assay for rapid and sensitive detection of the most frequent hemocromatosis mutations, C282Y and H63D. *Haematologica* 85, 990–991.

McKinney, J.T., Longo, N., Hahn, S.H., Matern, D., Rinaldo, P., Strauss, A.W., and Dobrowolski, S.F. (2004). Rapid, comprehensive screening of the human medium chain acyl-CoA dehydrogenase gene. *Mol. Genet. Metab.* 82, 112–120.

Montgomery, J., Wittwer, C.T., Kent, J.O., and Zhou, L. (2007a). Scanning the cystic fibrosis transmembrane conductance regulator gene using high-resolution DNA melting analysis. *Clin. Chem.* 53, 1891–1898.

Montgomery, J., Wittwer, C.T., Palais, R.A., and Zhou, L. (2007b). Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. *Nat. Protoc.* 2, 59–66.

Nellaker, C., Wallgren, U., and Karlsson, H. (2007). Molecular beacon-based temperature control and automated analyses for improved resolution of melting temperature analysis using SYBR I green chemistry. *Clin. Chem.* 53, 98–103.

Nomoto, K., Tsuta, K., Takano, T., Fukui, T., Fukui, T., Yokozawa, K., Sakamoto, H., Yoshida, T., Maeshima, A.M., Shibata, T., Furuta, K., Ohe, Y., and Matsuno, Y. (2006). Detection of EGFR mutations in archived cytologic specimens of non-small cell lung cancer using high-resolution melting analysis. *Am. J. Clin. Pathol.* 126, 608–615.

Odell, I.D., Cloud, J.L., Seipp, M., and Wittwer, C.T. (2005). Rapid species identification within the Mycobacterium chelonae-abscessus group by high-resolution melting analysis of hsp65 PCR products. *Am. J. Clin. Pathol.* 123, 96–101.

Palais, R.A., Liew, M.A., and Wittwer, C.T. (2005). Quantitative heteroduplex analysis for single nucleotide polymorphism genotyping. *Anal. Biochem.* 346, 167–175.

Palles, C., Johnson, N., Coupland, B., Taylor, C., Carvajal, J., Holly, J., Fentiman, I.S., Silva Idos, S., Ashworth, A., Peto, J., and Fletcher, O. (2008). Identification of genetic variants that influence circulating IGF1 levels: a targeted search strategy. *Hum. Mol. Genet.* 17, 1457–1464.

Paul, T.M., Steven, G., and Christopher, P.S. (2005). Comparison of SYTO9 and SYBR Green I for real-time polymerase chain reaction and investigation of the effect of dye concentration on amplification and DNA melting curve analysis. *Anal. Biochem.* 340, 24–34.

Pirulli, D., Boniotto, M., Puzzer, D., Spano, A., Amoroso, A., and Crovella, S. (2000). Flexibility of melting temperature assay for rapid detection of insertions, deletions, and single-point mutations of the AGXT gene responsible for type 1 primary hyperoxaluria. *Clin. Chem.* 46, 1842–1844.

Polakova, K.M., Lopotova, T., Klamova, H., and Moravcova, J. (2008). High-resolution melt curve analysis: initial screening for mutations in BCR-ABL kinase domain. *Leuk. Res.* 32, 1236–1243.

Pornprasert, S., Phusua, A., Suanta, S., Saetung, R., and Sanguansermsri, T. (2008). Detection of alpha-thalassemia-1 Southeast Asian type using real-time gap-PCR with SYBR Green1 and high resolution melting analysis. *Eur. J. Haematol.* 80, 510–514.

Price, E.P., Smith, H., Huygens, F., and Giffard, P.M. (2007). High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of Campylobacter jejuni. *Appl. Environ. Microbiol.* 73, 3431–3436.

Rasmussen, J.P., Saint, C.P., and Monis, P.T. (2007). Use of DNA melting simulation software for in silico diagnostic assay design: targeting regions with complex melting curves and confirmation by real-time PCR using intercalating dyes. *BMC Bioinformatics* 8, 107.

Reed, G.H., and Wittwer, C.T. (2004). Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. *Clin. Chem.* 50, 1748–1754.

Reed, G.H., Kent, J.O., and Wittwer, C.T. (2007). High-resolution DNA melting analysis for simple and efficient molecular diagnostics. *Pharmacogenomics* 8, 597–608.

Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.L., Drum, M.L., Iannuzzi, M.C., Collins, F.S., and Chui, L.C. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245, 1066–1073.

Ririe, K.M., Rasmussen, R.P., and Wittwer, C.T. (1997). Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal. Biochem.* 245, 154–160.

Robinson, B.S., Monis, P.T., and Dobson, P.J. (2006). Rapid, sensitive, and discriminating identification of Naegleria spp. by real-time PCR and melting-curve analysis. *Appl. Environ. Microbiol.* 72, 5857–5863.

Sang, F., and Ren, J. (2006). Capillary electrophoresis of double-stranded DNA fragments using a new fluorescence intercalating dye EvaGreen. *J. Sep. Sci.* 29, 1275–1280.

Seipp, M.T., Durtschi, J.D., Liew, M.A., Williams, J., Damjanovich, K., Pont-Kingdon, G., Lyon, E., Voelkerding, K.V., and Wittwer, C.T. (2007). Unlabeled oligonucleotides as internal temperature controls for genotyping by amplicon melting. *J. Mol. Diagn.* 9, 284–289.

Seipp, M.T., Pattison, D., Durtschi, J.D., Jama, M., Voelkerding, K.V., and Wittwer, C.T. (2008). Quadruplex genotyping of F5, F2, and MTHFR variants in a single closed tube by high-resolution amplicon melting. *Clin. Chem.* 54, 108–115.

Sestini, R., Provenzano, A., Bacci, C., Orlando, C., Genuardi, M., and Papi, L. (2008). NF2 mutation screening by denaturing high-performance liquid chromatography and high-resolution melting analysis. *Genet. Test.* 12, 311–318.

Slinger, R., Bellfoy, D., Desjardins, M., and Chan, F. (2007). High-resolution melting assay for the detection of gyrA mutations causing quinolone resistance in Salmonella enterica serovars Typhi and Paratyphi. *Diagn. Microbiol. Infect. Dis.* 57, 455–458.

Smith, G.D., Chadwick, B.E., Willmore-Payne, C., and Bentz, J.S. (2008). Detection of epidermal growth factor receptor gene mutations in cytology specimens from patients with non-small cell lung cancer utilising high-resolution melting amplicon analysis. *J. Clin. Pathol.* 61, 487–493.

Snell, C., Krypuy, M., Wong, E.M., Loughrey, M.B., and Dobrovic, A. (2008). BRCA1 promoter methylation in peripheral blood DNA of mutation negative familial breast cancer patients with a BRCA1 tumour phenotype. *Breast Cancer Res.* 10, R12.

Stephens, A.J., Inman-Bamber, J., Giffard, P.M., and Huygens, F. (2008). High-resolution melting analysis of the spa repeat region of Staphylococcus aureus. *Clin. Chem.* 54, 432–436.

Takano, T., Ohe, Y., Tsuta, K., Fukui, T., Sakamoto, H., Yoshida, T., Tateishi, U., Nokihara, H., Yamamoto, N., Sekine, I., Kunitoh, H., Matsuno, Y., Furuta, K., and Tamura, T. (2007). Epidermal growth factor receptor mutation detection using high-resolution melting analysis predicts outcomes in patients with advanced non small cell lung cancer treated with gefitinib. *Clin. Cancer Res.* 13, 5385–5390.

Takano, E.A., Mitchell, G., Fox, S.B., and Dobrovic, A. (2008). Rapid detection of carriers with BRCA1 and BRCA2 mutations using high resolution melting analysis. *BMC Cancer* 8, 59.

Tedde, A., Bagnoli, S., Cellini, E., Nacmias, B., Piacentini, S., and Sorbi, S. (2007). No association between the LRRK2 G2019S mutation and Alzheimer's disease in Italy. *Cell. Mol. Neurobiol.* 27, 877–881.

Tedde, A., Laura Putignano, A., Bagnoli, S., Congregati, C., Milla, M., Sorbi, S., Genuardi, M., and Papi, L. (2008). Interleukin-10 promoter polymorphisms influence susceptibility to ulcerative colitis in a gender-specific manner. *Scand. J. Gastroenterol.* 43, 712–718.

Toi, C.S., and Dwyer, D.E. (2008). Differentiation between vaccine and wild-type varicella-zoster virus genotypes by high-resolution melt analysis of single nucleotide polymorphisms. *J. Clin. Virol.* 43, 18–24.

Van Baal, S., Kaimakis, P., Phommarinh, M., Koumbi, D., Cuppens, H., Riccardino, F., Macek, M., Jr., Scriver, C.R., and Patrinos, G.P. (2007). FINDbase: a relational database recording frequencies of genetic defects leading to inherited disorders worldwide. *Nucleic Acids Res.* 35, D690–D695.

Vandersteen, J.G., Bayrak-Toydemir, P., Palais, R.A., and Wittwer, C.T. (2007). Identifying common genetic variants by high-resolution melting. *Clin. Chem.* 53, 1191–1198.

Vaughn, C.P., and Elenitoba-Johnson, K.S. (2004). High-resolution melting analysis for detection of internal tandem duplications. *J. Mol. Diagn.* 6, 211–216.

von Ahsen, N., Oellerich, M., and Schutz, E. (2001). Limitations of genotyping based on amplicon melting temperature. *Clin. Chem.* 47, 1331–1332.

White, H.E., Hall, V.J., and Cross, N.C.P. (2007). Methylation-sensitive high-resolution melting-curve analysis of the SNRPN gene as a diagnostic screen for Prader-Willi and Angelman syndromes. *Clin. Chem.* 53, 1960–1962.

Willmore, C., Holden, J.A., Zhou, L., Tripp, S., Wittwer, C.T., and Layfield, L.J. (2004). Detection of c-kit-activating mutations in gastrointestinal stromal tumors by high-resolution amplicon melting analysis. *Am. J. Clin. Pathol.* 122, 206–216.

Willmore-Payne, C., Holden, J.A., Tripp, S., and Layfield, L.J. (2005). Human malignant melanoma: detection of BRAF- and c-kit-activating mutations by high-resolution amplicon melting analysis. *Hum. Pathol.* 36, 486–493.

Willmore-Payne, C., Holden, J.A., Chadwick, B.E., and Layfield, L.J. (2006a). Detection of c-kit exons 11- and 17-activating mutations in testicular seminomas by high-resolution melting amplicon analysis. *Mod. Pathol.* 19, 1164–1169.

Willmore-Payne, C., Holden, J.A., and Layfield, L.J. (2006b). Detection of EGFR- and HER2-activating mutations in squamous cell carcinoma involving the head and neck. *Mod. Pathol.* 19, 634–640.

Willmore-Payne, C., Holden, J.A., and Layfield, L.J. (2006c). Detection of epidermal growth factor receptor and human epidermal growth factor receptor 2 activating mutations in lung adenocarcinoma by high-resolution melting amplicon analysis: correlation with gene copy number, protein expression, and hormone receptor expression. *Hum. Pathol.* 37, 755–763.

Wittwer, C.T., Herrmann, M.G., Moss, A.A., and Rasmussen, R.P. (1997a). Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques* 22, 130–131, 134–138.

Wittwer, C.T., Ririe, K.M., Andrew, R.V., David, D.A., Gundry, R.A., and Balis, U.J. (1997b). The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques* 22, 176–181.

Wittwer, C.T., Reed, G.H., Gundry, C.N., Vandersteen, J.G., and Pryor, R.J. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin. Chem.* 49, 853–860.

Wittwer, C.T., and Kusukawa, N. (2004). Real-time PCR. In *Diagnostic Molecular Microbiology; Principles and Applications* (Persing, D.H., Tenover, F.C., Versalovic, J., Tang, Y.W., Unger, E.R., Relman, D.A., and White, T.J., eds), pp. 71–84. ASM Press, Washington DC.

Wittwer, C.T., Zhou, L., Dujols, V.E., Holden, J.A., and Willmore-Payne, C. (2008). Nucleic acid melting analysis with saturation dyes. US Patent 7387887.

Wojdacz, T.K., and Dobrovic, A. (2007). Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Res.* 35, e41.

Wojdacz, T.K., Dobrovic, A., and Algar, E.M. (2008). Rapid detection of methylation change at H19 in human imprinting disorders using methylation-sensitive high-resolution melting. *Hum. Mutat*. in press.

Worm, J., Aggerholm, A., and Guldberg, P. (2001). In-tube DNA methylation profiling by fluorescence melting curve analysis. *Clin. Chem.* 47, 1183–1189.

Zhou, L., Myers, A.N., Vandersteen, J.G., Wang, L., and Wittwer, C.T. (2004a). Closed-tube genotyping with unlabeled oligonucleotide probes and a saturating DNA dye. *Clin. Chem.* 50, 1328–1335.

Zhou, L., Vandersteen, J., Wang, L., Fuller, T., Taylor, M., Palais, B., and Wittwer, C.T. (2004b). High-resolution DNA melting curve analysis to establish HLA genotypic identity. *Tissue Antigens* 64, 156–164.

Zhou, L., Wang, L., Palais, R., Pryor, R., and Wittwer, C.T. (2005). High-resolution DNA melting analysis for simultaneous mutation scanning and genotyping in solution. *Clin. Chem.* 51, 1770–1777.

Zhou, L., Errigo, R.J., Lu, H., Poritz, M.A., Seipp, M.T., and Wittwer, C.T. (2008). Snapback primer genotyping with a saturating DNA dye and melting analysis. *Clin. Chem*. 54, 1648–1656.

This page intentionally left blank

Chapter 16

DNA Microarrays and Genetic Testing

Lars Dyrskjøt, Karina Dalsgaard-Sørensen, Marie Stampe-Ostenfeld, Karin Birkenkamp-Demtroder, Kasper Thorsen, Claus L. Andersen, Mogens Kruhøffer, Jens L. Jensen and Torben F. Ørntoft

Department of Molecular Medicine, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, Aarhus N, Denmark

16.1 INTRODUCTION

The completion of the Human Genome Project and the simultaneous advances in microarray technology have had a large impact on the field of molecular biology and especially on cancer research. It is now possible to make systematic genome-wide searches for genes being differentially expressed, amplified, or deleted during cancer development. Earlier, this search was limited to only small numbers of candidate targets using, for example, Northern blotting assays. The use of microarray-based methods has boosted the assay capacity to include several thousand measurements. The genome-wide searches have identified several genes that can serve as diagnostic or prognostic molecular disease markers or as potential therapeutic targets.

The use of DNA microarray technology has revolutionized molecular biology since its introduction in the mid-1990s. Using DNA microarray analysis, it is now possible to make complete genome-wide screenings for, e.g., molecular disease markers. Especially in the field of cancer disease classification and outcome prediction, notable findings have been reported by large-scale gene expression profiling. Basically, a DNA microarray is an orderly arrangement of usually thousands of defined DNA molecules immobilized on a small glass surface. By convention the immobilized DNA molecules are called the probes and the sample under investigation is the target. Several microarray types exist for monitoring gene and microRNA (miRNA) expression, single nucleotide polymorphisms (SNPs), loss or gain of genomic material, and detection of splice variants. Furthermore, microarrays for resequencing of known gene sequences exist together with small interfering RNA (siRNA) microarrays for high-throughput functional testing.

This chapter introduces the most commonly used microarray types and platforms. The most novel findings using microarrays are described together with a summary of the future potentials in the use of microarrays in basic research and in a clinical setting.

16.2 DNA MICROARRAYS AND GENE EXPRESSION PROFILING

Microarrays for measuring gene expression levels are the most commonly used microarrays, and several important findings using this technology recently have been reported. The following sections describe the technology and analysis methods used for generating useful data from microarray experiments. Furthermore, the most important findings in the field of cancer research are described together with descriptions of the potential of using expression profiling for unraveling the affected molecular pathways, for example during disease progression.

16.2.1 DNA Microarray Technology

Several technologies and protocols exist for monitoring gene expression using microarray technology. In principle, labeled transcripts isolated from biological samples are hybridized to the DNA microarray probes for determination of the transcript abundance or relative expressions. In conventionally used protocols, total RNA from the biological sample is extracted and reverse transcribed into cDNA. Subsequently, an *in vitro* transcription of the cDNA is carried out with incorporation of modified nucleotides for later coupling with fluorescent molecules. In other protocols, modified nucleotides are incorporated directly into the cDNA product. This procedure, however, is limited in the amount of generated target due to the lack of an amplification step, which may be a problem when working with small amounts of starting material. The labeled target is hybridized to the DNA microarray slide for several hours to allow hybridization of the sample target to the microarray probes. Following the hybridization procedure and intensive washing of the slide to remove excess target molecules, scanning of the DNA microarray identifies the DNA probe hybridization levels, which reflects the gene expression levels in the samples investigated.

The DNA probes used for expression microarrays are either oligonucleotides or long PCR products amplified from cDNA clones. The advantage of using oligonucleotide probes is that they can be designed to ensure minimal cross-hybridization to other transcripts and, furthermore, that each gene will be covered with several probes. In commercially available systems (e.g. Affymetrix GeneChips) each gene is typically covered by up to 20–25 mer oligonucleotides and 20–25 mer oligonucleotides with a mismatched base in the middle position for measuring the amount of non-specific hybridization. The use of more than one probe for representing each gene reduces the problem with the non-functional probes, which can be a major obstacle when using only a single oligonucleotide probe per gene. Microarrays are manufactured by several companies (Agilent Technologies, Affymetrix, Applied Biosystems, Illumina).

When using PCR products from cDNA clones as probes it is more difficult to minimize cross-hybridization to other transcripts because of the probe length, which may be as long as 500–800 bp. Oligonucleotide and PCR probes are spotted directly on a glass surface using precision robotics and very accurate spotting pins, which are able to apply very small amounts of probe solution to a large number of glass slides.

Spotting of the probes introduces probe morphology variation between different microarray slides, and consequently these arrays typically are hybridized with both the biological sample under investigation and a common reference sample. Each sample is labeled with different fluorescent molecules (Cy3 and Cy5 are often used) and the microarray slide is scanned at two different wavelengths to obtain measurements of the transcript hybridization abundance of each sample. The use of a common reference sample makes comparisons between different slides with different probe morphology possible. The two-sample hybridization technique provides relative expression measures (ratios), which reflect on the relative expression of the genes in the two samples. However, when using commercially available microarrays or non-contact printed microarray slides the probes have highly uniform shapes, and sizes and it is possible to hybridize a single sample to each microarray slide and in this way obtain direct measures of the gene expression levels. Figure 16.1 illustrates the difference between the one- and two-sample microarray systems.

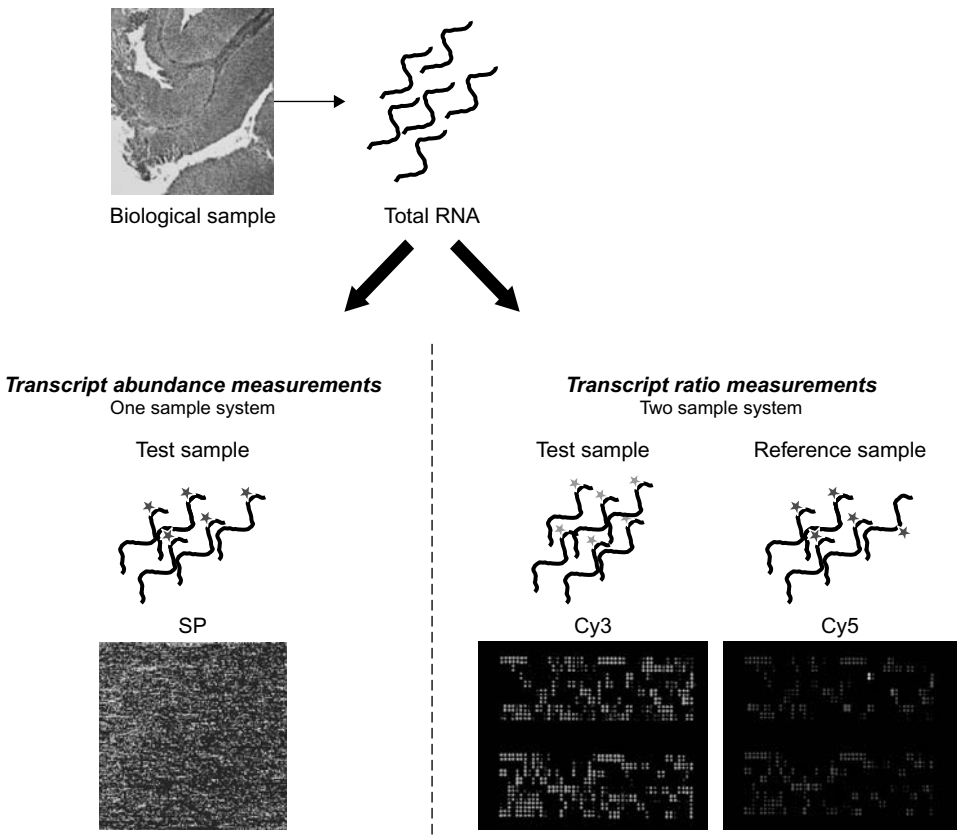


FIGURE 16.1 One-sample and two-sample microarray systems. In both systems total RNA is extracted from the biological sample under investigation. In one-sample systems like the Affymetrix GeneChip systems (left side of the figure) the total RNA is reverse transcribed into cDNA, which is transcribed into cRNA with incorporation of biotinylated nucleotides. These biotinylated nucleotides are hybridized to the array and subsequently coupled to streptavidin phycoerythrin (SP) conjugates. In this way, direct measurement of transcript abundance is performed. In traditional two-sample systems, the test and reference RNA is reverse transcribed into cDNA. The cDNA is transcribed into cRNA with the incorporation of aminoallyl-linked nucleotides. The test cRNA is coupled to Cy3 and reference cRNA is coupled to Cy5. Fluorescence measurements at the two different wavelengths gives the transcript ratios between the test and the references samples.

In addition to the profiling of mRNA transcipts, microarray technology is now also feasible for the profiling of microRNAs (miRNAs). Expression analysis of these noncoding 21–35 nucleotide molecules is challenging due to their small size and low abundance. Oligonucleotide microarray analysis is the most common high-throughput method for genome-wide analysis of miRNA expression. Several companies such as Rosetta Genetics Ltd, Illumina, Agilent Technologies, Geniom, and Exiqon offer microarray platforms for miRNA analysis. Usually, DNA oligonucleotides representing known human miRNAs are printed onto glass slides and subsequently hybridized to labeled miRNAs. After stringency washes, the slides are scanned and the array images analyzed. Some platforms employ oligonucleotides containing locked nucleic acid (LNA) nucleotides for improved miRNA hybridization (e.g. Exiqon). Other methods involve the use of bead-based flow cytometric technique (Lu *et al.*, 2005) or large-scale quantitative real-time PCR (Schmittgen *et al.*, 2008).

16.2.2 Normalization Strategies

The raw data obtained from a microarray experiment consists of an image of light intensities where some of the pixels are the spots (foreground) to which the labeled transcripts are hybridized, and where the remaining pixels between the spots constitute the background. The abundance of a particular transcript thus is not measured directly and is the result of a translation of the image of intensities into a number. Many sources of variation come into play when trying to evaluate the certainty of an expression level.

At the production stage of a DNA microarray, variations in the spots (morphology) cannot be avoided and the exact position of a spot is not known. The Affymetrix GeneChip microarray differs on this point by having spots of highly uniform size and shape. Even when two spots are labeled with the same DNA probe on the same slide there will be variations. At least the same amount of variation will be seen for slides produced in the same batch and more variation comes into play when using slides from different batches. Also the spots on a slide are produced by a set of pins and these may have different characteristics. The first step in the analysis is to find the spots on the slide and for each spot to subtract a background value. This is necessary since a light intensity of zero can never be found in the situation where the true expression level is zero. There are several competing packages for performing this part of the analysis (SPOT, QuantArray, etc.). The background subtraction is a very delicate process and a wrong correction can lead to non-linear effects in the later analysis.

Preparation of a sample for hybridization involves several steps and therefore also introduces a source of variation. It is believed, however, that these effects are mainly proportional and can be corrected by scaling all the measurements by the same number. The essence is that normalization between the Cy3 and Cy5 measurements is needed before relative abundances can be calculated. Based on the experience gathered so far the normalization is performed for each pin separately. Contrary to the belief mentioned previously, some authors find that a scaling is not sufficient for making the two channels comparable and they perform instead a non-linear scaling based on fitting a general curve to a plot of log differences against log averages (Yang *et al.*, 2002). Normalization of the sample under investigation against the reference sample presumes that the two samples behave similarly for a large fraction of the genes being spotted to the array. Thus, if most of the genes spotted to the array are believed to be involved in the process under study, a normalization cannot be based on the behavior of a large fraction of the genes. Instead, the study needs to include control genes that can be assumed to be similarly expressed in the two samples.

Having made a (successful) normalization between the biological sample under investigation and the common reference sample, relative abundance (as a measure of the expression level) can be compared directly between several arrays. For the comercially available microarray, where only one sample is hybridized to the array, it is necessary to make a normalization between the arrays. The principles for doing this are similar to those discussed earlier for the normalizations in the two-sample system.

On top of the experimentally related error terms described earlier, one has the biological variation for a homogeneous population. This biological variation is often large and influences the sample sizes required for a reliable observation of a differential expression between two groups.

16.2.3 Data Analysis Methods

Gene expression data analysis can be divided into two approaches: unsupervised and supervised methods (Golub *et al.*, 1999). Unsupervised analysis aims at identifying previously unknown relationships between samples. For this purpose hierarchical cluster analysis has been used extensively to group tumors according to similarity in expression profiles. This technique can also be applied to group genes with similar expression profiles across the tumor samples analyzed and in this way identify possible co-varying and/or functionally related genes. Hierarchical cluster analysis is a powerful tool for visualization of the expression patterns (Eisen *et al.*, 1998). Other unsupervised methods used are self-organizing maps (SOM), principal component analysis (PCA), and relevance networks. The unsupervised methods are suited for finding novel relationships between samples (or genes) based on similarities in the gene expression patterns. However, the possibility of identifying non-relevant groups exists because of the complexity in large-scale gene expression data.

Supervised analysis techniques are used for identifying differentially expressed genes between groups of samples; for example, where the histopathological stage or clinical outcome (survival, metastasis, recurrence, progression, etc.) is known. Optimal expression signatures for disease classification or outcome prediction are in this way generated from a set of training samples and the significance of the identified expression signature is tested using independent test samples. Standard statistical tests, such as the Student's *t*-test, are used to identify the genes that show the largest differential expression between the groups. Mathematical methods used for classifying samples based on the optimal gene expression signatures include maximum likelihood estimates, *k*-nearest neighbors, support vector machines, and weighted voting schemes, among others (Dudoit *et al.*, 2000).

The use of supervised methods in the analysis of gene expression data, however, involves the risk of oversimplifying the clinical groups under investigation. A single clinical class (e.g. poor outcome) may contain several molecular subclasses of tumors with distinct expression patterns, which may obscure and make the selection of differentially expressed genes between the clinical groups analyzed virtually impossible. When using supervised methods for selection of differentially expressed genes the significance of the selected genes can be accessed by permutation analysis. This usually is done by multiple permutations of the sample labels, followed by generation of statistics on the ability to select good differentially expressed genes in random classes, compared to the real data. Such tests determine the likelihood of obtaining the observed expression patterns by chance, which is a real problem when performing thousands of statistical tests.

16.2.4 Disease Diagnosis and Prognosis by Microarray Expression Profiling Analysis

A huge number of published studies involve microarray analysis for gene expression profiling, and several of these studies have demonstrated a potential clinical use of microarrays for diagnosis and for predicting the clinical outcome of patients with various types of cancers. This section reviews some of the most promising clinical studies published to date.

16.2.4.1 Breast Cancer

Microarray studies of tumors from patients with breast cancer have identified several very interesting disease characteristics as well as intriguing disease outcome predictions. To date, gene expression profiles associated with, for example, estrogen receptor protein expression, histological grade, lymph node status, ERBB2 (HER-2/neu) gene amplification, p53 mutational status, inflammatory breast cancer, and carcinoma-derived stromal signatures have been defined. Identification of the most prominent genes constituting the expression profiles associated with already known subtypes has significantly improved our understanding of the complex biology involved in the development of these different tumor subtypes and facilitated identification of new potential therapeutic targets. Moreover, class comparison studies using gene expression profiling highlight the capacity of this technology to refine clinical diagnostic or prognostic tools.

In one of the first studies by Perou and colleagues (2000), gene expression patterns in 65 breast tumors were studied using microarrays with probes for 8,102 human genes. By the use of unsupervised hierarchical cluster analysis the authors identified molecularly distinct subclasses of breast tumors. The subclasses were characterized by *ERBB2* overexpression and by similarities to basal epithelial cells, luminal epithelial cells, and adipose-enriched/normal breast cells. Additional studies of the identified subclasses of breast cancer revealed a subclass of the luminal epithelial cell-like tumors, which demonstrated a significant difference in disease outcome compared to the other groups.

One of the most interesting and clinically promising microarray expression profiling studies identified a 70-gene expression signature (Amsterdam signature) for predicting if patients diagnosed with breast cancer would develop disease metastases (van't Veer *et al.*, 2002). The expression patterns were found using a cohort of 78 young patients that had not received any treatment, and supervised learning methods were used for identification of the optimal genes for predicting disease outcome. The microarrays used in this study contained oligonucleotides probes for approximately 25,000 human genes. This clinically promising prognosis classifier was later validated on an independent tumor set consisting of 234 breast tumors using the same microarray platform (van de Vijver *et al.*, 2002). The validation confirmed a strong correlation between the gene expression signature for a poor prognosis and disease outcome. Other groups have subsequently reported signatures for disease outcome and tumor progression: 76-gene Rotterdam signature (clinical outcome) (Wang *et al.*, 2005), 21-gene recurrence score (clinical outcome) (Paik *et al.*, 2006), and 512-gene wound-response signature (tumor progression) (Chang *et al.*, 2005). These microarray-based studies of breast cancers hold much promise for tailoring personalized treatment regimens to patients suffering from breast cancer, as only the patients predicted to develop metastatic disease should be treated with adjuvant chemotherapy. Currently, the Dutch group is carrying out a multicenter prospective validation study (MINDACT trial) of the 70-gene Amsterdam signature in Europe that includes several thousand patients. Furthermore, the recurrence score is also being validated prospectively (TAILORX trial).

16.2.4.2 Bladder Cancer

For the study of bladder cancer development and progression, microarray gene expression profiling has also been applied with success. In one of the first studies using clinical material, Affymetrix GeneCchips with probes for approximately 5,000 human genes and ESTs were used to identify gene expression pattern changes between superficial and invasive tumors (Thykjaer *et al.*, 2001). The identified genes encoded oncogenes, growth factors, proteinases, and transcription factors together with proteins involved in cell cycle, cell adhesion, and immunology. This was the first study to identify genes that separated superficial from invasive bladder tumors. A later microarray-based study on bladder tumors showed advances in disease classification and outcome prediction (Dyrskjøt *et al.*, 2003). The authors of this study also used the GeneChips with probes for approximately 5,000-genes and ESTs for identification of a 32-gene expression pattern using 40 tumor samples for classifying tumors according to disease stage. This stage classifier was successfully validated on an independent test set consisting of 68 bladder tumors analyzed on a different array platform. The stage classifier did not only reproduce histopathological staging, but added important information regarding subsequent disease progression.

Prediction of disease progression from non-muscle-invasive to invasive stage would be of great benefit in the clinical management of patients diagnosed with early stage bladder tumors. In one study, a 45-gene molecular classifier was developed by comparing 29 non-muscle-invasive tumors (13 without later progression and 16 with later progression) using custom Affymetrix GeneChip arrays (Dyrskjøt *et al.*, 2005). The 45-gene classifier was tested on a series of 74 independent tumors using a two-color oligonucleotide array platform with only the genes of interest. The classification results showed a positive correlation to disease outcome ($P < 0.03$) with a positive predictive value of 0.3 and a negative predictive value of 0.95. The low positive predictive value may be explained by the fact that patients were continuously treated with transurethral resection and BCG installations. In another study of progression prediction (Wild *et al.*, 2005) the authors used 42 Ta tumors, where eight showed later progression to invasive bladder cancer and eight showed later CIS lesions to delineate a gene set optimal for predicting progression. Using cross-validation test, the predictor correctly classified 33 of the samples, which gives a sensitivity of 86% and a specificity of 71%. No independent test set validation results have been reported for this gene set. The consensus gene set of 11 genes resulting from the most commonly used genes in cross-validation loops show no overlap with the 45-gene set signature from Dyrskjøt and colleagues (Dyrskjøt *et al.*, 2005). The progression signature reported by Dyrskjøt and colleagues was recently validated in a large retrospective study using bladder tumors from a cohort of 404 patients diagnosed with bladder cancer in hospitals in Denmark, Sweden, France, England, and Spain (Dyrskjøt *et al.*, 2007). The molecular progression classifier was highly significantly correlated with progression-free survival ($P < 0.001$) and cancer-specific survival ($P = 0.001$). Furthermore, multivariate Cox's regression analysis showed the progression classifier to be an independent significant variable associated with disease progression after adjustment for known risk factors as age, sex, stage, grade, and treatment (hazard ratio 2.3, $P = 0.007$). Consequently, the retrospective multi-center validation study confirmed the potential clinical utility of the molecular classifier to predict the outcome of patients initially diagnosed with non-muscle invasive bladder cancer.

Gene expression profiles predictive of chemotherapy response have been published in several neoplasms. In a small study of muscle invasive bladder cancer, the response to neoadjuvant (in advance to surgical treatment) chemotherapy was investigated using cDNA microarrays (Takata *et al.*, 2005). Fourteen tumors were used to identify a signature of 14 predictive genes, which was validated on nine additional tumors. RT-PCR results showed good correlation with the microarray, warranting further validation in a larger series. In a recent study, Als and colleagues (2007) identified 55 genes that correlated significantly with survival following chemotherapy. The authors validated two of the protein products (emmprin and survivin) using immunohistochemistry on an independent sample set of 124 tumors. Multivariate analysis identified emmprin expression (hazard ratio, 2.23; $P < 0.0001$) and surviving expression (hazard ratio, 2.46; $P < 0.0001$) as independent prognostic markers for poor outcome, together with the presence of visceral metastases (hazard ratio, 2.62; $P < 0.0001$). In the clinical good prognostic group of patients without visceral metastases, both markers showed significant discriminating power as supplemental risk factors ($P < 0.0001$). Within this group of patients, the subgroups of patients with no positive, one positive, or two positive immunohistochemistry scores (emmprin and survivin) had estimated 5-year survival rates of 44.0%, 21.1%, and 0%, respectively. Response to chemotherapy could also be predicted with an odds ratio of 4.41 (95% confidence interval, 1.91–10.1) and 2.48 (95% confidence interval, 1.1–5.5) for emmprin and survivin, respectively. Consequently, emmprin and survivin proteins were identified as strong independent prognostic factors for response and survival after chemotherapy in patients with advanced bladder cancer.

16.2.4.3 Lung Cancer

Improvements in classification and prognosis have also been reported in studies of lung cancer by microarray expression profiling of 12,600 transcripts in 186 samples (Bhattacharjee *et al.*, 2001). Unsupervised hierarchical

cluster analysis grouped the samples according to established histological classes. In addition, the adenocarcinomas were found to contain four subclasses, and one of these had a less favorable outcome compared to the other subclasses of the adenocarcinomas.

Another interesting microarray-based finding was reported by Beer and colleagues (2002), who studied 86 primary early stage adenocarcinomas from patients with lung cancer. A 50-gene risk index was determined for identifying patients with a poor prognosis. The 50-gene risk index was validated using the data reported previously by Bhattacharjee and colleagues (2001). The authors classified the adenocarcinomas as low and high risk and found a close correlation to disease outcome, and concluded that high-risk patients may benefit from adjuvant treatment. In a more recent study, Lu *et al.* identified 64 genes capable of separating high-risk and low-risk groups with regard to overall survival. The authors performed a meta-analysis based on seven datasets and five different Affymetrix arrays. The authors suggest their signatures as a criteria for selecting patients to Cisplatin-based adjuvant chemotherapy in the high-risk group, and for stratifying patients according to risk in trials of adjuvant treatment (Lu *et al.*, 2006).

16.2.5 Diagnosis and Prognosis in Cancer Based on miRNA Array Profiling

miRNAs are a class of small non-coding RNA molecules of approximately 22 nucleotides that are involved in post-transcriptional regulation. miRNAs are transcribed as primary miRNAs (pri-miRNAs) and processed into shorter precursor miRNAs (pre-miRNAs) of 70–100 nt and eventually processed into mature single-stranded active miRNAs. Hundreds of miRNAs have been cloned and sequenced and current estimates suggest that up to 33% of all gene transcripts are negatively regulated by miRNAs. miRNAs bind to target sequences in the 3′-untranslated region (3′UTR) of mRNAs, which lead to translational repression or degradation of the target mRNA dependent upon the degree of complementarity. Accordingly, miRNAs have been found to influence gene regulatory processes such as development, differentiation, and disease such as in Tourette's syndrome (Abelson *et al.*, 2005), fragile X-syndrome (reviewed in Garber *et al.*, 2006), cardiac hypertrophy (Catalucci *et al.*, 2008), and particularly in cancer (Calin and Croce, 2006; Esquela-Kerscher and Slack, 2006). To date, oligonucleotide miRNA microarray analysis is the most common high-throughput technique for assessment of cancer-specific expression of hundreds of miRNAs and so far every tumor type analyzed shows a significantly different miRNA profile than corresponding normal tissue. Microarray analysis is usually accompanied by validation of the array results by quantitative real-time PCR or, more rarely, Northern blotting. Several studies have thus examined the miRNA expression in cancers such as colon, lung, breast, prostate, glioblastoma, leukemia, and hepatocellular carcinomas. The recent advances in microarray technology have contributed to the discovery of miRNAs as promising markers for diagnosis and prognosis of different cancers (Table 16.1) (reviewed in Calin and Croce, 2006; Davison *et al.*, 2006; Krichevsky, 2007).

One of the largest and earliest miRNA microarray profiling studies was conducted by Volinia *et al.* (2006), where 540 samples of six different solid tumors were examined by microarray analysis. Five μg of total RNA was hybridized to chips containing 40-mer oligonucleotide probes and the biotin-labeled transcripts were detected by streptavidin-alexa fluor conjugate and the slides were subsequently scanned. By performing a significance analysis of microarrays (SAM) with a false detection rate (FDR) $\leq 1\%$, 49 miRNAs were found differentially expressed among the solid tumors as compared to normal tissue. A six-miRNA tissue-specific cancer signature was also identified. This study detected approximately equal numbers of up- as well as down-regulated miRNAs, in contrast to another early large miRNA profiling work by Lu *et al.* (2005) who found miRNAs predominantly down-regulated in cancer. This study was, however, conducted by bead-based flow cytometric analysis, and the difference may therefore be ascribed to differences in technical platforms and analysis methods.

Based on the distinct miRNA profiles found in different types of cancer, a recent work by Rosetta Genomics Ltd coworkers suggested that miRNAs can accurately identify the tissue origin of cancer and thus be exploited in the classification of tissue origin of metastatic cancer of unknown primary origin. This cancer type accounts for 3–5% of all newly diagnosed cancers and represents a major challenge for diagnosis and selection of correct treatment. Thus a "decision-tree" classifier with 24 nodes was designed and, based on the miRNA expression of >600 tested human miRNAs, a series of binary decisions starting at node no. 1 and onwards led to high confidence accuracy of the tissue of origin (Rosenfeld *et al.*, 2008).

In a study on miRNA alterations in breast cancer, a miRNA signature was found within distinct subpopulations of luminal and myoepithelial cells that discriminates normal from breast tumor samples. This study did not find significant changes in miRNA expression correlating to ER/PR/HER2 status in breast cancer, although a trend of high miR-205 expression associated with favorable clinical outcome of ER-PR-HER2 was found. However, an earlier study did find an association between specific pathological features such as ER and PR expression, tumor stage, vascular invasion, and proliferation index (Iorio *et al.*, 2005). Microarray analysis of breast cancer cell lines have also identified metastasis-associated miRNAs that were subsequently manipulated in murine models of breast cancer,

TABLE 16.1 miRNA microarray profiling of cancers with diagnostic or prognostic value.

Cancer type	miRNA profiling	Importance	Reference
Lung cancer	6 miRNAs discriminates subtypes of cancer (adenocarcinoma/squamous cell carcinoma)	diagnostic/prognostic	Yanaihara *et al*. (2005)
	3-miRNA signature predicts outcome of patient with stage I adenocarcinoma		
Breast carcinoma	miRNA expression correlates with different pathological features (ER/PR status)	diagnostic/prognostic	Iorio *et al*. (2005)
	A trend of high miR-205 correlation with positive clinical outcome of ER-PR-HER2- patients		
	Metastasis-associated miRNAs identified by array analysis directly affects metastatic potential in a mouse model of cancer		Tavazoie *et al*. (2008)
Acute, chronic lymphotytic leukemia	A unique 13-miRNA expression signature predicts CLL prognosis related to ZAP70 and IgH mutation status	diagnostic	Calin *et al*. (2004, 2005)
Endocrine pancreatic tumors	3 miRNAs (miR-103, -107, -155) distinguish cancer from normal pancreas, ten miRNAs discriminate endocrine from acinar tumors, miR-21 correlates with liver metastasis	diagnostic/prognostic	Roldo *et al*. (2006)
Colon cancer	Differential expression between cancer/normal, high miR-21 associated with increased hazard ratio of 2.4 and poor outcome	diagnostic/prognostic	Schetter *et al*. (2008)
Glioblastoma	13 differentially expressed miRNAs between cancer/normal, among them miR-221 targets the cell cycle regulator p27Kip1	diagnostic	Ciafrè *et al*. (2005)
			Gillies and Lorimer (2007)
Esophageal squamous cell carcinoma	A 7-miRNA classifier discriminates cancerous tissue from adjacent normal, low miR-103/107 correlates with high survival, miRNAs discriminate gross pathological and differentiation classification	diagnostic/prognostic	Guo *et al*. (2008)
Papillary thyroid carcinomas	3 miRNAs (miR-221, -222, -146) can discriminate cancer tissue from normal thyroid	diagnostic	He *et al*. (2005)
Prostate cancer	51 miRNAs are differentially expressed in cancer out of 319 tested	diagnostic	Porkka *et al*. (2007)
Ovarian cancer	miRNAs differentially expressed in cancer vs. normal, specific biopathologic features (e.g. histotype and invasion) correlates with miRNA expression	diagnostic	Iorio *et al*. (2007)
Hepatocellular carcinoma	miRNA expression correlates with differentiation status	prognostic	Murakami *et al*. (2006)
Solid cancers	540 human samples of solid tumors and normal tissue examined, 49 miRNAs differentially expressed in solid tumors (equally up- and down-regulation), a six tissue-specific signature also generated	diagnostic	Volinia *et al*. (2006)
22 different tumor tissues/metastases	Molecular classification study for tumor origin based on miRNA expression (for classification of metastatic cancer of unknown origin)	diagnostic	Rosenfeld *et al*. (2008)

resulting in changed predisposition to metastasis (Tavazoie *et al.*, 2008).

In the analysis of miRNA alterations associated with human B cell chronic lymphocytic leukemia (CLL), unsupervised hierarchical clustering generated two clearly distinguishable miRNA signatures that were linked to a difference in the expression of ZAP-70, a tyrosine kinase that is a strong predictor of early disease progression (Calin *et al.*, 2004). A unique 13-miRNA expression signature was subsequently identified as a powerful indicator of CLL (Calin *et al.*, 2005). Different microarray analyses have demonstrated that miRNAs are involved in establishment, maintenance, and function of hematopoietic lineages and show distinct alterations in hematopoietic cancers (reviewed in Tili *et al.*, 2008).

Extensive collections of archival tumor biopsies and paired normal samples exist in hospital laboratories and biobanks often combined with clinicopathological information and patient outcome. However, long mRNA molecules are poorly preserved during cryostorage which impedes the usage of these samples in traditional expression microarray studies. A recent study on esophageal squamous cell carcinoma demonstrated how miRNAs are well preserved in >5 years cryoarchived tissues whereas the total RNA was extensively degraded. Accordingly, a 7-miRNA classifier could distinguish cancerous tissue from adjacent normal tissue, low miR103/107 was found correlated with high survival period, and miRNAs also discriminated between gross pathological and differentiation classification (Guo *et al.*, 2008). Thus, since accurate and informative analysis long after patient surgery is necessary to allow prognostic analysis, miRNA microarray analysis may prove more robust as a future microarray-based cancer prognostic tool due to the higher preservation of miRNAs as compared to mRNA transcripts.

16.2.6 Detection of Prevailing Molecular Pathways in Tumors Using Expression Microarrays

Analysis of differential expression in tumor tissues may provide new information about specific biological pathways involved in carcinogenic processes. The establishment of high-throughput gene expression platforms has moved the bottleneck from data production to data analysis. Gene expression profiling studies often contain hundreds of samples resulting in several million data points that must be coped with. Such data are increasingly becoming available for downloading from the internet, leading to even larger masses of data. Many bio-informatic companies and scientific research groups are continuously working on improving software for a wide variety of data analyses. Often, these programs are freely available on the internet. Several cancer studies have compared tumors with normal tissue or tumors at different stages in order to identify new tumor- or stage-specific markers. A thousand genes or more, for which the functions are unknown, may be reported as potential markers. The differentially expressed genes are compared to data from the literature for that particular disease or other similar diseases as an initial confirmation of the validity of data. Because of the molecular heterogeneity of most cancers, dimension-reduction methods like single value decomposition or various clustering methods are often applied as the next analysis step. Such analysis helps visualize patterns in the data that may help define new molecular subclasses in a disease.

In order to generate hypotheses concerning pathways in oncogenesis and disease progression, differentially expressed genes must also be put into a biological context. This may be done by extracting specific information related to gene interplay in signaling networks, called pathways. Once a hypothesis has been generated, specific genes may be chosen for further manipulations in a suitable model system in order to test the hypothesis. A common choice for this are cell lines, where gene expression can be manipulated by various methods, thus mimicking the conditions in tumor cells to a limited but defined extent.

There are several helpful tools available for pathway analysis either as free downloads, e.g. GenMAPP (www.genmapp.org), or as licensed software, e.g. MetaCore (www.genego.com), Pathway Assist (www.stratagene.com) or Ingenuity Pathway Analysis (IPA) (www.ingenuity.com). IPA is an all-in-one software application that enables researchers to model, analyze, and understand the complex biological and chemical systems at the core of life science research. IPA allows the direct coupling of protein or transcript expression data to a database providing information extracted by experts from the full text of the scientific literature. This includes information about genes, drugs, chemicals, cellular and disease processes, signaling and metabolic pathways as well as functional annotation of genes. Molecules differentially expressed can be grouped according to their cellular function or their involvement in canonical pathways. Moreover, expression of molecules can be related to different organisms and organs, as well as to the gene expression calls for the NCI-60 panel of cell lines. IPA comprises most of the pathways known so far. Adding new information will extend the pathways and networks which later on can be overlayed with the user's own expression data, derived from expression profiling data from tissues or cell line experiments. A color code can be defined to color all genes up-regulated in red or yellow, and all genes down-regulated in an experiment green or blue. The analysis software provides the possibility of changing exclusion limits as p-values, fold changes or $\log_2$ ratios. At last, the "IPA Path Designer" introduced in IPA6.0 provides the possibility of transforming networks and pathways into publication-quality pathway graphics. This kind of visualization of protein or transcript expression data representing

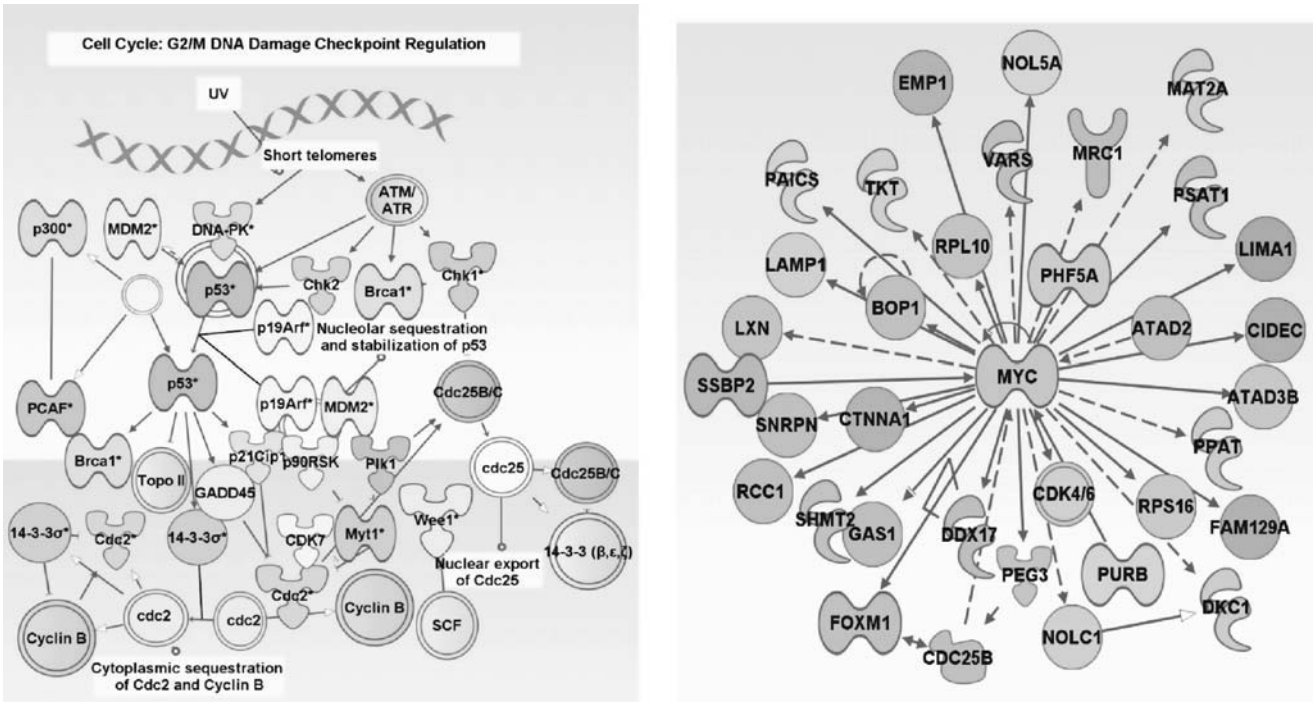


FIGURE 16.2 Ingenuity pathway analysis IPA 7.1 on microarray transcript profiling data comparing normal colon mucosa to colon adenocarcinomas. **A.** Canonical Pathway Cell Cycle: G1/S Check Point Regulation. **B.** A network centered around MYC.

biological pathways and groups of genes is extremely helpful to get an overview of the affected pathways in an intuitive way. Figure 16.2 shows an example of a canonical pathway derived from the IPA software.

16.3 APPLICATION OF CGH ARRAYS FOR STUDIES OF THE MALIGNANT CELL GENOME

Comparative genomic hybridization (CGH, see also Chapters 10 and 12) is a technique that detects and maps changes in the copy number of DNA sequences (Kallioniemi *et al.*, 1992). In CGH, DNA from a test (e.g. from a tumor) and reference genome (genomic DNA from a healthy/normal individual) are differentially labeled and hybridized to a representation of the genome. Originally, the representation was a metaphase chromosome spread, but over the past several years, microarray-based representations increasingly have been used. The fluorescence ratios of the test and reference hybridization signals are determined at different positions (the array elements) along the genome and provide information on the relative copy number of sequences in the test genome compared to the reference (a normal diploid) genome. The array representations can provide a number of advantages over the use of chromosomes, including higher resolution and dynamic range, direct mapping of the copy-number changes to the genome sequence, and high throughput. A variety of chromosomal aberrations underlie inherited diseases and cancer. Aberrations leading to changes in DNA copy number can be detected by CGH and include interstitial deletions and duplications, non-reciprocal translocations, and gene amplifications. It is important to note that CGH does not provide information on ploidy or location of the rearranged sequences responsible for the copy-number change. Furthermore, the capability of a CGH array to detect aberrations spanning small genomic regions depends on both the size and genomic spacing of the elements (clones) on the array.

Different applications of CGH arrays impose different performance requirements, so that certain approaches may be suitable for a particular application, whereas others are not. Accordingly, unique CGH arrays have been designed for various applications; for example, for screening of telomeric regions, screening of selected genomic regions known to be frequently involved in cancer, and genomewide screening. Three different types of array elements have been employed in the various approaches: bacterial artificial chromosomes (BACs), cDNA clones, and oligonucleotides. An analysis typically requires several hundred nanograms of genomic DNA from the specimen when using some BAC arrays (Snijders *et al.*, 2001), or one or more micrograms for cDNA arrays (Pollack *et al.*, 1999; Monni *et al.*, 2001). Whole genome amplification procedures are likely to reduce substantially the amount of required specimen.

The use of large insert clones like BACs for CGH arrays provides sufficiently intense signals such that single copy changes affecting individual clones on the array can be detected and aberration boundaries can be located within a fraction of a BAC length. However, propagating and printing BACs can be problematic. BACs are single copy vectors, hence the yield of BAC DNA is low and solutions of the high molecular weight DNA can be viscous, making it difficult to print. Since growing and processing large bacterial cultures is not practical for arrays of thousands of elements, a number of methods have been devised to generate representations of BACs from minute cultures, which can be set up in an industrial fashion.

DNA copy-number measurements can also be made using arrays containing elements made from cDNA or oligonucleotides. Typically, these arrays initially were produced to measure gene expression. The advantages of these arrays are that they are often readily available and contain large numbers of elements because they were produced to comprehensively assess the transcriptome. The disadvantages are that only large copy-number gains provide sufficient signal to determine the boundaries of the amplicons with high resolution.

The detection of low level copy-number changes (e.g. losses) requires calculating the running average of multiple clones (5–10 clones) along the genome, and frequently entails discarding measurements on a substantial fraction of elements because they do not provide adequate signals. Thus, the actual genomic resolution of the boundaries of single copy changes and the ability to detect focal single copy changes is considerably less than implied by the average genomic spacing between the clones on the array.

A number of studies have taken advantage of the higher resolution afforded by CGH arrays to more precisely map the boundaries of genomic copy-number alterations. Once these are defined, candidate tumor suppressors and oncogenes mapping within the region can readily be identified from the genome sequence databases. Investigation of expression levels of these candidate genes in tumors and model systems (cell lines, etc.) can then be used to determine which of the candidates are most likely to contribute to the disease phenotype and to be the "driver gene(s)" for the copy-number alteration. This approach to candidate gene identification is likely to be fruitful, since a correspondence between copy-number alterations and changes in gene expression was demonstrated in breast cancer as measured by genome-wide CGH arrays and global expression profiling on the same cDNA array (Hyman *et al.*, 2002; Pollack *et al.*, 2002).

Tumor classification based on copy-number profiles obtained using CGH arrays has been reported for several cancer types, including liposarcomas (Fritz *et al.*, 2002), renal cell carcinomas (Wilhelm *et al.*, 2002), and gastric adenocarcinomas (Weiss *et al.*, 2003). However, in a recent study of bladder tumors no significant relationship between copy-number alterations and tumor stage and grade was found (Veltman *et al.*, 2003). Instead, an intriguing link among particular genomic loci was revealed; that is, copy-number alteration at particular loci occurred together (e.g. gain of CCND1 and deletion of TP53), whereas alterations in the copy number of loci, harboring genes that function in the same pathway such as gains of CCND1 and E2F3, were found to be "complementary" as they did not occur in the same tumors. Together, the studies cited indicate that CGH arrays will be increasingly important in understanding tumor biology over the coming years.

16.4 WHOLE-GENOME SNP ARRAY ANALYSIS OF CANCER

Microsatellite markers have been used successfully for loss-of-heterozygosity (LOH) analysis and association studies. The instability of these markers in tumor tissue as well as the difficulty of upscaling and automating PCR-based microsatellite assays, however, make their usage for genome-wide studies impractical (Buschiazzo and Gemmell, 2006). Instead, focus has turned to SNP (single nucleotide polymorphism) markers and high-throughput genotyping of these using SNP microarrays. A SNP is defined as a DNA sequence variation at one specific position in the genome that occurs in at least 1% of the human population. Impressive advances with respect to identification and genomic mapping of millions of single nucleotide polymorphisms (SNPs) have been achieved within the last 5–10 years by the International SNP Consortium (Holden, 2002) and the International HapMap projects (2005 and 2007). As a result of these and other efforts, the most recent version of the human SNP database of the National Center for Biotechnology Information (NCBI) (dbSNP build 128, October 2007) contains approximately 9.3 million true SNPs, i.e. single nucleotide polymorphisms that map exactly once to a well-defined location in the genome (http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp).

Although SNPs are very abundant, they generally have a low average heterozygosity rate compared to microsatellites, meaning that on average three times as many SNPs are required to obtain a resolution equivalent to a microsatellite. This apparent drawback of SNP markers, however, has been more than sufficiently compensated for by rapid improvements in DNA oligonucleotide microarray manufacturing technologies making it possible now to genotype nearly 1 million SNPs on a single array (see below). This number is likely to increase even further as technologies continue to improve. Therefore, SNPs have currently become the markers of choice for whole-genome LOH and association studies (Beaudet and Belmont, 2008). Moreover, the ability to determine DNA copy-number variation based on relative hybridization signal intensities makes SNP array platforms highly attractive alternatives

to comparative genomic hybridization for determination of copy-number aberrations.

Genome-wide LOH analysis is typically performed by comparison of tumor DNA and normal genomic DNA (often from leukocytes) from the same individual by parallel SNP array analyses. If a matched normal control sample is unavailable, it is possible instead to use algorithms (Buschiazzo and Gemmell, 2006). The latter approach, however, is suboptimal and makes data interpretation more difficult due to extensive copy-number variation in the normal human genome (Redon *et al.*, 2006). SNP array genotyping determines if both alleles of a heterozygous SNP are present in the tumor. Thus, LOH regions are identified from stretches of adjacent heterozygous SNPs, while homozygous SNPs are non-informative (Heinrichs and Look, 2007). In contrast, for copy-number determination, which is inferred from the hybridization signal of each probe, all SNPs on the array are informative. Because genotyping and copy-number determination are integrated in the same analysis, a major advantage of SNP arrays over other high-throughput methods for genomic analysis, like CGH arrays, is their unique ability to identify uniparental disomy in tumor samples. Uniparental disomy is a common cancer-associated chromosomal aberration that occurs when one allele is lost and the other allele is duplicated, thereby causing LOH without copy-number change.

SNP array genotyping is also used to identify an association between specific polymorphic loci in the genome and the risk of disease development. Genome-wide association studies are performed by analysis of genomic DNA from a group of patients with disease (cases) compared to a control group of unaffected individuals, or as linkage studies in families with hereditary cancer. Statistically significant differences in allele frequencies of polymorphic markers (SNPs) between cases and controls indicate that the corresponding genomic region causes increased disease susceptibility. Association studies use either a direct approach, which aims to identify variants with a functional role, or an indirect method, which allows identification of polymorphic variants associated with a disease condition, but not necessarily functional themselves. A major advantage of the second approach, which is typically based on linkage disequilibrium, is that it can be performed without prior knowledge of candidate genes or regions (Kruglyak, 2008). Linkage disequilibrium (LD) is defined as the non-random association between two or more alleles so that some combinations, due to common descent, are more likely to occur together than others. LD in the human genome generally occurs in block-like structures, or haplotype blocks (The International HapMap Consortium, 2005 and 2007). In any given population, a typical SNP is therefore closely correlated with many of its neighboring SNPs, and it is possible to select a set of "tag SNPs" that can be used as proxies for a much larger set of genetically redundant SNPs. Analysis of tag SNPs can thereby reduce the total number of SNPs needed for full-genome coverage compared to randomly chosen SNPs (Barrett and Cardon, 2006). It has been shown that approximately 500,000 tag SNPs can provide good genomic coverage for LD-based association studies in most human populations (Kruglyak, 2008). For determination of copy-number variation or LOH analysis, however, a more equally dispersed set of SNPs may be preferred.

The following sections outline in more detail the technologies of modern commercial SNP array platforms from two major players on the market, Affymetrix and Illumina, and provide a few examples of significant discoveries in cancer research based on the application of these or similar technologies.

16.4.1 Affymetrix SNP Arrays

Since the first SNP analysis on high-density oligonucleotide arrays was developed by Affymetrix, the number of interrogated SNPs has increased dramatically from around 600 SNPs on each array (Wang *et al.*, 1998; Mei, *et al.*, 2000) to analysis of approximately 500,000 on two arrays (GeneChip Human Mapping 500K Array Set) and, most recently, over 900,000 SNPs on a single array (Affymetrix Genome-Wide Human SNP Array 6.0). The modern Affymetrix array types contain over 6 million $5 \times 5\mu m$ features on each array, with each feature consisting of more than 1 million copies of a specific 25-mer oligonucleotide probe. All oligonucleotides are synthesized directly and in parallel on the solid surface by photolithographic techniques.

Each array in the Mapping 500K set genotypes approximately 250,000 SNPs and together the two arrays in the set give a median physical distance of 2.5 kb (mean 5.8 kb) between SNPs with an average heterozygosity of 30%. On the Mapping 500K arrays every SNP is interrogated by 24 or 40 different oligonucleotides tiled around the SNP position. The oligonucleotide sequences differ according to perfect match and mismatch at the SNP position (all four nucleotides are represented at this position). There are also both mismatch and perfect match probes represented in the flanking sequences around the SNP. This probe redundancy improves the confidence of genotype calls. The most recent Affymetrix SNP Array (SNP Array 6.0) provides the highest physical coverage of the genome seen on the market so far with interrogation of 1.85 million markers on a single array. The SNP Array 6.0 contains probes for more than 906,600 SNPs along with 946,000 non-SNP markers for detection of copy-number variation, which give a median inter-marker distance of less than 700 bp. The average heterozygosity of the represented SNPs ranges from 25% to 29% dependent on the population (ethnic origin). Each SNP is interrogated by six to eight 25-mer oligonucleotide probes, and the concomitant development of a new genotyping algorithm has ensured high call rates even with fewer probes for each SNP compared to the 500K Array platform.

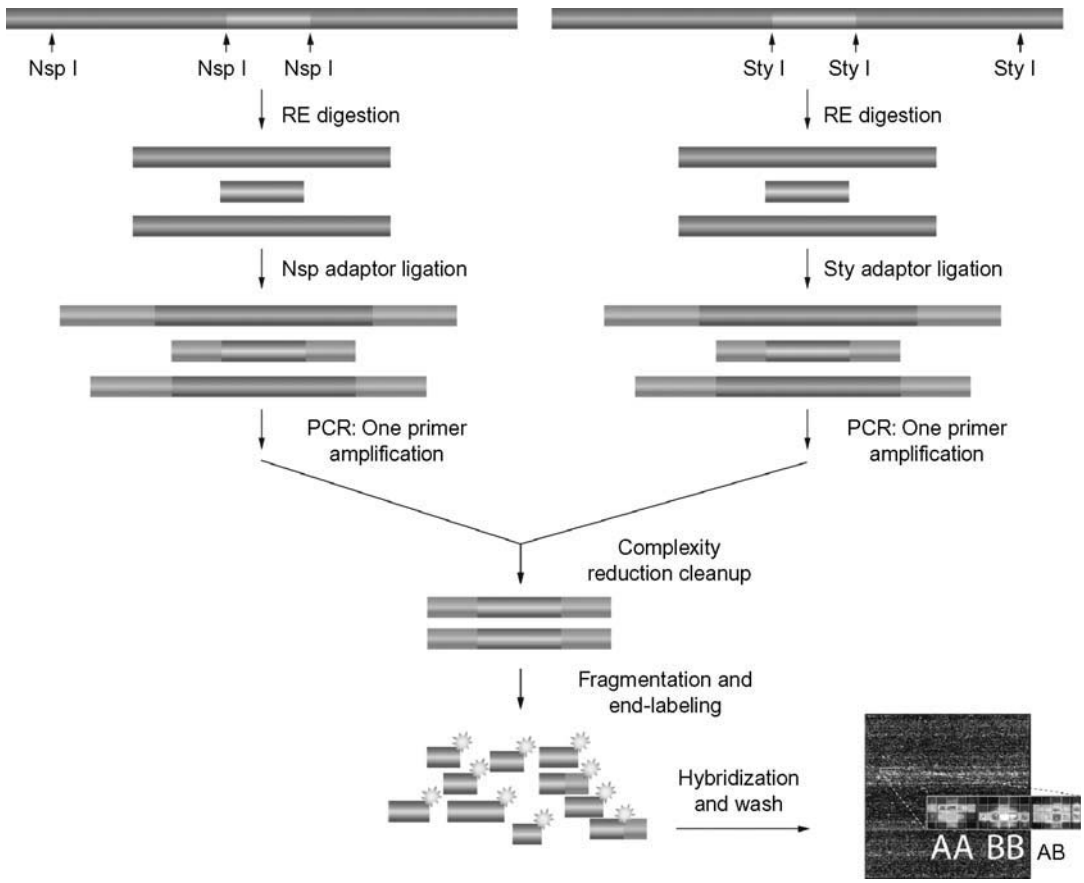


FIGURE 16.3 Overview of Affymetrix Genome-Wide Human SNP Array 6.0 Assay. (Image from Affymetrix datasheet for 6.0 SNP array.)

Whereas the earliest versions of the Affymetrix SNP arrays used multiplex PCR with specific primers for each of the PCR products to amplify genomic DNA, the complexity of genomic DNA targets is reduced for the modern array types by restriction enzyme digestion followed by ligation of a specific adapter (Fig. 16.3 – 500 K and 6.0). This adaptor allows the use of a single generic primer for amplification of all targets prior to hybridization. The PCR amplification step has been optimized to amplify sequences between 200 and 1,100 bp, so SNPs situated in restriction fragments below or above this size are not well represented on the arrays. For the 500 K Array Set, one array is used for Nsp I digested DNA and the other for Sty I digested DNA, whereas for the SNP 6.0 array, PCR products from the Nsp I and Sty I restriction enzyme digests are combined prior to fragmentation, end-labeling and hybridization.

The large number of SNPs represented on the modern array types make them suitable for use in both LOH, copy-number, and association studies. Former generations of Affymetrix SNP arrays, like 10 K and 100 K arrays (interrogating 10,000 and 100,000 SNPs, respectively), however, have been used predominantly for LOH analysis of tumors. Correlations between certain patterns of allelic imbalance and diverse clinical parameters (e.g. tumor stage, grade, risk of progression, or metastasis) have been reported for several malignancies, such as bladder cancer (Koed *et al.*, 2005), colon cancer (Andersen *et al.*, 2007), and prostate cancer (Tørring *et al.*, 2007). Moreover, higher resolution LOH and copy-number analysis of prostate cancer using the 500 K SNP array has mapped a variety of small deletions between the genes *ERG* and *TMPRSS2* on chromosome 21, which are responsible for oncogenic TMPRSS2-ERG fusion transcripts present in many prostate cancers (Liu *et al.*, 2007).

16.4.2 Illumina SNP Arrays

The high-density SNP genotyping systems developed by Illumina use self-assembling microarrays based on beads with coupled oligonucleotide probes, and take advantage of the recent discovery of over 2 million common tag SNPs (minor allele frequency of at least 0.05). Illumina BeadChips contain 12 sections (or stripes) of evenly spaced wells (5 μm spacing) that are created by a combination of photolithography and plasma etching on silicon wafers (Gunderson *et al.*, 2006). For each SNP to be analyzed, specific 50-mer oligonucleotide probes are chemically attached to 3 μm beads. The different bead types are then pooled and randomly dispersed into the wells on each

slide. A decoding process maps the location and identity of each bead on the array. Within each of the 12 sections on the slide it is possible to load 1.1 million beads and to decode 60,000 different bead types, thus each bead type is on average represented 20 times to ensure confident genotyping. The use of beads as microarray features ensures uniform spot size and provides a versatile platform that can be formatted into a variety of custom-designed layouts. Depending on the bead sets used it is possible on one array to genotype up to 720,000 SNPs in one sample (distinct bead sets added to the 12 sections) or, e.g., 60,000 SNPs in 12 different samples (same bead set added to all 12 sections).

Before hybridization to the BeadArray, genomic target DNA undergoes a whole-genome amplification step, followed by enzymatic fragmentation producing an average target DNA fragment size of 300 bp. Each SNP is then scored by an enzymatic extension assay using labeled nucleotides. Two different systems have been developed, named Infinium I and II, which rely on allele-specific primer extension (ASPE) and single base extension (SBE), respectively. In both cases, a sandwich-based immunohistochemical assay is subsequently used for visualization and to increase assay sensitivity. The ASPE assay uses two allele-specific probes differing only at the terminal 3′ base position (=interrogated SNP position) and provides a simple one-color readout. The SBE assay uses only one oligonucleotide probe for each SNP and gives a two-color readout, since A and T nucleotides are labeled in one color, while C and G nucleotides are labeled in another. Although the SBE system requires only half the number of probes compared to ASPE, it has the limitation that it only allows genotyping of 83% of all common bi-allelic SNPs and not the remaining 17%, which are AT or CG polymorphisms. These have to be scored either by ASPE or by SBE with swapped nucleotide labels (Steemers and Gunderson, 2007).

Illumina BeadArrays have been used successfully for genome-wide association studies, but can also be employed for both copy-number determination and LOH analysis. One of their latest products, the HumanHap550 BeadChip, contains probes for over 550,000 tag SNPs (ideal for genome-wide association studies) plus probes for numerous non-tag SNPs in order to achieve even spacing across the human genome, which allow high-resolution copy-number determination and LOH analysis (Steemers and Gunderson, 2007). Although the total number of SNPs that can be interrogated by current Illumina BeadChip technology (~720,000 SNPs per chip) is lower than that of the Affymetrix SNP Array 6.0 (~900,000 SNPs), the predominance of tag SNPs on the former may provide an advantage for genome-wide association studies based on LD, whereas the greater genomic coverage of the Affymetrix 6.0 array (1.8 million markers in total) could make it the optimal choice for high-resolution LOH and copy-number analyses. In a comparative study, it has been shown that 317,000 tag SNPs represented on the Illumina HumanHap300 BeadArray have comparable power to the Affymetrix 500 K Array Set in LD-based association studies (Barrett and Cardon, 2006).

The use of high-density tag SNP arrays (mainly Illumina Hap300 and Hap550) for genome-wide association studies has virtually revolutionized the field and led to the identification of strong susceptibility loci for several types of malignancies, including breast cancer (Hunter *et al.*, 2007; Stacey *et al.*, 2007), and colorectal cancer (Haiman *et al.*, 2007; Broderick *et al.*, 2007). Likewise, major breakthroughts in the understanding of prostate cancer genetics have been accomplished by the combined efforts of several studies (Witte, 2007), which have established a robust association between prostate cancer risk and five specific SNPs (three at 8q24, one at 17q12, and one at 17q24.3) and identified about ten additional susceptibility loci. Presence of at least four of the five high-risk SNPs at chromosome 8 and 17 is associated with a remarkable 4.47-fold increased risk of prostate cancer development (Zheng *et al.*, 2008).

16.4.3 Custom-Made SNP Arrays

Whereas the Affymetrix SNP array system is designed to cover the whole genome, the custom-made SNP arrays are designed to determine the frequency of specific SNPs, or sets of SNPs, of interest in specific genes. One of the most used examples of SNP determination on custom array is the method described by Fan and colleagues (2000) and Lindroos and colleagues (2002). There are slight differences in the exact method, but the method described here is an adaptation used in the authors' laboratory (Christensen and Koed, unpublished results).

Genomic DNA covering the SNPs of interest is amplified in a multiplex PCR amplification (see Fig. 16.4a). Surplus primers and nucleotides from the PCR amplification are degraded in order to avoid interference in the following single base-pair extension (SBE) reaction. A specific SBE primer that terminates one base upstream to the SNP is hybridized to the single-stranded PCR product and extended one base in the SBE reaction using differentially fluorescence labeled dideoxy nucleoside triphosphates. The SBE primer has a Tag sequence in the 5′A end. Complementary Tag (c-Tag) sequences are spotted on a glass slide with the help of a 3′A amino group. Subsequently, in a silicone mold, the ddNTP-labeled SBE-Tag primers are hybridized to the c-Tag spotted on the slide (see example in Fig. 16.4b). In each well the material from individual patients can be applied and each individual's composition of SNPs can be determined. Alternatively, pooled DNA from many individuals can be used in the PCR amplification and thereby the frequency of different SNPs can be determined.

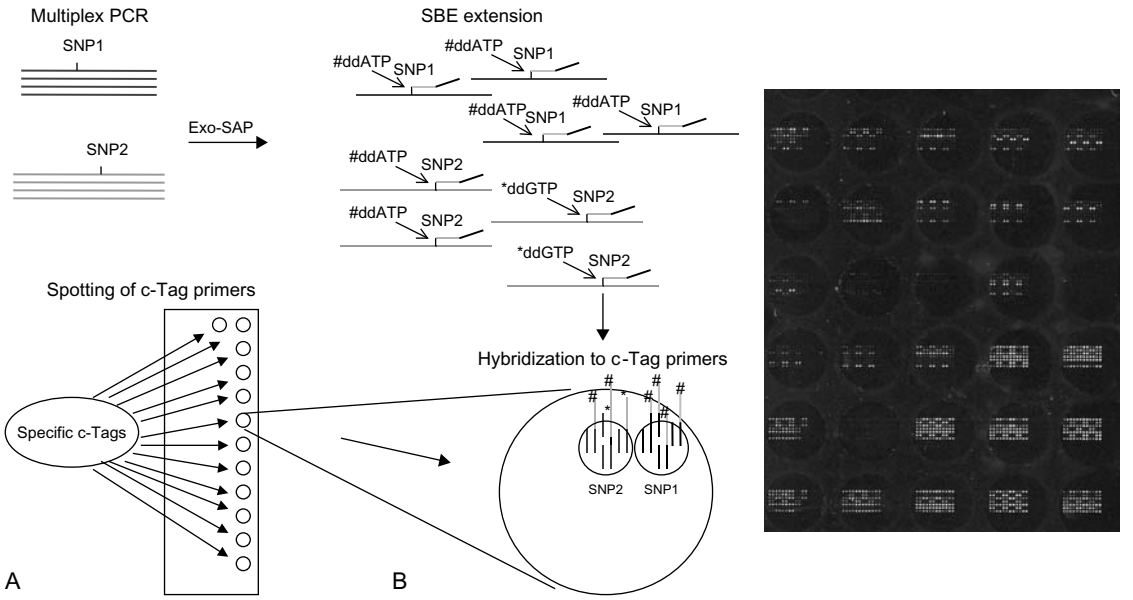


FIGURE 16.4 Custom-made SNP array. **A.** Overview of the SNP array. First, genomic DNA is copied by multiplex PCR (the authors' laboratory routinely runs a multiplex reaction with as many as ten different primer pairs; Christensen and Koed, unpublished results), primers and nucleotides are degraded by exonuclease 1 and shrimp alkaline phosphatase (Exo-SAP), single-base extension (SBE) is carried out in the presence of differentially fluorescence-labeled ddNTPs and specific single-base extension primers with specific Tags in the 5′A end. The Tags hybridize to complementary Tags (c-Tags) bound to glass slides in a definite pattern. **B.** Example of a spotted slide hybridized to labeled SBE primers (Christensen and Koed, unpublished results). The same c-Tags were spotted in each well, but each well represents the results from different multiplex PCR reactions.

Each glass slide contains 80 wells and in each well 100–200 c-Tags can be spotted. If a single bp extension reaction is performed in both the sense and the antisense direction, approximately 50–100 SNPs can be analyzed in each well.

16.5 DETECTION OF ALTERNATIVE SPLICING BY MICROARRAY ANALYSIS

Alternative splicing is the process in which the exons of a common pre-mRNA precursor are assembled to generate multiple different mature mRNA isoforms. It is estimated that about three-quarters of all human genes undergo alternative splicing (Clark *et al.*, 2007) making it a key mechanism in expanding a relatively limited number of genes into very complex transcriptomes and proteomes.

Several different microarray platforms enabling high-throughput analysis of alternative splicing currently exist. These platforms use two basic oligo-probe types to measure alternative splicing; exon body probes targeting the exon sequence and exon junction probes targeting the junction of two exons. Exon junction probes have the ability to detect the presence of specific exon/exon junctions but are restricted to known splice forms, whereas arrays containing exon body probes have the advantage of not being biased towards already known splice variants and therefore have more potential for detection of novel splice variants. Furthermore, the exon junction probes are position constrained making it harder to design ideal probes. Exon body probes can be placed more freely within the exon and are therefore easier to optimize.

Johnson and coworkers (2003) used a set of five arrays containing in total 125,000 36 nucleotide junction probes measuring exon/exon borders of 10,000 multi-exon genes in 52 tissues and cell lines. Thousands of potential tissue-specific alternative splicing events were detected, and 150 events were selected for experimental validation with a success rate of 49% (73 of 150). Pan and colleagues used a microarray platform utilizing both exon body and exon junction probes to monitor 3,126 preselected alternative splicing events in ten adult mouse tissues (Pan *et al.*, 2004). Six probes were designed for each splicing event, three exon body probes and three exon junction probes, and the results indicated that alternative splicing and transcription act independently on different sets of genes to define tissue-specific expression profiles.

The most current microarray platform, the Affymetrix Genechip Exon 1.0 ST Array, contains >5.5 million 25 nucleotide probes investigating the expression of ~1 million known and predicted exons. This exon array can be used for detection of alternative splicing, but since all probes on the exon array are exon body probes it can also be used as an expression array measuring the expression at the gene level. Clark and colleagues (2007) used this platform to identify tissue-specific alternative splicing in a panel of 16 diverse normal human tissues. Experimental validation was successful for 86% of the selected

candidates (72 of 86) and pairwise comparisons of tissues suggested that 73% of detected genes are alternatively spliced. Cancer-specific splicing has also been detected by this platform in glial brain tumors (French *et al.*, 2007), colon cancer (Gardina *et al.*, 2006), and in colon, bladder, and prostate cancer (Thorsen *et al.*, 2008). These studies indicate that it is possible to use alternative splice variants to classify tumor subgroups and that these splice variants may be used as novel biomarkers and drug targets. Exon array data have also been compared with the SNP data for well-characterized cell lines (HapMap) to detect the impact of genetic variation on alternative splicing (Kwan *et al.*, 2007, 2008).

Data analysis of splicing microarrays is complex as the expression of a specific exon in a mature mRNA isoform is dependent on both the overall expression of the gene as well as the inclusion rate of the exon. To identify alternative splicing it is therefore important to discriminate between changes in gene expression and changes in splicing of a specific gene. Furthermore, many different splice variants may exist for a single gene and data analysis must be able to discriminate these. Many algorithms for detection of alternative splicing have already been developed but experimental validation of the results is still necessary to validate splicing events and elucidate the sometimes many alternative splice variants of a gene.

16.6 RNAi ARRAYS – HIGH-THROUGHPUT FUNCTIONAL TESTING

Inhibition of gene expression using the RNA interference (RNAi) pathway mediated by small interfering RNAs (siRNA) has rapidly become the method of choice for studying gene function in mammalian cells. This technology allows the use of RNAi to study gene function in mammalian model systems in which classical methods are often limited and costly. One of the advantages of RNAi is that it is highly specific and that it can knock down a gene by 70–100%. RNAi does not rely on interference with the translation, unlike antisense oligonucleotides, but rather leads to degradation of the mRNA through the cells' own system. Although the treatment of cells can lead to unwanted induction of the interferon inflammatory pathway, much of this can be avoided by maintaining RNAi concentrations at a reasonable level.

One limitation of the technique has been the uncertainty in predicting the efficacy of siRNAs in silencing a gene. To overcome this problem, siRNA sequences may be monitored by their ability to reduce the expression of cognate target-reporter fusions with easily quantified readouts for the rapid and efficient identification of the most effective siRNA against any gene (Kumar *et al.*, 2003). A combination of this approach with microarray-based cell transfections has an unlimited potential in high-throughput screens for identifying effective siRNA probes for silencing genes in mammalian systems. This method facilitates the development of large-scale siRNA libraries for large-scale functional genomic studies.

The microarray technology often identifies a large number of genes of interest for further functional studies. Recently, the development of a new exciting tool has carried the RNAi technology into the microarray world, allowing highly parallel analysis used on a genome-wide scale (Mousses *et al.*, 2003). In this technology, siRNA probes are dissolved in a transfection matrix and spotted on poly-L lysine glass slides very much similar to cDNA microarrays. With a spot diameter of 100–500 mm and a center–center distance of 300–1,000 mm, there is space for 2,000 to 15,000 spots on a standard glass slide. Subsequently, the slide is overlaid with a monolayer of adherent cells in a cell tray and incubated to allow reverse transfection. For assessment of the effects of gene knock-down a highly magnified digital image is recorded with a charge-coupled device (CCD) camera. The gene silencing may be combined with other experiments, such as drug treatment, starvation, and heat shock, and scored with the only limitation that the effect must induce a visible change in phenotype.

16.7 FUTURE ASPECTS OF THE USE OF MICROARRAYS

The development of nano-technologies may bring the use of microarrays to a new level. As is stands today the whole human genome can be included in one microarray, looking at gene expression (see also next chapter). However, there are still unmet needs in terms of analysis of the variation occurring in the human genome and how this may alter cellular behavior in relation to physiological and disease processes. The number of splice variants per gene has been estimated to be at least 5–10 per gene and the total number of SNPs to be more than 5 million. To be able to analyze these very large numbers of variables, new and denser array formats are needed. It also poses a statistical problem on the analysis in terms of the requirements to the materials to be analyzed. Large tissue banks will have to be developed, for example if disease-related variations in the genome have to be identified.

Recent years have demonstrated that microarrays could have an important role in parallel analysis of cellular parameters in combination with drugs, genes, viruses, RNAi, and so on. The micro- or nano-format makes it possible to screen for a number of variables simultaneously, for example during treatment with a new drug candidate. The endpoints have to be measured in parallel in hundreds or thousands of small cell cultures. This is a challenge, as functional assays will have to be developed for this purpose such as apoptosis assays, cell cycle assays, cell kinetic assays, etc. Proteomic arrays have not been dealt

with in this chapter but are, of course, another array area where much is expected in the future. Some assays based on the function of molecules at DNA or peptide level are promising, such as those for phosphorylation of peptides or binding assays in which binding of a transcription factor to DNA is analyzed with respect to binding inhibitors. Such biochemical process arrays are very promising as they make it possible to screen large libraries of compounds for very specific endpoints.

Most array work has aimed at identifying changes in gene expression in clinical samples or *in vitro* in cell lines. The pathways and genes that change their expression are identified and this can be followed by *in vitro* testing of the effect of up- or down-regulation of the expression of the gene. This effect can again be analyzed by microarrays in a time course study of the gene expression and shifts in pathways. This situation in which a single parameter is changed and the effect is measured is quite straightforward. However, in the future more than one parameter will have to be altered, such as inhibiting one pathway and promoting another to get more information on the dynamic possibilities in cellular responses. To study such changes with appropriate controls and many time points is costly, and a single factor that will promote this development is the lower cost of arrays. A reduced microarray cost is needed to be able to use multiple arrays for analysis of complicated cellular networks, for example by using Bayesian network modeling. Apart from such increased use of arrays in defining cellular networks, it is expected that a clinical use of microarrays will be gradually introduced, especially in relation to cancer classification. In breast, lymphoma, leukemia, and bladder malignancies classifiers have already been published and are in clinical testing. They will make it possible to select the right treatment for the individual patient and to develop individual predictors of the disease course needed for individual follow-up programs. Such a use of microarrays may lead to great benefit for the society and to reduced spending on ineffective treatments, and selection of tailored and more effective treatments.

REFERENCES

Abelson, J.F., Kwan, K.Y., O'Roak, B.J., Baek, D.Y., Stillman, A.A., Morgan, T.M., Mathews, C.A., Pauls, D.L., Rasin, M.R., Gunel, M., Davis, N.R., Ercan-Sencicek, A.G., Guez, D.H., Spertus, J.A., Leckman, J.F., Dure, L.S., 4th, Kurlan, R., Singer, H.S., Gilbert, D.L., Farhi, A., Louvi, A., Lifton, R.P., Sestan, N., and State, M.W. (2005). Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science* 310, 317–320.

Ahrendt, S.A., Halachmi, S., Chow, J.T., Wu, L., Halachmi, N., Yang, S. C., Wehage, S., Jen, J., and Sidransky, D. (1999). Rapid p53 sequence analysis in primary lung cancer using an oligonucleotide probe array. *Proc. Natl. Acad. Sci. USA* 96, 7382–7387.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J.C., Botstein, D., Brown, P.O., and Staudt, L.M. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503–511.

Als, A.B., Dyrskjøt, L., von der Maase, H., Koed, K., Mansilla, F., Toldbod, H.E., Jensen, J.L., Ulhøi, B.P., Sengeløv, L., Jensen, K.M., and Orntoft, T.F. (2007). Emmprin and survivin predict response and survival following cisplatin-containing chemotherapy in patients with advanced bladder cancer. *Clin. Cancer Res.* 1, 4407–4414.

Andersen, C.L., Wiuf, C., Kruhøffer, M., Korsgaard, M., Laurberg, S., and Ørntoft, T.F. (2007). Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis* 28, 38–48.

Barrett, J.C., and Cardon, L.R. (2006). Evaluating coverage of genome-wide association studies. *Nat. Genet.* 38, 659–662.

Beaudet, A.L., and Belmont, J.W. (2008). Array-based DNA Diagnostics: Let the revolution begin. *Annu. Rev. Med.* 59, 113–129.

Beer, D.G., Kardia, S.L., Huang, C.C., Giordano, T.J., Levin, A. M., Misek, D.E., Lin, L., Chen, G., Gharib, T.G., Thomas, D.G., Lizyness, M.L., Kuick, R., Hayasaka, S., Taylor, J.M., Iannettoni, M. D., Orringer, M.B., and Hanash, S. (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.* 8, 816–824.

Beroukhim, R., Lin, M., Park, Y., Hao, K., Zhao, X., Garraway, L.A., Fox, E.A., Hochberg, E.P., Mellinghoff, I.K., Hofer, M.D., Descazeaud, A., Rubin, M.A., Meyerson, M., Wong, W.H., Sellers, W.R., and Li, C. (2006). Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide arrays. *PLoS Comput. Biol.* 2, e41.

Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E.J., Lander, E.S., Wong, W., Johnson, B.E., Golub, T. R., Sugarbaker, D.J., and Meyerson, M. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* 98, 13790–13795.

Broderick, P., Carvajal-Carmona, L., Pittman, A.M., Webb, E., Howarth, K., Rowan, A., Lubbe, S., Spain, S., Sullivan, K., Fielding, S., Jaeger, E., Vijayakrishnan, J., Kemp, Z., Gorman, M., Chandler, I., Papaemmanuil, E., Penegar, S., Wood, W., Sellick, G., Qureshi, M., Teixeira, A., Domingo, E., Barclay, E., Martin, L., Sieber, O., CORGI Consortium, Kerr, D., Gray, R., Peto, J., Cazier, J.B., Tomlinson, I., and Houlston, R.S. (2007). A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat. Genet.* 39, 1315–1317.

Buschiazzo, E., and Gemmell, N.J. (2006). The rise, fall and renaissance of microsatellites in eucaryotic genomes. *Bioessays* 28, 1040–1050.

Calin, G.A., Liu, C.G., Sevignani, C., Ferracin, M., Felli, N., Dumitru, C. D., Shimizu, M., Cimmino, A., Zupo, S., Dono, M., Dell'Aquila, M. L., Alder, H., Rassenti, L., Kipps, T.J., Bullrich, F., Negrini, M., and Croce, C.M. (2004). MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc. Natl. Acad. Sci. USA* 101, 11755–11760.

Calin, G.A., Ferracin, M., Cimmino, A., Di Leva, G., Shimizu, M., Wojcik, S.E., Iorio, M.V., Visone, R., Sever, N.I., Fabbri, M., Iuliano, R., Palumbo, T., Pichiorri, F., Roldo, C., Garzon, R., Sevignani, C., Rassenti, L., Alder, H., Volinia, S., Liu, C.G., Kipps, T.J., Negrini, M., and Croce, C.M. (2005). A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N. Engl. J. Med.* 353, 1793–17801.

Calin, G.A., and Croce, C.M. (2006). MicroRNA signatures in human cancers. *Nat. Rev. Cancer* 6, 857–866.

Catalucci, D., Latronico, M.V., and Condorelli, G. (2008). MicroRNAs control gene expression: importance for cardiac development and pathophysiology. *Ann. NY Acad. Sci.* 1123, 20–29.

Chang, H.Y., Nuyten, D.S., Sneddon, J.B., Hastie, T., Tibshirani, R., Sørlie, T., Dai, H., He, Y.D., van't Veer, L.J., Bartelink, H., van de Rijn, M., Brown, P.O., and van de Vijver, M.J. (2005). Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc. Natl. Acad. Sci. USA* 8, 3738–3743.

Ciafrè, S.A., Galardi, S., Mangiola, A., Ferracin, M., Liu, C.G., Sabatino, G., Negrini, M., Maira, G., Croce, C.M., and Farace, M.G. (2005). Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochem. Biophys. Res. Commun.* 34, 1351–1358.

Clark, T.A., Schweitzer, A.C., Chen, T.X., Staples, M.K., Lu, G., Wang, H., Williams, A., and Blume, J.E. (2007). Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol.* 8, R64.

Davison, T.S., Johnson, C.D., and Andruss, B.F. (2006). Analyzing microRNA expression using microarrays. *Methods Enzymol.* 411, 14–34.

Dudoit, S., Fridlyand, J., and Speed, T.P. (2000). Comparison of discrimination methods for the classification of tumors using gene expression data. Technical Report, Berkeley University.

Dyrskjøt, L., Thykjaer, T., Kruhoffer, M., Jensen, J.L., Marcussen, N., Hamilton-Dutoit, S., Wolf, H., and Orntoft, T.F. (2003). Identifying distinct classes of bladder carcinoma using microarrays. *Nat. Genet.* 33, 90–96.

Dyrskjøt, L., Zieger, K., Kruhøffer, M., Thykjaer, T., Jensen, J.L., Primdahl, H., Aziz, N., Marcussen, N., Møller, K., and Orntoft, T.F. (2005). A molecular signature in superficial bladder carcinoma predicts clinical outcome. *Clin. Cancer Res.* 11, 4029–4036.

Dyrskjøt, L., Zieger, K., Real, F.X., Malats, N., Carrato, A., Hurst, C., Kotwal, S., Knowles, M., Malmström, P.U., de la Torre, M., Wester, K., Allory, Y., Vordos, D., Caillault, A., Radvanyi, F., Hein, A.M., Jensen, J. L., Jensen, K.M., Marcussen, N., and Orntoft, T.F. (2007). Gene expression signatures predict outcome in non-muscle-invasive bladder carcinoma: a multicenter validation study. *Clin. Cancer Res.* 13, 3545–3551.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 95, 14863–14868.

Esquela-Kerscher, A., and Slack, F.J. (2006). Oncomirs – microRNAs with a role in cancer. *Nat. Rev. Cancer* 6, 259–269.

Fan, J.B., Chen, X., Halushka, M.K., Berno, A., Huang, X., Ryder, T., Lipshutz, R.J., Lockhart, D.J., and Chakravarti, A. (2000). Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res.* 10, 853–860.

French, P.J., Peeters, J., Horsman, S., Duijm, E., Siccama, I., van den Bent, M.J., Luider, T.M., Kros, J.M., van der Spek, P., and Sillevis Smitt, P.A. (2007). Identification of differentially regulated splice variants and novel exons in glial brain tumors using exon expression arrays. *Cancer Res.* 67, 5635–5642.

Fritz, B., Schubert, F., Wrobel, G., Schwaenen, C., Wessendorf, S., Nessling, M., Korz, C., Rieker, R.J., Montgomery, K., Kucherlapati, R., Mechtersheimer, G., Eils, R., Joos, S., and Lichter, P. (2002). Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. *Cancer Res.* 62, 2993–2998.

Garber, K., Smith, K.T., Reines, D., and Warren, S.T. (2006). Transcription, translation and fragile X syndrome. *Curr. Opin. Genet. Dev.* 16, 270–275.

Gardina, P.J., Clark, T.A., Shimada, B., Staples, M.K., Yang, Q., Veitch, J., Schweitzer, A., Awad, T., Sugnet, C., Dee, S., Davies, C., Williams, A., and Turpaz, Y. (2006). Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* 7, 325.

Gillies, J.K., and Lorimer, I.A. (2007). Regulation of p27Kip1 by miRNA 221/222 in glioblastoma. *Cell Cycle* 6, 2005–2009.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., and Lander, E.S. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537.

Gunderson, K.L., Kuhn, K.M., Steemers, F.J., Ng, P., Murray, S.S., and Shen, R. (2006). Whole-genome genotyping of haplotype tag single nucleotide polymorphisms. *Pharmacogenomics* 7, 641–648.

Guo, Y., Chen, Z., Zhang, L., Zhou, F., Shi, S., Feng, X., Li, B., Meng, X., Ma, X., Luo, M., Shao, K., Li, N., Qiu, B., Mitchelson, K., Cheng, J., and He, J. (2008). Distinctive microRNA profiles relating to patient survival in esophageal squamous cell carcinoma. *Cancer Res.* 68, 26–33.

Haiman, C.A., Le Marchand, L., Yamamato, J., Stram, D.O., Sheng, X., Kolonel, L.N., Wu, A.H., Reich, D., and Henderson, B.E. (2007). A common genetic risk factor for colorectal and prostate cancer. *Nat. Genet.* 39, 954–956.

He, H., Jazdzewski, K., Li, W., Liyanarachchi, S., Nagy, R., Volinia, S., Calin, G.A., Liu, C.G., Franssila, K., Suster, S., Kloos, R.T., Croce, C.M., and de la Chapelle, A. (2005). The role of microRNA genes in papillary thyroid carcinoma. *Proc. Natl. Acad. Sci. USA* 102, 19075–19080.

Heinrichs, S., and Look, A.T. (2007). Identification of structural aberrations in cancer by SNP array analysis. *Genome Biology* 8, 219.1–219.5.

Heller, M.J. (2002). DNA microarray technology: devices, systems, and applications. *Annu. Rev. Biomed. Eng.* 4, 129–153.

Holden, A.L. (2002). The SNP consortium: summary of a private consortium effort to develop an applied map of the human genome. *Biotechniques Suppl.* 22–24, 26.

Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S. E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., Wang, J., Yu, K., Chatterjee, N., Orr, N., Willett, W.C., Colditz, G.A., Ziegler, R.G., Berg, C.D., Buys, S.S., McCarty, C.A., Feigelson, H.S., Calle, E.E., Thun, M.J., Hayes, R.B., Tucker, M., Gerhard, D.S., Fraumeni, J.F., Jr., Hoover, R.N., Thomas, G., and Chanock, S.J. (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* 39, 870–874.

Hyman, E., Kauraniemi, P., Hautaniemi, S., Wolf, M., Mousses, S., Rozenblum, E., Ringner, M., Sauter, G., Monni, O., Elkahloun, A., Kallioniemi, O.P., and Kallioniemi, A. (2002). Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res.* 62, 6240–6245.

Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M., Ménard, S., Palazzo, J.P., Rosenberg, A., Musiani, P., Volinia, S., Nenci, I., Calin, G.A., Querzoli, P., Negrini, M., and Croce, C.M. (2005). MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* 65, 7065–7070.

Iorio, M.V., Visone, R., Di Leva, G., Donati, V., Petrocca, F., Casalini, P., Taccioli, C., Volinia, S., Liu, C.G., Alder, H., Calin, G.A., Ménard, S., and Croce, C.M. (2007). MicroRNA signatures in human ovarian cancer. *Cancer Res.* 67, 8699–86707.

Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P. M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., and Shoemaker, D.D. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302, 2141–2144.

Kennedy, G.C., Matsuzaki, H., Dong, S., Liu, W.M., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J., Liu, W., Yang, G., Di, X., Ryder, T., He, Z., Surti, U., Phillips, M.S., Boyce- Jacino, M.T., Fodor, S.P., and Jones, K.W. (2003). Large-scale genotyping of complex DNA. *Nat. Biotechnol.* 21, 1233–1237.

Koed, K., Wiuf, C., Christensen, L.L., Wikman, F.P., Zieger, K., Møller, K., von der Maase, H., and Orntoft, T.F. (2005). High-density single nucleotide polymorphism array defines novel stage and location-dependent allelic imbalances in human bladder tumors. *Cancer Res.* 65, 34–45.

Krichevsky, A.M. (2007). MicroRNA profiling: from dark matter to white matter, or identifying new players in neurobiology. *Scientific World Journal* 7, 155–166.

Kruglyak, L. (2008). The road to genome-wide association studies. *Nat. Rev. Genet.* 9, 314–318.

Kumar, R., Conklin, D.S., and Mittal, V. (2003). High-throughput selection of effective RNAi probes for gene silencing. *Genome Res.* 13, 2333–2340.

Kwan, T., Benovoy, D., Dias, C., Gurd, S., Provencher, C., Beaulieu, P., Hudson, T.J., Sladek, R., and Majewski, J. (2008). Genome-wide analysis of transcript isoform variation in humans. *Nat. Genet.* 40, 225–231.

Kwan, T., Benovoy, D., Dias, C., Gurd, S., Serre, D., Zuzan, H., Clark, T.A., Schweitzer, A., Staples, M.K., Wang, H., Blume, J.E., Hudson, T.J., Sladek, R., and Majewski, J. (2007). Heritability of alternative splicing in the human genome. *Genome Res.* 17, 1210–1218.

Laken, S.J., Petersen, G.M., Gruber, S.B., Oddoux, C., Ostrer, H., Giardiello, F.M., Hamilton, S.R., Hampel, H., Markowitz, A., Klimstra, D., Jhanwar, S., Winawer, S., Offit, K., Luce, M.C., Kinzler, K.W., and Vogelstein, B. (1997). Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat. Genet.* 17, 79–83.

Liu, W., Ewing, C.M., Chang, B.L., Li, T., Sun, J., Turner, A.R., Dimitrov, L., Zhu, Y., Sun, J., Kim, J.W., Zheng, S.L., Isaacs, W.B., and Xu, J. (2007). Multiple genomic alterations on 21q22 predict various TMPRSS2/ERG fusion transcripts in human prostate cancers. *Genes Chromosomes Cancer* 46, 972–980.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R., and Golub, T.R. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435, 834–838.

Lu, Y., Lemon, W., Liu, P.Y., Yi, Y., Morrison, C., Yang, P., Sun, Z., Szoke, J., Gerald, W.L., Watson, M., Govindan, R., and You, M. (2006). A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. *PLoS Med.* 3, e467.

Mei, R., Galipeau, P.C., Prass, C., Berno, A., Ghandour, G., Patil, N., Wolff, R.K., Chee, M.S., Reid, B.J., and Lockhart, D.J. (2000). Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Res.* 10, 1126–1137.

Monni, O., Barlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M.L., and Kallioniemi, A. (2001). Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. USA* 98, 5711–5716.

Mousses, S., Caplen, N.J., Cornelison, R., Weaver, D., Basik, M., Hautaniemi, S., Elkahloun, A.G., Lotufo, R.A., Choudary, A., Dougherty, E.R., Suh, E., and Kallioniemi, O. (2003). RNAi microarray analysis in cultured mammalian cells. *Genome Res.* 13, 2341–2347.

Murakami, Y., Yasuda, T., Saigo, K., Urashima, T., Toyoda, H., Okanoue, T., and Shimotohno, K. (2006). Comprehensive analysis of microRNA expression patterns in hepatocellular carcinoma and non-tumorous tissues. *Oncogene* 25, 2537–2545.

Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F.L., Watson, D., Bryant, J., Costantino, J.P., Geyer, C.E., Jr., Wickerham, D.L., and Wolmark, N. (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J. Clin. Oncol.* 24, 3726–3734.

Pan, Q., Shai, O., Misquitta, C., Zhang, W., Saltzman, A.L., Mohammad, N., Babak, T., Siu, H., Hughes, T.R., Morris, Q.D., Frey, B.J., and Blencowe, B.J. (2004). Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* 16, 929–941.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lonning, P.E., Borresen-Dale, A.L., Brown, P.O., and Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747–752.

Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., and Brown, P.O. (1999). Genome-wide analysis of DNA copynumber changes using cDNA microarrays. *Nat. Genet.* 23, 41–46.

Pollack, J.R., Sorlie, T., Perou, C.M., Rees, C.A., Jeffrey, S.S., Lonning, P.E., Tibshirani, R., Botstein, D., Borresen-Dale, A.L., and Brown, P.O. (2002). Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl. Acad. Sci. USA* 99, 12963–12968.

Porkka, K.P., Pfeiffer, M.J., Waltering, K.K., Vessella, R.L., Tammela, T. L., and Visakorpi, T. (2007). MicroRNA expression profiling in prostate cancer. *Cancer Res.* 67, 6130–6135.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T. D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C. R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K. W., Scherer, S.W., and Hurles, M.E. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454.

Roldo, C., Missiaglia, E., Hagan, J.P., Falconi, M., Capelli, P., Bersani, S., Calin, G.A., Volinia, S., Liu, C.G., Scarpa, A., and Croce, C.M. (2006). MicroRNA expression abnormalities in pancreatic endocrine and acinar tumors are associated with distinctive pathologic features and clinical behavior. *J. Clin. Oncol.* 24, 4677–4684.

Rosenfeld, N., Aharonov, R., Meiri, E., Rosenwald, S., Spector, Y., Zepeniuk, M., Benjamin, H., Shabes, N., Tabak, S., Levy, A., Lebanony, D., Goren, Y., Silberschein, E., Targan, N., Ben-Ari, A., Gilad, S., Sion-Vardy, N., Tobar, A., Feinmesser, M., Kharenko, O., Nativ, O., Nass, D., Perelman, M., Yosepovich, A., Shalmon, B., Polak-Charcon, S., Fridman, E., Avniel, A., Bentwich, I., Bentwich, Z., Cohen, D., Chajut, A., and Barshack, I. (2008). MicroRNAs accurately identify cancer tissue origin. *Nat Biotechnol.* 26, 462–469.

Rosenwald, A., Wright, G., Chan, W.C., Connors, J.M., Campo, E., Fisher, R.I., Gascoyne, R.D., Muller-Hermelink, H.K., Smeland, E. B., Giltnane, J.M., Hurt, E.M., Zhao, H., Averett, L., Yang, L., Wilson, W.H., Jaffe, E.S., Simon, R., Klausner, R.D., Powell, J., Duffey, P.L.,

Longo, D.L., Greiner, T.C., Weisenburger, D.D., Sanger, W.G., Dave, B.J., Lynch, J.C., Vose, J., Armitage, J.O., Montserrat, E., Lopez-Guillermo, A., Grogan, T.M., Miller, T.P., LeBlanc, M., Ott, G., Kvaloy, S., Delabie, J., Holte, H., Krajci, P., Stokke, T., and Staudt, L.M. (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* 346, 1937–1947.

Snijders, A.M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J.P., Gray, J.W., Jain, A.N., Pinkel, D., and Albertson, D.G. (2001). Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.* 29, 263–264.

Stacey, S.N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S.A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason, A., Aben, K.K., Strobbe, L.J., Albers-Akkers, M.T., Swinkels, D.W., Henderson, B.E., Kolonel, L.N., Le Marchand, L., Millastre, E., Andres, R., Godino, J., Garcia-Prats, M.D., Polo, E., Tres, A., Mouy, M., Saemundsdottir, J., Backman, V.M., Gudmundsson, L., Kristjansson, K., Bergthorsson, J.T., Kostic, J., Frigge, M.L., Geller, F., Gudbjartsson, D., Sigurdsson, H., Jonsdottir, T., Hrafnkelsson, J., Johannsson, J., Sveinsson, T., Myrdal, G., Grimsson, H.N., Jonsson, T., von Holst, S., Werelius, B., Margolin, S., Lindblom, A., Mayordomo, J.I., Haiman, C.A., Kiemeney, L.A., Johannsson, O.T., Gulcher, J.R., Thorsteinsdottir, U., Kong, A., and Stefansson, K. (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* 39, 865–869.

Steemers, F.J., and Gunderson, K.L. (2007). Whole genome genotyping technologies on the BeadArray platform. *Biotechnol. J.* 2, 41–49.

Takata, R., Katagiri, T., Kanehira, M., Tsunoda, T., Shuin, T., Miki, T., Namiki, M., Kohri, K., Matsushita, Y., Fujioka, T., and Nakamura, Y. (2005). Predicting response to methotrexate, vinblastine, doxorubicin, and cisplatin neoadjuvant chemotherapy for bladder cancers through genome-wide gene expression profiling. *Clin. Cancer Res.* 11, 2625–2636.

Tavazoie, S.F., Alarcón, C., Oskarsson, T., Padua, D., Wang, Q., Bos, P.D., Gerald, W.L., and Massagué, J. (2008). Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* 451, 147–152.

Tili, E., Michaille, J.J., and Calin, G.A. (2008). Expression and function of micro-RNAs in immune cells during normal or disease state. *Int. J. Med. Sci.* 5, 73–79.

The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature* 437, 1299–1320.

The International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861.

Thorsen, K., Sorensen, K.D., Brems-Eskildsen, A.S., Modin, C., Gaustadnes, M., Hein, A.M., Kruhoffer, M., Laurberg, S., Borre, M., Wang, K., Brunak, S., Krainer, A.R., Torring, N., Dyrskjot, L., Andersen, C.L., and Orntoft, T.F. (2008). Alternative splicing in colon, bladder, and prostate cancer identified by exon-array analysis. *Mol. Cell Proteomics*. (in press).

Thykjaer, T., Workman, C., Kruhoffer, M., Demtroder, K., Wolf, H., Andersen, L.D., Frederiksen, C.M., Knudsen, S., and Orntoft, T.F. (2001). Identification of gene expression patterns in superficial and invasive human bladder cancer. *Cancer Res.* 61, 2492–2499.

Tørring, N., Borre, M., Sørensen, K.D., Andersen, C.L., Wiuf, C., and Ørntoft, T.F. (2007). Genome-wide analysis of allelic imbalance in prostate cancer using the Affymetrix 50 K SNP mapping array. *Br. J. Cancer* 96, 499–506.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E.T., Friend, S. H., and Bernards, R. (2002). A gene expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 347, 1999–2009.

van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., and Friend, S.H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536.

Veltman, J.A., Fridlyand, J., Pejavar, S., Olshen, A.B., Korkola, J.E., DeVries, S., Carroll, P., Kuo, W.L., Pinkel, D., Albertson, D., Cordon-Cardo, C., Jain, A.N., and Waldman, F.M. (2003). Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. *Cancer Res.* 63, 2872–2880.

Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., Prueitt, R.L., Yanaihara, N., Lanza, G., Scarpa, A., Vecchione, A., Negrini, M., Harris, C.C., and Croce, C.M. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. USA* 103, 2257–22561.

Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lipshutz, R., Chee, M., and Lander, E.S. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280, 1077–1082.

Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J., Jatkoe, T., Berns, E.M., Atkins, D., and Foekens, J.A. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365, 671–679.

Weiss, M.M., Kuipers, E.J., Postma, C., Snijders, A.M., Siccama, I., Pinkel, D., Westerga, J., Meuwissen, S.G., Albertson, D.G., and Meijer, G.A. (2003). Genomic profiling of gastric cancer predicts lymph node status and survival. *Oncogene* 22, 1872–1879.

Wild, P.J., Herr, A., Wissmann, C., Stoehr, R., Rosenthal, A., Zaak, D., Simon, R., Knuechel, R., Pilarsky, C., and Hartmann, A. (2005). Gene expression profiling of progressive papillary noninvasive carcinomas of the urinary bladder. *Clin. Cancer Res.* 11, 4415–4429.

Wilhelm, M., Veltman, J.A., Olshen, A.B., Jain, A.N., Moore, D.H., Presti, J.C., Jr., Kovacs, G., and Waldman, F.M. (2002). Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer. *Cancer Res.* 62, 957–960.

Witte, J.S. (2007). Multiple prostate cancer risk variants on 8q24. *Nat. Genet.* 39, 579–580.

Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., Stephens, R.M., Okamoto, A., Yokota, J., Tanaka, T., Calin, G.A., Liu, C. G., Croce, C.M., and Harris, C.C. (2006). Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* 9, 189–198.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* 30, e15.

Zheng, S.L., Sun, J., Wiklund, F., Smith, S., Stattin, P., Li, G., Adami, H. O., Hsu, F.C., Zhu, Y., Bälter, K., Kader, A.K., Turner, A.R., Liu, W., Bleecker, E.R., Meyers, D.A., Duggan, D., Carpten, J.D., Chang, B. L., Isaacs, W.B., Xu, J., and Grönberg, H. (2008). Cumulative association of five genetic variants with prostate cancer. *N. Engl. J. Med.* 28(358), 910–918.

This page intentionally left blank

Chapter 17

Arrayed Primer Extension Microarrays for Molecular Diagnostics

Neeme Tonisson[1,3], Eneli Oitmaa[1], Kaarel Krjutskov[1], Janne Pullat[1], Ilona Lind[2,3], Merike Leego[1,2], Ants Kurg[1,2] and Andres Metspalu[1,2,4]

[1]*Institute of Molecular and Cell Biology, University of Tartu/Estonian Biocentre, Tartu, Estonia;*

[2]*Asper Biotech, Tartu, Estonia;*

[3]*Tartu University Hospital, Estonia;*

[4]*Estonian Genome Project, University of Tartu*

17.1 INTRODUCTION

Advances in genomics and human genetics have transferred DNA diagnostics firmly into medical practice in advanced hospitals and medical centers. There are several technological platforms in use and in this chapter we are introducing a microarray-based approach – namely Arrayed Primer Extension (APEX). This method is well suited for both applications: testing for known disease-causing mutations and analyzing single nucleotide polymorphisms (SNPs) in candidate regions where tens to thousand of markers need to be tested using custom-made arrays. The latter is suited for research laboratories. Regarding the testing of known mutations, APEX is not limited only to point mutations. The size of the deletions and insertions can be determined if analysis is carried out simultaneously from both DNA strands, since this gives the starting and ending points of the deletion or insertion in the known DNA sequence. Recently, an updated version of the method, namely APEX-2, has been introduced, which increased the degree of multiplexing by approximately 100-fold for SNP testing. So far, more than 600-plex reactions can be routinely performed in one tube, which reduces substantially the pipetting time and consumables needed.

APEX may be used to discover new mutations in the genes where this is needed (Krjutskov *et al*., 2008). For the *TP53* gene, a resequencing microarray has been developed, which also allows analyzing *de novo* mutations. It can be used for testing of somatic mutations directly from the tumor sample, with as little as 6% of mutated DNA on the background of normal sequence from the normal tissue (Le Calvez *et al*., 2005).

As previously mentioned, DNA testing is currently mostly used for molecular diagnostics of Mendelian diseases, where the results will help physicians to make decisions about diagnoses, risks, and treatment modalities. However, in the case of common complex diseases, the situation is different. It might be tempting to start individual testing for disease-related SNPs recently discovered through large-scale genome-wide association analyses as the ''risk'' alleles, but the predictive value of such testing is in most cases insufficient for personalized recommendations. Because most common diseases are caused by a complex interaction between the different genes, environmental factors and lifestyle elements, each of which will add only a minor part to the overall disease risk. Although we can predict that this kind of SNP testing will be reasonably justified one day, this may still be many years away.

Nowadays, DNA testing is still limited by the availability of suitable analyzing technology (e.g. lack of standardized automated mutation analyzers), comprehensive genetic information on disease-causing mutations, and susceptibility alleles for environmental and lifestyle factors. In order to make more progress in the present situation, we can point to two main factors, which need to be overcome. First, physicians of all specialties will need more comprehensive information about the recent developments in medical genetics. Second, broader availability of international quality control schemes and panels of test samples for complex phenotypes would be very helpful in order to ensure the high quality and proper standards in DNA testing. The availability of test samples is also crucial for any diagnostic laboratories and companies developing new assays.

17.2 APEX TECHNOLOGY OVERVIEW

17.2.1 Microarray Chemistry

The attachment chemistry of oligonucleotides on glass support must provide a reproducible coating process of the slides and functionality of the surface-bound DNA. The reliable chemical linkage must also be stable, produce specific binding, and eliminate undesired steric hindrance.

Genorama™ microarray slides from Asper Biotech have been employed for the APEX analyses. The binding chemistry of Genorama slides basically consists of aminosilane coating (Guo *et al.*, 1994) followed by 1,4-phenylene diisothiocyanate linker treatment to convert the amino groups to amino-reactive phenylisothiocyanate groups. Amino-modified oligonucleotide primers are diluted in alkaline spotting buffer (pH 9.0) and spotted onto the activated surface. The slides are later blocked with either ammonia vapors or 1% solution of ammonium hydroxide, washed with water, air dried, and stored at 4°C. In the authors' experience, the slides have been stable for at least six months. This thiocyanate functionalization has also been used as a reference method for other microarray chemistries (Lindroos *et al.*, 2002).

17.2.2 APEX Primer Design

Primers with 25- to 35-mer specific sequence have been mostly used in all APEX assays performed in the authors' laboratory. The primers are synthesized according to a consensus sequence with aminolink at their 5′ end. The 3′ ends are unmodified and free for extension by the DNA polymerase. The 3′ end of each primer is positioned one base before the base to be identified. For resequencing on the APEX assay, primers are designed with a single base shift on their sequence. For example, 1,000 primers are needed for identification of 1,000 bases in the sample DNA from one strand. As a routine, two primers have been used per each base pair, one for each DNA strand. Stable secondary structures (hairpins and oligonucleotide dimers) should be preferentially avoided. The potentially harmful structures are predictable by computer algorithms and the oligonucleotides may be designed beforehand to reduce stability of the secondary structures. In the case of the *TP53* gene APEX resequencing assay, 5.9% of the primers were redesigned by introducing a mismatch base to reduce the stability of the predicted dimers and avoid self-priming (Tonisson *et al.*, 2002). After this modification, 62% of these primers were found to generate signals only in the presence of target DNA and not from oligonucleotide dimers. The rest of modified primers did not give constant signals in the APEX reactions. Importantly, none of the modified primers were found to generate false-positive signals in the absence of the target DNA fitting well with the ''yes or no'' type general philosophy of APEX analysis.

For a part of the mutation sites analyzed, especially for more complicated mutations, enzymatic elongation of allele-specific primers has been implemented (Gemignani *et al.*, 2002). This modification of primer extension method adds flexibility to the primer design and increases the redundancy and thereby also the reliability of the information obtained. At the negative side, extra primers also somewhat increase the assay development and production costs.

17.2.3 APEX Reaction Conditions

APEX is a complex isothermal reaction consisting of target annealing to the oligonucleotide array and enzymatic primer extension reaction with fluorescent terminator nucleotides (dNTPs). Preparation of template (here synonymous to the target) DNA for primer extension starts from single or multiplex amplification by PCR (Fig. 17.1). The further step has routinely been a simultaneous asymmetrical fragmentation of the double-stranded DNA with uracil DNA glycosylase and degradation of the leftover dNTPs by shrimp alkaline phosphatase. An engineered thermostable DNA polymerase (Thermo Sequenase™, GE Healthcare Life Sciences, Milwaukee, WI, USA) has been used for primer extension (Pastinen *et al.*, 1997; Chen *et al.*, 1999; Kurg *et al.*, 2000). The use of four colors per reaction permits a straightforward simultaneous identification of all possible sequence variants within the same process (Kurg *et al.*, 2000; Lindroos *et al.*, 2001).

The primer extension reactions are routinely performed at 58°C constant temperature for 20 minutes (Kurg *et al.*, 2000; Tonisson *et al.*, 2002). Incorporation of labeled terminators is fast, but hybridization as an equilibrium process needs longer reaction time.

Thermo Sequenase™ and 20 to 50 picomoles of each fluorescent terminator per one APEX reaction have shown optimal efficiency and signal-to-noise ratio. The dye labels used are spectrally well separated (Table 17.1). The current set consists of (by spectral order): fluorescein, Cy3, Texas Red, and Cy5. All but fluorescein are stable to bleaching. Due to the bleaching of fluorescein, the SlowFade® Light antifade reagent (Invitrogen Molecular Probes, Eugene, OR, USA) has been used for imaging.

17.2.4 Imaging

A dedicated instrument, Genorama® QuattroImager total internal reflection (TIRF) microarray imaging system (patent application EP1088214, Asper Biotech, Tartu, Estonia; http://www.asperbio.com), has been developed for the fast analysis of APEX reactions. In the current version, four solid-state lasers, emitting on the wavelengths 473, 532, 594, and 653 nm, excite the fluorescence labels incorporated into the arrayed probes. The optical system consists of

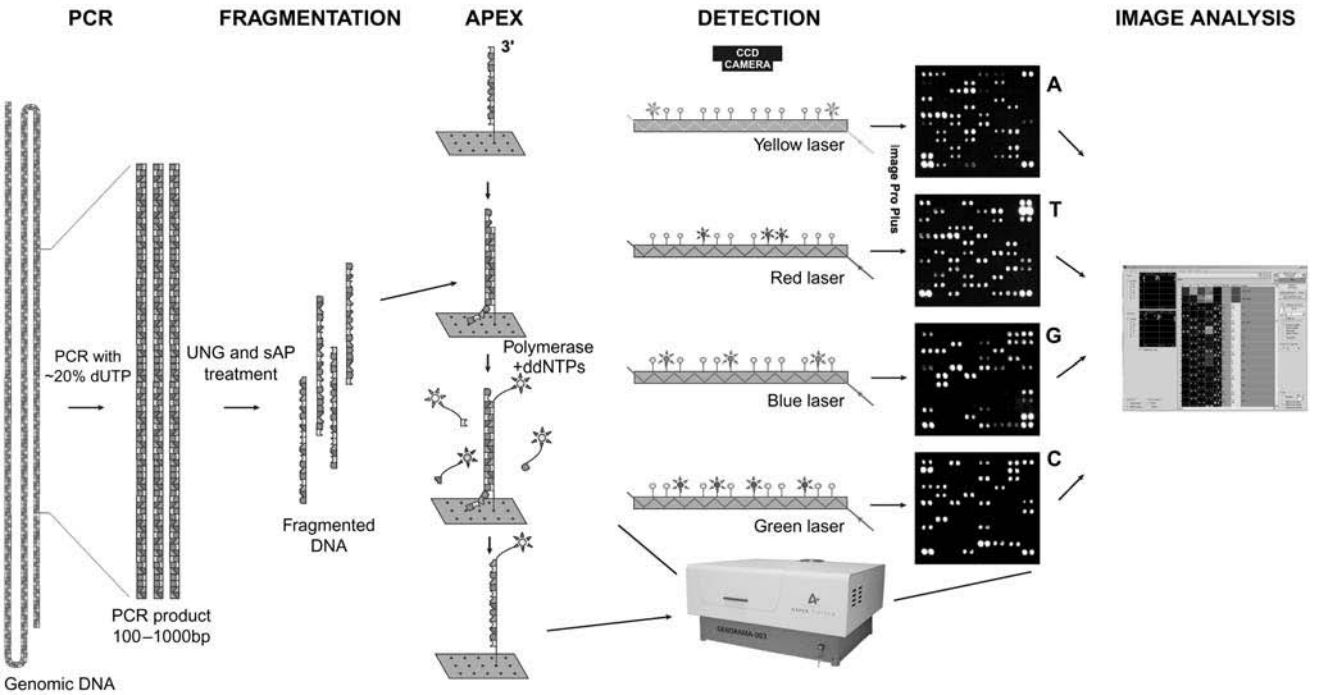


FIGURE 17.1 Schematic drawing of the conventional APEX methodology. The analyzed sequences are first amplified by PCR. A small fraction of dUTP in the reaction mixture allows for random asymmetric fragmentation with uracile DNA glycosylase at a later stage. The fragmented DNA is mixed with a suitable thermostable DNA polymerase and four fluorescently labeled ddNTP-s for APEX reaction. After the single base extension reaction and washing steps, the microarray signals will be quantitated and analyzed.

TABLE 17.1 Fluophores used in four-color primer extension on oligonucleotide microarray (Kurg *et al.*, 2000). The spectral data were obtained from http://www.perkinelmer.com/nucleotide_analogs.

Fluorescent dye	Excitation maximum (nm)	Extinction coefficient ($M^{-1}\cdot cm^{-1}$)	Emission maximum (nm)
Fluorescein	494	30,000	517
Cy™3	550	150,000	568
Texas Red	593	85,000	612
Cy™5	650	250,000	668

the light reflectors, rotating prism, and cylindrical lenses transforming a beam of each laser into a homogeneously illuminated stripe, which is introduced to the microarray slide from one edge. Due to the total internal reflection on the slide surfaces, the incoming beam can spread only inside the slide. Evanescent light field excites the incorporated fluorophores residing near the surface of the slide. A gated cooled charge-coupled device (CCD) camera is used to record the emitted fluorescence. The respective narrow-band interference filters fixed on the revolving wheel in front of the CCD camera depress a noise radiation of exciting laser scattered on the slide. Four images corresponding to four laser wavelengths are captured, one for each dye-labeled terminator nucleotide (Kurg *et al.*, 2000). Additional features of the system include automated laser beam adjustment and autofocusing of the CCD camera lens. Because of precise spectral separation of exciting and emitting wavelengths, the system demonstrates a high signal-to-noise ratio. TIRF excitation also gives a low image background.

Due to multiple variables involved (incorporation efficiency, fluorophore brightness, excitation power, and sensitivity of the detection system at a particular wavelength), the actual signal intensities have to be experimentally determined and used for empirical normalization (Lindroos *et al.*, 2002).

17.2.5 Analysis

The process of APEX data analysis is basically comparable to that of an automated four-color DNA sequencer consisting of intensity comparisons from different fluorescent labels on the same band/dot. The strongest signal is the base called. The specially developed Genorama™ Genotyping Software package for APEX image analysis is available from Asper Biotech. For mutation detection and resequencing, the sequence produced is compared with a reference sequence and diverging bases are indicated.

All divergences and heterozygous positions may be visually verified and the genotypes edited, if necessary. In its latest version, the analysis is recommended in the database (PicDB) format. This enables efficient normalization of the signal intensities all over the samples analyzed and implementation of expectation maximization algorithm (Hua *et al.*, 2007) for automated genotyping.

17.3 LARGE-SCALE MULTIPLEX ANALYSIS: APEX-2

APEX-2 is a further development of the above-described APEX method, permitting simultaneous single-tube amplification of a large number of SNPs and mutations analyzed in the target DNA (Krjutskov *et al.*, 2008). So far, its multiplexing ability has been tested at the 600-plex level.

With APEX-2, just two oligonucleotides are needed for amplification and single base extension (SBE), saving the valuable sample and reducing analysis costs and time. Each APEX-2 primer contains a specific region for genomic binding and a universal 5′ tail for amplification (Fig. 17.2). The binding to genomic DNA takes place adjacent to the studied position. Primers' 3′ end for both DNA strands reside just one base before the SNP or mutation to be addressed. As a final result of the amplification reaction, only the studied nucleotide is novel in the amplicon. The universal 5′ tail is uniform for all APEX-2 primers and guarantees an equal product amplification level for all the amplicons. Being unique against the human genome, synthesis of the complementary sequence for universal tail is a critical step for further amplification. The universal primer (identical to universal tail) can amplify only these products, which have been synthesized previously by specific binding on genome. The preferred amplification or relatively short products (100–200 bp) is also achieved by the short extension time used.

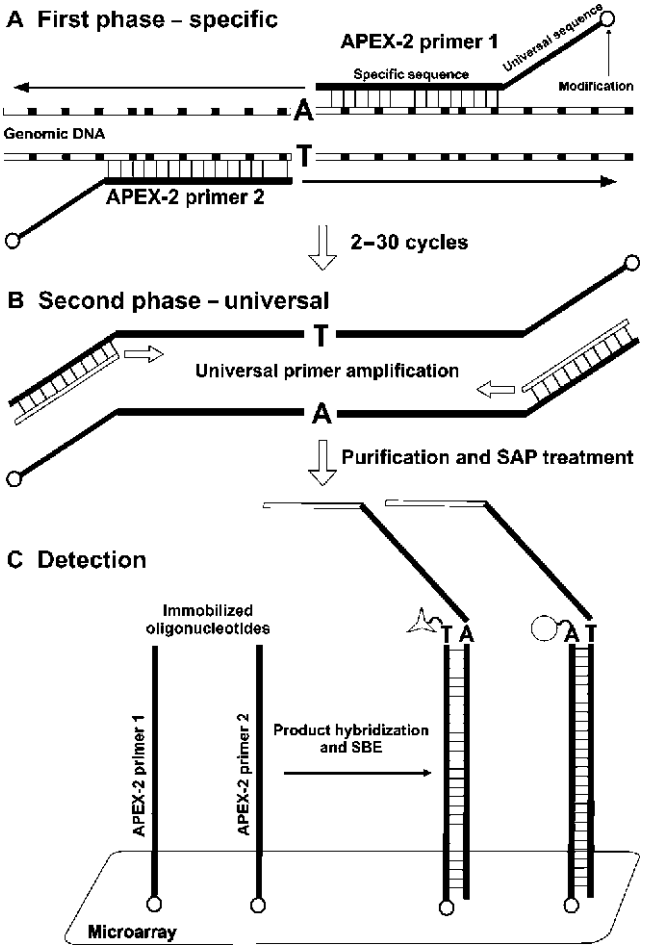


FIGURE 17.2 Large-scale multiplex amplification, APEX-2. APEX-2 primers (APEX-2 primers 1 and 2) bind to genomic DNA sequences immediately adjacent to the position of interest. The primers consist of target-specific and common flanking sequences. This allows for simultaneous multiplex PCR amplification of several hundreds of SNPs and causative mutation sites. After primer extension, the synthesized sequence contains the complement of the respective APEX-2 primer and the position of interest (SNP/causative mutation). The universal primer hybridizes to the 3′ end of the previously generated product. APEX-2 primers have a 5′NH_2 modification, enabling spotting and immobilization on the microarray. The purified universal phase PCR product anneals to the immobilized primers. The genotyping step is a four-color single-base extension reaction.

17.3.1 Primer Design and Marker Selection

The primer design of APEX-2 follows two main rules. First, the selected region is determined by the sequence around the variant position of interest and only the primer length can vary. The average melting temperature (T_m) for all primers is 60°C ± 2°C and primers are not shorter than 18 bp. Also, it is noticed that primers with low GC content, which require more than 35 bp to achieve the desired T_m (i.e. 58°C according to Primer 3 software; Rychlik and Rhoads, 1989) and product length higher than 120 bp, have increased failure rate. The universal 5′ tail consists of 21 bp and has a T_m of 54°C. The primer specificity should be evaluated *in silico* (Andreson *et al.*, 2006) to avoid multiple short products (up to 250 bp).

The primers are mixed together in equal concentration for multiplex PCR. The performance of a primer pair with low or missing signals on the microarray can sometimes be improved by a two-fold increase of both primers' concentration. A more precise optimization is usually not required.

All APEX-2 primers are amino modified at their 5′ end, to enable their covalent immobilization on the microarray surface. This has not caused any observable problems during PCR and column purification.

The selection of the genetic markers to be studied depends on the assay purpose. APEX-2 enables the detection of multiallelic SNPs and different types of mutations. For a single base insertion, the primers are designed similar

to a single base substitution and the fluorescent signal is complementary to the inserted nucleotide. In case of a singe base deletion, the first base of both primers' 3′ end is detected. For larger deletions or insertions (up to tens of nucleotides), APEX-2 primers detect also the first nucleotide of the studied region or the first 3′ base from the opposite primer sequence, respectively. Any SNP linked *in cis* to the sequence variation of interest within the short amplified sequence may cause either an increased blank rate or allele dropout during analysis. In this case, a single pair of primers for both alleles improves the call rate and the recommended amplicon should contain the two polymorphic positions and the sequence in between (Krjutskov *et al.*, 2008).

17.3.2 Advantages and Potential Disadvantages

APEX-2 approach is considered a significant advancement compared to most of the current PCR-based approaches. APEX-2 increases the multiplexing capacity of PCR by 100-fold, being thereby very time and cost effective. The first step of APEX is amplification of genomic regions adjacent to the SNP to produce sufficient template for detection. Similarly, after phase I of APEX-2, the studied position has already been incorporated into the template DNA. Phase II amplification produces a sufficient product quantity for the microarray detection. The single base extension step is similar to APEX SBE.

While other low- to medium-throughput techniques are currently also employed, they require additional steps circumvented by the APEX-2 method, such as including enzymatic treatment of PCR generated probes to remove unincorporated primers, electrophoresis (SNaPshot from Applied Biosystems®) or mass differences (iPLEX from Sequenom®) (Klito *et al.*, 2007). Further, APEX-2 differs from well-known molecular inversion probe (MIP) (Hardenbol *et al.*, 2003) and Illumina's Golden Gate reaction in some essential aspects: APEX-2 primers are linear oligonucleotides that contain neither cleavage sites nor *tag-ctag* system sequences.

The required quantity of genomic DNA per SNP or disease-causing mutation for APEX-2 (0.3 ng) is less than other multiplex PCR-based genotyping methods (Lahermo *et al.*, 2006) and comparable to the iPLEX (0.2–0.3 ng), MIP probe (0.17 ng), and Illumina's Golden Gate assays (0.3–1.4 ng).

Unsuitable for APEX-2 analysis are only the SNPs and mutations, (i) where multiple short amplicons could be generated during PCR, and (ii) with very similar flanking sequences, due to the cross-hybridization considerations. The experimental design allows for flexible assay setup. Projects like 124-plex forensic (SNPs over mitochondrion, Y-, and all autosomal chromosomes) have been successfully introduced (in preparation).

17.4 APEX WITH *IN SITU* SYNTHESIZED OLIGONUCLEOTIDE MICROARRAYS

Sequences of 10–30% oligonucleotide primers may need to be changed to convert all the sites analyzed (mutations/SNPs) into functional assays (Tonisson *et al.*, 2002; Jaakson *et al.*, 2003). This is a time-consuming and costly process because of the repeated orders to be made and because some oligonucleotides will be eventually used only a few times. Use of flexible *in situ* synthesized oligonucleotide microarrays seems to be an efficient and attractive method for the fast and cost-efficient pre-screening of an eventual high-throughput genotyping assay (Pullat *et al.*, 2007).

Usually, the initial microarray compilation is based on an *in silico* oligomer design and on theoretical calculations of interactions between all oligonucleotide probes and PCR fragments. The selected set of oligonucleotide probes is optimized thereafter by a subsequent experimental analysis. The respective oligomers that had performed well during the *in silico* selection process might be empirically found to perform poorly in real experimental reactions (Pullat *et al.*, 2007). The specificity and stability of DNA duplex formation is strongly dependent on the sequence base composition (Wetmur and Davidson, 1968; Breslauer *et al.*, 1986). Also, the target sequence on either side of the SNP position plays an important role since secondary structures may strongly affect the annealing behavior of a target sequence (Southern *et al.*, 1999). Therefore, it is frequently insufficient to predict the performance merely on the basis of theoretical calculations. The ability to easily change the oligonucleotide array layout is essential in the validation stages. Several microarray platforms have utilized light controlled *in situ* oligonucleotide probe synthesis with photolithographic masks, or by micro-mirrors instead (Fodor *et al.*, 1993; Singh-Gasson *et al.*, 1999; Beier and Hoheisel, 2000). The light-induced *in situ* synthesis controlled by a micro-mirror device combines very good synthesis yields of more than 99.5% per condensation (Fig. 17.3) (Beier and Hoheisel, 2000). This enables the production of high-density and -quality oligonucleotide microarrays with flexible layout and reproducible characteristics (Pullat *et al.*, 2007). The photolithographic synthesis of the entire oligonucleotide matrix is performed from photoprotected phosphoramidites by light irradiation. The synthesis using photolabile protecting groups has been successfully used to manufacture high-density microarrays in both the 3′→5′ and 5′→3′ directions (Albert *et al.*, 2003; Pullat *et al.*, 2008). All the designed oligonucleotide primers can be verified at once and redesigned if necessary. Empirical results from earlier experiments exploiting this technology can be immediately applied to the improvement of the next microarray (Bauer *et al.*, 2003). The maskless light-directed *in situ* microarray synthesis technology platform is a fast and cost-effective alternative for assay development, reducing the time needed for ordering new oligonucleotides. Although this platform is not well suited

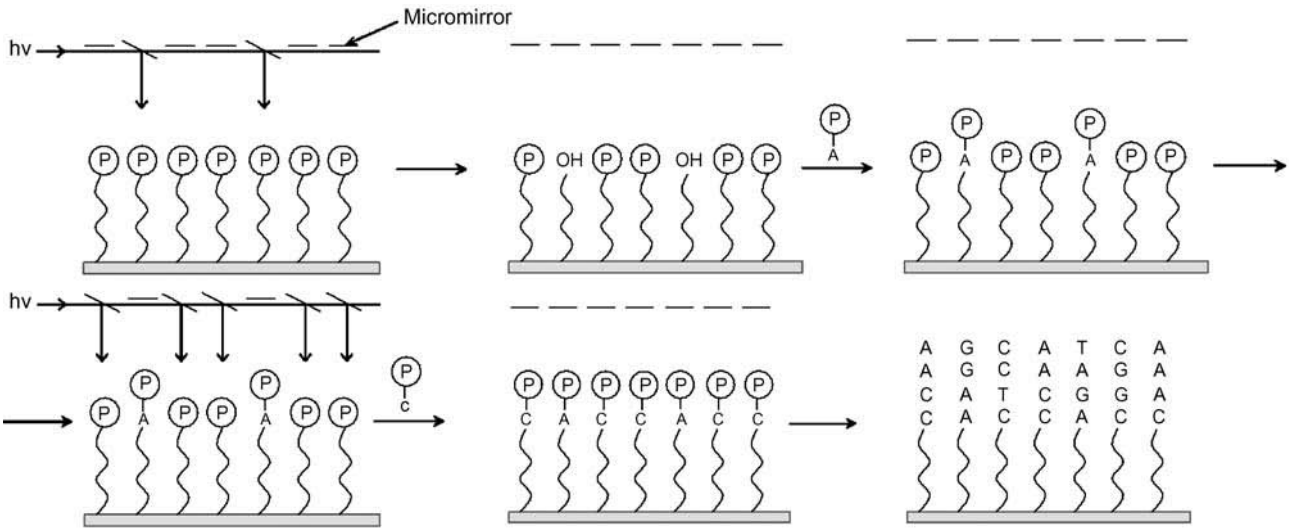


FIGURE 17.3 Maskless microarray *in situ* synthesis steps, which are involved in photolithographic addition of oligonucleotide bases. P: Photolabile protecting group.

for genomic association studies, it is ideal for analytical assays and assay development (Pullat *et al.*, 2008).

17.5 INTERNAL QUALITY CONTROL FOR APEX GENOTYPING

The routine quality control of APEX genotyping process in genetic testing laboratories involves the following steps:

- DNA sample shipment check (is the tube still closed, labeling same as in documentation provided?),
- DNA quality check, i.e. concentration measurement, purity, and absence of degradation,
- Amplification check, i.e. presence of PCR products and amplification specificity,
- Purification of PCR products and yield of the PCR products after purification,
- DNA fragmentation check before APEX reactions (not required for APEX-2),
- Microarray quality check, i.e. presence of primer spots and self-extension of internal controls,
- APEX reaction check, i.e. visual detection of signals over the microarray,
- Quality control of diagnostic results, i.e. mutations detected are at least partially reanalyzed by automated DNA sequencing.

Additional internal quality control is also done by random blind analysis of previously tested samples.

17.6 APEX IN DIAGNOSTIC APPLICATIONS

APEX microarrays have been designed and successfully used for molecular diagnostics of various disorders, such as single gene Mendelian disorders, e.g. cystic fibrosis (Schrijver *et al.*, 2005), Wilson disease (Gojova *et al.*, 2008), and autosomal recessive Stargardt disease (Jaakson *et al.*, 2003). The next, more complex field of use of the microarrays is diagnostic and carrier testing of Mendelian disorders with more than one gene responsible for a similar phenotype. Although these are still "single gene disorders" by inheritance, a number of genes need to be tested for identifying the causal gene and its pathogenic variants. This has, so far, become the field of use, where APEX microarrays have shown most promising results, compared to the more conventional mutation analysis methods. The currently available microarrays include tests for the β-thalassemias (covering mutations in the *HBB* and *HBD* genes of human β-globin gene cluster) (Kurg *et al.*, 2000; Gemignani *et al.*, 2002), hereditary sensorineural hearing loss (Gardner *et al.*, 2006), Usher syndrome (Cremers *et al.*, 2007), Leber congenital amaurosis (Zernant *et al.*, 2005), retinitis pigmentosa (Koenekoop *et al.*, 2007), hereditary breast and ovarian cancer, and others. APEX microarrays have also been used for the testing of somatic *TP53* gene mutations in cancer (Tonisson *et al.*, 2002; Kringen *et al.*, 2005; Le Calvez *et al.*, 2005) and comprehensive carrier testing of various diseases with increased prevalence among people with Ashkenazi Jewish heritage (Schrijver *et al.*, 2007).

17.6.1 Tests for "Single Gene" Disorders

17.6.1.1 Cystic Fibrosis

Cystic fibrosis (CF) is the most common life-threatening autosomal recessively inherited disorder among Caucasians, but it also occurs in other ethnicities with variable frequency. CF is caused by mutations in the gene coding for the Cystic Fibrosis Transmembrane Conductance

TABLE 17.2 CF prevalence and common disease alleles in Europe and the Americas (Bobadilla *et al.*, 2002).

Geographical location	Incidence per birth	Frequency of F508del	Other common disease alleles, comments
Europe	1:2,000 in United Kingdom (Dodge *et al.*, 1997) to 1:25,000 in Finland (Kere *et al.*, 1994)	100% in Faroe Islands (Schwartz *et al.*, 1995), 70% in central, northern, western, and north-eastern Europe (Estivill *et al.*, 1997), 20% in Turkey (Yilmaz *et al.*, 1995).	Additional 5–10 mutations account for 10–15% of CF alleles: G542X in Mediterranean region (Casals *et al.*, 1993), G551D in Ireland and Brittany (Schwarz *et al.*, 1995; Estivill *et al.*, 1997), N1303K in Western and Mediterranean countries (Estivill *et al.*, 1997), W1282X in Ashkenazi Jews (Kerem *et al.*, 1995), 394delTT in nordic countries (Schwartz *et al.*, 1994; Teder *et al.*, 2000), 3905insT in Switzerland (Hergersberg *et al.*, 1997), R1162X in Northeast Italy (Bonizzato *et al.*, 1995), *CFTR*dele2,3(21 kb) in Slavic populations (Dork *et al.*, 2000).
North America	1:3,500 based on derivative populations (Kosorok *et al.*, 1996); 1:15,000 African Americans (Macek *et al.*, 1997; Hamosh *et al.*, 1998); 1:80,000 Native Americans (Mercier *et al.*, 1994; Yee *et al.*, 2000); in Canada it varies a lot: 1:914 Saguenay-Lac St. Jean region of Quebec (Rozen *et al.*, 1992; De Braekeleer *et al.*, 1998)	Most common allele in North America, except in Hutterites (Zielenski *et al.*, 1993). In Canada p.F508del accounts for 60–70% of CF alleles, except Hutterite population	About 10 *CFTR* mutations with >0.5% frequency, account for 79.7% in the USA (Bobadilla *et al.*, 2002). In Canada Hutterite population frequency of M1101K is 69.0% and frequency of F508del is 31.0% (Kristidis *et al.*, 1992; Rozen *et al.*, 1992; Zielenski *et al.*, 1993; De Braekeleer *et al.*, 1998).
Latin America	Very heterogeneous, ranges from 1:3,900 in Cuba (Collazo *et al.*, 1995) to 1:8,500 in Mexico (Orozco *et al.*, 2000)	Most common allele	The spectrum of *CFTR* mutations largely mimics Southern European countries, but each country has its own set of private, rare mutations (Bobadilla *et al.*, 2002).

Regulator (*CFTR*; MIM # 602421), a chloride channel expressed at the membrane of epithelial cells (Riordan *et al.*, 1989; Rommens *et al.*, 1989). More than 1,300 CF-causing CFTR mutations have been found according to the Cystic Fibrosis Mutation Database (CFMDB; http://www.genet.sickkids.on.ca/*CFTR*/app).

Classical CF is characterized by chronic obstructive pulmonary disease, exocrine pancreatic insufficiency, elevated sweat electrolytes, and infertility (Davis *et al.*, 1996). A wide variability in the clinical outcome is found among patients which have only one or few of CF characteristic features, e.g. congenital bilateral absence of the vas deferens (CBAVD) (Chillon *et al.*, 1995), disseminated bronchiectasis (Pignatti *et al.*, 1995), chronic pancreatitis (Pezzilli *et al.*, 2003), diffuse panbronchiolitis (Dequeker *et al.*, 2002). The prevalence of common CFTR gene disease alleles in Europe and the Americas is shown in Table 17.2 and can be also found in the World Health Organization Molecular Genetic Epidemiology of Cystic Fibrosis (http://www.cfww.org/WHO/WHO_HGN_CF_WG_04.02.pdf) and the FINDbase database (http://www.findbase.org; van Baal *et al.*, 2007).

For the routine genetic testing and screening of *CFTR*, resequencing of the full *CFTR* gene is still too expensive. In developed countries, CF carrier screening panels include only the most common mutations. A typical cut-off would be to screen for all mutations with a frequency of 0.5%. In most European countries 90–95% of the CF alleles may be identified with such cut-off. In North America, as a very heterogeneous region, with even a 0.1% cut-off, no more than 88% of CF alleles would be identified. The core mutation panel of 25 mutations, recommended by the American College of Medical Genetics (ACMG) and the American College of Obstetricians and Gynecologists (ACOG, 2004), detects 84% of CF alleles in the patients in the USA. With 50 mutations, detection rate would increase to 86.3%. Using ACMG/ACOG recommended mutation panel for African Americans and Hispanics, the estimated detection rates of CF carrier status diminish to 69% and 57%, respectively. Extending ACMG/ACOG core panel with 19 mutations, specific for African Americans and Hispanics, the detection rate would increase to 72.7% (Grody *et al.*, 2001; Bobadilla *et al.*, 2002; Watson *et al.*, 2004).

To improve the detection of CF alleles in ethnically mixed populations, a *CFTR* genotyping microarray for 254 mutations in *CFTR* gene has been developed. The pan-ethnic microarray (patent application WO2005006951)

includes a large number of mutations that are frequent in the Caucasian population as well as non-Caucasians (Schrijver *et al.*, 2005). The microarray tests for 25 mutations are listed in ACMG/ACOG core panel, as well as most prevalent mutations in non-Caucasian populations. The mutations are selected from 26 exons and 15 introns, including single-nucleotide substitutions, insertions, deletions, and repeats, such as the IVS8-5T/7T/9T. IVS8-5T allele causes reduced levels of normal *CFTR* mRNA and this DNA variant is found in 79% of CBAVD patients (Cuppens *et al.*, 1998; Radpour *et al.*, 2007). Detection of 5T/7T/9T alleles is technically challenging. This has been solved with a set of allele-specific primers, which enables reliable detection from both DNA strands. In addition to CF carrier screening, the developed pan-ethnic CF APEX assay can serve as the initial research tool to screen mutations in CFTR gene for a variety of diseases other than classic CF (e.g. CBAVD), followed by resequencing of the CFTR gene, if results are negative (Schrijver *et al.*, 2005).

17.6.2 Autosomal Recessive Stargardt Disease

Stargardt disease type 1 (STGD1, MIM # 248200) is a relatively common autosomal recessive macular dystrophy with estimated incidence 1 in 10,000 individuals. The disease is characterized by juvenile to young onset, central visual impairment, progressive bilateral atrophy of the macular retinal pigment epithelium (RPE) and neuroepithelium, with the frequent appearance of orange–yellow flecks distributed around the macula and/or the midretinal periphery following by characteristic progressive ''beaten bronze'' atrophy of the retinal pigment epithelium of the fovea. The causative gene, *ABCA4* or *ABCR*, is the family member of ATP binding cassette (ABC) transporters' ABCA subfamily. *ABCA4* has a restricted tissue distribution, the gene expression is observed in the retina by Northern blot analysis and specifically to the photoreceptor cell layer of the retina by *in situ* hydridization analysis (Anderson *et al.*, 1995; Allikmets *et al.*, 1997).

A detailed overview of genes and mutations involved in various retinal disorders, including type I Stargardt disease, may be found in the Retinal information network (http://www.sph.uth.tmc.edu/Retnet), Retina international locus specific database for retinal diseases (http://www.retina-international.com), and NEIBank database of genes and proteins expressed in the eye and visual system (http://neibank.nei.nih.gov/cgi-bin/eyeDiseaseGenes.cgi).

The initial version of *ABCA4* gene microarray for 400 mutations was 54–78% effective in detecting at least one of the two STGD1 patients mutations, depending on geographical cohorts (Jaakson *et al.*, 2003). The current ABCA4 microarray detects over 500 mutations, and is >98% effective in detecting patients' variations (Koenekoop *et al.*, 2007).

ABCA4 involvement in autosomal recessive cone-rod dystrophy (arCRD) and retinitis pigmentosa (arRP) has been studied in patients from the Netherlands and Germany. Genotyping of 90 arRP patients revealed sequence variations in five individuals, suggesting that *ABCA4* mutations are only a minor cause (2–5%) of arRP, not exceeding the contribution of most other arRP genes. In 18 of 54 patients with arCRD, 27 putatively pathologic *ABCA4* alleles were identified in total and if combined with earlier data, the results suggested that *ABCA4* mutations will be present in approximately 67% of arCRD cases (Klevering *et al.*, 2004). The list of other retinal diseases caused by or associated with *ABCA4* mutations includes *fundus flavimaculatus* and age-related macular degeneration. Both these can be screened with the *ABCA4* microarray as well (Koenekoop *et al.*, 2007).

17.6.3 Tests for Mendelian Disorders with Multiple Genes Causing Similar Phenotypes

17.6.3.1 Hereditary Sensorineural Hearing Loss

Hereditary sensorineural hearing loss (SNHL) is a very common disease that occurs in 1 out of every 1,000 to 2,000 live births. SNHL includes syndromic (Pendred syndrome, Waardenburg syndrome, Usher syndrome, etc.) and non-syndromic (without other associated findings) forms. About 80% of SNHL is non-syndromic and the pattern of inheritance can be autosomal recessive (DFNB, 75–80%), autosomal dominant (DFNA, 20–25%), X-linked (DFN, 1–1.5%), and mitochondrial (1–3%) (Van Camp *et al.*, 1997). DFNB1-linked deafness (MIM # 220290) accounts for half of congenital hearing loss cases (Zelante *et al.*, 1997). The genes responsible for DFNB1, GJB2 (MIM # 121011) and GJB6 (MIM # 604418), encode the gap junction proteins connexin 26 and connexin 30 (Gonzalez *et al.*, 2006).

Different mutations of *GJB2* gene account for up to 50% of autosomal recessive non-syndromic hearing loss in many world populations (Kenneson *et al.*, 2002), but also the dominantly inherited form of deafness (DFNA3; MIM # 601544). Over 100 mutations have been detected in the *GJB2* gene, mostly associated with early-onset deafness of different severity. Among these, c.35delG frameshift represents 28 to 85% of all *GJB2* mutations in Caucasians, depending on the population (Estivill *et al.*, 1998a; Gasparini *et al.*, 2000). The hearing loss in individuals with biallelic *GJB2* mutations ranges from mild to profound and is most commonly non-progressive (Denoyelle *et al.*, 1999; Murgia *et al.*, 1999). Numerous groups have studied *GJB2* genotype–phenotype correlations. For example, c.35delG homozygotes have significantly more severe hearing loss than do most of the c.35delG/non-c.35delG compound heterozygotes, although exceptions exist. Individuals with two non-c.35delG mutations have less severe hearing loss (Cryns *et al.*, 2004; Snoeckx *et al.*, 2005). A 309 kb deletion

(*GJB6*-D13S1830, previously referred as 342 kb deletion) truncating the *GJB6* gene was shown to be the accompanying mutation in up to 50% of deaf *GJB2* heterozygotes in different populations (del Castillo *et al.*, 2002, 2005).

Dominant or recessive non-syndromic forms of hearing loss (DFNA2; MIM # 600101) may be caused by mutations in the *GJB3* gene (MIM # 603324) encoding connexin 31 (Xia *et al.*, 1998; Liu *et al.*, 2000). Mutations in the *GJA1* gene encoding connexin 43 (MIM # 21014) cause a common form of recessively inherited non-syndromic hearing loss in African Americans (Liu *et al.*, 2001).

Inherited hearing loss may also be caused by mutations in the *SLC26* anion transport gene family. Mutations in the *SLC26A4* (pendrin) gene (MIM # 605646) are associated both with autosomal recessive non-syndromic recessive deafness (DFNB4; MIM # 600791) (Baldwin *et al.*, 1995; Li *et al.*, 1998) and with autosomal recessive Pendred syndrome (PDS; MIM # 274600). This is the most common syndromic form of deafness (up to 10% of all hereditary hearing loss) (Everett *et al.*, 1997; Reardon *et al.*, 1999). Mutations in *SLC26A5* (Prestin) gene (MIM # 604943) lead to non-syndromic hearing loss (Liu *et al.*, 2003).

Mutations in mitochondrial DNA (mtDNA) have been found to cause inherited syndromic, non-syndromic, and ototoxic hearing loss (Fischel-Ghodsian, 1999, 2003). Up to 3% of patients with SNHL have mutations in the mtDNA (Gardner *et al.*, 2006). A few mutations in the mitochondrial 12S rRNA gene and $tRNA^{Ser(UCN)}$ gene have been described. For example, homoplasmic 1555A > G mutation in the 12S rRNA gene is the main cause of aminoglycoside-induced and non-syndromic deafness in families of different ethnic backgrounds (Prezant *et al.*, 1993; Estivill *et al.*, 1998b). The 7511T > C mutation in the mitochondrial $tRNA^{Ser(UCN)}$ gene is known to cause non-syndromic deafness in families from different ethnic groups (Sue *et al.*, 1999; Chapiro *et al.*, 2002; Ishikawa *et al.*, 2002).

As described above, non-syndromic hearing loss is extraordinarily heterogeneous. Therefore, a comprehensive APEX microarray for the detection of 198 different causative mutations causing non-syndromic and syndromic SNHL has been developed (patent application US2007134691). The assay detects mutations in *GJB2*, *GJB3*, *GJB6*, *GJA*, *SLC26A4*, *SLC26A5*, and the mitochondrial, 12S rRNA and $tRNA^{Ser(UCN)}$, genes. The 198 mutations were selected from multiple sources including the Connexin Deafness home page (http://davinci.crg.es/deafness), mitochondrial mutation literature (Fischel-Ghodsian, 2003) and the Mitomap database (http://www.mitomap.org), the Hereditary Hearing Loss home page (http://dnalab-www.uia.ac.be/dnalab/hhh), and the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk). The APEX SNHL microarray is a comprehensive molecular diagnostic test for inherited hearing loss and thereby can facilitate the management and family counseling of hearing loss (Gardner *et al.*, 2006).

17.6.3.2 Usher Syndrome

Usher syndrome is inherited in an autosomal recessive manner and is responsible for over half of the cases involving deafness and blindness (Williams, 2008). It affects about 1 in 23,000 individuals in the USA, 1 in 29,000 in Scandinavia, and 1 in 25,000 in Germany (Boughman *et al.*, 1983; Otterstedde *et al.*, 2001). Three clinical subtypes can be distinguished. Patients with Usher syndrome type I (USH1; MIM # 276900) show severe to profound congenital hearing loss, arRP, and vestibular areflexia. Usher syndrome type II (USH2; MIM # 276905) is characterized by moderate to severe hearing loss, RP, and normal or variable vestibular function. Usher syndrome III (USH3; MIM # 605472) represents with progressive hearing loss, RP, and variable vestibular function. USH1 and USH2 are the most common forms of Usher syndrome, with USH3 contributing a large proportion of cases only in isolated areas, such as Finland (Pakarinen *et al.*, 1995) and Birmingham, UK (Hope *et al.*, 1997). RP, which is clinically similar to non-syndromic retinitis pigmentosa, develops in all types of Usher syndrome (Smith *et al.*, 1994). In some cases, mutations in causative genes give rise to non-syndromic deafness or autosomal recessive RP only (Williams, 2008). Eight of the nine USH genes are represented by 410 mutations and 18 SNPs on the current Usher syndrome APEX microarray (Table 17.3). At present, *DFNB31* gene, involved in USH2D (Ebermann *et al.*, 2007) and non-syndromic hearing loss, is not tested with the above-mentioned array.

For USH1, the microarray has revealed the highest percentage of patients with either one or two mutations, 51% for the European patients, 30% for the patients from the USA, and 46% when both groups are combined. For USH2, the percentages of patients with variants were 25%, 24%, and 25%, respectively. For USH3 or USHA (atypical Usher syndrome), the sensitivity of the microarray has not been reliably assessed. The Usher syndrome microarray is particularly useful for the analysis of patients with USHA, who have been shown to carry mutations in the *MYO7A*, *CDH23*, and *USH2A* genes (Cremers *et al.*, 2007).

17.6.3.3 Leber Congenital Amaurosis

Leber congenital amaurosis (LCA; MIM # 204000) is the earliest and most severe form of all the inherited retinal dystrophies responsible for congenital blindness (Perrault *et al.*, 1999). Its incidence is 2–3 per 100,000 births and it accounts for 10–18% of cases of congenital blindness among children in institutes for the blind and for 5% of all retinal dystrophies (Fazzi *et al.*, 2003). Hitherto, LCA is considered as an autosomal recessive genetically heterogeneous condition (Hanein *et al.*, 2004).

The currently recognized criteria for a diagnosis of LCA are onset of blindness or poor vision (frequently before 6 months of age), sluggish papillary reactions, roving eye

TABLE 17.3 Genes involved in Usher syndrome and the APEX microarray content (Ouyang *et al.*, 2004, 2005; Roux *et al.*, 2006; Cremers *et al.*, 2007; Ebermann *et al.*, 2007; Williams, 2008).

Usher syndrome type	Subtype	% of USH types among subtypes	Gene	Known disease alleles	Mutations and SNPs tested on microarray
USH1	USH1B	39–55	*MYO7A*	118	135
	USH1C	6–7	*USH1C*	8	9
	USH1D	19–35	*CDH23*	52	58
	USH1F	11–19	*PCDH15*	7	12
	USH1G	Rare (7)	*USH1G*	4	6
USH2	USH2A	80	*USH2A*	88	191
	USH2C	15	*GPR98*	6	6
	USH2D	5	*DFNB31*	3	0
USH3	USH3A	Unknown	*CLRN1*	9	11

movements/nystagmus, oculodigital signs (eye poking, eye rubbing, etc.), extinguished or severely reduced scotopic and photopic electroretinogram (ERG), absent or abnormal VEPs, and variable fundus (normal, marbled, etc.) (Fazzi *et al.*, 2003).

With the current LCA APEX microarray, 494 alleles are tested from 12 identified LCA genes and some other genes causing early-onset RP (such as *LRAT*, *MERTK*, *TULP1*, and *CRB1*) (Table 17.4).

Among 58 unrelated patients with LCA, who predominantly live in Belgium and the Netherlands, the microarray study has identified the molecular causes in one-third of the cases (Yzer *et al.*, 2006). In a different study involving 298 individuals, the screening efficiency of the LCA microarray was 20.3% in Canadian, 23.8% in the USA, and 20.7% in patients from the Netherlands (Zernant *et al.*, 2005).

17.6.4 Detection of *TP53* Gene Mutations in Cancer

TP53 tumor suppressor gene is mutated in various types of cancer and the knowledge about mutations has multiple implications for cancer detection and prognosis. Over 24,800 somatic and 390 germ-line *TP53* gene mutations have been reported in the IARC TP53 Mutation Database, http://www-p53.iarc.fr. An APEX microarray has been developed for the rapid and sensitive detection and identification of mutations in the *TP53* gene. The mutations are typically missense (73.4% of somatic and 77.9% of germ-line mutations (IARC TP53 Mutation Database) occur at many codons (mostly within exons 4 to 9)). The mutations frequently result in inactivation of p53 protein that is consequently unable to transactivate a wide panel of genes involved in cell cycle arrest, apoptosis, and DNA repair. Loss of these functions favors DNA replication and cell proliferation in adverse conditions where genomic and genetic alterations may accumulate. Moreover, some mutants have not only lost the wild-type function, but also acquired new, ''gain-of-function'' properties (Petitjean *et al.*, 2007). The extent of these effects (loss of suppressor as well as gain of promoter function) is dependent upon the consequence of the mutation on p53 protein structure and activity.

Screening for TP53 mutations gene has yet to become a routine in clinical or epidemiological practice. A strong limitation to routine analysis of *TP53* gene mutations resides in the fact that many tumors contain an excess of wild-type TP53 as compared to mutant, resulting from the presence of intact alleles in tumor as well as in non-cancerous cells (stroma, inflammatory cells, blood vessels). The first version of TP53 APEX microarray included oligonucleotide probes for exons 2 to 9 of the *TP53* gene (Tonisson *et al.*, 2002). In its current version, the microarray covers the entire *TP53* coding region (exons 2 to 11) and exon/intron borders. The detection limits of the microarray for a small fraction of mutated DNA were shown to be around 3 to 12%, depending on the mutation (Tonisson *et al.*, 2002; Le Calvez *et al.*, 2005). In three evaluation studies, the APEX microarray has shown a comparable sensitivity to temporal temperature gradient gel electrophoresis (TTGE; see also Chapter 6) and denaturing high-performance liquid chromatography (DHPLC) and an increased sensitivity compared to automated DNA sequencing (Tonisson *et al.*, 2002; Kringen *et al.*, 2005; Le Calvez *et al.*, 2005). Nevertheless, DNA sequencing is sometimes required for the correct identification

TABLE 17.4 Overview of genes involved in LCA and/or early-onset RP and their coverage by APEX microarray (den Hollander *et al.*, 2008).

Disease	Gene	MIM disease alleles	Mutations and SNPs tested on microarray	Mutation frequency in LCA (%)
LCA1, CRD6	*GUCY2D*	9	78	11.7
LCA2, RP20	*RPE65*	8	90	6.0
LCA3	*RDH12*	13	35	2.7
LCA4, autosomal dominant CRD	*AIPL1*	4	33	5.3
LCA5	*LCA5*	4	2	1.8
LCA6	*RPGRIP1*	8	49	4.2
LCA7, RP12	*CRB1*	0	88	9.9
LCA8	*CRX*	10	32	1.0
LCA10	*CEP290*	0	58	15.0
LCA11, autosomal dominant RP	*IMPDH1*	5	0	8.3
LCA12	*RD3*	1	0	0.1
RP14	*TULP1*	8	18	0.8
Retinal dystrophy, early-onset severe	*LRAT*	2	3	0.5
RP38, CORD	*MERTK*	4	8	0.6

of mutated sequence. In its future versions, this microarray could be updated with probes specific for most common small deletions and insertions of *TP53*.

17.6.5 Carrier Testing for Ashkenazi Jewish Diseases

A number of serious and lethal genetic disorders occur at higher frequencies among members of certain ethnic, racial, or demographic groups than in the general population. One example of a genetic isolate is Ashkenazi Jews, who have inhabited Northern and Eastern Europe since the 9th century and account for over 90% of Jews living in the USA today (Ostrer, 2001). In the Ashkenazi Jews population it has been estimated that one in four individuals is a carrier of one of several genetic conditions like Tay-Sachs disease, Canavan disease, Niemann-Pick disease, Gaucher's disease, familial dysautonomia, Bloom syndrome, Fanconi anemia, cystic fibrosis, and mucolipidosis type 4 (Victor Center for Jewish Genetic Diseases, Philadelphia, PA). The high frequency of more than 20 known recessive disease alleles in Ashkenazi Jewish populations compared to non-Ashkenazi Jews has raised multiple hypotheses ranged from a population ''bottleneck'' (Behar *et al.*, 2004) and founder effects (Risch *et al.*, 1995, 2003; Goldstein *et al.*, 1999; Slatkin, 2004) to the comparative contribution of heterozygote advantage (Cochran *et al.*, 2006). A web database with broad information coverage of Jewish disease heritage is available from Victor Center for Jewish Genetic Diseases, Philadelphia, PA: http://www.jewishvirtuallibrary.org/jsource/Health/genetics.html.

The American College of Obstetricians and Gynecologists currently recommends that couples of Ashkenazi Jewish ancestry should be offered carrier screening for four inherited disorders, Tay-Sachs disease, cystic fibrosis, Canavan disease, and familial dysautonomia, based on carrier frequencies of 1:40 or less (ACOG committee opinion, 2004). Additional severely disabling or fatal disorders are frequent in the Ashkenazi Jewish population, but the genetic testing for most of these is performed sequentially or for a small subset of conditions. There is a considerable variability in the diseases, prices, and labels used to describe the Ashkenazi Jews carrier testing panels. A total of 27 panels for three to nine diseases with very variable prizes were offered, as described by Leib and coworkers (2005). For comprehensive carrier screening and disease detection purposes we have developed an APEX-based genetic test for 22 inherited disorders with total of 77 disease-causing mutations prevalent in Ashkenazi Jews and other Jewish populations (patent application WO2008051604, Table 17.5). After publishing results from the initial APEX microarray version (Schrijver *et al.*, 2007), probes for the most common mutations related to cystic fibrosis, alpha 1-anti-trypsin deficiency, nemaline myopathy, Usher syndrome type 1F, familial hyperinsulinemia, familial hypercholesterolemia, and maple syrup urine disease type 3 were included to increase the coverage of the testing panel for Ashkenazi Jews and Jewish couples of mixed ancestry.

17.7 CONCLUSIONS

Among the variety of genotyping and mutation detection methods available, APEX microarrays have shown their good potential and usefulness for molecular diagnostics and carrier screening of multiple different genetic disorders. The ability of APEX microarrays to detect somatic mutations can be used in the molecular diagnostics of cancer.

New assays can be added with reduced effort by utilizing the flexibility of *in situ* synthesis platforms. APEX-2 opens up new frontiers for utilizing the microarrays in

TABLE 17.5 List of disorders and efficiency of carrier screening with the Ashkenazi Jews APEX microarray (adapted from Schrijver and coworkers (2007)).

Disease	Gene	No. of mutations	Carrier frequency	Mutation detection with carrier screening	MIM	References
Tay-Sachs disease	*HEXA*	8	1:28	>93%	272800	Mahuran *et al.* (1990); Triggs-Raine *et al.* (1992), GeneTests
Bloom syndrome	*BLM*	1	1:102	>97%	210900	Ellis and German (1996); Kaneko and Kondo (2004); Amor-Gueret (2006)
Canavan disease	*ASPA*	4	1:40	>99%	271900	Kaul *et al.* (1994), GeneTests
Factor XI deficiency	*F11*	3	1:23	>96%	264900	Asakai *et al.* (1991); Bolton-Maggs (1996)
Familial dysautonomia	*IKBKAP*	2	1:30	>99%	223900	Dong *et al.* (2002), GeneTests
Familial Mediterranean fever	*MEFV*	1	1:5	>90%	249100	Aksentijevich *et al.* (1999), GeneTests
Fanconi anemia type C	*FANCC*	4	1:80	>99%	227645	Verlander *et al.* (1994); Whitney *et al.* (1994), GeneTests
Gaucher disease	*GBA*	7	1:10	>90%	230800	Diaz *et al.* (2000); Charrow (2004); Strom *et al.* (2004), GeneTests
Glycogen storage disease Ia (von Gierke)	*G6PC*	1	1:71	~94%	232200	Parvari *et al.* (1997a); Ekstein *et al.* (2004)
Glycogen storage disease type 3a (N. African Jews)	*AGL*	1	1:35	All N. African Jews in Israel	232400	Parvari *et al.* (1997b)
Maple syrup urine disease type 1B	*BCKHDB*	3	1:80	~99%	248600	Edelmann *et al.* (2001); Henneke *et al.* (2003)
Mucolipidosis type 4	*MCOLN1*	2	1:100	>95%	252650	Bargal *et al.* (2000, 2001); Edelmann *et al.* (2002), GeneTests
Niemann-Pick type A	*SMPD1*	4	1:80	~95%	257200	Levran *et al.* (1992, 1993)
Non-syndromic sensorineural hearing loss	*GJB2*	2	1:25	>60%	220290	Morell *et al.* (1998); Lerer *et al.* (2000)
Torsion dystonia	*DYT1*	2	1:900	>95%	128100	Ozelius *et al.,* (1997); Charrow (2004)
Cystic fibrosis	*CFTR*	31	1:29	>97%	219700	Kerem *et al.* (1995); Quint *et al.* (2005)
Alpha 1-anti-trypsin deficiency	*SERPINA1*	1		~96%	107400	de Serres (2002); Blanco *et al.* (2006)
Nemaline myopathy	*NEB*	1	1: 108	~99%	256030	Anderson *et al.* (2004)
Usher syndrome type 1F	*PCDH15*	1	1:72	~64%	602083	Ben-Yosef *et al.* (2003)
Familial hyperinsulinemia	*ABCC8*	2	1:167	88%	256450	Nestorowicz *et al.* (1996)
Familial hypercholesterolemia	*LDLR*	1	1:333	~64% (Ashkenazi Jews with Lithuanian origin)	143890	Meiner *et al.* (1991); Durst *et al.* (2001); Risch *et al.* (2003)
Maple syrup urine disease type 3 (lipoamide dehydrogenase deficiency)	*DLD*	2	1:94	~99%	238331	Shaag *et al.* (1999)

custom genotyping applications. The future applications of the APEX microarrays will probably also include copy-number variation testing.

ACKNOWLEDGEMENTS

The authors and research have been supported by European Union through the European Regional Development Fund, in the frame of Centre of Excellence in Genomics, EU FP7 project "ECOGENE", targeted financing from Estonian government SF0180142, and Estonian Science Foundation grant 7473.

REFERENCES

ACOG committee opinion. No. 298, Aug. (2004). Prenatal and preconceptional carrier screening for genetic diseases in individuals of Eastern European Jewish descent (2004). *Obstet. Gynecol.* 104 425–428.

Aksentijevich, I., Torosyan, Y., Samuels, J., Centola, M., Pras, E., Chae, J.J., Oddoux, C., Wood, G., Azzaro, M.P., Palumbo, G., Giustolisi, R., Pras, M., Ostrer, H., and Kastner, D.L. (1999). Mutation and haplotype studies of familial Mediterranean fever reveal new ancestral relationships and evidence for a high carrier frequency with reduced penetrance in the Ashkenazi Jewish population. *Am. J. Hum. Genet.* 64, 949–962.

Albert, T.J., Norton, J., Ott, M., Richmond, T., Nuwaysir, K., Nuwaysir, E.F., Stengele, K.P., and Green, R.D. (2003). Light-directed 5′→3′ synthesis of complex oligonucleotide microarrays. *Nucleic Acids Res.* 31, e35.

Allikmets, R., Singh, N., Sun, H., Shroyer, N.F., Hutchinson, A., Chidambaram, A., Gerrard, B., Baird, L., Stauffer, D., Peiffer, A., Rattner, A., Smallwood, P., Li, Y., Anderson, K.L., Lewis, R.A., Nathans, J., Leppert, M., Dean, M., and Lupski, J.R. (1997). A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat. Genet.* 15, 236–246.

Amor-Gueret, M. (2006). Bloom syndrome, genomic instability and cancer: the SOS-like hypothesis. *Cancer Lett.* 236, 1–12.

Anderson, K.L., Baird, L., Lewis, R.A., Chinault, A.C., Otterud, B., Leppert, M., and Lupski, J.R. (1995). A YAC contig encompassing the recessive Stargardt disease gene (STGD) on chromosome 1p. *Am. J. Hum. Genet.* 57, 1351–1363.

Anderson, S.L., Ekstein, J., Donnelly, M.C., Keefe, E.M., Toto, N.R., LeVoci, L.A., and Rubin, B.Y. (2004). Nemaline myopathy in the Ashkenazi Jewish population is caused by a deletion in the nebulin gene. *Hum. Genet.* 115, 185–190.

Andreson, R., Reppo, E., Kaplinski, L., and Remm, M. (2006). GENOMEMASKER package for designing unique genomic PCR primers. *BMC Bioinformatics* 7, 172.

Asakai, R., Chung, D.W., Davie, E.W., and Seligsohn, U. (1991). Factor XI deficiency in Ashkenazi Jews in Israel. *N. Engl. J. Med.* 325, 153–158.

Baldwin, C.T., Weiss, S., Farrer, L.A., De Stefano, A.L., Adair, R., Franklyn, B., Kidd, K.K., Korostishevsky, M., and Bonne-Tamir, B. (1995). Linkage of congenital, recessive deafness (DFNB4) to chromosome 7q31 and evidence for genetic heterogeneity in the Middle Eastern Druze population. *Hum. Mol. Genet.* 4, 1637–1642.

Bargal, R., Avidan, N., Ben-Asher, E., Olender, Z., Zeigler, M., Frumkin, A., Raas-Rothschild, A., Glusman, G., Lancet, D., and Bach, G. (2000). Identification of the gene causing mucolipidosis type IV. *Nat. Genet.* 26, 118–123.

Bargal, R., Avidan, N., Olender, T., Ben Asher, E., Zeigler, M., Raas-Rothschild, A., Frumkin, A., Ben-Yoseph, O., Friedlender, Y., Lancet, D., and Bach, G. (2001). Mucolipidosis type IV: novel MCOLN1 mutations in Jewish and non-Jewish patients and the frequency of the disease in the Ashkenazi Jewish population. *Hum. Mutat.* 17, 397–402.

Bauer, A., Beckmann, B., Busold, C., Brandt, O., Kusnezow, W., Pullat, J., Aign, V., Fellenberg, K., Fleischer, R., Jacob, A., Frohme, M., and Hoheisel, J.D. (2003). Use of complex DNA and antibody microarrays as tools in functional analyses. *Comp. Funct. Genomics* 4, 520–524.

Behar, D.M., Hammer, M.F., Garrigan, D., Villems, R., Bonne-Tamir, B., Richards, M., Gurwitz, D., Rosengarten, D., Kaplan, M., Della Pergola, S., Quintana-Murci, L., and Skorecki, K. (2004). MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. *Eur. J. Hum. Genet.* 12, 355–364.

Beier, M., and Hoheisel, J.D. (2000). Production by quantitative photolithographic synthesis of individually quality checked DNA microarrays. *Nucleic Acids Res.* 28, E11.

Ben-Yosef, T., Ness, S.L., Madeo, A.C., Bar-Lev, A., Wolfman, J.H., Ahmed, Z.M., Desnick, R.J., Willner, J.P., Avraham, K.B., Ostrer, H., Oddoux, C., Griffith, AJ., and Friedman, T.B. (2003). A mutation of PCDH15 among Ashkenazi Jews with the type 1 Usher syndrome. *N. Engl. J. Med.* 348, 1664–1670.

Blanco, I., de Serres, F.J., Fernandez-Bustillo, E., Lara, B., and Miravitlles, M. (2006). Estimated numbers and prevalence of PI*S and PI*Z alleles of alpha1-antitrypsin deficiency in European countries. *Eur. Respir. J.* 27, 77–84.

Bobadilla, J.L., Macek, M., Jr, Fine, J.P., and Farrell, P.M. (2002). Cystic fibrosis: a worldwide analysis of CFTR mutations – correlation with incidence data and application to screening. *Hum. Mutat.* 19, 575–606.

Bolton-Maggs, P.H. (1996). Factor XI deficiency. *Baillieres Clin. Haematol.* 9, 355–368.

Bonizzato, A., Bisceglia, L., Marigo, C., Nicolis, E., Bombieri, C., Castellani, C., Borgo, G., Zelante, L., Mastella, G., Cabrini, G., Bisceglia, L., Zelante, L., Gasparini, P., Marigo, C., Bombieri, C., and Pignatti, P.F. (1995). Analysis of the complete coding region of the CFTR gene in a cohort of CF patients from north-eastern Italy: identification of 90% of the mutations. *Hum. Genet.* 95, 397–402.

Boughman, J.A., Vernon, M., and Shaver, K.A. (1983). Usher syndrome: definition and estimate of prevalence from two high-risk populations. *J. Chronic. Dis.* 36, 595–603.

Breslauer, K.J., Frank, R., Blocker, H., and Marky, L.A. (1986). Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA* 83, 3746–3750.

Casals, T., Nunes, V., Palacio, A., Gimenez, J., Gaona, A., Ibanez, N., Morral, N., and Estivill, X. (1993). Cystic fibrosis in Spain: high frequency of mutation G542X in the Mediterranean coastal area. *Hum. Genet.* 91, 66–70.

Chapiro, E., Feldmann, D., Denoyelle, F., Sternberg, D., Jardel, C., Eliot, M.M., Bouccara, D., Weil, D., Garabedian, E.N., Couderc, R., Petit, C., and Marlin, S. (2002). Two large French pedigrees with non syndromic sensorineural deafness and the mitochondrial DNA T7511C mutation: evidence for a modulatory factor. *Eur. J. Hum. Genet.* 10, 851–856.

Charrow, J. (2004). Ashkenazi Jewish genetic disorders. *Fam. Cancer* 3, 201–206.

Chen, X., Levine, L., and Kwok, P.Y. (1999). Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Res.* 9, 492–498.

Chillon, M., Casals, T., Mercier, B., Bassas, L., Lissens, W., Silber, S., Romey, M.C., Ruiz-Romero, J., Verlingue, C., Claustres, M., Nunes, V., Férec, C., and Estivill, X. (1995). Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. *N. Engl. J. Med.* 332, 1475–1480.

Cochran, G., Hardy, J., and Harpending, H. (2006). Natural history of Ashkenazi intelligence. *J. Biosoc. Sci.* 38, 659–693.

Collazo, T., Magarino, C., Chavez, R., Suardiaz, B., Gispert, S., Gomez, M., Rojo, M., and Heredero, L. (1995). Frequency of delta-F508 mutation and XV2C/KM19 haplotypes in Cuban cystic fibrosis families. *Hum. Hered.* 45, 55–57.

Cremers, F.P., Kimberling, W.J., Kulm, M., de Brouwer, A.P., van Wijk, E., te Brinke, H., Cremers, C.W., Hoefsloot, L.H., Banfi, S., Simonelli, F., Fleischhauer, J.C., Berger, W., Kelley, P.M., Haralambous, E., Bitner-Glindzicz, M., Webster, A.R., Saihan, Z., De Baere, E., Leroy, B.P., Silvestri, G., McKay, G.J., Koenekoop, R.K., Millan, J.M., Rosenberg, T., Joensuu, T., Sankila, E.M., Weil, D., Weston, M.D., Wissinger, B., and Kremer, H. (2007). Development of a genotyping microarray for Usher syndrome. *J. Med. Genet.* 44, 153–160.

Cryns, K., Orzan, E., Murgia, A., Huygen, P.L., Moreno, F., del Castillo, I., Chamberlin, G.P., Azaiez, H., Prasad, S., Cucci, R.A., Leonardi, E., Snoeckx, R.L., Govaerts, P.J., Van de Heyning, P.H., Van de Heyning, C.M., Smith, R.J., and Van Camp, G. (2004). A genotype-phenotype correlation for GJB2 (connexin 26) deafness. *J. Med. Genet.* 41, 147–154.

Cuppens, H., Lin, W., Jaspers, M., Costes, B., Teng, H., Vankeerberghen, A., Jorissen, M., Droogmans, G., Reynaert, I., Goossens, M., Nilius, B., and Cassiman, J.J. (1998). Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes. The polymorphic (Tg)m locus explains the partial penetrance of the T5 polymorphism as a disease mutation. *J. Clin. Invest.* 101, 487–496.

Davis, P.B., Drumm, M., and Konstan, M.W. (1996). Cystic fibrosis. *Am. J. Respir. Crit. Care. Med.* 154, 1229–1256.

De Braekeleer, M., Mari, C., Verlingue, C., Allard, C., Leblanc, J.P., Simard, F., Aubin, G., and Ferec, C. (1998). Complete identification of cystic fibrosis transmembrane conductance regulator mutations in the CF population of Saguenay Lac-Saint-Jean (Quebec, Canada). *Clin. Genet.* 53, 44–46.

de Serres, F.J. (2002). Worldwide racial and ethnic distribution of alpha1-antitrypsin deficiency: summary of an analysis of published genetic epidemiologic surveys. *Chest.* 122, 1818–1829.

del Castillo, F.J., Rodriguez-Ballesteros, M., Alvarez, A., Hutchin, T., Leonardi, E., de Oliveira, C.A., Azaiez, H., Brownstein, Z., Avenarius, M.R., Marlin, S., Pandya, A., Shahin, H., Siemering, K. R., Weil, D., Wuyts, W., Aguirre, L.A., Martin, Y., Moreno-Pelayo, M.A., Villamar, M., Avraham, K.B., Dahl, H.H., Kanaan, M., Nance, W.E., Petit, C., Smith, R.J., Van Camp, G., Sartorato, E.L., Murgia, A., Moreno, F., and del Castillo, I. (2005). A novel deletion involving the connexin-30 gene, del(GJB6-d13s1854), found in trans with mutations in the GJB2 gene (connexin-26) in subjects with DFNB1 non-syndromic hearing impairment. *J. Med. Genet.* 42, 588–594.

del Castillo, I., Villamar, M., Moreno-Pelayo, M.A., del Castillo, F.J., Alvarez, A., Telleria, D., Menendez, I., and Moreno, F. (2002). A deletion involving the connexin 30 gene in nonsyndromic hearing impairment. *N. Engl. J. Med.* 346, 243–249.

den Hollander, A.I., Roepman, R., Koenekoop, R.K., and Cremers, F. P. (2008). Leber congenital amaurosis: genes, proteins and disease mechanisms. *Prog. Retin. Eye Res.* 27, 391–419.

Denoyelle, F., Marlin, S., Weil, D., Moatti, L., Chauvin, P., Garabedian, E.N., and Petit, C. (1999). Clinical features of the prevalent form of childhood deafness, DFNB1, due to a connexin-26 gene defect: implications for genetic counselling. *Lancet* 353, 1298–1303.

Dequeker, E., Accurso, F., Cabeza, S., Cassiman, J.-J., Corey, M., Davidson, A., Döring, G., Heidet, L., Heijerman, H., Kotsimbos, T., Mastella, G., Morrison, C., Pignatti, P.F., Strandvik, B., Tsui, L.-C., and Dodge, J. (2002). Classification of cystic fibrosis and related disorders. *J. Cyst. Fibros.* 1, 5–8.

Diaz, G.A., Gelb, B.D., Risch, N., Nygaard, T.G., Frisch, A., Cohen, I. J., Miranda, C.S., Amaral, O., Maire, I., Poenaru, L., Caillaud, C., Weizberg, M., Mistry, P., and Desnick, R.J. (2000). Gaucher disease: the origins of the Ashkenazi Jewish N370S and 84GG acid beta-glucosidase mutations. *Am. J. Hum. Genet.* 66, 1821–1832.

Dodge, J.A., Morison, S., Lewis, P.A., Coles, E.C., Geddes, D., Russell, G., Littlewood, J.M., and Scott, M.T. (1997). Incidence, population, and survival of cystic fibrosis in the UK, 1968–95. UK Cystic Fibrosis Survey Management Committee. *Arch. Dis. Child.* 77, 493–496.

Dong, J., Edelmann, L., Bajwa, A.M., Kornreich, R., and Desnick, R.J. (2002). Familial dysautonomia: detection of the IKBKAP IVS20(+6T>C) and R696P mutations and frequencies among Ashkenazi Jews. *Am. J. Hum. Genet.* 110, 253–257.

Dork, T., Macek, M., Jr, Mekus, F., Tummler, B., Tzountzouris, J., Casals, T., Krebsova, A., Koudova, M., Sakmaryova, I., Macek, M., Sr, Vavrova, V., Zemkova, D., Ginter, E., Petrova, N.V., Ivaschenko, T., Baranov, V., Witt, M., Pogorzelski, A., Bal, J., Zekanowsky, C., Wagner, K., Stuhrmann, M., Bauer, I., Seydewitz, H.H., Neumann, T., and Jakubiczka, S. (2000). Characterization of a novel 21-kb deletion, CFTRdele2,3(21 kb), in the CFTR gene: a cystic fibrosis mutation of Slavic origin common in Central and East Europe. *Hum. Genet.* 106, 259–268.

Durst, R., Colombo, R., Shpitzen, S., Avi, L.B., Friedlander, Y., Wexler, R., Raal, F.J., Marais, D.A., Defesche, J.C., Mandelshtam, M.Y., Kotze, M.J., Leitersdorf, E., and Meiner, V. (2001). Recent origin and spread of a common Lithuanian mutation. G197del LDLR, causing familial hypercholesterolemia: positive selection is not always necessary to account for disease incidence among Ashkenazi Jews. *Am. J. Hum. Genet.* 68, 1172–1188.

Ebermann, I., Scholl, H.P., Charbel Issa, P., Becirovic, E., Lamprecht, J., Jurklies, B., Millan, J.M., Aller, E., Mitter, D., and Bolz, H. (2007). A novel gene for Usher syndrome type 2: mutations in the long isoform of whirlin are associated with retinitis pigmentosa and sensorineural hearing loss. *Hum. Genet.* 121, 203–211.

Edelmann, L., Dong, J., Desnick, R.J., and Kornreich, R. (2002). Carrier screening for mucolipidosis type IV in the American Ashkenazi Jewish population. *Am. J. Hum. Genet.* 70, 1023–1027.

Edelmann, L., Wasserstein, M.P., Kornreich, R., Sansaricq, C., Snyderman, S.E., and Diaz, G.A. (2001). Maple syrup urine disease: identification and carrier-frequency determination of a novel founder mutation in the Ashkenazi Jewish population. *Am. J. Hum. Genet.* 69, 863–868.

Ekstein, J., Rubin, B.Y., Anderson, S.L., Weinstein, D.A., Bach, G., Abeliovich, D., Webb, M., and Risch, N. (2004). Mutation frequencies for glycogen storage disease Ia in the Ashkenazi Jewish population. *Am. J. Med Genet. A* 129A, 162–164.

Ellis, N.A., and German, J. (1996). Molecular genetics of Bloom's syndrome. *Hum. Mol. Genet.* 5, 1457–1463.

Estivill, X., Bancells, C., and Ramos, C. (1997). Geographic distribution and regional origin of 272 cystic fibrosis mutations in European populations. The Biomed CF Mutation Analysis Consortium. *Hum. Mutat.* 10, 135–154.

Estivill, X., Fortina, P., Surrey, S., Rabionet, R., Melchionda, S., D'Agruma, L., Mansfield, E., Rappaport, E., Govea, N., Mila, M., Zelante, L., and Gasparini, P. (1998a). Connexin-26 mutations in sporadic and inherited sensorineural deafness. *Lancet* 351, 394–398.

Estivill, X., Govea, N., Barcelo, E., Badenas, C., Romero, E., Moral, L., Scozzri, R., D'Urbano, L., Zeviani, M., and Torroni, A. (1998b). Familial progressive sensorineural deafness is mainly due to the mtDNA A1555G mutation and is enhanced by treatment of aminoglycosides. *Am. J. Hum. Genet.* 62, 27–35.

Everett, L.A., Glaser, B., Beck, J.C., Idol, J.R., Buchs, A., Heyman, M., Adawi, F., Hazani, E., Nassir, E., Baxevanis, A.D., Sheffield, V. C., and Green, E.D. (1997). Pendred syndrome is caused by mutations in a putative sulphate transporter gene (PDS). *Nat. Genet.* 17, 411–422.

Fazzi, E., Signorini, S.G., Scelsa, B., Bova, S.M., and Lanzi, G. (2003). Leber's congenital amaurosis: an update. *Eur. J. Paediatr. Neurol.* 7, 13–22.

Fischel-Ghodsian, N. (1999). Mitochondrial deafness mutations reviewed. *Hum. Mutat.* 13, 261–270.

Fischel-Ghodsian, N. (2003). Mitochondrial deafness. *Ear. Hear.* 24, 303–313.

Fodor, S.P., Rava, R.P., Huang, X.C., Pease, A.C., Holmes, C.P., and Adams, C.L. (1993). Multiplexed biochemical assays with biological chips. *Nature* 364, 555–556.

Gardner, P., Oitmaa, E., Messner, A., Hoefsloot, L., Metspalu, A., and Schrijver, I. (2006). Simultaneous multigene mutation detection in patients with sensorineural hearing loss through a novel diagnostic microarray: a new approach for newborn screening follow-up. *Pediatrics* 118, 985–994.

Gasparini, P., Rabionet, R., Barbujani, G., Melchionda, S., Petersen, M., Brondum-Nielsen, K., Metspalu, A., Oitmaa, E., Pisano, M., Fortina, P., Zelante, L., and Estivill, X. (2000). High carrier frequency of the 35delG deafness mutation in European populations. Genetic Analysis Consortium of GJB2 35delG. *Eur. J. Hum. Genet.* 8, 19–23.

Gemignani, F., Perra, C., Landi, S., Canzian, F., Kurg, A., Tonisson, N., Galanello, R., Cao, A., Metspalu, A., and Romeo, G. (2002). Reliable detection of beta-thalassemia and G6PD mutations by a DNA microarray. *Clin. Chem.* 48, 2051–2054.

Gojova, L., Jansova, E., Kulm, M., Pouchla, S., and Kozak, L. (2008). Genotyping microarray as a novel approach for the detection of ATP7B gene mutations in patients with Wilson disease. *Clin. Genet.* 73, 441–452.

Goldstein, D.B., Reich, D.E., Bradman, N., Usher, S., Seligsohn, U., and Peretz, H. (1999). Age estimates of two common mutations causing factor XI deficiency: recent genetic drift is not necessary for elevated disease incidence among Ashkenazi Jews. *Am. J. Hum. Genet.* 64, 1071–1075.

Gonzalez, J.R., Wang, W., Ballana, E., and Estivill, X. (2006). A recessive Mendelian model to predict carrier probabilities of DFNB1 for non-syndromic deafness. *Hum. Mutat.* 27, 1135–1142.

Grody, W.W., Cutting, G.R., Klinger, K.W., Richards, C.S., Watson, M. S., and Desnick, R.J. (2001). Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. *Genet. Med.* 3, 149–154.

Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R., and Smith, L.M. (1994). Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res.* 22, 5456–5465.

Hanein, S., Perrault, I., Gerber, S., Tanguy, G., Barbet, F., Ducroq, D., Calvas, P., Dollfus, H., Hamel, C., Lopponen, T., Munier, F., Santos, L., Shalev, S., Zafeiriou, D., Dufier, J.L., Munnich, A., Rozet, J.M. and Kaplan, J. (2004). Leber congenital amaurosis: comprehensive survey of the genetic heterogeneity, refinement of the clinical definition, and genotype-phenotype correlations as a strategy for molecular diagnosis. *Hum. Mutat.* 23, 306–317.

Hamosh, A., FitzSimmons, S.C., Macek, M., Jr, Knowles, M.R., Rosenstein, B.J., and Cutting, G.R. (1998). Comparison of the clinical manifestations of cystic fibrosis in black and white patients. *J. Pediatr.* 132, 255–259.

Hardenbol, P., Baner, J., Jain, M., Nilsson, M., Namsaraev, E.A., Karlin-Neumann, G.A., Fakhrai-Rad, H., Ronaghi, M., Willis, T.D., Landegren, U., and Davis, R.W. (2003). Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat. Biotechnol.* 21, 673–678.

Henneke, M., Flaschker, N., Helbling, C., Muller, M., Schadewaldt, P., Gartner, J., and Wendel, U. (2003). Identification of twelve novel mutations in patients with classic and variant forms of maple syrup urine disease. *Hum. Mutat.* 22, 417.

Hergersberg, M., Balakrishnan, J., Bettecken, T., Chevalier-Porst, F., Bragger, C., Burger, R., Einschenk, I., Liechti-Gallati, S., Morris, M., Schorderet, D., Thonney, F., Moser, H., and Malik, N. (1997). A new mutation, 3905insT, accounts for 4.8% of 1173 CF chromosomes in Switzerland and causes a severe phenotype. *Hum. Genet.* 100, 220–223.

Hope, C.I., Bundey, S., Proops, D., and Fielder, A.R. (1997). Usher syndrome in the city of Birmingham – prevalence and clinical classification. *Br. J. Ophthalmol.* 81, 46–53.

Hua, J., Craig, D.W., Brun, M., Webster, J., Zismann, V., Tembe, W., Joshipura, K., Huentelman, M.J., Dougherty, E.R., and Stephan, D. A. (2007). SNiPer-HD: improved genotype calling accuracy by an expectation-maximization algorithm for high-density SNP arrays. *Bioinformatics* 23, 57–63.

Ishikawa, K., Tamagawa, Y., Takahashi, K., Kimura, H., Kusakari, J., Hara, A., and Ichimura, K. (2002). Nonsyndromic hearing loss caused by a mitochondrial T7511C mutation. *Laryngoscope* 112, 1494–1499.

Jaakson, K., Zernant, J., Kulm, M., Hutchinson, A., Tonisson, N., Glavac, D., Ravnik-Glavac, M., Hawlina, M., Meltzer, M.R., Caruso, R. C., Testa, F., Maugeri, A., Hoyng, C.B., Gouras, P., Simonelli, F., Lewis, R.A., Lupski, J.R., Cremers, F.P., and Allikmets, R. (2003). Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. *Hum. Mutat.* 22, 395–403.

Kaneko, H., and Kondo, N. (2004). Clinical features of Bloom syndrome and function of the causative gene, BLM helicase. *Expert Rev. Mol. Diagn.* 4, 393–401.

Kaul, R., Gao, G.P., Aloya, M., Balamurugan, K., Petrosky, A., Michals, K., and Matalon, R. (1994). Canavan disease: mutations among Jewish and non-Jewish patients. *Am. J. Hum. Genet.* 55, 34–41.

Kenneson, A., Van Naarden Braun, K., and Boyle, C. (2002). GJB2 (connexin 26) variants and nonsyndromic sensorineural hearing loss: a HuGE review. *Genet. Med.* 4, 258–274.

Kere, J., Estivill, X., Chillon, M., Morral, N., Nunes, V., Norio, R., Savilahti, E., and de la Chapelle, A. (1994). Cystic fibrosis in a low-incidence population: two major mutations in Finland. *Hum. Genet.* 93, 162–166.

Kerem, E., Kalman, Y.M., Yahav, Y., Shoshani, T., Abeliovich, D., Szeinberg, A., Rivlin, J., Blau, H., Tal, A., Ben-Tur, L., Springer, C., Augarten, A., Godfrey, S., Lerer, I., Branski, D., Friedman, M., and Kerem, B. (1995). Highly variable incidence of cystic fibrosis and different mutation distribution among different Jewish ethnic groups in Israel. *Hum. Genet.* 96, 193–197.

Klevering, B.J., Yzer, S., Rohrschneider, K., Zonneveld, M., Allikmets, R., van den Born, L.I., Maugeri, A., Hoyng, C.B., and Cremers, F.P. (2004). Microarray-based mutation analysis of the ABCA4 (ABCR) gene in autosomal recessive cone-rod dystrophy and retinitis pigmentosa. *Eur. J. Hum. Genet.* 12, 1024–1032.

Klito, N.G., Tan, Q., Nyegaard, M., Brusgaard, K., Thomassen, M., Skouboe, C., Dahlgaard, J., and Kruse, T.A. (2007). Arrayed primer extension in the "array of arrays" format: a rational approach for microarray-based SNP genotyping. *Genet. Test.* 11, 160–166.

Koenekoop, R.K., Lopez, I., den Hollander, A.I., Allikmets, R., and Cremers, F.P. (2007). Genetic testing for retinal dystrophies and dysfunctions: benefits, dilemmas and solutions. *Clin. Experiment. Ophthalmol.* 35, 473–485.

Kosorok, M.R., Wei, W.H., and Farrell, P.M. (1996). The incidence of cystic fibrosis. *Stat. Med.* 15, 449–462.

Kringen, P., Bergamaschi, A., Due, E.U., Wang, Y., Tagliabue, E., Nesland, J.M., Ahman, A., Tõnisson, N., and Borresen-Dale, A.L. (2005). Evaluation of arrayed primer extension for TP53 mutation detection in breast and ovarian carcinomas. *Biotechniques* 39, 755–761.

Kristidis, P., Bozon, D., Corey, M., Markiewicz, D., Rommens, J., Tsui, L. C., and Durie, P. (1992). Genetic determination of exocrine pancreatic function in cystic fibrosis. *Am. J. Hum. Genet.* 50, 1178–1184.

Krjutskov, K., Andreson, R., Magi, R., Nikopensius, T., Khrunin, A., Mihailov, E., Tammekivi, V., Sork, H., Remm, M., and Metspalu, A. (2008). Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays. *Nucleic Acids Res.* 36, e75.

Kurg, A., Tonisson, N., Georgiou, I., Shumaker, J., Tollett, J., and Metspalu, A. (2000). Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. *Genet. Test.* 4, 1–7.

Lahermo, P., Liljedahl, U., Alnaes, G., Axelsson, T., Brookes, A.J., Ellonen, P., Groop, P.H., Hallden, C., Holmberg, D., Holmberg, K., Keinanen, M., Kepp, K., Kere, J., Kiviluoma, P., Kristensen, V., Lindgren, C., Odeberg, J., Osterman, P., Parkkonen, M., Saarela, J., Sterner, M., Stromqvist, L., Talas, U., Wessman, M., Palotie, A., and Syvanen, A.C. (2006). A quality assessment survey of SNP genotyping laboratories. *Hum. Mutat.* 27, 711–714.

Le Calvez, F., Ahman, A., Tonisson, N., Lambert, J., Temam, S., Brennan, P., Zaridze, D.G., Metspalu, A., and Hainaut, P. (2005). Arrayed primer extension resequencing of mutations in the TP53 tumor suppressor gene: comparison with denaturing HPLC and direct sequencing. *Clin. Chem.* 51, 1284–1287.

Leib, J.R., Gollust, S.E., Hull, S.C., and Wilfond, B.S. (2005). Carrier screening panels for Ashkenazi Jews: is more better?. *Genet. Med.* 7, 185–190.

Lerer, I., Sagi, M., Malamud, E., Levi, H., Raas-Rothschild, A., and Abeliovich, D. (2000). Contribution of connexin 26 mutations to nonsyndromic deafness in Ashkenazi patients and the variable phenotypic effect of the mutation 167delT. *Am. J. Med. Genet.* 95, 53–56.

Levran, O., Desnick, R.J., and Schuchman, E.H. (1992). Identification and expression of a common missense mutation (L302P) in the acid sphingomyelinase gene of Ashkenazi Jewish type A Niemann-Pick disease patients. *Blood* 80, 2081–2087.

Levran, O., Desnick, R.J., and Schuchman, E.H. (1993). Type A Niemann-Pick disease: a frameshift mutation in the acid sphingomyelinase gene (fsP330) occurs in Ashkenazi Jewish patients. *Hum. Mutat.* 2, 317–319.

Li, X.C., Everett, L.A., Lalwani, A.K., Desmukh, D., Friedman, T.B., Green, E.D., and Wilcox, E.R. (1998). A mutation in PDS causes nonsyndromic recessive deafness. *Nat. Genet.* 18, 215–217.

Lindroos, K., Liljedahl, U., Raitio, M., and Syvanen, A.C. (2001). Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries. *Nucleic Acids Res.* 29, E69.

Lindroos, K., Sigurdsson, S., Johansson, K., Ronnblom, L., and Syvanen, A.C. (2002). Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. *Nucleic Acids Res.* 30, e70.

Liu, X.Z., Ouyang, X.M., Xia, X.J., Zheng, J., Pandya, A., Li, F., Du, L. L., Welch, K.O., Petit, C., Smith, R.J., Webb, B.T., Yan, D., Arnos, K. S., Corey, D., Dallos, P., Nance, W.E., and Chen, Z.Y. (2003). Prestin, a cochlear motor protein, is defective in non-syndromic hearing loss. *Hum. Mol. Genet.* 12, 1155–1162.

Liu, X.Z., Xia, X.J., Adams, J., Chen, Z.Y., Welch, K.O., Tekin, M., Ouyang, X.M., Kristiansen, A., Pandya, A., Balkany, T., Arnos, K.S., and Nance, W.E. (2001). Mutations in GJA1 (connexin 43) are associated with non-syndromic autosomal recessive deafness. *Hum. Mol. Genet.* 10, 2945–2951.

Liu, X.Z., Xia, X.J., Xu, L.R., Pandya, A., Liang, C.Y., Blanton, S.H., Brown, S.D., Steel, K.P., and Nance, W.E. (2000). Mutations in connexin31 underlie recessive as well as dominant non-syndromic hearing loss. *Hum. Mol. Genet.* 9, 63–67.

Macek, M., Jr, Mackova, A., Hamosh, A., Hilman, B.C., Selden, R.F., Lucotte, G., Friedman, K.J., Knowles, M.R., Rosenstein, B.J., and Cutting, G.R. (1997). Identification of common cystic fibrosis mutations in African-Americans with cystic fibrosis increases the detection rate to 75%. *Am. J. Hum. Genet.* 60, 1122–1127.

Mahuran, D.J., Triggs-Raine, B.L., Feigenbaum, A.J., and Gravel, R.A. (1990). The molecular basis of Tay-Sachs disease: mutation identification and diagnosis. *Clin Biochem.* 23, 409–415.

Meiner, V., Landsberger, D., Berkman, N., Reshef, A., Segal, P., Seftel, H.C., van der Westhuyzen, D.R., Jeenah, M.S., Coetzee, G.A., and Leitersdorf, E. (1991). A common Lithuanian mutation causing familial hypercholesterolemia in Ashkenazi Jews. *Am. J. Hum. Genet.* 49, 443–449.

Mercier, B., Raguenes, O., Estivill, X., Morral, N., Kaplan, G.C., McClure, M., Grebe, T.A., Kessler, D., Pignatti, P.F., Marigo, C., Bombieri, C., Audrézet, M.P., Verlingue, C., and Férec, C. (1994). Complete detection of mutations in cystic fibrosis patients of Native American origin. *Hum. Genet.* 94, 629–632.

Morell, R.J., Kim, H.J., Hood, L.J., Goforth, L., Friderici, K., Fisher, R., Van Camp, G., Berlin, C.I., Oddoux, C., Ostrer, H., Keats, B., and Friedman, T.B. (1998). Mutations in the connexin 26 gene (GJB2) among Ashkenazi Jews with nonsyndromic recessive deafness. *N. Engl. J. Med.* 339, 1500–1505.

Murgia, A., Orzan, E., Polli, R., Martella, M., Vinanzi, C., Leonardi, E., Arslan, E., and Zacchello, F. (1999). Cx26 deafness: mutation analysis and clinical variability. *J. Med. Genet.* 36, 829–832.

Nestorowicz, A., Wilson, B.A., Schoor, K.P., Inoue, H., Glaser, B., Landau, H., Stanley, C.A., Thornton, P.S., Clement, J.P., Bryan, J., Aguilar-Bryan, L., and Permutt, M.A. (1996). Mutations in the sulonylurea receptor gene are associated with familial hyperinsulinism in Ashkenazi Jews. *Hum. Mol. Genet.* 5, 1813–1822.

Orozco, L., Velazquez, R., Zielenski, J., Tsui, L.C., Chavez, M., Lezana, J. L., Saldana, Y., Hernandez, E., and Carnevale, A. (2000). Spectrum of CFTR mutations in Mexican cystic fibrosis patients: identification of five novel mutations (W1098C, 846delT, P750L, 4160insGGGG and 297-1G→A). *Hum. Genet.* 106, 360–365.

Ostrer, H. (2001). A genetic profile of contemporary Jewish populations. *Nat. Rev. Genet.* 2, 891–898.

Otterstedde, C.R., Spandau, U., Blankenagel, A., Kimberling, W.J., and Reisser, C. (2001). A new clinical classification for Usher's syndrome

based on a new subtype of Usher's syndrome type I. *Laryngoscope* 111, 84–86.

Ozelius, L.J., Hewett, J.W., Page, C.E., Bressman, S.B., Kramer, P. L., Shalish, C., de Leon, D., Brin, M.F., Raymond, D., Corey, D.P., Fahn, S., Risch, N.J., Buckler, A.J., Gusella, J.F., and Breakefield, X. O. (1997). The early-onset torsion dystonia gene (DYT1) encodes an ATP-binding protein. *Nat. Genet.* 17, 40–48.

Ouyang, X.M., Hejtmancik, J.F., Jacobson, S.G., Li, A.R., Du, L.L., Angeli, S., Kaiser, M., Balkany, T., and Liu, X.Z. (2004). Mutational spectrum in Usher syndrome type II. *Clin. Genet.* 65, 288–293.

Ouyang, X.M., Yan, D., Du, L.L., Hejtmancik, J.F., Jacobson, S.G., Nance, W.E., Li, A.R., Angeli, S., Kaiser, M., Newton, V., Brown, S. D., Balkany, T., and Liu, X.Z. (2005). Characterization of Usher syndrome type I gene mutations in an Usher syndrome patient population. *Hum. Genet.* 116, 292–299.

Pakarinen, L., Tuppurainen, K., Laippala, P., Mantyjarvi, M., and Puhakka, H. (1995). The ophthalmological course of Usher syndrome type III. *Int. Ophthalmol.* 19, 307–311.

Parvari, R., Lei, K.J., Bashan, N., Hershkovitz, E., Korman, S.H., Barash, V., Lerman-Sagie, T., Mandel, H., Chou, J.Y., and Moses, S.W. (1997a). Glycogen storage disease type 1a in Israel: biochemical, clinical, and mutational studies. *Am. J. Med. Genet.* 72, 286–290.

Parvari, R., Moses, S., Shen, J., Hershkovitz, E., Lerner, A., and Chen, Y. T. (1997b). A single-base deletion in the 3′-coding region of glycogen-debranching enzyme is prevalent in glycogen storage disease type IIIA in a population of North African Jewish patients. *Eur. J. Hum. Genet.* 5, 266–270.

Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L., and Syvanen, A.C. (1997). Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res.* 7, 606–614.

Perrault, I., Rozet, J.M., Gerber, S., Ghazi, I., Leowski, C., Ducroq, D., Souied, E., Dufier, J.L., Munnich, A., and Kaplan, J. (1999). Leber congenital amaurosis. *Mol. Genet. Metab.* 68, 200–208.

Petitjean, A., Achatz, M.I., Borresen-Dale, A.L., Hainaut, P., and Olivier, M. (2007). TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene* 26, 2157–2165.

Pezzilli, R., Morselli-Labate, A.M., Mantovani, V., Romboli, E., Selva, P., Migliori, M., Corinaldesi, R., and Gullo, L. (2003). Mutations of the CFTR gene in pancreatic disease. *Pancreas* 27, 332–336.

Pignatti, P.F., Bombieri, C., Marigo, C., Benetazzo, M., and Luisetti, M. (1995). Increased incidence of cystic fibrosis gene mutations in adults with disseminated bronchiectasis. *Hum. Mol. Genet.* 4, 635–639.

Prezant, T.R., Agapian, J.V., Bohlman, M.C., Bu, X., Oztas, S., Qiu, W.Q., Arnos, K.S., Cortopassi, G.A., Jaber, L., Rotter, J.I., Shohat, M., and Fischel-Ghodsian, N. (1993). Mitochondrial ribosomal RNA mutation associated with both antibiotic-induced and non-syndromic deafness. *Nat. Genet.* 4, 289–294.

Pullat, J., Fleischer, R., Becker, N., Beier, M., Metspalu, A., and Hoheisel, J.D. (2007). Optimization of candidate-gene SNP-genotyping by flexible oligonucleotide microarrays; analyzing variations in immune regulator genes of hay-fever samples. *BMC Genomics* 8, 282.

Pullat, J., Kusnezow, W., Jaakson, K., Beier, M., Hoheisel, J.D., and Metspalu, A. (2008). Arrayed primer extension on in situ synthesized 5′>3′ oligonucleotides in microchannels. *New Biotechnol.* 25, 133–141.

Quint, A., Lerer, I., Sagi, M., and Abeliovich, D. (2005). Mutation spectrum in Jewish cystic fibrosis patients in Israel: implication to carrier screening. *Am. J. Med. Genet. A* 136, 246–248.

Radpour, R., Gourabi, H., Gilani, M.A., and Dizaj, A.V. (2007). Molecular study of (TG)m(T)n polymorphisms in Iranian males with congenital bilateral absence of the vas deferens. *J. Androl.* 28, 541–547.

Reardon, W., Coffey, R., Chowdhury, T., Grossman, A., Jan, H., Britton, K., Kendall-Taylor, P., and Trembath, R. (1999). Prevalence, age of onset, and natural history of thyroid disease in Pendred syndrome. *J. Med. Genet.* 36, 595–598.

Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.L., Drumm, M.L., Iannuzzi, M.C., Collins, F.C., and Tsui, L.C. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245, 1066–1073.

Risch, N., de Leon, D., Ozelius, L., Kramer, P., Almasy, L., Singer, B., Fahn, S., Breakefield, X., and Bressman, S. (1995). Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat. Genet.* 9, 152–159.

Risch, N., Tang, H., Katzenstein, H., and Ekstein, J. (2003). Geographic distribution of disease mutations in the Ashkenazi Jewish population supports genetic drift over selection. *Am. J. Hum. Genet.* 72, 812–822.

Rommens, J.M., Iannuzzi, M.C., Kerem, B., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordam, J.R., Tsui, L.C., and Collins, F.C. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* 245, 1059–1065.

Rozen, R., De Braekeleer, M., Daigneault, J., Ferreira-Rajabi, L., Gerdes, M., Lamoureux, L., Aubin, G., Simard, F., Fujiwara, T.M., and Morgan, K. (1992). Cystic fibrosis mutations in French Canadians: three CFTR mutations are relatively frequent in a Quebec population with an elevated incidence of cystic fibrosis. *Am. J. Med. Genet.* 42, 360–364.

Roux, A.F., Faugere, V., Le Guedard, S., Pallares-Ruiz, N., Vielle, A., Chambert, S., Marlin, S., Hamel, C., Gilbert, B., Malcolm, S., and Claustres, M. (2006). Survey of the frequency of USH1 gene mutations in a cohort of Usher patients shows the importance of cadherin 23 and protocadherin 15 genes and establishes a detection rate of above 90%. *J. Med. Genet.* 43, 763–768.

Rychlik, W., and Rhoads, R.E. (1989). A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res.* 17, 8543–8551.

Schrijver, I., Kulm, M., Gardner, P.I., Pergament, E.P., and Fiddler, M.B. (2007). Comprehensive arrayed primer extension array for the detection of 59 sequence variants in 15 conditions prevalent among the (Ashkenazi) Jewish population. *J. Mol. Diagn.* 9, 228–236.

Schrijver, I., Oitmaa, E., Metspalu, A., and Gardner, P. (2005). Genotyping microarray for the detection of more than 200 CFTR mutations in ethnically diverse populations. *J. Mol. Diagn.* 7, 375–387.

Schwarz, M.J., Malone, G.M., Haworth, A., Cheadle, J.P., Meredith, A. L., Gardner, A., Sawyer, I.H., Connarty, M., Dennis, N., Seller, A., Harris, A., Taylor, R., Dear, S., Middleton-Price, H., McMahon, C., Mayall, E., McMahon, R., Barton, D.E., Giles, M., Lindley, V., Plaha, D.S., Price, S., Sharif, A., Cross, G.S., Dalton, A., Taylor, G., Wallace, A., Tassabehji, M., Whittaker, J.L., Butler, R., Curtis, A., Pinkett, R., Gilfillan, A.J., Brock, D.J., Higgins, G.S., Lanyon, G., Miedzybrodzka, Z., Davidson, M., Graham, C.A., and Hill, A. J. (1995). Cystic fibrosis mutation analysis: report from 22 U.K. regional genetics laboratories. *Hum. Mutat.* 6, 326–333.

Schwartz, M., Anvret, M., Claustres, M., Eiken, H.G., Eiklid, K., Schaedel, C., Stolpe, L., and Tranebjaerg, L. (1994). 394delTT: a Nordic cystic fibrosis mutation. *Hum. Genet.* 93, 157–161.

Schwartz, M., Sorensen, N., Brandt, N.J., Hogdall, E., and Holm, T. (1995). High incidence of cystic fibrosis on the Faroe Islands: a molecular and genealogical study. *Hum. Genet.* 95, 703–706.

Shaag, A., Saada, A., Berger, I., Mandel, H., Joseph, A., Feigenbaum, A., and Elpeleg, O.N. (1999). Molecular basis of lipoamide dehydrogenase deficiency in Ashkenazi Jews. *Am. J. Med. Genet.* 82, 177–182.

Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., and Cerrina, F. (1999). Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.* 17, 974–978.

Slatkin, M. (2004). A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. *Am. J. Hum. Genet.* 75, 282–293.

Smith, R.J., Berlin, C.I., Hejtmancik, J.F., Keats, B.J., Kimberling, W. J., Lewis, R.A., Moller, C.G., Pelias, M.Z., and Tranebjaerg, L. (1994). Clinical diagnosis of the Usher syndromes. Usher Syndrome Consortium. *Am. J. Med. Genet.* 50, 32–38.

Snoeckx, R.L., Huygen, P.L., Feldmann, D., Marlin, S., Denoyelle, F., Waligora, J., Mueller-Malesinska, M., Pollak, A., Ploski, R., Murgia, A., Orzan, E., Castorina, P., Ambrosetti, U., Nowakowska-Szyrwinska, E., Bal, J., Wiszniewski, W., Janecke, A.R., Nekahm-Heis, D., Seeman, P., Bendova, O., Kenna, M.A., Frangulov, A., Rehm, H.L., Tekin, M., Incesulu, A., Dahl, H.H., du Sart, D., Jenkins, L., Lucas, D., Bitner-Glindzicz, M., Avraham, K.B., Brownstein, Z., del Castillo, I., Moreno, F., Blin, N., Pfister, M., Sziklai, I., Toth, T., Kelley, P.M., Cohn, E.S., Van Maldergem, L., Hilbert, P., Roux, A.F., Mondain, M., Hoefsloot, L.H., Cremers, C.W., Lopponen, T., Lopponen, H., Parving, A., Gronskov, K., Schrijver, I., Roberson, J., Gualandi, F., Martini, A., Lina-Granade, G., Pallares-Ruiz, N., Correia, C., Fialho, G., Cryns, K., Hilgert, N., Van de Heyning, P., Nishimura, C.J., Smith, R.J., and Van Camp, G. (2005). GJB2 mutations and degree of hearing loss: a multicenter study. *Am. J. Hum. Genet.* 77, 945–957.

Southern, E., Mir, K., and Shchepinov, M. (1999). Molecular interactions on microarrays. *Nat. Genet.* 21, 5–9.

Strom, C.M., Crossley, B., Redman, J.B., Quan, F., Buller, A., McGinniss, M.J., and Sun, W. (2004). Molecular screening for diseases frequent in Ashkenazi Jews: lessons learned from more than 100,000 tests performed in a commercial laboratory. *Genet. Med.* 6, 145–152.

Sue, C.M., Tanji, K., Hadjigeorgiou, G., Andreu, A.L., Nishino, I., Krishna, S., Bruno, C., Hirano, M., Shanske, S., Bonilla, E., Fischel-Ghodsian, N., DiMauro, S., and Friedman, R. (1999). Maternally inherited hearing loss in a large kindred with a novel T7511C mutation in the mitochondrial DNA tRNA(Ser(UCN)) gene. *Neurology* 52, 1905–1908.

Teder, M., Klaassen, T., Oitmaa, E., Kaasik, K., and Metspalu, A. (2000). Distribution of CFTR gene mutations in cystic fibrosis patients from Estonia. *J. Med. Genet.* 37, E16.

Tonisson, N., Zernant, J., Kurg, A., Pavel, H., Slavin, G., Roomere, H., Meiel, A., Hainaut, P., and Metspalu, A. (2002). Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene. *Proc. Natl. Acad. Sci. USA* 99, 5503–5508.

Triggs-Raine, B.L., Mules, E.H., Kaback, M.M., Lim-Steele, J.S., Dowling, C.E., Akerman, B.R., Natowicz, M.R., Grebner, E.E., Navon, R., Welch, J.P., Greenberg, C.R., Thomas, G.H., and Gravel, R.A. (1992). A pseudodeficiency allele common in non-Jewish Tay-Sachs carriers: implications for carrier screening. *Am. J. Hum. Genet.* 51, 793–801.

van Baal, S., Kaimakis, P., Phommarinh, M., Koumbi, D., Cuppens, H., Riccardino, F., Macek, M., Jr, Scriver, C.R., and Patrinos, G.P. (2007). FINDbase: a relational database recording frequencies of genetic defects leading to inherited disorders worldwide. *Nucleic Acids Res.* 35, D690–D695.

Van Camp, G., Willems, P.J., and Smith, R.J. (1997). Nonsyndromic hearing impairment: unparalleled heterogeneity. *Am. J. Hum. Genet.* 60, 758–764.

Verlander, P.C., Lin, J.D., Udono, M.U., Zhang, Q., Gibson, R.A., Mathew, C.G., and Auerbach, A.D. (1994). Mutation analysis of the Fanconi anemia gene FACC. *Am. J. Hum. Genet.* 54, 595–601.

Watson, M.S., Cutting, G.R., Desnick, R.J., Driscoll, D.A., Klinger, K., Mennuti, M., Palomaki, G.E., Popovich, B.W., Pratt, V.M., Rohlfs, E.M., Strom, C.M., Richards, C.S., Witt, D.R., and Grody, W.W. (2004). Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. *Genet. Med.* 6, 387–391.

Wetmur, J.G., and Davidson, N. (1968). Kinetics of renaturation of DNA. *J. Mol. Biol.* 31, 349–370.

Whitney, M.A., Jakobs, P., Kaback, M., Moses, R.E., and Grompe, M. (1994). The Ashkenazi Jewish Fanconi anemia mutation: incidence among patients and carrier frequency in the at-risk population. *Hum. Mutat.* 3, 339–341.

Williams, D.S. (2008). Usher syndrome: animal models, retinal function of Usher proteins, and prospects for gene therapy. *Vision Res.* 48, 433–441.

Xia, J.H., Liu, C.Y., Tang, B.S., Pan, Q., Huang, L., Dai, H.P., Zhang, B. R., Xie, W., Hu, D.X., Zheng, D., Shi, X.L., Wang, D.A., Xia, K., Yu, K.P., Liao, X.D., Feng, Y., Yang, Y.F., Xiao, J.Y., Xie, D.H., and Huang, J.Z. (1998). Mutations in the gene encoding gap junction protein beta-3 associated with autosomal dominant hearing impairment. *Nat. Genet.* 20, 370–373.

Yee, K., Robinson, C., Hurlock, G., Moss, R.B., and Wine, J.J. (2000). Novel cystic fibrosis mutation L1093P: functional analysis and possible Native American origin. *Hum. Mutat.* 15, 208.

Yilmaz, E., Erdem, H., Ozguc, M., Coskun, T., Ozcelik, U., Gocmen, A., and Ozalp, I. (1995). Study of 12 mutations in Turkish cystic fibrosis patients. *Hum. Hered.* 45, 175–177.

Yzer, S., Leroy, B.P., De Baere, E., de Ravel, T.J., Zonneveld, M.N., Voesenek, K., Kellner, U., Ciriano, J.P., de Faber, J.T., Rohrschneider, K., Roepman, R., den Hollander, A.I., Cruysberg, J.R., Meire, F., Casteels, I., van Moll-Ramirez, N.G., Allikmets, R., van den Born, L. I., and Cremers, F.P. (2006). Microarray-based mutation detection and phenotypic characterization of patients with Leber congenital amaurosis. *Invest. Ophthalmol. Vis. Sci.* 47, 1167–1176.

Zelante, L., Gasparini, P., Estivill, X., Melchionda, S., D'Agruma, L., Govea, N., Mila, M., Monica, M.D., Lutfi, J., Shohat, M., Mansfield, E., Delgrosso, K., Rappaport, E., Surrey, S., and Fortina, P. (1997). Connexin26 mutations associated with the most common form of non-syndromic neurosensory autosomal recessive deafness (DFNB1) in Mediterraneans. *Hum. Mol. Genet.* 6, 1605–1609.

Zernant, J., Kulm, M., Dharmaraj, S., den Hollander, A.I., Perrault, I., Preising, M.N., Lorenz, B., Kaplan, J., Cremers, F.P., Maumenee, I., Koenekoop, R.K., and Allikmets, R. (2005). Genotyping microarray (disease chip) for Leber congenital amaurosis: detection of modifier alleles. *Invest. Ophthalmol. Vis. Sci.* 46, 3052–3059.

Zielenski, J., Fujiwara, T.M., Markiewicz, D., Paradis, A.J., Anacleto, A.I., Richards, B., Schwartz, R.H., Klinger, K.W., Tsui, L.C., and Morgan, K. (1993). Identification of the M1101K mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) gene and complete detection of cystic fibrosis mutations in the Hutterite population. *Am. J. Hum. Genet.* 52, 609–615.

Chapter 18

Application of Proteomics to Disease Diagnostics

Samir Hanash, Ji Qiu and Vitor Faca
Department of Molecular Diagnostics, Fred Hutchinson Cancer Research Center, Seattle, WA

18.1 INTRODUCTION

Progress in the application of proteomics to disease diagnostics has been made at a rapid pace. This progress is manifested notably by reduced emphasis on two-dimensional gel methodology for discovery and increased emphasis on mass spectrometry resulting from innovations in instrumentation and experimental design and increased emphasis on protein microarrays given their enhanced content.

Proteomics is particularly useful for disease diagnostics as a substantial proportion of diagnostic tests in the clinical laboratory target proteins. This chapter reviews in particular the application of proteomics from the perspective of cancer diagnostics in view of the expertise of the authors. The emphasis is on the application of sound strategies rather than detailed description of methodologies and instrumentation.

18.2 MASS SPECTROMETRY CENTRIC STRATEGIES

Mass spectrometry has evolved from a tool to identify and characterize isolated proteins, such as from 2D gels and from mass peak-based profiling as in the application of matrix assisted laser desorption ionization (MALDI) to clinical samples, to a platform for interrogating complex proteomes by matching mass spectra to sequence databases to derive protein identifications (Cox and Mann, 2007). At present, a trend has emerged to replace laborious 2D gel-based protein separations for proteomic profiling, which have limited sensitivity and proteome coverage, with liquid chromatography-based protein and peptide separations in combination with mass spectrometry-based profiling (see also Chapter 21), which is viewed as more sensitive, more reliable, and potentially rich in information content. However, as mass spectrometry technologies improve, yielding ever more complex sets of data, so does the need for more elaborate informatics and statistical analyses tools to deal with data complexity.

Only a few years ago, the use of mass spectrometry for clinical applications was largely equated rather erroneously with the use of MALDI-based mass spectrometry profiling of tissues and biological fluids, namely Surface Enhanced Laser Desorption Ionization (SELDI), largely as a result of early publications that showed substantial promise of this approach, despite the limited resolution achieved and the substantial difficulty in identifying proteins and peptides corresponding to potentially informative mass peaks. Numerous studies reported mass profiles for unidentified proteins that were proposed to be diagnostic for several common types of cancer but whose validity was subsequently questioned (Ransohoff, 2005). In a multi-institutional rigorous prostate cancer validation study a classifier developed with a training set failed to distinguish between cancer cases and controls in a validation set (McLerran *et al.*, 2008). The prior highly promising results likely resulted from the use of suboptimal samples with built-in biases for discovery.

Even with substantial improvements in sensitivity and mass accuracy, the complexities of tissue and biological fluid proteomes substantially exceed the sampling capabilities of mass spectrometry instruments to fully identify and provide measures for all protein and peptide constituents in a single analysis. Therefore current strategies to achieve in-depth coverage require sample fractionation followed by separate analyses of individual fractions (Misek *et al.*, 2005; Tang *et al.*, 2005; Wang and Hanash, 2005; Faca *et al.*, 2007), or capture of protein or peptide subsets, such as glycosylated proteins or phosphopeptides that are analyzed exclusively (Anderson *et al.*, 2004; Wu *et al.*, 2006; Bernhard *et al.*, 2007; Zhou *et al.*, 2007). It follows that there is a trade-off between in-depth analysis but with limited throughput and limited coverage of proteome with

enhanced throughput (Fig. 18.1). The choice depends on the relative abundance of candidate markers sought and the dynamic range of protein abundance in the biological material under investigation. While the emphasis in the recent past has been on achieving depth of analysis and reliable identifications using mass spectrometry, current trends include increased emphasis on characterization of post-translational modifications and other types of isoforms as may result from alternative splicing (Cox and Mann, 2007), development of means to further increase peak capacity (Liu *et al.*, 2007), and development of high-throughput strategies for validation studies for a limited number of informative proteins and peptides in complex mixtures (Stahl-Zeng *et al.*, 2007).

18.3 DISCOVERY APPROACHES vs. APPROACHES FOR VALIDATION AND CLINICAL IMPLEMENTATION

Ideally, discovery technologies should allow an unbiased quantitative analysis of all proteins and their isoforms in a biological sample for the purpose of identification of the most informative markers. Thus depth of analysis is particularly important (Diamandis, 2004). The procedures and instrumentation for discovery need not be of general applicability in the clinical laboratory and may remain in the domain of the research/discovery laboratories. On the other hand, technologies for marker evaluation need to allow analysis of a predetermined set of candidates using accurate and reproducible procedures that are relatively easily implementable in various laboratories, that may be transferable to the clinical laboratory at a relatively reduced cost per sample, per marker, and that can interrogate a large number of samples efficiently.

There is increasing interest in evaluating mass spectrometry as a means to quantify proteins of interest. For single proteins, the approach is referred to as single reaction monitoring (SRM), and the multiplex version is referred to as multiple reaction monitoring (MRM). These approaches rely on accurate mass spectrometer filters to select specific peptides, fragment them, and monitor intensity of specific fragments in order to quantify these species in the sample. The use of stable isotopes of peptides as internal standards provides a very accurate and reliable way to obtain absolute quantitation of the peptides of interest. The multiplexing features of the MRM approach can be improved by fractionating the sample prior to MS analysis, coupling the mass spectrometer to a chromatographic system and performing a liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) experiment. This way, several peptides can be monitored simultaneously in a single run (Anderson and Hunter, 2006). The contributions of MRM to biomarker development remain to be determined and at the present time there are limited data available with respect to extent of multiplexing and the limits of sensitivity are still not reaching biomarker levels in crude plasma. Moreover, this technology is applicable only to previously well-characterized peptides.

There are currently several approaches available for further testing of identified candidate biomarkers based on capture agents. Sandwich ELISA assays represent a

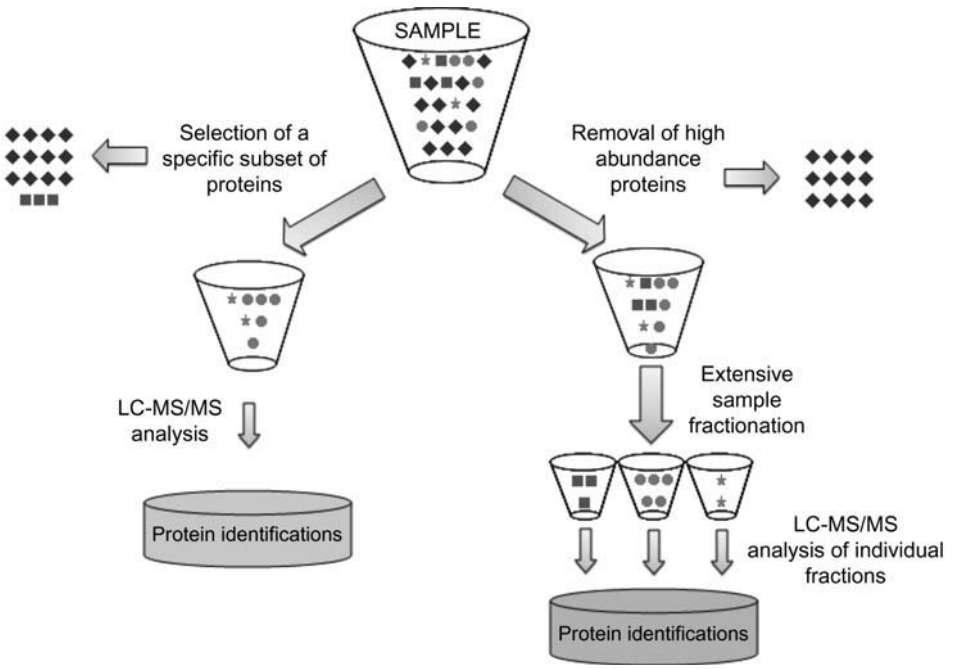


FIGURE 18.1 Proteomic strategies to identify low abundance proteins.

well-established approach. However, numerous assays formats are currently in use. Some rely on bead-based assays and others utilize flat surfaces. Beads have the advantage of being dispersed throughout the assay fluid, which facilitates reagent mixing. In contrast, concentrating capture reagent on a spot on a flat surface decreases reagent interactions. Therefore, flat-surface assays have slower reaction kinetics, although the final equilibrium constants apparently are the same (Kusnezow *et al.*, 2006). Another factor is the typical amount of capture reagent used in the two systems. The low amount of capture reagent in the microarray format will only bind about 2% of the analyte molecules at saturation, which, however, are concentrated into a small area of high signal density (Saviranta *et al.*, 2004). In contrast, the much larger amount of capture reagent in bead assays will capture a higher percentage of analyte molecules present in low concentration in samples, but may also adversely affect the lower limit of detection because nonspecific background binding is proportional to the amount of capture antibody present.

Approaches to improve sensitivity include immuno-PCR, which has proven to be a sensitive method for measuring protein abundance with a wide dynamic range. It is possible to perform the sample and antibody incubation steps using beads or microtiter plates. Limitations with this methodology include an elaborate, multi-step protocol and the need for specialized equipment and reagents (Niemeyer *et al.*, 2005). The above considerations are presented not to be exhaustive but rather to make the point that discovery strategies have different requirements than strategies focused on testing markers already discovered.

18.4 QUANTITATIVE APPROACHES TO PROFILING USING MASS SPECTROMETRY

The quantitative analysis of proteomes has reached a mature stage, with a rich repertoire of technologies available. Quantitative methods currently available can be grouped into label-free methods, which compare pure peptide ion intensities between MS analysis, and isotope labeling methods, which compare relative peptide ion intensities in the MS analysis of mixture of peptides differentially labeled with stable isotopes. Isotope-labeling methods are more accurate, since isotope pairs behave identically in MS analysis, with several different approaches available. Cells grown in culture can be differentially isotope labeled *in vivo* using medium enriched with ^{15}N, or stable isotope labeled amino acids. The isotope labeling of samples can be performed during tryptic digestion with ^{18}O, or after enzymatic digestion with a vast variety of tags. Labeled peptides can even be isolated and enriched from peptide mixtures, as with the Isotope Coded Affinity Tag (ICAT) technology (Gygi *et al.*, 1999). However, if samples to be compared have to be processed separately until the labeling step, large quantitation errors may be introduced by artifactual variations. Quantitative proteomics is reviewed in Sechi and Oda (2003), Julka and Regnier (2004), and Ong and Mann (2005).

Isotopic labeling of intact proteins in a mixture has the advantage of reducing artifactual quantitative variation and allowing differential expression of protein isoforms to be uncovered. Although the application of isotope labeling of intact protein in the literature is limited, this approach is catching more attention and two new commercial kits are available exclusively for this application. The method termed isotope-coded protein label (ICPL) labels lysine residues of proteins with nicotinic acid isotopes, permitting the fractionation of intact proteins in a high-throughput mode and obtaining high-protein identification coverage (Schmidt *et al.*, 2005). Another method under development uses the already commercialized Isobaric Tags for Relative and Absolute Quantitation (iTRAQ – Applied Biosystems) to tag free amines in intact proteins. The particularity of isobaric tags consists of the acquisition of relative quantitation from fragment ions generated by Collision-Induced Dissociation (CID), but the greatest advantage of this method is the possibility of multiplex analysis of up to eight different samples mixed together. For both of these methods, there are the disadvantages of changes in isoeletric point of protein and losing the tryptic sites for peptide cleavage due to labeled Lys residues, which obligates the use of alternative enzymes.

One approach we have implemented for intact protein-based analysis consists of cysteine alkylation with acrylamide isotopes (Sechi, 2002; Faca *et al.*, 2006). Acrylamide is a small reagent (mass = 71) that does not introduce significant mass shift or charge changes in the protein and does not negatively affect protein solubility. The alkylation reaction is performed using standard protein solubilization conditions with a virtually 100% yield. Additionally, the reagents are relatively inexpensive, making it practical to perform experiments starting with large amounts of protein as needed for extensive fractionation and in-depth analysis (Faca *et al.*, 2006). After labeling two different samples with the light (acrylamide) and heavy acrylamide isotopes (D_3-acrylamide or 1,2,3-$^{13}C_3$-acrylamide), samples are mixed and submitted to multi-dimensional protein fractionation.

We utilized isotope acrylamide labeling of intact proteins to identify changes in human serum associated with cancer. Using a two-dimensional liquid chromatography fractionation approach (anion-exchange followed by reverse-phase chromatography) following mixing of differentially labeled samples, intact proteins were fractionated in about 130 fractions that were digested with trypsin and individually analyzed by liquid chromatography-mass spectrometry. Well over 1,500 proteins have been confidently identified and relative quantitation information is obtained for more than 40%. As a remarkable characteristic of this approach, proteins and protein fragments, as well

as their post-translational modifications, can be separated and quantified.

18.5 MASS SPECTROMETRY-BASED PROFILING OF BIOLOGICAL FLUIDS TO IDENTIFY CANDIDATE BIOMARKERS

Various types of biological fluids, notably plasma and urine, but also fluids more proximal to disease tissues, such as effusions, represent an important biomarker discovery source. This is particularly advantageous given the ease with which the same fluids used for discovery may be interrogated for diagnosis in the clinical setting, unlike, for example, strategies whereby inferences are made from a tissue source to be applied to a different diagnostic source, namely plasma. However, studies of biological fluids present several challenges related to their complexity and vast dynamic range of protein abundance. Biological fluids, notably plasma, may contain several thousand proteins with concentrations ranging from as high as 20–50 mg/ml for serum albumin to femtomolar concentrations for some known biomarkers (States *et al.*, 2006). Plasma consists of vast assemblies of proteins and complexes that reflect the physiologic or pathologic state of cells, tissues and organs, with many of the plasma proteins found in multiple forms (Misek *et al.*, 2005). Some forms result from alternative splicing at the RNA level and others result from cleavages and post-translational modifications (Nedelkov *et al.*, 2005).

The reliability and reproducibility of comprehensive plasma and serum proteome profiles are exquisitely dependent on sample preparation to the point that differences in patterns may be elicited by no more than differences in the gauge of the needle used to draw blood. Therefore it is critically important to standardize sample collection, storage, and distribution for proteomic analysis. Plasma is preferable to serum; the choice is related to observations that clotting, in addition to depleting factors, may release breakdown products from blood cells that are damaged during the clotting process. There are currently several agents for preventing blood clotting at the time of sample collection, including heparin, citrate, and EDTA. It is currently unclear to what extent anticoagulants have a significant impact on mass spectrometry. Data to date (Mei *et al.*, 2003) suggest significant matrix effects caused by heparin and not by EDTA. In addition, heparin may bind to many proteins affecting their detection and identification. Furthermore, the anionic nature of heparin and the acidic properties of EDTA and citrate may have an impact on protein separations through ion exchange columns in LC systems.

Serum and plasma are rich in numerous proteases which may cause protein degradation both *in vivo*, before sample collection, and *in vitro*, after sample collection (Zhang *et al.*, 2004). This raises the issue of merits of cocktails of various protease inhibitors for maintaining sample integrity. Although quite a few studies of plasma proteins have included inhibitors, there is a need to assess which protease inhibitors, if any, have utility without analytical interferences. It should be kept in mind that EDTA is a known metallo-protease inhibitor; therefore the choice of EDTA as an anticoagulant would have added benefit as protease inhibitor.

The development and application of methods for specifically and efficiently removing high-abundance proteins from various biological fluids, in particular serum and plasma, is being broadly pursued to enhance discovery of candidate disease biomarkers through the application of proteomics technologies. By removing the high-abundance plasma proteins, the tremendous dynamic range in protein concentration is effectively reduced, and a significantly greater amount of low-abundance proteins can be detected in downstream analyses. Antibody-based immunoaffinity subtraction systems are at the forefront of removal methods due to their high efficacy and superior reproducibility. However, despite the increasing popularity of immunoaffinity subtraction systems that permit better detection of low-abundance proteins (Pieper *et al.*, 2003; Chromy *et al.*, 2004; Echan *et al.*, 2005), questions still arise with regard to their comparable efficiency, reproducibility, and selectivity.

18.6 MICROARRAY-BASED PROFILING

Various microarray formats in which protein capture agents, notably antibodies, or recombinant or natural proteins, or cell and tissue lysates, are arrayed and interrogated are slowly complementing mass spectrometry for biomarker discovery and validation (Kingsmore, 2006). Protein microarrays provide a high-throughput, sensitive, and low-volume sample consumption platform for various assays. There are generally two major classes of protein microarrays. One is intended for protein profiling in which multiple protein capture agents, most commonly various types of antibodies, are spotted to assay the abundance of corresponding antigens or epitopes in a biological sample. Alternatively, multiple biological samples are spotted to assay for proteins that interact with a specific analyte in the biological samples applied to the array, such as interaction with particular lectins that recognize a subtype of glycoproteins. Another class is intended for functional protein analysis in which large numbers of purified proteins are spotted to study their biochemical properties.

Unlike DNA microarrays, which provide only one measure of gene expression, namely RNA levels, many different features of proteins may be addressed by different types of protein microarrays, including determination of their functional states, as may be deduced for example

from their extent of phosphorylation. Many cellular processes are regulated by reversible protein phosphorylation. A systematic profiling of functional states of proteins involved in signal transduction pathways provides insight into disease development and treatment. The development of reagents that allow assessment of protein modification such as phosphorylation, glycosylation, or other functionally relevant protein changes has substantial utility for clinical investigations and diagnostics (Speer *et al.*, 2007). The phosphorylation status of proteins can be measured using specific anti-phosphoprotein antibodies arrayed in microarrays, permitting a high-throughput and multiplexing interrogation of the state of entire signaling pathways (Gulmann *et al.*, 2006). For example, protein pathway analysis of human rhabdomyosarcoma with this methodology revealed a strong association between activation (phosphorylation) of multiple components of the Akt/mTOR pathway and a poor disease outcome or overall survival (Petricoin *et al.*, 2007). One recently developed alternative to antibody microarrays consists of peptide microarrays to detect protein kinase activity in cell lysates by arraying substrate peptides for kinases (Shigaki *et al.*, 2007).

Perhaps more than any other application of proteomics, the discovery of markers for early disease detection, notably cancer detection, represents a hot topic that is highly promising and at the same time highly challenging. For most tumor types, the earliest stages of tumor development remain poorly understood and the involved tissue ill defined or inaccessible, thus complicating strategies for early detection. An ideal screening test for the early detection of cancer is a non-invasive test based on analysis of blood or other biological fluids that is applied to subjects at risk for cancer. A promising approach for early cancer detection is through harnessing the immune response directed against tumor antigens. The identification of a panel of antigenic markers that are tumor specific and that elicit immunoreactivity early in tumor development and at a high frequency would provide an effective strategy for cancer screening. There is also equal interest in identifying proteins associated with particular autoimmune disorders.

Microarray formats have been successfully utilized for the discovery of tumor antigens that induce autoantibodies. Recombinant protein microarrays were utilized to screen for autoantibodies in ovarian cancer (Hudson *et al.*, 2007). Microarrays containing spotted proteins derived from tumors or tumor cell lines allow analysis of proteins and peptides in their post-translationally modified states, as they occur in cancer cells (Fig. 18.2). Modifications, such as glycosylation, may be immunogenic, and therefore there is merit in utilizing approaches that preserve such epitopes. Such microarrays containing the repertoire of natural proteins expressed in tumor cells have the potential to substantially accelerate the pace of discovery of tumor antigens and could provide a molecular signature for immune responses in different types of cancer (Madoz-Gurpide *et al.*, 2008).

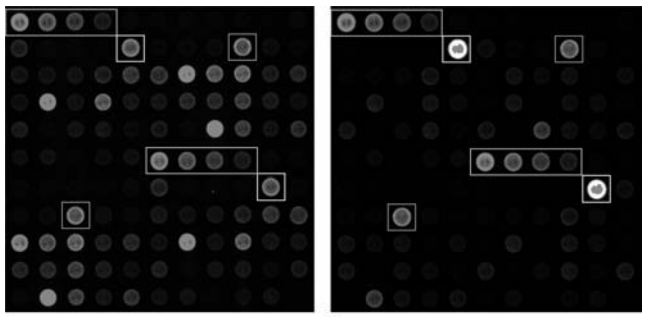


FIGURE 18.2 Sample natural protein microarray images. A549 natural protein microarray was prepared from fractions generated from extensive two-dimensional separation of lung adenocarcinoma cell line A549 cell lysates with anion exchange HPLC followed by reverse-phase HPLC. A total of 1,820 final fractions were spotted in replicate in cyclic. Partial images for two A549 natural protein microarrays hybridized with a cancer serum sample **A** and a normal control serum sample **B**, respectively, were shown.

18.7 CONCLUDING REMARKS

An important aspect of the application of proteomics to the identification of disease biomarkers pertains to the substantial need for informatics resources. The proteomics field is now capturing the attention of computer scientists and bioinformaticians. New computational tools are being developed in several different areas, ranging from better algorithms for protein identification and measurements of statistical confidence in identification, to *de novo* peptide sequencing and whole genome databank searches.

A useful repertoire of proteomics technologies is currently available for disease-related applications, though further technological innovations would be beneficial to increase sensitivity, reduce sample size requirements, increase throughput, and more effectively uncover various types of protein alterations, such as post-translational modifications. The application of high-throughput procedures to the discovery of biomarkers leads to large datasets of multi-dimensional data and a certain percentage of false discoveries. It becomes crucial, therefore, to carefully design validation studies and characterize in a representative population the sensitivity and specificity (Pepe, 2003) of each biomarker candidate. Additionally, it is more likely that a robust predictor of a disease state is a combination of markers (Simon, 2006).

REFERENCES

Anderson, L., and Hunter, C.L. (2006). Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol. Cell. Proteomics* 5, 573–588.

Anderson, N.L., Anderson, N.G., Haines, L.R., Hardie, D.B., Olafson, R.W., and Pearson, T.W. (2004). Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J. Proteome Res.* 3, 235–244.

Bernhard, O.K., Kapp, E.A., and Simpson, R.J. (2007). Enhanced analysis of the mouse plasma proteome using cysteine-containing tryptic glycopeptides. *J. Proteome Res.* 6, 987–995.

Chromy, B.A., Gonzales, A.D., Perkins, J., Choi, M.W., Corzett, M.H., Chang, B.C., Corzett, C.H., and McCutchen-Maloney, S.L. (2004). Proteomic analysis of human serum by two-dimensional differential gel electrophoresis after depletion of high-abundant proteins. *J. Proteome Res.* 3, 1120–1127.

Cox, J., and Mann, M. (2007). Is proteomics the new genomics? *Cell* 130, 395–398.

Diamandis, E.P. (2004). Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. *J. Natl. Cancer Inst.* 96, 353–356.

Echan, L.A., Tang, H.Y., Ali-Khan, N., Lee, K., and Speicher, D.W. (2005). Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and plasma. *Proteomics* 5, 3292–3303.

Faca, V., Coram, M., Phanstiel, D., Glukhova, V., Zhang, Q., Fitzgibbon, M., McIntosh, M., and Hanash, S. (2006). Quantitative analysis of acrylamide labeled serum proteins by LC-MS/MS. *J. Proteome Res.* 5, 2009–2018.

Faca, V., Pitteri, S., Newcomb, L., Glukhova, V., Phanstiel, D., Krasnoselsky, A., Zhang, Q., Struthers, J., Wang, H., Eng, J., Fitzgibbon, M., McIntosh, M., and Hanash, S. (2007). Contribution of protein fractionation to depth of analysis of the serum and plasma proteomes. *J. Proteome Res.* 6, 3558–3565.

Gulmann, C., Sheehan, K.M., Kay, E.W., Liotta, L.A., and Petricoin, E. F., 3rd (2006). Array-based proteomics: mapping of protein circuitries for diagnostics, prognostics, and therapy guidance in cancer. *J. Pathol.* 208, 595–606.

Gygi, S., Rist, B., Gerber, S., Turecek, F., Gelb, M., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999.

Hudson, M.E., Pozdnyakova, I., Haines, K., Mor, G., and Snyder, M. (2007). Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays. *Proc. Natl. Acad. Sci. USA* 104, 17494–17499.

Julka, S., and Regnier, F. (2004). Quantification in proteomics through stable isotope coding: a review. *J. Proteome Res.* 3, 350–363.

Kingsmore, S.F. (2006). Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nat. Rev. Drug Discov.* 5, 310–320.

Kusnezow, W., Syagailo, Y.V., Goychuk, I., Hoheisel, J.D., and Wild, D. G. (2006). Antibody microarrays: the crucial impact of mass transport on assay kinetics and sensitivity. *Expert Rev. Mol. Diagn.* 6, 111–124.

Liu, X., Valentine, S.J., Plasencia, M.D., Trimpin, S., Naylor, S., and Clemmer, D.E. (2007). Mapping the human plasma proteome by SCX-LC-IMS-MS. *J. Am. Soc. Mass Spectrom.* 18, 1249–1264.

Madoz-Gurpide, J., Kuick, R., Wang, H., Misek, D.E., and Hanash, S. M. (2008). Integral protein microarrays for the identification of lung cancer antigens in sera that induce a humoral immune response. *Mol. Cell. Proteomics* 7, 268–281.

McLerran, D., Grizzle, W.E., Feng, Z., Thompson, I.M., Bigbee, W.L., Cazares, L.H., Chan, D.W., Dahlgren, J., Diaz, J., Kagan, J., Lin, D. W., Malik, G., Oelschlager, D., Partin, A., Randolph, T.W., Sokoll, L., Srivastava, S., Srivastava, S., Thornquist, M., Troyer, D., Wright, G. L., Zhang, Z., Zhu, L., and Semmes, O.J. (2008). SELDI-TOF MS whole serum proteomic profiling with IMAC surface does not reliably detect prostate cancer. *Clin. Chem.* 54, 53–60.

Mei, H., Hsieh, Y., Nardo, C., Xu, X., Wang, S., Ng, K., and Korfmacher, W.A. (2003). Investigation of matrix effects in bioanalytical high-performance liquid chromatography/tandem mass spectrometric assays: Application to drug discovery. *Rapid Commun. Mass Spectrom.* 34, 375–379.

Misek, D.E., Kuick, R., Wang, H., Deng, B., Zhao, R., Galchev, V., Tra, J., Pisano, M.R., Amunugama, R., Allen, D., Strahler, J., Andrews, P., Omenn, G.S., and Hanash, S.M. (2005). A wide range of protein isoforms in serum and plasma uncovered by a quantitative Intact Protein Analysis System (IPAS). *Proteomics* 5, 3343–3352.

Nedelkov, D., Kiernan, U.A., Niederkofler, E.E., Tubbs, K.A., and Nelson, R.W. (2005). Investigating diversity in human plasma proteins. *Proc. Natl. Acad. Sci. USA* 102, 10852–10857.

Niemeyer, C.M., Adler, M., and Wacker, R. (2005). Immuno-PCR: high sensitivity detection of proteins by nucleic acid amplification. *Trends Biotechnol.* 23, 208–216.

Ong, S.E., and Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* 1, 252–262.

Pepe, M. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press.

Petricoin, E.F., 3rd, Espina, V., Araujo, R.P., Midura, B., Yeung, C., Wan, X., Eichler, G.S., Johann, D.J., Jr., Qualman, S., Tsokos, M., Krishnan, K., Helman, L.J., and Liotta, L.A. (2007). Phosphoprotein pathway mapping: Akt/mammalian target of rapamycin activation is negatively associated with childhood rhabdomyosarcoma survival. *Cancer Res.* 67, 3431–3440.

Pieper, R., Gatlin, C.L., Makusky, A.J., Russo, P.S., Schatz, C.R., Miller, S.S., Su, Q., McGrath, A.M., Estock, M.A., Parmar, P.P., Zhao, M., Huang, S.T., Zhou, J., Wang, F., Esquer-Blasco, R., Anderson, N.L., Taylor, J., and Steiner, S. (2003). The human serum proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. *Proteomics* 3, 1345–1364.

Ransohoff, D.F. (2005). Bias as a threat to the validity of cancer molecular-marker research. *Nat. Rev. Cancer* 5, 142–149.

Saviranta, P., Okon, R., Brinker, A., Warashina, M., Eppinger, J., and Geierstanger, B.H. (2004). Evaluating sandwich immunoassays in microarray format in terms of the ambient analyte regime. *Clin. Chem.* 50, 1907–1920.

Schmidt, A., Kellermann, J., and Lottspeich, F. (2005). A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* 5, 4–15.

Sechi, S. (2002). A method to identify and simultaneously determine the relative quantities of proteins isolated by gel electrophoresis. *Rapid Commun. Mass Spectrom.* 16, 1416–1424.

Sechi, S., and Oda, Y. (2003). Quantitative proteomics using mass spectrometry. *Curr. Opin. Chem. Biol.* 7, 70–77.

Shigaki, S., Yamaji, T., Han, X., Yamanouchi, G., Sonoda, T., Okitsu, O., Mori, T., Niidome, T., and Katayama, Y. (2007). A peptide microarray for the detection of protein kinase activity in cell lysate. *Anal. Sci.* 23, 271–275.

Simon, R. (2006). Validation of pharmacogenomic biomarker classifiers for treatment selection. *Cancer Biomark.* 2, 89–96.

Speer, R., Wulfkuhle, J., Espina, V., Aurajo, R., Edmiston, K.H., Liotta, L. A., and Petricoin, E.F., 3rd (2007). Development of reverse phase protein microarrays for clinical applications and patient-tailored therapy. *Cancer Genomics Proteomics* 4, 157–164.

Stahl-Zeng, J., Lange, V., Ossola, R., Aebersold, R., and Domon, B. (2007). High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol. Cell. Proteomics* 6, 1809–1817.

States, D.J., Omenn, G.S., Blackwell, T.W., Fermin, D., Eng, J., Speicher, D.W., and Hanash, S.M. (2006). Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat. Biotechnol.* 24, 333–338.

Tang, H.Y., Ali-Khan, N., Echan, L.A., Levenkova, N., Rux, J.J., and Speicher, D.W. (2005). A novel four-dimensional strategy combining protein and peptide separation methods enables detection of low-abundance proteins in human plasma and serum proteomes. *Proteomics* 5, 3329–3342.

Wang, H., and Hanash, S. (2005). Intact-protein based sample preparation strategies for proteome analysis in combination with mass spectrometry. *Mass Spectrom. Rev.* 24, 413–426.

Wu, S.L., Kim, J., Bandle, R.W., Liotta, L., Petricoin, E., and Karger, B. L. (2006). Dynamic profiling of the post-translational modifications and interaction partners of epidermal growth factor receptor signaling after stimulation by epidermal growth factor using Extended Range Proteomic Analysis (ERPA). *Mol. Cell. Proteomics* 5, 1610–1627.

Zhang, R., Barker, L., Pinchev, D., Marshall, J., Rasamoelisolo, M., Smith, C., Kupchak, P., Kireeva, I., Ingratta, L., and Jackowski, G. (2004). Mining biomarkers in human sera using proteomic tools. *Proteomics* 4, 244–256.

Zhou, Y., Aebersold, R., and Zhang, H. (2007). Isolation of N-linked glycopeptides from plasma. *Anal. Chem.* 79, 5826–5837.

This page intentionally left blank

Chapter 19

RNA-Based Variant Detection: The Protein Truncation Test

Johan T. den Dunnen
Leiden Genome Technology Center, Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

19.1 SUMMARY

Only changes in the DNA sequence manifesting deleterious effects at a functional level provide pathogenic (disease-causing) variants. Consequently, DNA variant scanning techniques applied on a protein level, directly interrogating protein function, would be most informative. However, due to a lack of powerful protein-based methods and of detailed functional knowledge, most currently applied techniques try to resolve variants indirectly, i.e. at the DNA level. In this respect, DNA variant scanning techniques applied at the intermediate level – RNA – have several advantages.

The Protein Truncation Test has been the most widely applied RNA-based technique, revealing nearly exclusively disease-causing variants. The PTT assay generates a DNA copy of the RNA (the cDNA) that is translated into protein, immediately revealing variants that cause premature termination of protein translation, i.e. truncating variants. In general, RNA-based assays save work since a condensed target, i.e. the cDNA instead of all the exons, is analysed. PTT is able to scan regions as large as 3.5 kb in one analysis. PTT points to the site of the variant and its detection efficiency is close to 100%. In addition, as an RNA-based technique, it reveals variants that influence the processing of the transcript (e.g. splicing, poly-adenylation) as well as the absolute level of transcription. Since processes like nonsense-mediated mRNA decay and non-random X-inactivation may obscure the results obtained, an essential quality check to be performed is to verify whether transcripts of both alleles were amplified. When PTT reveals no deleterious change, the cDNA fragments generated can be used to apply standard DNA-based variant scanning techniques, e.g. to detect variants that do not truncate protein translation (including missense changes).

19.2 INTRODUCTION

Many different changes may occur at the DNA level, including changes at the nucleotide sequence level (nucleotide substitutions, deletions, duplications, insertions), changes in amount (deletions and duplications at the gene or exon level) and changes in position (insertions, inversions, translocations, and transpositions). Although these changes can be detected readily at the DNA level, their ultimate outcome for the organism is difficult to predict. In the cell, DNA is transcribed to RNA and then translated into protein. It is at this last level that the consequences of a change will become evident. Changes at DNA level may yield proteins that are non-functional, comprise altered deleterious function, and are produced at too high or low levels and/or in the wrong tissue or at the wrong time during development.

Ultimately, only changes in the DNA sequence manifesting deleterious effects at a functional level provide pathogenic (disease-causing) variants, i.e. the changes that are identified in a diagnostic setting. Consequently, variant scanning techniques applied at the protein level, directly interrogating protein function, would be most informative. Although theoretically simple, in practice variant analysis at the protein level is very complicated. The protein of interest can, in general, be difficult to obtain; the tissue where this protein is expressed might not be available, isolation can mostly give low yields and the isolate may contain several co-purified contaminants. In addition, unlike PCR for DNA there are no techniques to amplify proteins to yields required for analysis. Even when purified, it is very difficult to test a protein on a functional level when – as for many proteins – its exact function is unknown. As a result, due to the lack of powerful methods and of

functional knowledge, most techniques try to resolve variants indirectly at the DNA level.

RNA is the intermediate between DNA and protein. Isolation of RNA is rather simple and RNA can be amplified using reverse transcription and PCR (RT-PCR). Variant scanning techniques applied at the RNA level have several advantages. First, especially true for large multi-exon genes, an RNA-based variant scanning technique saves a lot of work since it only analyses the condensed protein-coding region. Second, only analysis at the RNA level is able to reveal variants that influence either RNA processing (i.e. splicing, poly-adenylation) or its level of expression. Third, since the analysis generates a DNA copy of the mRNA (the cDNA), its translation is a way of studying the encoded protein. Finally, some variants that were missed at the DNA level, in particular certain deletions and duplications, may be revealed more easily at the RNA level. However, RNA-based analysis has some intrinsic problems. First, in diagnostic laboratories DNA is the standard and from older samples only DNA has been stored. Second, working with RNA is laborious, technically more demanding, and RNA degradation is a continuous threat. Finally, differences in expression from the two chromosomes in the cell complicate analysis, in particular the process of nonsense-mediated mRNA decay (NMD; Dietz and Kendzior, 1994) that tends to degrade the copy carrying the mutated allele. Consequently, when RNA-based assays are applied, additional controls are required proving that transcripts of both chromosomes have been analysed.

19.3 THE PROTEIN TRUNCATION TEST

The "Protein Truncation Test" (PTT; Roest *et al*., 1993; Van Der Luijt *et al*., 1994), also known as the In-Vitro Synthesized Protein (IVSP) assay (Powell *et al*., 1993), was developed specifically to zoom in on variants that compromise normal protein translation. PTT was originally developed to study Duchenne muscular dystrophy (DMD; Den Dunnen *et al*., 1989; Roest *et al*., 1993), a neuromuscular disorder caused by variants in the human *DMD* gene located on chromosome Xp21. About two-thirds of the pathogenic changes detectable then were large intragenic deletions (~60%) and duplications (~5–10%) that could simply be detected using Southern blot analysis or a multiplex-PCR (Den Dunnen *et al*., 1989; Den Dunnen and Beggs, 2006). However, the complexity and size of the gene, 79 exons spread over 2.3 Mb of DNA, made all efforts to scan the remaining patients for small variants (nucleotide changes) a daunting task. In addition, initial analysis showed that the gene contained many irrelevant variants (polymorphisms), making the available methods to pick up sequence variants (mobility shift assays) lighting up many false positive signals. It was noted, however, that the common denominator of the deleterious variants identified

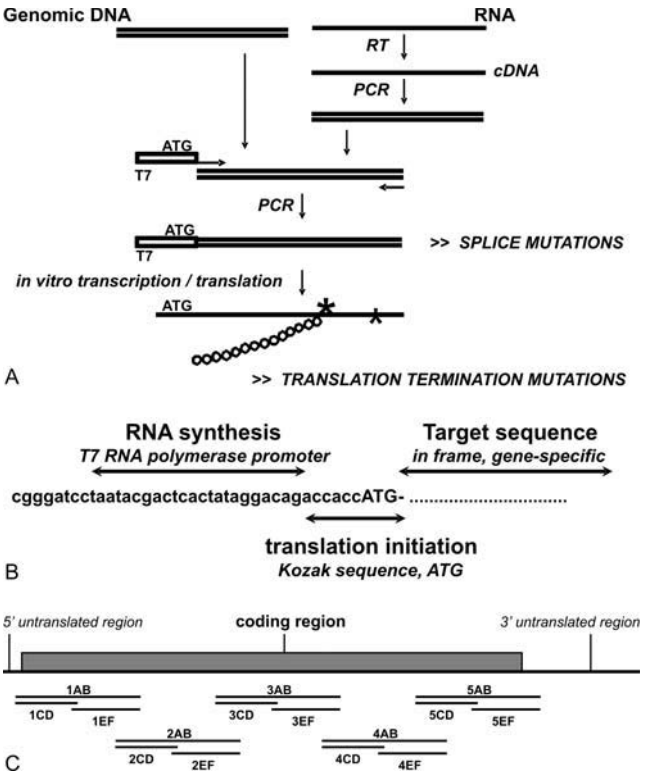


FIGURE 19.1 The Protein Truncation Test. **A.** PTT schematic. Either genomic DNA or RNA can be used as a template. During PCR a tailed primer is used that adds an RNA polymerase promoter (T7) and a translation initiation sequence (ATG). During *in vitro* transcription/translation the RNA polymerase promoter is used to initiate transcription, the translation initiation sequence is used to start protein translation. When genomic DNA is used, PTT starts directly with PCR amplification using tailed primers. RNA templates are first reverse transcribed (RT) to generate a DNA copy (cDNA) and then amplified using PCR. Mutations affecting RNA splicing are detected when PCR fragments are analyzed, e.g. using agarose gel-electrophoresis. Mutations affecting protein translation will be revealed upon size separation of the translation products. **B.** Tailed PTT primer containing an RNA polymerase promoter, a translation initiation sequence (Kozak and ATG) and a gene target sequence. **C.** Gene target sequence separated in five large (first round PCR) and ten smaller (second round PCR) properly overlapping segments.

was that nearly all were truncating, i.e. causing premature termination of protein translation (Monaco *et al*., 1988). This observation initiated the desire to develop a technology that would zoom in on these truncating variants only; the trigger to develop the Protein Truncation Test (Roest *et al*., 1993).

In the original PTT assay (Fig. 19.1a) RNA is isolated and copied into cDNA using RT-PCR (Fig. 19.2a). Subsequently, the cDNA is used to generate RNA (Fig. 19.2b), the RNA is translated into radioactively labeled protein using ^{3}H-Leucine incorporation, the proteins are size separated using gel electrophoresis, the gel is blotted and finally exposed to X-ray film. After exposure, shorter than normal-size proteins point to the presence of a premature translation terminating variants (Fig. 19.2c). One of the primers used to generate the RT-PCR products contains

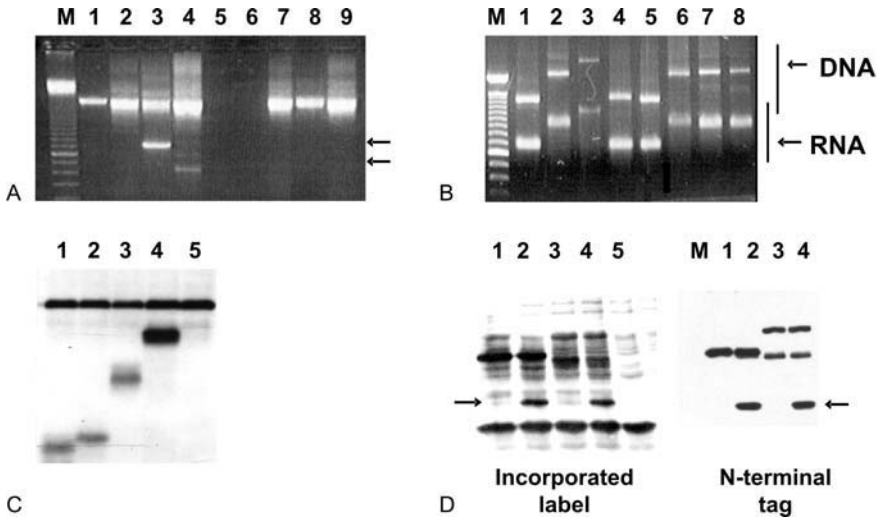


FIGURE 19.2 Results from PTT analysis. **A.** Agarose gel-electrophoresis of RNA-derived PTT-PCR fragments (M = marker, 1 = PCR control, 6 = no template control). Amplification of sample 5 failed, while samples 3 and 4 show shorter fragments derived from abnormally spliced RNA molecules (indicated by arrows). **B.** PCR fragments obtained using tailed PTT primers were *in vitro* transcribed and separated using agarose gel-electrophoresis. Note that the input (double-stranded) DNA fragments migrate slower than the derived (single-stranded) RNA molecules. **C.** PTT-result; ^{3}H-leucine-labeled *in vitro* transcription/translation products, separated using PAGE, blotted and analyzed using X-ray autoradiography. Note that lanes 1 to 4 contain, next to the normal product, shorter translation products indicating the presence of premature translation termination mutations. **D.** Comparison of label incorporation versus N-terminal tagging. Label incorporation was performed using botinylated-lysine, the N-terminal protein tag was detected using an antibody. Lanes 1 and 3 control, 2 and 4 mutant (arrow = truncation fragment). Note that label incorporation shows a background of many translation products that are not visible using N-tag detection; the background products thus probably derive from secondary translation initiation.

a 5′ tail sequence encoding both an RNA-polymerase promoter (usually T7) and a eukaryotic translation initiation (or Kozak) sequence lined up in frame with the open reading frame of the gene to be analyzed (Fig. 19.1b). After RT-PCR, when the resulting cDNA fragments are analyzed using gel electrophoresis, variants that affect RNA processing (i.e. deletions, duplications, and variants affecting splicing) will be detected because they produce fragments of altered size (Fig. 19.2a). Although originally developed for RNA analysis, PTT can also be performed on genomic DNA. PTT scanning of the *APC* (Van Der Luijt *et al*., 1994) and *BRCA1/2* (Hogervorst *et al*., 1995) genes, containing large exons and a significant fraction of pathogenic variants, has been effective and very popular.

19.4 METHODOLOGY

In general, the gene of interest will be too large to permit PTT analysis using one translated fragment. Consequently, the gene has to be split into partly overlapping 1 to 2 Kb segments together spanning the complete coding region (Fig. 19.1c). Although proteins can be translated efficiently from templates up to 4 Kb, it is often difficult to amplify such fragments unless specific precautions are taken (Whittock *et al*., 1997; Den Dunnen, 2004). Furthermore, optimal results to detect small mobility shifts with SDS-PAGE are obtained using fragments of 1 to 2 Kb. To prevent false negative results, the overlapping segments must be designed very carefully (Fig. 19.1c). Important rules include: (i) the primer-derived ATG translation initiation codon should be in frame with the coding sequence; (ii) flanking segments should contain sufficient overlap to ensure detection of variants close to their ends; (iii) to prevent a frame shift mutation resulting in translation termination, the reverse primer should not be selected near the end of a region where a large open reading frame is present in either of the shifted reading frames; and (iv) to minimize the risk of not amplifying an exon deletion allele, amplification of at least one segment should cover the entire exon.

For RNA isolation, usually a simple RNA-protocol suffices, e.g. based on RNA-sol. As RNA source, the affected tissue is always preferred, but in general is often very difficult to obtain; at best, a tiny amount of a needle biopsy is available. Blood is the most frequently used RNA source, while cultured cells (generally fibroblasts) provide another valuable source. Since expression is usually low, a nested PCR reaction will be required to amplify sufficient material to allow a transcription/translation reaction. During cell culturing, cylcoheximide treatment can be used to reduce the effect of NMD, equaling expression level of both alleles (Lamande *et al*., 1998; Den Dunnen, 2004). Other RNA sources that have been used successfully are buccal swaps and excrements (e.g. stool).

RT-PCR can be performed using random or specifically primed cDNA synthesis (Fig. 19.1a) and different labs have different preferences. To reduce the danger of contamination, a continuous threat in PCR-based diagnostics, assays

that include a one-tube RT-PCR reaction are preferred. *In vitro* transcription and translation can be performed using either a coupled or two separate reactions (Fig. 19.2b). Coupled transcription/translation is simple while performing separate reactions gives flexibility and advantages to sort out methodological problems when these occur. *In vitro* translation has been nearly exclusively performed using rabbit reticulocyte lysates. When translation of a fragment is problematic, i.e. low translation yields, a wheat germ lysate may provide a good alternative (Hope and Struhl, 1985).

Initially, detection of translation products could only be achieved using incorporation of radioactive-labeled amino acids (Powell *et al*., 1993; Roest *et al*., 1993). Later, biotin-labeled amino acids were released (Fig. 19.2d), facilitating detection with fluorescently labeled antibodies (Gite *et al*., 2003). Recently, fluorescently labeled amino acids became available facilitating in-gel detection of the translation products (Traverso *et al*., 2003). Since these labeled amino acids are available in four different colors and they obviate the need to blot the protein gel after electrophoretic separation, their use is very attractive and at the same time reduces the overall analysis time.

N- and C-terminal tagging of the translation products, i.e. the addition of a specific protein tag sequence encoded by the tailed primer(s) used during RT-PCR amplification has revolutionized the PTT assay (Fig. 19.2d). First, N-terminal tagging (Rowan and Bodmer, 1997) gives a significant reduction in background since no signals are obtained from secondary translation initiation products (Fig. 19.2d). Second, the signal obtained from the truncated and the normal allele are equal (label incorporation yields stronger signals for larger translation products). Third, detection, purification, and/or concentration of the translation products can be done in multiple ways. Fourth, C-terminal tags can be used to detect run-off translation products, i.e. those products that do not encounter a translational stop codon in the shifted reading frame (Kahmann *et al*., 2002). Finally, determining the ratio of signals obtained from antibodies directed against the N- and C-terminally tagged translation products can be used to detect truncating changes without electrophoretic separation. This gives attractive opportunities to increase throughput and facilitates automation of a PTT assay (Traverso *et al*., 2003).

19.5 ALTERNATIVE DETECTION SYSTEMS

Thus far, the translation products of the PTT have been analyzed mainly using size separation, mostly polyacrylamide gel electrophoresis. When the focus is other than the length of the product, i.e. aiming at the detection of truncating variants, alternative approaches can be used as well. These include protein band shift assays like 2D gels and isoelectric focusing (IEF), analyses that might reveal variants yielding amino acid substitutions when these affect protein structure/folding. Recently, the use of mass spectrometry (MALDI-TOF, see also Chapter 21) was reported, facilitating the high-throughput analysis of exon-sized translation products (Garvin *et al*., 2000). Notably, based on the molecular weight, MALDI-TOF analysis will not only reveal premature translation terminating variants but amino acid substitutions as well.

An alternative way to detect truncating variants is through expression cloning (Ishioka *et al*., 1997; Suzuki *et al*., 1998). One application focuses on the reading frame *per se*, i.e. detecting premature translation termination through protein fusion products, e.g. by simply scoring the fraction of blue/white clones obtained through gene fusion cloning in β-galactosidase vectors. Blue clones should contain fragments with an open reading frame, white clones inserts with translation terminating changes; an increased fraction of white clones thus pointing at the presence of a truncating mutation in the fragment analyzed. Another approach uses the generation of deleterious/toxic fusion products yielding clones only when they derive from transcripts containing protein truncating variants (Ishioka *et al*., 1997); increased clone numbers pointing to truncating variants, and sequence analysis revealing the deleterious variant.

Theoretically, *in vitro* transcription/translation (PTT) can be used to generate the full-length protein, in particular when analyzing smaller genes. For genes where a functional test of the protein is possible, an assay could then be designed that uses these translation products to directly test their function, e.g. their DNA binding properties or enzymatic activity. To our knowledge this approach has not yet been used in practice although variants containing an intermediate expression cloning step were reported. For example, to test the DNA binding properties of p53-derived PCR products, Ishioka *et al*. (1993) cloned these in yeast, and assayed them for transcriptional activity in a simple growth assay. Similarly, variants containing segments of RT-PCR products have been used in complementation cloning in yeast (PEX genes) and to follow protein trafficking upon overexpression in mammalian cells (Wang *et al*., 2006).

19.6 FALSE POSITIVES/FALSE NEGATIVES

A false positive PTT result, i.e. the detection of a truncating fragment that does not derive from a causative variant, only rarely occurs. The most prominent false positive derives from secondary translation initiation (Fig. 19.2d); these products can be easily detected since they will appear in all samples analyzed for a specific gene fragment. Another false positive might result from PCR errors introducing a truncating variant during amplification. Such an error is unlikely and will be detected by performing independent duplicate assays (i.e. from a second sample of the same patient or, when this is not available, starting with a

second RT reaction). Theoretically, another false positive is a truncating variant that does not cause disease. Although truncating but seemingly neutral variants in a heterozygous state have been reported (dopamine D5 receptor (Sobell *et al*., 1995)), they are rare. A unique "false positive" was discovered by Laken *et al*. (1997), a variant in the *APC* gene causing colon cancer predisposition through the creation of a hypermutable region. This variant, a nucleotide substitution at the DNA level, caused a range of deletions and insertions at the RNA level yielding a weak smear of truncated proteins that could only be detected after cloning and sequencing individual cDNA fragments.

False negatives, i.e. a truncating change is present but not detected, can occur due to several causes. In most cases they result from a failure of the amplification of the mutated allele, e.g. when using an RNA sample that is of insufficient quality to ensure reliable analysis. Another prominent cause is amplification of only one of the two alleles expressed in the sample analyzed. The most obvious causes behind amplification failure include the presence of a sequence variant (polymorphism) in one of the primers used, a genomic deletion or inversion encompassing the sequence analyzed and a prominent difference in the level of expression of the two alleles. The latter effect can be direct, e.g. as a consequence of non-random X-inactivation or variants affecting transcription levels, or indirect as a consequence of differences in RNA stability and/or the specific degradation of one of the two transcripts by processes like NMD (Dietz and Kendzior, 1994). False negatives may also derive from transcription/translation failure of either or both alleles, e.g. as a consequence of using a bad-quality translation kit or a low-quality forward primer. Regarding the latter, it is notable that forward PTT primers are rather long and hence if produced at low yields the 5′ end that contains the RNA polymerase promoter sequence is significantly affected (Den Dunnen, 2004).

Another cause of false negative results is when truncated proteins cannot be detected. This occurs when (a) the translation products migrate near the separation limit of the gel (both large and small), (b) background fragments obscure truncated fragments, (c) no label is incorporated, (d) the sequence change does not produce a stop codon in the shifted reading frame, and (e) when variants change the translation initiation (ATG) and termination codon. A proper and thoughtful design of the PTT assay is instrumental in preventing most of these false negative results, in particular ensuring sufficient overlap between flanking segments (see section 21.4). In addition, as for the analysis of DNA samples and especially when pathogenic changes are not detected, it is absolutely essential to assess whether transcripts/translation products of both alleles were analyzed. When a variant (single nucleotide polymorphism; SNP) in either of the two alleles is known, this quality check is rather simple to perform; when a variant is not known this check can be very difficult.

When truncating variants occur near the extreme ends of a protein (N- or C-terminal), the question is whether this has any deleterious effects. Early N-terminal truncations, including those affecting the ATG translation initiation codon, either produce no protein or cause secondary translation initiation. When no protein is produced the variant should be deleterious by reducing expression to 50%. Very short N-terminal proteins may theoretically have (dominant) negative effects but in practice are often very instable and nonfunctional. Sometimes, early truncations lead to secondary translation initiation and relatively mild phenotypes (Puel *et al*., 2006). Similarly, late C-terminal truncations produce a near normal protein and are unlikely to be disease causing (e.g. BRCA2, Mazoyer *et al*., 1996). When the translation stop codon is affected, proteins get an extended C-terminal tail which in most cases has deleterious consequences.

19.7 APPLICATIONS

The PTT assay became a big success, although less for the disease it was originally designed for, i.e. Duchenne and Becker muscular dystrophy (DMD/BMD). In blood, expression of the DMD gene is extremely low, making reliable amplification cumbersome. Consequently, samples from a muscle biopsy are preferred, but more difficult to obtain routinely. For males, variant detection is sensitive enough, but for female carrier detection the presence of the other allele that is normally expressed often causes detection problems. Not only is expression of the mutated allele often reduced by NMD, it can also be affected by non-random X-inactivation. For other genes, in particular tumor suppressor genes, PTT has become an attractive and very reliable methodology to reveal pathogenic changes, especially when it can be performed using a DNA template. The latter is the case when a gene of interest contains a large exon in which many disease-causing variants are found, like in the *APC*, *BRCA1*, and *BRCA2* genes (Van Der Luijt *et al*., 1994; Hogervorst *et al*., 1995).

Potential target genes for PTT include: tumor suppressor genes (colon cancer (*FAP* and *HNPCC*), breast cancer, neurofibromatosis, *ATM*, *TSC*, etc.), genes involved in diseases mainly caused by truncating variants (muscular dystrophy (B/DMD, EMD), polycystic kidney disease and many others), and genes involved in diseases with a complex variant spectrum of which a significant fraction is caused by truncating changes (cystic fibrosis (non-p.F508del variants), collagen genes, Hunter syndrome and OTC-deficiency). The flexibility and dynamics of the PTT technique is elegantly demonstrated by the work of Traverso *et al*. (2003). These authors developed the PTT into a sensitive, semi-automated, high-throughput assay that was applied in various situations, including the early detection of rare truncating variants in stool samples of potential colon cancer patients.

19.8 CONCLUSIONS

Although technically demanding and laborious, PTT has several attractive characteristics; unlike other techniques it pinpoints the position of a variant, it has an excellent sensitivity and a low false positive rate, and, most importantly, it nearly exclusively highlights pathogenic variants. In addition, since it is an RNA-based technique, PTT facilitates the detection of variants that affect RNA processing. The main technical hurdles experienced with PTT relate to the fact that it uses an RNA target, including the difficulties arising from the potential differential allelic expression and stability of the transcripts derived from the two alleles present in diploid organisms.

REFERENCES

Den Dunnen, J.T. (2004). Protein truncation test, current protocols in human genetics. (Dracopoli, N.C., Haines, J.L., Korf, B.R., Moir, D.T., Morton, C.C., Seidman, C.E., Seidman, J.G., Smith, D.R., eds), John Wiley & Sons, Inc, New York, pp. 9.11.1–9.11.18.

Den Dunnen, J.T., and Beggs, A.H. (2006). Multiplex PCR for identifying dystrophin gene deletions. *Curr. Protocols Hum. Genet.*, 9.3.1–9.3.17. Unit 9.3.

Den Dunnen, J.T., Grootscholten, P.M., Bakker, E., Blonden, L.A.J., Ginjaar, H.B., Wapenaar, M.C., Van Paassen, H.M.B., Van Broeckhoven, C., Pearson, P.L., and Van Ommen, G.J.B. (1989). Topography of the DMD gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. *Am. J. Hum. Genet.* 45, 835–847.

Dietz, H.C., and Kendzior, R.J., Jr. (1994). Maintenance of an open reading frame as an additional level of scrutiny during splice site selection. *Nat. Genet.* 8, 183–188.

Garvin, A.M., Parker, K.C., and Haff, L. (2000). MALDI-TOF based mutation detection using tagged in vitro synthesized peptides. *Nat. Biotechnol.* 18, 95–97.

Gite, S., Lim, M., Carlson, R., Olejnik, J., Zehnbauer, B., and Rothschild, K. (2003). A high-throughput nonisotopic protein truncation test. *Nat. Biotechnol.* 21, 194–197.

Hogervorst, F.B.L., Cornelis, R.S., Bout, M., Van Vliet, M., Oosterwijk, J.C., Olmer, R., Bakker, E., Klijn, J.G.M., Vasen, H.F.A., Meijers-Heijboer, H., Menko, F.H., Cornelisse, C.J., Den Dunnen, J. T., Devilee, P., and Van Ommen, G.J.B. (1995). Rapid detection of BRCA1 mutations by the protein truncation test. *Nat. Genet.* 10, 208–212.

Hope, I.A., and Struhl, K. (1985). GCN4 protein synthesized in vitro, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast. *Cell* 43, 177–188.

Ishioka, C., Frebourg, T., Yan, Y.X., Vidal, M., Friend, S.H., Schmidt, S., and Iggo, R. (1993). Screening patients for heterozygous p53 mutations using a functional assay in yeast. *Nat. Genet.* 5, 124–129.

Ishioka, C., Suzuki, T., FitzGerald, M., Krainer, M., Shimodaitra, H., Shimada, A., Nomizu, T., Isselbacher, K.J., Haber, D., and Kanamaru, R. (1997). Detection of heterozygous truncating mutations in the BRCA1 and APC genes by using a rapid screening assay in yeast. *Proc. Natl. Acad. Sci. USA* 94, 2449–2453.

Kahmann, S., Herter, P., Kuhnen, C., Muller, K.M., Muhr, G., Martin, D., Soddemann, M., and Muller, O. (2002). A non-radioactive protein truncation test for the sensitive detection of all stop and frameshift mutations. *Hum. Mutat.* 19, 165–172.

Laken, S.J., Petersen, G.M., Gruber, S.B., Oddoux, C., Ostrer, H., Giardiello, F.M., Hamilton, S.R., Hampel, H., Markowitz, A., Klimstra, D., Jhanwar, S., Winawer, S., Offit, K., Luce, M.C., Kinzler, K.W., and Vogelstein, B. (1997). Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat. Genet.* 17, 79–83.

Lamande, S.R., Bateman, J.F., Hutchison, W., McKinlay Gardner, R.J., Bower, S.P., Byrne, E., and Dahl, H.H. (1998). Reduced collagen VI causes Bethlem myopathy: a heterozygous COL6A1 nonsense mutation results in mRNA decay and functional haploinsufficiency. *Hum. Mol. Genet.* 7, 981–989.

Mazoyer, S., Dunning, A.M., Serova, O., Dearden, J., Puget, N., Healey, C.S., Gayther, S.A., Mangion, J., Stratton, M.R., Lynch, H.T., Goldgar, D.E., Ponder, B.A., and Lenoir, G.M. (1996). A polymorphic stop codon in BRCA2. *Nat. Genet.* 14, 253–254.

Monaco, A.P., Bertelson, C.J., Liechti-Gallati, S., Moser, H., and Kunkel, L.M. (1988). An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics* 2, 90–95.

Powell, S.M., Petersen, G.M., Krush, A.A.J., Booker, S., Jen, J., Giardello, F.M., Hamilton, S.R., Vogelstein, B., and Kinzler, K.W. (1993). Molecular diagnosis of Familial Adenomatous Polyposis. *New Engl. J. Med.* 329, 1982–1987.

Puel, A., Reichenbach, J., Bustamante, J., Ku, C.L., Feinberg, J., Doffinger, R., Bonnet, M., Filipe-Santos, O., Beaucoudrey, L., Durandy, A., Horneff, G., Novelli, F., Wahn, V., Smahi, A., Israel, A., Niehues, T., and Casanova, J.L. (2006). The NEMO mutation creating the most-upstream premature stop codon is hypomorphic because of a reinitiation of translation. *Am. J. Hum. Genet.* 78, 691–701.

Roest, P.A.M., Roberts, R.G., Sugino, S., Van Ommen, G.J.B., and Den Dunnen, J.T. (1993). Protein truncation test (PTT) for rapid detection of translation-terminating mutations. *Hum. Mol. Genet.* 2, 1719–1721.

Rowan, A.J., and Bodmer, W.F. (1997). Introduction of a myc reporter tag to improve the quality of mutation detection using the protein truncation test. *Hum. Mutat.* 9, 172–176.

Sobell, J.L., Lind, T.J., Sigurdson, D.C., Zald, D.H., Snitz, B.E., Grove, W.M., Heston, L.L., and Sommer, S.S. (1995). The D5 dopamine receptor gene in schizophrenia: identification of a nonsense change and multiple missense changes but lack of association with disease. *Hum. Mol. Genet.* 4, 507–514.

Suzuki, T., Ishioka, C., Kato, S., Mitachi, Y., Shimodaira, H., Sakayori, M., Shimada, A., Asamura, M., and Kanamaru, R. (1998). Detection of APC mutations by a yeast-based protein truncation test (YPTT). *Gene. Chromosome. Canc.* 21, 290–297.

Traverso, G., Diehl, F., Hurst, R., Shuber, A., Whitney, D., Johnson, C., Levin, B., Kinzler, K.W., and Vogelstein, B. (2003). Multicolor in vitro translation. *Nat. Biotechnol.* 21, 1093–1097.

Van Der Luijt, R.B., Meera Kahn, P., Vasen, H., Van Leeuwen, C., Tops, C., Roest, P.A.M., Den Dunnen, J.T., and Fodde, R. (1994). Rapid detection of translation-terminating mutations at the Adenomatous Polyposis Coli (APC) gene by direct Protein Truncation Test. *Genomics* 20, 1–4.

Wang, T., Zhou, A., Waters, C.T., O'Connor, E., Read, R.J., and Trump, D. (2006). Molecular pathology of X linked retinoschisis: mutations interfere with retinoschisin secretion and oligomerisation. *Br. J. Ophthalmol.* 90, 81–86.

Whittock, N.V., Roberts, R.G., Mathew, C.G., and Abbs, S.J. (1997). Dystrophin point mutation screening using a multiplexed Protein Truncation Test. *Genet. Test.* 1, 115–123.

Chapter 20

Protein Diagnostics by Proximity Ligation: Combining Multiple Recognition and DNA Amplification for Improved Protein Analyses

Karl-Johan Leuchowius, Ola Söderberg, Masood Kamali-Moghaddam, Malin Jarvius, Irene Weibrecht, Katerina Pardali and Ulf Landegren

Department of Genetics and Pathology, Uppsala University, Rudbeck Laboratory SE-751 85, Uppsala, Sweden

20.1 INTRODUCTION

The ability to investigate large numbers of proteins with excellent precision in extensive series of patient samples, first in research and then clinically, will greatly impact healthcare through improved opportunities for molecular diagnostics. Having the total human genome sequence information in our hands, we now face exciting opportunities for analyses of all the protein products that are encoded in our genomes. Efficient protein assays will yield improved functional understanding of physiological and pathological processes, and measurements of the amounts of specific proteins and their modifications and interactions can provide new approaches to molecular diagnostics. The field of protein diagnostics is attracting great interest because of its potential to allow early diagnosis, serving to reveal subclinical, still treatable diseases. It can also help selection of the optimal treatment and monitoring disease progression, thus offering important benefits in healthcare. Sufficiently efficient and robust methods will permit such analyses to be performed at the point-of-care, or even in households, to diagnose infections and demonstrate signs of disease. It is clear, however, that methods that have been available for protein detection until recently fail to permit the parallel and highly resolving analyses that will now be needed for extensive analyses of proteins, both in research and clinically. In addition, high-throughput techniques for proteome analyses are required in order to take advantage of the large collections of human samples that are being organized in the form of extensive biobanks (Yuille *et al*., 2008). This represents a valuable resource for the identification of diagnostic protein markers, provided appropriate methods become available.

Proteome analyses are complicated by the enormous concentration ranges of protein classes in samples, such as serum, and even within individual cells, placing stringent demands on the ability to distinguish proteins that may differ by many orders of magnitude in concentration. This selectivity problem is further compounded by the often closely similar but functionally different variants of proteins that arise due to mechanisms like differential splicing, processing, and post-translational modification (PTM). Moreover, protein function is manifested not only by changes of levels of proteins, but also by the rearrangement of proteins in complexes, necessitating measurement of interactions among sets of proteins, as well as of individual protein molecules.

While mass spectrometry has emerged as a powerful research tool for mapping the protein composition in biological samples (see also next chapter), assays based on reagents with affinity for specific proteins remain the most promising avenue for protein-based diagnostics. However, this approach is complicated by the difficulty of generating suitable affinity reagents. This is in contrast to the situation for nucleic acid analyses, where the simple rules directing base pairing can be employed to design hybridization probes, and the combination of such hybridization probes with nucleic acid-specific enzymes serves to improve the selectivity and sensitivity of detection, i.e. via amplification reactions. It is gratifying to note that several initiatives are under way aiming to raise comprehensive sets of affinity reagents against large subsets of, and ultimately all,

human proteins (Taussig *et al*., 2007). It is unlikely, however, that completely monospecific affinity reagents can be developed against all human proteins, given the risk for cross-reactivity for antigenic epitopes in other, perhaps more abundant, proteins. Moreover, individual binders can only assess epitopes that represent a small part of a protein, while multiple diagnostic determinants may be distributed in remote locations on a target protein or protein complex. Accordingly, test architectures must be developed that require simultaneous binding to multiple epitopes on target molecules for enhanced selectivity and reduced risks of cross-reactivity.

In this chapter, we will provide a brief description of current methods for protein analysis, as a background to a description of a general molecular strategy for protein analysis, the so-called proximity ligation assays; we also discuss future developments of this technique.

20.2 BINDING THE PROTEOME

Binding reagents are of central importance in a majority of methodologies used to study proteins. The complexity of protein binding currently precludes prediction of suitable binding reagents, in contrast to the situation for DNA binding by hybridization probes. Protein binding reagents are therefore typically obtained by processes of *in vivo* or *in vitro* selection from large repertoires of molecules with constant and variable sequence elements. Depending on how reagents are generated and selected, they may preferentially recognize linear or context-dependent determinants. As a consequence, they may differ in their suitability for assays where proteins are mostly in a native configuration, such as in analyses of tissue lysates or serum protein in solution, or ones where proteins have been fixed and often denatured, such as in immunohistochemistry and Western blots.

Polyclonal and monoclonal antibodies are the most commonly used source of protein binders. The diversity of antibodies generated by rearrangements of gene segments, in combination with somatic hypermutations or gene conversion, provides a wide repertoire of antibodies for selection of high-affinity binders against any target protein. Besides *in vivo* immunizations, antibody fragments can be selected from large libraries in, e.g., bacteriophages, or expressed by ribosome or RNA display. These *in vitro* selection techniques have also permitted the development of a growing list of so-called scaffold proteins, with variable sequences that can form protein-binding clefts and surfaces (Nygren and Uhlen, 1997; Binz *et al*., 2005). In addition, DNA or RNA aptamers can be selected from nucleic acid libraries to bind proteins with high affinity (Tuerk and Gold, 1990; Bock *et al*., 1992). Moreover, there is a large repertoire of natural ligands that can be used to bind receptors, and lectins can be employed to detect protein glycosylation.

A number of projects have been initiated to establish comprehensive repertoires of binding agents against large numbers of, and eventually all, human proteins. Examples of such initiatives include the Human Proteome Atlas (http://www.proteinatlas.org), raising polyclonal affinity-purified antibodies against human proteins, the National Cancer Institute preparing monoclonal antibodies against proteins of special interest in malignancy, and ProteomeBinders (http://www.proteomebinders.org), coordinating activities among a large number of mainly European groups, developing protein binding reagents. It is therefore reasonable to assume that reliable clonal and replenishable binding reagents that can be widely shared among labs will be increasingly available for protein detection in years to come. Accordingly, assays must now be established that allow such reagents to be applied for extensive analyses of the proteome.

20.3 CURRENT AFFINITY-BASED PROTEIN DETECTION ASSAYS

Early protein detection assays depended on target binding by single antibodies, such as the radio-immunoassays developed in the late 1950s (Yalow and Berson, 1960). Sandwich immune assays, first published almost ten years later, improved detection selectivity, sensitivity, and convenience by requiring binding by two antibodies, one immobilized and another detectable one added in solution after binding of antigen (Wide *et al*., 1967; Engvall and Perlman, 1971). In a different manner, Western blot evaluates proteins both with respect to migration during gel electrophoresis and binding by an antibody, thereby improving selectivity over single-binder assays (Towbin *et al*., 1979). The immobilization of different antibodies in array format can permit simultaneous analysis of several labeled proteins in a sample. The selectivity of detection in such array-based assays can be improved by using a sandwich format where secondary, labeled antibodies are added in solution. The rapidly increasing risks of cross-reactive binding among non-cognate antibody pairs tend to limit such multiplex assays to a few tens of analytes, however (Fig. 20.1) (Nielsen and Geierstanger, 2004). Imaging protein assays are also of interest by demonstrating the distribution of target proteins in cells and tissues. Such assays can be performed by staining specific proteins with the help of labeled antibodies that are added to tissue sections in immunohistochemical reactions.

To enable detection of bound antibodies in these various assays, the antibodies can be modified by covalently attaching detectable moieties such as fluorophores, heavy metals, radioisotopes, or enzymes. Fluorescent dyes generate a signal proportional to the amount of bound antibodies, and the signal can be enhanced by increasing the amount of fluorophores present per antibody. However, for low-abundance proteins or in cases of high fluorescent

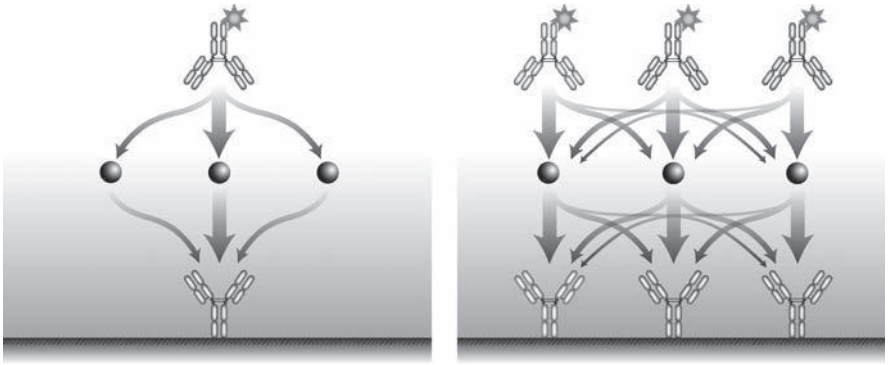


FIGURE 20.1 The risk of cross-reactive binding among non-cognate antibody pairs increases when the number of proteins detected simultaneously in a sandwich immunoassay is increased. **A.** When proteins are detected in singleplex, the risk of cross-reactive binding is low since the target proteins must be recognized by both the capture antibody and the detection antibody for a positive identification. **B.** In the case of multiplex detection, several types of target proteins are captured simultaneously and can therefore give rise to false positive signals if the detection antibody binds non-specifically to other captured proteins.

background, signal amplification may be required for detection. Reactions involving antibody-conjugated enzymes result in signal amplification when substrate molecules are converted into a detectable product. The strength of the signal increases over time, potentially allowing detection of lower concentrations of antigens, although often also increasing non-specific signals in parallel and thus limiting any improvement of signal over background.

The advantages of polymerase-based DNA amplification have been exploited in protein detection assays by conjugating DNA molecules to antibodies. Thereby, immuno-PCR can be used to exponentially amplify detection signals from bound antibodies with attached DNA strands by using PCR (Sano *et al*., 1992). The mechanism lends itself to multiplex assays where each antibody is conjugated to a unique DNA oligonucleotide that can be amplified by PCR and identified. The amount of each PCR product correlates to the amount of antibody bound to an antigen. For localized detection in, e.g., tissue sections, another method of DNA amplification, rolling circle amplification (RCA), can be used. A very processive DNA polymerase such as the bacteriophage phi 29 DNA polymerase can generate a long single-stranded concatemer of complements of a circular DNA strand used to template the reaction. The DNA molecule that forms can reach tens of micrometers in length, but it is typically rolled up in a ball of DNA of submicrometer dimensions. The RCA technique has been used in immuno-RCA (Schweitzer *et al*., 2000; Wiltshire *et al*., 2000; Gusev *et al*., 2001), where oligonucleotides are conjugated to antibodies and used to prime RCA reactions of subsequently added circular DNA strand. By hybridizing oligonucleotides labeled with fluorophores to the repeated sequences in the rolling circle products, the bundles of DNA can be easily visualized by fluorescence microscopy as brightly fluorescent spots, and the location of the binding event can be deduced. A main consideration for all kinds of binders is their selectivity for a target molecule. For positive identification of a protein, the binder should ideally interact with the correct protein with high affinity, and it should exhibit minimal cross-reactivity to other proteins. As discussed above, the selectivity of the detection reaction can be increased by adding a second criterion, such as binding by a second antibody or by distinguishing target proteins according to molecular weight or isoelectric charge – as used in sandwich immune assays, Western blots, and after 2D-gel electrophoresis, respectively. In a related manner, proximity ligation reactions depend on proximal binding by two or more proximity probes to increase the selectivity, while the sensitivity of the assays are ensured using DNA amplification.

20.4 PROXIMITY LIGATION ASSAYS FOR DETECTING PROTEINS IN SOLUTION

In proximity ligation assays (PLA), single-stranded DNA oligonucleotides are conjugated to antibodies, and the conjugates are referred to as proximity probes. One antibody is coupled to an oligonucleotide via the 3′-end and another antibody is coupled to a second oligonucleotide via the 5′-end. When two proximity probes with different DNA strands bind the same target protein or protein complex, a subsequently added connector oligonucleotide can hybridize to the ends of the conjugated DNA strands and guide their joining by enzymatic ligation. This creates a DNA molecule that can then be amplified and quantified by, for instance, real-time PCR, whereas remaining non-ligated DNA strands cannot be amplified. By taking measures to ensure that minimal ligation products form in the absence of target, sub-pM levels of proteins can be detected by signal amplification through real-time PCR in a homogeneous assay (Fredriksson *et al*., 2002; Gullberg *et al*., 2004)

(Fig. 20.2). In a solid-phase variant of PLA, the target molecules to be detected are first captured by antibodies immobilized on a surface – as in sandwich ELISA – while other components are removed by washes. Thereafter, a pair of proximity probes is added and after renewed washes the ligation reaction is performed, followed by amplification. The method provides further increased sensitivity of protein detection, and even single infectious units of microbial pathogens can be demonstrated using this solid-phase format of PLA (Gustafsdottir *et al*., 2006). Generally, this assay format has particular advantages for detecting rare molecules in dilute samples or in samples that contain components that can interfere with the detection reaction, while the benefit of the homogeneous variant of PLA is the requirement for very small amounts of sample, and the possibility for parallel detection in a simple procedure with no need for washes (Fredriksson *et al*., 2007).

In another variant of PLA, 3PLA, the proximity probes are designed to require recognition by three proximity probes and two ligation events to form a complete PCR template. The requirement for simultaneous binding by several affinity binders decreases background due to

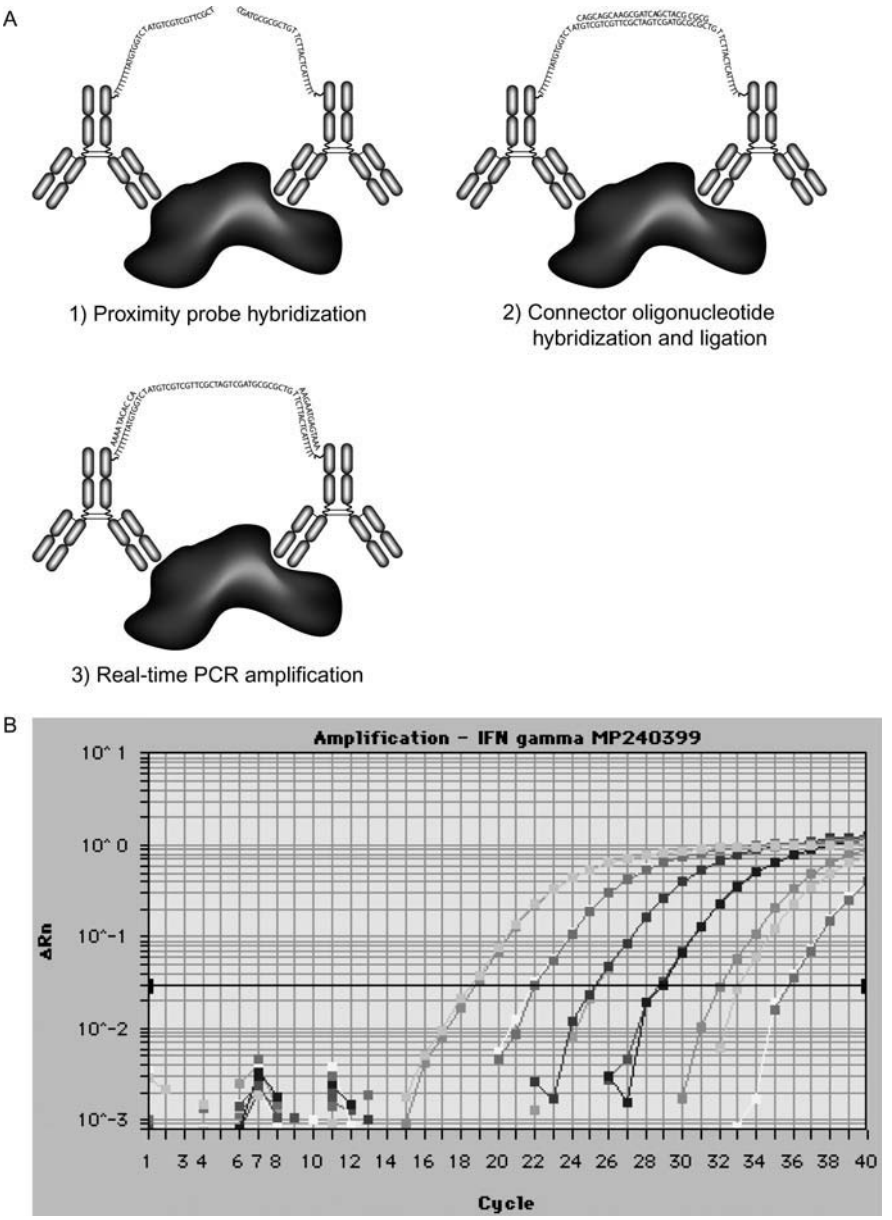


FIGURE 20.2 A. Schematic description of a homogeneous proximity ligation assay. (1) The target protein is recognized by the affinity-binding parts of two different proximity ligation probes, which brings the DNA strands of the probes in proximity. (2) Next, a connector oligonucleotide is hybridized to the ends of the DNA strands allowing these to be joined by enzymatic ligation. (3) The newly formed DNA template is then used for PCR with real-time detection of the accumulating amplification products. **B.** An illustration of a typical PLA-based real-time PCR result that can be used for quantification of the target protein, here, tumor necrosis factor α (TNFα).

target-independent ligation or cross-reactivity for related target molecules, and hence even lower concentrations of the correct target protein can be detected over background signals. Furthermore, assays that depend on three binding events can be used for detection of complex protein interactions or modifications involving three target epitopes (Schallmeiner *et al*., 2007). Recently, PLA has been successfully combined with array readout for parallel protein detection in a small-scale experiment, and the method is also suitable for identification of pairs of interacting proteins among a large set of proteins (Ericsson *et al*., 2008). The proximity ligation technique has also been adapted to examine interactions between proteins and specific DNA sequences (Gustafsdottir *et al*., 2007), illustrating the versatility of this new detection mechanism.

20.5 *IN SITU* PROXIMITY LIGATION ASSAYS

Information about the subcellular location of protein and protein complexes, and about their cell-to-cell variations within a tissue, can provide important insights into the functional status of cells. PCR is not suitable as an amplification method for localized detection in cells or tissues, since the PCR products tend to diffuse freely. By instead combining a PLA with RCA, a prominent localized detection signal can be obtained at the site of specific recognition of individual target molecules, and with minimal background. For this purpose, samples are first incubated with pairs of proximity probes. If these probes bind in proximity, then their attached DNA strands can serve as templates in a ligation reaction joining the ends of two oligonucleotides that are included in the ligation reaction. This gives rise to DNA circles that can next be replicated by RCA, using one of the antibody-bound oligonucleotides as a primer. The localized amplification products are then visualized using fluorescently labeled oligonucleotides that hybridize to the resulting RCA product (Soderberg *et al*., 2006) (Fig. 20.3). The intensity of the brightly fluorescent spots, representing individual RCA products, reduces problems with background fluorescence. The individual fluorescent signals can be quantified digitally with the help of a freely available software package ''BlobFinder'' developed by Allalou and Wählby (http://www.cb.uu.se/~amin/BlobFinder) for objective evaluation of the results.

The requirement for dual recognition by the proximity probes greatly improves the selectivity of the method compared to immuno-RCA, as individual detection probes and other nucleic acids fail to give rise to circular DNA strands and, as a consequence, no RCA products form unless dual recognition is achieved. The two proximity probes can be selected to bind to the same molecule for increased selectivity of detection, but they can also target two different proteins in a complex. In the standard format of the assay, determinants located within several tens of nanometers of each other can give rise to detection signals. Such an approach was used to study the heterodimerization between c-Myc and its binding partner Max (Soderberg *et al*., 2006; Vermeer *et al*., 2008), and between inositol 1,4,5-triphosphate receptors and protein kinase B/Akt (Szado *et al*., 2008). In analogy to what has been described for homogeneous PLA, *in situ* PLA assays can be designed to require more than two recognition events for the formation of an amplifiable ligation product. The approach can be used to visualize multicomponent interactions, as exemplified by the visualization of c-Myc/Max transcription factor complexes located in proximity of the transcribing RNA polymerase II (Soderberg *et al*., 2006).

In situ PLA has also been used to study post-translational modifications by using one antibody directed against platelet-derived growth factor receptor beta (PDGFRβ), and one antibody binding the modification, in this case a phosphorylated tyrosine residue of the ligand stimulated receptor (Jarvius *et al*., 2007). In these experiments, the samples were first incubated with a polyclonal rabbit antiserum against the receptor and a mouse monoclonal antibody directed against the phosphorylated residue. Next, two oligonucleotide-conjugated antibodies directed against rabbit and mouse immunoglobulin, respectively, were added to perform an *in situ* PLA. With this approach, standard oligonucleotide-conjugated secondary antibodies can be used for *in situ* PLA as long as the pair of primary antibodies is derived from two different animal species. This eliminates the need for constructing antibody-oligonucleotide conjugates for all primary antibodies, thereby saving time, costs, and effort.

20.6 CONCLUSION AND FUTURE PERSPECTIVES

As discussed herein, opportunities for protein diagnostics are excellent. Increasingly, the development of new therapies will be accompanied by analyses of molecular surrogate markers, both nucleic acids and proteins, to evaluate clinical responses. Some of these assays will likely also follow the drugs into clinical routine use in the form of theranostic approaches. Molecular diagnostics can also be expected to become more broadly accessible through the development of simple new detection formats that vastly increase the scope for point-of-care diagnostics. Development of reagents and test architectures proceed along several lines. We believe that the combination of a wide variety of protein-binding reagents with information-carrying, amplifiable DNA strands could serve a central element of many new diagnostic approaches. The mechanism provides valuable opportunities for increasing both the selectivity and sensitivity of detection. In contrast to more standard approaches, multiple proteins can potentially be investigated in parallel without

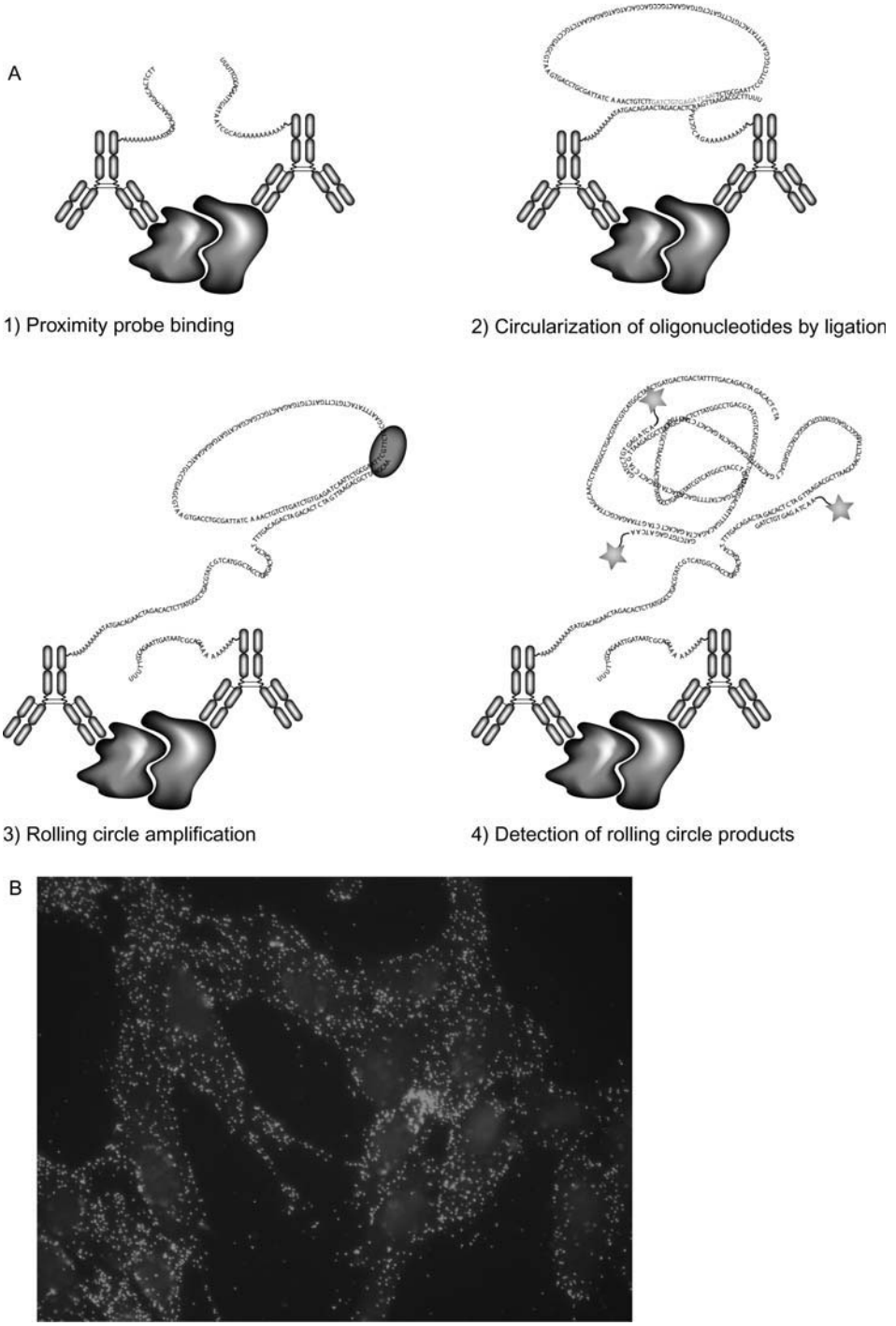


FIGURE 20.3 A. A schematic presentation of *in situ* PLA. Dual binding by a pair of proximity probes (antibodies with attached DNA strands) serve to template the formation of circular DNA molecules by ligation, provided that the antibodies have bound in sufficient proximity. The circular DNA molecules that form are then amplified by rolling-circle amplification (RCA) primed by one of the proximity probes. The localized amplification reaction gives rise to a concatemeric amplification product covalently attached to one of the proximity probes, and the RCA product can subsequently be detected by hybridization of fluorescence-labeled oligonucleotides. **B.** *In situ* PLA using secondary species-specific proximity probes in combination with a pair of primary antibodies, directed against platelet-derived growth factor receptor β and one of its phosphorylated tyrosine residues, respectively, was performed to visualize the phosphorylated receptor. Phosphorylated receptors are visualized as red dots in cells that have been counterstained with antibodies directed against actin (green) and with the DNA binding Hoechst dye (blue) to visualize the cytoplasm and the nucleus, respectively (Jarvius *et al.*, 2007). *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

rapidly growing problems with cross-reactivity by ensuring that only cognate pairs of proximity probes give rise to detection signals. Proximity ligation lends itself both for assays where protein levels are measured, and where their distribution is visualized in tissues, illustrating functional effects of signal transduction, developmental processes, as well as effects of disease and of therapeutic intervention.

There are exciting opportunities for further development of the technique. At this point it is not clear what level of multiplexing of protein measurement can be achieved by proximity ligation but the assay is compatible with a wide range of readout formats, including separate real-time amplification reactions (Fredriksson *et al*., 2007), or novel microarray detection mechanisms (Ericsson *et al*., 2008), and it should also be suitable for digital recording of individual proximity probe ligation products using new parallel DNA sequencing techniques (Schuster, 2008). Also the *in situ* forms of the assay can likely be adapted to permit simultaneous analysis of numerous proteins and their different interactions in the same sample for a better understanding of cellular responses. The diagnostic value of measuring multiple protein interactions and post-translational protein modifications currently is largely unexplored, but this must be viewed as one of many existing opportunities that currently emerge.

ACKNOWLEDGEMENTS

The work in our laboratory is supported by grants from the Wallenberg Foundation, by the EU Integrated Project ''ENLIGHT'', Cancerfonden, and by the Research Councils of Sweden for natural science and medicine.

REFERENCES

Binz, H.K., Amstutz, P., and Pluckthun, A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* 23, 1257–1268.

Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H., and Toole, J.J. (1992). Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* 355, 564–566.

Engvall, E., and Perlman, P. (1971). Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* 8, 871–874.

Ericsson, O., Jarvius, J., Schallmeiner, E., Howell, M., Nong, R.Y., Reuter, H., Hahn, M., Stenberg, J., Nilsson, M., and Landegren, U. (2008). A dual-tag microarray platform for high-performance nucleic acid and protein analyses. *Nucleic. Acids. Res.* 36, e45.

Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gustafsdottir, S.M., Ostman, A., and Landegren, U. (2002). Protein detection using proximity-dependent DNA ligation assays. *Nat. Biotechnol.* 20, 473–477.

Fredriksson, S., Dixon, W., Ji, H., Koong, A.C., Mindrinos, M., and Davis, R.W. (2007). Multiplexed protein detection by proximity ligation for cancer biomarker validation. *Nat. Methods* 4, 327–329.

Gullberg, M., Gustafsdottir, S.M., Schallmeiner, E., Jarvius, J., Bjarnegard, M., Betsholtz, C., Landegren, U., and Fredriksson, S. (2004). Cytokine detection by antibody-based proximity ligation. *Proc. Natl. Acad. Sci. USA* 101, 8420–8424.

Gusev, Y., Sparkowski, J., Raghunathan, A., Ferguson, H., Jr., Montano, J., Bogdan, N., Schweitzer, B., Wiltshire, S., Kingsmore, S.F., Maltzman, W., and Wheeler, V. (2001). Rolling circle amplification: a new approach to increase sensitivity for immunohistochemistry and flow cytometry. *Am. J. Pathol.* 159, 63–69.

Gustafsdottir, S.M., Nordengrahn, A., Fredriksson, S., Wallgren, P., Rivera, E., Schallmeiner, E., Merza, M., and Landegren, U. (2006). Detection of individual microbial pathogens by proximity ligation. *Clin. Chem.* 52, 1152–1160.

Gustafsdottir, S.M., Schlingemann, J., Rada-Iglesias, A., Schallmeiner, E., Kamali-Moghaddam, M., Wadelius, C., and Landegren, U. (2007). In vitro analysis of DNA-protein interactions by proximity ligation. *Proc. Natl. Acad. Sci. USA* 104, 3067–3072.

Jarvius, M., Paulsson, J., Weibrecht, I., Leuchowius, K.J., Andersson, A.C., Wahlby, C., Gullberg, M., Botling, J., Sjoblom, T., Markova, B., Ostman, A., Landegren, U., and Soderberg, O. (2007). In situ detection of phosphorylated platelet-derived growth factor receptor beta using a generalized proximity ligation method. *Mol. Cell. Proteomics* 6, 1500–1509.

Nielsen, U.B., and Geierstanger, B.H. (2004). Multiplexed sandwich assays in microarray format. *J. Immunol. Methods* 290, 107–120.

Nygren, P.A., and Uhlen, M. (1997). Scaffolds for engineering novel binding sites in proteins. *Curr. Opin. Struct. Biol.* 7, 463–469.

Sano, T., Smith, C.L., and Cantor, C.R. (1992). Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science* 258, 120–122.

Schallmeiner, E., Oksanen, E., Ericsson, O., Spangberg, L., Eriksson, S., Stenman, U.H., Pettersson, K., and Landegren, U. (2007). Sensitive protein detection via triple-binder proximity ligation assays. *Nat. Methods* 4, 135–137.

Schuster, S.C. (2008). Next-generation sequencing transforms today's biology. *Nat. Methods* 5, 16–18.

Schweitzer, B., Wiltshire, S., Lambert, J., O'Malley, S., Kukanskis, K., Zhu, Z., Kingsmore, S.F., Lizardi, P.M., and Ward, D.C. (2000). Inaugural article: immunoassays with rolling circle DNA amplification: a versatile platform for ultrasensitive antigen detection. *Proc. Natl. Acad. Sci. USA* 97, 10113–10119.

Soderberg, O., Gullberg, M., Jarvius, M., Ridderstrale, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G., and Landegren, U. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* 3, 995–1000.

Szado, T., Vanderheyden, V., Parys, J.B., De Smedt, H., Rietdorf, K., Kotelevets, L., Chastre, E., Khan, F., Landegren, U., Soderberg, O., Bootman, M.D., and Roderick, H.L. (2008). Phosphorylation of inositol 1,4,5-trisphosphate receptors by protein kinase B/Akt inhibits Ca2+ release and apoptosis. *Proc. Natl. Acad. Sci. USA* 105, 2427–2432.

Taussig, M.J., Stoevesandt, O., Borrebaeck, C.A., Bradbury, A.R., Cahill, D., Cambillau, C., de Daruvar, A., Dubel, S., Eichler, J., Frank, R., Gibson, T.J., Gloriam, D., Gold, L., Herberg, F.W., Hermjakob, H., Hoheisel, J.D., Joos, T.O., Kallioniemi, O., Koegl, M., Konthur, Z., Korn, B., Kremmer, E., Krobitsch, S., Landegren, U., van der Maarel, S., McCafferty, J., Muyldermans, S., Nygren, P.A., Palcy, S., Pluckthun, A., Polic, B., Przybylski, M., Saviranta, P., Sawyer, A., Sherman, D.J., Skerra, A., Templin, M., Ueffing, M., and Uhlen, M. (2007). ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nat. Methods* 4, 13–17.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* 76, 4350–4354.

Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249, 505–510.

Vermeer, M., van Doorn, R., Dijkman, R., Mao, X., Whittaker, S., van Voorst-Vader, P.C., Gerritsen, M.P., Geerts, M.L., Gellrich, S., Söderberg, O., Leuchowius, K.J., Landegren, U., Out-Luiting, J.J., Knijnenburg, J., IJszenga, M., Szuhai, M., Willemze, R., and Tensen, C.P. (2008). Novel and highly recurrent chromosomal alterations in Sézary syndrome. *Cancer Res.* 68, 2689–2698.

Wide, L., Bennich, H., and Johansson, S.G. (1967). Diagnosis of allergy by an in-vitro test for allergen antibodies. *Lancet* 2, 1105–1107.

Wiltshire, S., O'Malley, S., Lambert, J., Kukanskis, K., Edgar, D., Kingsmore, S.F., and Schweitzer, B. (2000). Detection of multiple allergen-specific IgEs on microarrays by immunoassay with rolling circle amplification. *Clin. Chem.* 46, 1990–1993.

Yalow, R.S., and Berson, S.A. (1960). Immunoassay of endogenous plasma insulin in man. *J. Clin. Invest* 39, 1157–1175.

Yuille, M., van Ommen, G.J., Brechot, C., Cambon-Thomsen, A., Dagher, G., Landegren, U., Litton, J.E., Pasterk, M., Peltonen, L., Taussig, M., Wichmann, H.E., and Zatloukal, K. (2008). Biobanking for Europe. *Brief. Bioinform.* 9, 14–24.

Chapter 21

Mass Spectrometry and its Applications to Functional Proteomics

Thomas Köcher[1] and Giulio Superti-Furga[2]

[1]*IMP – Research Institute of Molecular Pathology, Vienna, Austria;*

[2]*Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria*

21.1 INTRODUCTION

Mass spectrometry has become an indispensable tool in biological research (Aebersold and Mann, 2003; Domon and Aebersold, 2006). Apart from the ability to characterize sugars (Harvey, 2003), lipids (Han and Gross, 2005), and other small molecules by mass spectrometry, the possibility of identifying proteins in an unbiased, sensitive, and rapid way (Ferguson and Smith, 2003; Steen and Mann, 2004) has led to important, otherwise impossible, insights in biology. Together with other high-throughput technologies, modern mass spectrometry triggered conceptual changes in the way we understand modern basic and applied biological research and thus is one of the key technologies enabling the realization of systems biology (Weston and Hood, 2004). The concept of systems biology can be defined by the use and the integration of systematically acquired large-scale genomic, metabolomic, or proteomic data sets to construct mathematical and predictive models of a biological system (Ideker *et al*., 2001). Such systems might be a protein complex, a specific pathway, a cell, a physiological system, such as the immune system, or even a whole organism.

The aim of systems biology is to understand these systems, to build mathematical models of these systems, to further refine and interrogate them and to predict their behavior to external signals perturbing it. Although understanding the function of single genes and proteins is important, this concept aims at understanding the additional properties of the structure and dynamics of a biological system as a whole. Naturally, the features of a system are not just defined by combining the properties of the involved individual proteins.

The first step enabling systemic approaches in biology was to define the part lists of these systems by sequencing the genomes of most model organisms. The next level of analysis would be provided by a systematic view into the system by building a diagram of the interconnections between these parts. Next, it will be necessary to define the regulatory networks and the physicochemical nature of the interactions such as their affinities or kinetics. Increased experimental complexity will be encountered at each further level of the analysis. Obviously, the systematic analysis of biological systems and the subsequent construction of models require the acquisition of standardized comprehensive sets of quantitative data of the constituting parts of the system under study.

The idea of systems biology has a potentially high impact upon pharmaceutical and medical research. The striking applications include detailed models of regulatory events, especially signal-transduction cascades and modules to provide critical and systematic insights for mechanism-based drug discovery (Fishman and Porter, 2005).

An important part of this new and systemic approach in biology has emerged based on the capabilities of modern mass spectrometry, especially the ability to identify thousands of proteins in a single experiment. This discipline within biology, termed proteomics, is devoted to the study of the proteome, the protein equivalent of the genome, present in protein complexes (Alberts, 1998), organelles (Andersen and Mann, 2006), cells, organs, or even whole organisms. Other techniques (Ito *et al*., 2001; Uetz *et al*., 2000) have been used to create protein–protein interaction data sets but the vast majority of proteomic data sets are being generated by mass spectrometric platforms of ever-increasing performance. The success of state-of-the-art mass spectrometry was mainly triggered by the two separate inventions of soft ionization techniques capable of efficiently ionizing proteins and peptides, electrospray ionization (ESI) (Fenn *et al*., 1989), and matrix-assisted laser desorption/ionization (MALDI) (Karas *et al*., 1987).

There are two major and fundamentally different approaches to studying the proteome. From a historical point

of view, expression-based proteomics is the original concept of proteomics which can be defined as the attempt to catalog the expression of all proteins present in a specific biological system such as a cell. Such experiments are often performed in a differential and quantitative manner, by the analysis of biological systems, differing in specific physiological or pathophysiological phenotypes. This strategy has been successful (Andersen and Mann, 2006; Anderson and Anderson, 2002; Andrade *et al*., 2007) and was also extended to the analysis of the dynamic behavior of post-translational modifications (PTMs) of proteins (Blagoev *et al*., 2004; Kruger *et al*., 2008). However, it is still not possible to completely cover complex proteomes such as those present in body fluids, although the recent years have seen a dramatic improvement in sensitivity, throughput, and dynamic range of signal detection of modern mass spectrometers. The huge dynamic range of protein expression in such biological systems, with proteins present at low copy-numbers frequently hidden by abundant proteins is still preventing substantial progress towards the ambitious aim of characterizing the complete expressed proteome. In the human plasma proteome, protein concentrations are spanning over ten orders of magnitude (Anderson and Anderson, 2002). An important field within expression-based proteomics, called peptidomics, studies the low mass regions of expressed gene products such as neuropeptides, which are not covered by routine mass spectrometric protocols (Hummon *et al*., 2006). Expression-based proteomics is also limited by the challenges associated with precise protein quantitation (Ong and Mann, 2005; Zhang *et al*., 2004). Furthermore, the interpretation of quantitative data from patients is complicated by the natural variation of protein expression, caused by genetic differences and the environmental factors affecting each individual. Additionally, most cellular and pathological processes are influenced by post-translational modifications (PTMs), still difficult to analyze in a comprehensive and quantitative manner (Jensen, 2004; Mann and Jensen, 2003). When used in a biomedical context, expression-based proteomics might lead to correlative relationships between diseases and protein profiles. However, owing to the dynamic range of protein expression, the obtained results are seldom pointing to regulating proteins such as transcription factors or kinases, expressed at low copy-numbers. The correlative nature of the results implies that extensive validation is required.

The other approach was termed functional proteomics. The naming derives from the concept that the association of proteins to protein complexes would also suggest their common involvement in a biological function. Also, it was hypothesized that many if not all cellular functions are not performed by proteins acting in an independent fashion but instead executed by protein complexes, acting like molecular machines (Alberts, 1998). The composition of these protein complexes is dynamic, and many of their components associate and dissociate depending on regulatory events such as phosphorylation. Naturally, the physicochemical binding properties of the complex members are neither equal nor constant, with subsets of proteins tending always to act together by founding core structures of a protein complex (Le Hir and Andersen, 2008). The essence of protein complexes relies on the cooperative effects within its components (Whitty, 2008). In addition, many proteins can have multiple and regulatory functions in more than one protein complex (Gavin *et al*., 2006).

A plethora of different methods for studying protein–protein interactions was developed, ranging from genetic methods such as the two-hybrid assay (Chien *et al*., 1991; Fields and Song, 1989), immunoprecipitations, and affinity purifications (Terpe, 2003). During the last decade, a number of crucial breakthroughs have been achieved in the large-scale characterization of protein complexes of entire organisms (Gavin *et al*., 2002; Ho *et al*., 2002) or signal transduction pathways (Blagoev *et al*., 2003; Bouwmeester *et al*., 2004). In addition, the composition of a constantly increasing number of selected molecular machines has been elucidated (Riedel *et al*., 2006; Vanacova *et al*., 2005). Finally, the proof that the proposed protein complexes really exist was delivered by the three-dimensional structures of molecular assemblies (Groll *et al*., 2005; Le Hir and Andersen, 2008; Lorentzen *et al*., 2005; Nickell *et al*., 2006) revealing features that strongly support the hypothesis functional proteomics was built upon.

The success of the first experimental results of functional proteomics and the connected conceptual changes in biology resulted in the tendency that the characterization of the molecular partners of a protein has become a critical part of most biological studies. An additional but related line of research investigates the interaction of small molecules with proteins such as drugs (Bantscheff *et al*., 2007; Rix *et al*., 2007). There is also the connected field of structural proteomics, exploring the structure of the proteome by methods such as high-throughput protein crystallography (Stevens *et al*., 2001).

The true value of large-scale proteomics data sets will emerge from the integration of data from the different systematic approaches. So far, integrative approaches have just begun to provide additional insights into the function and the evolution of biological systems (Ideker *et al*., 2001; Sharan and Ideker, 2006; Tornow and Mewes, 2003).

Taking a look at current large-scale interaction data, the available protein interaction maps and their compilations are currently incomplete and are lacking the required dynamic resolution. Either proteome maps are derived from specific single cellular states, or, more commonly, they represent a superposition of many different conditions. However, it is conceivable that future maps will not only describe the cellular activities in a comprehensive and potentially even quantitative fashion but also resolve the time axis by studying the changes in protein composition of a complex during biological events such as the cell cycle.

Driven by the vision of a future quantitative biology, the quantitation of proteins is an emerging trend in proteomics.

In order to obtain such data sets, various methods have been developed to allow the quantitative analysis of proteomics samples. In functional genomics, these tools are not only used to distinguish complex components from contaminating proteins (Blagoev *et al*., 2003; Mann, 2006; Ranish *et al*., 2003) but also used to determine the stoichiometry of the proteins present in a complex (Hochleitner *et al*., 2005). More importantly, they can also be used to determine the dynamic composition of protein complexes such as changes reacting upon stimulation or inhibition.

Summarizing, functional proteomics and expression-based proteomics provide complementary views onto the ensemble of proteins and their associations within a biological system. The previous advances in proteomic methodologies and mass spectrometric technologies have influenced biological research already to a great extent. However, the practical impact of the technology on medicine and the drug discovery process is still relatively low (Butcher, 2005). Application of proteomic approaches to medically oriented research have only begun, which will lead to important improvements in our understanding of pathological processes and ultimately in clinical practice (Anderson and Anderson, 2002; Araujo *et al*., 2007; Fishman and Porter, 2005; Weston and Hood, 2004).

21.2 FUNCTIONAL PROTEOMICS

It was a fundamental and surprising discovery that not only a selected subset of the proteins are associated to protein complexes but rather a vast majority of the proteins act in ensembles which can be seen as molecular machines (Alberts, 1998). In order to elucidate the function of a protein it is critical to define the interacting proteins because the function of a specific protein is not only defined by its sequence and structure but also by its partners and the associated protein network. Logically, protein–protein interaction experiments have become a crucial part of most biological research studies. By the identification of the interacting proteins, a protein can be mapped in a relatively straightforward manner to biological pathways, with the complication that certain proteins participate in more than one complex. As already mentioned, a number of functional proteomics technologies are available to this purpose. The common theme of all methods is the fishing of prey proteins by their physical interaction with the bait protein.

The yeast two-hybrid approach (Chien *et al*., 1991; Fields and Song, 1989) maps binary interactions between proteins by genetic means. Its premise is that many transcription factors can be split into two fragments and still activate transcription when the two fragments are indirectly connected. Using this method, plasmids are generated expressing a fusion protein in which the DNA-binding domain fragment is fused to one target protein. Another plasmid is engineered, producing a protein fused to the activation domain. The assay is performed with a whole cDNA library representing most of the proteins expressed in a particular organism or tissue. If the bait and prey proteins interact then the two domains of the transcription factor are indirectly connected and transcription of the reporter gene can occur. A problematic aspect of the assay is its relatively high false positive rate, mirrored by the low overlap of the results obtained in several large-scale studies.

Phage display technology (Rodi and Makowski, 1999) screens for the interaction of immobilized bait proteins with proteins expressed on the surface of phage particles. The potential preys are expressed by introducing cDNA libraries into the genome of the phages. Interacting proteins are identified indirectly by the phages binding to the immobilized protein. These proteins can then be identified by amplification of the phages followed by sequencing the DNA sequences of the putative interacting molecules.

Protein microarrays (Cahill, 2001; Zhu and Snyder, 2003) are conceptually similar to DNA microarrays but instead of DNA, proteins are spotted onto a chip surface. These proteins can be recombinant proteins (Zhu *et al*., 2001), samples from patients (Paweletz *et al*., 2001), or antibodies (Huang *et al*., 2001b). Protein microarrays have been used to detect protein–protein, protein–nucleic acid interactions and biochemical functions such as kinase activity (Zhu *et al*., 2000). Detection of the monitored features is based on fluorescent (Hamelinck *et al*., 2005) or chemoluminescent probes, radioisotope labeling (Zhu *et al*., 2000), or mass spectrometry (Davies *et al*., 1999).

Although all these methods can identify protein–protein interactions, the experimental conditions of these methods are very far from the *in vivo* situation. This might be a critical limitation because PTMs, localization to a specific compartment or interaction with other members of the complex, might be required for binding. Also, all of the above methods are only able to map binary interactions; therefore, the results generated by these methods nicely complement the results obtained by the characterization based on protein complex purification.

A completely different approach aims for the characterization of protein complexes by biochemical means. The aim of biochemical purification of protein complexes is the characterization of their components by mass spectrometry (Aebersold and Mann, 2003). As opposed to techniques such as the yeast two-hybrid assay, these methods are performed in a near to physiological context. The interactions between proteins are not studied in an artificial environment such as a chip surface, but in their correct *in vivo* situation. Consequently, interactions can be characterized in the proper cell type and the proteins will be correctly post-translationally modified. Instead of finding a number of potentially binary interactors, the entire protein complex will be identified, and eventually even parts of other molecular machines interacting with the complex associated with the bait protein. A complication in the data interpretation process is that the proteins

present in a purified complex may represent the superposition of different subcomplexes or variations of one complex. These various forms may represent the different temporal or spatial variants of the same entity. It should be noted that neither binary interactions can be easily defined by protein complex data nor information about a protein complex can be automatically retrieved from binary data sets.

The analysis can be extended to the characterization of biological processes, signaling pathways, or even cells by using more entry points. In the analysis of pathways, a typical number of entry points might be a number between five (Major *et al*., 2007) or as many as 32 (Bouwmeester *et al*., 2004) of bait proteins. In the large-scale analysis of whole organisms, a much larger number of entry points have to be used, optimally all expressed proteins in a given cell. A larger number of entry points result in redundant information but greatly increase confidence in the data set and allow easier data interpretation.

All protein purification strategies share their dependence on mass spectrometric methods to identify the isolated proteins. Currently, state-of-the-art mass spectrometry is the only method capable of analyzing thousands of proteins rapidly and with the required high sensitivity (Aebersold and Mann, 2003).

Most protein complex characterization methods isolate endogenous protein complexes from cells. However, a very straightforward *in vitro* approach identifies interacting proteins by their capture on immobilized recombinant proteins. The other approaches for purifying endogenous protein complexes can be divided into immunoprecipitations, biochemical purification methods, and affinity chromatography.

A traditional method for purifying protein complexes are biochemical purifications using techniques such as size exclusion chromatography or ultracentrifugation. However, usually these methods only allow for partial enrichment and the numerous purification steps which might be required can result in disassembly of the protein complex. An individual purification strategy must be tested and developed for each type of complex. The strategy is limited to abundant and stable protein complexes such as the ribosome, the proteasome, or the spliceosome (Hua *et al*., 1996; Huang *et al*., 2001a). Intact protein complexes can also be separated by native gel-electrophoresis prior to the final mass spectrometric analysis (Meyer *et al*., 2007).

A very versatile method is based on immunochemical capture of the bait protein and its interactors. In immunoprecipitations, protein complexes are isolated from a cell lysate by using an immobilized antibody to a component of a putative complex (Harlow and Lane, 1988). First, nonspecifically binding proteins are removed by washing steps, followed by elution of the protein complex by harsh conditions. The main limitation of this method is the need for a specific and efficient antibody. If such an antibody can be generated, immunoprecipitations of protein complexes can be performed from all types of biological sources such as tissue samples from patients. The second limitation concerns the necessity of eluting the protein complex from the antibody without massive amounts of the antibody coeluting from the column. Ideally, this can be partially achieved by covalently cross-linking the antibody and mild or antigen-specific conditions for elution. The approach has not been applied to large-scale studies but projects to generate antibodies to the entire human proteome might enable such studies in the future (Uhlen and Ponten, 2005). A significant background of contaminating proteins is characteristic for immunochemical purifications, requiring extensive validation experiments.

Affinity purification-based techniques exploit the biochemical properties of a tag genetically fused to the sequence of the bait protein. This tag is then used for isolating the target protein, but also to co-purify the other components of the respective protein complex.

The target-protein and the peptide-tag coding sequences are fused by using standard cloning techniques, followed by expression of the tagged protein in a model organism. A large number of tags and their respective affinity materials have been described. They include histidine (*His*) tags, glutathione S-transferase (GST) tags, Flag tags, the calmodulin-binding peptide, the streptavidin-binding peptide, or the *in vivo* biotinylation of the target tagged peptide using coexpression of the BirA ligase (Waugh, 2005). Each tag is specifically captured by an affinity material such as beads-bound nickel ions binding *His* tags. In addition to the individual tags, several combinations of them have been also used in various configurations.

As an example, the spliceosomal U1 small nuclear ribonucleoprotein was one of the first protein complexes characterized by tagging technology in conjunction with mass spectrometry (Neubauer *et al*., 1997). Here, a *His* tag fused to a known component of the protein complex was used and the complex was isolated by nickel–nitrilotriacetic acid affinity chromatography.

A dramatic improvement of the technology was achieved by the development of the tandem affinity purification (TAP) tag technology (Puig *et al*., 2001; Rigaut *et al*., 1999). In this method, two sequential enrichment steps are used, exploiting two different tags. Originally developed for yeast, the TAP tag is composed of a protein A tag, followed by a tobacco etch virus (TEV) protease cleavage site and a calmodulin binding peptide. The tag can be fused to the N- or C-terminus of the target protein. It might be required to generate both variants of the construct because one variant of the fusion protein can have an impaired function or interacting properties. In yeast, bait proteins are expressed by replacing the endogenous gene by homologous recombination with its tagged construct (Rigaut *et al*., 1999). This strategy not only allows for the expression of the tagged protein under its native promoter but also simultaneously results in abolishing expression of the native gene product. As a consequence, the quantities

of the bait are similar to the native conditions, avoiding overexpression of the construct. Also, the tagged protein does not have to compete for binding its interaction partners with the native protein, resulting in higher total yields of the recovered protein complex.

After cell lysis, the target complex is enriched on an immunoglobulin gamma (IgG) affinity resin and contaminants are removed by washing steps. Very mild conditions are used to release the complex from the beads by cleaving the protein A tag from the fusion protein with the highly specific TEV protease. The next affinity purification is performed with calmodulin bound to another column exploiting the high affinity of the calmodulin-binding peptide to calmodulin in the presence of calcium. From this material, the protein complex can be easily eluted by addition of a metal ion chelating agent such as ethylene glycol tetraacetic acid. In contrast to other protocols, all binding and elution steps are performed in mild buffer conditions. Other two-step approaches have been used by combining other tags, such as the combination of protein G, a TEV protease cleavage site, and streptavidin-binding peptide (Burckstummer *et al*., 2006; Drakas *et al*., 2005; Forler *et al*., 2003; Knuesel *et al*., 2003; Puig *et al*., 2001).

In mammalian or other higher eukaryotic systems, expression of the tagged protein in a cell-line cannot be achieved as easily as in yeast. Especially, it is not practical to replace the native gene by the tagged construct and express the construct in the natural chromosomal context. For these systems, retroviral-mediated gene transfer (Bouwmeester *et al*., 2004) or transient transfection (Westermarck *et al*., 2002) of the bait is the method of choice. As already mentioned, the expression level of the tagged protein of interest is critical and overexpression of the bait protein should be avoided. Apart from disturbing the biological system under study by many possibly unknown ways, the problems likely associated with overexpression are caused by misfolded proteins, resulting in their association with heat shock proteins. They may also be mislocalized or aberrantly interact with other proteins than their natural partners. In addition to overexpression, transient transfection might also induce a cellular shock caused by the surge of newly translated target proteins. The potentially reduced recovery rate, caused by competition with the endogenous gene product, can be avoided by RNA interference-mediated knock-down (Forler *et al*., 2003).

The decision to choose a specific tagging strategy is difficult. Numerous successful constructs have been published and each protocol seems to have advantages and disadvantages. Also, a specific purification protocol might work best for a specific protein complex in one model organism depending on unknown variables, such as the strength of interactions. Already, the fundamental decision whether a one-step or a two-step procedure might be optimal for the protein complex under study seems not to be easy to take. Purifications from eukaryotic cells have been reported by both one-step and two-step purifications, showing a clear trend of higher recovery rates for one-step purifications. On the other hand, they also lead to a higher number of contaminating proteins. Two-step procedures, such as the TAP-tag, tend to yield cleaner complexes but weaker interactors might not be detected (Burckstummer *et al*., 2006). The difference in recovery rates can be estimated to be approximately five times higher for methods using a single step of purification. However, a shorter list of potentially interacting proteins obtained with a two-step procedure might exceed the disadvantage of lower recovery rates because data validation by additional biological experiments is the most time-consuming step.

The epitope tag can interfere with the function of the protein; therefore, it might be required to use both N- and C-terminal fusion proteins for the experiment. It was estimated that the function of the fusion protein is impaired significantly in 10 to 15% of the cases (Bauch and Superti-Furga, 2006). However, even in cases where the bait protein is still functional, its structure might be changed and its binding properties might be impaired, resulting in lower recovery rates and false negative results.

A relatively new approach in mass spectrometry and functional proteomics is the characterization of unseparated and undigested protein complexes in their native form with mass spectrometry (Benesch *et al*., 2006). Similarly to other top-down approaches in mass spectrometry (Kelleher *et al*., 1999), these experiments are currently limited to very abundant and relatively homogeneous complexes, such as ribosomes (Benjamin *et al*., 1998), proteasomes (Sharon *et al*., 2006), or exosomes (Hernandez *et al*., 2006). Following this approach, dedicated instrumentation, manual data interpretation, and large amounts of the purified protein complex are required.

Another interesting method relates to the main limitation that loosely bound interacting proteins might be lost during the protein purification process. Stringent washing conditions or dual tagging strategies further exacerbate this problem. Another shortcoming of the technology is that the topology of the protein complex is not directly amenable. To overcome these difficulties, cross-linking methodology has been successfully applied to the study of protein complexes (Rappsilber *et al*., 2000). These approaches can be combined with regular biochemical purification methods (Vasilescu *et al*., 2004). Studying the topology between proteins, identified cross-linked peptides contain amino acid sequence stretches from directly interacting proteins. Identification of these peptides poses a challenge for analysis and bioinformatic data interpretation but proves that there is a direct interaction and gives clues about their binding interfaces between the interacting partners. Several successful experiments have been reported, but topology mapping experiments require relatively large amounts of the purified complex, the specific detection of the cross-linked peptides, and bioinformatic tools of their detection.

21.3 SAMPLE PREPARATION AND SEPARATION TECHNOLOGIES

Regardless of the type of proteomics experiment, mass spectrometric analysis aiming for protein identification is optimally performed with proteolytic peptide mixtures generated from the isolated proteins (Fig. 21.1). There are a number of reasons for following this analytical strategy. First, mass spectrometers have a lower limit of detection for peptides than for proteins. In electrospray ionization, ionized proteins are simultaneously present at many different charge states, diluting the overall signal strength for a single mass to charge (m/z) value. Second, different isoforms and potentially present post-translational modifications will further dilute the signal intensity of a protein; whereas many peptides will be generated from a common and unmodified region of these different forms of the protein and lead to a maximum of signal strength. Third, medium size peptides can be readily extracted from sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels or separated by high-performance liquid chromatography (HPLC) or isoelectric focusing. Finally, the cleavage pattern generated by specific proteases such as trypsin provides additional information crucial for the protein identification process.

In parallel with improvements of the existing protocols and methods relying on the indirect identification of proteins via the analysis of their proteolytic peptides, there is an emerging trend of analyzing undigested intact proteins (Kelleher *et al*., 1999). As mentioned, such an analysis comes with a number of disadvantages, but allows theoretically the complete characterization of a protein, including the precise primary structure including eventually occurring post-translational modifications (Siuti and Kelleher, 2007). Although this approach is tedious and less sensitive, a number of successful analyses have been performed (Thomas *et al*., 2006). Currently, the approach requires higher sample quantities and is restricted to proteins with a relatively low molecular weight. Increasing the global information about the protein studied, top-down analysis might also be combined with an independent analysis of the proteolytic peptides generated from the same protein,

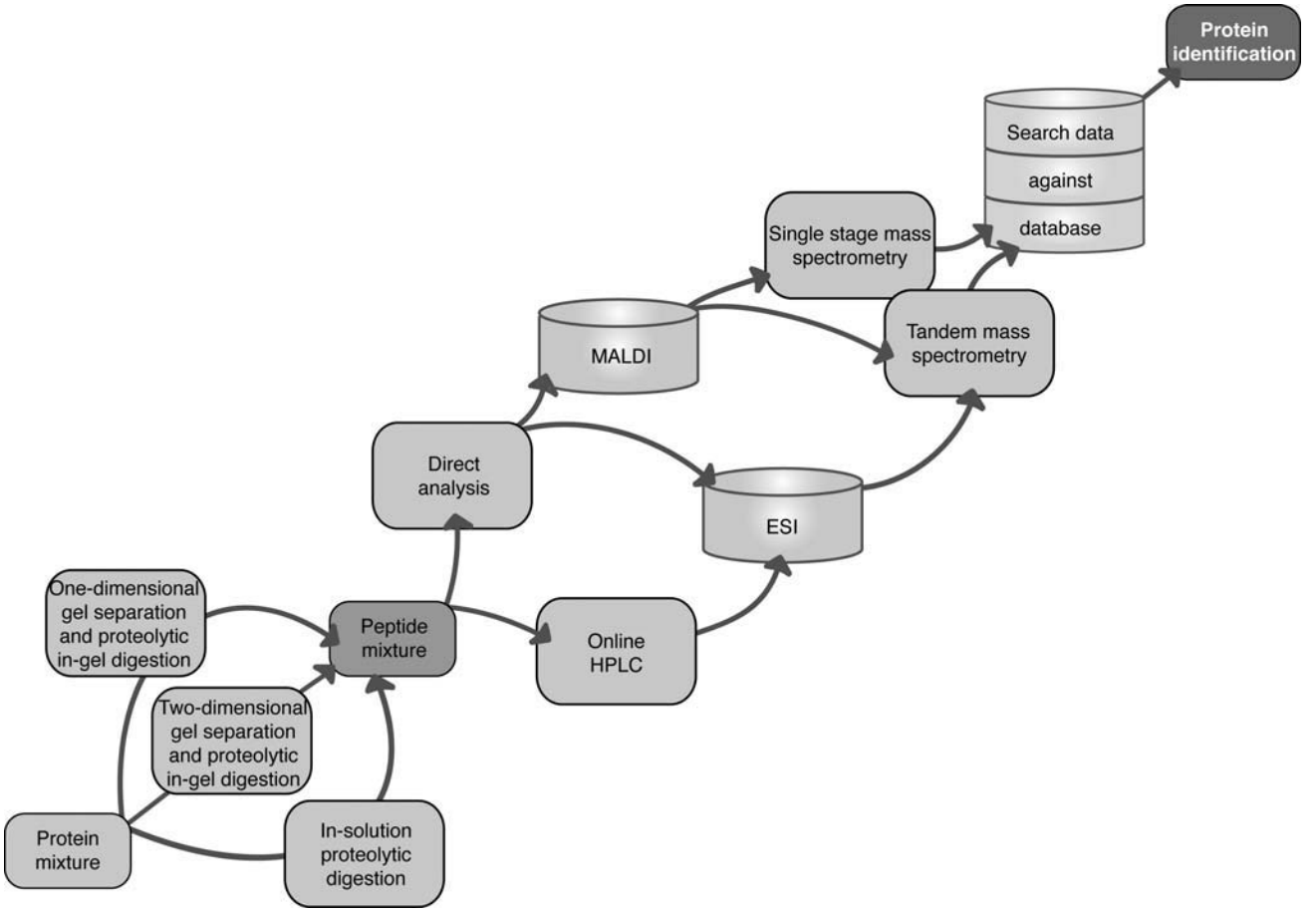


FIGURE 21.1 Main routes to protein identification by mass spectrometry are depicted. Depending on the complexity of the sample, increasing numbers of separation steps are required. Samples containing few proteins present at similar concentrations can be digested without a separation step and the resulting peptides can be analyzed directly. Very complex samples require separation of proteins by methods such as SDS-PAGE gel electrophoreses and/or online HPLC separation of the peptide mixture prior to the mass-spectrometric analysis.

thus reducing some of the current shortcomings of the top-down technology (Kjeldsen *et al*., 2003).

Traditionally, many analytical routines start with the separation of the isolated protein mixtures by SDS-PAGE gel electrophoresis (Fig. 21.1). The highly complex protein mixtures typical for expression-based proteomics experiments might even require maximal separation power achievable with two-dimensional gel electrophoresis (Gorg *et al*., 2004). In functional proteomics experiments, the complexity of the samples and the dynamic range of the quantities of the proteins present in the sample are much lower; therefore, proteins are rather separated in only one dimension. The gels are usually stained with a mass spectrometry compatible dye such as silver (Shevchenko *et al*., 1996), fluorescence-based dyes such as SYPRO ruby (Berggren *et al*., 2000), or Coomassie. As a positive side effect, SDS-PAGE separation removes potentially interfering buffer components such as detergents from the sample, and the complexity of the protein mixture is decreased by separating the proteins according to their molecular weight. Additionally, the staining pattern of the gel can be used as semiquantitative assessment of the experiment, for example to evaluate the quantity of the bait protein versus the interacting components or the comparison with a control sample. Depending on the experiment, individual protein bands of interest are excised or the entire lane is cut into slices. These gel slices are first incubated with reducing and alkylating reagents in order to cleave eventually present cysteine bonds (Shevchenko *et al*., 2006). After several washing steps the proteins are in-gel digested with a specific protease, such as Lys-C, chymotrypsin, or in most cases trypsin, to produce peptides for mass spectrometry analysis (Shevchenko *et al*., 2006). The digested peptides are extracted from the gel pieces and after an optional sample clean-up and desalting step (Erdjument-Bromage *et al*., 1998) analyzed by mass spectrometry. The combined extraction and in-gel digestion efficiency of typical tryptic peptides from a gel is only about 20% and is highly dependent on the primary structure of the peptide (Havlis and Shevchenko, 2004).

Although SDS-PAGE offers a number of practical features such as a first quality assessment of the biological experiment and an efficient protein separation, it is also time consuming and results in the multiplication of mass spectrometric analysis time because each gel-band is analyzed separately. In order to circumvent these disadvantages, protein mixtures can be digested in solution without prior separation of individual components and analyzed by mass spectrometry (Link *et al*., 1999). The main advantages of in-solution digestion protocols are the reduction of the total time required for analysis and a higher recovery of peptides when compared to in-gel digestion. However, without initial separation of the proteins in the mixture, online separation of the generated complex peptide mixture is generally required. Buffer components, such as detergents, might prohibit a direct analysis of the sample because they can interfere with the ionization process. If these conditions cannot be avoided during the purification protocol, the proteins can be precipitated with trichloroacetic acid and acetone, cleaned and dissolved in a digestion buffer containing the appropriate protease. Naturally, any additional sample preparation steps reduce the overall sensitivity of the analysis. In-solution digests of protein mixtures absolutely require separation of the generated peptide mixtures by HPLC. If required, an option to improve the total peak capacity of the separation is two-dimensional liquid chromatography (Link *et al*., 1999; Washburn *et al*., 2001). Obviously, the combination of chromatographic methods should ensure a maximum of orthogonality. Most commonly, peptide mixtures are separated by strong ion exchange in the first dimension followed by reverse phase chromatography (Washburn *et al*., 2001). As an alternative, combinations of two reversed phase separations, operated at two different pH conditions, can be preformed. These techniques can be applied by using one HPLC system without collecting fractions. There is also the possibility of using isoelectric focusing for the first dimension, with the individual fractions loaded on the reversed phase HPLC system online coupled to the mass spectrometer. In the first dimension, immobilized metal affinity chromatography (IMAC) or other affinity chromatography methods might be also used for enrichment of specific peptides, such as phosphopeptides (Posewitz and Tempst, 1999).

The proteolytic peptide mixtures can be either directly analyzed by mass spectrometry (Wilm and Mann, 1996; Wilm *et al*., 1996) or separated by HPLC before mass-spectrometric analysis (Fig. 21.1). Simple samples can be analyzed by directly electrospraying the unseparated peptide mixture into the mass spectrometer (Wilm *et al*., 1996). Such an analysis is faster when compared to liquid chromatography-mass spectrometry (LC-MS/MS) because the loading of a sample onto a liquid chromatography system and subsequent separation is a time-consuming procedure, with gradient time frames of 30 minutes up to several hours. On the other hand, the use of online-coupled HPLC systems not only results in a dramatically higher number of detected peptides but also greatly facilitates automation. Typically, an even greater number of proteins can be identified using two-dimensional peptide separations such as combining strong ion exchange with conventional reversed phase separation (Washburn *et al*., 2001).

It is of high importance to aim for a good chromatographic resolution because the resulting sharp elution profiles will ensure high ion counts, resulting in lower limits of detection. It is also critical to match the expected peak capacity of the complete separation strategy with the expected complexity of the sample. In practice, laboratories aim to achieve chromatographic resolutions between 10 and 20 seconds at full width half maximum (FWHM). Another important aspect of the LC-MS/MS instrumentation is the flow rate, which inversely affects the ionization

efficiency and thus the sensitivity of the analysis (Haskins *et al*., 2001; Luo *et al*., 2005; Wilm and Mann, 1996). In general, low-complexity samples can be analyzed simply and rapidly without peptide separation, whereas complex samples require chromatographic separation before mass-spectrometric analysis (Fig. 21.1).

21.4 MASS SPECTROMETRY

The aim of the mass-spectrometric analysis is to characterize the proteins present in the isolated samples. Theoretically, this could be achieved by measuring the intact masses of these proteins with high precision and searching these data against a database. However, the measured intact molecular mass is rarely identical with the mass predicted from the genomic information; therefore, this approach does not lead to correct protein identifications. However, if the ionized intact protein is fragmented in the mass spectrometer into its characteristic fragments, the additional sequence-specific information allows protein identifications with extremely high confidence.

Currently, protein identifications are usually based on the analysis of their proteolytic peptide mixtures. The identification is based on peptide masses alone, where the mass lists of experimental data sets are compared to the theoretical mass lists of tryptic peptides based on databases. The preferable option is to fragment as many peptides as possible, record the *m*/*z* (mass divided by charge) values of the fragment ions and to search the mass list of fragment data combined with the mass of the corresponding precursor ion. This approach is feasible because peptides tend to fragment in a specific and characteristic way (Steen and Mann, 2004). In most instruments, peptides can be very efficiently fragmented by collisional-activated dissociation (CAD), also called collisional-induced dissociation (CID). When realized in a triple quadrupole mass spectrometer, peptides collide with neutral gas molecules and at the same time a low electric potential is applied. Under these conditions, peptides fragment in a defined way with the main fragmentation channel leading to cleavage of the peptide bond to form so called b- and y-ions (Roepstorff and Fohlman, 1984). The specific nature of these fragment ions allows searching of such data against a database and programs have been developed automatically comparing experimental data sets against data stored in protein databases (Eng *et al*., 1994; Perkins *et al*., 1999). There is another main type of ion process leading to the fragmentation of peptides. By reaction of low energy electrons with peptide ions in a radical process, called electron capture dissociation (ECD), the amine bond of the peptide backbone is preferentially cleaved, leading to c- and z-ions (Zubarev, 2004; Zubarev *et al*., 1998). Recently, another technique was invented where electrons are transferred onto peptides by an ion–ion reaction from small organic anions, called electron transfer dissociation (ETD), leading to a fragmentation pattern similar to ECD (Syka *et al*., 2004). Both techniques can be used to identify proteins. In addition, post-translational modifications, such as phosphorylation, acetylation, or methylation, can be identified, with ECD and ETD being the techniques of choice to characterize labile modifications (Jensen, 2004).

Although many different types of mass spectrometers exist, which are capable of generating the required data for protein identification, the general configuration of their components is very similar. First, the analyte molecules are ionized and transferred into the gas phase. This is done in the ion source of the instrument exploiting one of the two soft ionization techniques, electrospray ionization (ESI) (Fenn *et al*., 1989) or matrix-assisted laser desorption/ionization (MALDI) (Karas *et al*., 1987) (Fig. 21.2a, b). Apart from ion transfer components, the second and essential component in a mass spectrometer is the mass analyzer. Ion traps (Paul traps), quadrupole mass analyzers, time-of-flight analyzers (ToF), Fourier transform ion cyclotron resonance (Penning traps), and the Orbitrap are the most commonly used mass analyzers, all analyzing the *m*/*z* ratios of the analyte molecules (Fig. 21.2c–f). Some of these devices can perform multiple rounds of ion manipulation, such as ion traps, which can perform in theory an unlimited number of ion selection, fragmentation, and ion analysis experiments. On other types of instruments, several mass analyzers might be operated in series. For example, triple quadrupole mass spectrometers have three quadrupole mass analyzers aligned with each other along the axis of the ion beam. On such an instrument, tandem mass spectrometry (MS/MS) is performed by using the first quadrupole for selecting ions of interest, the second quadrupole for fragmentation, and in the third quadrupole the product ions are analyzed. There are variations of mass spectrometers termed hybrid instruments, such as quadrupole ToF instruments, combining different mass analyzers. Ingenious solutions have been developed, such as orthogonal ion extraction in order to fit together devices working with constant ion beams such as quadrupoles with devices, such as ToF analyzers (Chernushevich *et al*., 2001). The last and equally essential component of each mass spectrometer is the detector, such as microchannel plate detectors (MCPs), where the ions are detected. Many different combinations of ion sources with mass spectrometers have been constructed, with specific combinations easier to realize such as MALDI ion sources generating pulses of ions and ToF analyzers.

In the most common instrumental designs, ESI ion sources are installed on mass spectrometers capable of performing tandem mass spectrometry (MS/MS) experiments. Currently, ion traps (Paul traps), quadrupole time-of-flight instruments (Q-ToF), Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers, and the Orbitrap are the most frequently used types of instrumentation used in high-end protein analysis (Fig. 21.2c–f). A critical factor for the unambiguous identification of proteins is the mass accuracy achievable with the mass spectrometer (Clauser *et al*., 1999),

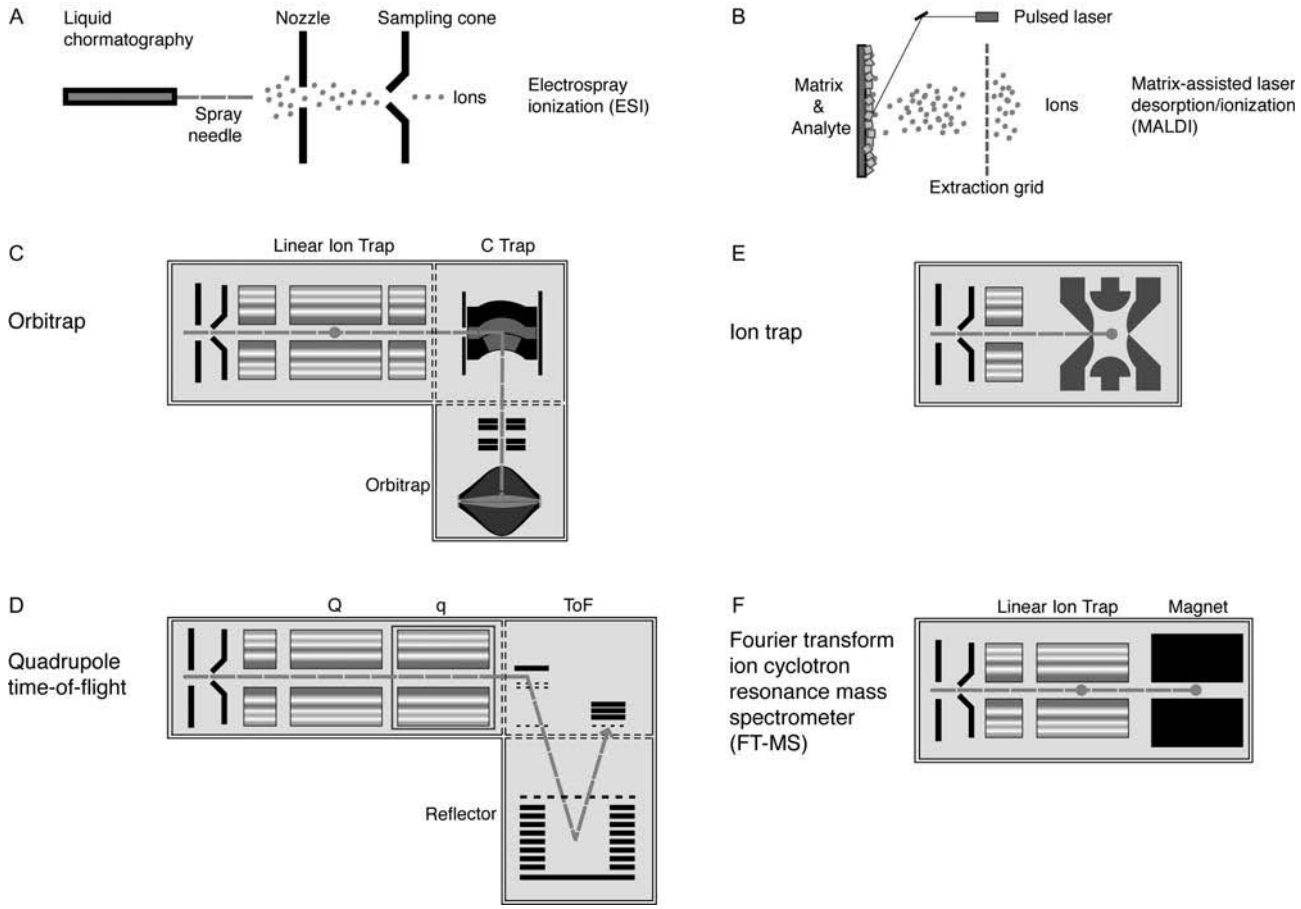


FIGURE 21.2 Electrospray ionization (ESI, A) and matrix-assisted laser desorption/ionization (MALDI, B) are the two soft ionization techniques capable of ionizing macromolecules and transferring them into the gas phase with high efficiency. Schemes of the ionization processes are shown in panels **A** and **B.** Some important mass spectrometers are depicted schematically in panels C to F. Their general configuration is very similar. First, ions are generated with one of the two soft ionization sources. Second, the ions are separated in a mass analyzer, followed by their detection. In tandem mass spectrometry, the initially detected ions are selected individually, fragmented, and their fragmentation products are separated and detected. **C.** The linear ion trap-Orbitrap mass spectrometer is a hybrid instrument assembling a linear ion trap with the Orbitrap. The *m*/*z* values of the ions are measured in the Orbitrap, with high resolution and high mass accuracy. Mass analysis is based on trapping ions in an electrostatic field. Trapped ions are oscillating around the ring electrode of the instrument, with their frequency inversely proportional to their *m*/*z*. Ions of interest are then isolated in the linear ion trap and fragmented. The fragment ions can be detected either in the linear ion trap or in the Orbitrap, each device associated with different duty cycles, sensitivities, and mass accuracies. **D.** The three-dimensional ion trap (Paul trap) captures the ions in an electromagnetic field. By changing the field, selected ions can be ejected from the trap and measured in a detector, generating a mass spectrum. For tandem mass spectrometry experiments, ions of a particular *m*/*z* can be selected and then fragmented in the trap, followed by analyzing the fragments to generate a tandem mass spectrum. **E.** The quadrupole time-of-flight (ToF) instrument combines the front part of a triple quadrupole instrument with a reflector ToF section for measuring the masses of the ions with a higher mass accuracy. Quadrupole mass analyzers can separate ions based on the stability of their trajectories in an electromagnetic field, analogous to ion traps. In a ToF analyzer the ions are accelerated to an identical kinetic energy and are then separated along a field-free flight tube as a result of their different velocities. For tandem mass spectrometry, the *m*/*z* values of the ions are first measured in the ToF part of the instrument, to obtain their molecular masses. Second, ions are selected in the first quadrupole (Q) and fragmented in the collision cell (q). The product ions are then recorded in the ToF analyzer. **F.** The Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer is an ion trapping instrument, exploiting the ion trapping ability of magnetic fields. Upon excitation, ions move in phase and the frequency of their image current can be detected, again inversely proportional to their *m*/*z*. The figure depicts a hybrid instrument combining a linear ion trap with an FT-ICR mass spectrometer. The FT-ICR mass spectrometer is used for measuring the intact peptide ions with high mass accuracy and resolution, while the linear ion trap is used for fragmentation and, eventually, for measuring the masses of the product ions.

with FT-ICR mass spectrometers allowing the highest mass resolution and precision (Marshall *et al*., 1998). High sensitivity and high-sequencing speed might, nevertheless, compensate for lower mass accuracy by generating more MS/MS spectra per protein. Although all of the above-mentioned mass spectrometers can be used for protein identification, a specific analytical question might be answered optimally using specific mass-spectrometric instrumentation. For example, sequencing of labile modifications might favor choosing ECD for fragmentation, which in turn requires FT-ICR technology (Zubarev, 2004). Q-ToF instruments now have mass accuracies for typical tryptic peptides of approximately

10 parts per million (ppm), the Orbitrap can achieve mass accuracies of better than 2 ppm, and FT-ICR instrumentation can be optimized for sub-ppm mass accuracy.

It should be kept in mind that the low mass accuracies typically achieved on ion traps of triple quadrupole instruments can also lead to the unambiguous identification of proteins. Using these instruments, it might be necessary to employ fragmentation data from more peptides to ensure correct protein identifications, while it might be very critical to assess the used parameters of the search software by validation against a decoy database (Elias and Gygi, 2007).

As mentioned above, ESI-MS experiments are typically performed by directly coupling an HPLC system to the mass spectrometer. Such experiments are performed in a data-dependent acquisition mode, whereby precursor ion selection and MS/MS acquisition are automatically triggered by the instrumentation software. The *m*/*z* values of the peptides eluting at a given time from the column are recorded, and the peptide with the most intense signal is automatically selected, fragmented and a product ion spectrum is recorded. This procedure can be repeated with the other eluting peptides, but is limited by the duty cycle of the mass spectrometer, detection limits, and the *m*/*z* value of the peptides.

An alternative method to ESI-MS is MALDI-MS, usually performed with ToF instruments recording only the mass spectrum of the peptide ions. The obtained data sets can then be searched against a database. This analytical strategy is rapid but protein identification, based on single stage MS data, relies on dramatically less statistical information as compared to tandem MS data sets. This can be problematic when analyzing protein mixtures or samples containing modified proteins, common in samples from higher eukaryotes. Instruments such as MALDI–Q-ToFs (Shevchenko *et al*., 2000) or MALDI-ToF/ToFs (Medzihradszky *et al*., 2000) can operate in an MS/MS mode. Although these instruments are capable of precursor ion selection, tryptic peptide mixtures from complex samples are usually separated offline with an HPLC system and the resulting fractions are spotted on the MALDI target. The MALDI ion sources generate mostly singly charged ions which might result in weaker fragment ion spectra when compared to those generated by multiply charged electrospray ions.

Generally, the time required for analysis is a critical factor for most proteomics laboratories. Although MALDI-ToF offers the fastest route to protein identification, samples with high complexity or with peptide concentrations spanning over a wider dynamic range can seldom be analyzed to a sufficient analytical depth. In any case, MALDI-ToF-based protein identification requires prior separation of proteins by gel electrophoresis or HPLC, both time-consuming techniques. Shotgun ESI LC-MS/MS analysis of in-solution digested protein samples (Washburn *et al*., 2001) might be the method leading to faster and unambiguous identification of protein complexes, with the advantage of being easily automatable.

21.5 QUANTITATIVE PROTEOMICS

Proteins can be relatively quantified using mass spectrometry (Bantscheff *et al*., 2007) by comparing two or more samples containing variable amounts of a protein. The result of such an experiment is a ratio of the measured protein abundance in the analyzed samples. On the other hand, some biological questions require the absolute quantification of proteins (Bantscheff *et al*., 2007), such as in experiments analyzing the stoichiometry of a protein complex. Such data can be generated by spiking a peptide mixture with isotopically labeled already quantified peptides (Gerber *et al*., 2003).

In expression-based proteomics experiments, the relative quantitation of the proteins is essential, especially when comparing two different biological conditions or patient samples with healthy individuals. Functional proteomics experiments are less dependent on quantitative data. However, quantitative results can aid in discriminating *bona fide* components of the complex from contaminating proteins (Ranish *et al*., 2004, 2003) and can characterize the stoichiometry of protein complexes. Mass spectrometry is becoming increasingly popular for quantifying and identifying proteins at the same time. It can also be used to quantify the differential modifications between samples, such as phosphorylation (Kruger *et al*., 2008). Unfortunately, mass spectrometric results are not quantitative *per se*, because the signal strength of a peptide ion depends not only on its concentration but also on the ionization efficiency and on other parameters which are not known *a priori*. In addition, the various steps before mass spectrometric analysis, such as extraction from gel slices, introduce sample losses depending on the physicochemical properties of the peptides.

Several solutions have been found to address this issue. The most accurate results can be achieved by incorporating a stable isotopic label into the peptides. This can be done at various points during the entire workflow of the experiment. Obviously, the sooner the label is introduced and the two samples can be mixed, the less unspecific sample losses during the biochemical and analytical manipulations will distort the result. Stable isotope labeling with amino acids in cell culture (SILAC) (Mann, 2006) is a method for *in vivo* incorporation of the isotopic labels into the proteins at the starting point of the experiment. First, the cell populations are grown separately in media containing amino acids with different isotopic compositions. In the next step, the biological experiment is performed and before starting further protocols, such as purification of protein complexes, the two samples are mixed. In the final mass spectrometric experiment, each peptide will appear as a pair of peaks with the intensity ratio reflecting the ratio of quantities.

Labeling at a later stage can be performed before tryptic digestion such as by labeling the cysteines of the proteins with the isotope-coded affinity tag (ICAT) (Gygi *et al*., 1999). In the final mass spectrometric experiment each peptide will appear as a pair of peaks with the intensity ratio

reflecting the ratio of quantities. Using an even later stage of labeling, isobaric tags for relative and absolute quantification (iTRAQ) label the N-terminal amino group and the lysine residues of each peptide (Ross *et al*., 2004). These tags are different in the sense that each sample is labeled with a tag of the same mass unit. In contrast to ICAT or SILAC, the observed *m*/*z* values are equally shifted for all differentially modified samples. The differences between the tags appear only after fragmentation, because each tag generates a unique reporter ion and the ratios between the signal intensities of these reporter ions reflect the ratios of the relatively quantified samples.

Quantification can also be achieved without introducing an isotopic label. Spectral counting exploits the fact that abundant proteins are identified with a higher number of tandem mass spectra than low abundance proteins. Using this strategy, the samples are analyzed independently with the same analytical protocol. The protein abundance in each sample is approximated from the number of MS/MS spectra for each protein present in the sample, normalized for the number of theoretical tryptic peptides for each protein (Ishihama *et al*., 2005).

The third possibility is to compare different biological samples in separate mass spectrometry experiments acquired under identical conditions. First, the chromatographic runs have to be aligned by means of computational tools (Jaffe *et al*., 2006), followed by comparing the signal intensities of the peptides and computing protein ratios from the peptide ratios. The approach is compatible with applications that require profiling of multiple biological samples, such as proteomics-based candidate biomarker discovery. The drawbacks of this method include the increased computational complexity and the required technical and biological repeats to account for the higher statistical variation of this method.

For all quantitative methods, a significant complication relates to the ambiguity in assigning peptides to proteins (Nesvizhskii and Aebersold, 2005; Rappsilber and Mann, 2002). If peptides are shared between different proteins of a protein group, its quantification will no longer be a reliable measure of the abundance of the corresponding proteins. Consequently, the quality and reliability of protein identification determines the quality of the corresponding quantification.

21.6 DATA INTERPRETATION, VALIDATION, STANDARDIZATION, AND BIOINFORMATIC ASPECTS

As already mentioned, proteins are identified either on the list of proteolytic peptide masses or on the combined sets of peptide masses and fragment ion masses. In contrast to the plethora of methods to analyze these peptide mixtures, the approaches to analyze the obtained data sets are fairly similar. First, the raw data files are processed by the software supplied by the manufacturing company of the mass spectrometer. Additional software might be applied for other processing steps (Gentzel *et al*., 2003). Typical software tools apply smoothing, centroiding, deisotoping, and charge-state deconvolution algorithms to the acquired spectra. The processed data sets are at one point translated into mass lists and are searched against a protein database.

Manual interpretation of tandem mass spectra and searching defined information such as sequence tags is possible (Mann and Wilm, 1994), but antiquated and not adequate for the large data sets generated by LC-MS/MS. The diverse LC-MS/MS platforms have in common that thousands of tandem mass spectra are acquired per experiment. Their assignment to peptide sequences and the subsequent identification of proteins poses an enormous computational challenge. Obviously, the identification process is essential for proteomics and should be done with as little false positive or false negative assignments as possible. Currently, the two most commonly used algorithms use either a probabilistic approximation such as the search engine MASCOT (Perkins *et al*., 1999) or a mathematical correlation method such as SEQUEST (Eng *et al*., 1994). The output of these software tools is a list of identified proteins and peptides, each assigned with a characteristic score, reflecting various aspects of match between the experimental data and the data stored in the database. Although these and other search engines differ in their mathematical approach and their exact statistical methods (Colinge *et al*., 2003; Craig and Beavis, 2004), the most crucial factor reducing the obtained false positive negative rate is the experimentally obtained mass accuracy. The scores obtained for individual protein identifications facilitate the discrimination between correct and false positive hits. Consequently, the greatest care should be taken in defining the thresholds of the minimum scores and the allowed mass tolerances for the precursor ion and the fragment ions. A valid approach for validation of the chosen parameters is to search the obtained data sets against a decoy protein database (Elias and Gygi, 2007), obviously at the cost of increasing the total search time. An alternative and faster method is used by the program PeptideProphet (Keller *et al*., 2002). By applying a statistical model and using a discriminant function calculated from the scores obtained from the original data set probabilities for protein identifications are calculated. The software also uses an algorithm that applies a statistical model; the specific protein from a protein group (Rappsilber and Mann, 2002) is present in the sample (Nesvizhskii and Aebersold, 2005; Nesvizhskii *et al*., 2003). Grouping of identified proteins is important because a set of identified peptides might match to a group of similar proteins, which are often homologous proteins or different isoforms of the same reading frame (Nesvizhskii and Aebersold, 2005; Rappsilber and Mann, 2002). Several algorithms have been developed to group these proteins, facilitating easier

assessment and interpretation of data. Auxiliary information such as retention times of the reversed phase separation prior to mass spectrometry might be used in the validation process (Nesvizhskii and Aebersold, 2005).

Also, the initial protein identification process might be followed by further filtering the obtained protein lists by setting additional thresholds such as a minimum peptide length or a specific number of peptides to consider a protein identification to be correct. There is no consensus in the field about the specific requirements for protein identification such as minimum mass accuracies or acceptable false positive rates. In addition, a controversy exists over the question whether protein identifications by tandem mass spectrometry data should be based on a minimum of two identified tryptic peptides or if protein identifications based on a single tandem mass spectrum with very high statistical significance or MS/MS/MS (Olsen and Mann, 2004) data of a single peptide might be sufficient.

These questions are directly related to the issues of data standardization in mass spectrometry and proteomics because they have a direct impact on the quality of future databases of proteomics data sets and on the assessment of already published data. The progressively growing amount of protein–protein interaction data and expression-based proteomics data has to be standardized in a similar fashion as DNA-based data sets. Naturally, the different types of mass spectrometers and ionization processes, and the different types of chromatographic systems, search engines and other software tolls, make the assessment of published data increasingly difficult. For systems biology, integration of large-scale data sets is of utmost importance. Without common standards in data formats, data acquisition, data interpretation, and data storage sharing of data sets will be always limited.

There are efforts to define the standards for generating and publishing mass-spectrometric data and other parameters of proteomics experiments. Although this process is far from complete and different groups of researchers already defined their own standardization protocols (Brazma *et al*., 2006), large-scale data sets require a common format of publishing and storing of proteomic data (Taylor *et al*., 2007).

Some fundamental experimental information which should be reported includes the mass-spectrometric instrumentation and resolution, mass accuracy, software for data interpretation, HPLC flow rates, and composition or type of MALDI sample plates.

Naturally an analogous standardization should be applied for the protein identification process. Currently, there is no consensus about specific minimum requirements for protein identification such as an acceptable false positive rate. In addition, there are a number of search engines used in the protein identification process, and the results obtained by the different algorithms are not identical. Standardization in this area would result in the detailed descriptions of the software used, input parameters such as the database queried, the restrictions applied to the search, the cleavage agents, and the mass tolerances used. Specific requirements for publishing the output of the identification process, such as the accession codes of the identified proteins, the protein scores or the obtained sequence coverage, have also been defined. In addition, any publication reporting proteomics data should contain a statistical analysis of the data, such as the determination of the false positive rate.

Standardization should also be applied to the final output of proteomic protein–protein interaction studies (Orchard *et al*., 2007). In recent years, several public databases have been created for storing functional proteomics data sets. Currently, the data are curated manually, and most of the data are still extracted from the literature. Obviously, common standards would facilitate uploading protein–protein interaction data sets into such databases. A number of parameters are obvious for the format. For example, they should use unambiguous protein identifiers and clear descriptions of experimental conditions. Identical requirements should apply to the publication of large-scale data sets.

21.7 DATA VALIDATION AND INTERPRETATION

An unwanted side aspect of the constantly increasing sensitivity of mass spectrometry is that various contaminating proteins, such as keratins or highly abundant proteins, are also identified in most protein complex purifications.

In the case of small-scale experiments, these contaminants do not impose substantial problems, because a manual interpretation of the data sets based on biological knowledge and experience is still possible. There is also the option to perform additional experiments for validation, such as co-localization studies and gain-of-function or loss-of-function experiments. The already mentioned quantitative methods in mass spectrometry can also facilitate this process (Mann, 2006; Ong and Mann, 2005; Zhang *et al*., 2004). The components of the protein complex can be distinguished from contaminant proteins by comparison with a control experiment (Ranish *et al*., 2003). This approach could also be applied to large-scale studies. In such experiments, the subjective and individual evaluation of the results is no longer possible, simply because the amount of data does not allow removal of contaminating proteins based on individual judgments. Proteins co-purifying with protein complexes are either abundant proteins or proteins with the tendency to enrich during the specific purification procedure. Therefore, a possible solution to remove contaminating proteins in high-throughput data is to quantitatively compare them against core proteomes (Schirle *et al*., 2003). These core proteomes have been defined as the subset of proteins which will always be identified in a mass-spectrometric analysis of a cell lysate. They represent the highly abundant proteins in a specific cell type.

By taking the conditions of the purification protocol into account, likely contaminants are defined as repetitively identified proteins. If a protein is identified in a larger number of pull-down assays than a specific cut-off value (Krogan *et al*., 2006) or is present in pull-downs of unrelated proteins (Bouwmeester *et al*., 2004) then it will be removed automatically from the data set.

Judging the data published from the yeast studies using the TAP tagging technique, contaminating proteins appear to be a relatively moderate problem. In part, the technology allows a relatively clean isolation of protein complexes in yeast. In addition, the redundancy obtained with tagging several if not all members of protein complex, characteristic for large-scale studies enables a robust statistical and probabilistic analysis (Gavin *et al*., 2006; Krogan *et al*., 2006). However, in experiments using cell-lines derived from higher organisms, contaminating proteins have proved to be a problematic issue stressing the necessity of validation experiments or aiming for a similar or better coverage of technical repeats and entry points.

Large-scale experiments require a completely automated interpretation routine. This interpretation can be based on mathematical *a priori* models. For example, protein complexes can be simplified as static entities, by the creation of networks in which the edges represent proteins shared between complexes (Gavin *et al*., 2002). This does not fully reflect the dynamics of complex composition and showed to be very sensitive to false positive results, even after surviving the initial statistical filtering. Another simplified model assumes that each identified protein in a pull-down interacts with each other protein in the same complex, large networks can be computed in which each protein is represented as an entity connected to various degrees with other proteins, in a network of binary interactions similar to those generated by two-hybrid screens (Ito *et al*., 2001; Uetz *et al*., 2000). Protein complexes can also be represented as modular entities consisting of core components that are always grouped together and alternative attachments of one or more other proteins (Gavin *et al*., 2006). Decomposed this way, the modular organization of the interactome allows for the prediction of evolutionarily more conserved elements represented by the core elements and more variable and less conserved attachments as well as for diversification of function based on a combination of limited sets of components (Gavin *et al*., 2006).

21.8 LIMITATIONS AND FUTURE PERSPECTIVE OF FUNCTIONAL PROTEOMICS

During the last decade, technical developments in protein complex purification, mass spectrometry and bioinformatics enabled large-scale protein–protein interaction studies. Studies in yeast (Gavin *et al*., 2006; Krogan *et al*., 2006) and animals have been reported, and there are plans to tackle the human proteome. The underlying aims of these efforts surpass the intrinsic value of an infant building plan of a cell. Proteomic studies are an important part of the toolbox of stem biology, trying to elucidate the dynamic nature of the proteome composition and complex assembly and disassembly, the absolute and relative quantitation of the parts of these complexes, and the proteome as a whole. The final aim of systems biology is to build predictive models of biological systems similar to the construction plans of engineered machines.

Current experimental limitations of the technology still hamper the mapping of transient or weak protein–protein interactions, the analysis of many membrane proteins, the detection of proteins present at low copy-numbers in the presence of highly abundant proteins, and the comprehensive characterization of post-translational modifications.

Given the tremendous and continuous improvements of mass spectrometry, HPLC technology, bioinformatics, and protein purification techniques it is foreseeable that all these limitations will be overcome. It is evident that the most rewarding challenge will be the application of these technologies to medicine (see also Chapter 18). Although efforts in this direction have been made, medical proteomics is still in its infancy, and medicine and pharmaceutical sciences will profit enormously from these emerging trends (Goh *et al*., 2007). Apart from the far distant possibility of using comprehensive and quantitative models of pathological systems in clinical practice, the possibility of identification and characterization of prognostic molecules and the discovery of new targets for therapeutics are close visions of future directions in medicine (Weston and Hood, 2004).

REFERENCES

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198–207.

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92, 291–294.

Andersen, J.S., and Mann, M. (2006). Organellar proteomics: turning inventories into insights. *EMBO Rep.* 7, 874–879.

Anderson, N.L., and Anderson, N.G. (2002). The human plasma proteome – history, character, and diagnostic prospects. *Mol. Cell. Proteomics* 1, 845–867.

Andrade, E.C., Krueger, D.D., and Nairn, A.C. (2007). Recent advances in neuroproteomics. *Curr. Opin. Mol. Ther.* 9, 270–281.

Araujo, R.P., Liotta, L.A., and Petricoin, E.F. (2007). Proteins, drug targets and the mechanisms they control: the simple truth about complex networks. *Nat. Rev. Drug Discov.* 6, 871–880.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* 389, 1017–1031.

Bauch, A., and Superti-Furga, G. (2006). Charting protein complexes, signaling pathways, and networks in the immune system. *Immunol. Rev.* 210, 187–207.

Benesch, J.L.P., Aquilina, J.A., Ruotolo, B.T., Sobott, F., and Robinson, C.V. (2006). Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies. *Chem. Biol.* 13, 597–605.

Benjamin, D.R., Robinson, C.V., Hendrick, J.P., Hartl, F.U., and Dobson, C.M. (1998). Mass spectrometry of ribosomes and ribosomal subunits. *Proc. Natl. Acad. Sci. USA* 95, 7391–7395.

Berggren, K., Chernokalskaya, E., Steinberg, T.H., Kemper, C., Lopez, M.F., Diwu, Z., Haugland, R.P., and Patton, W.F. (2000). Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. *Electrophoresis* 21, 2509–2521.

Blagoev, B., Kratchmarova, I., Ong, S.E., Nielsen, M., Foster, L.J., and Mann, M. (2003). A proteomics strategy to elucidate functional protein–protein interactions applied to EGF signaling. *Nat. Biotechnol.* 21, 315–318.

Blagoev, B., Ong, S.E., Kratchmarova, I., and Mann, M. (2004). Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.* 22, 1139–1145.

Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., Hopf, C., Huhse, B., Mangano, R., Michon, A.M., Schirle, M., Schlegl, J., Schwab, M., Stein, M.A., Bauer, A., Casari, G., Drewes, G., Gavin, A.C., Jackson, D.B., Joberty, G., Neubauer, G., Rick, J., Kuster, B., and Superti-Furga, G. (2004). A physical and functional map of the human TNF-alpha NF-kappa B signal transduction pathway. *Nat. Cell Biol.* 6, 97–105.

Brazma, A., Krestyaninova, M., and Sarkans, U. (2006). Standards for systems biology. *Nat. Rev. Genet.* 7, 593–605.

Burckstummer, T., Bennett, K.L., Preradovic, A., Schutze, G., Hantschel, O., Superti-Furga, G., and Bauch, A. (2006). An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat. Methods* 3, 1013–1019.

Butcher, E.C. (2005). Can cell systems biology rescue drug discovery? *Nat. Rev. Drug Discov.* 4, 461–467.

Cahill, D.J. (2001). Protein and antibody arrays and their medical applications. *J. Immunol. Methods* 250, 81–91.

Chernushevich, I.V., Loboda, A.V., and Thomson, B.A. (2001). An introduction to quadrupole-time-of-flight mass spectrometry. *J. Mass Spectrom.* 36, 849–865.

Chien, C.T., Bartel, P.L., Sternglanz, R., and Fields, S. (1991). The 2-hybrid system – a method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl. Acad. Sci. USA* 88, 9578–9582.

Clauser, K.R., Baker, P., and Burlingame, A.L. (1999). Role of accurate mass measurement (+/− 10 ppm) in protein identification strategies employing MS or MS MS and database searching. *Anal. Chem.* 71, 2871–2882.

Colinge, J., Masselot, A., Giron, M., Dessingy, T., and Magnin, J. (2003). OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics* 3, 1454–1463.

Craig, R., and Beavis, R.C. (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20, 1466–1467.

Davies, H., Lomas, L., and Austen, B. (1999). Profiling of amyloid beta peptide variants using SELDI ProteinChip (R) arrays. *Biotechniques* 27, 1258–1261.

Domon, B., and Aebersold, R. (2006). Review – mass spectrometry and protein analysis. *Science* 312, 212–217.

Drakas, R., Prisco, M., and Baserga, R. (2005). A modified tandem affinity purification tag technique for the purification of protein complexes in mammalian cells. *Proteomics* 5, 132–137.

Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 4, 207–214.

Eng, J.K., McCormack, A.L., and Yates, J.R. (1994). An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5, 976–989.

Erdjument-Bromage, H., Lui, M., Lacomis, L., Grewal, A., Annan, R.S., McNulty, D.E., Carr, S.A., and Tempst, P. (1998). Examination of micro-tip reversed-phase liquid chromatographic extraction of peptide pools for mass spectrometric analysis. *J. Chromatogr. A.* 826, 167–181.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass-spectrometry of large biomolecules. *Science* 246, 64–71.

Ferguson, P.L., and Smith, R.D. (2003). Proteome analysis by mass spectrometry. *Annu. Rev. Biophys. Biomolecul. Struct.* 32, 399–424.

Fields, S., and Song, O.K. (1989). A novel genetic system to detect protein protein interactions. *Nature* 340, 245–246.

Fishman, M.C., and Porter, J.A. (2005). Pharmaceuticals – a new grammar for drug discovery. *Nature* 437, 491–493.

Forler, D., Kocher, T., Rode, M., Gentzel, M., Izaurralde, E., and Wilm, M. (2003). An efficient protein complex purification method for functional proteomics in higher eukaryotes. *Nat. Biotechnol.* 21, 89–92.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., Edelmann, A., Heurtier, M.A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J.M., Kuster, B., Bork, P., Russell, R.B., and Superti-Furga, G. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440, 631–636.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147.

Gentzel, M., Kocher, T., Ponnusamy, S., and Wilm, M. (2003). Preprocessing of tandem mass spectrometric data to support automatic protein identification. *Proteomics* 3, 1597–1610.

Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., and Gygi, S.P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. USA* 100, 6940–6945.

Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., and Barabasi, A.L. (2007). The human disease network. *Proc. Natl. Acad. Sci. USA* 104, 8685–8690.

Gorg, A., Weiss, W., and Dunn, M.J. (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 4, 3665–3685.

Groll, M., Bochtler, M., Brandstetter, H., Clausen, T., and Huber, R. (2005). Molecular machines for protein degradation. *Chembiochem* 6, 222–256.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999.

Hamelinck, D., Zhou, H.P., Li, L., Verweij, C., Dillon, D., Feng, Z.D., Costa, J., and Haab, B.B. (2005). Optimized normalization for antibody microarrays and application to serum-protein profiling. *Mol. Cell. Proteomics* 4, 773–784.

Han, X.L., and Gross, R.W. (2005). Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. *Mass Spectrom. Rev.* 24, 367–412.

Harlow, E., and Lane, D. (1988). *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory, New York.

Harvey, D.J. (2003). Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates and glycoconjugates. *Int. J. Mass Spectrom.* 226, 1–35.

Haskins, W.E., Wang, Z.Q., Watson, C.J., Rostand, R.R., Witowski, S.R., Powell, D.H., and Kennedy, R.T. (2001). Capillary LC-MS2 at the attomole level for monitoring and discovering endogenous peptides in microdialysis samples collected in vivo. *Anal. Chem.* 73, 5005–5014.

Havlis, J., and Shevchenko, A. (2004). Absolute quantification of proteins in solutions and in polyacrylamide gels by mass spectrometry. *Anal. Chem.* 76, 3029–3036.

Hernandez, H., Dziembowski, A., Taverner, T., Seraphin, B., and Robinson, C.V. (2006). Subunit architecture of multimeric complexes isolated directly from cells. *EMBO Rep.* 7, 605–610.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sørensen, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W., Figeys, D., and Tyers, M. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* 415, 180–183.

Hochleitner, E.O., Kastner, B., Frohlich, T., Schmidt, A., Luhrmann, R., Arnold, G., and Lottspeich, F. (2005). Protein stoichiometry of a multiprotein complex, the human spliceosomal U1 small nuclear ribonucleoprotein – absolute quantification using isotope-coded tags and mass spectrometry. *J. Biol. Chem.* 280, 2536–2542.

Hua, S.B., To, W.Y., Nguyen, T.T., Wong, M.L., and Wang, C.C. (1996). Purification and characterization of proteasomes from Trypanosoma brucei. *Mol. Biochem. Parasitol.* 78, 33–46.

Huang, L., Jacob, R.J., Pegg, S.C.H., Baldwin, M.A., Wang, C.C., Burlingame, A.L., and Babbitt, P.C. (2001a). Functional assignment of the 20 S proteasome from Trypanosoma brucei using mass spectrometry and new bioinformatics approaches. *J. Biol. Chem.* 276, 28327–28339.

Huang, R.P., Huang, R.C., Fan, Y., and Lin, Y. (2001b). Simultaneous detection of multiple cytokines from conditioned media and patient's sera by an antibody-based protein array system. *Anal. Biochem.* 294, 55–62.

Hummon, A.B., Amare, A., and Sweedler, J.V. (2006). Discovering new invertebrate neuropeptides using mass spectrometry. *Mass Spectrom. Rev.* 25, 77–98.

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., and Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* 4, 1265–1272.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* 98, 4569–4574.

Jaffe, J.D., Mani, D.R., Leptos, K.C., Church, G.M., Gillette, M.A., and Carr, S.A. (2006). PEPPeR, a platform for experimental proteomic pattern recognition. *Mol. Cell. Proteomics* 5, 1927–1941.

Jensen, O.N. (2004). Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* 8, 33–41.

Karas, M., Bachmann, D., Bahr, U., and Hillenkamp, F. (1987). Matrix-assisted ultraviolet-laser desorption of nonvolatile compounds. *Int. J. Mass Spectrom. Ion Process* 78, 53–68.

Kelleher, N.L., Lin, H.Y., Valaskovic, G.A., Aaserud, D.J., Fridriksson, E.K., and McLafferty, F.W. (1999). Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *J. Am. Chem. Soc.* 121, 806–812.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74, 5383–5392.

Kjeldsen, F., Haselmann, K.F., Budnik, B.A., Sorensen, E.S., and Zubarev, R.A. (2003). Complete characterization of posttranslational modification sites in the bovine milk protein PP3 by tandem mass spectrometry with electron capture dissociation as the last stage. *Anal. Chem.* 75, 2355–2361.

Knuesel, M., Wan, Y., Xiao, Z., Holinger, E., Lowe, N., Wang, W., and Liu, X.D. (2003). Identification of novel protein–protein interactions using a versatile mammalian tandem affinity purification expression system. *Mol. Cell. Proteomics* 2, 1225–1233.

Krogan, N.J., Cagney, G., Yu, H.Y., Zhong, G.Q., Guo, X.H., Ignatchenko, A., Li, J., Pu, S.Y., Datta, N., Tikuisis, A.P., Punna, T., Peregrín-Alvarez, J.M., Shales, M., Zhang, X., Davey, M., Robinson, M.D., Paccanaro, A., Bray, J.E., Sheung, A., Beattie, B., Richards, D.P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M.M., Vlasblom, J., Wu, S., Orsi, C., Collins, S.R., Chandran, S., Haw, R., Rilstone, J.J., Gandi, K., Thompson, N.J., Musso, G., St Onge, P., Ghanny, S., Lam, M.H., Butland, G., Altaf-Ul, A.M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J.S., Ingles, C.J., Hughes, T.R., Parkinson, J., Gerstein, M., Wodak, S.J., Emili, A., and Greenblatt, J.F. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* 440, 637–643.

Kruger, M., Kratchmarova, I., Blagoev, B., Tseng, Y.H., Kahn, C.R., and Mann, M. (2008). Dissection of the insulin signaling pathway via quantitative phosphoproteomics. *Proc. Natl. Acad. Sci. USA* 105, 2451–2456.

Le Hir, H., and Andersen, G.R. (2008). Structural insights into the exon junction complex. *Curr. Opin. Struct. Biol.* 18, 112–119.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* 17, 676–682.

Lorentzen, E., Walter, P., Fribourg, S., Evguenieva-Hackenberg, E., Klug, G., and Conti, E. (2005). The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat. Struct. Mol. Biol.* 12, 575–581.

Luo, Q.Z., Shen, Y.F., Hixson, K.K., Zhao, R., Yang, F., Moore, R.J., Mottaz, H.M., and Smith, R.D. (2005). Preparation of 20-mu m-i.d. silica-based monolithic columns and their performance for proteomics analyses. *Anal. Chem.* 77, 5028–5035.

Major, M.B., Camp, N.D., Berndt, J.D., Yi, X.H., Goldenberg, S.J., Hubbert, C., Biechele, T.L., Gingras, A.C., Zheng, N., Maccoss, M.J., Angers, S., and Moon, R.T. (2007). Wilms tumor suppressor WTX negatively regulates WNT/beta-catenin signaling. *Science* 316, 1043–1046.

Mann, M. (2006). Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* 7, 952–958.

Mann, M., and Jensen, O.N. (2003). Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* 21, 255–261.

Mann, M., and Wilm, M. (1994). Error tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* 66, 4390–4399.

Marshall, A.G., Hendrickson, C.L., and Jackson, G.S. (1998). Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom. Rev.* 17, 1–35.

Medzihradszky, K.F., Campbell, J.M., Baldwin, M.A., Falick, A.M., Juhasz, P., Vestal, M.L., and Burlingame, A.L. (2000). The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* 72, 552–558.

Meyer, B., Wittig, I., Trifilieff, E., Karas, M., and Schagger, H. (2007). Identification of two proteins associated with mammalian ATP synthase. *Mol. Cell. Proteomics* 6, 1690–1699.

Nesvizhskii, A.I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data – the protein inference problem. *Mol. Cell. Proteomics* 4, 1419–1440.

Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75, 4646–4658.

Neubauer, G., Gottschalk, A., Fabrizio, P., Seraphin, B., Luhrmann, R., and Mann, M. (1997). Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl. Acad. Sci. USA* 94, 385–390.

Nickell, S., Kofler, C., Leis, A.P., and Baumeister, W. (2006). A visual approach to proteomics. *Nat. Rev. Mol. Cell Biol.* 7, 225–230.

Olsen, J.V., and Mann, M. (2004). Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. USA* 101, 13417–13422.

Ong, S.E., and Mann, M. (2005). Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* 1, 252–262.

Orchard, S., Salwinski, L., Kerrien, S., Montecchi-Palazzi, L., Oesterheld, M., Stumpflen, V., Ceol, A., Chatr-aryamontri, A., Armstrong, J., Woollard, P., Salama, J.J., Moore, S., Wojcik, J., Bader, G.D., Vidal, M., Cusick, M.E., Gerstein, M., Gavin, A.C., Superti-Furga, G., Greenblatt, J., Bader, J., Uetz, P., Tyers, M., Legrain, P., Fields, S., Mulder, N., Gilson, M., Niepmann, M., Burgoon, L., De Las Rivas, J., Prieto, C., Perreau, V.M., Hogue, C., Mewes, H.W., Apweiler, R., Xenarios, I., Eisenberg, D., Cesareni, G., and Hermjakob, H. (2007). The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat. Biotechnol.* 25, 894–898.

Paweletz, C.P., Charboneau, L., Bichsel, V.E., Simone, N.L., Chen, T., Gillespie, J.W., Emmert-Buck, M.R., Roth, M.J., Petricoin, E.F., and Liotta, L.A. (2001). Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* 20, 1981–1989.

Perkins, D.N., Pappin, D.J.C., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567.

Posewitz, M.C., and Tempst, P. (1999). Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* 71, 2883–2892.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* 24, 218–229.

Ranish, J.A., Hahn, S., Lu, Y., Yi, E.C., Li, X.J., Eng, J., and Aebersold, R. (2004). Identification of TFB5, a new component of general transcription and DNA repair factor IIH. *Nat. Genet.* 36, 707–713.

Ranish, J.A., Yi, E.C., Leslie, D.M., Purvine, S.O., Goodlett, D.R., Eng, J., and Aebersold, R. (2003). The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* 33, 349–355.

Rappsilber, J., and Mann, M. (2002). What does it mean to identify a protein in proteomics?. *Trends Biochem. Sci.* 27, 74–78.

Rappsilber, J., Siniossoglou, S., Hurt, E.C., and Mann, M. (2000). A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal. Chem.* 72, 267–275.

Riedel, C.G., Katis, V.L., Katou, Y., Mori, S., Itoh, T., Helmhart, W., Galova, M., Petronczki, M., Gregan, J., Cetin, B., Mudrak, I., Ogris, E., Mechtler, K., Pelletier, L., Buchholz, F., Shirahige, K., and Nasmyth, K. (2006). Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I. *Nature* 441, 53–61.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030–1032.

Rix, U., Hantschel, O., Duernberger, G., Rix, L.L.R., Planyavsky, M., Fernbach, N.V., Kaupe, I., Bennett, K.L., Valent, P., Colinge, J., Köcher, T., and Superti-Furga, G. (2007). Chemical proteomic profiles of the BCR-ABL inhibitor imatinib, nilotinib, and dasatinib, reveal novel kinase and nonkinase targets. *Blood* 110, 4055–4063.

Rodi, D.J., and Makowski, L. (1999). Phage-display technology – finding a needle in a vast molecular haystack. *Curr. Opin. Biotechnol.* 10, 87–93.

Roepstorff, P., and Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass-spectra of peptides. *Biomed. Mass Spectrom.* 11, 601–602.

Ross, P.L., Huang, Y.L.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D.J. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–1169.

Schirle, M., Heurtier, M.A., and Kuster, B. (2003). Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry. *Mol. Cell. Proteomics* 2, 1297–1305.

Sharan, R., and Ideker, T. (2006). Modeling cellular machinery through biological network comparison. *Nat. Biotechnol.* 24, 427–433.

Sharon, M., Taverner, T., Ambroggio, X.I., Deshaies, R.J., and Robinson, C.V. (2006). Structural organization of the 19S proteasome lid: insights from MS of intact complexes. *PLoS Biol.* 4, 1314–1323.

Shevchenko, A., Loboda, A., Ens, W., and Standing, K.G. (2000). MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Anal. Chem.* 72, 2132–2141.

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 1, 2856–2860.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal. Chem.* 68, 850–858.

Siuti, N., and Kelleher, N.L. (2007). Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* 4, 817–821.

Steen, H., and Mann, M. (2004). The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* 5, 699–711.

Stevens, R.C., Yokoyama, S., and Wilson, I.A. (2001). Global efforts in structural genomics. *Science* 294, 89–92.

Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., and Hunt, D. F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci. USA* 101, 9528–9533.

Taylor, C.F., Paton, N.W., Lilley, K.S., Binz, P.-A., Julian, R.K., Jones, A.R., Zhu, W., Apweiler, R., Aebersold, R., Deutsch, E.W., Dunn, M.J., Heck, A.J., Leitner, A., Macht, M., Mann, M., Martens, L., Neubert, T.A., Patterson, S.D., Ping, P., Seymour, S.L., Souda, P., Tsugita, A., Vandekerckhove, J., Vondriska, T.M., Whitelegge, J. P., Wilkins, M.R., Xenarios, I., Yates, J.R., 3rd, and Hermjakob, H. (2007). The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* 25, 887–893.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 60, 523–533.

Thomas, C.E., Kelleher, N.L., and Mizzen, C.A. (2006). Mass spectrometric characterization of human histone H3: a bird's eye view. *J. Proteome Res.* 5, 240–247.

Tornow, S., and Mewes, H.W. (2003). Functional modules by relating protein interaction networks and gene expression. *Nucleic Acids Res.* 31, 6283–6289.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J.M. (2000). A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. *Nature* 403, 623–627.

Uhlen, M., and Ponten, F. (2005). Antibody-based proteomics for human tissue profiling. *Mol. Cell. Proteomics* 4, 384–393.

Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol.* 3, 986–997.

Vasilescu, J., Guo, X.C., and Kast, J. (2004). Identification of protein–protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics* 4, 3845–3854.

Washburn, M.P., Wolters, D., and Yates, J.R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 19, 242–247.

Waugh, D.S. (2005). Making the most of affinity tags. *Trends Biotechnol.* 23, 316–320.

Westermarck, J., Weiss, C., Saffrich, R., Kast, J., Musti, A.M., Wessely, M., Ansorge, W., Seraphin, B., Wilm, M., Valdez, B.C., and Bohmann, D. (2002). The DEXD/H-box RNA helicase RHII/Gu is a co-factor for c-Jun-activated transcription. *EMBO J.* 21, 451–460.

Weston, A.D., and Hood, L. (2004). Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J. Proteome Res.* 3, 179–196.

Whitty, A. (2008). Cooperativity and biological complexity. *Nat. Chem. Biol.* 4, 435–439.

Wilm, M., and Mann, M. (1996). Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* 68, 1–8.

Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* 379, 466–469.

Zhang, H., Yan, W., and Aebersold, R. (2004). Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr. Opin. Chem. Biol.* 8, 66–75.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M., and Snyder, M. (2001). Global analysis of protein activities using proteome chips. *Science* 293, 2101–2105.

Zhu, H., Klemic, J.F., Chang, S., Bertone, P., Casamayor, A., Klemic, K.G., Smith, D., Gerstein, M., Reed, M.A., and Snyder, M. (2000). Analysis of yeast protein kinases using protein chips. *Nat. Genet.* 26, 283–289.

Zhu, H., and Snyder, M. (2003). Protein chip technology. *Curr. Opin. Chem. Biol.* 7, 55–63.

Zubarev, R.A. (2004). Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* 15, 12–16.

Zubarev, R.A., Kelleher, N.L., and McLafferty, F.W. (1998). Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* 120, 3265–3266.

This page intentionally left blank

Section II

Applications of Molecular Diagnostics and Related Issues

This page intentionally left blank

Chapter 22

Pharmacogenetics and Pharmacogenomics: Impact on Drug Discovery and Clinical Care

Minoli Perera[1] and Federico Innocenti[2]

[1]*Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA;*

[2]*Hematology/Oncology, Department of Medicine, University of Chicago, Chicago, IL, USA*

22.1 INTRODUCTION

Adverse drug effects account for approximately 100,000 deaths per year in the USA (Lazarou *et al*., 1998). It is the hope of many that adverse reactions and therapeutic failures will be a thing of the past with the advent of personalized medicine. Advances made over the last 30 years in molecular biology, molecular genetics and genomics, and in the development and refinement of associated methods and technologies have had a major impact on the understanding of the action of drugs and other biologically active xenobiotics. The tools that have been developed to allow these advances, and the knowledge of fundamental principles underlying cellular function, have become quintessential and indeed indispensable for almost any kind and field of biological research, including future progress in biomedicine and health care. Over the past few decades it has been known that genetic variability affects drug response and efficacy (Vesell and Page, 1968a, b). With the end of the Human Genome Project (Venter *et al*., 2001) and the wealth of information readily available online (Marsh *et al*., 2002), the promise of personalized medicine looms large, but the generalization into clinical applications of pharmacogenomics has been more challenging. This is mainly due to the complexity of both disease (disease heterogeneity, i.e. several phenotypes in a single disease) and genetic involvement (several genes contributing to a single phenotype), as well as the difficulty in prospective validation of pharmacogenetic markers in controlled trials.

In pharmacogenetics, most research involves determining the effect of polymorphisms (changes in the genetic code that occur in at least 1% of the population) on a given drug response. A polymorphism may be a single nucleotide change in the DNA sequence (single nucleotide polymorphism, SNP), an insertion or deletion of nucleotides or a change in the number of gene copies (copy-number variation or CNV; see also Chapter 12). While the focus of early pharmacogenetic investigations was oriented towards functional polymorphisms found in exons (i.e. coding regions), functional variation can be found in any area of the gene including introns, untranslated regions (UTRs), upstream of the transcriptional start site, and sometimes in nearby genes (Brenner and Duggan, 2004).

This chapter aims to provide the necessary information in regard to how the discipline of genetics and genomics can impact drug discovery and development, and more broadly to the practice of health care. Particular emphasis will be placed on examining the role of pharmacogenetics and pharmacogenomics with regard to common complex diseases. Readers will find this chapter a useful source for standard definitions for some of the terms that are in wide use today, almost always sorely missing from both academic and public policy-related documents on the topic. We will also describe in detail the use of pharmacogenetics in current clinical care, those pharmacogenetic candidates ''not quite ready for prime time'', the ethical and legal implications of pharmacogenetics, and finally we will provide an overview of the publicly available web-based genetic resources.

22.2 DEFINITIONS AND HISTORY

Pharmacogenetic research started from the observation that not all patients responded the same way to the same medication and that genetic differences between individuals may account for these observed differences (Vogel, 1959). Current drug development practices consider patient populations to come from a relatively homogeneous group (i.e. all hypertensive subjects will respond similarly to the

same medication). When differences in drug response or effect is anticipated, as in the case of organ (kidney and liver) impairment, specific studies are conducted. However, genetic variation may contribute an additional amount of the variability to drug response. So while drugs in Europe and the USA are usually tested in Caucasian patients with defined disease states, doses for the medication are marketed for all patients, without distinction of their geographical origin and the heterogeneity of certain diseases. Thus, the contribution of genetics to the variability in drug response is what the field of pharmacogenetics and pharmacogenomics aims to answer.

22.2.1 Pharmacogenetics Versus Pharmacogenomics

The term pharmacogenetics was first coined by Vogel in 1959 to describe the inheritance of an aberrant drug metabolism. Pharmacogenetics has since been defined as the study of the variation in drug response as it relates to an individual's genetic makeup. Pharmacogenomics usually refers to a broader use of genome-wide association studies (GWAS) and potential complex interactions as well as alteration in gene expression that correlates to drug response. Both sciences deal with the germ-line heritable effects of the patient's genetic variation on drug response, and their goals are overlapping.

22.2.2 History of Pharmacogenetics

Prior to the discovery of DNA and the effect of genetic polymorphisms, scientists made observation on the heterogeneous nature of human drug response. Garrod and coworkers (1931) stated there was such a thing as human biochemical individuality. His foresight into the tenants on pharmacogenetics can be seen in the following passage: ''Even against chemical poisons taken by mouth, or by other channels, there are some means of defense. Every active drug is a poison, when taken in large enough doses; and in some subjects a dose which is innocuous to the majority of people has toxic effects, whereas others show exceptional tolerance to the same drug'' (Garrod, 1931; Snyder *et al*., 1949). In 1932, the first large-scale study testing the ability of family members to taste phenylthiocarbamide was conducted and showed a clear autosomal recessive inheritance to what was then called ''taste blindness'' (Fox, 1932). This was also the first study to show an interethnic difference in the frequency of this phenotype. In the 1950s several case studies showed the role of genetic variation in the metabolism and response to drugs, such as variation in isoniazid acetylation (Hughes *et al*., 1954) and decreased cholinesterase activity in the action of succinylcholine, which leads to prolonged sedation (Kalow and Staron, 1957). These case studies raised the real possibility of pharmacologically heritable differences that may be clinically important.

In the paper by Arno Motulsky entitled ''Drug reactions, enzymes and biochemical genetics'' in 1957, the author outlined the genetic basis for the adverse reactions to primiquine and succinylcholine as well as barbiturate-precipitated attacks and hereditary hyperbilirubinemia (Motulsky, 1957). Even with these landmark studies and papers, the most well-known drug metabolizing defect discovered to affect drug metabolism was found by Matsunaga and coworkers (Matsunaga *et al*., 1989); the poor debrisoquine metabolizer phenotype found in 8% of Caucasian livers was attributed to a deficiency in CYP2D6. The phenotype for this deficiency was found independently by two groups. In London, Dr. Robert Smith and some of his coworkers ingested 32 mg of debrisoquine (an obsolete anti-hypertensive drug); Dr. Smith experienced prolonged orthostatic hypotension hours after the dose (Smith, 1986). This effect was not seen in his coworkers. A subsequent study in medical students and three families solidified the ''poor'' and ''extensive'' metabolizer phenotype (Mahgoub *et al*., 1977). During this same time, Michael Eichelbaum was studying the pharmacokinetics on sparteine (an antiarrhythmic agent) when two individuals experienced nausea, diploplia, and headaches (Eichelbaum *et al*., 1979). It was later confirmed that these two independent observations were the result of the same genetic deficiency. However, the molecular genetics of this apparently simple phenotypic defect was found to be complex, with frameshift mutation, splicing defects, and a stop codon resulting in absent enzyme activity of the CYP2D6 enzyme. SNPs within the gene result in low enzyme activity, and gene duplication results in high enzyme activity. Because of its importance in the metabolism of several prescribed drugs and the dramatic clinical manifestation seen with aberrant metabolism, *CYP2D6* has become one of the most studied pharmacogenetic genes (see section 22.4.4.2 for the most recent application of this phenotype).

These case studies and investigations laid the ground work for what is now the modern pharmacogenetic discipline. The advances in the field have led to Food and Drug Administration (FDA) label changes (see section 22.4.3), a guidance for industry and FDA staff on the pharmacogenetic and genetic tests for heritable markers (see http://www.fda.gov/cdrh/oivd/guidance/1549.pdf), and to unique methodology and the beginning of understanding of the genetic basis of complex diseases.

22.2.3 Analytical Methods in Pharmacogenetics

Currently, a variety of approaches are used to associate genotype (the specific nucleotide combination at a given position in the genome) with phenotype (an observable

physical or biochemical characteristic), such as decreased drug clearance or increased drug toxicity, as well as altered efficacy of a medication. Figure 22.1 provides some common definitions used in pharmacogenetics and pharmacogenomics. Some commonly used strategies include: (1) candidate gene approach (i.e. phenotype/genotype or phenotype/haplotype association studies) in which SNPs are correlated to phenotype, (2) candidate pathway-based studies, which investigate several genes with respect to the pharmacology of a specific drug, disease, or biological pathway for variations that may contribute to phenotypic differences, or (3) GWAS, in which thousands of SNPs are spaced throughout the genome. Each of these methods has advantages and disadvantages. GWAS comprehensively survey the genetic variation in the entire genome, but their power to detect moderate associations is greatly limited because 500,000 to 1,000,000 SNPs may be typed leading to multiple statistical tests, thus requiring correction for false positive associations. In addition, they require large sample sets to achieve adequate power. However, these studies do provide a unique opportunity to discover novel genes not previously known to effect drug response. The candidate gene approach narrows the focus to a few important genes/SNPs and therefore has higher power. Because there is a rationale to the selection of these genes, the SNPs found are usually biologically plausible. However, candidate gene studies may miss the real causative SNP and are limited to the available knowledge.

Within the candidate gene approach, the haplotype/phenotype approach in which haplotypes (i.e. groups of SNPs closely linked on one chromosome and inherited as a unit) are correlated to phenotype holds the added advantage of assaying for unknown variation in linkage disequilibrium (LD) with the genotyped SNPs. LD refers to the non-random association of alleles at two or more loci, meaning alleles that are located close to each other on the same chromosome have the tendency to occur and be inherited together. This also means that a small number of SNPs can be used to capture information about a larger region. This is commonly referred to as a haplotype tagging approach. Regardless of the approach used, all positive findings need to be replicated in independent data sets to confirm their validity. There are several current reviews of statistical methods in pharmacogenetics (Balding, 2006; Montana, 2006).

22.3 PHARMACOGENETICS IN DRUG DEVELOPMENT

The advent of pharmacogenetics has had a mixed response in the pharmaceutical industry, with many executives expressing concern that the technology remains unproven, expensive, and disruptive to the overall revenue in drug sales (Danzon and Towse, 2002; Hopkins *et al*., 2006). There are foreseeable benefits to the addition of pharmacogenetics into the drug development process. Pharmaceutical companies want to get their new drugs out to market sooner with reduced risk of failure. A prime example of the benefit pharmacogenetics could have on the drug development processes is seen in the case of troglitazone (an oral hypoglycemic agent use to treat type II diabetes). This drug was taken off the market due to severe and sometimes fatal hepatotoxicity. A pharmacogenetic case control study of troglitazone-induced hepatotoxicity showed strong correlation between elevated liver enzymes and functional mutations in GSTT1 and GSTM1 (phase II metabolic enzymes; Watanabe *et al*., 2003). Early evaluation of these pharmacogenetic variants may have saved a therapeutically important drug. Currently, most large pharmaceutical companies have an internal pharmacogenetics program, but these programs

Genotype – The specific nucleotide combination at a given position in the genome.

Phenotype – An observable physical or biochemical characteristic, such as decreased drug clearance or increased drug toxicity.

Allele – One of the variant forms of a gene at a particular location (i.e. locus) on a chromosome.

Polymorphism – Difference in DNA sequence among individuals. Genetic variations occurring in more than 1% of a population.

Single nucleotide polymorphism (SNP) – A genetic marker resulting from variation in sequence at a particular position within a DNA sequence. SNPs can result from a base transition (A for G, T for C), transversion (G or A for T or C), or single-base deletion.

Haplotype – Groups of SNPs closely linked on one chromosome and inherited as a unit.

Linkage disequilibrium (LD) – Non-random association of alleles at two or more loci, meaning alleles that are located close to each other on the same chromosome have the tendency to occur and be inherited together.

FIGURE 22.1 Common definitions in pharmacogenetics.

are utilized to weed out poor candidates in the drug development process. Therefore, a large amount of the publicly available pharmacogenetic data in the field still comes from academia. Part of the reluctance from the drug industry comes from the concern that the FDA will delay the registration process. Steps within the FDA were taken in March 2005 to encourage the voluntary submission of pharmacogenetic data with the guidelines emphasizing the neutral effect on the registration processes (US Department of Health and Human Services, 2005). Hopefully this will encourage the collection of pharmacogenetic data on the large Phase III cohorts used in the approval process. The European Agency for the Evaluation of Medicinal Products (EMEA) has created new internal structures to support pharmacogenetic development, with guidance to sponsors provided through briefing meetings (European Medicines Agency, 2007). Yet within Europe the German, Irish, Dutch, and UK national agencies have received few requests from sponsors to consider pharmacogenetic data.

22.3.1 Pharmacokinetics and Pharmacodynamics

Most pharmacogenetic investigation involves genes or pathways related to pharmacokinetics and pharmacodynamics. Pharmacokinetic effects are due to interindividual differences in absorption, distribution, metabolism (with regard to both activation of pro-drugs, inactivation of the active molecule), or excretion of the drug. The pharmacokinetic genes are obvious targets for pharmacogenetic research given that drug levels are easily measured and, for some drugs, correlated to drug response (i.e. therapeutic drug monitoring). Differential effects caused by variation in pharmacokinetic genes are due to the presence at the intended site of action of either inappropriate concentrations of the pharmaceutical agent, or inappropriate metabolites, or both, resulting either in lack of efficacy or toxic effects or both. Pharmacogenetics, as it relates to pharmacokinetics, has been recognized as an entity for more than 100 years, going back to the observation commonly credited to Archibald Garrod that a subset of psychiatric patients treated with the hypnotic, sulphonal, developed porphyria. Since then, the underlying genetic causes for many of the previously known differences in enzymatic activity have been elucidated, most prominently with regard to the P450 enzyme family, and these have been the subject of many reviews (Evans and Relling, 2004; Ingelman-Sundberg *et al*., 2007; see also http://www.imm.ki.se/CYPalleles; Table 22.1). However, such pharmacokinetic effects are also seen with membrane transporters, such as in the case of differential activity of genetic variants of the *ABCB1* gene that affect the effective intracellular concentration of antiretrovirals (Fellay *et al*., 2002), anticancer agents (Nooter *et al*., 1990), and anticonvulsants (Kwan and Brodie, 2005), or of the purine-analog-metabolizing enzyme thiomethyl-purinetransferase (Dubinsky *et al*., 2000). Table 22.1 outlines several examples of phase I and phase II drug metabolizing enzymes of pharmacogenetic relevance for drug therapy.

Pharmacodynamic effects, in contrast, may lead to interindividual differences in a drug's effects despite the presence of appropriate concentrations of the active drug compound at the intended site of action. Genetic variation in how the target molecule or another (downstream) member of the target molecule's mechanistic pathway can respond to the medicine modulates the effects of the drug. Table 22.2 gives examples of some of the pharmacodynamic genes that have been shown to effect drug response.

Figure 22.2 is an example of how genetic variation in a complex disease affects treatment; in this example hypertension (phenotype, blood pressure). In this example the phenotype is composed of three mechanisms, in which only one is dysfunctional and causes symptomatic disease. A palliative treatment may address one of the pathways that, though not dysregulated, contributes to the overall deviant physiology (see Fig. 22.2; lane F); the respective pharmacogenetic/pharmacodynamic scenario would occur if this particular pathway, due to a genetic variant, was not responsive to the drug chosen (see Fig. 22.2; lane G). If a genetic variant changes the overall contribution of each mechanism to the overall phenotype as in lane H (i.e. M2 is impaired because of a genetic polymorphism, M3 is upregulated to maintain homeostasis), a treatment targeting M2 will not yield the expected results (see Fig. 22.2; lane F), but will produce little change in the phenotype (see Fig. 22.2; lanes I and J).

22.4 PHARMACOGENETICS IN CLINICAL CARE

22.4.1 Population Differences

Difference in drug response between ethnic populations was seen in the early 1920s, when differential cardiac effects of atropine were reported between Caucasians and African Americans (Paskind, 1921). Similar population-specific effects were seen with antimalarial medication used in African American soldiers (Motulsky, 1960), and pupillary dilation after various mydriatic eye drops used in Caucasians, Asians, and African Americans (Chen and Poth, 1929).

These differences in drug metabolism and response may have arisen for several different reasons, mainly, different allele frequencies and/or haplotype structures between populations derived from (1) population-specific mutations which affect the enzyme function, (2) mutations in a specific population, which garners an evolutionary advantage, or (3) mutations arising after populations have separated. Much of the inter-ethnic variation we see has to do with the

TABLE 22.1 Examples of pharmacogenetic effects in pharmacokinetic genes.

Enzymes	Effected drugs/substances
Phase I enzymes	
Aldehyde-dehydrogenase	Acetaldehyde
Alcohol-dehydrogenase	Ethanol
CYP1A2	Caffeine
CYP2A6	Nicotine, coumarin
CYP2C9	Phenytoin, warfarin, tolbutamide, glipizide, fluvastatin
CYP2C19	Mephenytoin, omeprazole, pantoprazole, amitryptyline, imipramine
CYP2D6	Dextromethorphan, debrisoquine, sparteine, tamoxifen, nortryptyline, clomipramine
CYP2E1	Chloroxazone, caffeine
CYP3A4	Erythromycin
CYP3A5	Tacrolimus, saquinavir
Serum cholinesterase	Benzoylcholine, butyrylcholine
Paraoxonase/arylesterase	Paraoxon
Phase II enzymes	
Acetyltransferase (NAT2)	Isoniazid, sulfamethazine, caffeine, procainamide, dapsone
Dihydropyrimidine-dehydrogenase	5-Fluorouracil
Organic anion transport protein 1B1 (SLCO1B1)	Pravastatin
Thiomethyltransferase	2-Mercaptoethanol, D-penicillamine, captopril
Thiopurine-methyltransferase (TPMT)	6-Mercaptopurine, 6-thioguanine, azathioprine
UDP-glucuronosyl-transferase 1A1 (UGT1A1)	Bilirubin, irinotecan
UDP-glucuronosyl-transferase 2B7 (UGT2B7)	Oxazepam, ketoprofen, estradiol, morphine

early human migration out of Africa into Asia and Europe (Tishkoff and Williams, 2002). The genetic variation seen outside Africa is generally thought to be a subset of the variation found within Africa (Cavalli-Sforza and Feldman, 2003). This migration caused populations to be genetically isolated and allowed population-specific mutations to arise. Mutations which cause decreased enzyme function, such as glucose-6-phosphate dehydrogenase deficiency, posed an evolutionary advantage. In this case, this variation confers protection against malaria, and therefore rose in frequency in African populations (Motulsky, 1960). It is also widely recognized that populations of recent African ancestry tend to have considerably more genetic variation (i.e. more sites within the genes are likely to be polymorphic), and the extent of LD can be much lower than in populations of recent European or Asian ancestry (Tishkoff and Verrelli, 2003), meaning that the length of the haplotypes found in African Americans is significantly shorter that Caucasians and would require typing more haplotype tagging SNPs to obtain similar assessment of the genetic variation. Therefore, pharmacogenetic studies based on Caucasian subjects may not always translate into an observed affect in another population, and should be taken into consideration when designing pharmacogenetic studies.

The genetic definition of race or ethnicity has usually corresponded to the continental or subcontinental regions which include the following populations: sub-Saharan Africans, Europeans, western Asians, northern Africans, eastern Asians, Polynesians and other inhabitants of Oceania, and Native Americans (Rosenberg *et al*., 2002; Bamshad *et al*., 2003). However, 5–15% of genetic variation occurs between these groups, with the remaining variation occurring within such groups (Jorde *et al*., 2000). This makes classification of individuals into distinct racial

TABLE 22.2 Examples of pharmacogenetic effects in pharmacodynamic genes.

Gene	Affected drugs/disease
5-Lipoxygenase	Monolukast (asthma)
ApoE2	LDL-C level
ACE	ACE inhibitors (hypertension)
Beta 2 adrenergic receptor	Albuterol (asthma)
Beta 1 adrenergic receptor	Beta blockers (hypertension)
Corticotropin-releasing hormone	Antidepressants
Dopamine receptor	Selective serotonin reuptake inhibitors (depression)
EGFR	Gefitinib, cetuximab (cancer)
ER alpha	Bone mineral density
FCGR3A	Infliximab (Crohn's disease)
HMGCoA reductase	Statins (cholesterol lowering drugs)
NEF3	Response to antipsychotics
TNF	Anti-TNF therapy (Crohn's disease and rheumatoid arthritis)
Serotonin receptor	Selective serotonin reuptake inhibitors (depression)
VKORC1	Warfarin dose (anticoagulation)

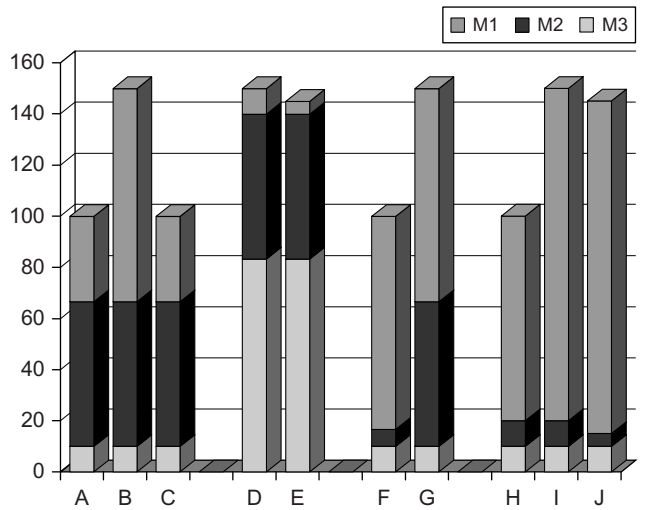


FIGURE 22.2 Three molecular mechanisms (M1, M2, M3) contribute to the phenotype. **Lane A**: Normal physiology. **Lane B**: Diseased physiology caused by disruption of M1. **Lane C**: Diseased physiology after treatment targeted at M1. **Lane D**: Diseased caused by disruption of M3. **Lane E**: Diseased physiology identical to that seen in lane D, after treatment targeted at M1: treatment does not address cause. **Lane F**: Diseased physiology identical to that seen in lane B, after treatment targeted at M2. While the treatment does not address the abberent pathway, the pheotype is returned to normal. **Lane G**: Diseased physiology identical to that seen in lane B, after treatment targeted at M2. However, a genetic variant in M2 results in a less than optimal response and the phenotype remains above normal levels. **Lane H**: Normal physiology. However, a genetic variant causes differential contribution of M1 and M2 to the normal phenotype. **Lane I**: Diseased physiology seen in lane B; however, with the differential contribution of M1 and M2 caused by a genetics variant. After treatment targeting M1, no significant change is seen in the phenotype. **Lane J**: Diseased physiology identical to that in lane I after treatment targeting M2. No significant change is seen in the phenotype. Solid bars: Normal function; stippling: pathological dysfunction; hatching: therapeutic modulation (from Lindpaintner, 2003; with permission).

categories difficult and often inaccurate. This is evident in the case of people who self-identify as African Americans. This ''race'' has some European ancestry, as measured through genetic markers, ranging from an estimated 7% for a sample of Jamaicans to 23% for a sample of African Americans from New Orleans (Parra *et al*., 1998). A recent commentary highlights the difficulty of the categorization of Asians (Po, 2007). Drug response and efficacy may vary widely in these populations on patients; however, categorical distinction in both pharmacogenetic studies and drug recommendations are rarely made.

It should be noted that several pharmacogenetic-relevant genes show population-specific differences. These differences may explain differences in drug tolerability, efficacy, and disease susceptibility. A specific example can be seen with the drug metabolizing enzyme CYP3A5. A common splice variant in the *CYP3A5* gene results in a nonfunctional enzyme (named *CYP3A5*3*). This variant is quite common in European populations, but less common in Africans and African Americans. This means that African populations have a higher frequency of the functional enzyme than non-African populations (Kuehl *et al*., 2001). Among several examples of the clinical role of this variant, this polymorphism has been associated with increased tacrolimus dose in transplant patients (Op den Buijsch *et al.*, 2007). This may explain differences seen in organ transplantation between Caucasians and African Americans (Dirks *et al*., 2004). It should be noted that accounting for population differences has been conducted on relatively few drugs. Therefore the gap seen in the pharmacogenetic world and that in the clinical trial and clinical practice setting is yet to be bridged.

22.4.2 Complex Phenotypes

With the completion of the Human Genome Project, researchers have found that SNPs are ubiquitous, occurring approximately once in every 1,000–3,000 base pairs (Sachidanandam *et al*., 2001). However, determining which of these SNPs is responsible for variation in drug response has been a daunting task. This association process becomes increasingly complicated when more than one gene is involved in the pharmacologic effect seen.

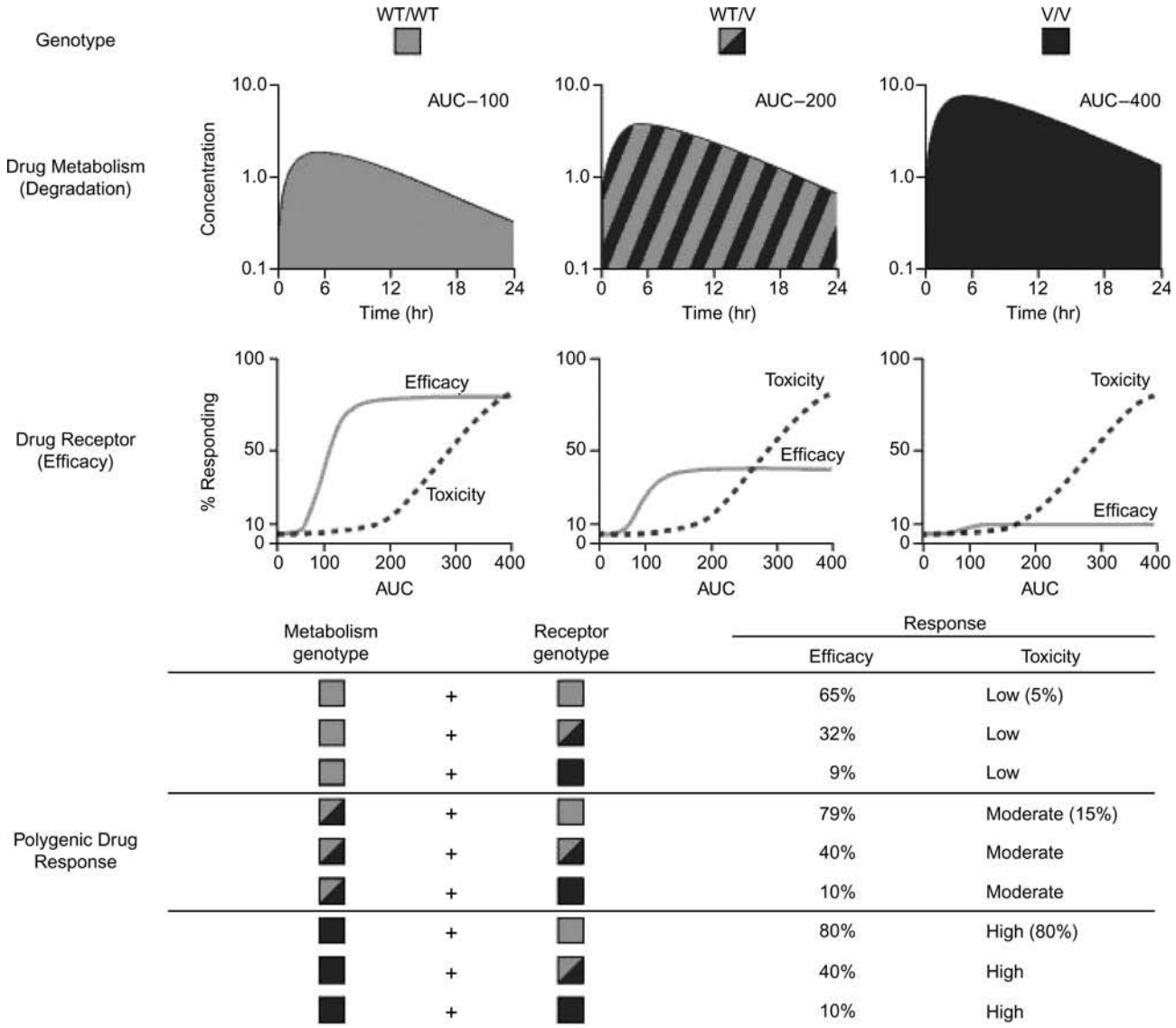


FIGURE 22.3 This figure from Evans and McLeod (2003) depicts the different roles genetic polymorphisms can play in drug response. The first row shows the effect of genetic variation in drug metabolizing enzymes, with an increasing area under the curve (AUC) associated with the variant allele (V). The second row shows the effect of genetic variation on drug targets, with decreased efficacy associated with the V allele. The combination of both types of variation leads to the therapeutic effect and/or the toxicity seen in patients. In essences, these two SNPs in combination can result in nine different phenotypes (Evans and McLeod, 2003).

Figure 22.3 depicts the range of effects that can be seen from polymorphisms in genes that affect the pharmacokinetics (i.e. drug metabolizing enzymes) and/or the drug target. The combination of these effects leads to a wide range of phenotypes. These added complexities require not only knowledge of gene–drug interaction but also of the entire pathway affected by the drug.

Common complex diseases, and thus the vast majority of what is to be clinically applied genetics, behave almost fundamentally differently from rare, classic, monogenic, Mendelian diseases: whereas in the latter the impact of the genetic variant is typically categorical in nature due to high penetrance, in the former case the presence of a disease-associated genetic variant is merely of probabilistic value, raising (or lowering) the likelihood of disease occurrence to some extent, but never predicting it in a black-and-white fashion. This becomes evident in a pathway-based view of the problem of drug response. While differences in drug metabolism and mutations within drug targets contribute to the variability seen, other factors such as polymorphism in regulatory genes (i.e. transcription factors), down-regulation of essential hormone mediators, and abnormalities in promoter sites, to name just a few, contribute to the drug response variability. These variables still do not take into account the effect of environmental factors. Therefore, for a pharmacogenetic test to be viable in this atmosphere, all these variables must, in some form, be accounted for. An example of the role non-genetic factors along with genetic

factors play in pharmacogenetic testing can be seen in warfarin dosing discussed in section 22.4.3.3.

Communicating this difference to a public that has long been misled into a perception of everything genetic being of deterministic, Mendelian quality, represents a second, no less important and difficult challenge. Unless this effort is successful, by engaging in a true dialog with all stakeholders, in providing the basis for informed discourse and sensible decision-making on the societal level, the full potential of our deepening understanding of biology and of these technological advances will not be recognized.

Inroads into understanding the genetic basis of complex diseases have been made through the advent of the GWAS. While criticized by some for being non-hypothesis driven (i.e. no *a priori* hypothesis is given on which gene and/or variant contributes to the disease in question), GWAS are revolutionary because they permit interrogation of the entire human genome in thousands of unrelated individuals, unconstrained by prior hypotheses regarding genetic associations. The past 5 years have been flush with replicated findings in GWAS. These include complex diseases such as type II diabetes (Zeggini *et al*., 2008), obesity (Loos *et al*., 2008), and Crohn's disease (Rioux *et al*., 2007) as well as prostate (Eeles *et al*., 2008), breast (Easton *et al*., 2007), colon (Tenesa *et al*., 2008), and lung cancer (Hung *et al*., 2008). GWAS has also identified important pharmacogenetic associations, as seen in the recent paper which showed an association between an SNP in *SLCO1B1* (encoding an anionic-transporting polypeptide) and myopathy after statin use (a cholesterol lowering agent) (Link *et al*., 2008). These investigations have led to the association of important susceptibility loci, such as the *NOD2* in Crohn's disease and *FTO* for obesity. One of the most important outcomes of these studies has been the discovery of new biological associations in genes or regions previously unrecognized to have a role in each disease. It is interesting to note that many of the associations found have relatively small effect, with typical odd ratios of about 1.5 to 2.0. These studies have also found that a substantial amount of the association to disease arose from regions outside of the coding gene regions. This is not surprising given that non-coding variation may effect gene regulation and is preponderant of coding regions in the human genome. Several groups are currently working on GWAS applied to clinical pharmacogenetics and these results are expected to be available in the public domain in the near future.

22.4.3 Pharmacogenetic Information in Drug Labels and FDA Label Changes in Dosing

Currently, the field of clinical pharmacogenetics has shown the greatest utility in oncology, where many drugs show a narrow therapeutic index. In oncology, the two major examples of the advancing role of pharmacogenetics in the prevention of drug toxicities are *TPMT* and 6-mercaptopurine (6-MP) along with *UGT1A1* and irinotecan, both of which resulted in drug label changes. The most recent FDA label change has been the addition of pharmacogenetic testing recommendations for *CYP2C9* and *VKORC1* to the drug label of warfarin. In general, these changes state a recommendation to test and not a requirement.

Approximately 10% of labels for drugs approved by the FDA contain pharmacogenetic information. In the context of drug labels, this information can be classified on the basis of their specific use such as clinical response and differentiation, risk identification, dose selection guidance, susceptibility, resistance and differential disease diagnosis, and polymorphic drug targets. It should be noted that many of the labels containing pharmacogenetic information do not provide recommendations for a specific action (i.e. genetic testing). A complete list of valid genetic biomarkers can be found at the FDA website (http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm).

22.4.3.1 UGT1A1 and Irinotecan

Currently, there are over 60 known polymorphisms in the *UGT1A1* gene (Hasegawa *et al*., 2006), many of which have functional consequences (Tukey and Strassburg, 2000). Because of its role in bilirubin clearance, several investigators identified polymorphisms within *UGT1A1* that contribute to a common benign familial condition of decreased bilirubin glucuronidation (Gilbert's syndrome) (Bosma *et al*., 1995; Monaghan *et al*., 1996).

The *UGT1A1*28* polymorphism, which is an insertion of an extra thymine-adenine (TA) repeat in the promoter, is associated with enzyme expression inversely related to repeat length. This insertion polymorphism affects the TATA box upstream of *UGT1A1* transcription initiation site, which is responsible for the binding of general transcription factor IID, which plays an important role in the initiation of transcription (Bosma *et al*., 1995). Therefore, individuals that are homozygous for seven TA repeats (i.e. $(TA)_7$, or *UGT1A1*28*) have a 70% reduction in UGT1A1 gene expression compared to those possessing the $(TA)_6$ allele (Beutler *et al*., 1998). *UGT1A1*28* is thought to explain up to 40% of the variability in *in vitro* enzyme activity of UGT1A1 (Peterkin *et al*., 2007). Two additional alleles at this location have been identified in persons of African ancestry, $(TA)_5$ (*UGT1A1*36*) and $(TA)_8$ (*UGT1A1*37*). Another common functional polymorphism found in exon 1 of the gene, denoted as *UGT1A1*6*, is found in Asian populations, and results in the substitution of an arginine for a glycine (Aono *et al*., 1995). The allelic frequency of *UGT1A1*6* in Asians ranges from 18 to 23% (Akaba *et al*., 1998; Takeuchi *et al*., 2004), with a 40% reduction in enzyme activity as compared to the wild-type enzyme (Jinno *et al*., 2003). A complete list of all

known functional polymorphisms can be found at the UGT nomenclature home page (Bock *et al*., 2005).

Recent evidence points to a more complex picture of the metabolism by UGT1A1 and functional *UGT1A* polymorphisms. Because of the structure of the *UGT1A* locus (unique exons 1 for each isoform with common exons 2 through 5), haplotypes may exist between the coding regions at this locus that affect the function of one or more of the UGT1A enzymes. A haplotype effect over the effect of *UGT1A1*28* alone at this gene locus has been recently postulated (Paoluzzi *et al*., 2004; Sai *et al*., 2004; Innocenti *et al*., 2005; Lankisch *et al*., 2005).

Irinotecan is approved for the treatment of metastatic colorectal cancer, and is often used in combination with 5-fluorouracil, leucovorin, bevacizumab, and/or cetuximab. Irinotecan is a pro-drug, with therapeutic activity found in its active metabolite, 7-ethyl-10-hydroxycamptothecan (SN-38), which is a potent topoisomerase I inhibitor (Gupta *et al*., 1994). The SN-38 AUC has been correlated to neutropenia (Pitot *et al*., 2000). The major route of clearance for SN-38 is glucuronidation to SN-38 glucuronide (SN-38G) by UGT1A1, although it is possible that UGT1A9 and UGT1A7 (extrahepatic) may also have a role (Gagne *et al*., 2002).

The initial clinical association between SN-38 glucuronidation and irinotecan toxicity was described by Gupta and coworkers (Gupta *et al*., 1994). The evidence that this was due to variability in *UGT1A1* came from two case reports in patients with Gilbert's syndrome (Wasserman *et al*., 1997), as well as studies by Iyer and coworkers (1999), demonstrating that SN-38 is a substrate for UGT1A1. Furthermore, the extent of metabolism was inversely correlated to the number of TA repeats in the promoter region (Iyer *et al*., 1999). The association between *UGT1A1**28 and the risk of severe diarrhea and/or neutropenia in patients receiving irinotecan was seen in a retrospective study by Ando and coworkers (Ando *et al*., 2000). Iyer and coworkers (2002) conducted the first prospective trial of irinotecan pharmacogenetics, which showed an association between *UGT1A1*28* and decreased SN-38 glucuronidation rate. This study was amended to increase the dose of irinotecan to 350 mg/m^2 (Innocenti *et al*., 2004). Under the revised protocol, by Innocenti *et al*. (2004), the investigators found a significant association of *UGT1A1*28* with incidence of grade 4 neutropenia. Following this initial work, several studies found association of *UGT1A1*28* and pharmacokinetic parameters (i.e. decreased SN-38G/SN-38 area under the plasma concentration/time curve (AUC) ratio, and SN-38 AUCs) (Paoluzzi *et al*., 2004; Sai *et al*., 2004; Araki *et al*., 2006) and drug toxicity such as neutropenia (Ando *et al*., 2000; Rouits *et al*., 2004; Toffoli *et al*., 2006), and diarrhea (Marcuello *et al*., 2004; Massacesi *et al*., 2006). For a recent review on these studies, see Kim and Innocenti (2007).

This evidence led to the inclusion by the FDA of *UGT1A1*28* as a risk factor for severe neutropenia in patients receiving irinotecan (http://www.fda.gov/cder/foi/label/2005/020571s024,027,028lbl.pdf). In August 2005, the FDA approved the marketing of the Invader UGT1A1 assay (Third Wave Technologies, Madison, WI) as a diagnostic test for the *UGT1A1*28* allele. *UGT1A1* genotype results can also be obtained at off-site facilities, such as Mayo Medical Laboratories or ARUP Laboratories which process whole blood samples and return genotype data directly to the physician.

22.4.3.2 TPMT and 6-Mercaptopurine

6-Mercaptopurine (6-MP) and azathioprine are used to treat a variety of autoimmune disorders and to prevent organ rejection as well as in the treatment of childhood acute lymphoblastic leukemia (ALL), rheumatoid arthritis, and inflammatory bowl diseases. Azathioprine is converted into 6-MP, and thiopurine methyltransferase (TPMT) is involved in the methylation reactions of 6-mercaptopurine, preventing the formation of its active metabolites. The clinical importance of the pharmacogenetics of *TPMT* has been clearly established in relation to the prediction of severe, and rarely fatal, hematological toxicity in childhood ALL patients with reduction or loss of TPMT activity. In the absence of the information on the TPMT status of patients, the drug dosage is adjusted by assessing the myelosuppressive effect of the therapy, as it is required that patients experience moderate myelosuppression. In some patients, bone marrow depression is severe, exposing them to the risk of life-threatening infections. In other patients, myelosuppression is less severe, but requires the interruption of therapy or dose reductions, until the bone marrow has recovered.

Up to 17 variant alleles in the *TPMT* gene have been identified, most of them correlated with a low TPMT activity. The non-synonymous variants *3C (719A>G), *3A (460G>A and 719A>G) and *2 (238G>C) are common (2–10%) and account for over 95% of the alleles associated with reduced TPMT activity (Wang and Weinshilboum, 2006). About 10% of ALL patients are heterozygous and 0.3% are homozygous for the deficient alleles. As a consequence, heterozygous ALL patients are at intermediate risk of exaggerated bone marrow toxicity. In these patients, a 35% dose reduction of 6-MP has been demonstrated to be safe (Relling *et al*., 1999). In ALL patients, who are homozygous for defective alleles, a 90% dose reduction is required (Evans *et al*., 1991, 2001).

It is important to understand the impact of the heterozygous state (which occurs in 10% of patients) on toxicity and survival compared to the patients without inactivating *TPMT* variants. Dose adjustment of 6-MP in heterozygous ALL patients preserved the efficacy of 6-MP (Relling *et al*., 2006) and another study has shown that *TPMT* heterozygosity has been associated with a 2.9-fold lower risk of minimal residual disease (an important prognostic factor of early response) compared to patients with no deficient

TPMT allele (Stanulla *et al*., 2005). The identification of TPMT-deficient patients at the beginning of the maintenance treatment might lead to optimization of therapy by (1) avoiding severe myelosuppression and (2) tailoring the 6-MP dose to maintain an adequate level of exposure throughout the course of the treatment, maintaining the dose intensity of the regimen. The label of 6-MP has been recently revised to contain information about substantial dose reduction in *TPMT* homozygous patients (Yong and Innocenti, 2007).

22.4.3.3 CYP2C9/VKORC1 and Warfarin

Because of its narrow therapeutic index and serious side-effect profile, warfarin has been the most recent drug to change its drug label to recommend genotype-guided dosing. Several studies have looked at the effect of *CYP2C9* variants on warfarin dose and on adverse events (Aithal *et al*., 1999; Veenstra *et al*., 2005a; Moridani *et al*., 2006), as CYP2C9 is the main enzyme of the inactivation of the active S enantiomer of warfarin (Breckenridge *et al*., 1974). *CYP2C9* is part of the 4-gene *CYP2C* gene cluster that is located on chromosome 11 (Gray *et al*., 1995). Most studies have focused on the common functional variants in *CYP2C9*, named *CYP2C9*2* and *CYP2C9*3*. These variants require significantly lower maintenance doses of warfarin, approximately 0.85 mg for *CYP2C9*2*, 1.92 mg for *CYP2C9*3*, and 1.47 mg for both alleles, from the mean daily dose (Sanderson *et al*., 2005), and have higher risk of serious bleeds. Both of these alleles have low allele frequency (approximately 10%) in Caucasians, and even lower frequency in African Americans (2–5%) and are absent in Chinese and Japanese subjects (Wang *et al*., 1995; Nasu *et al*., 1997). The *CYP2C9*5* allele was found in healthy African American volunteers, and has been hypothesized to slow the metabolism of warfarin (Dickmann *et al*., 2001).

In addition to *CYP2C9* genotypes, recent studies have shown that variation in the *VKORC1* gene accounts for 20 to 30% of the observed population variation in warfarin dose (Rieder *et al*., 2005; Sconce *et al*., 2005; Veenstra *et al*., 2005b). Vitamin K epoxide reductase (VKOR), the therapeutic target of warfarin, is encoded by the gene *VKORC1*, in which rare mutations have been associated with warfarin resistance (Rost *et al*., 2004). The recent publication by Rieder and coworkers (2005) highlights the growing effort to incorporate this new gene into warfarin dosing. The investigators found ten common SNPs in three populations that formed nine haplotypes. An association study was then conducted (excluding Asians and African Americans) to determine the effect of these haplotypes on warfarin dose. The results showed that two haplotypes were associated with low warfarin dose (haplogroup A) and three haplotypes were associated with high warfarin dose (haplogroup B). As mentioned previously, these two genes are an ideal example of the range of phenotypes seen when genetic variation affects both pharmacokinetics and the drug target. From this study we can see that patients possessing the wild-type *CYP2C9* and the *VKORC1* B/B haplogroup required the highest doses, while patients possessing a variant *CYP2C9* allele and *VKORC1* haplogroup A/A required the lowest doses.

Several algorithms have been created that determine the maintenance dose of warfarin using both genetic (*VKORK1* and *CYP2C9*) and non-genetic (age, weight, co-administered medications) factors (Sconce *et al*., 2005; Takahashi *et al*., 2006; Caldwell *et al*., 2007; Millican *et al*., 2007). These algorithms explain approximately 50% of the variability in warfarin dose (Sconce *et al*., 2005). On August 16, 2007, the data from these studies led to changes in the warfarin drug label to include the consideration of *CYP2C9* and *VKORC1* genotype in warfarin dosing. No consensus was reached on how to implement such testing (i.e. which polymorphisms to include and which algorithm to recommend in the drug label) therefore clinical implementation of this test has remained difficult.

22.4.4 "Not Ready for Prime Time" Tests

There are a few drugs in which the research points to a benefit of pharmacogenetic testing in dosing and/or treatment. However, further validation and replication would solidify the case to label changes in these drugs.

22.4.4.1 Abacavir and HLA*B5710 Testing

Pharmacogenetic effects may account for not only differential efficacy, but may also contribute to differential occurrence of adverse effects. An example of this scenario is provided by the pharmacogenetic association between *HLA*B5710* and abacavir. Abacavir is an nucleoside analog inhibitor of reverse transcriptase used to treat HIV-1 infections. Usual formulations are abacavir alone or in combination with other anti-retrovirals sush as lamivudine and zidovudine. These treatments are highly effective, though 2 to 9% of patients receiving the drug develop hypersensitivity reactions which result in discontinuation of the medication (Hetherington *et al*., 2001). This reaction is dose independent, and usually occurs within the first 6 weeks of therapy. Symptoms include: fever, rash, gastrointestinal and respiratory symptoms. Treatment consists of permenant cessation of abacavir and supportive care. Death may occur with reintroduction to abacavir.

Initial indications that the hypersensitivity reaction may be genetic came during the clinical trial of abacavir. First, only a small percentage of patients were affected, and second, a clear racial difference in risk was seen with black patients at lower risk for the adverse event. Using a candidate gene approach which included genes involved in immune response and drug metabolism, *HLA*B5701* was

identified as having a strong statistical association with the hypersensitivity reaction (Hetherington *et al*., 2002). The finding was also independently found in a Western Australian cohort (Mallal *et al*., 2002). A clinical study was undertaken to determine the utility of *HLA*B5701* testing. The PREDICT-1 study randomized patients to standard of care or prospective pharmacogenetic screening prior to therapy (Mallal *et al*., 2008). Those patients that were screened and were shown to be *HLA*B5701* positive were excluded for abacavir regimens. The prospective screening of patients reduced the overall incidence of clinically diagnosed hypersensitivity reaction from 7.8% in the standard of care arm to 3.4% in the pre-screened arm.

The same association has not been shown in a black population, presumably because of ethnic differences in LD patterns in the MHC (Hughes *et al.*, 2004). Further prospective clinical trials are needed to assess the utility of this test in clinical care.

24.4.4.2 Tamoxifen and CYP2D6

An example of how pharmacogenetic testing may affect drug selection can be seen in the case of tamoxifen. Tamoxifen is the standard endocrine therapy for estrogen receptor (ER) positive breast cancer in pre-menopausal women. Both the parent drug and its metabolite (endoxifen and 4-hydroxy tamoxifen) have pharmacologic activity, with the metabolites of tamoxifen having 30 to 100 times greater affinity to the ER. Both of these metabolites are formed through the metabolic conversion of tamoxifen by the hepatic enzyme CYP2D6.

Currently, there are over 60 known alleles in the *CYP2D6* gene (Sims, 2008). These include polymorphisms that abolish (*4, *5), decrease (*10, *17, and *41), or increase (caused by gene duplication) function. These alleles usually classify people into the following phenotypic groups: poor metabolizers (PM), intermediate metabolizers (IM), extensive metabolizers (EM), and ultrarapid metabolizers (UM).

With respect to tamoxifen dosing, investigators have looked at the association of these alleles with endoxifen concentration. Borges and coworkers (2006) showed that individuals carrying at least one null (*4) of reduced function allele (*10) and had similar endoxifen concentrations, while those with multiple copies of the gene have significantly higher endoxifen plasma levels. These differences in active metabolite concentrations have translated into differences in drug efficacy. Goetz and coworkers (2005) showed that women carrying the *CYP2D6*4/*4* genotype had shortened relapse-free time and worse disease-free survival compared to women that did not carry this allele. These findings were confirmed by Schroth and coworkers (2007), who showed that tamoxifen-treated women that carried a null or reduced function allele had significantly more recurrence, shorter relapse-free survival, and worse event-free survival compared to functional allele carrriers. In later studies it was noted that the co-administration of CYP2D6 inhibitors (common in breast cancer therapy for treatment of hot flashes and depression) converted the EM phenotype to a PM phenotype (Jin *et al*., 2005). When both these factors were considered in treatment outcomes, patients with decreased metabolism (either by genetic variant and/or co-administered drugs) had a significantly shorter time to recurrence and worse relapse-free survival (Goetz *et al*., 2007).

Because of these findings, the FDA advisory committee met October 18, 2006 to determine if recommendations should be added to the drug label of tamoxifen. The consensus of the advisory committee was that the tamoxifen label should be updated to reflect the increased risk for disease recurrence in postmenopausal women with ER-positive breast cancer who have the *CYP2D6* PM genotype or women taking drugs which interact with CYP2D6. Consensus was not reached on whether *CYP2D6* testing should be recommended. The final decision from the FDA regarding the label change is pending (Goetz *et al*., 2008).

Current consensus on pharmacogenetic testing is that *CYP2D6* testing may be most beneficial in settings in which alternative therapies are known to be equivalent or superior to tamoxifen monotherapy (e.g. the use of aromatase inhibitors, AI, in postmenopausal women; Goetz *et al*., 2007). For example, in these patients, PM and IM phenotype patients should be considered for alternative therapy. In addition, the optimal sequence of hormonal therapy (tamoxifen for some duration followed by an AI) for patients considered to be EM or UM needs to be investigated (Goetz *et al*., 2008).

22.5 USEFUL RESOURCES IN PHARMACOGENETICS

The advent of pharmacogenetic studies has led to the development of several databases, by which investigators can glean valuable genotype and phenotype information. The HapMap Project leads the way as a source of large genotype information. The HapMap Project evolved naturally from the Human Genome Project which made the sequence for the entire human genome publicly available. The goal of the international HapMap Project is to identify and catalog genetic similarities and differences in human beings, and to develop a haplotype map of the entire human genome that can be easily utilized in genetic association studies. Four world populations/racial groups were genotyped: 30 trios (two parents and a child) from Nigeria (YRI), 30 trios from the USA and of European ancestry (CEU), 45 unrelated Japanese (JPT), and 45 unrelated Han Chinese (CHB) (see http://www.hapmap.org). These populations were chosen to broadly represent African, European, and Asian ancestry and to identify most of the common haplotypes that exist in populations worldwide. The hope is that the information in

HapMap can be used to elucidate genes that affect health, disease, and individual responses to medications and environmental factors. By using this haplotype map, researchers will use the LD pattern seen in human population in pharmacogenetic studies. However, while the HapMap is sufficient for cataloging common variation, rarer alleles are not well represented. In addition, regions with complex and large regions of LD, such as the HLA region, may not be well represented by these data.

Another publicly available resource for pharmacogenetic information is the Pharmacogenetics and Pharmacogenomics Knowledge Base (PharmGKB; http://www.pharmgkb.org). The purpose of this database is to curate information that establishes knowledge about the relationships among drugs, diseases, and genes, including their variations and gene products (Klein *et al*., 2001). In addition to genes annotated to the drugs they may affect, PharmGKB also contains drug-specific pathways and genes found in these pathways. While quite thorough, this database is limited in the number of genes/drugs that are included.

Since the volume of pharmacogenetic databases and resources is quite large and variable, Table 22.3 summaries a few of the widely used databases publicly available.

22.6 ETHICAL IMPLICATIONS

Integration of pharmacogenetic information into both health care decisions and patient records brings up several ethical considerations. The most looming of these are privacy and confidentiality issues. Should genetic information be treated as any other diagnostic test, or is special protection needed for this data? Is patient consent required prior to testing for all pharmacogenetic mutations, even those used to guide dosing or therapy or only for those that predict disease susceptibility? Many of these concerns arise from lumping pharmacogenetic tests into the category of ''genetic tests''. Most pharmacogenetic tests are used for routine clinical decision making (i.e. *CYP2D6* SNP testing to determine enzyme activity to guide drug dosing), while genetic tests that are predictive of disease susceptibility may have an impact on the patient's life, family, or third-party interests (i.e. *BRCA1* and *BRCA2* genotyping for risk of early onset breast cancer). Therefore, not all ''genetic'' tests are created equal (Roses, 2000; Green and Botkin, 2003). The views on the amount of protection needed and appropriate safeguards to be used are a much-debated area (Hedgecoe, 2006). Some in the field have suggested that since the intent

TABLE 22.3 Commonly used resources in pharmacogenetics.

Resource	Home page web address	
dbGAP	http://www.ncbi.nlm.nih.gov/sites/entrez?Db = gap	Developed to archive and distribute the results of studies that have investigated the interaction of genotype and phenotype
dbSNP	http://www.ncbi.nlm.nih.gov/projects/SNP/index.html	Serves as a central repository for both single base nucleotide substitutions and short deletion and insertion polymorphisms
ENCODE Project	http://www.genome.gov/10005107	Project to identify and characterize functional elements in the human genome
International HapMap Project	http://www.hapmap.org/index.html.en	International repository of genetic variation that can be used to find genes that affect health, disease, and individual responses to medications and environmental factors
NIEHS SNPs	http://egp.gs.washington.edu	An SNP discovery resource focused on examining the relationships between environmental exposures, inter-individual sequence variation in human genes and disease risk in US populations
Perlegen Database	http://genome.perlegen.com	Access to several human genomic datasets released by Perlegen Sciences also available in HapMap
SeattleSNPs	http://pga.mbt.washington.edu	An SNP discovery resource focused on genes involved in pathways that underlie inflammatory responses in humans
UCSC Genome Browser	http://genome.ucsc.edu	Contains reference sequence and working draft assemblies for a large collection of genomes
Wellcome Trust Case Control Consortium	http://www.wtccc.org.uk/info/access_to_data_samples.shtml	Identifies genome sequence variants influencing major causes of human morbidity and mortality, through implementation and analysis of large-scale genome-wide association studies

of pharmacogenetic testing is not to predict risk or disease, the need for consent and genetic counseling is unnecessary. Requiring the same consent and regulatory requirements as other genetic tests may block the integration of pharmacogenetics into clinical care and may potentially deprive patients of the benefits this type of testing makes possible. Others feel that patient consent and an open dialog between patients and clinicians on the benefits and potential risks of pharmacogenetic testing would make patients more willing to use pharmacogenetic testing (Robertson, 2001; Schubert, 2004; van Delden *et al*., 2004).

22.6.1 Pharmacogenetic Legislation

Concerns such as these have been the impetus for the current bill regarding discrimination and genetic testing, dubbed GINA (Genetic Information Non-discrimination Act). This bill hopes to prohibit discrimination on the basis of genetic information with respect to health insurance and employment and hopes this protection will encourage the public to confidently partake in genetic testing for both health and research purposes. Currently, the bill has passed the US House of Representatives and the Senate and was recently signed into law (Genetic Alliance Organization, 2008). Similar steps have been taken in Europe with less concrete results. The 1997 European Convention on Human Rights and Biomedicine includes an article prohibiting discrimination on genetic grounds (Biomedicine, 1997). However, only four of the present EU member states have ratified it, namely Spain, Portugal, Greece, and Denmark. The Universal Declaration on the Human Genome and Human Rights, adopted by United Nations Educational Scientific and Cultural Organization (UNESCO) in 1997, though not legally binding, states that no one should be subject to discrimination on genetic grounds, and that genetic information should be confidential (Biomedicine, 1997).

22.6.2 Secondary Information

A second concern regarding pharmacogenetic testing is the issue of secondary information, garnered as a consequence of pharmacogenetic testing. For example, a patient is genotyped today for a SNP that is predictive of drug response; however, at some future date science may be able to use this genotype to predict a predisposition to a disease (Netzer and Biller-Andorno, 2004). One example in which this may be the case is the hepatic arylamine N-acetyltransferase-2 (NAT2). This enzyme shows polymorphic activity, in which one allele codes for fast acetylation and several alleles code for impaired or slow acetylation. Patients with a slow acetylation genotype may be at increased risk of adverse drug responses (Tanigawara *et al*., 2002; Soejima *et al*., 2007). N-acetylation status has also been associated with increased risk of bladder cancer (Cartwright *et al*., 1982), and lung cancer (Seow *et al*., 1999). The same case can be made for *UGT1A1*28*. Several investigators have looked at the association between the number of TA repeats and breast cancer risk because of its role in estradiol glucuronidation (Guillemette *et al*., 2000, 2001; Sparks *et al*., 2004). One study in 200 African American women showed a 1.8-fold increase in breast cancer risk in pre-menopausal women with the $(TA)_7$ and $(TA)_8$ alleles specifically in estrogen receptor negative (ER) breast cancers (Guillemette *et al*., 2000). However, a larger case-control study in Caucasian women showed no association (Guillemette *et al*., 2001). Clearly, further studies are needed; however, these findings may play a role in patients' willingness to be genotyped. Currently, there are no significant examples in which employment or insurance discrimination has occurred based on genetic data. Current trends in public policy and legislation provide individuals with increased protection from discrimination based on genetic information obtained for clinical purposes, as reflected in the legislative action outlined in section 22.6.1. However, because of the ever-changing nature of pharmacogenetics, further regulation may be needed as the knowledge in genetic susceptibility advances.

22.7 CONCLUSIONS AND FUTURE PERSPECTIVES

The current scientific literature has solidified genotype associations with drug response/efficacy in a few cases, which was further supported by the changes in drug labeling instituted by the FDA. However, the utility of such testing is still uncertain and far from optimal. Further prospective clinical studies as well as discovery research incorporating more complex genetic tools such as GWA and haplotype analysis are needed to provide utility to genetic information in both the clinical and drug development setting. The effective utilization of such testing will also rely heavily on clinician education, understanding, and professional judgment. We stand on the cusp of personalized medicine. The success of this endeavor will depend on more than knowing the genetic code, but in a deeper understanding of the intricacies that regulate and underlie the code.

REFERENCES

Aithal, G.P., Day, C.P., Kesteven, P.J., and Daly, A.K. (1999). Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Lancet* 353, 717–719.

Akaba, K., Kimura, T., Sasaki, A., Tanabe, S., Ikegami, T., Hashimoto, M., Umeda, H., Yoshida, H., Umetsu, K., Chiba, H., Yuasa, I., and Hayasaka, K. (1998). Neonatal hyperbilirubinemia and mutation of the bilirubin uridine diphosphate-glucuronosyltransferase gene: a common missense mutation among Japanese, Koreans and Chinese. *Biochem. Mol. Biol. Int.* 46, 21–26.

Ando, Y., Saka, H., Ando, M., Sawa, T., Muro, K., Ueoka, H., Yokoyama, A., Saitoh, S., Shimokata, K., and Hasegawa, Y. (2000). Polymorphisms of UDP-glucuronosyltransferase gene and irinotecan toxicity: a pharmacogenetic analysis. *Cancer Res.* 60, 6921–6926.

Aono, S., Adachi, Y., Uyama, E., Yamada, Y., Keino, H., Nanno, T., Koiwai, O., and Sato, H. (1995). Analysis of genes for bilirubin UDP-glucuronosyltransferase in Gilbert's syndrome. *Lancet* 345, 958–959.

Araki, K., Fujita, K., Ando, Y., Nagashima, F., Yamamoto, W., Endo, H., Miya, T., Kodama, K., Narabayashi, M., and Sasaki, Y. (2006). Pharmacogenetic impact of polymorphisms in the coding region of the UGT1A1 gene on SN-38 glucuronidation in Japanese patients with cancer. *Cancer Sci.* 97, 1255–1259.

Balding, D.J. (2006). A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.* 7, 781–791.

Bamshad, M.J., Wooding, S., Watkins, W.S., Ostler, C.T., Batzer, M.A., and Jorde, L.B. (2003). Human population genetic structure and inference of group membership. *Am. J. Hum. Genet.* 72, 578–589.

Beutler, E., Gelbart, T., and Demina, A. (1998). Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism?. *Proc. Natl. Acad. Sci. USA* 95, 8170–8174.

Biomedicine (1997). Council of Europe: Convention on Human Rights and Biomedicine. 2008, from http://fds.oup.com/www.oup.co.uk/pdf/bt/cassese/cases/part3/ch16/1121.pdf.

Bock, W.K., Guillemette, C., Mackenzie C., and Nebert, P.I. (2005). UGT nomenclature home page. http://galien.pha.ulaval.ca/alleles/ugt1a/UGT1A1.htm (assessed May 25, 2008).

Borges, S., Desta, Z., Li, L., Skaar, T.C., Ward, B.A., Nguyen, A., Jin, Y., Storniolo, A.M., Nikoloff, D.M., Wu, L., Hillman, G., Hayes, D.F., Stearns, V., and Flockhart, D.A. (2006). Quantitative effect of CYP2D6 genotype and inhibitors on tamoxifen metabolism: implication for optimization of breast cancer treatment. *Clin. Pharmacol. Ther.* 80, 61–74.

Bosma, P.J., Chowdhury, J.R., Bakker, C., Gantla, S., de Boer, A., Oostra, B.A., Lindhout, D., Tytgat, G.N., Jansen, P.L., Oude Elferink, R.P., and Chowdhury, N.R. (1995). The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. *N. Engl. J. Med.* 333, 1171–1175.

Breckenridge, A., Orme, M., Wesseling, H., Lewis, R.J., and Gibbons, R. (1974). Pharmacokinetics and pharmacodynamics of the enantiomers of warfarin in man. *Clin. Pharmacol. Ther.* 15, 424–430.

Brenner, C., and Duggan, D. (2004). *Oncogenomics: Molecular Approaches to Cancer*. Wiley-Liss, Hoboken.

Caldwell, M.D., Berg, R.L., Zhang, K.Q., Glurich, I., Schmelzer, J.R., Yale, S.H., Vidaillet, H.J., and Burmester, J.K. (2007). Evaluation of genetic factors for warfarin dose prediction. *Clin. Med. Res.* 5, 8–16.

Cartwright, R.A., Glashan, R.W., Rogers, H.J., Ahmad, R.A., Barham-Hall, D., Higgins, E., and Kahn, M.A. (1982). Role of N-acetyltransferase phenotypes in bladder carcinogenesis: a pharmacogenetic epidemiological approach to bladder cancer. *Lancet* 2, 842–845.

Cavalli-Sforza, L.L., and Feldman, M.W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.* 33(Suppl), 266–275.

Chen, K.K., and Poth, E.J. (1929). Racial differences as illustrated by the mydriatic action of cocaine, euphthalmine, and ephedrine. *J. Pharmacol. Exp. Ther.* 36, 363–400.

Danzon, P., and Towse, A. (2002). The economics of gene therapy and of pharmacogenetics. *Value Health* 5, 5–13.

Dickmann, L.J., Rettie, A.E., Kneller, M.B., Kim, R.B., Wood, A.J., Stein, C.M., Wilkinson, G.R., and Schwarz, U.I. (2001). Identification and functional characterization of a new CYP2C9 variant (CYP2C9*5) expressed among African Americans. *Mol. Pharmacol.* 60, 382–387.

Dirks, N.L., Huth, B., Yates, C.R., and Meibohm, B. (2004). Pharmacokinetics of immunosuppressants: a perspective on ethnic differences. *Int. J. Clin. Pharmacol. Ther.* 42, 701–718.

Dubinsky, M.C., Lamothe, S., Yang, H.Y., Targan, S.R., Sinnett, D., Theoret, Y., and Seidman, E.G. (2000). Pharmacogenomics and metabolite measurement for 6-mercaptopurine therapy in inflammatory bowel disease. *Gastroenterology* 118, 705–713.

Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C.S., Bowman, R., Meyer, K.B., Haiman, C.A., Kolonel, L.K., Henderson, B.E., Le Marchand, L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C.Y., Wu, P.E., Wang, H.C., Eccles, D., Evans, D.G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M.R., Rahman, N., Chenevix-Trench, G., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska, B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K.Y., Noh, D.Y., Ahn, S.H., Hunter, D.J., Hankinson, S.E., Cox, D.G., Hall, P., Wedren, S., Liu, J., Low, Y.L., Bogdanova, N., Schurmann, P., Dork, T., Tollenaar, R.A., Jacobi, C.E., Devilee, P., Klijn, J.G., Sigurdson, A.J., Doody, M.M., Alexander, B.H., Zhang, J., Cox, A., Brock, I.W., MacPherson, G., Reed, M.W., Couch, F.J., Goode, E.L., Olson, J.E., Meijers-Heijboer, H., van den Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R.L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J.L., McCredie, M., Southey, M., Giles, G.G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y.D., Spurdle, A.B., Beesley, J., Chen, X., Mannermaa, A., Kosma, V.M., Kataja, V., Hartikainen, J., Day, N.E., Cox, D.R., and Ponder, B.A. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447, 1087–1093.

Eeles, R.A., Kote-Jarai, Z., Giles, G.G., Olama, A.A., Guy, M., Jugurnauth, S.K., Mulholland, S., Leongamornlert, D.A., Edwards, S.M., Morrison, J., Field, H.I., Southey, M.C., Severi, G., Donovan, J.L., Hamdy, F.C., Dearnaley, D.P., Muir, K.R., Smith, C., Bagnato, M., Ardern-Jones, A.T., Hall, A.L., O'Brien, L.T., Gehr-Swain, B.N., Wilkinson, R.A., Cox, A., Lewis, S., Brown, P.M., Jhavar, S.G., Tymrakiewicz, M., Lophatananon, A., Bryant, S.L., Horwich, A., Huddart, R.A., Khoo, V.S., Parker, C.C., Woodhouse, C.J., Thompson, A., Christmas, T., Ogden, C., Fisher, C., Jamieson, C., Cooper, C.S., English, D.R., Hopper, J.L., Neal, D.E., and Easton, D.F. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.* 40, 316–321.

Eichelbaum, M., Spannbrucker, N., Steincke, B., and Dengler, H.J. (1979). Defective N-oxidation of sparteine in man: a new pharmacogenetic defect. *Eur. J. Clin. Pharmacol.* 16, 183–187.

European Medicines Agency (2007). Pre-authorisation Evaluation of Medicines for Human Use.

Evans, W.E., Hon, Y.Y., Bomgaars, L., Coutre, S., Holdsworth, M., Janco, R., Kalwinsky, D., Keller, F., Khatib, Z., Margolin, J., Murray, J., Quinn, J., Ravindranath, Y., Ritchey, K., Roberts, W., Rogers, Z.R., Schiff, D., Steuber, C., Tucci, F., Kornegay, N., Krynetski, E.Y., and Relling, M.V. (2001). Preponderance of thiopurine S-methyltransferase deficiency and heterozygosity among patients intolerant to mercaptopurine or azathioprine. *J. Clin. Oncol.* 19, 2293–2301.

Evans, W.E., Horner, M., Chu, Y.Q., Kalwinsky, D., and Roberts, W.M. (1991). Altered mercaptopurine metabolism, toxic effects, and dosage requirement in a thiopurine methyltransferase-deficient child with acute lymphocytic leukemia. *J. Pediatr.* 119, 985–989.

Evans, W.E., and McLeod, H.L. (2003). Pharmacogenomics – drug disposition, drug targets, and side effects. *N. Engl. J. Med.* 348, 538–549.

Evans, W.E., and Relling, M.V. (2004). Moving towards individualized medicine with pharmacogenomics. *Nature* 429, 464–468.

Fellay, J., Marzolini, C., Meaden, E.R., Back, D.J., Buclin, T., Chave, J.P., Decosterd, L.A., Furrer, H., Opravil, M., Pantaleo, G., Retelska, D., Ruiz, L., Schinkel, A.H., Vernazza, P., Eap, C.B., and Telenti, A. (2002). Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study. *Lancet* 359, 30–36.

Fox, A.L. (1932). The relationship between chemical constitution and taste. *Proc. Natl. Acad. Sci. USA* 18, 115–120.

Gagne, J.F., Montminy, V., Belanger, P., Journault, K., Gaucher, G., and Guillemette, C. (2002). Common human UGT1A polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38). *Mol. Pharmacol.* 62, 608–617.

Garrod, A.E. (1931). *The Inborn Factors in Disease: An Essay.* The Clarendon Press, Oxford.

Genetic Alliance Organization (2008). Genetic Information Nondiscrimination Act. 2008, from http://www.geneticalliance.org/ws_display.asp?filter = policy.leg.nondiscrim.

Goetz, M.P., Kamal, A., and Ames, M.M. (2008). Tamoxifen pharmacogenomics: the role of CYP2D6 as a predictor of drug response. *Clin. Pharmacol. Ther.* 83, 160–166.

Goetz, M.P., Knox, S.K., Suman, V.J., Rae, J.M., Safgren, S.L., Ames, M.M., Visscher, D.W., Reynolds, C., Couch, F.J., Lingle, W.L., Weinshilboum, R.M., Fritcher, E.G., Nibbe, A.M., Desta, Z., Nguyen, A., Flockhart, D.A., Perez, E.A., and Ingle, J.N. (2007). The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res. Treat.* 101, 113–121.

Goetz, M.P., Rae, J.M., Suman, V.J., Safgren, S.L., Ames, M.M., Visscher, D.W., Reynolds, C., Couch, F.J., Lingle, W.L., Flockhart, D.A., Desta, Z., Perez, E.A., and Ingle, J.N. (2005). Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. *J. Clin. Oncol.* 23, 9312–9318.

Gray, I.C., Nobile, C., Muresu, R., Ford, S., and Spurr, N.K. (1995). A 2.4-megabase physical map spanning the CYP2C gene cluster on chromosome 10q24. *Genomics* 28, 328–332.

Green, M.J., and Botkin, J.R. (2003). "Genetic exceptionalism" in medicine: clarifying the differences between genetic and nongenetic tests. *Ann. Intern. Med.* 138, 571–575.

Guillemette, C., De Vivo, I., Hankinson, S.E., Haiman, C.A., Spiegelman, D., Housman, D.E., and Hunter, D.J. (2001). Association of genetic polymorphisms in UGT1A1 with breast cancer and plasma hormone levels. *Cancer Epidemiol. Biomarkers Prev.* 10, 711–714.

Guillemette, C., Millikan, R.C., Newman, B., and Housman, D.E. (2000). Genetic polymorphisms in uridine diphospho-glucuronosyltransferase 1A1 and association with breast cancer among African Americans. *Cancer Res.* 60, 950–956.

Gupta, E., Lestingi, T.M., Mick, R., Ramirez, J., Vokes, E.E., and Ratain, M.J. (1994). Metabolic fate of irinotecan in humans: correlation of glucuronidation with diarrhea. *Cancer Res.* 54, 3723–3725.

Hasegawa, Y., Ando, Y., Ando, M., Hashimoto, N., Imaizumi, K., and Shimokata, K. (2006). Pharmacogenetic approach for cancer treatment-tailored medicine in practice. *Ann. NY Acad. Sci.* 1086, 223–232.

Hedgecoe, A.M. (2006). Context, ethics and pharmacogenetics. *Stud. Hist. Philos. Biol. Biomed. Sci.* 37, 566–582.

Hetherington, S., Hughes, A.R., Mosteller, M., Shortino, D., Baker, K.L., Spreen, W., Lai, E., Davies, K., Handley, A., Dow, D.J., Fling, M.E., Stocum, M., Bowman, C., Thurmond, L.M., and Roses, A.D. (2002). Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. *Lancet* 359, 1121–1122.

Hetherington, S., McGuirk, S., Powell, G., Cutrell, A., Naderer, O., Spreen, B., Lafon, S., Pearce, G., and Steel, H. (2001). Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. *Clin. Ther.* 23, 1603–1614.

Hopkins, M.M., Ibarreta, D., Gaisser, S., Enzing, C.M., Ryan, J., Martin, P.A., Lewis, G., Detmar, S., van den Akker-van Marle, M.E., Hedgecoe, A.M., Nightingale, P., Dreiling, M., Hartig, K.J., Vullings, W., and Forde, T. (2006). Putting pharmacogenetics into practice. *Nat. Biotechnol.* 24, 403–410.

Hughes, A.R., Mosteller, M., Bansal, A.T., Davies, K., Haneline, S.A., Lai, E.H., Nangle, K., Scott, T., Spreen, W.R., Warren, L.L., and Roses, A.D. (2004). Association of genetic variations in HLA-B region with hypersensitivity to abacavir in some, but not all, populations. *Pharmacogenomics* 5, 203–211.

Hughes, H.B., Biehl, J.P., Jones, A.P., and Schmidt, L.H. (1954). Metabolism of isoniazid in man as related to the occurrence of peripheral neuritis. *Am. Rev. Tuberc.* 70, 266–273.

Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., Bencko, V., Foretova, L., Janout, V., Chen, C., Goodman, G., Field, J.K., Liloglou, T., Xinarianos, G., Cassidy, A., McLaughlin, J., Liu, G., Narod, S., Krokan, H.E., Skorpen, F., Elvestad, M.B., Hveem, K., Vatten, L., Linseisen, J., Clavel-Chapelon, F., Vineis, P., Bueno-de-Mesquita, H.B., Lund, E., Martinez, C., Bingham, S., Rasmuson, T., Hainaut, P., Riboli, E., Ahrens, W., Benhamou, S., Lagiou, P., Trichopoulos, D., Holcatova, I., Merletti, F., Kjaerheim, K., Agudo, A., Macfarlane, G., Talamini, R., Simonato, L., Lowry, R., Conway, D.I., Znaor, A., Healy, C., Zelenika, D., Boland, A., Delepine, M., Foglio, M., Lechner, D., Matsuda, F., Blanche, H., Gut, I., Heath, S., Lathrop, M., and Brennan, P. (2008). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 452, 633–637.

Ingelman-Sundberg, M., Sim, S.C., Gomez, A., and Rodriguez-Antona, C. (2007). Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacol. Ther.* 116, 496–526.

Innocenti, F., Liu, W., Chen, P., Desai, A.A., Das, S., and Ratain, M.J. (2005). Haplotypes of variants in the UDP-glucuronosyltransferase1A9 and 1A1 genes. *Pharmacogenet. Genomics* 15, 295–301.

Innocenti, F., Undevia, S.D., Iyer, L., Chen, P.X., Das, S., Kocherginsky, M., Karrison, T., Janisch, L., Ramirez, J., Rudin, C.M., Vokes, E.E., and Ratain, M.J. (2004). Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *J. Clin. Oncol.* 22, 1382–1388.

Iyer, L., Das, S., Janisch, L., Wen, M., Ramirez, J., Karrison, T., Fleming, G.F., Vokes, E.E., Schilsky, R.L., and Ratain, M.J. (2002). UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. *Pharmacogenomics J.* 2, 43–47.

Iyer, L., Hall, D., Das, S., Mortell, M.A., Ramirez, J., Kim, S., Di Rienzo, A., and Ratain, M.J. (1999). Phenotype-genotype correlation of in vitro SN-38 (active metabolite of irinotecan) and bilirubin glucuronidation in human liver tissue with UGT1A1 promoter polymorphism. *Clin. Pharmacol. Ther.* 65, 576–582.

Jin, Y., Desta, Z., Stearns, V., Ward, B., Ho, H., Lee, K.H., Skaar, T., Storniolo, A.M., Li, L., Araba, A., Blanchard, R., Nguyen, A., Ullmer, L., Hayden, J., Lemler, S., Weinshilboum, R.M., Rae, J.M., Hayes, D.F., and Flockhart, D.A. (2005). CYP2D6 genotype, antidepressant

use, and tamoxifen metabolism during adjuvant breast cancer treatment. *J. Natl. Cancer. Inst.* 97, 30–39.

Jinno, H., Tanaka-Kagawa, T., Hanioka, N., Saeki, M., Ishida, S., Nishimura, T., Ando, M., Saito, Y., Ozawa, S., and Sawada, J. (2003). Glucuronidation of 7-ethyl-10-hydroxycamptothecin (SN-38), an active metabolite of irinotecan (CPT-11), by human UGT1A1 variants, G71R, P229Q, and Y486D. *Drug Metab. Dispos.* 31, 108–113.

Jorde, L.B., Watkins, W.S., Bamshad, M.J., Dixon, M.E., Ricker, C.E., Seielstad, M.T., and Batzer, M.A. (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* 66, 979–988.

Kalow, W., and Staron, N. (1957). On distribution and inheritance of atypical forms of human serum cholinesterase, as indicated by dibucaine numbers. *Can. J. Biochem. Physiol.* 35, 1305–1320.

Kim, T.W., and Innocenti, F. (2007). Insights, challenges, and future directions in irinogenetics. *Ther. Drug Monit.* 29, 265–270.

Klein, T.E., Chang, J.T., Cho, M.K., Easton, K.L., Fergerson, R., Hewett, M., Lin, Z., Liu, Y., Liu, S., Oliver, D.E., Rubin, D.L., Shafa, F., Stuart, J.M., and Altman, R.B. (2001). Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. *Pharmacogenomics J.* 1, 167–170.

Kuehl, P., Zhang, J., Lin, Y., Lamba, J., Assem, M., Schuetz, J., Watkins, P.B., Daly, A., Wrighton, S.A., Hall, S.D., Maurel, P., Relling, M., Brimer, C., Yasuda, K., Venkataramanan, R., Strom, S., Thummel, K., Boguski, M.S., and Schuetz, E. (2001). Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. *Nat. Genet.* 27, 383–391.

Kwan, P., and Brodie, M.J. (2005). Potential role of drug transporters in the pathogenesis of medically intractable epilepsy. *Epilepsia* 46, 224–235.

Lankisch, T.O., Vogel, A., Eilermann, S., Fiebeler, A., Krone, B., Barut, A., Manns, M.P., and Strassburg, C.P. (2005). Identification and characterization of a functional TATA box polymorphism of the UDP glucuronosyltransferase 1A7 gene. *Mol. Pharmacol.* 67, 1732–1739.

Lazarou, J., Pomeranz, B.H., and Corey, P.N. (1998). Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 279, 1200–1205.

Lindpaintner, K. (2003). Pharmacogenetics and pharmacogenomics in drug discovery and development: an overview. *Clin. Chem. Lab. Med.* 41, 398–410.

Link, E., Parish, S., Armitage, J., Bowman, L., Heath, S., Matsuda, F., Gut, I., Lathrop, M., and Collins, R. (2008). SLCO1B1 variants and statin-induced myopathy – a genomewide study. *N. Engl. J. Med.* 359, 789–799.

Loos, R.J., Lindgren, C.M., Li, S., Wheeler, E., Zhao, J.H., Prokopenko, I., Inouye, M., Freathy, R.M., Attwood, A.P., Beckmann, J.S., Berndt, S.I., Jacobs, K.B., Chanock, S.J., Hayes, R.B., Bergmann, S., Bennett, A.J., Bingham, S.A., Bochud, M., Brown, M., Cauchi, S., Connell, J.M., Cooper, C., Smith, G.D., Day, I., Dina, C., De, S., Dermitzakis, E.T., Doney, A.S., Elliott, K.S., Elliott, P., Evans, D.M., Sadaf Farooqi, I., Froguel, P., Ghori, J., Groves, C.J., Gwilliam, R., Hadley, D., Hall, A.S., Hattersley, A.T., Hebebrand, J., Heid, I.M., Lamina, C., Gieger, C., Illig, T., Meitinger, T., Wichmann, H.E., Herrera, B., Hinney, A., Hunt, S.E., Jarvelin, M.R., Johnson, T., Jolley, J.D., Karpe, F., Keniry, A., Khaw, K.T., Luben, R.N., Mangino, M., Marchini, J., McArdle, W.L., McGinnis, R., Meyre, D., Munroe, P.B., Morris, A.D., Ness, A.R., Neville, M.J., Nica, A.C., Ong, K.K., O'Rahilly, S., Owen, K.R., Palmer, C.N., Papadakis, K., Potter, S., Pouta, A., Qi, L., Randall, J.C., Rayner, N.W., Ring, S.M., Sandhu, M.S., Scherag, A., Sims, M.A., Song, K., Soranzo, N., Speliotes, E.K., Syddall, H.E., Teichmann, S.A., Timpson, N.J., Tobias, J.H., Uda, M., Vogel, C.I., Wallace, C., Waterworth, D.M., Weedon, M.N., Willer, C.J., Wraight,, Yuan, X., Zeggini, E., Hirschhorn, J.N., Strachan, D.P., Ouwehand, W.H., Caulfield, M.J., Samani, N.J., Frayling, T.M., Vollenweider, P., Waeber, G., Mooser, V., Deloukas, P., McCarthy, M.I., Wareham, N.J., Barroso, I., Jacobs, K.B., Chanock, S.J., Hayes, R.B., Lamina, C., Gieger, C., Illig, T., Meitinger, T., Wichmann, H.E., Kraft, P., Hankinson, S.E., Hunter, D.J., Hu, F.B., Lyon, H.N., Voight, B.F., Ridderstrale, M., Groop, L., Scheet, P., Sanna, S., Abecasis, G.R., Albai, G., Nagaraja, R., Schlessinger, D., Jackson, A.U., Tuomilehto, J., Collins, F.S., Boehnke, M., and Mohlke, K.L. (2008). Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat. Genet.* 40, 768–775.

Mahgoub, A., Idle, J.R., Dring, L.G., Lancaster, R., and Smith, R.L. (1977). Polymorphic hydroxylation of Debrisoquine in man. *Lancet* 2, 584–586.

Mallal, S., Nolan, D., Witt, C., Masel, G., Martin, A.M., Moore, C., Sayer, D., Castley, A., Mamotte, C., Maxwell, D., James, I., and Christiansen, F.T. (2002). Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* 359, 727–732.

Mallal, S., Phillips, E., Carosi, G., Molina, J.M., Workman, C., Tomazic, J., Jagel-Guedes, E., Rugina, S., Kozyrev, O., Cid, J.F., Hay, P., Nolan, D., Hughes, S., Hughes, A., Ryan, S., Fitch, N., Thorborn, D., and Benbow, A. (2008). HLA-B*5701 screening for hypersensitivity to abacavir. *N. Engl. J. Med.* 358, 568–579.

Marcuello, E., Altes, A., Menoyo, A., Del Rio, E., Gomez-Pardo, M., and Baiget, M. (2004). UGT1A1 gene variations and irinotecan treatment in patients with metastatic colorectal cancer. *Br. J. Cancer* 91, 678–682.

Marsh, S., Kwok, P., and McLeod, H.L. (2002). SNP databases and pharmacogenetics: great start, but a long way to go. *Hum. Mutat.* 20, 174–179.

Massacesi, C., Terrazzino, S., Marcucci, F., Rocchi, M.B., Lippe, P., Bisonni, R., Lombardo, M., Pilone, A., Mattioli, R., and Leon, A. (2006). Uridine diphosphate glucuronosyl transferase 1A1 promoter polymorphism predicts the risk of gastrointestinal toxicity and fatigue induced by irinotecan-based chemotherapy. *Cancer* 106, 1007–1016.

Matsunaga, E., Zanger, U.M., Hardwick, J.P., Gelboin, H.V., Meyer, U.A., and Gonzalez, F.J. (1989). The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine 4-hydroxylase deficiency in DA rats. *Biochemistry* 28, 7349–7355.

Millican, E., Jacobsen-Lenzini, P.A., Milligan, P.E., Grosso, L., Eby, C., Deych, E., Grice, G., Clohisy, J.C., Barrack, R.L., Burnett, R.S., Voora, D., Gatchel, S., Tiemeier, A., and Gage, B.F. (2007). Genetic-based dosing in orthopaedic patients beginning warfarin therapy. *Blood* 110, 1511–1515.

Monaghan, G., Ryan, M., Seddon, R., Hume, R., and Burchell, B. (1996). Genetic variation in bilirubin UPD-glucuronosyltransferase gene promoter and Gilbert's syndrome. *Lancet* 347, 578–581.

Montana, G. (2006). Statistical methods in genetics. *Brief. Bioinform.* 7, 297–308.

Moridani, M., Fu, L., Selby, R., Yun, F., Sukovic, T., Wong, B., and Cole, D.E. (2006). Frequency of CYP2C9 polymorphisms affecting warfarin metabolism in a large anticoagulant clinic cohort. *Clin. Biochem.* 39, 606–612.

Motulsky, A.G. (1957). Drug reactions, enzymes and biochemical genetics. *JAMA* 165, 835–837.

Motulsky, A.G. (1960). Metabolic polymorphisms and the role of infectious diseases in human evolution. *Hum. Biol.* 32, 28–62.

Nasu, K., Kubota, T., and Ishizaki, T. (1997). Genetic analysis of CYP2C9 polymorphism in a Japanese population. *Pharmacogenetics* 7, 405–409.

Netzer, C., and Biller-Andorno, N. (2004). Pharmacogenetic testing, informed consent and the problem of secondary information. *Bioethics* 18, 344–360.

Nooter, K., Sonneveld, P., Oostrum, R., Herweijer, H., Hagenbeek, T., and Valerio, D. (1990). Overexpression of the mdr1 gene in blast cells from patients with acute myelocytic leukemia is associated with decreased anthracycline accumulation that can be restored by cyclosporin-A. *Int. J. Cancer* 45, 263–268.

Op den Buijsch, R.A., Christiaans, M.H., Stolk, L.M., de Vries, J.E., Cheung, C.Y., Undre, N.A., van Hooff, J.P., van Dieijen-Visser, M.P., and Bekers, O. (2007). Tacrolimus pharmacokinetics and pharmacogenetics: influence of adenosine triphosphate-binding cassette B1 (ABCB1) and cytochrome (CYP) 3A polymorphisms. *Fundam. Clin. Pharmacol.* 21, 427–435.

Paoluzzi, L., Singh, A.S., Price, D.K., Danesi, R., Mathijssen, R.H., Verweij, J., Figg, W.D., and Sparreboom, A. (2004). Influence of genetic variants in UGT1A1 and UGT1A9 on the in vivo glucuronidation of SN-38. *J. Clin. Pharmacol.* 44, 854–860.

Parra, E.J., Marcini, A., Akey, J., Martinson, J., Batzer, M.A., Cooper, R., Forrester, T., Allison, D.B., Deka, R., Ferrell, R.E., and Shriver, M.D. (1998). Estimating African American admixture proportions by use of population-specific alleles. *Am. J. Hum. Genet.* 63, 1839–1851.

Paskind, H.A. (1921). Some diffrences in response to artopine in white and coloured races. *J. Lab. Clin. Med.* 7, 104–108.

Peterkin, V.C., Bauman, J.N., Goosen, T.C., Menning, L., Man, M.Z., Paulauskis, J.D., Williams, J.A., and Myrand, S.P. (2007). Limited influence of UGT1A1*28 and no effect of UGT2B7*2 polymorphisms on UGT1A1 or UGT2B7 activities and protein expression in human liver microsomes. *Br. J. Clin. Pharmacol.* 64, 458–468.

Pitot, H.C., Goldberg, R.M., Reid, J.M., Sloan, J.A., Skaff, P.A., Erlichman, C., Rubin, J., Burch, P.A., Adjei, A.A., Alberts, S.A., Schaaf, L.J., Elfring, G., and Miller, L.L. (2000). Phase I dose-finding and pharmacokinetic trial of irinotecan hydrochloride (CPT-11) using a once-every-three-week dosing schedule for patients with advanced solid tumor malignancy. *Clin. Cancer Res.* 6, 2236–2244.

Po, A.L. (2007). Personalised medicine: who is an Asian?. *Lancet* 369, 1770–1771.

Relling, M.V., Hancock, M.L., Rivera, G.K., Sandlund, J.T., Ribeiro, R.C., Krynetski, E.Y., Pui, C.H., and Evans, W.E. (1999). Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus. *J. Natl. Cancer Inst.* 91, 2001–2008.

Relling, M.V., Pui, C.H., Cheng, C., and Evans, W.E. (2006). Thiopurine methyltransferase in acute lymphoblastic leukemia. *Blood* 107, 843–844.

Rieder, M.J., Reiner, A.P., Gage, B.F., Nickerson, D.A., Eby, C.S., McLeod, H.L., Blough, D.K., Thummel, K.E., Veenstra, D.L., and Rettie, A.E. (2005). Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. *N. Engl. J. Med.* 352, 2285–2293.

Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W., Shugart, Y.Y., Griffiths, A.M., Targan, S.R., Ippoliti, A.F., Bernard, E.J., Mei, L., Nicolae, D.L., Regueiro, M., Schumm, L.P., Steinhart, A.H., Rotter, J.I., Duerr, R.H., Cho, J.H., Daly, M.J., and Brant, S.R. (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.* 39, 596–604.

Robertson, J.A. (2001). Consent and privacy in pharmacogenetic testing. *Nat. Genet.* 28, 207–209.

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., and Feldman, M.W. (2002). Genetic structure of human populations. *Science* 298, 2381–2385.

Roses, A.D. (2000). Pharmacogenetics and the practice of medicine. *Nature* 405, 857–865.

Rost, S., Fregin, A., Ivaskevicius, V., Conzelmann, E., Hortnagel, K., Pelz, H.J., Lappegard, K., Seifried, E., Scharrer, I., Tuddenham, E. G., Muller, C.R., Strom, T.M., and Oldenburg, J. (2004). Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature* 427, 537–541.

Rouits, E., Boisdron-Celle, M., Dumont, A., Guerin, O., Morel, A., and Gamelin, E. (2004). Relevance of different UGT1A1 polymorphisms in irinotecan-induced toxicity: a molecular and clinical study of 75 patients. *Clin. Cancer Res.* 10, 5151–5159.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., and Altshuler, D. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928–933.

Sai, K., Saeki, M., Saito, Y., Ozawa, S., Katori, N., Jinno, H., Hasegawa, R., Kaniwa, N., Sawada, J., Komamura, K., Ueno, K., Kamakura, S., Kitakaze, M., Kitamura, Y., Kamatani, N., Minami, H., Ohtsu, A., Shirao, K., Yoshida, T., and Saijo, N. (2004). UGT1A1 haplotypes associated with reduced glucuronidation and increased serum bilirubin in irinotecan-administered Japanese patients with cancer. *Clin. Pharmacol. Ther.* 75, 501–515.

Sanderson, S., Emery, J., and Higgins, J. (2005). CYP2C9 gene variants, drug dose, and bleeding risk in warfarin-treated patients: a HuGEnet systematic review and meta-analysis. *Genet. Med.* 7, 97–104.

Schroth, W., Antoniadou, L., Fritz, P., Schwab, M., Muerdter, T., Zanger, U.M., Simon, W., Eichelbaum, M., and Brauch, H. (2007). Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. *J. Clin. Oncol.* 25, 5187–5193.

Schubert, L. (2004). Ethical implications of pharmacogenetics – do slippery slope arguments matter? *Bioethics* 18, 361–378.

Sconce, E.A., Khan, T.I., Wynne, H.A., Avery, P., Monkhouse, L., King, B.P., Wood, P., Kesteven, P., Daly, A.K., and Kamali, F. (2005). The impact of CYP2C9 and VKORC1 genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. *Blood* 106, 2329–2333.

Seow, A., Zhao, B., Poh, W.T., Teh, M., Eng, P., Wang, Y.T., Tan, W.C., Lee, E.J., and Lee, H.P. (1999). NAT2 slow acetylator genotype is associated with increased risk of lung cancer among non-smoking Chinese women in Singapore. *Carcinogenesis* 20, 1877–1881.

Sims, S. (2008). CYP2D6 allele nomenclature. May 2008, from http://www.cypalleles.ki.se/cyp2d6.htm.

Smith, R.L. (1986). Introduction: human genetic variations in oxidative drug metabolism. *Xenobiotica* 16, 361–365.

Snyder, L.H., Clarke, H., and Moore, C.V. (1949). Studies in human inheritance; further data on the linkage of the genes for sickle cells and the M-N blood types. *Ohio J. Sci.* 49, 32.

Soejima, M., Sugiura, T., Kawaguchi, Y., Kawamoto, M., Katsumata, Y., Takagi, K., Nakajima, A., Mitamura, T., Mimori, A., Hara, M., and Kamatani, N. (2007). Association of the diplotype configuration at

the N-acetyltransferase 2 gene with adverse events with co-trimoxazole in Japanese patients with systemic lupus erythematosus. *Arthritis Res. Ther.* 9, R23.

Sparks, R., Ulrich, C.M., Bigler, J., Tworoger, S.S., Yasui, Y., Rajan, K.B., Porter, P., Stanczyk, F.Z., Ballard-Barbash, R., Yuan, X., Lin, M.G., McVarish, L., Aiello, E.J., and McTiernan, A. (2004). UDP-glucuronosyltransferase and sulfotransferase polymorphisms, sex hormone concentrations, and tumor receptor status in breast cancer patients. *Breast Cancer Res.* 6, R488–R498.

Stanulla, M., Schaeffeler, E., Flohr, T., Cario, G., Schrauder, A., Zimmermann, M., Welte, K., Ludwig, W.D., Bartram, C.R., Zanger, U.M., Eichelbaum, M., Schrappe, M., and Schwab, M. (2005). Thiopurine methyltransferase (TPMT) genotype and early treatment response to mercaptopurine in childhood acute lymphoblastic leukemia. *JAMA* 293, 1485–1489.

Takahashi, H., Wilkinson, G.R., Nutescu, E.A., Morita, T., Ritchie, M.D., Scordo, M.G., Pengo, V., Barban, M., Padrini, R., Ieiri, I., Otsubo, K., Kashima, T., Kimura, S., Kijima, S., and Echizen, H. (2006). Different contributions of polymorphisms in VKORC1 and CYP2C9 to intra- and inter-population differences in maintenance dose of warfarin in Japanese, Caucasians and African-Americans. *Pharmacogenet. Genomics* 16, 101–110.

Takeuchi, K., Kobayashi, Y., Tamaki, S., Ishihara, T., Maruo, Y., Araki, J., Mifuji, R., Itani, T., Kuroda, M., Sato, H., Kaito, M., and Adachi, Y. (2004). Genetic polymorphisms of bilirubin uridine diphosphate-glucuronosyltransferase gene in Japanese patients with Crigler-Najjar syndrome or Gilbert's syndrome as well as in healthy Japanese subjects. *J. Gastroenterol. Hepatol.* 19, 1023–1028.

Tanigawara, Y., Kita, T., Aoyama, N., Gobara, M., Komada, F., Sakai, T., Kasuga, M., Hatanaka, H., Sakaeda, T., and Okumura, K. (2002). N-acetyltransferase 2 genotype-related sulfapyridine acetylation and its adverse events. *Biol. Pharm. Bull.* 25, 1058–1062.

Tenesa, A., Farrington, S.M., Prendergast, J.G., Porteous, M.E., Walker, M., Haq, N., Barnetson, R.A., Theodoratou, E., Cetnarskyj, R., Cartwright, N., Semple, C., Clark, A.J., Reid, F.J., Smith, L.A., Kavoussanakis, K., Koessler, T., Pharoah, P.D., Buch, S., Schafmayer, C., Tepel, J., Schreiber, S., Volzke, H., Schmidt, C.O., Hampe, J., Chang-Claude, J., Hoffmeister, M., Brenner, H., Wilkening, S., Canzian, F., Capella, G., Moreno, V., Deary, I.J., Starr, J.M., Tomlinson, I.P., Kemp, Z., Howarth, K., Carvajal-Carmona, L., Webb, E., Broderick, P., Vijayakrishnan, J., Houlston, R.S., Rennert, G., Ballinger, D., Rozek, L., Gruber, S.B., Matsuda, K., Kidokoro, T., Nakamura, Y., Zanke, B.W., Greenwood, C.M., Rangrej, J., Kustra, R., Montpetit, A., Hudson, T.J., Gallinger, S., Campbell, H., and Dunlop, M.G. (2008). Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat. Genet.* 40, 631–637.

Tishkoff, S.A., and Verrelli, B.C. (2003). Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu. Rev. Genomics Hum. Genet.* 4, 293–340.

Tishkoff, S.A., and Williams, S.M. (2002). Genetic analysis of African populations: human evolution and complex disease. *Nat. Rev. Genet.* 3, 611–621.

Toffoli, G., Cecchin, E., Corona, G., Russo, A., Buonadonna, A., D'Andrea, M., Pasetto, L.M., Pessa, S., Errante, D., De Pangher, V., Giusto, M., Medici, M., Gaion, F., Sandri, P., Galligioni, E., Bonura, S., Boccalon, M., Biason, P., and Frustaci, S. (2006). The role of UGT1A1*28 polymorphism in the pharmacodynamics and pharmacokinetics of irinotecan in patients with metastatic colorectal cancer. *J. Clin. Oncol.* 24, 3061–3068.

Tukey, R.H., and Strassburg, C.P. (2000). Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu. Rev. Pharmacol. Toxicol.* 40, 581–616.

US Department of Health and Human Services, Food and Drug Administration (2005), Guidance for Industry: Pharmacogenomic 2008.

van Delden, J., Bolt, I., Kalis, A., Derijks, J., and Leufkens, H. (2004). Tailor-made pharmacotherapy: future developments and ethical challenges in the field of pharmacogenomics. *Bioethics* 18, 303–321.

Veenstra, D.L., Blough, D.K., Higashi, M.K., Farin, F.M., Srinouanprachan, S., Rieder, M.J., and Rettie, A.E. (2005a). CYP2C9 haplotype structure in European American warfarin patients and association with clinical outcomes. *Clin. Pharmacol. Ther.* 77, 353–364.

Veenstra, D.L., You, J.H., Rieder, M.J., Farin, F.M., Wilkerson, H.W., Blough, D.K., Cheng, G., and Rettie, A.E. (2005b). Association of vitamin K epoxide reductase complex 1 (VKORC1) variants with warfarin dose in a Hong Kong Chinese patient population. *Pharmacogenet. Genomics* 15, 687–691.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M.,

Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). The sequence of the human genome. *Science* 291, 1304–1351.

Vesell, E.S., and Page, J.G. (1968a). Genetic control of drug levels in man: antipyrine. *Science* 161, 72–73.

Vesell, E.S., and Page, J.G. (1968b). Genetic control of drug levels in man: phenylbutazone. *Science* 159, 1479–1480.

Vogel, F. (1959). Moderne Probleme der Humangenetic. *Ergebnisse der Inneren Medizin und Kinderheilkunde* 12, 65–126.

Wang, L., and Weinshilboum, R. (2006). Thiopurine S-methyltransferase pharmacogenetics: insights, challenges and future directions. *Oncogene* 25, 1629–1638.

Wang, S.L., Huang, J., Lai, M.D., and Tsai, J.J. (1995). Detection of CYP2C9 polymorphism based on the polymerase chain reaction in Chinese. *Pharmacogenetics* 5, 37–42.

Wasserman, E., Myara, A., Lokiec, F., Goldwasser, F., Trivin, F., Mahjoubi, M., Misset, J.L., and Cvitkovic, E. (1997). Severe CPT-11 toxicity in patients with Gilbert's syndrome: two case reports. *Ann. Oncol.* 8, 1049–1051.

Watanabe, I., Tomita, A., Shimizu, M., Sugawara, M., Yasumo, H., Koishi, R., Takahashi, T., Miyoshi, K., Nakamura, K., Izumi, T., Matsushita, Y., Furukawa, H., Haruyama, H., and Koga, T. (2003). A study to survey susceptible genetic factors responsible for troglitazone-associated hepatotoxicity in Japanese patients with type 2 diabetes mellitus. *Clin. Pharmacol. Ther.* 73, 435–455.

Yong, W.P., and Innocenti, F. (2007). Translation of pharmacogenetic knowledge into cancer therapeutics. *Clin. Adv. Hematol. Oncol.* 5, 698–706.

Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F., Marchini, J.L., Hu, T., de Bakker, P.I., Abecasis, G.R., Almgren, P., Andersen, G., Ardlie, K., Bostrom, K.B., Bergman, R.N., Bonnycastle, L.L., Borch-Johnsen, K., Burtt, N.P., Chen, H., Chines, P.S., Daly, M.J., Deodhar, P., Ding, C.J., Doney, A.S., Duren, W.L., Elliott, K.S., Erdos, M.R., Frayling, T.M., Freathy, R.M., Gianniny, L., Grallert, H., Grarup, N., Groves, C.J., Guiducci, C., Hansen, T., Herder, C., Hitman, G.A., Hughes, T.E., Isomaa, B., Jackson, A.U., Jorgensen, T., Kong, A., Kubalanza, K., Kuruvilla, F.G., Kuusisto, J., Langenberg, C., Lango, H., Lauritzen, T., Li, Y., Lindgren, C.M., Lyssenko, V., Marvelle, A.F., Meisinger, C., Midthjell, K., Mohlke, K.L., Morken, M.A., Morris, A.D., Narisu, N., Nilsson, P., Owen, K.R., Palmer, C.N., Payne, F., Perry, J.R., Pettersen, E., Platou, C., Prokopenko, I., Qi, L., Qin, L., Rayner, N.W., Rees, M., Roix, J.J., Sandbaek, A., Shields, B., Sjogren, M., Steinthorsdottir, V., Stringham, H.M., Swift, A.J., Thorleifsson, G., Thorsteinsdottir, U., Timpson, N.J., Tuomi, T., Tuomilehto, J., Walker, M., Watanabe, R.M., Weedon, M.N., Willer, C.J., Illig, T., Hveem, K., Hu, F.B., Laakso, M., Stefansson, K., Pedersen, O., Wareham, N.J., Barroso, I., Hattersley, A.T., Collins, F.S., Groop, L., McCarthy, M. I., Boehnke, M., and Altshuler, D. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* 40, 638–645.

This page intentionally left blank

Chapter 23

Nutrigenomics: Integrating Genomic Approaches into Nutrition Research

Lynette R. Ferguson[1,3], Martin Philpott[1,3] and Matthew P.G. Barnett[2,3]

[1]*Discipline of Nutrition, Faculty of Medical and Health Science, The University of Auckland, Private Bag 92019, Auckland, New Zealand;*

[2]*Food, Metabolism and Microbiology, Food and Textiles Group, AgResearch Limited, Palmerston North, New Zealand;*

[3]*Nutrigenomics New Zealand*

23.1 INTRODUCTION

While a number of formal definitions exist, in essence nutrigenomics (sometimes called nutritional genomics) considers the interactions between foods or dietary supplements and an individual's genome, and the consequent downstream effects on their phenotype. The field has the potential to provide tailored nutritional advice or develop specialist food products for populations or for individuals. It recognizes that appropriate dietary advice for one individual may be inappropriate, or actually harmful, to another. The potential is comparable to that of its sister field of pharmacogenomics, which considers individualized drug therapy. While pharmacogenomics is considered to have "come of age" (Allison, 2008), nutrigenomics is still considered as an emerging science. Joost and coworkers (2007) point to several challenges in the field to be overcome. The Nutrigenomics New Zealand model (http://www.nutrigenomics.org.nz) provides an approach to integrating the various available technologies for systems biology, ultimately directed towards the development of novel and targeted foods. However, this example also provides a cohesive and integrated model that can be applied to other research objectives. Inflammatory bowel disease (IBD), specifically Crohn's disease (CD), has been used as proof of principle.

One of many challenges in the field is combining the new developments in systems biology modeling with classical nutritional approaches to develop a more comprehensive model, which takes all relevant mechanisms into account, including phenotype. Nutrigenomics offers a multi-level systems biology model that is both flexible and rigorous. Because some of the newer tools provide a sensitivity that has been previously lacking, this field can be applied to maintenance of good health, as well as to disease states.

A necessary initial step establishes the nature of genetic variations relating to the disease or health endpoint of interest, using the techniques of genetic epidemiology (Burton *et al.*, 2005). While many studies thus far have measured variation primarily in the form of single nucleotide polymorphisms (SNPs) and haplotypes, future work is increasingly considering copy-number variants (CNVs) (Redon *et al.*, 2006). Accurate dietary assessment is also essential. The information generated by these two approaches can be used as a basis for the design of a cell-line (or other high-throughput model) that mimics the genotype, and allows large numbers of food fractions or known dietary components to be tested for their ability to overcome the phenotypic effects of the variant SNP or copy-number (Philpott *et al.*, 2007). Food components thus identified are subjected to further testing in both *in vitro* and *in vivo* models. Animal models permit the study of defined nutrients, foods and food components on the expression of disease, health- or performance-related genes. Food component-induced changes in gene and protein expression are measured in appropriate tissues collected from 2×2 mouse experiments (Park *et al.*, 1997), using microarrays and proteomics techniques to identify disease-, health-, or performance-related genes. As well as testing whether protective foods identified *in vitro* can extrapolate to the more complex *in vivo* situations, these studies also enable identification of biomarkers (Ferguson, 2008). These are then used as surrogate endpoints to study effects of dietary interventions in small animal models of human metabolism, or limited human clinical trials. Fundamental to the success of this approach is the ability to maintain large data sets and analyze complex multidimensional interactions.

23.2 THE NATURE OF GENETIC VARIATION

23.2.1 Genetic Epidemiology

Genetic epidemiology is "a science which deals with the etiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations" (Morton, 1982). This field has provided initial proof of principle, that genotype does profoundly influence response to diets or to nutrient in humans (e.g. Ordovas, 2006). A large number of studies have related variants in individual genes to disease risk, with varying degrees of success to this time. However, it will be important that future studies consider more than just individual gene variants, one at a time. The genetic basis of many complex diseases appears to be a combination of relatively small variations in the DNA, typically in the form of SNPs or CNVs, which increase disease susceptibility, rather than directly cause disease. For example, many of the genes associated with susceptibility to asthma and allergic response increase the risk of developing the disease by around 1.2-fold. However, carrying more than one such variant increases the risk significantly more (Vercelli, 2008). Genetic epidemiology considers inherited causes of disease, by looking at inheritance patterns, identifying the causal genes, and establishing which variants are associated with disease susceptibility (Burton *et al.*, 2005). This field also provides a necessary first step to studying gene–gene and gene–environment interactions, including nutrigenomics (Fig. 23.1).

Familial studies have traditionally been used for linkage analysis, but with the increased sensitivity of methods for analyzing genetic variation, such as high-density SNP microarrays, population studies are being increasingly used to establish associations between genes and disease risk. Whichever method is being used, the key steps involve consideration of the following questions:

- Is there a genetic component to the disorder?
- What is the relative size of that component in relation to other causes of disease susceptibility?
- Which gene(s) are associated with the effect?
- Which are the functional variants in those genes?
- How will those variants interact with diet and/or other environmental factors to cause the phenotype?

Linkage studies consider the segregation of genes among family members, and depend upon the co-segregation of two loci (one of which is the disease locus). They are dependent upon having access to a number of first degree relatives for analysis. While association studies may also be family based (transmission/disequilibrium test – TDT), they are increasingly being applied to population databases in whole genome association studies (WGAs). For example, the Wellcome Trust Case Control Consortium involved 17,000 subjects, and follow-up studies also examine high numbers of individuals from different populations (Wellcome Trust Case Control Consortium, 2007).

Twin studies provide important evidence on the genetic basis of disease, by comparing monozygotic (MZ) twins (who share all their genetic variants) with dizygotic (DZ) twins (who share half of their genetic variants). If there is evidence of higher concordance between MZ as compared with DZ twins, this implies a genetic basis for the disease in question (MacGregor *et al.*, 2000). Disease concordance less than 100% between MZ twins implies an environmental component, exactly as seen in nutrigenomics. Such evidence has been used to suggest gene–environment interactions in the etiology of many diseases including cancer (Risch, 2001), celiac disease (Greco *et al.*, 2002), and schizophrenia (Sullivan *et al.*, 2003).

Once there is evidence that the disease has a genetic basis, it becomes necessary to establish which genes are important. Much of the traditional work in this area has involved family studies, which consider allele frequencies in closely related subjects. Most powerfully, these involve multigenerational studies of large family pedigrees, but parent/child trios may be a pragmatic solution. Segregation analysis can be used in order to establish Mendelian inheritance patterns or non-classical patterns of inheritance, including mitochondrial diseases and genomic imprinting. These data can be subjected to linkage analysis in order to localize the crude chromosomal regions likely to contain the genes relevant to the disease. For example, the Online Mendelian Inheritance in Man (OMIM) database (http://www.ncbi.nlm.nih.gov/sites/entrez?db = omim) typically identifies chromosome regions, based on linkage analysis. This yields broad chromosome regions harboring many genes, where finer resolution comes from recombination events (meioses) in families. This approach has often been used in the past since it does not require many markers, but it is poor in terms of finding specific variants. It is important that linkage studies are not confused with association studies, since they consider different questions. A comparison of these two approaches is provided in Table 23.1.

While linkage studies rely on recombinational analyses, association studies are based on linkage disequilibrium (LD). This assumes that the SNP or CNV being studied is sufficiently close to the disease gene to result in an allelic association within the population (Carlson *et al.*, 2004; Morton, 2005). CD must be considered as a successful example of a complex disease in which both linkage and association studies have been used to locate and find disease susceptibility genes (Ferguson *et al.*, 2007a). Nevertheless, these studies are subject to the same pitfalls as conventional epidemiology, including selection bias, information bias, and confounding (Clayton *et al.*, 2005).

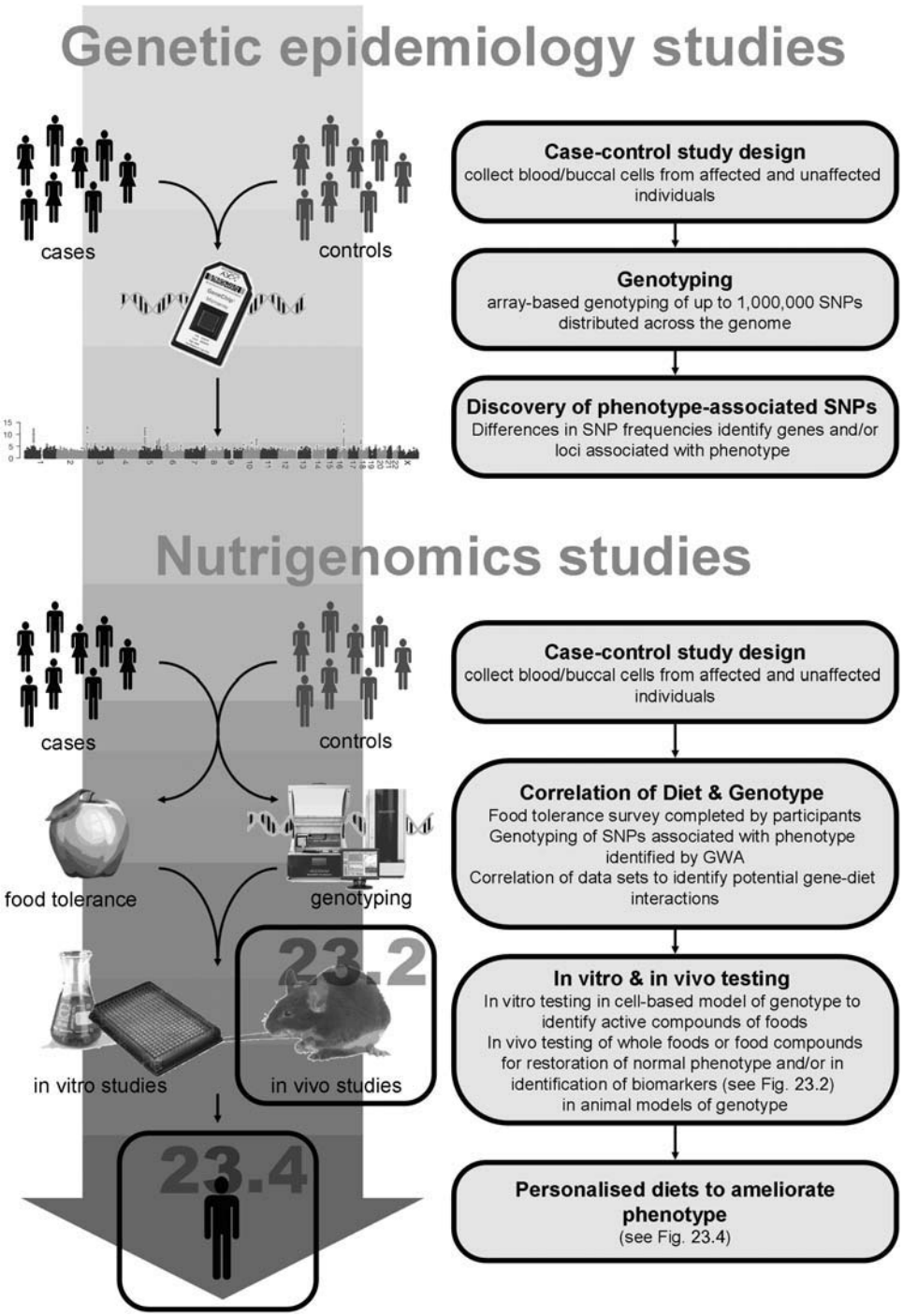


FIGURE 23.1 Depiction of how genetic epidemiology studies (top), particularly WGA studies, can form the basis of nutrigenomics studies (bottom). Text boxes provide a description of the workflow of both study types.

23.2.2 Single Nucleotide Polymorphisms (SNPS)

23.2.2.1 SNP Discovery

Fundamental to the success of genetic epidemiology is the identification of novel associations between SNPs in certain genes and a particular phenotype. Whether established by WGA studies or identified by decades' worth of linkage and candidate gene studies, large amounts of information on the genetic basis of many diseases are readily accessible through databases such as OMIM.

The increasing density and falling cost of DNA microarrays has made WGAs the current method of choice for SNP discovery. A number of microarray platforms allowing

TABLE 23.1 A comparison of linkage versus association studies.

Linkage studies	Association studies
Family based	Families or unrelated individuals
Matching/ethnicity generally unimportant	Matching/ethnicity crucial
Few markers for genome coverage (e.g. 300–400 STRs)	Many markers required for genome coverage (10^5–10^6 SNPs)
Good for initial detection; poor for fine-mapping	Poor for initial detection; good for fine-mapping
Powerful for rare variants	Powerful for common variants; rare variant identification generally impossible
Identifies chromosome regions rather than individual genes	Identifies individual genes

TABLE 23.2 Attributes of a desirable SNP genotyping method for nutrigenomics studies.

Attribute	Characteristics
Simple assay design	An assay should be easily and quickly developed from sequence information and require as few manual steps as possible.
High throughput	Newer methods require high numbers of SNPs to be assayed in (preferably) thousands of samples.
Robustness	The method should tolerate suboptimal DNA quality and minor variations in assay conditions such as reagent volumes, incubation periods, and temperatures.
Unambiguous genotype calling	Genotype calling should ideally be automated if large numbers of samples and/or genotypes are being assayed.
Affordability	When calculating the cost per genotype, the expense of equipment, reagents, consumables, and personnel time should be allowed for.
Reliability	The assay must produce accurate and reliable results.

genome-wide association studies are available, but the field is currently dominated by solutions provided by Affymetrix and Illumina (see also Chapter 16). Affymetrix SNP chips use photolithography to synthesize 25-mer oligonucleotide probes *in situ* onto silicon wafers, with the current Human SNP Array 6.0 having probes for >900,000 SNPs and a similar number of CNVs. Assays are performed by the hybridization of biotinylated cRNA, followed by washing, staining, and fluorescent detection. Illumina whole-genome genotyping assays use silica micro-beads, each uniquely labeled with a 50-mer oligonucleotide probe, with the current Human1M-Duo BeadChip having >1,000,000 markers that are informative for both SNPs and CNVs. Assays are performed by hybridization of unlabeled genomic DNA, followed by single base extension with labeled nucleotides, washing, and fluorescent detection.

Irrespective of the platform used, array-based WGAs provide a powerful new mechanism for unselected investigation of disease relationships. Furthermore, an enormous amount of effort is also being put into bringing the cost of whole-genome sequencing down into the range of US$1,000 (see also Chapter 24). Such an approach would provide all the genetic variation of an individual and may represent the future of WGAs.

23.2.2.2 SNP Analysis for Nutrigenomics

Because of the rapid proliferation of published WGA studies for many common diseases, SNP discovery is likely to be less important in nutrigenomics than the ability to genotype a population for which dietary and phenotype information is available, allowing correlations to be made. Over 100 different genotyping methods have been described to date, several of which have been discussed in detail in earlier chapters of this book. The ideal genotyping assay should have the qualities listed in Table 23.2.

Appropriate methods exist for genotyping a small number of samples for a single SNP through to thousands of samples for thousands of SNPs, or even higher. A number of the most common genotyping methods have been previously compared for accuracy and reliability, covering the range from low to medium-high throughput (Huebner *et al.*, 2007). The specific example chosen was of a triallelic SNP, since there is reason to believe that these may be more common than hitherto suspected (Huebner *et al.*, 2007).

Restriction fragment length polymorphism (RFLP), allelic discrimination PCR (ADPCR), TaqMan® (Applied Biosystems), LightCycler melting curve analysis (Roche Applied Science), Sequenom iPlex (Sequenom), and DNA sequencing were used to genotype the triallelic SNP G2677/T/A (rs2032582) in the MDR1 gene of 73 samples. Triallelic SNPs are increasingly being reported in the genome but are often initially reported as diallelic, due to limitations in common genotyping methods. In order to test the various available genotyping techniques, assay designs for genotyping analysis were based on the assumption that there were only two alleles (G and T), and the ability to detect the rare third A allele was assessed.

RFLP and LightCycler methodologies reported genotypes for all 73 samples, whereas the other methods had missing genotype rates between 4.1 and 5.5% (Table 23.3). More importantly, however, genotyping error rates were high for RFLP, ADPCR, and TaqMan®, due largely to their inability to correctly call the third allele. If the A allele is removed from the calculations, then the genotyping error rates for these assays are greatly improved, particularly for TaqMan® assays. Furthermore, if the presence of the third allele was known in advance, the ADPCR, and TaqMan® assays could be redesigned to successfully detect this allele. LightCycler, Sequenom, and sequencing methods were all capable of detecting the third allele without prior knowledge of its existence, although for LightCycler and Sequenom methods, this ability initially required examination of the individual spectra rather than relying on automated genotype calling. The LightCycler assay suffered from high genotyping error rates, irrespective of whether the A allele was included in the analysis, negating the low rate of missing genotypes and the ability to detect the third allele. In comparison, the Sequenom and sequencing approaches had much lower genotyping error rates. The observed genotype frequencies for the Sequenom assay most closely matched the true genotype frequency, demonstrating that while a high genotype calling rate is important, it is more important that the genotypes are being called correctly.

Based partly on results from these studies, a combination of genotyping approaches would seem appropriate, with the method determined by the throughput required. RFLP remains the simplest way of assaying a small number of samples for a small number of SNPs. The equipment and most of the reagents required to perform these assays will almost certainly be pre-existing in any molecular biology laboratory, making the only significant cost the restriction enzyme. However, RFLP assays require considerable manual input and the cost per genotype is linear, restricting its use to studies requiring low numbers of genotypes. These might include initial or exploratory studies, but the method would be impractical for the large nutrigenomics studies required to associate genotype with diet. TaqMan® assays are PCR-based homogeneous assays performed in 96- or 384-well plates with fluorescent detection in a RT-PCR instrument. This allows much greater throughput than RFLP, making TaqMan® assays ideal where an individual or relatively small number of SNPs need to be genotyped in larger populations of ~100 to several thousand. Due to the availability of a range of kit sizes, cost per genotype tends to go down with increasing number of samples, although

TABLE 23.3 Comparison of six common SNP genotyping approaches for genotyping a tri-allelic SNP (adapted from Huebner and coworkers, 2007). Numbers represent percentages.

Method	Missing genotypes	Genotyping errors (G/T/A alleles)	Genotyping errors (G/T alleles)
RFLP	0	12.3	3.0
ADPCR	5.5	16.9	6.8
TaqMan®	4.1	11.1	1.6
LightCycler	0	10.6	9.5
Sequenom	5.5	0	0
Sequencing	5.5	1.5	1.5

each SNP must be assayed individually, meaning cost per genotype is linear with increasing SNPs.

Increasingly, it is desirable to genotype the large numbers of SNPs identified in WGA studies in a study population where specific dietary information can also be collected. To achieve this at an acceptable cost per genotype, multiplexed assays become necessary. The Sequenom iPlex Gold genotyping assay is based on multiplex PCR followed by single base primer extension and allelic discrimination by MALDI TOF MS. Online tools available at Sequenom RealSNP.com (https://www.realsnp.com) and/or Assay Design 3.1 software are used to design PCR primers and extension primers for multiplex reactions, currently up to 40-plex. Primer pools are created and PCR amplification of DNA samples to be genotyped is performed in a 384-well format, before unincorporated nucleotides are deactivated by shrimp alkaline phosphatase (SAP) treatment. Primer extension is then carried out on a thermocycler using the iPlex enzyme, followed by the addition of ion exchange resin to remove ions that would otherwise interfere with MS detection. Reactions are then spotted onto SpectroCHIPs, which are small silicon wafers with an array of 384 spots preloaded with MALDI matrix optimized for DNA analysis. Finally, assay definitions and plate definitions are set up in Typer 4.0 software and spectra acquired from the SpectroCHIPs in the Sequenom Compact bench top MALDI TOF MS. The acquired genotype data can be reviewed in the Typer 4.0 software, before export for statistical analysis. In this way, the Sequenom platform is capable of generating 150,000 genotypes per day at a cost per genotype as low as 3.5¢ (US).

23.2.3 Copy-Number Variants

A CNV has been defined as a copy-number change involving a DNA fragment that is approximately 1 kb or larger (Feuk *et al.*, 2006). Changes involving the insertion or deletion of transposable elements are specifically excluded from the definition (Freeman *et al.*, 2006). There is reason to believe that most humans carry considerably more genetic variation in the form of CNVs than had previously been suspected, and that these may be a greater source of human variation than SNPs (Redon *et al.*, 2006; Shen *et al.*, 2008). It is of interest that CNVs appear to be enriched within genes that are important in molecular–environmental interactions, potentially influencing immune defense and disease resistance or susceptibility of humans (Barber *et al.*, 2005).

Whole-genome-based analyses such as array comparative genomic hybridization (CGH; see also Chapter 12) or the high-density SNP genotyping arrays from Affymetrix and Illumina will identify variants, and establish their location. Using high-density 500 K SNP genotyping arrays, Redon and coworkers (2006) revealed around 1,200 CNVs of varying sizes (approximately 1 kb to over 3 Mb). Such initial studies need to be validated through repeated analysis, which may involve larger numbers of patient samples and/or a different population, using a similar technology (Freeman *et al.*, 2006). However, other assays are necessary to test specific hypotheses. Quantitative PCR-based methods such as TaqMan® analysis may be employed, but multiplex ligation-dependent probe amplification (MLPA; see also Chapters 12 and 13) is also increasingly used to identify and screen gene-based deletions and duplications (den Dunnen and White, 2006). However, Shen and coworkers (2008) point out that the use of DNA probes developed for polymorphic nucleotides is not necessary for CNV identification and detection, and SNP-based assay design may even hinder successful CNV detection. They designed and evaluated a high density array that used non-polymorphic oligonucleotide probes for CNV detection, effectively uncoupling CNV detection from SNP genotyping. They were able to detect nearly 200 CNVs, of which approximately half appeared novel. These were independently validated using quantitative PCR.

23.3 NUTRITIONAL EPIDEMIOLOGY

A fundamental assumption of nutrigenomics is that multigenic traits can be modified by a combination of nutritional and other environmental factors. If the field is to be successfully applied, accurate nutritional assessment must be a starting point. Traditional methods of analysis are well documented (e.g. Willett, 2005). However, current assessment tools, including either food frequency questionnaires, diet dairy or diet recall methods, are notoriously unreliable. Bingham and coworkers (1994) systematically compared eight dietary assessment methods in a sample of 160 women (a simple 24 h recall, a structured 24 h recall with portion size assessments using photographs, two food-frequency questionnaires, a 7-day estimated record or open-ended food diary, a structured food-frequency (menu) record, and a structured food-frequency (menu) record with portion sizes assessed using photographs), against weighed dietary records. These authors found that 7-day estimated records or food diaries gave the most accurate information for standard nutrients. However, a nutritional role is increasingly attributed to bioactive dietary components that are not associated with traditional nutrition. More than 25,000 bioactive food components have been identified, and may impact in a genotype-specific manner through their modes of interaction with the cell (Ferguson and Philpott, 2007; Ferguson *et al.*, 2007b). No food composition database provides sufficient information on levels of these compounds in common foods that could be used to estimate their consumption in the diet. Accurate nutritional assessment provides a major challenge to the field at present.

Nutrition researchers require methods of assessing food intake that will measure what people actually eat, instead of what they remember eating. New methods are under development (e.g. cell phones are being used to report what people are eating on a real-time basis, using the camera feature of the cell phone to take a picture of the meal). While this may produce an accurate record, there will be other challenges in the analysis.

Traditional food-frequency data is useful for determining typical intake of macro- and micro-nutrients, but it gives no indication of what short-term effects they might have on phenotype. In the case of CD, where diet can result in rapid changes in phenotype, knowing what foods people know they can or cannot eat is the most valuable information of all. Thus, a qualitative measure of food tolerance can be used, in preference to more conventional methods.

23.4 EXPERIMENTAL MODELS

23.4.1 Cell-Based Assays

While genomics technologies make it possible to simultaneously examine the expression of tens of thousands of genes or detect hundreds of thousands of SNPs and metabolomics is moving towards the identification of the multitude of compounds present in foods (see below), the tools to test the ability of foods to modulate phenotype against a specific genotype are still in their infancy. A genotype-specific cell-based assay (Philpott *et al.*, 2007) has been previously reported that allows for the screening of food extracts or compounds for their ability to overcome the defects imparted by the 3020insC polymorphism in the NOD2 gene, which is associated with CD. This example utilizes technology that can be readily applied to other genetic variants, providing the SNP leads to a clearly defined phenotypic difference between cells carrying the wild-type and cells carrying the variant allele.

In this assay, HEK293T cells (which do not natively express NOD2) are cotransfected with an NF-κB-SEAP reporter construct, β-galactosidase transfection control plasmid, and a vector constitutively expressing either the wild-type or mutant NOD2. Wild-type NOD2, but not mutant NOD2, responds to the specific ligand muramyl dipeptide (MDP) by triggering NF-κB activation, which can be assessed by SEAP production in this system. Food extracts are then screened for their ability to restore normal phenotype (i.e. SEAP production) to the mutant cells. Serial fractionation and retesting of any extracts producing a positive response in the assay allows the identification of the active compound responsible. The use of robotics and high well-density plates, along with the initial use of crude extracts that contain many compounds at potentially active concentrations, allows large numbers of food compounds to be rapidly screened.

23.4.2 Animal Models

Because of the complex interactions between foods, microflora, and host cells that occur within the intestine, *in vitro* models are limited in their ability to help understand these interactions. It is difficult to foresee any *in vitro* assay ever being able to represent this complexity; since they commonly exclude some of the important interactions, the role of the intestinal microflora being an excellent example. Furthermore, even the most complex *in vitro* models will never contain every different cell type present in the intestinal or other tissues. Therefore, *in vivo* models are required to gain a better understanding of changes in phenotype from healthy to disease states, and the effects of foods and food components on this process.

A large number of animal models are available for specific purposes, which may conveniently be categorized into the following groups:

- Induced (experimental) disease models
- Spontaneous (mutant) disease models
- Genetically modified disease models
  - Transgenic
  - Knockouts
  - Chemically induced

Such models are currently available for all of the most common human diseases. In the following example, one specific group of such models is considered, for colonic inflammation.

The authors' studies aim to understand how different foods or food components might interact with a particular genotype to cause the chronic intestinal inflammation which is a hallmark of CD. A systematic approach is needed where intestinal parameters are measured at several time points, during which homeostasis of the intestinal system is challenged. The impacts on the biological gene and protein networks are measured, and physiology modeling systems biology approaches used to integrate the key drivers of the effects of foods on intestinal wellness (Fig. 23.2).

Many different rodent models of intestinal inflammation could have been selected (Jurjus *et al.*, 2004). This includes inducible colitis models and gene knockout or transgenic mouse and rat models. All animal models of intestinal disease should exhibit some key disease attributes (similarity in morphology changes, inflammation status, pathophysiology and time course). Our studies are specifically looking at nutrient–gene interactions, so we have chosen to work with mouse models in which a well-defined genetic variation is associated with intestinal inflammation. At least ten rodent models of IBD involving single gene manipulations exist with some of the symptoms characteristic of IBD (Kang *et al.*, 2008). Two in particular have been selected.

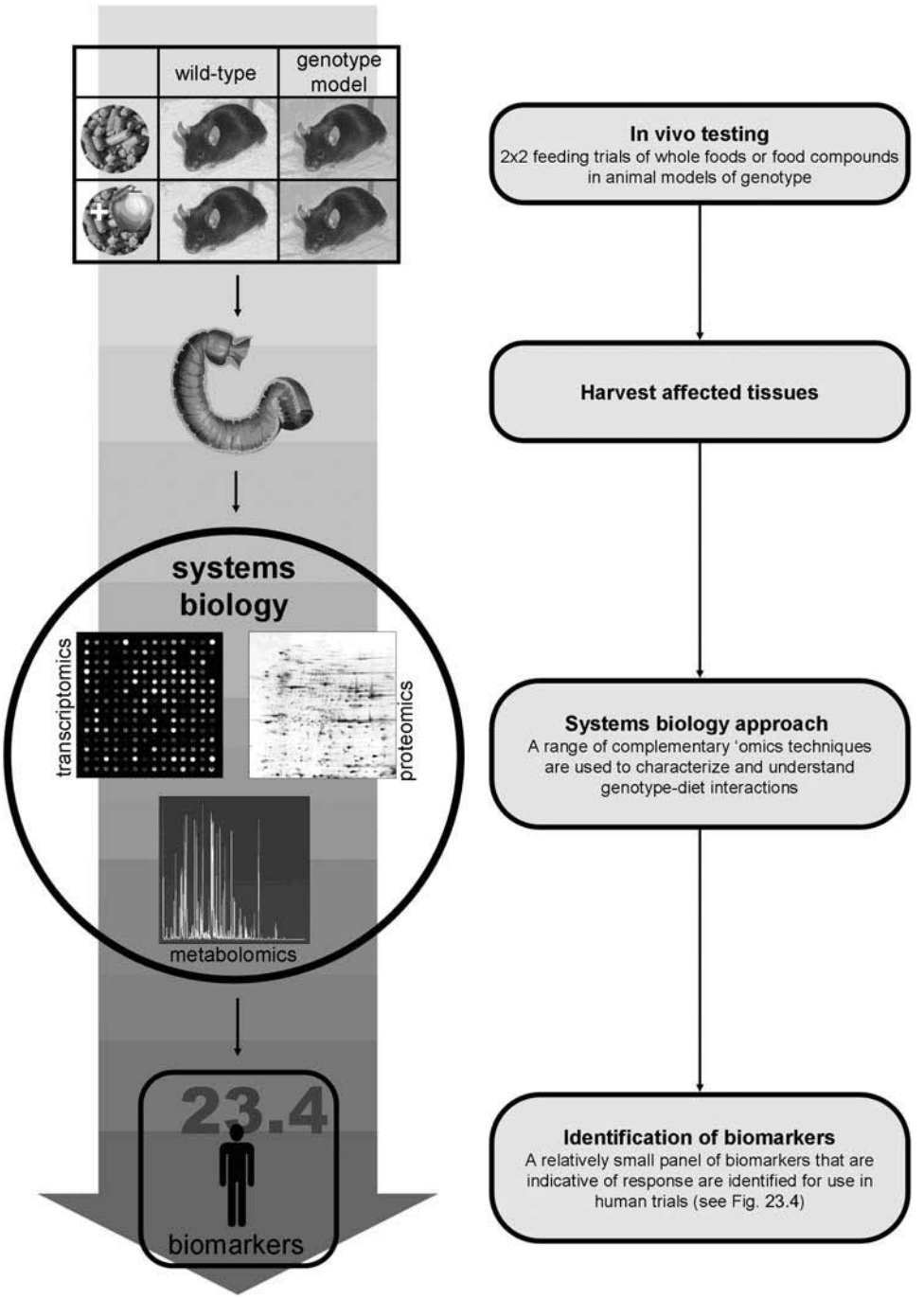


FIGURE 23.2 Systems biology approach employed for animal model studies to understand how different foods or food compounds might interact with a particular genotype. One output of this approach is a defined set of biomarkers for use in human trials (see also Fig. 23.4).

23.4.2.1 The Interleukin-10 Gene-Deficient (IL10$^{-/-}$) Mouse

IL10 is an anti-inflammatory cytokine and a defect in the IL10 gene leads to an imbalance in the inflammatory responses. While IL10$^{-/-}$ mice (C57BL/6J/129-Ola background) have been reported to develop Crohn's-like colitis by 12 weeks of age when raised under conventional conditions (Berg *et al.*, 1996), the level of inflammation observed is influenced by the background strain (see also Chapter 31 for inbred strains). In the case of the C57BL/6J background, more consistent inflammation is observed when the IL10$^{-/-}$ mice are inoculated with *Enterococcus faecalis* (Balish and

Warner, 2002) or *Enterococcus* isolates (both faecalis and faecium) in combination with conventional intestinal flora derived from healthy C57BL/6J mice raised under conventional conditions (Roy *et al.*, 2007; Barnett *et al.*, 2008). As is the case in human IBD, the precise mechanism that results in inflammation in IL10$^{-/-}$ mice is unclear, although evidence suggests an inappropriate inflammatory response to normal intestinal flora, through activation of CD4^{+} th1 cells and the depletion of their inhibitors, i.e. the regulatory T cells (Kuhn *et al.*, 1993).

Although IL10$^{-/-}$ mice have IBD-like symptoms, variants in this gene have not been consistently associated with IBD in human populations (Klein *et al.*, 2000; Castro-Santos *et al.*, 2006). However, variants of the IL23 receptor have shown association with IBD in a number of population groups, including those from New Zealand (Roberts *et al.*, 2007), the Netherlands (Weersma *et al.*, 2008), and Finland (Lappalainen *et al.*, 2008). It has been shown that the T cell-mediated colitis that develops in IL10$^{-/-}$ mice may be dependent on IL23 (Yen *et al.*, 2006). This mouse model has been used by our group (Roy *et al.*, 2007; Knoch *et al.*, 2008) and others to identify food components which may ameliorate or prevent human IBD. Figure 23.3 shows a biological network of the genes of the most significantly

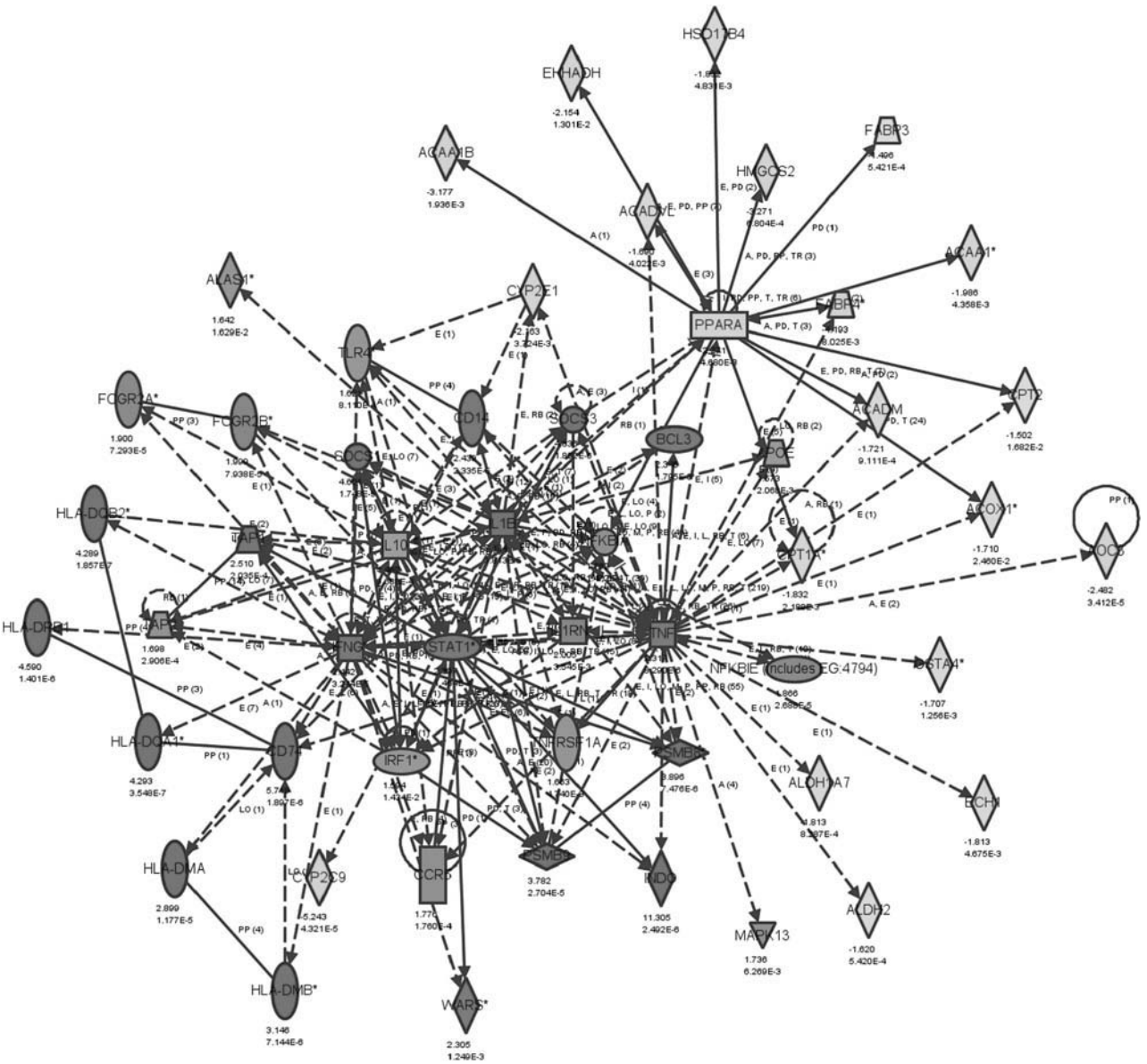


FIGURE 23.3 A biological network of the genes of the most significantly altered pathways in the IL10$^{-/-}$ mice (C57BL/6J) compared to control C57BL/6J mice (Knoch *et al.*, 2008). The network was generated by Ingenuity Pathway Analysis software (Ingenuity Systems) using all molecules from significantly affected canonical pathways. Connections were applied based on known interactions between these genes within the Ingenuity Pathways Knowledge Base.

altered pathways in the IL10$^{-/-}$ mice (C57BL/6J) compared to control C57BL/6J mice (Knoch *et al.*, 2008). This model has been used to test the efficacy of potential food components including polyunsaturated fatty acids (Roy *et al.*, 2007; Knoch *et al.*, 2008), fish oils (Hegazi *et al.*, 2006), flaxseed oil (Cohen *et al.*, 2005), and some probiotics (Madsen, 2001; Schultz *et al.*, 2002; McCarthy *et al.*, 2003).

23.4.2.2 The Multidrug Resistance Gene-Deficient (mdr1a$^{-/-}$) Mouse

MDR1 belongs to a family of transmembrane transporters, known as ATP-binding cassette transporters, and a number of mutations that lead to reduced MDR1 activity have been linked to IBD (Brant *et al.*, 2003; Schwab *et al.*, 2003; Ho *et al.*, 2006). In the intestinal tract, MDR1 is expressed on the luminal surface of epithelial cells, pumping toxins from inside the cells back into the intestinal lumen (Ho *et al.*, 2003; Bilsborough and Viney, 2004); thus, a failure to "clear" these toxic compounds may be triggering and maintaining the spontaneous intestinal inflammation observed in *mdr1a*$^{-/-}$ mice (Panwala *et al.*, 1998; Banner *et al.*, 2004; Wilk *et al.*, 2005; Dommels *et al.*, 2007). Furthermore, antibiotic treatment has been shown to both prevent and therapeutically reverse inflammation in *mdr1a*$^{-/-}$ mice, demonstrating both a role for bacterial flora in the initiation and progression of inflammation in these mice (Panwala *et al.*, 1998), and the importance of considering the bacterial flora in any gut health studies. The *mdr1a*$^{-/-}$ model is another tool we have used to understand the development of intestinal inflammation (Dommels *et al.*, 2007), and to test the efficacy of food components such as curcumin and rutin (Nones *et al.*, 2008a) and green tea (Nones *et al.*, 2008b) on the prevention of this phenotype.

For gene-mutation IBD rodent models, as well as monitoring the effect of food components on the physical and biochemical signs of inflammation, gene (Roy *et al.*, 2007; Knoch *et al.*, 2008) and protein (Werner and Haller, 2007) expression, metabolite profiling and epigenomic changes in the colon cells can also be assessed to understand the mechanisms by which the changes in inflammation occur. The expression of genes implicated in human IBD, e.g. those involved in processes such as immune response, inflammation, antigen presentation and xenobiotic metabolism (Dommels *et al.*, 2007; Roy *et al.*, 2007; Barnett *et al.*, 2008; Knoch *et al.*, 2008; Nones *et al.*, 2008a, b), is of particular interest.

23.4.3 Human Clinical Trials

The ultimate proof that a nutritional intervention is beneficial to health utilizes a double blind, placebo-controlled clinical trial. Optimal study design would utilize randomly chosen subjects, who are prospectively examined and compared to well-matched control individuals, using genomically based diagnostic techniques. However, such studies are not cost effective in terms of either time or money. Furthermore, an increasing number of such trials have provided negative or ambiguous results, even where there is a very strong reason to believe that the foods in question are beneficial. Examples would include omega-3 polyunsaturated fatty acids, where the meta-analysis of several large international studies has failed to support a benefit against cardiovascular disease (Hooper *et al.*, 2006). Similarly, despite good theoretical grounds for the hypothesis, most human intervention studies on antioxidants have given disappointing results (Moller and Loft, 2004, 2006). More generally, most human clinical studies on randomly selected populations have failed to demonstrate predicted outcomes from dietary interventions alone. One possible explanation is that many of the studies have merely provided a better description of homeostasis, showing the robustness of health and the more analytically elusive subtleties of multiple minor changes. Two approaches are yielding fruit, and should not be considered mutually exclusive.

23.4.3.1 Subject Pre-Selection or Stratification According to Genotype

A flow diagram showing how a trial involving subject preselection or stratification according to genotype might optimally work is provided in Fig. 23.4. To date, studies that have stratified populations into specific subpopulations which are uniquely responsive to particular diets or nutrients have not yet been completed. However, Kornman and coworkers (2007) stratified their population of individuals susceptible to cardiovascular disease risk according to genotype, and tested a botanic formulation that included rosehips, a blueberry and blackberry mixture, and a grapevine extract in their nutrigenetics proof of concept clinical trial. They studied healthy adults with elevated C-reactive protein as an indication of inflammation, and stratified them according to being positive or negative for known variations that related to risk of chronic disease in the interleukin-1 (IL-1) gene. Thirty-nine IL-1 positive subjects (who carried genetic variants) and 40 IL-1 negative subjects were randomized to the candidate botanical formulation or a placebo control. These authors were able to show that a 12-week period of dosing with their proprietary formula reduced inflammation in the IL-1 positive individuals. These individuals carried genetic variations that predisposed them to overexpression of IL-1, and are known to be susceptible to early heart disease. The effect of the botanic preparation was much reduced in the IL-1 negative subjects.

23.4.3.2 Challenging Homeostasis

Ultimately, the goal of nutrigenomics is to ensure optimum health. However, quantification of health cannot be achieved by applying "absence of disease" approaches.

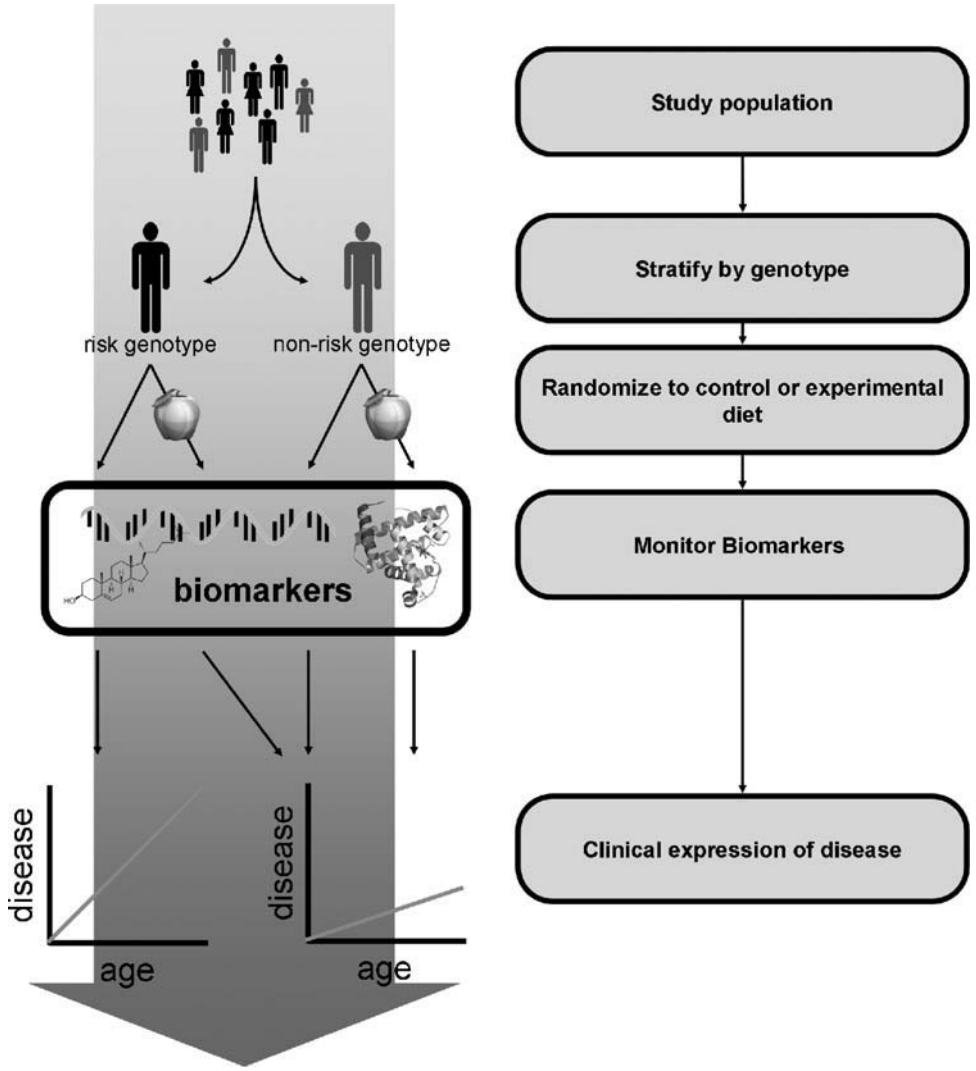


FIGURE 23.4 Study design for human trials of foods where subjects are pre-selected or stratified according to genotype. Study participants are randomized on to a control diet or an experimental diet that includes the food shown to ameliorate the phenotype associated with a particular genotype in animal models (see also Fig. 23.2). Biomarkers previously identified in the same animal studies are monitored to rapidly establish if the experimental diet can restore the non-risk-genotype profile, without the need to follow study participants until the development of disease. These results can be extrapolated to predict the long-term protection against disease development.

Intermediate processes such as metabolic stress, oxidative stress, or inflammatory stress may be more amenable to biomarker approaches. However, these processes strongly influence each other, and their complexity means that there is no completely satisfactory simple biomarker. Inter-individual variation may be larger than the effect imposed by the treatment, partly because the "confounders" cannot be adequately controlled. Genetics, lifestyle, age, stress, concurrent infections, etc., all determine the health status of an individual, and need to be quantified.

For these reasons, newer clinical trial designs look to challenging homeostasis. The concept considers that only after stressing or perturbing homeostasis can the robustness ("health") of the system be demonstrated and potentially quantified. A classic example in nutrition research is the oral glucose tolerance test, whereby individuals who have been fasted overnight are provided with a high glucose bolus, and the level of glucose in their bloodstream measured immediately and again after 2 hours. Other examples are actively being developed in order to accurately target a range of parameters of health or disease, including metabolism, oxidation, inflammation, and psychological stress (Kaput, 2008). For example, Kleemann and coworkers (2007) have shown that increased dietary cholesterol intake increased both atherosclerosis and liver inflammation in an animal model. This happened in a relatively short time span, and could be sensitively measured using a combined transcriptomics and metabolomics analysis. Theoretically at

least, a short-term human intervention study that increased cholesterol for a short time (challenges homeostasis) could provide information as to how other dietary factors may alleviate such stress.

23.5 DEFINING THE PHENOTYPE

The term "deep phenotyping" has recently been proposed, meaning "…more complete descriptions of the physical state of individuals (and, by aggregation, groups of individuals) than has been possible in the past…" (Tracy, 2008). The justifications for this are three-fold (Tracy, 2008):

- Translational research is necessary for translation of dietary recommendations, but insufficient to provide definitive proof of efficacy.
- Phenotype variability dilutes statistical power and strength of association.
- There are differences in pathophysiology over time within an individual.

A good example is provided by the field of cardiovascular disease, in which three phases can be identified, whereby it has become increasingly possible to define a serum lipid profile that differentiates healthy individuals from those at significantly increased risk of developing disease (Fig. 23.5). This modified lipid profile which integrates not only serum cholesterol but also HDL and LDL cholesterol provides a basis whereby effects of a dietary intervention can be readily measured. The aim is to move individuals from the "high risk" to "low risk" profile, with good predictive powers for benefit or otherwise.

More powerfully, deep phenotyping involves the application of "omics" technologies to better understand the underlying biology and metabolism associated with disease, thereby enhancing diagnosis and potentially identifying improved phenotypic markers of disease state. International standards are being applied in the development of guidelines and standards for these technologies. For example, *Nature Biotechnology* has an online community consultation initiative (http://www.nature.com/nbt/consult/index.html) which is intended to encourage researchers to participate across country barriers.

Because deep phenotyping relies on a suite of new and emerging technologies, the costs can be considerable. For example, one cholesterol measurement might cost US$5–10, whereas an appropriately planned and performed microarray-based expression experiment could cost in the order of US$500–1,000 per sample. Specialized tissue collection and complex bioinformatic analysis only adds to this cost. However, the end result of preliminary analyses using these methods may be a relatively small number of measurements (be they gene or protein expression events, or blood/urine metabolites) which, when combined, give a very accurate and precise diagnosis of a disease. Most importantly, such measures may be possible before there is any obvious disease onset, providing the ability to prevent or treat symptoms, and therefore improving the prognosis for the individual in question. Therefore, techniques that are becoming routinely applicable to nutrigenomics research hold much promise in the field of disease diagnosis.

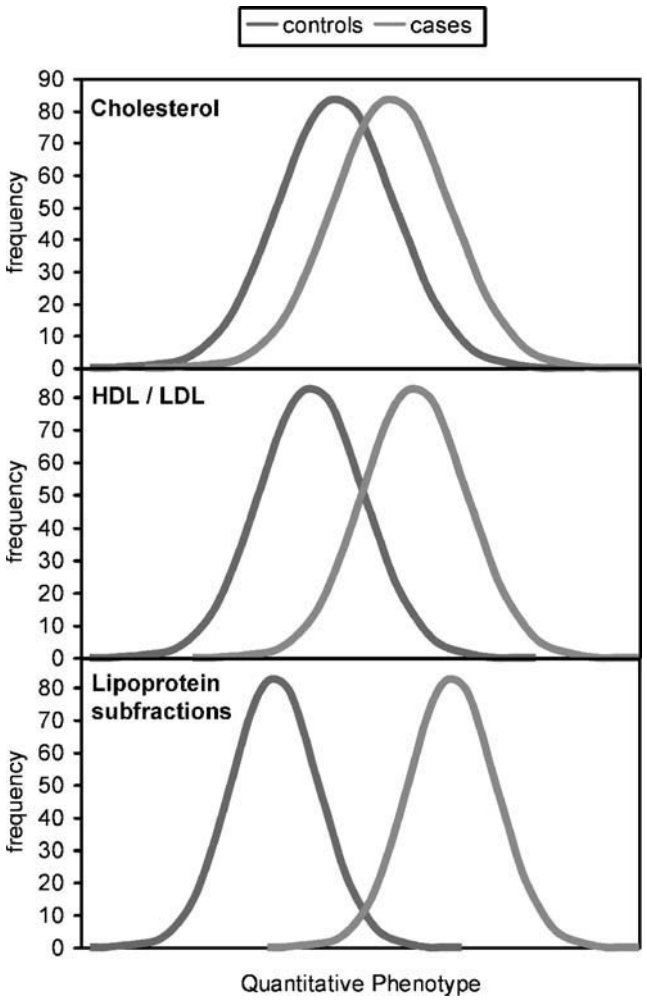


FIGURE 23.5 Example of how deep phenotyping can be used to better define disease risk. Since the 1950s it has been recognized that cholesterol levels are associated with increased risk of cardiovascular disease, but that the predictive power of cholesterol levels was relatively poor (top panel). In the 1980s it was demonstrated that a wider lipid profile investigating HDL and LDL was an improved predictor of cardiovascular disease (middle panel). More recently, detailed examination of lipoprotein subfractions has been used to establish much more accurate cardiovascular disease risk prediction (lower panel). Dietary intervention can then be used to alter an individual's lipid profile to shift them from the high risk to the low risk profile.

23.5.1 Transcriptomics

The transcriptome is the set of all messenger RNA (mRNA) molecules, or "transcripts", produced in one or a population of cells (Duffus *et al.*, 2007). The transcriptome

reflects the genes that are being actively expressed at any given time because it includes all *mRNA* transcripts in the cell. Analyses often use high-throughput techniques based on DNA microarrays.

While originally developed using arrays of cDNA "spotted" onto glass microscope slides, the field of transcriptomics (as in the case of WGAs) is now dominated by solutions provided by companies such as Affymetrix, Agilent Technologies, and Illumina. Affymetrix arrays, like the SNP chips, consist of *in situ* synthesized 25-mer oligonucleotide probes applied using photolithography technology. Agilent Technologies also use *in situ* hybridized oligos, but these are synthesized using piezo-electric (inkjet) technology, and have a length of 60 bases. Illumina BeadChips use 50-mer oligonucleotides immobilized onto microbeads, which then self-assemble on to ordered microwell-etched substrates. A comparison of some features of these three key microarray technologies is shown in Table 23.4.

For genome-wide measurement of gene expression changes associated with intestinal inflammation in both $IL10^{-/-}$ (Roy *et al.*, 2007) and $mdr1a^{-/-}$ mice (Dommels *et al.*, 2007), Nutrigenomics New Zealand has used Agilent Technologies microarrays. Briefly, cyanine-3 and cyanine-5-labeled cRNA is synthesized from purified colon RNA and hybridized to Agilent mouse whole genome arrays. The slides are scanned, using either a GenePix 4200A scanner (Molecular Devices Corporation, Sunnyvale, CA, USA) with spots identified and quantified using GenePix 6.0 software (Molecular Devices Corporation, Sunnyvale, CA, USA), or using an Agilent Technologies microarray scanner with the digitized data being aligned and spot intensity data downloaded as text files using Feature Extraction software. Linear models for microarray analysis (Bioconductor) (Smyth, 2005) are then used for normalization and identification of differentially expressed mRNA with multiple comparisons to minimize false discoveries using internationally recognized standards (Benjamini and Hochberg, 1995). Bioconductor (http://www.bioconductor.org) is open source and open development software for the analysis and comprehension of genomic data.

23.5.2 Proteomics

Proteomics is the analysis of the protein complement present in a cell, organ, or organism at any given time (Matt *et al.*, 2008; see also Chapter 18). While the genome provides information about the potential nature of the cellular proteins, the actual protein composition ("the proteome") ultimately determines the phenotype. Fundamental technologies for the separation of proteins and/or peptides are one- and two-dimensional gel electrophoresis and one- and two-dimensional liquid chromatography, typically coupled with mass spectrometry (see also Chapter 21). Matt and coworkers (2008) provide a comprehensive description of current technologies for proteome fractionation and separation of biological samples.

In order to allow direct comparisons with gene expression measurements, the same tissue sample can be extracted to provide a source for both RNA and protein. This is possible using TriZOL reagent or, more recently, kits such as the All-in-One Purification Kit (Norgen Biotek), which purifies RNA, DNA, and proteins from the same piece of tissue.

The model of intestinal inflammation again provides an example of workflow in proteomics. The proteomic method that has been commonly employed within Nutrigenomics New Zealand is 2D gel electrophoresis using DIGE technology, by which proteins differentially expressed during intestinal inflammation can be recognized, and subsequently identified using mass spectrometry technology (LCMS, MS/MS). More recently, an online 2D multidimensional liquid chromatography method combined with MS differential analysis software has been utilized (Cooney *et al.*, 2007). A number of the proteins identified in this study as being differentially expressed in intestinal inflammation may be candidates for colonic biomarkers associated with CD, particularly when combined with gene and metabolite data.

23.5.3 Metabolomics

Despite its wide application in pharmacology and toxicology, the use of metabolomics is relatively new to human nutrition (Gibney *et al.*, 2005). The ultimate goal is to

TABLE 23.4 Features of three currently available expression array technologies.

Supplier	Affymetrix	Agilent	Illumina
Feature	Oligonucleotide	Oligonucleotide	Oligonucleotide
Length	25-mer	60-mer	50-mer
Color	One color	One or two color	One color
Number of unique probes	~5,500,000 (×1 sample/chip)	~44,000 (×8 samples/slide)	~48,000 (×12 samples/slide, 15-fold redundancy)

understand the effects of dietary components on human metabolic regulation. Knowledge is only slowly accumulating as to how changes in the nutrient content of the human diet lead to changes in metabolic profiles. These are complex, because they result from at least three different sets of signals:

- Nutrients and bioactive compounds absorbed from the diet,
- Xenobiotics that are absorbed and metabolized from their environmental sources,
- The large-bowel microflora also produces significant metabolic signals.

Metabolomics is the most recent of the "omics" technologies, and therefore less well advanced than either genomics or proteomics. Metabolic profiling uses pattern-recognition statistics on assigned and unassigned metabolite signals, and the collection of comprehensive data sets of identified metabolites. The technology has the potential to distinguish between different dietary treatments, which would not have been targeted with conventional techniques. The field faces significant technical challenges due to the sheer number of potential metabolites, and their widely varying chemical and physical properties. Thus, the ultimate success of this approach will be heavily dependent upon the development of libraries of small molecules to aid in metabolite identification. However, a key advantage to metabolomics is the potential for relatively non-invasive tissue collection (e.g. fecal, urine, blood, or buccal swabs).

Two key technologies for metabolomics research are NMR or mass spectrometry based. For example, urine metabolites associated with inflammation in the *mdr1a*$^{-/-}$ mouse model have been identified (Dryland *et al.*, 2008) and GC-MS technology has been applied in the case of the IL10$^{-/-}$ model. It is anticipated that either of these approaches will be applicable to human samples.

23.5.4 Biomarkers

A biomarker provides a surrogate endpoint that indicates a probability of disease. The use of biomarkers to predict disease states (e.g. blood pressure for cardiovascular disease) has been used effectively to direct clinicians' recommendations for intervention and treatment. Similarly, biomarkers can be used as indicators of disease risk reduction following dietary manipulation (Ferguson, 2008). The importance of identifying biomarkers is to identify signs of disease susceptibility as early as possible, so that interventions have a good probability to impact the progression of disease and improve outcomes. A desirable nutrigenomics research agenda should include investigation of a set of changes that lead to a desirable health endpoint. These changes should be investigated in observational studies in carefully selected populations, with nutrient/dietary pattern exposure data calibrated using specific validated exposure biomarkers. This would be a substantial intervention development/initial testing research enterprise, including small-scale human feeding studies with biomarkers of disease risk as outcomes; before there is collaborative agreement of basic, clinical, and population scientists to assess readiness for dietary intervention randomized controlled trials (RCTs).

It is beyond the scope of this chapter to comprehensively cover biomarkers, but more details are provided by Ferguson (2008). While there is value in developing and validating biomarkers for specific disease states such as cancer (Sawyers, 2008), there may be more value in general biomarkers of disease processes such as inflammation (Virani *et al.*, 2008), since this is involved in a considerable number of different diseases. Important recent advances utilize genomic technologies for biomarker discovery. For example, Matt and coworkers (2008) discuss how the broad application of proteomic technologies has the potential to accelerate our understanding of the molecular mechanisms underlying disease, and may facilitate the discovery of new drug targets and diagnostic disease markers. These authors describe the current technologies for proteome fractionation and separation of biological samples, based on their own laboratory workflow for biomarker discovery and validation. Hassan and coworkers (2008) also discuss the use of full tissue microarrays as an emerging standard for biomarker identification and validation.

23.6 INTEGRATING COMPLEX DATA SETS: DATA MANAGEMENT, BIOINFORMATICS, AND STATISTICS

The model that Nutrigenomics New Zealand has selected for tailoring foods according to genotype requires an intelligent choice of food sources, fractionating these foods to class fractions, and chemically characterizing food sources and food components for use in high-throughput nutrient sensor arrays. For example, initial screens suggested that kiwifruit and berryfruit might be good sources of foods to counter CD in specific populations (Philpott *et al.*, 2007; Sutherland *et al.*, 2008). High-throughput screens have been used to generate data for a food-component efficacy database, which is integrated into the bioinformatics platform, and then used to choose material for testing against other diet/genome targets appropriate to health, performance and/or disease. Very large data sets are generated by these means.

A major initiative is necessary to deliver the bioinformatics, biostatistics, and database resources for nutrigenomics studies. This typically involves establishment of a relational database which can be interrogated for relationships between genetic variation and disease states, plus the impact of diet in human populations. It also requires the development of formal infrastructural platforms for bioinformatics and database resources, in order to analyze microarray and proteomics

data on gene expression patterns, identify clusters of affected genes, discover human analogs, and thereby extrapolate this information to human health or disease. Reducing the dimensionality of these complex data sets remains a significant analytical challenge (Kaput and Dawson, 2007).

23.7 CONCLUSIONS

The Nutrigenomics New Zealand model provides an approach towards personalized, genotype-based nutrition that has the potential to provide food products and personalized advice to benefit health at the individual or population level. There is convincing evidence that SNPs in certain genes may profoundly influence the biological response to nutrients. However, effects of single-gene variants on risk or risk factor levels of a complex disease tend to be small and inconsistent. Increased sensitivity of current biological measurements, plus methods of integrating information on combinations of relevant SNPs or CNVs in different genes, will become necessary to move the field to a higher dimension. Many of the challenges are in bioinformatics, especially in relation to reducing the complexity of multidimensional data sets (Kaput and Dawson, 2007). To date, there are only sporadic examples of clinical trials utilizing these technologies, and we have not investigated potential adverse effects of a genotype-derived dietary intervention. There are a considerable number of issues to be addressed before genomic approaches can become an acceptable method to guide food development or nutritional recommendations.

ACKNOWLEDGEMENTS

Nutrigenomics New Zealand is a collaboration between AgResearch Ltd, Crop and Food Research, HortResearch, and The University of Auckland, with funding through the Foundation for Research Science and Technology (FRST). M.P.G. Barnett is funded by FRST Postdoctoral Fellowship AGRX0504.

REFERENCES

Allison, M. (2008). Is personalized medicine finally arriving? *Nat. Biotechnol.* 26, 509–517.

Balish, E., and Warner, T. (2002). Enterococcus faecalis induces inflammatory bowel disease in interleukin-10 knockout mice. *Am. J. Pathol.* 160, 2253–2257.

Banner, K.H., Cattaneo, C., Le Net, J.L., Popovic, A., Collins, D., and Gale, J.D. (2004). Macroscopic, microscopic and biochemical characterisation of spontaneous colitis in a transgenic mouse, deficient in the multiple drug resistance 1a gene. *Br. J. Pharmacol.* 143, 590–598.

Barber, J.C., Maloney, V., Hollox, E.J., Stuke-Sontheimer, A., du Bois, G., Daumiller, E., Klein-Vogler, U., Dufke, A., Armour, J.A., and Liehr, T. (2005). Duplications and copy number variants of 8p23.1 are cytogenetically indistinguishable but distinct at the molecular level. *Eur. J. Hum. Genet.* 13, 1131–1136.

Barnett, M., McNabb, W., Cookson, A., Zhu, S., Davy, M., Knoch, B., Nones, K., Hodkingson, A., and Roy, N. (2008). Changes in colon gene expression associated with increased colon inflammation in interleukin-10 gene deficient mice inoculated with Enterococcus species. *Physiol. Genomics*. (submitted).

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* 57, 289–300.

Berg, D.J., Davidson, N., Kuhn, R., Muller, W., Menon, S., Holland, G., Thompson-Snipes, L., Leach, M.W., and Rennick, D. (1996). Enterocolitis and colon cancer in interleukin-10-deficient mice are associated with aberrant cytokine production and CD4(+) TH1-like responses. *J. Clin. Invest.* 98, 1010–1020.

Bilsborough, J., and Viney, J.L. (2004). Out, out darn toxin: the role of MDR in intestinal homeostasis. *Gastroenterology* 127, 339–340.

Bingham, S.A., Gill, C., Welch, A., Day, K., Cassidy, A., Khaw, K.T., Sneyd, M.J., Key, T.J., Roe, L., and Day, N.E. (1994). Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. *Br. J. Nutr.* 72, 619–643.

Brant, S.R., Panhuysen, C.I., Nicolae, D., Reddy, D.M., Bonen, D.K., Karaliukas, R., Zhang, L., Swanson, E., Datta, L.W., Moran, T., Ravenhill, G., Duerr, R.H., Achkar, J.P., Karban, A.S., and Cho, J.H. (2003). MDR1 Ala893 polymorphism is associated with inflammatory bowel disease. *Am. J. Hum. Genet.* 73, 1282–1292.

Burton, P.R., Tobin, M.D., and Hopper, J.L. (2005). Key concepts in genetic epidemiology. *Lancet* 366, 941–951.

Carlson, C.S., Eberle, M.A., Kruglyak, L., and Nickerson, D.A. (2004). Mapping complex disease loci in whole-genome association studies. *Nature* 429, 446–452.

Castro-Santos, P., Suarez, A., Lopez-Rivas, L., Mozo, L., and Gutierrez, C. (2006). TNFalpha and IL-10 gene polymorphisms in inflammatory bowel disease. Association of -1082 AA low producer IL-10 genotype with steroid dependency. *Am. J. Gastroenterol.* 101, 1039–1047.

Clayton, D.G., Walker, N.M., Smyth, D.J., Pask, R., Cooper, J.D., Maier, L.M., Smink, L.J., Lam, A.C., Ovington, N.R., Stevens, H. E., Nutland, S., Howson, J.M., Faham, M., Moorhead, M., Jones, H. B., Falkowski, M., Hardenbol, P., Willis, T.D., and Todd, J.A. (2005). Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.* 37, 1243–1246.

Cohen, S.L., Moore, A.M., and Ward, W.E. (2005). Flaxseed oil and inflammation-associated bone abnormalities in interleukin-10 knockout mice. *J. Nutr. Biochem.* 16, 368–374.

Cooney, J.M., Jensen, D.J., Laing, W., Barnett, M.P.G., McNabb, W.C., and Roy, N.C. (2007). Proteomic differential expression of colonic biomarkers associated with Crohn's disease in a mouse model using on-line 2D-LCMS and DeCyder MS. *Proc. 55th American Society of Mass Spectrometry Conference.*

den Dunnen, J.T., and White, S.J. (2006). MLPA and MAPH: sensitive detection of deletions and duplications. *Curr. Protoc. Hum. Genet.* Chapter 7, Unit 7.14.

Dommels, Y.E., Butts, C., Zhu, S.T., Davy, M., Martell, S., Hedderly, D., Barnett, M.P., Broadley, K., and Roy, N.C. (2007). Characterization of inflammation in mdr1a −/− mice and identification of related gene expression changes in colonic epithelial cells. *Genes Nutr.* 2, 209–223.

Dryland, P.A., Love, D.R., Walker, M.F., Dommels, Y., Butts, C., Rowan, D., Roy, N.C., Helsby, N., Browning, B.L., Zhu, S., Copp, B.R., and Ferguson, L.R. (2008). Allantoin as a biomarker of inflammation in an inflammatory bowel disease mouse model: NMR analysis of urine. *Open Bioactive Compounds J.* 1, 6.

Duffus, J.H., Nordberg, M., and Templeton, D.M. (2007). IUPAC glossary of terms used in toxicology. *Pure Appl. Chem.* 79, 192.

Ferguson, L.R. (2008). Biomarkers as endpoints in intervention studies. In *Molecular Epidemiology of Chronic Diseases: New Techniques and Approaches* (Vineis, P., Garte, S., and Wild, C., eds), pp. 255–266. John Wiley and Sons Ltd., Hoboken, USA.

Ferguson, L.R., and Philpott, M. (2007). Nutrigenomics and chronic inflammation. In *Personalized Nutrition: Principles and Applications* (Kok, F., Bouwman, L., and Desiere, F., eds), pp. 49–59. Taylor and Francis CRC Press, Routledge, USA.

Ferguson, L.R., Shelling, A.N., Browning, B.L., Huebner, C., and Petermann, I. (2007a). Genes, diet and inflammatory bowel disease. *Mutat. Res.* 622, 70–83.

Ferguson, L.R., Philpott, M., and Dryland, P. (2007b). Nutrigenomics in the whole-genome scanning era: Crohn's disease as example. *Cell Mol. Life Sci.* 64, 3105–3118.

Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. *Nat. Rev. Genet.* 7, 85–97.

Freeman, J.L., Perry, G.H., Feuk, L., Redon, R., McCarroll, S.A., Altshuler, D.M., Aburatani, H., Jones, K.W., Tyler-Smith, C., Hurles, M.E., Carter, N.P., Scherer, S.W., and Lee, C. (2006). Copy number variation: new insights in genome diversity. *Genome Res.* 16, 949–961.

Gibney, M.J., Walsh, M., Brennan, L., Roche, H.M., German, B., and van Ommen, B. (2005). Metabolomics in human nutrition: opportunities and challenges. *Am. J. Clin. Nutr.* 82, 497–503.

Greco, L., Romino, R., Coto, I., Di Cosmo, N., Percopo, S., Maglio, M., Paparo, F., Gasperi, V., Limongelli, M.G., Cotichini, R., D'Agate, C., Tinto, N., Sacchetti, L., Tosi, R., and Stazi, M.A. (2002). The first large population based twin study of coeliac disease. *Gut* 50, 624–628.

Hassan, S., Ferrario, C., Mamo, A., and Basik, M. (2008). Tissue microarrays: emerging standard for biomarker validation. *Curr. Opin. Biotechnol.* 19, 19–25.

Hegazi, R.A., Saad, R.S., Mady, H., Matarese, L.E., O'Keefe, S., and Kandil, H.M. (2006). Dietary fatty acids modulate chronic colitis, colitis-associated colon neoplasia and COX-2 expression in IL-10 knockout mice. *Nutrition* 22, 275–282.

Ho, G.T., Moodie, F.M., and Satsangi, J. (2003). Multidrug resistance 1 gene (P-glycoprotein 170): an important determinant in gastrointestinal disease? *Gut* 52, 759–766.

Ho, G.T., Soranzo, N., Nimmo, E.R., Tenesa, A., Goldstein, D.B., and Satsangi, J. (2006). ABCB1/MDR1 gene determines susceptibility and phenotype in ulcerative colitis: discrimination of critical variants using a gene-wide haplotype tagging approach. *Hum. Mol. Genet.* 15, 797–805.

Hooper, L., Thompson, R.L., Harrison, R.A., Summerbell, C.D., Ness, A.R., Moore, H.J., Worthington, H.V., Durrington, P.N., Higgins, J.P., Capps, N.E., Riemersma, R.A., Ebrahim, S.B., and Davey Smith, G. (2006). Risks and benefits of omega 3 fats for mortality, cardiovascular disease, and cancer: systematic review. *BMJ* 332, 752–760.

Huebner, C., Petermann, I., Browning, B.L., Shelling, A.N., and Ferguson, L.R. (2007). Triallelic single nucleotide polymorphisms and genotyping error in genetic epidemiology studies: MDR1 (ABCB1) G2677/T/A as an example. *Cancer Epidemiol. Biomarkers Prev.* 16, 1185–1192.

Joost, H.G., Gibney, M.J., Cashman, K.D., Gorman, U., Hesketh, J.E., Mueller, M., van Ommen, B., Williams, C.M., and Mathers, J.C. (2007). Personalised nutrition: status and perspectives. *Br. J. Nutr.* 98, 26–31.

Jurjus, A.R., Khoury, N.N., and Reimund, J.M. (2004). Animal models of inflammatory bowel disease. *J. Pharmacol. Toxicol. Methods* 50, 81–92.

Kang, S.S., Bloom, S.M., Norian, L.A., Geske, M.J., Flavell, R.A., Stappenbeck, T.S., and Allen, P.M. (2008). An antibiotic-responsive mouse model of fulminant ulcerative colitis. *PLoS Med.* 5, e41.

Kaput, J. (2008). Nutrigenomics research for personalized nutrition and medicine. *Curr. Opin. Biotechnol.* 19, 110–120.

Kaput, J., and Dawson, K. (2007). Complexity of type 2 diabetes mellitus data sets emerging from nutrigenomic research: a case for dimensionality reduction? *Mutat. Res.* 622, 19–32.

Kleemann, R., Verschuren, L., van Erk, M.J., Nikolsky, Y., Cnubben, N.H., Verheij, E.R., Smilde, A.K., Hendriks, H.F., Zadelaar, S., Smith, G.J., Kaznacheev, V., Nikolskaya, T., Melnikov, A., Hurt-Camejo, E., van der Greef, J., van Ommen, B., and Kooistra, T. (2007). Atherosclerosis and liver inflammation induced by increased dietary cholesterol intake: a combined transcriptomics and metabolomics analysis. *Genome Biol.* 8, R200.

Klein, W., Tromm, A., Griga, T., Fricke, H., Folwaczny, C., Hocke, M., Eitner, K., Marx, M., Runte, M., and Epplen, J.T. (2000). The IL-10 gene is not involved in the predisposition to inflammatory bowel disease. *Electrophoresis* 21, 3578–3582.

Knoch, B., Barnett, M., McNabb, W., Park-Ng, Z., Zhu, S., Nones, K., Dommels, Y., Knowles, S., and Roy, N. (2008). Genome-wide analysis of genes involved in eicosapentaenoic acid-induced modulation of colonic inflammation. *J. Nutrigenet. Nutrigenomics*. (accepted).

Kornman, K., Rogus, J., Roh-Schmidt, H., Krempin, D., Davies, A.J., Grann, K., and Randolph, R.K. (2007). Interleukin-1 genotype-selective inhibition of inflammatory mediators by a botanical: a nutrigenetics proof of concept. *Nutrition* 23, 844–852.

Kuhn, R., Lohler, J., Rennick, D., Rajewsky, K., and Muller, W. (1993). Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* 75, 263–274.

Lappalainen, M., Halme, L., Turunen, U., Saavalainen, P., Einarsdottir, E., Farkkila, M., Kontula, K., and Paavola-Sakki, P. (2008). Association of IL23R, TNFRSF1A, and HLA-DRB1*0103 allele variants with inflammatory bowel disease phenotypes in the finnish population. *Inflamm. Bowel Dis.* (in press).

MacGregor, A.J., Snieder, H., Schork, N.J., and Spector, T.D. (2000). Twins. Novel uses to study complex traits and genetic diseases. *Trends Genet.* 16, 131–134.

Madsen, K.L. (2001). Inflammatory bowel disease: lessons from the IL-10 gene-deficient mouse. *Clin. Invest. Med.* 24, 250–257.

Matt, P., Fu, Z., Fu, Q., and Van Eyk, J.E. (2008). Biomarker discovery: proteome fractionation and separation in biological samples. *Physiol. Genomics* 33, 12–17.

McCarthy, J., O'Mahony, L., O'Callaghan, L., Sheil, B., Vaughan, E.E., Fitzsimons, N., Fitzgibbon, J., O'Sullivan, G.C., Kiely, B., Collins, J.K., and Shanahan, F. (2003). Double blind, placebo controlled trial of two probiotic strains in interleukin 10 knockout mice and mechanistic link with cytokine balance. *Gut* 52, 975–980.

Moller, P., and Loft, S. (2004). Interventions with antioxidants and nutrients in relation to oxidative DNA damage and repair. *Mutat. Res.* 551, 79–89.

Moller, P., and Loft, S. (2006). Dietary antioxidants and beneficial effect on oxidatively damaged DNA. *Free Radic. Biol. Med.* 41, 388–415.

Morton, N.E. (1982). *Outline of Genetic Epidemiology*. S. Karger Publishing, Basel, Switzerland.

Morton, N.E. (2005). Linkage disequilibrium maps and association mapping. *J. Clin. Invest.* 115, 1425–1430.

Nones, K., Dommels, Y., Martell, S., Butts, C., McNabb, W., Park-Ng, Z., Zhu, S., Hedderley, D., Barnett, M., and Roy, N. (2008a). The effects of dietary curcumin and rutin on colon inflammation and gene expression in mdr1a − /− mice, a model of inflammatory bowel diseases. *Br. J. Nutr.* (accepted).

Nones, K., Dommels, Y., Martell, S., Butts, C., McNabb, W., Park-Ng, Z., Barnett, M., Zhu, S.-T., Hedderley, D., and Roy, N. (2008b). Effect

of green tea extract on colon inflammation and gene expression in mdr1a −/− mice. *J. Nutrigenet. Nutrigenomics* 1, 82.

Ordovas, J.M. (2006). Genetic interactions with diet influence the risk of cardiovascular disease. *Am. J. Clin. Nutr.* 83(2), 443S–446S.

Panwala, C.M., Jones, J.C., and Viney, J.L. (1998). A novel model of inflammatory bowel disease: mice deficient for the multiple drug resistance gene, mdr1a, spontaneously develop colitis. *J. Immunol.* 161, 5733–5744.

Park, E.I., Paisley, E.A., Mangian, H.J., Swartz, D.A., Wu, M.X., O'Morchoe, P.J., Behr, S.R., Visek, W.J., and Kaput, J. (1997). Lipid level and type alter stearoyl CoA desaturase mRNA abundance differently in mice with distinct susceptibilities to diet-influenced diseases. *J. Nutr.* 127, 566–573.

Philpott, M., Mackay, L., Ferguson, L.R., Forbes, D., and Skinner, M. (2007). Cell culture models in developing nutrigenomics foods for inflammatory bowel disease. *Mutat. Res.* 622, 94–102.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., and Hurles, M.E. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454.

Risch, N. (2001). The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol. Biomarkers Prev.* 10, 733–741.

Roberts, R.L., Gearry, R.B., Hollis-Moffatt, J.E., Miller, A.L., Reid, J., Abkevich, V., Timms, K.M., Gutin, A., Lanchbury, J.S., Merriman, T.R., Barclay, M.L., and Kennedy, M.A. (2007). IL23R R381Q and ATG16L1 T300A are strongly associated with Crohn's disease in a study of New Zealand Caucasians with inflammatory bowel disease. *Am. J. Gastroenterol.* 102, 2754–2761.

Roy, N., Barnett, M., Knoch, B., Dommels, Y., and McNabb, W. (2007). Nutrigenomics applied to an animal model of inflammatory bowel diseases: transcriptomic analysis of the effects of eicosapentaenoic acid- and arachidonic acid-enriched diets. *Mutat. Res.* 622, 103–116.

Sawyers, C.L. (2008). The cancer biomarker problem. *Nature* 452, 548–552.

Schultz, M., Veltkamp, C., Dieleman, L.A., Grenther, W.B., Wyrick, P.B., Tonkonogy, S.L., and Sartor, R.B. (2002). *Lactobacillus plantarum* 299V in the treatment and prevention of spontaneous colitis in interleukin-10-deficient mice. *Inflamm. Bowel Dis.* 8, 71–80.

Schwab, M., Schaeffeler, E., Marx, C., Fromm, M.F., Kaskas, B., Metzler, J., Stange, E., Herfarth, H., Schoelmerich, J., Gregor, M., Walker, S., Cascorbi, I., Roots, I., Brinkmann, U., Zanger, U.M., and Eichelbaum, M. (2003). Association between the C3435T MDR1 gene polymorphism and susceptibility for ulcerative colitis. *Gastroenterology* 124, 26–33.

Shen, F., Huang, J., Fitch, K.R., Truong, V.B., Kirby, A., Chen, W., Zhang, J., Liu, G., McCarroll, S.A., Jones, K.W., and Shapero, M.H. (2008). Improved detection of global copy number variation using high density, non-polymorphic oligonucleotide probes. *BMC Genet.* 9, 27.

Smyth, G.K. (2005). Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (Carey, G.R.V., Dudoit, S., Irizarry, R., and Huber, W., eds), pp. 397–420. Springer, New York.

Sullivan, P.F., Kendler, K.S., and Neale, M.C. (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* 60, 1187–1192.

Sutherland, J., Miles, M., Hedderley, D., Li, J., Devoy, S., Sutton, K., and Lauren, D. (2008). In vitro effects of food components on selected probiotic and pathogenic bacteria. *J. Food Sci. Nutr.* (accepted).

Tracy, R.P. (2008). "Deep phenotyping": characterizing populations in the era of genomics and systems biology. *Curr. Opin. Lipidol.* 19, 151–157.

Vercelli, D. (2008). Discovering susceptibility genes for asthma and allergy. *Nat. Rev. Immunol.* 8, 169–182.

Virani, S.S., Polsani, V.R., and Nambi, V. (2008). Novel markers of inflammation in atherosclerosis. *Curr. Atheroscler. Rep.* 10, 164–170.

Weersma, R.K., Zhernakova, A., Nolte, I.M., Lefebvre, C., Rioux, J.D., Mulder, F., van Dullemen, H.M., Kleibeuker, J.H., Wijmenga, C., and Dijkstra, G. (2008). ATG16L1 and IL23R are associated with inflammatory bowel diseases but not with celiac disease in the Netherlands. *Am. J. Gastroenterol.* 103, 621–627.

Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678.

Werner, T., and Haller, D. (2007). Intestinal epithelial cell signalling and chronic inflammation: from the proteome to specific molecular mechanisms. *Mutat. Res.* 622, 42–57.

Wilk, J.N., Bilsborough, J., and Viney, J.L. (2005). The mdr1a − /− mouse model of spontaneous colitis: a relevant and appropriate animal model to study inflammatory bowel disease. *Immunol. Res.* 31, 151–159.

Willett, W. (2005). Toward optimal health: Walter Willett, M.D., DR.P.H. discusses dietary guidelines. *J. Women's Health (Larchmt)* 14, 679–683.

Yen, D., Cheung, J., Scheerens, H., Poulet, F., McClanahan, T., McKenzie, B., Kleinschek, M.A., Owyang, A., Mattson, J., Blumenschein, W., Murphy, E., Sathe, M., Cua, D.J., Kastelein, R.A., and Rennick, D. (2006). IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6. *J. Clin. Invest.* 116, 1310–1316.

This page intentionally left blank

Chapter 24

Novel Next-Generation DNA Sequencing Techniques for Ultra High-Throughput Applications in Bio-Medicine

Wilhelm J. Ansorge
Ecole Polytechnique Federal Lausanne, EPFL, Lausanne, Switzerland

24.1 INTRODUCTION

The scientific journal *Nature Methods* (Schuster *et al*., 2008, editorial and articles) selected, as the method of the year 2007, the next generation high-throughput DNA sequencing techniques, which are opening new fascinating opportunities in biomedicine. Until a few years ago, the methods used for the sequencing were the enzymatic technique described by Sanger *et al*. (1977) and the chemical degradation method described in the same year by Maxam and Gilbert (1977), which was used in sequence cases not easily resolved with the enzymatic technique. The Sanger method has been used in the first automated sequencing project using fluorescent labeling, the sequence determination of a complete gene loci, namely the *HPRT* gene, performed using the system developed at the European Molecular Biology Laboratory in Heidelberg (http://www.embl.org). In that sequencing project, the important concept of paired-end sequencing was introduced for the first time (Edwards *et al*., 1990). The successful and unambiguous sequencing of a real 60 kb genomic DNA region, rich in Alu repeats in both directions and other sequence pitfalls, was an achievement that demonstrated not only the applicability of automated sequencing, using fluorescent dyes, to analyze DNA sequence of whole genomes, but also the technical feasibility of the Human Genome Sequencing project. Advantages of this approach included: (a) handling of non-toxic biochemicals, (b) pioneering automated sequencing with fluorescent labeling (Ansorge *et al*., 1986, 1987), (c) introduction of end-pairing sequence determination (Edwards *et al*., 1990), (d) possibility of robotic sample preparation, and (e) extensive use and handling of large amounts of data with computer systems. The two laboratories, where the first automated DNA sequencers were produced simultaneously, were the laboratory of Leroy Hood at Caltech (Smith *et al*., 1986), which was later commercialized by Applied Biosystems (ABI), and the laboratory of the author at EMBL (Ansorge *et al*., 1986, 1987), which was commercialized by Pharmacia-Amersham, later General Electric Healthcare.

When the international community decided to complete the sequence determination of the whole human genome, this goal triggered the development of techniques allowing higher sequencing throughput. In Japan, the work on fluorescent DNA sequencing technology resulted in the development of a high-throughput capillary array DNA sequencer, carried out by the team of Hideki Kambara (http://www.hitachi.com/rd/fellow_kambara.html) in Hitachi laboratories. Two companies, ABI (commercializing the system of Kambara) and Amersham (taking over and further developing the system set up in the USA by scientists at the Molecular Dynamics Company), at that time commercialized automated sequencing using parallel analysis in systems of up to 384 capillaries. Together with partial miniaturization of the robotic sample preparation, large efforts in automation of laboratory processes, accompanied by advances in new enzymes and biochemicals, the Sanger technique made it possible to determine, by two consortia in parallel, the sequence of the human genome (Venter *et al*., 2001, International Human Genome Sequencing Consortium, 2001).

The great limitation in the sequencing protocols, as the experts and developers of the DNA sequencing techniques realized, was the need for gels or polymers used as sieving separation media for the fluorescently labeled DNA fragments. Thus, there were thoughts about developing techniques without gels, which would allow sequence determination on great numbers (millions) of samples in parallel. One of the first was probably the development of such a technique at the EMBL Laboratory in Heidelberg (at that time one of the two world leaders in that field) in the years 1988–1990, and described in a patent application

in 1990 (Ansorge, 1991). In this application, a large-scale DNA sequencing technique without gels was described for the first time, namely ''Sequencing-by-Synthesis, addition and detection of the incorporated base''. This approach proposed and described the use of the so-called reversible terminator's nucleotides, for speed and efficiency. This gel-free method was proposed in 1990, which was the year when the first generation of automated sequencers had just been launched in the market, and, notably, 15 years before this principle was employed in the nowadays next-generation systems. At that time, the technique consisted of detecting the next-added fluorescently labeled base in the growing chain of DNA by means of a sensitive charge-coupled device (CCD) camera, in the presence of the enzyme DNA polymerase, on a great number of samples attached either to a planar support (DNA chips) or beads. This allowed simultaneous analysis of a large number of samples in parallel, without gels or polymers, minimizing the reaction volumes in a miniaturized microsystem. The principle described in the patent is in part very similar to that used today in the very promising so-called next-generation DNA sequencing devices, with many additional original developments commercialized by the companies Illumina-Solexa (http://www.illumina.com), Helicos (http://www.helicosbio.com), and others, as mentioned in the following paragraphs.

Since the year 2000, focused developments have been taking place at several laboratories. Other institutions, particularly the European laboratories with less visionary management and understanding of the technology significance, considered the capillary systems as the top point and stopped developments of even the most promising novel sequencing techniques, turning the attention exclusively to arrays. But in the USA, grants for development and testing of novel sequencing technologies were provided by the important funding agencies. These efforts signaled that developments were well under way to bring novel, non-gel based, high-throughput approaches to maturity. The next-generation devices and platforms available in mid-2008 on the market, as well as some interesting parallel developments, are described in more detail below. The European Union has recently initiated significant support for work on novel high-throughput DNA sequencing techniques (e.g. READNA consortium; http://www.cng.fr/READNA).

One of the limiting factors of the new technology remains the overall high cost for generation of the sequence with very high throughput, even though the cost per base is lower several order of magnitudes, compared with conventional DNA sequencing performed with the improved protocols of the technique described by Sanger and coworkers (1977). Further cost reduction by two orders of magnitude is needed to be able to realize the potential of personal genomics (see section 24.3 below), which sets the goal for cost of genome sequence to be around US$1,000.

Reduction of the sequencing errors is another point, and in this respect the Sanger sequencing method is competitive and will be used in the immediate future.

24.2 NEXT-GENERATION DNA SEQUENCING PLATFORMS

Novel DNA sequencing techniques provide high sequencing speed and throughput that were considered only a distant dream a few years ago. Genome sequencing projects that took several years with the Sanger technique could nowadays be completed in a few weeks. Another advantage of the next-generation DNA sequencing platforms is the determination of the sequence data from amplified single DNA fragments, avoiding the need for cloning of DNA fragments in plasmids.

Novel fields and applications in biology and medicine (beyond the genomic sequencing as the original development goal and application) are becoming a reality, e.g. personal genomics with detailed analysis of individual genomic stretches, precise analysis of RNA transcripts for gene expression (surpassing and replacing in several aspects analysis carried out by various microarray platforms, for instance in reliable and precise quantization of transcripts), and as a reliable tool for the identification and analysis of DNA regions that interact with regulating proteins in functional regulation of gene expression. The next-generation DNA sequencing technologies offer novel and rapid ways for genome-wide characterization and analysis of profiles of mRNAs, small RNAs, transcription-factor binding sites, genetic variations, chromatin structure and DNA methylation patterns, microbiology, and metagenomics.

As mentioned above, one of the present limitations is the high sequencing cost, but the impressive results obtained so far with the new technology in various projects are very convincing, and without any doubt further developments will be concentrated on lowering the cost. Other limitations are: (a) the short reading length in some cases, (b) non-uniform confidence in base calling in the sequence reads, particularly the deteriorating 3′ sequence quality in technologies with short reading capabilities, and (c) generally lower reading accuracy in homopolar stretches of identical bases. The huge amount of generated data by these high-throughput techniques (over 1 gigabase (Gb) per run) in the form of short readings presents another challenge to software developers, and novel, more efficient computer algorithms and more effort are being devoted to this task in bioinformatics laboratories.

24.2.1 The GS 454 Life Sciences Approach

The principle of the pyrophosphate detection, as the basis of the device, was described already in 1985 (Nyren and Lundin, 1985), and a first system using this principle in a new method for DNA sequencing was reported in 1988 (Hyman, 1988; see also Chapter 8). The technique was further developed into a routinely functioning technique by the teams of M. Ronaghi, M. Uhlen, and P. Nyren (Ronaghi *et al*., 1996), leading to a technique commercialized for the analysis of 96 samples in parallel in a microtiter plate.

In 2005, 454 Life Sciences (http://www.454.com) introduced the GS device developed by a very enthusiastic and competent team, as the first system of the next-generation DNA sequencer on the market.

In this approach, random libraries of DNA fragments are generated by shearing an entire genome and isolating single DNA molecules by limiting dilution. Subsequently, specialized common adapters are added to the fragments so that they are captured each on one bead, and amplified, within the droplets of an emulsion, using adapter-specific primers. The DNA fragment amplification is necessary to obtain the light signal intensity sufficiently strong for reliable base detection in the following sequencing reaction steps. Sequencing-by-synthesis is then simultaneously performed in open wells of a fiber-optic slide using a modified pyrosequencing protocol (Margulies *et al.*, 2005) which is designed to take advantage of the small scale of the wells. The slide, containing approximately 1.6 million wells, is loaded with beads and mounted in a flow chamber designed to create a 300 μm high channel, above the well openings, through which the sequencing reagents flow. The unetched base of the slide is in optical contact with a second fiber-optic imaging bundle bonded to a charge-coupled device (CCD) sensor, allowing the capture of emitted photons from the bottom of each individual well (Fig. 24.1).

The sequencing reaction is dominated by mass transport effects, therefore well depth is selected on the basis of a number of competing requirements: (1) wells need to be deep enough for the DNA-carrying beads to remain in the wells in the presence of convective transport past the wells; (2) they must be sufficiently deep to provide adequate isolation against diffusion of by-products from a well in which incorporation is taking place to a well where no incorporation is occurring; and (3) they must be shallow enough to allow rapid diffusion of nucleotides into the wells and rapid washing out of remaining nucleotides at the end of each flow cycle to enable high-sequencing throughput and reduced reagent use. After the flow of each nucleotide, a wash containing apyrase is used to ensure that nucleotides do not remain in any well before the next nucleotide is introduced. Nucleotide incorporation is detected by the associated release of inorganic pyrophosphate and the generation of photons. Wells containing template-carrying beads are identified by detecting a known four-nucleotide "key" sequence at the beginning of the read. Raw signals are background subtracted, normalized, and corrected. The normalized signal intensity at each nucleotide flow, for a particular well, indicates the number of nucleotides, if any, that were incorporated.

The method has recently increased the claimed achieved reading length to 400–500 bp, with paired-end reads, and as such is being applied to genomes (bacterial, animals, human) sequencing. One of the spectacular results obtained with this system was the personal human genome sequence of James Watson (Wheeler *et al.*, 2008), see also section 24.3 below. Another application of the system was the identification of the culprit in the recent honey-bee disease epidemics. Relatively high cost of operation, and generally lower reading accuracy in homopolar stretches of identical bases, are being reported as a few drawbacks of the method.

The next upgrade in the system, the 454 FLX Titanium run, will quintuple the data output from 100 Mb to about 500 Mb (or more). The new picotiter plate in the device uses

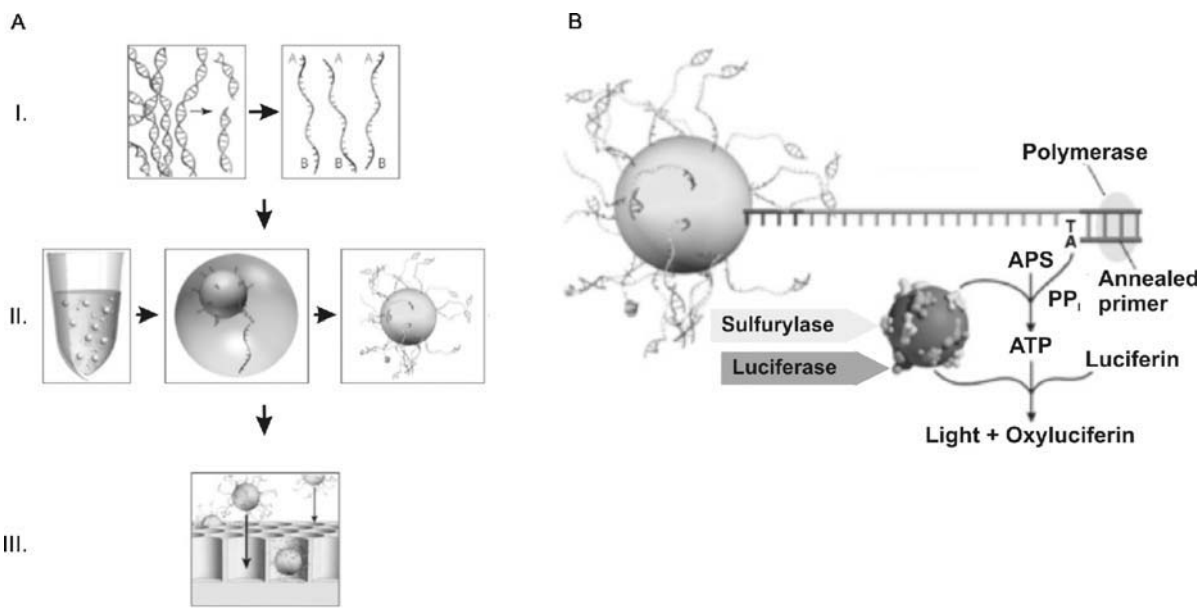


FIGURE 24.1 **A.** Outline of the GS 454 DNA sequencer workflow. Library construction (I) ligates 454-specific adapters to DNA fragments (indicated as **A** and **B**) and couples amplification beads with DNA in an emulsion PCR to amplify fragments before sequencing (II). The beads are loaded into the picotiter plate (III). **B.** Schematic drawing illustrating the pyrosequencing reaction that occurs on nucleotide incorporation to report sequencing-by-synthesis (see also text for details). Figure adapted from http://www.454.com.

much smaller beads, about 1 micron in diameter. Up until now, the primary analysis (image processing and base calling) of 454 data was able to be performed in a few hours on a moderately powerful computer. With the increased data output, primary analysis requires a small cluster.

Further developments and a list of relevant publications with applications of the 454 Life Sciences next-generation sequencer can be found at http://www.454.com/index.asp and in Schuster and coworkers (2008).

24.2.2 The Illumina-Solexa Genome Analyzer

The Solexa (later acquired by Illumina; http://illumina.com), next-generation sequencing platform was commercialized in 2006. The principle of the system is based on ingenious sequencing-by-synthesis chemistry, with novel reversible terminator nucleotides for the four bases each labeled with a different fluorescent dye, and a special DNA polymerase enzyme able to incorporate them.

DNA fragments are ligated on both ends to adapters, and after denaturation immobilized at one end on a solid support, formed by the surface of a flow cell. The surface of the support is coated densely with the adapters, as well as with the complementary adapters. Each single-stranded fragment, immobilized at one end on the surface, creates a ''bridge'' structure by bending and hybridizing its free end to the complementary adapter on the surface of the support (Fig. 24.2). In the reaction mixture containing the PCR amplification reagents, the adapters on the surface act as primers for the following PCR amplification. Again, the amplification is necessary to get the light signal intensity sufficiently strong for reliable base detection in the sequencing-by-synthesis reaction steps.

After a sufficient number of solid-phase amplification PCR cycles and denaturation, clusters of about 1,000 copies of single-stranded DNA fragments are randomly created on the surface, also known as DNA polonies (resembling cell colonies, after polymerase amplification). For the sequencing reactions, a DNA synthesis mixture containing primers, four reversible terminator nucleotides each labeled with a different fluorescent dye, and a DNA polymerase able to incorporate them in the synthesized strand, is supplied onto the surface. After incorporation in the DNA strand, the terminator nucleotide is detected by CCD camera, identified by its fluorescent dye. Subsequently, the terminator group at the 3′ end of the base and the fluorescent dye are removed from the base and the next synthesis cycle is repeated. The sequence read length achieved in the repetitive reactions is about 35 nucleotides. In parallel, the sequence in at least 40 million polonies (clusters) can be simultaneously determined, resulting in a very high sequence throughput, in the order of Gb per support.

Although the reading length of 30–35 bases may appear very short, the Solexa technology has found many interesting applications in: (a) reliable and quantitative characterization of gene expression, by sequencing of RNA transcripts, (b) characterization of mRNA, methylated, or unmethylated fragments, (c) determination of protein binding sites on DNA or RNA (ChIP-Seq technique, allowing mapping of DNA sites occupied by a DNA binding protein).

In 2008, Illumina launched an upgrade to the Genome Analyzer II that triples output compared to the previous Genome Analyzer. The company introduced a paired-end module for the sequencer, and with new optics and camera components that allow the system to image DNA clusters more efficiently over larger areas, the new instrument triples the output per paired-end run from 1 to 3 Gb. GAII uses a 4-megapixel camera to scan 800 images across the eight channels of the newly designed wider flow cells.

The system generates at least 1.5 Gb of single-read data per run, at least 3 Gb of data in a paired-end run, and more than 750 Mb of data per day, recording data from more than 50 million reads per flow cell. The run time for a 36-cycle run decreased to two days for a single-read run, and four days for a paired-end run.

Illumina currently supports 36 base-pair reads on the system, but the platform enables reads longer than 50 base pairs. The system will reach these longer reads when the chemistry is further optimized. The company will add a module called ''Integrated Primary Analysis and Reporting'' to the system, which enables quality control in real time and will provide real-time image processing to accelerate data migration and analysis. The module will be of particular interest to small and medium-sized laboratories with limited data hardware and network resources. Paired-end sequencing has demonstrated its value in sequencing of model organisms for which there is no reference, as well as for the full characterization of whole human genome sequencing.

Further developments of Illumina's Genome Analyzer and list of relevant publications with applications can be found at http://www.solexa.com and in Schuster and coworkers (2008).

24.2.3 The ABI SOLiD System

The ABI SOLiD (Sequencing by Oligo Ligation and Detection) next-generation sequencing system, a commercial platform using a unique sequencing chemistry based upon ligation and catalyzed by DNA ligase, was introduced on the market in the autumn of 2007.

The specific process couples oligonucleotide adaptor-linked DNA fragments with 1 μm magnetic beads, on which complementary oligonucleotides are attached. Libraries of DNA fragments are generated by fragmentation or mated pairing (Fig. 24.3). Each bead–DNA complex is then amplified by emulsion PCR. Again, the amplification is necessary to obtain the light signal intensity sufficiently strong for reliable detection and determination of the sequence in the following reaction steps.

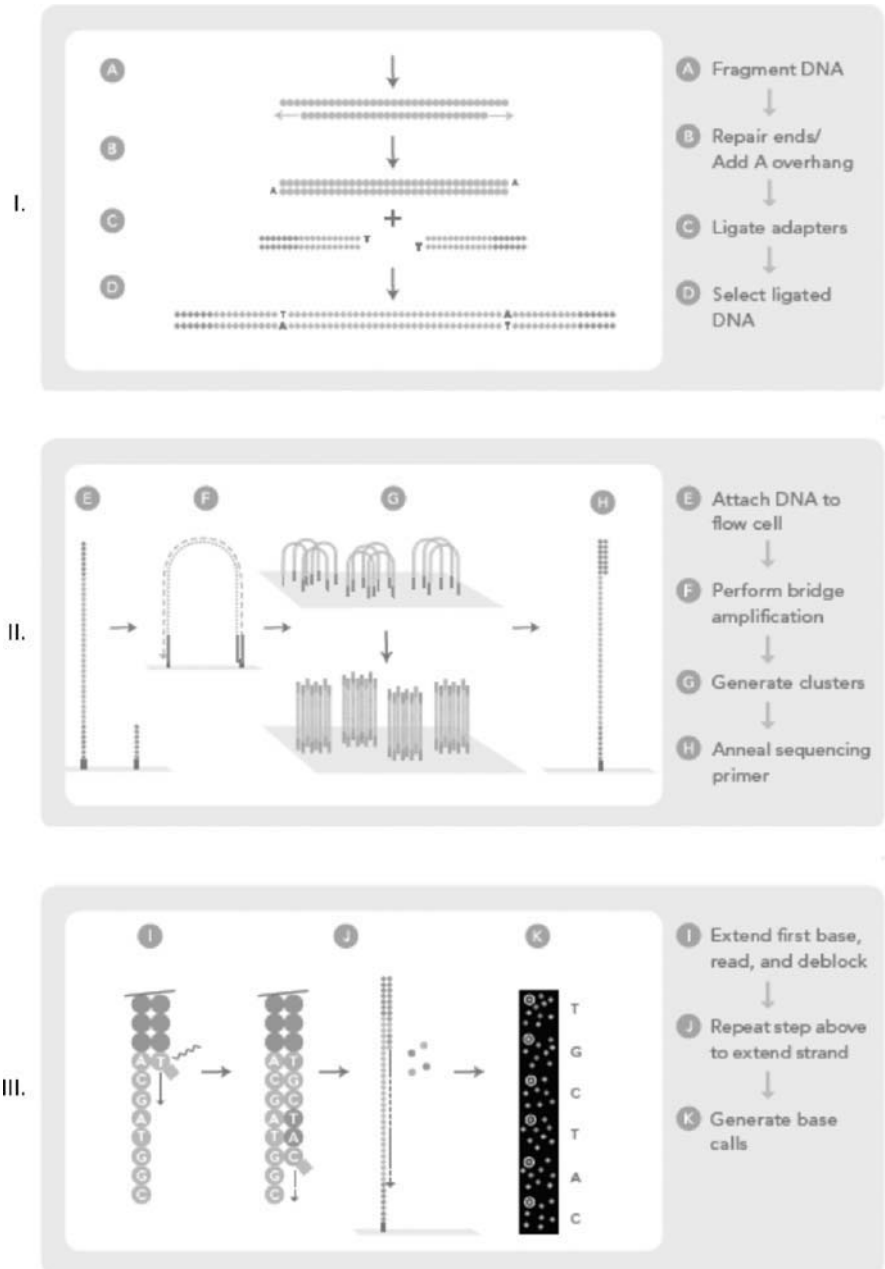


FIGURE 24.2 Outline of the Illumina Genome Analyzer workflow. Similar fragmentation and adapter ligation steps take place (I), prior to applying the library onto the solid surface of a flow cell. Attached DNA fragments form ''bridge molecules'' that are subsequently amplified, via an isothermal amplification process, leading to a cluster of identical fragments that are subsequently denatured for sequencing primer annealing (II). Finally, amplified DNA fragments are subjected to sequencing-by-synthesis using 3 blocked labeled nucleotides (III). Figure adapted from the Genome Analyzer brochure (available from http://www.solexa.com).

After amplification, the beads are covalently attached to the surface of a specially treated glass slide that is placed into a fluidics cassette within the SOLiD sequencer. The ligation-based sequencing process starts with the annealing of a universal sequencing primer that is complementary to the SOLiD-specific adapters on the library fragments. The addition of a limited set of semi-degenerate octamer oligonucleotides and DNA ligase is automated by the instrument. DNA ligation occurs when a matching octamer hybridizes to the DNA fragment sequence adjacent to the universal primer 3′ end. In these octamers, the fourth and fifth base doublet contains a unique fluorescent label at the end of the octamer, which allows sequence determination following fluorescent label detection. Subsequently, the fluorescent label is enzymatically removed together with the three last bases of the octamer, therefore enabling a subsequent ligation round. Multiple cycles of ligation, detection, and cleavage are performed to determine the

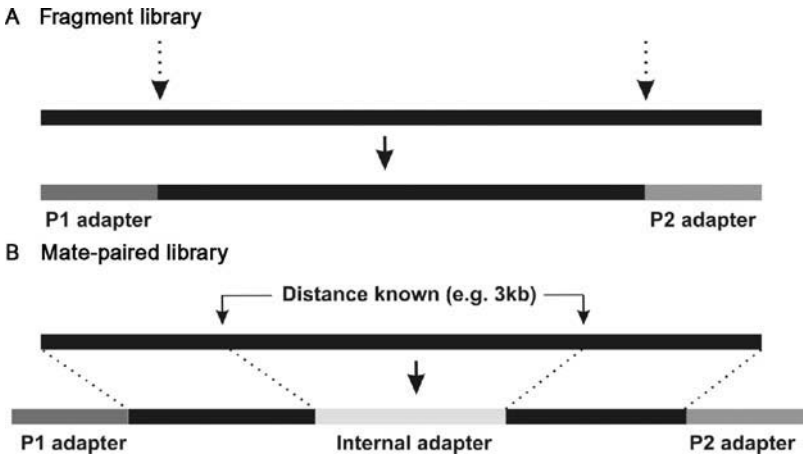


FIGURE 24.3 Library preparation for DNA sequencing using the SOLiD DNA sequencing platform. **A.** *Fragment library*: After whole genome DNA is randomly fragmented (indicated by the dashed arrows), two different 25 bp DNA adapters (P1 and P2) are ligated at the 5′ and 3′ ends of the DNA fragments generated, respectively. **B.** *Mate-paired library*: In this case, DNA fragments that are separated from another DNA fragment of known length (e.g. 3 kb for this example) are ligated so that they encompass an internal adapter. Subsequently, two different DNA adapters are ligated at the 5′ and 3′ ends, respectively, similarly to the situation described in (A).

DNA sequence of bases 9 and 10, 14 and 15, and so on until the eventual read length is achieved, presently about 35 bp. Following this series of reactions, sequencing process continues in the same way with another primer, this time shorter by one base than the previous one, allowing determining in the successive cycles the DNA sequence of bases 3 and 4, 8 and 9, 13 and 14, and so on (Fig. 24.4). Therefore, five rounds of sequence reset are sufficient to address the DNA sequence of each library fragment.

The unique attribute of this ligation-based approach and the octamer labeling is that an extra quality check of read accuracy is enabled. This essentially relies on the known fixed nucleotide identities in the octamer sequences to identify erroneous base calls from true nucleotide variations during the data analysis step. Also, due to the fact that each base is determined with a different fluorescent label, the error rate in the sequence is reduced. In parallel, the sequence in more than 50 million bead clusters can be simultaneously determined, resulting in a very high sequence throughput, on the order of Gb per run.

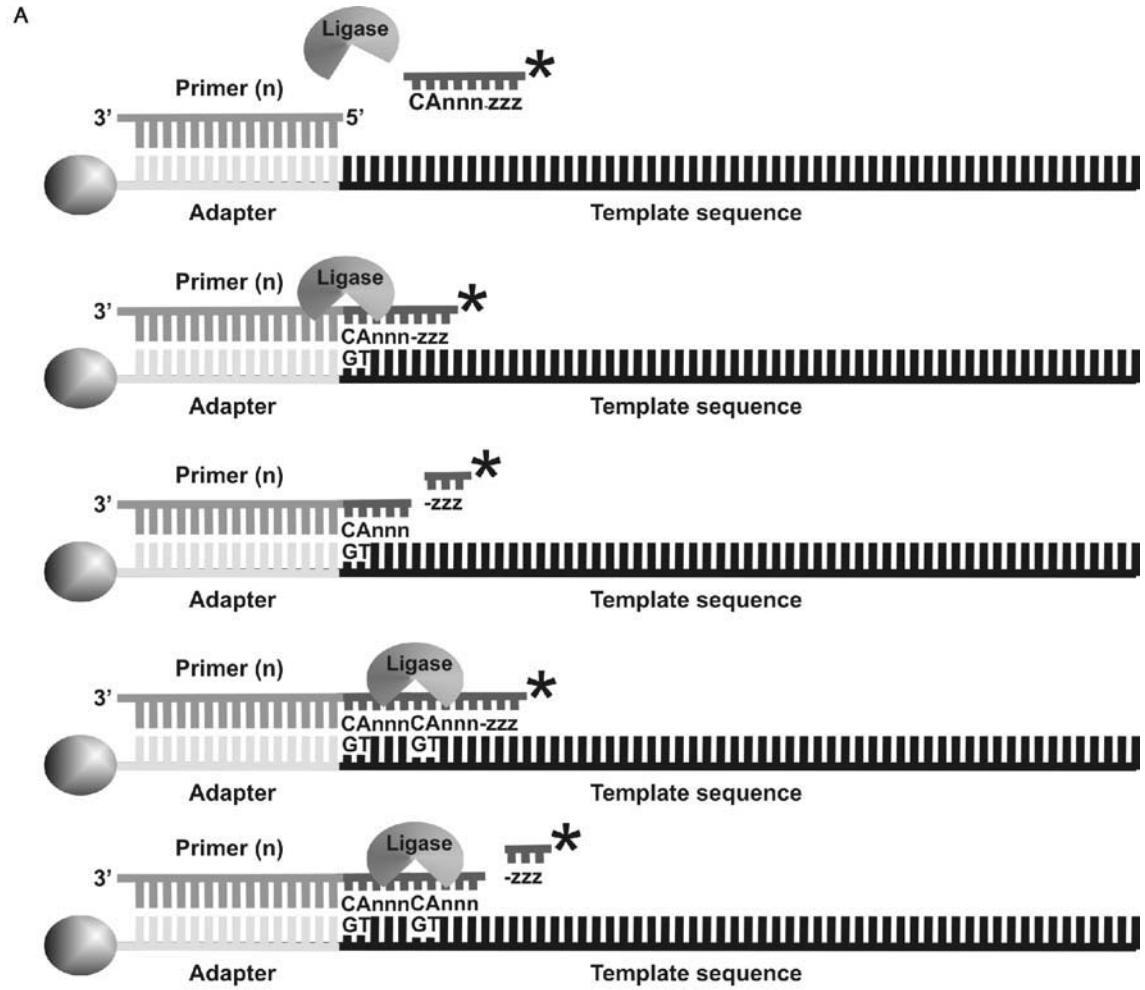
In 2008, Applied Biosystems launched an updated version of SOLiD, called SOLiD 2.0 platform, which will increase the output of the instrument from 3 to 10 Gb per run and reduce the overall run time. The SOLiD upgrade, which does not involve changes to the hardware of the sequencer, consists of new chemistry, fine-tuned software, and an improved workflow, replacing the existing emulsion maker with a smaller device, improving the sample preparation workflow. Combined with an optimized emulsion PCR protocol, this change would also increase the bead load with template, thereby reducing the overall run time on the new system to 4.5 days instead of 8.5 days on the existing machine for a fragment library, and 8 days instead of 12 to 15 days for a mate-paired library. Also, it is estimated that employing one instead of two oligonucleotide ligation probes sets will shorten the ligation reactions and decrease the sequencing error rate by 30%.

More information about the ABI SOLiD system, and list of publications with applications, can be found at: http://marketing.appliedbiosystems.com/images/Product/Solid_Knowledge/flash/102207/solid.html and in Schuster and coworkers (2008).

24.2.4 Helicos Single-Molecule Sequencing Device, HeliScope

The next generation systems, discussed in the previous paragraphs, required the emulsion PCR amplification step of DNA fragments, to ensure sufficiently strong light signal to be detected by the CCD cameras, allowing accurate DNA sequence determination. The PCR amplification has revolutionized the DNA analysis, but in some instances it may introduce sequence errors in the copied DNA strands, favor a certain sequence over another, or change the relative frequency and abundance of various DNA fragments that existed in the original *in vivo* situation before amplification.

The ultimate miniaturization into nanotechnology scale, and the accompanying ultimate minimal use of biochemicals, would be achieved if the sequence could be determined directly from one single DNA molecule, without the need for PCR amplification and its potential for distortion of abundance levels. This would require a very sensitive light detection system and physical arrangement, capable of light detection and identification from a single dye molecule. Techniques for detection and analysis of single photons have been under intensive development over the past decades, and several commercially available sensitive systems have been produced and tested. One of the first techniques for sequencing from a single DNA molecule



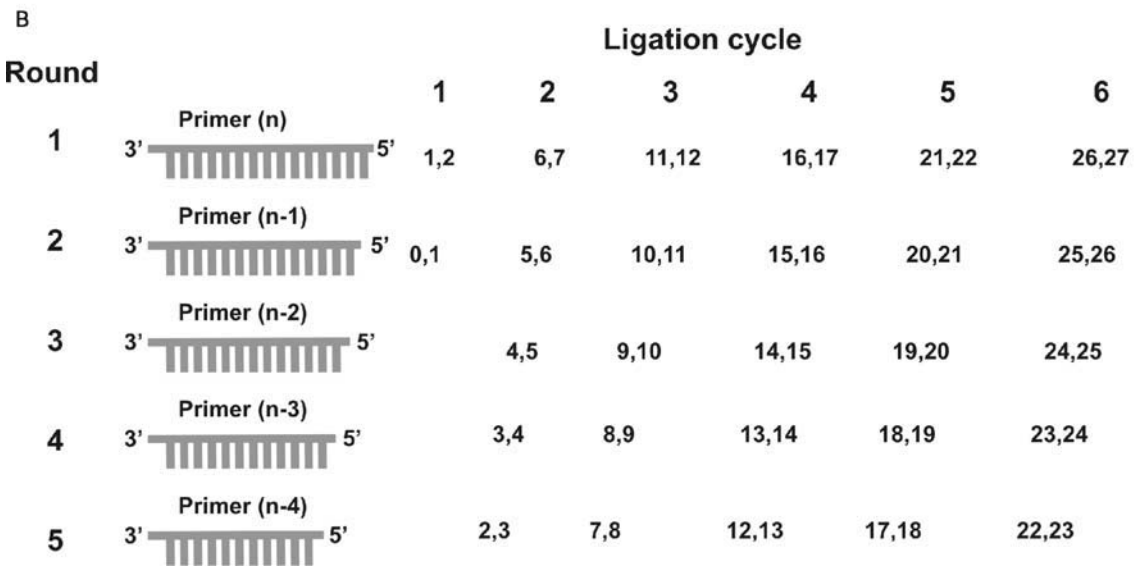


FIGURE 24.4 Sequencing-by-ligation, using the SOLiD DNA sequencing platform. **A.** Primers hybridize to the P1 adapter within the library template. A set of four fluorescently labeled di-base probes compete for ligation to the sequencing primer. These probes have partly degenerate DNA sequence (indicated by ''n'' and ''z'') and for simplicity, only one probe is shown in this example (labeling is denoted here by an asterisk). Specificity of the di-base probe is achieved by interrogating the 1st and 2nd base in each ligation reaction (CA in this case for the complementary strand). Following ligation, the fluorescent label is enzymatically removed together with the three last bases of the octamer. **B.** Sequence determination, using the SOLiD DNA sequencing platform is performed in multiple ligation cycles, using different primers, each one shorter from the previous one by a single base. The number of ligation cycles (six for this example) determines the eventual read length, while for each sequence tag, five rounds of primer reset occur (from primer (n) to primer (n-4)). The dinucleotide positions on the template sequence that are interrogated each time are depicted underneath each ligation cycle and are separated by 5 bp from the dinucleotide position interrogated in the subsequent ligation cycle.

was described by the team of S. Quake in 2003 (Braslavsky *et al*., 2003), and licensed by Helicos Biosciences.

In late 2007, the Helicos Company (http://www.helicosbio.com) introduced the first commercial DNA sequencing system operating without the PCR or cloning amplification, determining the sequence from one single DNA molecule. In the Helicos system nucleic acid fragments are hybridized to primers that are covalently anchored in random positions on a solid surface (glass cover slip) in a flow cell. After attachment of the single DNA fragments to the solid surface, labeled nucleotides together with primers and polymerase enzyme are added. The next base incorporated in the synthesized DNA strand is determined, as its light signal that it emits is analyzed by very sensitive photon detections systems in the sequencing-by-synthesis technique (similar to Fig. 24.2, but on only one DNA fragment, without amplification). The device allows the analysis of many millions of single DNA fragments simultaneously, resulting in the sequence throughput in the Gb range, without the need for PCR amplification.

Although this system is still in its first years of operation, it has been tested and validated in several applications with promising results, and improvements in operation have been continued in collaboration with the users. A very promising application of the technique is in the pre-natal Trisomia-21 test (Down-Syndrome) using only the maternal blood sample, potentially replacing the up-to-present standard test associated with some risk to the fetus (Fan *et al.*, 2008).

The Helicos system was used to sequence the roughly 7,000-nucleotide genome of the M13 virus (http://www.helicosbio.com). In this project, the team used single-molecule sequencing on more than 280,000 primer-template duplexes at once. In total, the M13 genomic DNA went through 224 sequencing cycles. The approach reportedly enabled 100% coverage of the double-stranded DNA form of the M13 phage's 6,407-nucleotide genome, by sequencing to a depth of more than 150-fold. Read lengths averaged about 23 bases, and could be increased by using more cycles. There were still some limitations in the single-molecule technology used in the project, based on the first generation of the company's single molecule chemistry. For example, it has been shown that in a run of bases of the same kind, in the homopolar regions, multiple fluorophore incorporations could decrease emissions, sometimes below the level of detection. When errors did occur, most of them were deletions.

Recently, Helicos announced that it has developed a new generation of 'one-base-at-a-time' nucleotides which allow more accurate homopolymer sequencing, and lower overall error rates. With future improvements the method will allow an affordable, high-throughput, comparative sequencing of thousands of human genomes. More information and detailed explanation about the Helicos system, schema of operation, its further developments and list of publications with applications can be found on the company website http://www.helicosbio.com/.

24.2.5 NEXT-GENERATION TECHNIQUES UNDER DEVELOPMENT

Development of novel DNA sequencing techniques is taking place in many laboratories worldwide. In the laboratory of George Church at Cambridge, MA (Shendure *et al*., 2005), the so-called multiplex polony technology has been developed, using also the sequencing-by-synthesis method described above. Several hundred sequencing templates are deposited on agarose thin layers, and the sequence in each of them is determined in parallel. This presents a several order of magnitudes increase in the number of samples which can be analyzed simultaneously. Further advantages are the large reduction of reaction volumes, smaller amounts of biochemicals needed, and the resulting lower cost of sequencing. The laboratory continues development of their platform, and offers this technique to academic laboratories at relatively low cost, using off-the-shelf optics, hardware, and biochemical reagents.

A very promising approach, attempting to use real-time single-molecule DNA sequence determination for sequencing without the need for cloning, is the technique being developed by the company VisiGen Biotechnologies Inc. (http://visigenbio.com). In this very unique approach, the researchers have produced a specially engineered DNA polymerase (acts as a ''real-time sensor'' for modified nucleotides), with incorporated donor fluorescent dye close to the active site involved in selection of the nucleotides for integration in the growing DNA chain during synthesis. All four nucleotides to be integrated have been modified, each with a different acceptor fluorescent dye. During the synthesis of the new strand, the polymerase selects the next nucleotide to be integrated in the DNA strand (complementarily according to the base pairing principle). When the correct nucleotide is found, selected, and entered in the active site of the enzyme, the donor dye label in the polymerase comes into very close proximity with the acceptor dye label on the nucleotides and energy is transferred from donor to acceptor dye giving rise to Fluorescent Resonance Energy Transfer (FRET) light signal. The frequency of this signal varies in dependency on the fluorescent dye label incorporated in the nucleotides, and by recording frequencies of the emitted FRET signals it will be possible to determine sequence of the bases, at the speed that the polymerase can integrate the nucleotides during the synthesis process (usually a few hundred per second). The acceptor fluorophore dye is removed during nucleotide incorporation, which ensures that there are no DNA modifications that might slow down the polymerase during synthesis.

The company is working on refining its chemistry and detection technologies. In particular, the team is working on slowing down the nucleotide incorporation step in order to detect a stronger fluorescence signal. By multiplexing a large number of samples in parallel, the sequence throughput could be in the gigabase range. VisiGen Biotechnologies

plans to offer a service based on its real-time single-molecule nano-sequencing technology by the end of 2009, and to follow that with the launch of equipment and reagents in another 18 months. The technology could eventually enable researchers to sequence an entire human genome in less than a day for under $1,000. The company is currently working on its first version of the instrument, which can generate around 4 Gb of data per day. At that throughput, the technology could sequence 44 human genomes per year at ten-fold coverage for around $1,000 per genome. With future modifications there is a potential to increase the current throughput of 1 Mb per second to 50 Mb per second.

The single-molecule approach requires no amplification, which eliminates a large part of the cost relative to current sequencing technologies. In addition, read lengths for the instrument are expected to be around 1 kb, which is longer than any current next-generation sequencing platform. For example (see above), the GS FLX system from Roche's 454 Life Sciences generates reads of around 250 base pairs, and the company expects to extend that to 400 bases in the next version. Illumina's Genome Analyzer and Applied Biosystems' SOLiD sequencers both have reads of around 30–35 base pairs.

Another company in the USA, Pacific Biosciences (http://www.pacificbiosciences.com), announced recently that it is working on a next-generation DNA sequencing instrument that it believes will eventually be able to produce 100 Gb of sequence data per hour, or a diploid human genome at one-fold coverage in about 4 minutes. They plan to sell its first DNA sequencing systems in 2010. The price of the instrument will likely be in the range of those next-generation sequencers sold by 454/Roche, Illumina, and Applied Biosystems, that is approximately $400,000 to $600,000. The company's single-molecule, real-time, or SMRT, technology is based on zero mode waveguides (ZMWs) that were originally developed in the Nanobiotechnology Center at Cornell University. In principle, ZMWs are nanometer-scale holes in a 100-nanometer metal film deposited on a clear substrate. Due to the behavior of light aimed at such a small chamber, the observation volume is only 20 zeptoliters, enabling researchers to measure the fluorescence of nucleotides incorporated by a single DNA polymerase enzyme into a growing DNA strand in real time. On a prototype system, PacBio researchers have so far observed read lengths of a little over 1,500 bases and a rate of 10 bases per second, and have been able to analyze up to 3,000 ZMWs in parallel.

Another single-molecule sequencing technique may develop from studies on translocation of DNA through various artificial nanopores. The work in this field was pioneered at Harvard University by Daniel Branton, George Church, and Jene Golovchenko, at the University of California Santa Cruz by David Deamer and Mark Akeson, and at the National Institute of Standards and Technology by John Kasianowicz.

An example of such studies, with summary of previous work on this subject, is the recent work by Trepagnier and coworkers (2007). Artificial nanopores have been considered for rapid DNA sequencing. The approach is based on the modulation of the ionic current through the pore as a DNA molecule traverses it, revealing characteristics and parameters (such as diameter, length, conformation) of the molecule. One of the limitations to single base resolution in these nanopore-based DNA sequencing approaches is the insufficient control of the translocation speed of the molecule, e.g. during electrophoresis of the DNA molecules through the nanopore (speed around 30 kb per millisecond). This limitation was overcome by integration of an optical trapping system in the nanopore translocation experiments, lowering the translocation speed by several hundred-fold. For the demonstration, a known DNA fragment was used, which was attached to a polystyrene bead with a diameter of 10 micrometers, via streptavidin-biotin bond. The bead was placed into an optical trap, and the translocation speed was reduced about 200 times, thus giving more time for analysis of the DNA molecule passing through the nanopore. It was also possible to control the motion of the DNA molecule, and return it to its starting point before the translocation, thus repeated measurement and analysis were possible.

Another nanopore approach, studied in collaboration by several teams under a European Union grant on nano-DNA sequencing, coordinated by R. Zikic from Belgrade University and L. Forro from the Polytechnic School EPFL in Lausanne, is the development of a nano-electronic device for high-throughput single-molecule DNA sequencing, with the potential to determine long genomic sequences. The approach is based on the electrical characterization of individual nucleotides, while DNA passes through a nanopore, in principle similar to the one described by Trepagnier and coworkers (2007). However, in this new approach nanotube side electrodes (developed at the EPFL in Lausanne, Switzerland) are integrated in the nanopore. A lithographically fabricated nanogap is produced with a single-nanometer precision, and will allow characterization of the tunneling conductance across DNA bases, and electrical response of DNA molecule translocation between the two carbon nanotube electrodes. The control of the translocation rate of DNA through the nanopore is a critical step; besides the standard techniques (applied voltage, viscosity change, and DNA charge at various pH values), it will be varied by an optical tweezers system, with the aim of monitoring single-base resolution. Further improvements and modifications of the technique, such as increasing the number of parameters measured during the translocation of the DNA enabling single base resolution, could lead to a rapid nanopore-based DNA sequencing technique.

Sequenom (http://www.sequenom.com) announced that it has exclusively licensed technology from Harvard University that it will use to develop a nanopore-based sequencing platform, which will be faster and cheaper than currently

available technologies. The technology, which is able to detect a single DNA strand, as it passes through a nanopore, was developed by Amit Meller from Boston University's Department of Biomedical Engineering. The company expects the technology to enable DNA sequencing, as well as whole-genome genotyping and RNA and epigenetic analysis. In the near term the company expects the nanopore technology to deliver large-scale genotyping solutions, while in the long term it has the potential to provide a commercially viable, rapid, sub-thousand dollar human genome sequencing solution. The technology from Harvard was also licensed by the Oxford Nanopores Company in the UK, aiming at the development of a high-throughput DNA sequencing device.

BioNanomatrix and Complete Genomics (http://bionanomatrix.com) announced in 2007 the formation of a joint venture that will develop technology able to sequence a human genome in eight hours for less than $100. The proposed sequencing platform will use Complete Genomics' sequencing chemistry and BioNanomatrix' nanofluidic technology. They plan to adapt DNA sequencing chemistry with linearized nanoscale DNA imaging, to create a system that can read DNA sequences longer than 100 Kb quickly and with an accuracy exceeding the current industry standard. The next-generation sequencing technologies, as desribed above in section 24.2, have lowered the cost of sequencing the human genome (size 3 Gb) to around $100,000, and recent advances promise to reduce this cost significantly in the coming years, down to as little as $1,000 per individual. However, the two companies target with their design a price of $100 for this project, which would make it possible to sequence everyone's genome.

A very different and interesting approach to single-molecule DNA sequencing, using RNA polymerase (RNAP), has been presented recently by Greenleaf and Block (2006). In the planned method, the RNA polymerase enzyme is attached to one polystyrene bead, while the distal end of a DNA fragment is attached to another polystyrene bead. Each of the two beads is placed in an optical trap, and the pair of optical traps levitates the beads. The RNA polymerase interacts with the DNA fragment, and the transcriptional motion of RNA polymerase along the template changes the length of the DNA between the two beads. This leads to displacement of the two beads which can be registered with a precision in the angstrom range, and could give single-base resolution on a single DNA molecule. In order to distinguish between the four bases in the sequence, the single-molecule test is run four times (on four copies of the DNA fragment), each time with one of the nucleotides being present in much lower concentration than the other three nucleotides. This way, the RNA polymerase is induced to pause at every DNA position that requires the addition of the limiting nucleotide. By aligning on a subnanometer scale the four displacement records – using for calibration purposes the known sequences flanking the unknown fragment to be sequenced, in an analogous role to the primers used in Sanger sequencing – it is possible to deduce the sequence information. In the published work, the group has correctly identified 30 out of 32 bases in about 2 minutes. Since the processivity of the RNA polymerase is several Kb, the readings could be in principle that long. But the induced pauses of the enzyme, associated with misincorporation errors, are not always sufficiently regular to allow error-less sequencing. At present it is also not clear to what extent this ''motion-resolved'' sequencing can be parallelized and miniaturized, allowing sequencing of many DNA molecules in parallel and increasing the sequence throughput. But the technique demonstrates that the polymerization movement of a nucleic acid enzyme along the DNA fragment, and the very sensitive optical trap method, may allow extracting sequence information directly from a single DNA molecule.

24.3 PERSONAL GENOMICS AND THE 1,000 GENOMES PROJECT

The cost of genome sequencing, an overwhelmingly important factor in future studies, is getting low enough to make personal genomics a reality. Reduction of the cost by two orders of magnitude is needed to be able to realize the potential of personal genomics, which set the goal for cost of genome sequence to $1,000.

The first published human genome, sequenced by a post-Sanger DNA sequencing platform, was Jim Watson's genome sequence. The results of the study, which used Roche 454's sequencing technology, were assessed by scientists at Roche's Life Sciences and at Baylor College of Medicine.

The scientists sequenced both chromosome sets of Watson's genome, mapping unpaired sequence reads from 454's Genome Sequencer FLX to the human reference genome. The data were generated within two months for less than $1 million. Their report summarizes the identification of 3.3 million single nucleotide polymorphisms in the genome, and more than 600,000 of these were previously unknown. About 10,500 of the SNPs cause amino acid substitutions, potentially altering the function of proteins. They detected more than 200,000 small insertions and deletion polymorphisms, as well as a small number of copy-number variations, which result in local gains or losses in chromosomal regions (Wheeler *et al*., 2008). The sequencing data were shown to be of high quality, including regions of the human genome that were not contained in the human reference genome sequence (which was sequenced seven years ago with the Sanger technique); the data contained about 50 potential new genes.

Almost simultaneously, the completion of Craig Venter's genome sequence was announced (Levy *et al.*, 2008), carried out in the Craig Venter Institute, comparing sequences obtained with resequencing technology of his genome with the human reference genome. In mid-2008,

the genome sequence of two other individuals, namely a Chinese male and a Dutch female, has been announced.

The analysis of the first four available human genomes has demonstrated how difficult it still is to draw medically or biologically relevant conclusions from individual sequences. More genomes need to be sequenced, to learn how genotype correlates with phenotype. Researches are preparing a concrete plan for a project to sequence 1,000 human genomes, which will allow creation of a reference standard for analysis of human genomic variations which is expected to contribute to studies of human disease. The companies Illumina, Roche 454 Life Sciences, and Applied Biosystems decided to help with the project, and to generate the equivalent of 25 human genomes each per year, over a period of three years. This significant sequence contribution will enable the team to analyze the human genome with deeper sequencing and shorten its completion time. The 1,000 Genomes Project will identify variants present at a frequency of 1% over most of the genome, and as low as 0.5% within genes.

At this time, the immediate applications and relevance of the next-generation sequencing techniques in the medical field have been demonstrated by their ability to detect cancer alleles by deep sequencing of genomic DNA in cancerous tissues (isolated carefully by laser microdissection and capture techniques), which would have presented a very tedious task for the Sanger technique. Similar genome-wide analysis is planned for Stem cell-lines.

24.4 RNA SEQUENCING, ANALYSIS OF GENE EXPRESSION

High throughput of the next-generation sequencing technology, producing rapidly a huge number of short sequencing reads, made possible the analysis of a complex sample containing mixture of a large number of nucleic acids, by sequencing simultaneously the content of the sample. This is now possible without the tedious and time-consuming bacterial cloning, avoiding disadvantages associated with it. Next-generation sequencing may be applied to the characterization of mRNAs, methylated DNA, DNA, or RNA regions bound by certain proteins, and other DNA or RNA regions taking part in gene expression and regulation. Next-generation platforms are particularly suitable for analysis of gene expression in functional genomics, taking advantage of a very important methodology concept developed by Velculescu *et al*. (1995), and known as serial analysis of gene expression (SAGE). The original SAGE technique has demonstrated its novelty and powerful analysis, but it was limited in applications due to the need for difficult ligation of a huge number of short transcript DNA, subsequent cloning and tedious sequencing by the Sanger technique. Using the next-generation technology, the concept of the SAGE method now allows the analysis of RNA transcripts in a biological sample by obtaining short sequence tags 20–35 bp long directly from each transcript in the sample. Using computer informatics tools, these sequence tags allow the location of each of them to be determined and mapped in the reference genome. When mapped to a genome location, one can make genome-wide statistical analysis and distribution of the sequence tags, characterizing the gene expression in the starting biological or medical sample. With this technique, the transcripts are characterized through their sequence, in contrast to the probe hybridization employed in DNA microarray techniques, with the inherent difficulties in cross-hybridization and quantitation of hybridization. One obvious advantage of the sequence approach is the very reliable genome-wide quantitation of the transcript numbers and gene expression profile, yielding digital results. Owing to the huge number of samples sequenced simultaneously, these DNA sequencing techniques can detect even low-abundance RNAs, small RNAs, or detect the presence of mosaicism contained within a sample. Another advantage of the next-generation sequencing approach is that it does not require prior knowledge of the genome sequence. The sequence tag is mapped to the data collected in the projects in the database, and thus even organisms where the genome sequence is not yet known can be characterized by this RNA sequence method. An example of an application of the technique is the recent work on transcriptome profiling in stem cells by Cloonan *et al*. (2008).

24.5 RNA-SEQ STUDY INTO ALTERNATIVE SPLICING IN HUMAN CELLS

The next-generation sequencing technology has provided new insights into the complex alternative splicing patterns in the human genome. Sultan and coworkers (2008) used a technique known as the RNA-Seq technique, i.e. sequencing of complementary cDNA derived from RNA, to map and compare the transcriptomes in two different human cell types, and analyzed the type and frequency of alternative splicing events in these cells. It is well known that cells use alternative splicing to create alternative RNA transcripts from the same stretch of DNA. Until now, experimental tools to directly examine the alternative splicing were not available. These authors used the RNA-Seq technique to investigate both gene expression and alternative splicing in the same data set. They compared gene expression and splicing profiles in two cell-lines, human embryonic kidney cells and human B cells, because of the expected difference in expression and splicing profiles. With this approach they found more than 10,000 transcribed genes in each cell type, a number which is about 25% higher than that found by microarray platforms. About one-third of these genes were unknown. Nearly 100,000 splice junctions were identified (about 5% were new) in about 3,000 genes. The most common form of alternative splicing in both cell types was the

exon skipping. Instead of predictions, the RNA-Seq technique allows direct identification of the splicing junctions. Additional cell types will be analyzed by this method in the future, comparison of the gene expression and splicing data will show how the two are related in different cell types.

24.6 SEQUENCING OF CHROMATIN IMMUNOPRECIPITATED FRAGMENTS (CHIP-SEQ)

Chromatin immunoprecipitation or ChIP is a technique for identification and localization of DNA sites with which DNA binding proteins can interact and bind. The DNA binding protein to be analyzed is cross-linked to its DNA binding sites, and the complex is immunoprecipitated by an antibody recognizing specifically the protein. The DNA binding site is isolated together with the protein. In the microarray-based procedure, the mapping of the DNA site fragments is done by hybridizing the mixture of the immunoprecipitated DNA fragments to a tilling microarray that covers most of the genome under analysis. The latter became known as the ChIP-on-chip technique. With this technique, RNA polymerase and associated factors' binding sites were analyzed, including specific transcription factors, histones, and proteins involved in DNA repair or replication.

The next-generation sequencing technology allowed the microarrays in the mapping step to be replaced with high-throughput sequencing of the DNA binding sites, and their direct mapping to a reference genome. Detailed description of this ChIP-Seq method can be found in Robertson and coworkers (2007).

Any organism with a known genome sequence can be analyzed by a sequence-based method, without the inaccuracies inherent in microarray platforms caused by cross-hybridization or by difficulties in quantitation. The sequence of the binding site is mapped with a high resolution to regions shorter than 40 bases, not achievable by microarray mapping. Moreover, the ChIP-Seq approach is not biased and allows the identification of unknown protein binding sites, since in the case of ChIP-on-chip approach, the sequence of the DNA fragments tiled onto the microarray is predetermined, e.g. promoter arrays, exon arrays, etc.

24.7 GENOMICS AND MEDICINE, PROSPECTS FOR FUTURE DNA SEQUENCING

It is expected that the availability of ultra-deep sequencing of the human genomic DNA will transform the medical field in the near future, e.g. with the diagnostics sequencing allowing the analysis of disease causes, and providing novel approaches to the development of new drugs. It may become a promising tool in analysis of mental and developmental disorders like schizophrenia and autism (Morrow *et al*., 2008; Geschwind, 2008; Sutcliffe, 2008). It is anticipated that DNA sequencing of whole genomes for clinical purposes using these new technologies will likely occur in the next couple of decades. It will be increasingly common for patients who are suspected of or diagnosed with specific diseases to have a full DNA fingerprint taken. But the goal of sequencing healthy individuals to identify disease risks is very controversial and refused by the legal systems in many countries. At present it seems more acceptable and directly beneficent to sequence and analyse genomes of seniors who reached old age in a very good health. Resequencing of parts of the human genome in individuals to understand specific diseases associated, e.g., with SNPs is presently tested in many laboratories.

The novel sequencing technology will also be useful in microbial genomics, e.g. in the metagenomics, measuring the genetic diversity encoded by microbial life in organisms inhabiting a common environment (Hugenholtz and Tyson, 2008). Many microbial sequencing projects have already been completed or are in preparation, and several comparative genome analyses are under way to link the genotype and phenotype at the genomic level.

An important application is planned by the US Department of Energy's Joint Genome Institute (JGI), which will focus its sequencing efforts on new plant and microbial targets that may be of use in development of alternative energy sources. From the plants, the eucalyptus tree was selected for the first analysis, since its biomass production and carbon sequestration capacities make the tree of interest for alternative energy production and global carbon cycling. The JGI will also sequence the genome of the marine red alga, which seems to play an important environmental role in removing carbon dioxide from the atmosphere.

The proposed Human Microbiome Project, or as some call it the ''Second Human Genome Project'', emphasizes metagenomics – particularly in the human environment. Sequencing will be a big part of the project. The human microbiome, the collection of microbes in and on the human body, is one of the factors in understanding human health and disease. Changes in microbial communities in the body have been linked, for example, to immune system function, or to obesity and cancer. Some connections among these would not be surprising, taking into consideration the estimate that the human body contains ten times as many bacterial cells as human cells (which would mean roughly 1,000 bacterial genes for every human gene). Although scientists have been aware of the microbes associated with the human body for hundreds of years, the viral, bacterial, fungal, and single-celled eukaryote communities associated with humans remain largely unknown and poorly understood. The benefits of understanding the human microbiome may lead to the development of pro- and pre-biotic approaches for maintaining human health; it may make people aware of beneficial microbes and of the effect of antibiotics in the alteration of normal flora. In the

future, each person's microbiome could eventually become his or her medical biometric.

24.8 CONCLUSIONS

Next-generation high-throughput DNA sequencing techniques are creating fascinating opportunities in life sciences. Novel fields and applications in biology and medicine are becoming a reality, beyond the genomic sequencing as the original development goal and application. For example: personal genomics with detailed analysis of individual genomic stretches; precise analysis of RNA transcripts for gene expression (surpassing and replacing in several aspects analysis carried out by various microarray platforms, for instance in reliable and precise quantification of transcripts); and as a tool for identification and analysis of DNA regions that interact with regulatory proteins in functional regulation of gene expression. The next-generation sequencing technologies offer novel and rapid ways for genome-wide characterization and profiling of mRNAs, small RNAs, transcription factor regions, structure of chromatin, and DNA methylation patterns, microbiology and metagenomics. Presently, commercially available very high-throughput DNA sequencing platforms, as well as techniques under development, were described, and their applications in bio-medical fields discussed.

Genomics, proteomics and medical research benefit from recent advances and novel techniques for high-throughput analysis (e.g. DNA and protein microarrays, quantitative PCR, novel DNA sequencing techniques, mass spectrometry, and others).

The effort to find a cost-effective and rapid technique for whole genome sequencing has resulted in several novel approaches. The efficiency and throughput were improved by increasing the number of sequenced samples by several orders of magnitude. Another aim of efficient analysis is to sequence single molecules of DNA, an ultimate goal for a system miniaturization with minimum consumption of biochemicals. Moreover, it would allow the sequence information to be obtained without the need for amplification of DNA, associated with additional costs and potential source of inaccuracies. There are several approaches to the single-molecule sequencing, one of them commercially distributed by Helicos (section 24.2.4 above). At present, with the exception of the Roche 454 system, the result is usually a large number of short (several tens of bases long) sequences. Devices with these short readings have found many applications, e.g. in gene expression, or resequencing of genome regions, as described in the sections above.

But for genomic sequencing, and for analysis of the evermore important structural genetic variations in genomes, e.g. copy-number variation (CNVs; selected by the journal *Science* – in December 2007, volume 318, editorial and articles starting on page 1842 – as the most important scientific breakthrough of the year in 2007), as well as for chromosomal translocations, inversions, large deletions, insertions, and duplications, it would be a great advantage if sequence read length on the original single DNA molecule could be increased to several thousand bases and more per second. Ideally, the goal would be the sequence determination of a whole chromosome from a single original DNA molecule. Hopes for future development of such a novel very long read sequencing technique may give several physical techniques (e.g. AFM, electron microscopy, various spectroscopic techniques, soft X-rays, nanopores and nano-edges, to name just some of those most obvious), with many developments and improvements still needed.

ACKNOWLEDGEMENTS

I am grateful to Dr. George P. Patrinos for discussions and help with preparation of the figures.

REFERENCES

Ansorge, W., Sproat, B.S, Stegemann, J., and Schwager, C. (1986). A non-radioactive automated method for DNA sequence determination. *J. Biochem. Biophys. Methods* 13, 315–323.

Ansorge, W., Sproat, B.S, Stegemann, J., Schwager, C., and Zenke, M. (1987). Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res.* 15, 4593–4602.

Ansorge, W., (1991). *Process for Sequencing Nucleic Acids Without Gel Sieving Media on Solid Support and DNA Chips (Verfahren zur Sequenzierung von Nukleinsauren ohne Gele)*. EMBL German Patent Application DE 41 41 178 A1, and corresponding worldwide patent applications, Applied for 13.12.1991.

Braslavsky, I., Hebert, B., Kartalov, E., and Quake, S.R. (2003). Sequence information can be obtained from single DNA molecule. *Proc. Natl. Acad. Sci. USA* 100, 3960–3964.

Cloonan, N., Forrest, A., Kolle, G., Gardiner, B., Faulkner, G., Brown, M., Taylor, D., Steptoe, A., Wani, S., Bethel, G., Robertson, A., Perkins, A., Bruce, S., Lee, C., Ranade, S., Peckham, H., Manning, J., McKernan, K., and Grimmond, S.M. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* 5, 613–619.

Edwards, A., Voss, H., Rice, P., Civitello, A., Stegemann, J., Schwager, C., Zimmermann, J., Erfle, H., Caskey, C.T., and Ansorge, W. (1990). Automated DNA sequencing of the human HPRT locus. *Genomics* 6, 593–608.

Fan, H.C., Blumenfeld, Y.J., Chitkara, U., Hudgins, L., and Quake, S.R. (2008). Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc. Natl. Acad. Sci. USA,* 105, 16266–16271.

Geschwind, D.H. (2008). Autism – family connections. *Nature* 454, 838–839.

Greenleaf, W.J., and Block, S.M. (2006). Single-molecule, motion-based dna sequencing using RNA polymerase. *Science* 313, 801.

Hugenholtz, P., and Tyson, G.W. (2008). Metagenomics. *Nature* 455, 481–483.

Hyman, E.D. (1988). A new method of sequencing DNA. *Anal. Biochem.* 174, 423–436.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., Lin, Y., MacDonald, J.R., Pang, A.W., Shago, M., Stockwell, T.B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S.A., Busam, D.A., Beeson, K.Y., McIntosh, T.C., Remington, K.A., Abril, J.F., Gill, J., Borman, J., Rogers, Y.H., Frazier, M.E., Scherer, S.W., Strausberg, R.L., and Venter, J.C. (2007). The diploid genome sequence of an individual human. *PLoS Biol.* 5, e254.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., and Rothberg, J.M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.

Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA,* 74, 560–564.

Morrow, E.M., Yoo, S.Y., Flavell, S.W., Kim, T.K., Lin, Y., Hill, R.S., Mukaddes, N.M., Balkhy, S., Gascon, G., Hashmi, A., Al-Saad, S., Ware, J., Joseph, R.M., Greenblatt, R., Gleason, D., Ertelt, J.A., Apse, K.A., Bodell, A., Partlow, J.N., Barry, B., Yao, H., Markianos, K., Ferland, R.J., Greenberg, M.E., and Walsh, C.A. (2008). Identifying autism loci and genes by tracing recent shared ancestry. *Science* 321, 218–223.

Nyren, P., and Lundin, A. (1985). Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Anal. Biochem.* 151, 504–509.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., Thiessen, N., Griffith, O.L., He, A., Marra, M., Snyder, M., and Jones, S. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods,* 4, 651–657.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M., and Nyren, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 242, 84–89.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.

Schuster, S.C., Chi, K.R., Wold, B., and Myers, R.M. (2008). Method of the year, next-generation DNA sequencing, functional genomics and medical applications. *Nat. Methods* 5, 11–21.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309, 1728–1732.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B., and Hood, L.E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* 321, 674–679.

Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., Schmidt, D., O'Keeffe, S., Haas, S., Vingron, M., Lehrach, H., and Yaspo, M.L. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321, 956–960.

Sutcliffe, J.S. (2008). Insights into the pathogenesis of autism. *Science* 321, 208–209.

Trepagnier, E.H., Radenovic, A., Sivak, D., Geissler, P., and Liphardt, J. (2007). Controlling DNA capture and propagation through artificial nanopores. *Nano Lett.* 7, 2824–2830.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). *Science* 270, 484–487.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). The sequence of the human genome. *Science* 291, 1304–1351.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C.L., Irzyk, G.P., Lupski, J.R., Chinault, C., Song, X.Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D.M., Margulies, M., Weinstock, G.M., Gibbs, R.A., and Rothberg, J.M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872–876.

Chapter 25

Locus-Specific and National/Ethnic Mutation Databases: Emerging Tools for Molecular Diagnostics

George P. Patrinos[1,2]

[1]*Department of Pharmacy, School of Health Sciences, University of Patras, Patras, Greece;*

[2]*Erasmus University Medical Center, Faculty of Medicine and Health Sciences, Department of Bioinformatics, Rotterdam, The Netherlands*

25.1 INTRODUCTION

Research into the genetic basis of disease has advanced in scale and sophistication, leading to very high rates of data production in many laboratories, while DNA diagnostics and electronic health care records are increasingly common features of modern medical practice. Therefore, it should be possible to integrate all of this information in order to establish a detailed understanding of how genome sequence differences impact human health. Bioinformatics, usually defined as informatics tools used in molecular biology, is an important scientific discipline that emerged in the post-genomic era from the recent developments in the field of human genomics. Bioinformatics can be significantly helpful in, e.g., (1) accommodating huge data size (billions of base pairs), (2) stratifying large data diversity (repeated sequences, protein domains, interspecies DNA sequence comparison), and (3) integration of published references.

In the last decade, major advances have been made in the cloning and characterization of genes involved in human diseases. Concurrently, advances in technology have led to the identification of numerous mutations in these genes, ranging from point mutations to large rearrangements (see previous chapters). It rapidly has become clear that the knowledge and organization of these alterations in structured repositories will be of great importance not only for diagnosis but also for clinicians and researchers. Genetic or mutation databases are referred to as online repositories of mutation data, described for single (*locus-specific*) or more (*general*) genes or specifically for a population or ethnic group (*national/ethnic*). The main applications of mutation databases are: (1) to facilitate diagnosis at the DNA level and to define an optimal strategy for mutation detection, (2) to provide information about mutation-specific phenotypic patterns, and (3) to correlate locus-specific variant information with genome-wide features, such as repetitive elements, gene structures, interspecies conservation, mutation hotspots, recombination frequencies, and so on.

With great vision, Victor McKusick made, in 1966, the first serious efforts towards summarizing DNA variations and their clinical consequences when he published the Mendelian Inheritance in Man (MIM), a paper compendium of information on genetic disorders and genes (McKusick, 1966). This is now distributed electronically (Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/omim) by the National Center for Biotechnology Information (NCBI) and updated on a daily basis (Hamosh *et al*., 2005). Also, the first database collecting mutations from a single gene was published in 1976. It included 200 mutations from the globin gene in a book format and has led to the HbVar database for hemoglobin variants and thalassemia mutations (Hardison *et al*., 2002; Patrinos *et al*., 2004).

A decade later, David Cooper began listing mutations in genes to determine which mutational change was the most common (Cooper *et al*., 1998). In the mid-1990s the Human Genome Organization-Mutation Database Initiative (HUGO-MDI) was created, in order to organize this new domain of genetics and mutation analysis (Cotton *et al*., 1998), which then evolved into the Human Genome Variation Society (HGVS: http://www.hgvs.org). Today, the stated objective of the HGVS is "…to foster discovery and characterization of genomic variations, including population distribution and phenotypic associations".

Nowadays, this field is rapidly expanding and there are several mutation databases available on the internet. However, not all mutation databases fulfill quality requirements, while

others have been built by researchers "on the side" for their own use. Most importantly, funding opportunities for database-related projects have always been very limited. This chapter aims to summarize the current state-of-the-art in the field of mutation databases. In particular, the main activities relating to mutation databases will be highlighted, in order to: (i) describe the existing and emerging database types in this domain, (ii) emphasize their potential applications in modern medical genetics, and (iii) comment upon the key elements that are still missing and holding back the field.

25.2 MODELS FOR DATABASE MANAGEMENT

In its strict definition, a database is a collection of records, each of which contains one or more fields (i.e. pieces of data) about some entity (e.g. DNA sequences, mutations), which has a regular structure and which is organized in such a way that the desired information can easily be retrieved. The creation of databases relies on the model that the curator, i.e. the person or group of persons responsible for developing, updating and ultimately maintaining a mutation database, will choose for setting one up. There are mainly three types of database management models, from low to higher degrees of complexities, each with its own advantages and drawbacks.

Information contained within plain text websites is the simplest "database" format. This structure cannot be considered as a database in a strict sense, but it is often used for several database projects. The advantages for such a model are the development and maintenance simplicity, since no specific software is required. However, there are no true data querying options, apart from the standard searching tool provided by the respective internet browser, while the database becomes very difficult to maintain when the data sets get expanded.

Flat-file databases are the simplest type. They have for a long time been the dominant type, and can still be useful, particularly for small-scale and simple applications. These databases have modest querying capacity, and can accommodate small to moderately big data sets. On the contrary, when the data sets are significantly expanded, they become problematic to query and maintain. Also, their development requires average computing skills, even though they are based on simple software.

Relational databases are far more efficient for dealing with large volumes of information than flat-file databases. A relational database is based on data organization in a series of inter-related tables. Also, information can be retrieved in an extremely flexible manner by using structured data queries. Although interest in this model was initially confined to academia, subsequently, relational databases became the dominant type for high-performance applications because of their efficiency, ease of use, and ability to perform a variety of useful tasks that had not been originally envisioned. The dominant query language for relational databases is the semi-standardized structured query language (SQL). There are different SQL variants, the main ones being Microsoft SQL (http://www.microsoft.com/sql), MySQL (http://www.mysql.com), and postgreSQL (http://www.postgresql.org). Although critics claim that SQL is not consistent with the relational model, it works extremely well in practice and no replacement is on the horizon. The requirement of specialized software for developing a relational database can potentially be a disadvantage, since significant computer proficiency is required.

25.3 MUTATION DATABASE TYPES

The various depositories that fall under the banner of "mutation databases" can be categorized into three types: *general* (or *central*) *mutation* databases (GMDs), *locus-specific* databases (LSDBs) and *national/ethnic mutation* databases (NEMDBs).

GMDs attempt to capture all described mutations in all genes, but with each being represented in only limited detail. The included phenotype descriptions are generally quite cursory, making GMDs of little value for those wishing to understand the subtleties of phenotypic variability. GMDs tend to include only mutations of large effect that result in Mendelian patterns of inheritance, while sequence variations associated with no, minor, or uncertain clinical consequences are rarely cataloged. Thus, GMDs provide a good overview of patterns of clinically relevant mutations and polymorphisms, but almost no fine detail to aid proper understanding. The best current example of a GMD would be the Human Gene Mutation Database (http://www.hgmd.org; Stenson *et al*., 2003), which by June 2008 contained over 57,000 different records in 2,185 different genes in the public release and over 79,000 different records in 3,000 different genes in the HGMD Professional Release (V8.1). New entries are accumulating in HGMD at an average rate of approximately 5,000 per annum. Experts in specific fields do not maintain these databases, but they integrate data on mutations, their distribution, and references. Only published mutations are collected, mostly using automated data (text) mining routines. Each mutation is entered only once in order to avoid confusion between recurrent and identical-by-descent lesions and the phenotypic description associated to the mutation is very limited, preventing any study on phenotypic variability. These databases are frequently referred as "mile wide and inch deep", as they included mutations from many genes but with a limited description. A detailed survey on the GMDs currently available has been recently compiled by George *et al*. (2008).

In contrast to GMDs, LSDBs are concerned with just one or a few specific genes (Claustres *et al*., 2002), usually related to a single disease entity. They aim to be highly

curated repositories of published and unpublished mutations within those genes, and as such they provide a much-needed complement to the core databases. Data quality and completeness is typically high, with up to 50% of stored records pertaining to otherwise unpublished mutations. The data are also very rich and informative and the annotation of each mutant includes a full molecular and phenotypic description. Therefore, these databases are referred to as ''inch wide and mile deep''. For example, LSDBs will typically present each of the multiple discoveries of recurrent mutational events, so allowing mutation hotspots to be identified: and when these mutations occur upon different chromosomal backgrounds (linked to other mutations) such that they result in several, or different, disease features, these correlations are also recorded. A good example of an LSDB would be the HbVar database (http://globin.bx.psu.edu/hbvar; Hardison *et al*., 2002; Patrinos *et al*., 2004), a relational database of hemoglobin variants and thalassemia mutations, providing information on pathology, hematology, clinical presentation, and laboratory findings for a wealth of DNA alterations. Gene/protein variants are annotated with respect to biochemical data, analytical techniques, structure, stability, function, literature references, and qualitative and quantitative distribution in ethnic groups and geographic locations (Patrinos *et al*., 2004). As is common in LSDBs, entries can be accessed through summary listings or user-generated queries, which can be highly specific. A listing of the over 650 currently available LSDBs can be found at http://www.hgvs.org, http://www.hgmd.org and in the literature (Cotton *et al*., 2007).

Finally, NEMDBs are repositories documenting the genetic composition of an ethnic group and/or population, the genetic defects, leading to various inherited disorders and their frequencies calculated on a population-specific basis. The emergence of the NEMDBs is justified from the fact that the spectrum of mutations observed for any gene or disease will often differ between population groups across the planet, and also between distinct ethnic groups within a geographical region. Not only do NEMDBs help elaborate the demographic history of human population groups, they are also a prerequisite to the optimization of national DNA diagnostic services. That is, they will provide essential reference information for use in the design of targeted mutation detection efforts for clinical use, and they may also serve to enhance awareness among health care professionals, bio-scientists, patients, and the general public about the range of most-common genetic disorders (and their environmental correlates) suffered by particular population groups.

Beyond the aforementioned main database types, DNA variation is also recorded in various polymorphism databases, such as dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP; Wheeler *et al*., 2008), the HAPMAP Data Coordination Center (DCC: http://www.hapmap.org; International HapMap Consortium, 2003), and the Human Genome Variation Database (HGVbase: http://www.hgvbase.org; Fredman *et al*., 2004). These resources are not yet perfect in design or content (Marsh *et al*., 2002; Aerts *et al*., 2002; Dvornyk *et al*., 2004), but they do make available a very extensive list of normally occurring human genome variation. They are mentioned briefly here because they are important in helping to complete the picture for any gene or region of interest, by summarizing all the neutral variants that are typically not included in GMDs, LSDBs, and NEMDBs. In brief, GMDs, LSDBs, and NEMDBs share the same primary purpose of representing DNA variations that have definitive or likely phenotypic effect. They tackle this goal from very different perspectives, and there is clearly a need for these three types of resource in the various disciplines of human genetics and genomics, particularly genetic testing.

In the following paragraphs, the basic aspects of LSDBs and NEMDBs will be analyzed in detail, in relation to their applications in molecular diagnostics.

25.4 LSDBs IN MOLECULAR GENETIC TESTING

In the previous chapters, it has been nicely demonstrated that detection of DNA sequence variation can be very efficiently performed using a plethora of techniques. LSDBs can facilitate molecular diagnosis of inherited diseases in various ways. For example, LSBDs can assist in ascertaining whether a DNA variation is indeed a causative mutation, leading to a genetic disease, or a neutral polymorphism. Similarly, some high-quality LSDBs provide detailed phenotypic information that is related to disease-causing mutations. Ultimately, LSDBs can assist in the selection of the optimal mutation detection strategy.

25.4.1 Identification of Causative Mutations

Usually, in diagnostic laboratories, if a missense mutation is detected, additional experiments need to be conducted prior to concluding that the mutation in question is in fact the causative mutation in the family. In the absence of a functional test, the segregation of the mutation in the affected family members, the absence of this variation in a panel of at least 100 control samples, the prediction of the biochemical nature of the substitution, the region where the mutation is located, and the degree of conservation among species are some of the arguments for a causative mutation. This approach is often both time consuming and cost effective. Therefore, the use of an LSDB can provide researchers with valuable information to help in such a decision process.

If the mutation has been reported as a causative one, the full description of the mutation and the corresponding literature is provided in the LSDB. Furthermore, a comprehensive LSDB includes not only the reference sequence of the gene but also the description of structural domains and data about interspecies conservation for each protein residue.

Also, as many publications include a large set of data, it is often possible to observe typing errors or use of a wrong nomenclature, due to reference to an old sequence (up to 10% of errors for some publications). The use of LSDBs can be helpful to this end. Several LSDBs include an automatic nomenclature system, based on a reference sequence. In other words, variant entry is done such that the variant nucleotide is automatically checked against the reference sequence at the respective variant position and named based on the official (HGVS) nomenclature (den Dunnen and Antonarakis, 2001). Interestingly, a dedicated module, namely Mutation Analyzer (Mutalyzer) sequence variation nomenclature checker (http://www.lovd.nl/mutalyzer; Wildeman *et al*., 2008) has been constructed to enable unambiguous and correct sequence variant descriptions and to avoid mistakes and uncertainties that may lead to undesired errors in clinical diagnosis. Mutalyzer handles most variation types, i.e. substitutions, deletions, duplications, etc., and follows current HGVS recommendations. Input is a GenBank accession number or an uploaded reference sequence file in GenBank format with user-modified annotation, a HUGO gene nomenclature committee (HGNC) gene symbol, and the variant (single or in a batch file). Mutalyzer generates variant descriptions at DNA level, the level of all annotated transcripts and the deduced outcome at protein level.

Finally, a handful of LSDBs include data presentation tools to visualize their content in a graphical display. Recently, Smith and Cotton (2008) developed a generic visualization toolkit that can be employed by LSBDs to generate graphical models of gene sequence with corresponding variations and their consequences. The VariVis software package can run on any web server capable of executing Perl CGI scripts and can interface with numerous database management systems and even flat-file databases. VariVis produces two easily understandable graphical depictions of any gene sequence and matches these with variant data. The toolkit has already been tested in A_1ATVar, an LSDB for *SERPINA1* gene variants, leading to α_1-antitrypsin deficiency (Fig. 25.1; Zaimidou *et al*., 2009) and can be integrated into generic database management systems used for LSDB's development (see also section 25.6).

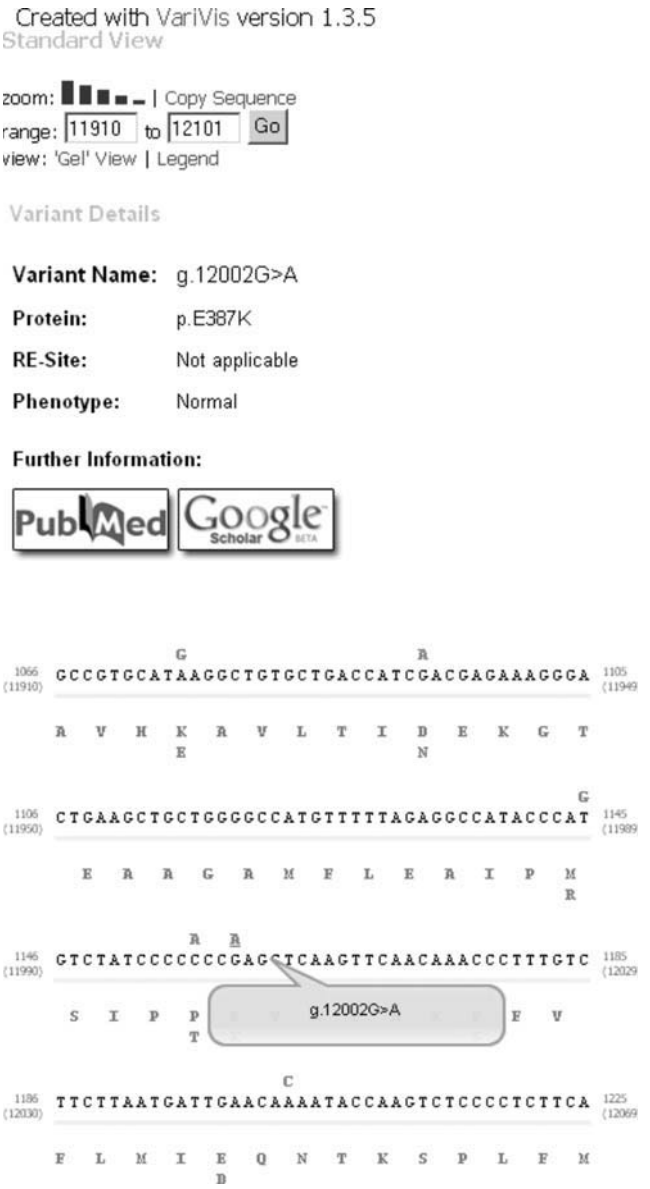


FIGURE 25.1 The VariVis visualization toolkit and its implementation in the A_1ATVar LSDB, depicting mutations in exon 5 of the *SERPINA1* gene, leading to α1-antitrypsin deficiency. Further information for this variant can be also obtained from PubMed and Google Scholar literature databases. Adapted from Zaimidou *et al*. (2009).

25.4.2 Linking Genotype Information with Phenotypic Patterns

LSDBs are far more than just inert repositories, as they include analyzing tools, which exploit computing power to answer complex queries, such as phenotypic heterogeneity and genotype/phenotype correlations. The vast majority of LSDBs provide phenotypic descriptions in abstract format. A typical example is provided in Fig. 25.2, where genotype information of each *SERPINA1* gene variant, leading to α_1-antitrypsin deficiency, is displayed next to certain phenotypic descriptions, e.g. variant category, organ affected, etc. Similar displays are also available for the LOVD-based LSDBs (see section 25.6). Also, in other databases, such as HbVar, phenotypic descriptions are slightly more detailed, e.g. providing information on the clinical presentation of thalassemia carriers and patients together with their hematological indices for every hemoglobin variant and/or thalassemia mutation. All LSDBs attempt to enforce controlled vocabulary to facilitate straightforward data querying. Therefore, phenotype data in the vast majority of LSDBs are presented in a very basic way, such as in the form of free text entries and/or with very little detail. There is a definite need for this situation to be improved, and a general wish for the comprehensive analysis of phenotypes to occur, a goal termed "phenomics" (Gerlai, 2002; Hall, 2003; Scriver, 2004), supported by the necessary informatics solutions.

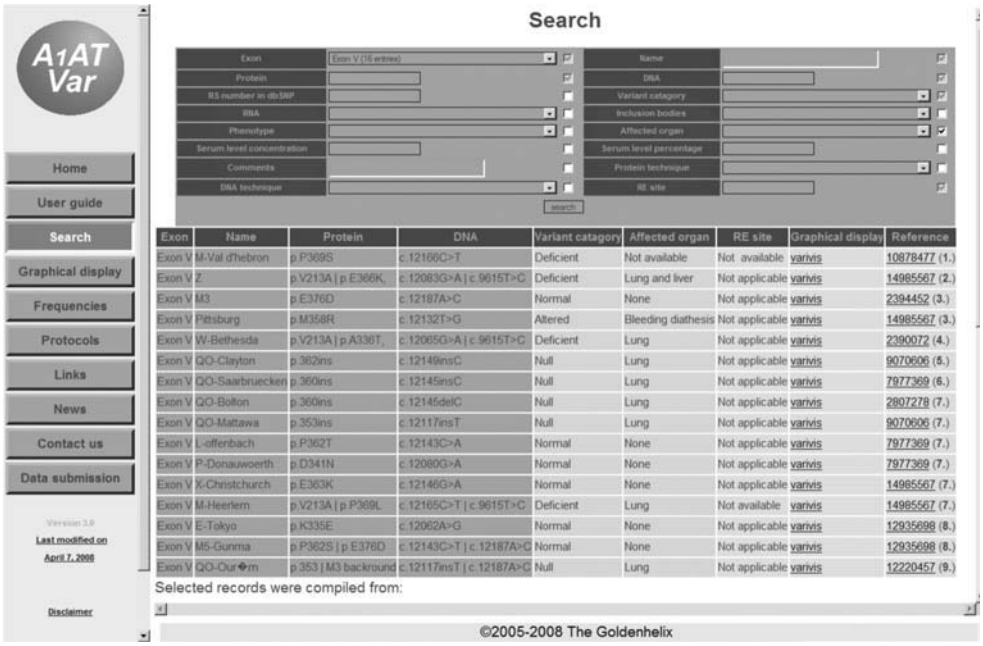


FIGURE 25.2 Data querying in A_1ATVar database for *SERPINA1* gene variants. This example illustrates a query for all *SERPINA1* gene variants located in exon 5. Query output includes 16 records (also indicated in the drop-down menu) for which the desired information is provided (for RNA levels and phenotype) in addition to the standard columns display (selected by default and shown in gray). Adapted from Zaimidou *et al*. (2008). *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

Also, it is possible for patients with the same mutation to have a completely different phenotype. This valuable information could be useful for predictive medicine. A good example is given by the distribution of mutations of the *FBN1* gene (MIM# 134797) that are associated with Marfan syndrome (MFS) and a spectrum of conditions phenotypically related to MFS, including dominantly inherited ectopia lentis, severe neonatal MFS, and isolated typical features of MFS. MFS, the founding member of heritable disorders of connective tissue, is a dominantly inherited condition characterized by tall stature and skeletal deformities, dislocation of the ocular lens, and propensity to aortic dissection (Collod-Beroud and Boileau, 2002). The syndrome is characterized by considerable variation in the clinical phenotype between families and also within the same family. Severe neonatal MFS has features of the MFS and of congenital contractural arachnodactyly present at birth, along with unique features such as loose, redundant skin, cardiac malformations, and pulmonary emphysema (Collod-Beroud and Boileau, 2002). A specific pattern of mutations is observed in exons 24 to 26 in association with the neonatal MFS. In fact, 73.1% of mutations are located in this region in the neonatal form of the disease, but only 4.8% of mutations associated with a classical MFS are located in these exons (FBN1 database: http://www.umd.be; Collod *et al*., 1996; Collod-Beroud *et al*., 2003).

Recently, an interesting project attempted to connect human phenotype and clinical data in various LSDBs with data on genome sequences, evolutionary history, and function from the ENCODE project and other resources in genome browsers. PhenCode (Phenotypes for ENCODE; http://www.bx.psu.edu/phencode; Giardine *et al*., 2007a) is a collaborative, exploratory project to help understand phenotypes of human mutations in the context of sequence and functional data from genome projects. The project initially focused on a few selected LSDBs, namely HbVar (*HBA2*, *HBA1*, and *HBB* genes) and PAHdb (*PAH* gene; Scriver *et al*., 2000), etc. Interesting mutations found in a genome browser can be tracked by following links back to the LSDBs for more detailed information. Alternatively, users can start with queries on mutations or phenotypes at an LSDB and then display the results at the genome browser to view complementary information such as functional data (e.g. chromatin modifications and protein binding from the ENCODE consortium), evolutionary constraint, regulatory potential, and/or any other tracks they choose. PhenCode provides a seamless, bidirectional connection between LSDBs and ENCODE data at genome browsers, which allows users to easily explore phenotypes associated with functional elements and look for genomic data that could explain clinical phenotypes, thus helping to fulfill the promise of the Human Genome Project to improve human health. Therefore, PhenCode is not only helpful to clinicians for diagnostics, it also serves biomedical researchers by integrating multiple types of information and facilitating the generation of testable hypotheses to improve our understanding of both the functions of genomic DNA and the mechanisms by which it achieves those functions. These and other types of data provide new opportunities to better explain phenotypes.

25.4.3 Selection of the Optimal Mutation Detection Strategy

As LSDBs collect all published and unpublished mutations, they are very useful to define an optimal mutation screening

strategy. Therefore, an overview of the distribution of mutations at the exonic level can help to focus on specific exons, where most of the mutations are located. The best example is given by the study of the TP53 gene involved in up to 50% of human cancers (Soussi, 2000). This gene is composed of 11 exons from which ten are transcribed in a 393 amino acids protein. The distribution of the 23,868 mutations reported from 2,201 references in the TP53 database (http://www.umd.be; last assessed in June 2008) shows that approximately 95% of mutations are located in four out of the 11 exons of the gene (exons 5–8; Beroud and Soussi, 2003; Soussi and Beroud, 2003). This observation has led 39% of research groups to search for mutations only in these exons, whereas 13% performed a complete scanning of the gene (Soussi and Beroud, 2001). Although this strategy is cost effective, one needs to be careful in case of a negative result and should perform a complete scanning in order to avoid bias, as described in Soussi and Beroud (2001).

Similarly, summary listings of mutations documented in LSDBs will help to choose the best technical approach for mutation screening. For example, if most mutations are nonsense mutations, the protein truncation test could be one of the best approaches, whereas sequencing is considered the gold standard for mutation detection. Certain LSBDs may contain additional information about primers and technical conditions to help new research groups or diagnostic laboratories to establish their own diagnostic procedures. A very nice website is http://www.dmd.nl, about various genes involved in muscular dystrophies, from where a list of primers can be used for multiplex PCR to detect deletions of the DMD gene involved in Duchenne and Becker myopathies. Furthermore, few other LSDBs have summary listings of complete mutation detection methods for mutation screening in the underlying gene(s). XPRbase (http://www.goldenhelix.org/xprbase), a companion database to HbVar, consisting of mutation detection protocols to screen for globin gene mutations (Giardine *et al*., 2007b; Fig. 25.3) and UMD database (http://www.umd.be; Beroud *et al*., 2000) are typical examples.

Finally, in a few LSDBs, explicit information is provided regarding the mutation pattern associated with a population/ethnic group and/or geographical region. This information can be extremely helpful in stratifying mutation detection strategies. In other words, in ethnic groups with a more or less homogeneous mutation pattern, mutation screening efforts can be either targeted to those genomic regions that the majority of mutations have been reported, or to a specific mutation detection technique, i.e. ARMS or restriction enzyme analysis, thus saving time and resources. Coupling LSDBs with NEMDBs resources (see also next paragraph), if available, would further facilitate these efforts (Fig. 25.4). However, as previously explained, extreme caution should be taken in the case of a negative result that would require complete mutation scanning in order to avoid bias.

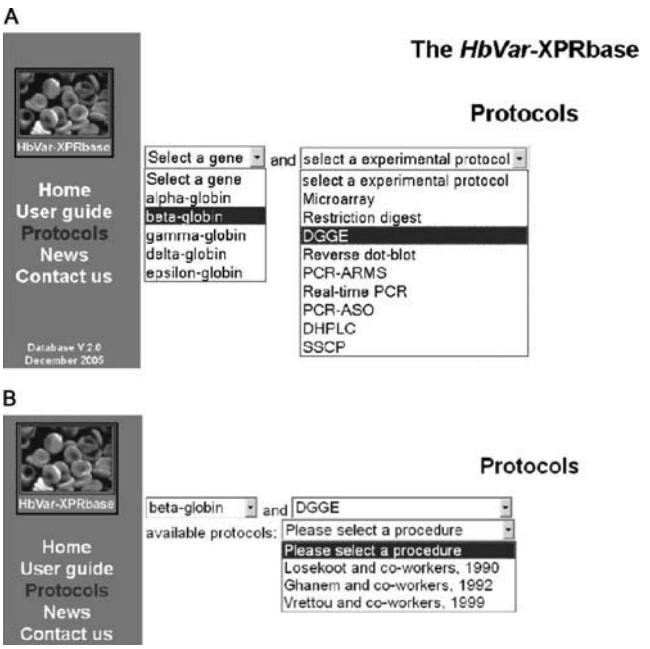


FIGURE 25.3 The protocols page from XPRbase, a companion database forHbVar. **A.** Users can select from two drop-down menus, which give the protocols that are available for mutation screening in every globin gene, for example "Find all DGGE protocols available for β-globin gene mutation screening". **B.** This query returns three protocols, each one of which consists of the primer sequence, amplification conditions, reaction ingredients, and the source reference(s), hyperlinked to PubMed, from where the publication(s) can be retrieved (Giardine *et al*., 2007b).

25.4.4 Criteria for LSDB Selection

Over 700 LSDBs are available today on the internet, which sometimes makes it difficult to choose the "best" LSDB. For example, there are at least six different LSDBs for *TP53* gene mutations, from which two are recognized as reference databases (http://p53.free.fr and http://www-p53.iarc.fr/index.html). The same is true for a few other LSDBs. This fact generates confusion to potential users as to which LSDB to choose. This can be particularly worrying, since not all databases conform to the proposed quality guidelines, or are curated or updated frequently, while they are rather diverse in terms of content and structure. The most efficient way for an interested user to directly select the reference database for its purposes is to consult the HGVS portal, where summary listings of all available LSDBs are provided and get frequently updated (http://www.hgvs.org/dblist/glsdb.html).

25.5 NEMDBs: A NEW TREND

NEMDBs are continuously updated mutation depositories, recording extensive information over the described genetic heterogeneity of an ethnic group or population. These resources have recently emerged, mostly driven by the need to document the varying mutation spectrum observed for any gene (or multiple genes) associated with a genetic disorder, among different population and ethnic groups (Patrinos, 2006).

A

Population	Disease	Gene	Mutation	HGVS nomenclature	No Chromosomes	Frequency (%)
Greek Cypriot	Familiar Mediterranean Fever	MEFV	E148Q	p.E148Q	9	5.80
Greek Cypriot	Familiar Mediterranean Fever	MEFV	E167D-F479L	p.E167D and p.F479L	34	21.80
Greek Cypriot	Familiar Mediterranean Fever	MEFV	M680I	p.M680I	2	1.30
Greek Cypriot	Familiar Mediterranean Fever	MEFV	M694I	p.M694I	3	1.90
Greek Cypriot	Familiar Mediterranean Fever	MEFV	M694V	p.M694V	33	21.10
Greek Cypriot	Familiar Mediterranean Fever	MEFV	V726A	p.V726A	43	27.60
Greek Cypriot	Familiar Mediterranean Fever	MEFV	R761H	p.R761H	5	3.20
Greek Cypriot	Familiar Mediterranean Fever	MEFV	Unknown	N.A.	27	17.30

B

Population	Disease	Gene	Mutation	HGVS nomenclature	No Chromosomes	Frequency (%)
Greek Cypriot	Beta-thalassemia	HBB	Cd 6 (-A)	c.20delA	1	0.14
Greek Cypriot	Beta-thalassemia	HBB	Cd 8 (-AA)	c.25 to 26delAA	3	0.42
Greek Cypriot	Beta-thalassemia	HBB	IVS I-1 (G>A)	c.93G>A	36	5.41
Greek Cypriot	Beta-thalassemia	HBB	IVS I-6 (T>C)	c.98T>C	39	5.83
Greek Cypriot	Beta-thalassemia	HBB	IVS I-110 (G>A)	c.202G>A	563	79.96
Greek Cypriot	Beta-thalassemia	HBB	Cd 39 (C>T)	c.248C>T	20	2.83
Greek Cypriot	Beta-thalassemia	HBB	IVS II-745 (C>G)	c.1190C>G	36	5.41
Turkish Cypriot	Beta-thalassemia	HBB	IVS I-1 (G>A)	c.93G>A	17	7.3
Turkish Cypriot	Beta-thalassemia	HBB	IVS I-6 (T>C)	c.98T>C	18	7.8
Turkish Cypriot	Beta-thalassemia	HBB	IVS I-110 (G>A)	c.202G>A	172	74.1
Turkish Cypriot	Beta-thalassemia	HBB	Cd 39 (C>T)	c.248C>T	2	0.9
Turkish Cypriot	Beta-thalassemia	HBB	IVS II-745 (C>G)	c.1190C>G	15	6.5
Turkish Cypriot	Beta-thalassemia	HBB	Unknown	N.A.	6	3.4

FIGURE 25.4 NEMDBs, coupled to LSDBs, can significantly facilitate molecular genetic testing efforts. **A.** Data querying in the Cypriot NEMDB for the mutational spectrum leading to Familial Mediterranean Fever in the Cypriot population indicates that five *MEFV* gene mutations account for approximately 95% of the mutant alleles. **B.** On the contrary, β-thalassemia has an extremely homogeneous mutational pattern in both the Greek Cypriot and Turkish Cypriot ethnic groups, with only one mutation accounting for approximately 75–80% of the mutant chromosomes. The latter can provide useful insights in the selection of the most appropriate mutation screening strategy. Also, this figure demonstrates that the Greek Cypriot and Turkish Cypriot groups, although genetically related to each other, differ substantially in their mutation spectra compared to the Greek and Turkish populations, respectively (not shown), despite being in close physical proximity and their cultural relationships.

In general, the NEMDBs available to date can be divided in two subcategories. The ''National Genetic'' (or Disease Mutation) databases, the first ones that appeared online, record the extant genetic composition of a population or ethnic group but with limited or no description of mutation frequencies. The first NEMDB to come online was the Finnish database (http://www.findis.org; Sipila and Aula, 2002), which although rich in information provided very limited querying capacity, particularly for allelic frequencies. On the other hand, the ''National Mutation Frequency'' databases provide comprehensive information only of those inherited disorders whose disease-causing mutation spectrum is well defined. The Hellenic and Cypriot NEMDBs (available at http://www.goldenhelix.org/hellenic and http://www.goldenhelix.org/cypriot respectively; Patrinos *et al*., 2005a; Kleanthous *et al*., 2006) are trying to do somewhat better from the outset, by introducing a specialized database management system (see also next paragraph) that enables both basic query formulation and restricted-access data entry so that all records are manually curated to ensure high and consistent data quality (Patrinos *et al*., 2005a).

In order to provide a simple and expandable system for worldwide population-specific mutation frequency data documentation, the latter NEMDB group was used as the basis for the design of the Frequency of INherited Disorders database (FINDbase; http://www.findbase.org), a relational database that currently records frequencies of causative mutations and pharmacogenetic markers worldwide (van Baal *et al*., 2007). FINDbase offers a user-friendly query interface, providing instant access to the list and frequencies of the different mutations and query outputs can be either in a table or graphical format (Fig. 25.5). Data entry and database curation are being performed online by three different user groups, namely administrator, national coordinator and curator, via a password-protected interface.

Given their scope, to maximize the utility of NEMDBs the mode by which their content is provided needs to provide a seamless integration with related content in LSDBs and GMDs (Patrinos and Brookes, 2005). Furthermore, as is always desirable for specialized databases, extensive links to other external information (e.g. to OMIM and to various types of genome sequence annotation) would ideally be provided as part of the necessary tying together of the growing network of genomic databases.

25.5.1 NEMDBs and Genetic Testing

NEMDBs can be helpful in multiple ways in a molecular diagnostic setting. First of all, they can help optimize national molecular diagnostic services by providing essential reference

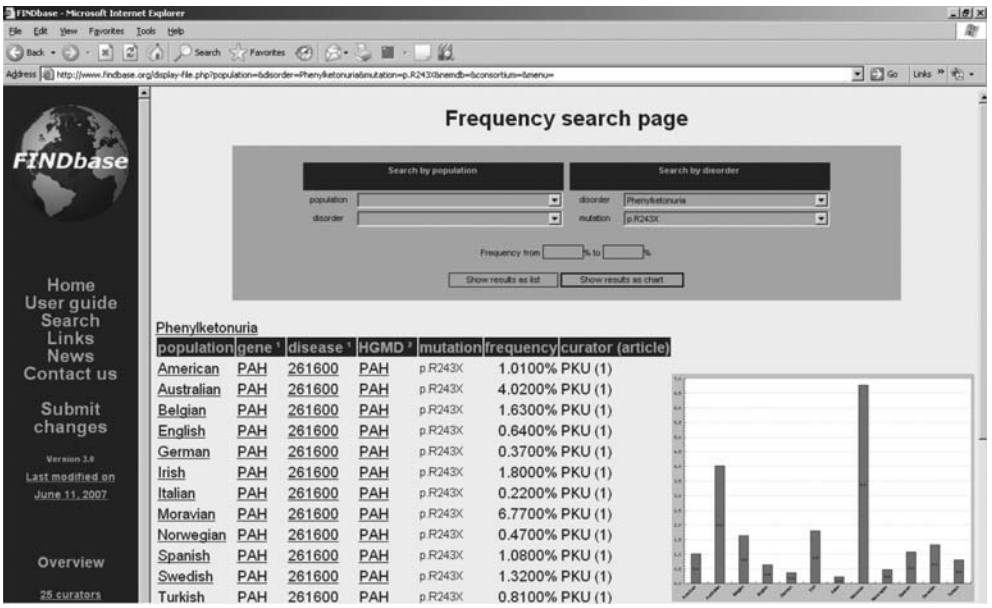


FIGURE 25.5 Use of FINDbase database to compare the different allelic frequencies in various populations. This example illustrates the query "Frequencies of the mutant p.R243X allele, leading to phenylketonuria in all populations." Query outcome includes 12 populations in which the p.R243X is found, along with the corresponding allelic frequencies, in both tabular and graphical format for comparison (see also van Baal *et al*., 2007).

information for the design and implementation of regional or nationwide mutation screening efforts (see also section 25.4.3). In addition, NEMDBs can enhance awareness among clinicians, bioscientists, and the general public about the range of most common genetic disorders suffered by certain populations and/or ethnic groups. Based on populations' mutation spectra, customized genetic tests can be designed. Several diagnostic companies have designed kits for β-thalassemia, which qualitatively detect, e.g. the most common Mediterranean and Asian β-thalassemia mutations (reviewed in Patrinos *et al*., 2005b). Most importantly, those databases can also assist in interpreting diagnostic test results in countries with heterogeneous populations, particularly where interpretation of test results in minority ethnic groups may be ambiguous or problematic (Zlotogora *et al*., 2007).

25.5.2 NEMDBs and Society

Apart from their importance in a clinical laboratory setting, NEMDBs can also contribute toward the elucidation of populations' origin and migration. History of a certain population is tightly linked with the history of its allele(s). Therefore, NEMDBs, particularly those including data from many population groups, can serve as the platform for comparative genomic studies that can reciprocally provide insights into, e.g., the demographic history of human populations, patterns of their migration and admixture, gene/mutation flow, etc. NEMDBs can therefore be particularly helpful to determine the recent microevolution of human populations via precise ethnic identification that in turn is extremely important for the rapid transition of modern societies, through migration to multi-ethnicity.

To this end, in order for an NEMDB to be maximally comprehensive and accurate, it is vital that certain parameters are strictly observed. In the case of recurrent mutation events (e.g. sickle cell mutation), caution should be taken to record precisely the underlying genomic background on which a mutation has occurred. Furthermore, mutation frequencies should be calculated based on the most representative study that involves sufficient numbers of patients and controls. Estimation of absolute mutation frequencies based on multiple reports has the inherent danger of including redundant cases that can alter the calculated allele frequencies. Finally, the very delicate issue of anonymity should be adequately preserved, by including data only at a summary rather than individual level, so that NEMDB data contents consist only of number of chromosomes rather than sensitive personal details of their carriers.

25.6 DATABASE MANAGEMENT SYSTEMS FOR LSDBs AND NEMDBs

In is noteworthy that although over 79,000 different mutations are recorded in HGMD for 3,000 different genes, only just over 600 LSDBs are available today, including some redundant ones for specific genes. From this, it is clear that a great number of LSDBs still do not exist for many genes. The HGVS produced guidelines and recommendations for mutation nomenclature (den Dunnen and Antonarakis, 2001; den Dunnen and Paalman, 2003), content, structure (Scriver *et al*., 1999), curation and deployment of LSDBs (Claustres *et al*., 2002; Cotton *et al*., 2008), and NEMDBs (Patrinos, 2006).

To facilitate interested parties and research groups to develop and curate their own LSDBs and NEMDBs, generic

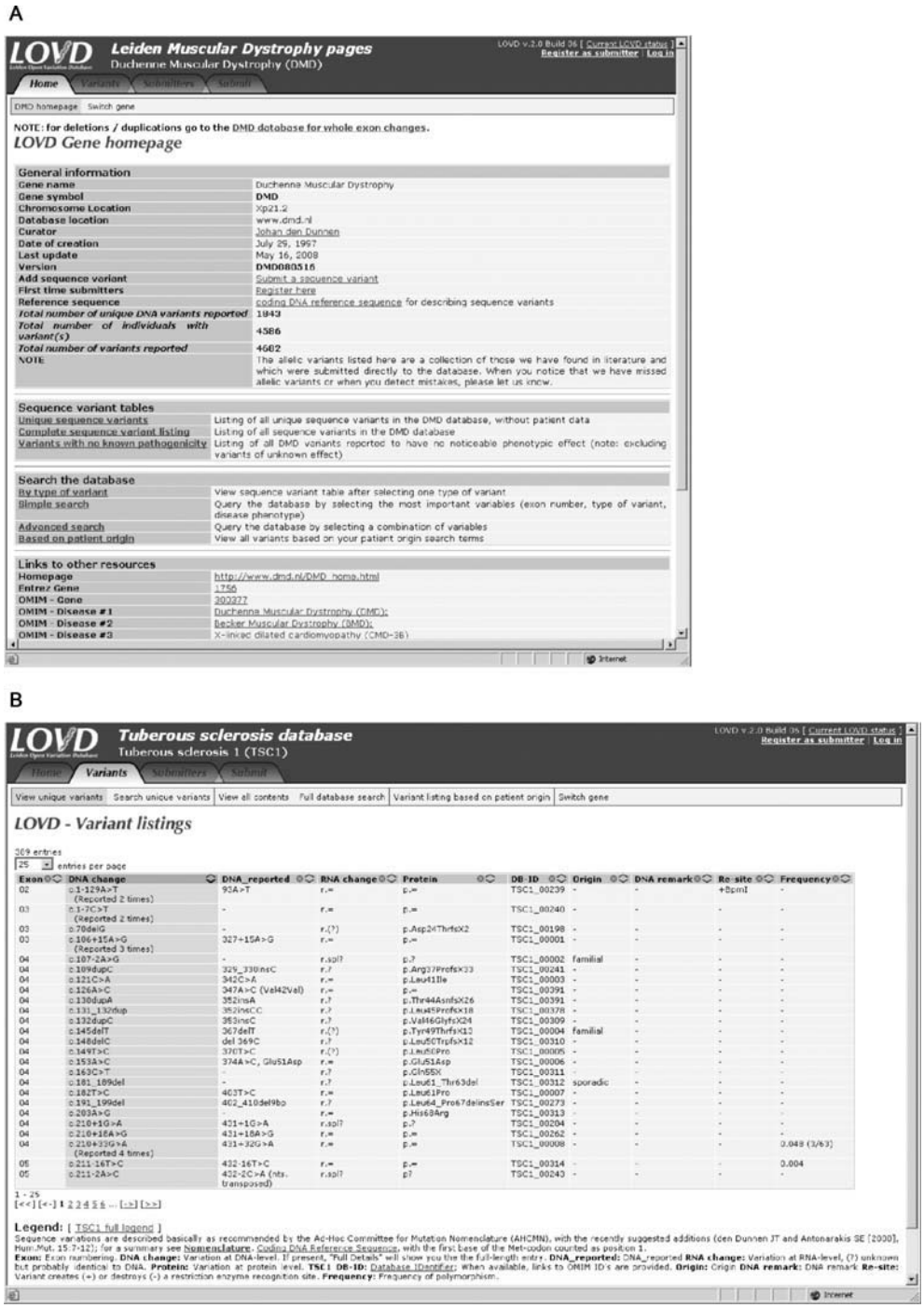


FIGURE 25.6 Examples of use of the LOVD DBMS. **A.** Home (summary) page of the LOVD-based Duchenne Muscular Dystrophy (DMD)-specific database. **B.** Variant page of the LOVD-based TSC1-specific database, including detailed information of all *TSC1* gene variants, leading to tuberous sclerosis in a tabular format.

tools, known as database management systems (DBMS), have been made available for this purpose. Today, several off-the-shelf freely available and user-friendly softwares are available for LSDB development and curation. These include MUTbase (Riikonen and Vihinen, 1999), Universal Mutation database (UMD; Beroud *et al*., 2000), Mutation Storage and Retrieval (MutStaR; Brown and McKie, 2000), and Leiden-Open (source) Variation Database (LOVD, Fokkema *et al*., 2005; Fig. 25.6). As far as NEMDBs are concerned, ETHnic and National database Operating System (ETHNOS; Patrinos *et al*., 2005a) is currently available in both flat-file and relational database format, on which the first generation NEMDBs and FINDbase were developed. Table 25.1 summarizes the existing DBMS available today.

TABLE 25.1 Database management systems (DBMS) available today for the development and curation of LSDBs and NEMDBs. Their URL addresses are provided, together with the corresponding number of installed databases (installations). Information is collected from the corresponding websites (last visited: June 2008). N.A.: Not available.

DBMS	URL address	Installations	References
LSDBs			
MUTbase	http://bioinf.uta.fi/MUTbase	N.A.	Riikonen and Vihinen (1999)
UMD	http://www.umd.be	18	Beroud *et al*. (2000)
MuStaR	http://www.hgu.mrc.ac.uk/Softdata/Mustar	4	Brown and McKie (2000)
LOVD	http://www.lovd.nl	20	Fokkema *et al*. (2005)
NEMDBs			
ETHNOS	http://www.goldenhelix.org	12	Patrinos *et al*. (2005)

These user-friendly DBMSs are designed to promote the creation of more and better LSDBs, by reducing or eliminating the requirement for substantial knowledge of computing and bioinformatics for interested parties to establish an LSDB from scratch. In addition, the use of off-the-shelf solutions will positively impact on data uniformity. Contrary to the NEMDB structure and data content that is relatively uniform, since the majority of them are ETHNOS based, the LSDB domain suffers from extreme structure and data content heterogeneity that impact their overall quality. It is expected that adoption of a generic DBMS and migration of most of the existing LSDBs to this DBMS will drastically solve the issue and increase data uniformity among these resources.

Finally, the use of a database management system that can be run on any platform will reduce the risk of the database being ''lost''. If database curation for some reason, e.g. lack of funding, is interrupted, data will then be transferred directly between platforms or locations and they will remain accessible to all. Potential curators will be encouraged to set up LSDBs, with the choice of using these software packages locally on their own workstations or, most importantly, having their databases hosted on a central server on the internet.

25.7 FUTURE CHALLENGES

Notwithstanding the technical challenges, perhaps more difficult to overcome will be problems associated with the way database research is organized, motivated, and rewarded. For example, forming consensus opinions and truly committed consortia in order to create standards is far from easy in the highly competitive world of science. This may partly explain why leading bioinformatics activities today are often conducted in large specialized centers (e.g. the European Bioinformatics Institute, and the US National Center for Biotechnology Information) where the political influence and critical mass is such that what they produce automatically becomes the *de facto* standard. These groups, however, cannot build all the necessary GMDs, LSDBs, and NEMDBs that are needed, but they could help others (biological domain experts) to do it and then integrate all their efforts (Stein, 2002).

In 2006, a global initiative, the Human Variome Project (HVP), was initiated, aiming to catalog all human genetic variation and to make that information freely available to researchers, clinicians, and patients worldwide (Ring *et al*., 2006). The HVP (http://www.humanvariomeproject.org) envisions achieving improved health outcomes by facilitating the unification of human genetic variation and its impact on human health (Horaitis *et al*., 2007). It will support the use of human variation information in clinical and research environments across the world by developing the resources required to undertake the following key tasks: (1) to capture and archive all human gene variation associated with human disease developed from gene-specific curation in a central location with mirror sites in other countries to maximize data security and integrity that allows searching across all genes using a common interface, (2) to establish systems that ensure adequate curation of human variation knowledge from LSDBs, NEMDBs, or disease-specific database perspective to improve accuracy, reduce errors and develop a comprehensive data set comprising all human genes, (3) to facilitate the development of software to collect and exchange human variation data in a federation of GMDs, LSDBs, NEMDBs, and disease-specific databases, (4) to create a support system for research laboratories for the collection of genotypic and phenotypic data together using the defined reference sequence in a free, unrestricted, and open access system and create a simple mechanism

for logging discoveries, (5) to develop ethical standards that ensure open access to all human variation data that are to be used for global public good and address the needs of ''indigenous'' communities under threat of dilution in emerging countries, (6) to provide support to developing countries to build capacity and to fully participate in the collection, analysis, and sharing of genetic variation information, and (7) to establish a communication and education program to collect and spread knowledge related to human variation knowledge to all countries of the world.

This kind of distributed and coordinated effort would also, ideally, be managed in close partnership with specialized journals (Patrinos and Wajcman, 2004) to ensure that contributors not only have the means but also the incentives to publish their efforts. In 2008, *Nature Genetics* proposed a new publication scheme, termed microattribution, which could produce the required incentive for high-quality community annotation of the human genome. In brief, microattribution aims to produce a publication workflow that is open to all journals and that draws on the expertise of all those with a stake in understanding variation at a particular region of the human genome (Axton, 2008). Similarly, database journals, i.e. databases inter-relating with scientific journals, can also provide publication incentives. To this end, *Human Genomics and Proteomics*, inter-related with FINDbase (http://www.sage-hindawi.com/journals/hgp), stands as a representative example of a database journal. In brief, population-specific data sets, e.g. documenting the molecular basis of inherited disorders in different populations that could otherwise be kept unpublished and eventually risk being lost, are submitted to *Human Genomics and Proteomics* and once accepted become part of the main FINDbase data collection and their description gets deposited in PubMed literature database and registered under a unique PubMed ID. Such innovative publication modality would not only provide authors with a certain degree of credit for their contribution but also yield a nonprofit model for sustainable database funding.

Finally, the most fundamental hurdle of all that retards the field is that of limited funding. Because of this, almost all mutation databases in existence today have been built by researchers ''on the side'' for their own use, with a small degree of sponsorship/funding at best. To advance beyond this cottage industry state of affairs, projects need to be increased in scale, quality and durability, and this can only happen if strategically minded funding agencies make available substantial targeted funds. The new databases that thereby emerge will then need long-term support for general maintenance and further development. To solve this, the projects may ultimately need to be run as self-sustaining ''businesses'' that charge for data access, and/or it might be possible to develop novel forms of joint academic–corporate funding.

With great vision, the European Commission announced in 2007 the 1st call for proposals for the 7th Framework Program (FP7; thematic area HEALTH), that included a topic on ''…unifying human and model organism genetic variation databases''. This topic focused on developing a data and analysis structure by creating a hierarchy of bioinformatics grid-linked databases, tools and standards, centered on a generalized existing or novel genome browser. GEN2PHEN (http://www.gen2phen.org) is a large-scale integrating project, with 19 participating academic and corporate entities, which was funded from this call of proposals and aims to unify human genetic variation databases towards increasingly holistic views into genotype-to-phenotype (G2P) data, and to link this system into other biomedical knowledge sources via genome browser functionality. The GEN2PHEN strategy is designed so as to: (1) analyze the G2P field and investigate current needs and practices, (2) develop key standards for the G2P field, (3) create generic database components, services, and integration infrastructures for the G2P domain, (4) create data search and presentation solutions for G2P knowledge, (5) facilitate the populating of research and diagnostic G2P databases, (6) build a major G2P internet portal, (7) deploy GEN2PHEN solutions to the community, and (8) address system durability and long-term financing. It is expected that similar funding opportunities may arise from other funding bodies, such as the National Institutes of Health in the USA.

25.8 CONCLUSIONS

It is widely accepted that LSDBs and NEMDBs are increasingly becoming valuable tools in molecular diagnostics. However, the current array of GDBs, LSDBs, and NEMDBs is limited in number and in their degree of interconnection to capture all that is known and being discovered regarding pathogenic DNA mutations. The main reason for this deficiency is that the modern research ethos fails to provide adequate incentives (i.e. publication options, peer recognition, funding) to encourage researchers to build new databases. Initiatives designed to make it technically easier to set up and use such databases are welcomed, such as specialized software (see section 25.6) or interactive user interfaces (e.g. Genewindow; http://genewindow.nci.nih.gov), as are those aiming to directly transfer data from clinical diagnostics laboratories into these depositories. Apparently, the biomedical community must first appreciate the overwhelming need for improved genetic/mutation database systems and the most adequate solution will then presumably follow.

REFERENCES

Aerts, J., Wetzels, Y., Cohen, N., and Aerssens, J. (2002). Data mining of public SNP databases for the selection of intragenic SNPs. *Hum. Mutat.* 20, 162–173.

Axton, M. (2008). Human Variome microattribution reviews. *Nat. Genet.* 40, 1.

Beroud, C., and Soussi, T. (2003). The UMD-p53 database: new mutations and analysis tools. *Hum. Mutat.* 21, 176–181.

Beroud, C., Collod-Beroud, G., Boileau, C., Soussi, T., and Junien, C. (2000). UMD (Universal Mutation Database): a generic software to build and analyze locus-specific databases. *Hum. Mutat.* 15, 86–94.

Brown, A.F., and McKie, M.A. (2000). MuStaR and other software for locus-specific mutation databases. *Hum. Mutat.* 15, 76–85.

Claustres, M., Horaitis, O., Vanevski, M., and Cotton, R.G. (2002). Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. *Genome Res.* 12, 680–688.

Collod, G., Beroud, C., Soussi, T., Junien, C., and Boileau, C. (1996). Software and database for the analysis of mutations in the human FBN1 gene. *Nucleic Acids Res.* 24, 137–140.

Collod-Beroud, G., and Boileau, C. (2002). Marfan syndrome in the third millennium. *Eur. J. Hum. Genet.* 10, 673–681.

Collod-Beroud, G., Le Bourdelles, S., Ades, L., Ala-Kokko, L., Booms, P., Boxer, M., Child, A., Comeglio, P., De Paepe, A., Hyland, J.C., Holman, K., Kaitila, I., Loeys, B., Matyas, G., Nuytinck, L., Peltonen, L., Rantamaki, T., Robinson, P., Steinmann, B., Junien, C., Beroud, C., and Boileau, C. (2003). Update of the UMD-FBN1 mutation database and creation of an FBN1 polymorphism database. *Hum. Mutat.* 22, 199–208.

Cooper, D.N., Ball, E.V., and Krawczak, M. (1998). The human gene mutation database. *Nucleic Acids Res.* 26, 285–287.

Cotton, R.G., McKusick, V., and Scriver, C.R. (1998). The HUGO mutation database initiative. *Science* 279, 10–11.

Cotton, R.G., Phillips, K., and Horaitis, O. (2007). A survey of locus-specific database curation. Human Genome Variation Society. *J. Med. Genet.* 44, e72.

Cotton, R.G., Auerbach, A.D., Beckmann, J.S., Blumenfeld, O.O., Brookes, A.J., Brown, A.F., Carrera, P., Cox, D.W., Gottlieb, B., Greenblatt, M.S., Hilbert, P., Lehvaslaiho, H., Liang, P., Marsh, S., Nebert, D.W., Povey, S., Rossetti, S., Scriver, C.R., Summar, M., Tolan, D.R., Verma, I.C., Vihinen, M., and den Dunnen, J.T. (2008). Recommendations for locus-specific databases and their curation. *Hum. Mutat.* 29, 2–5.

den Dunnen, J.T., and Antonarakis, S.E. (2001). Nomenclature for the description of human sequence variations. *Hum. Genet.* 109, 121–124.

den Dunnen, J.T., and Paalman, M.H. (2003). Standardizing mutation nomenclature: why bother? *Hum. Mutat.* 22, 181–182.

Dvornyk, V., Long, J.R., Xiong, D.H., Liu, P.Y., Zhao, L.J., Shen, H., Zhang, Y.Y., Liu, Y.J., Rocha-Sanchez, S., Xiao, P., Recker, R.R., and Deng, H.W. (2004). Current limitations of SNP data from the public domain for studies of complex disorders: a test for ten candidate genes for obesity and osteoporosis. *BMC Genet.* 5, 4.

Fokkema, I.F., den Dunnen, J.T., and Taschner, P.E. (2005). LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach. *Hum. Mutat.* 26, 63–68.

Fredman, D., Munns, G., Rios, D., Sjoholm, F., Siegfried, M., Lenhard, B., Lehvaslaiho, H., and Brookes, A.J. (2004). HGVbase: a curated resource describing human DNA variation and phenotype relationships. *Nucleic Acids Res.* 32, D516–D519.

George, R.A., Smith, T.D., Callaghan, S., Hardman, L., Pierides, C., Horaitis, O., Wouters, M.A., and Cotton, R.G. (2008). General mutation databases: analysis and review. *J. Med. Genet.* 45, 65–70.

Gerlai, R. (2002). Phenomics: fiction or the future? *Trends Neurosci.* 25, 506–509.

Giardine, B., Riemer, C., Hefferon, T., Thomas, D., Hsu, F., Zielenski, J., Sang, Y., Elnitski, L., Cutting, G., Trumbower, H., Kern, A., Kuhn, R., Patrinos, G.P., Hughes, J., Higgs, D., Chui, D., Scriver, C., Phommarinh, M., Patnaik, S.K., Blumenfeld, O., Gottlieb, B., Vihinen, M., Väliaho, J., Kent, J., Miller, W., and Hardison, R.C. (2007a). PhenCode: connecting ENCODE data with mutations and phenotype. *Hum. Mutat.* 28, 554–562.

Giardine, B., van Baal, S., Kaimakis, P., Riemer, C., Miller, W., Samara, M., Kollia, P., Anagnou, N.P., Chui, D.H., Wajcman, H., Hardison, R.C., and Patrinos, G.P. (2007b). HbVar database of human hemoglobin variants and thalassemia mutations: 2007 update. *Hum. Mutat.*28, 206.

Hall, J.G. (2003). A clinician's plea. *Nat. Genet.* 33, 440–442.

Hamosh, A., Scott, A.F., Amberger, J., Bocchini, C., and McKusick, V.A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 33, D514–D517.

Hardison, R.C., Chui, D.H., Giardine, B., Riemer, C., Patrinos, G.P., Anagnou, N., Miller, W., and Wajcman, H. (2002). HbVar: A relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. *Hum. Mutat.* 19, 225–233.

Horaitis, O., Talbot, C.C., Jr., Phommarinh, M., Phillips, K.M., and Cotton, R.G. (2007). A database of locus-specific databases. *Nat. Genet.* 39, 425.

International HapMap Consortium (2003). The international HapMap project. *Nature* 426, 789–796.

Kleanthous, M., Patsalis, P.C., Drousiotou, A., Motazacker, M., Christodoulou, K., Cariolou, M., Baysal, E., Khrizi, K., Pourfarzad, F., Moghimi, B., van Baal, S., Deltas, C.C., Najmabadi, S., and Patrinos, G.P. (2006). The cypriot and iranian national mutation databases. *Hum. Mutat.* 27, 598–599.

Marsh, S., Kwok, P., and McLeod, H.L. (2002). SNP databases and pharmacogenetics: great start, but a long way to go. *Hum. Mutat.* 20, 174–179.

McKusick, V.A. (1966). *Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders*, 1st edition. Johns Hopkins University Press, Baltimore, MD.

Patrinos, G.P. (2006). National and ethnic mutation databases: recording populations' genography. *Hum. Mutat.* 27, 879–887.

Patrinos, G.P., and Wajcman, H. (2004). Recording human globin gene variation. *Hemoglobin* 28, v–vii.

Patrinos, G.P., and Brookes, A.J. (2005). DNA, disease and databases: disastrously deficient. *Trends Genet.* 21, 333–338.

Patrinos, G.P., Giardine, B., Riemer, C., Miller, W., Chui, D.H., Anagnou, N.P., Wajcman, H., and Hardison, R.C. (2004). Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. *Nucleic Acids Res.* 32, D537–D541.

Patrinos, G.P., van Baal, S., Petersen, M.B., and Papadakis, M.N. (2005a). The hellenic national mutation database: a prototype database for mutations leading to inherited disorders in the hellenic population. *Hum. Mutat.* 25, 327–333.

Patrinos, G.P., Kollia, P., and Papadakis, M.N. (2005b). Molecular diagnosis of inherited disorders: lessons from hemoglobinopathies. *Hum. Mutat.* 26, 399–412.

Riikonen, P., and Vihinen, M. (1999). MUTbase: maintenance and analysis of distributed mutation databases. *Bioinformatics* 15, 852–859.

Ring, H.Z., Kwok, P.Y., and Cotton, R.G. (2006). Human Variome Project: an international collaboration to catalogue human genetic variation. *Pharmacogenomics* 7, 969–972.

Scriver, C.R. (2004). After the genome – the phenome? *J. Inherit. Metab. Dis.* 27, 305–317.

Scriver, C.R., Nowacki, P.M., and Lehvaslaiho, H. (1999). Guidelines and recommendations for content, structure, and deployment of mutation databases. *Hum. Mutat.* 13, 344–350.

Scriver, C.R., Waters, P.J., Sarkissian, C., Ryan, S., Prevost, L., Cote, D., Novak, J., Teebi, S., and Nowacki, P.M. (2000). PAHdb: a locus-specific knowledgebase. *Hum. Mutat.* 15, 99–104.

Sipila, K., and Aula, P. (2002). Database for the mutations of the Finnish disease heritage. *Hum. Mutat.* 19, 16–22.

Smith, T.D., and Cotton, R.G. (2008). VariVis: a visualisation toolkit for variation databases. *BMC Bioinformatics* 9, 206.

Soussi, T. (2000). The p53 tumor suppressor gene: from molecular biology to clinical investigation. *Ann. NY Acad. Sci.* 910, 121–137.

Soussi, T., and Beroud, C. (2001). Assessing TP53 status in human tumours to evaluate clinical outcome. *Nat. Rev. Cancer* 1, 233–240.

Stein, L. (2002). Creating a bioinformatics nation. *Nature* 417, 119–120.

Stenson, P.D., Ball, E.V., Mort, M., Phillips, A.D., Shiel, J.A., Thomas, N.S., Abeysinghe, S., Krawczak, M., and Cooper, D.N. (2003). Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.* 21, 577–581.

van Baal, S., Kaimakis, P., Phommarinh, M., Koumbi, D., Cuppens, H., Riccardino, F., Macek, M., Jr., Scriver, C.R., and Patrinos, G.P. (2007). *FINDbase*: a relational database recording frequencies of genetic defects leading to inherited disorders worldwide. *Nucleic Acids Res.* 35, D690–D695.

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L.Y., Helmberg, W., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Miller, V., Ostell, J., Pruitt, K.D., Schuler, G.D., Shumway, M., Sequeira, E., Sherry, S.T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R.L., Tatusova, T.A., Wagner, L., and Yaschenko, E. (2008). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 36, D13–D21.

Wildeman, M., van Ophuizen, E., den Dunnen, J.T., and Taschner, P.E. (2008). Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum. Mutat.* 29, 6–13.

Zaimidou, S., van Baal, S., Smith, T.D., Mitropoulos, K., Ljujic, M., Radojkovic, D., Cotton, R.G., and Patrinos, G.P. (2009). A_1ATVar: a relational database of human *SERPINA1* gene variants leading to alpha1-antitrypsin deficiency. *Hum. Mutat.* **30**, 308–313.

Zlotogora, J., van Baal, S., and Patrinos, G.P. (2007). Documentation of inherited disorders and mutation frequencies in the different religious communities in Israel in the Israeli National Genetic Database. *Hum. Mutat.* 28(10), 944–949.

This page intentionally left blank

Chapter 26

Molecular Diagnostic Applications in Forensic Science

Bruce Budowle[1], John V. Planz[2], Rowan Campbell[2] and Arthur J. Eisenberg[2]

[1]Federal Bureau of Investigation, Laboratory Division, Quantico, VA, United States of America;

[2]DNA Identity Laboratory, University of North Texas Health Science Center, Ft. Worth, TX, United States of America

26.1 INTRODUCTION

Forensic human identity testing is focused primarily on determining the source of biological samples found at crime scenes. The biological samples encountered in criminal cases or in mass disasters can be compromised by the presence of environmental contaminants, they can be degraded, and typically they are limited in quantity. Furthermore, the samples are diverse, including such materials as bloodstains, saliva stains, semen stains, bone, teeth, hair, or muscle tissue, and are found on a variety of substrates. Therefore, there are a number of challenges to consider when developing and implementing analytical methods that genetically characterize biological evidence. These include:

- Extracting DNA of sufficient quantity and quality to enable analysis
- Developing robust analytical technology so that reliable results can be obtained
- Establishing reliable interpretation guidelines
- Conveying the results in a court of law

Historically, polymorphic protein genetic markers were used to potentially differentiate individuals (Murch and Budowle, 1986; Saferstein, 1995; Budowle and Brown, 2001). The classic protein polymorphic system is the ABO blood group system, which had been used for decades for identity testing. There are four common ABO blood types: A, B, O, and AB. For example, if a bloodstain found at a crime scene is typed as B, then only B-type individuals can be possible sources of the sample. Those who have blood types A, O, and AB (slightly more than 90% of the population) are excluded as possible sources of the bloodstain. Other protein genetic marker systems included serum and red cell proteins, such as group-specific component, transferrin, hemoglobin, phosphoglucomutase-1, acid phosphatase, esterase D, and others. However, protein-based genetic systems were limited in use because the stability or persistence is low for these proteins in biological samples exposed to the environment, the discrimination power of the systems is low, and the protein markers did not exist at typeable levels in all tissues of an individual. Typing genetic polymorphisms at the DNA level, rather than the protein level, obviates these limitations to a much greater degree.

The ability to type DNA from biological evidence is one of the most important developments in forensic science since the advent of latent fingerprint analysis. Any biological material that contains nucleated cells, including blood, semen, saliva, hair, bones, and teeth, potentially can be typed for DNA polymorphisms. The human DNA typing techniques and the battery of available genetic markers are more sensitive, more specific, and better resolving than the classic protein genetic markers just described. Most importantly, DNA technology affords the forensic scientist the ability to exclude from consideration individuals who have been falsely associated with a biological sample and to reduce the number of potential contributors of the sample to a few (if not one) individuals. The molecular biology methods available include a wide range of genetic markers and a variety of typing strategies. The methods that have been or are used routinely for human identity testing include restriction fragment length polymorphism (RFLP) typing of variable number of tandem repeat (VNTR) loci (Wyman and White, 1980; Jeffreys *et al*., 1985a, b; Budowle and Baechtel, 1990), and amplification of target DNA molecules by the polymerase chain reaction (PCR, Saiki *et al*., 1985) with subsequent typing of specified genetic markers (Saiki *et al*., 1989; Kasai *et al*., 1990; Budowle *et al*., 1991, 1996a, b; Comey and Budowle, 1991; Edwards *et al*., 1991; Sullivan *et al*., 1991a, b; Holland *et al*., 1993; Wilson *et al*., 1995a, b). A subclass of VNTR loci, known as short tandem

repeat (STR) or microsatellite loci, is the primary class of genetic markers used for human identity testing. Such loci are highly abundant in the human genome, and the forensically informative ones, residing on autosomal chromosomes, are hypervariable. The STR loci are composed of tandemly repeated sequences, each of which is two to seven base pairs (bp) in length. Loci containing repeat sequences consisting of four bp (or tetranucleotides) are used routinely for human identification (Edwards *et al*., 1991; Budowle *et al*., 1998; Chakraborty *et al*., 1999). Some pentanucleotide repeat loci are also used (Budowle *et al*., 2001c). The number of common alleles at a single forensically important STR locus ranges usually from three to eight. Because the region containing the repeats is generally quite small (an STR region with 15 copies of a 4 bp repeat is 60 bp in length), they are amenable to amplification by the PCR. Amplified alleles, however, are somewhat larger – generally less than 350 bp – due to the incorporation, into the amplicon, of regions that flank the repeat sequence and spacing requirements for multiplex analysis. Because of the use of the PCR, less than 1 ng of human template DNA is needed for STR analysis, with less than 200 pg of template DNA (an equivalent of 33 human diploid cells) at times yielding reliable typing results. The main benefit of amplifying small-sized amplicons is that many degraded DNA samples can be successfully amplified at the STR loci. Another advantage is that a number of STR loci can be amplified simultaneously in a multiplex PCR. To facilitate analysis, by reducing labor and consumption of evidence, five to 15 STR loci can be amplified in a single PCR and subsequently simultaneously typed. Commercial kits are available to assist in typing multiple STR loci (Lins *et al*., 1998; Micka *et al*., 1999; Moretti *et al*., 2001; Budowle *et al*., 2001a; Holt *et al*., 2002; Krenke *et al*., 2002, 2005). The process for typing the amplified STRs entails separating the fragments, usually by capillary electrophoresis (CE), and detecting fluorescently labeled products in real time. Profile patterns from different samples, that is, evidence and reference samples, are compared and interpretations of match, exclusion, or inconclusive are made. The multiple locus STR profiles (typically 13 loci in the USA) can be very informative.

Although the battery of 13 autosomal STR loci offers a high degree of discrimination power, such that only one or a few individuals may carry a particular multilocus profile, there are some samples from forensic cases where these core autosomal STR loci may not be informative. Forensic biological samples are often found at scenes of violent crimes and typically these are composed of more than one donor. Identification of genotypes at STR loci from a male perpetrator may not be possible in a mixed sample composed of low levels of male DNA. When there is a large background of female DNA in a mixed male/female sample, the female DNA competes for the reagents during STR amplification so that little or no male contribution is observed. By amplifying loci on male-specific DNA, for example Y-chromosome STR loci, the competition for reagents as occurs with autosomal markers is substantially reduced. Vaginal swabs, fingernail scrapings, diluted stains, and the like may provide Y-specific genetic profiles where the autosomal STR loci from the portion of the samples derived from a male perpetrator may be too low to detect or at a level too low to interpret reliably (de Knijff *et al*., 1997; Jobling *et al*., 1997; Prinz *et al*., 1997; Kayser *et al*., 1997a, b, 1998; Honda *et al*., 1999; Corach *et al*., 2001; Dekairelle and Hoste, 2001; Gusmão *et al*., 2001; Ballantyne and Hall, 2003; Budowle *et al*., 2003b; Sinha *et al*., 2003a, b).

The Y-chromosome genetic markers are the most recent class of DNA-based markers to gain interest as tools for human identity testing (Chakraborty, 1985; de Knijff *et al*., 1997; Jobling *et al*., 1997; Kayser *et al*., 1997a, 1998; Roewer *et al*., 1996, 2000, 2001). Because of their more recent history in forensic analyses, this chapter describes the general characteristics of Y-STRs in forensic analyses, some analytic procedures for typing Y-STR loci, and general interpretation issues to consider. Although only Y-chromosome and briefly mitochondrial DNA (mtDNA) markers are described, you will gain a better appreciation of forensic science as it applies to human identification.

26.2 GENERAL CHARACTERISTICS OF Y-CHROMOSOME MARKERS

Y-chromosome DNA resides in the nucleus, but markers residing on the Y-chromosome differ in some respects from the autosomal loci (de Knijff *et al*., 1997; Jobling *et al*., 1997; Gill *et al*., 2000; Budowle *et al*., 2003b; Kayser, 2003). The Y-chromosome is the smallest chromosome of the human genome (about 60 million bp), and unlike the autosomal chromosomes, it is transmitted solely paternally, from father to son. Most of the DNA in the Y-chromosome is non-recombinant, with only the most distal portions of the chromosome able to recombine with the X-chromosome. Therefore, barring mutation, the Y-linked DNA types are identical for all paternal relatives, including male siblings. This characteristic can be helpful in human identity cases, such as those involving the analysis of the remains of a missing person, where known paternal relatives can provide reference samples for direct comparison to the unknown Y-linked DNA types. Moreover, reference samples from living individuals can be used to aid in the identification of paternal relatives that are several generations removed. Such was the case with the study of the paternal lineage of Thomas Jefferson and that of male offspring of Sally Hemmings (Foster *et al*., 1998). However, due to its mode of inheritance, all paternal relatives of an accused male suspect will share the same Y-marker profile and this must be considered when evaluating and presenting evidence.

Both Y-STR loci and single nucleotide polymorphisms (SNPs) have been used in evolutionary studies and are available for identity testing. SNP-based Y-markers will

not be discussed further; these Y-markers have yet to be used significantly for forensic analyses and have been used predominately for deciphering prehistoric human migration routes of global dispersal of modern humans (Underhill *et al*., 2001; Wells *et al*., 2001; Cruciani *et al*., 2002; Lell *et al*., 2002; Mountain *et al*., 2002). Y-STR polymorphic loci are being used more routinely, and commercially available kits are available for typing forensically informative loci.

The Y-STR loci have certain characteristics that make them useful for forensic casework analyses. These include:

1. Y-markers can be useful in DNA analysis of violent crime samples, because males are involved in the majority of violent crimes (Bureau of Justice Statistics, 2001).
2. More than 100 Y-linked STR loci have been discovered (Kayser, 2003), and many are sufficiently polymorphic for forensic applications.
3. Y-STR loci reside on the non-recombinant portion of the Y-chromosome. The Y-markers are transmitted from father to male progeny as a haplotype. A haplotype is a specific array of alleles/loci observed in an individual. Because there is no independent assortment of the Y-loci, they can be very informative in multigenerational paternal lineage studies.
4. Because of the haploid nature of Y-STR loci, for most loci, only one allele per locus is displayed per individual, as opposed to two alleles for autosomal loci. Thus, interpretation of profiles in mixed samples is simplified. For example, fewer alleles per locus may make it easier to elucidate the number of male contributors in a multimale contributor sample.
5. Those Y-STR loci selected for forensic applications are male specific. DNA from a female victim does not contribute to the Y-linked DNA profile. Thus, again interpretation in mixture cases is simplified by having fewer alleles to evaluate.
6. For autosomal STR typing, about 200 pg–1 ng of template DNA are used. When the male portion of the mixed sample is too low, for example 0.3–10% of the total DNA, using 200 pg–1 ng of template DNA in the PCR will not provide meaningful information for autosomal STRs. The amount of male DNA is too low to render a detectable profile. However, much larger amounts of total female/male DNA can be placed in a PCR when typing Y-STR loci. For example, suppose that 400 ng of a sample of DNA were composed of 399 ng from a female contributor and 1 ng from a male contributor. If a total of 1 ng of DNA from this sample was used for analysis, then only the female's type would be detected (if autosomal STRs were analyzed). However, putting all 400 ng into a PCR and typing for male-specific loci would enable genetic identity profiling of the male perpetrator. Quantification assays have been developed so that relative quantities of Y-chromosomal DNA (and hence male DNA) in a cocktail of nuclear DNA can be estimated.
7. Stochastic affects due to too few template molecules in the PCR are less of an issue for interpretation of Y-STR loci than for autosomal loci. For autosomal loci, a heterozygote can appear as a "pseudohomozygote" if too few template molecules are amplified during PCR (Budowle *et al*., 2001b). For Y-STR loci, this is less of an issue. Too few copies of template DNA will result in only either the allele being amplified or the locus yielding no result. Thus, the minimum template requirements for the PCR of Y-STR loci can be lower than for autosomal STR loci.
8. The analytic tools and apparatus for typing Y-STR loci are the same as those used for typing autosomal loci. Therefore, no new equipment needs to be acquired in a forensic laboratory to implement Y-STR typing.

The International Y-STR User Group has recommended use of nine Y-STR loci for identity testing. These are the loci: DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393, defined as the European Minimal Haplotype (EMH) (Kayser *et al*., 1997b; Roewer *et al*., 2001). In the USA, the Scientific Working Group on DNA Analysis Methods (SWGDAM) also recommended the use of these nine loci as well as the loci DYS438 and DYS439 (Ayub *et al*., 2000). So that the forensic community could analyze these 11 loci in a single multiplex system, two multiplex kits have been developed. The PowerPlex® Y System kit (Promega, Madison, WI) enables amplification of these 11 loci and the additional locus DYS437 (Ayub, 2000; Krenke, 2003) for a total of 12 YSTR loci. The Y-PLEXTM 12 kit (Reliagene, New Orleans, LA) also enables amplification of the SWGDAM 11 loci plus the amelogenin locus, which is used routinely in forensic DNA typing for gender determination (Budowle *et al*., 1996a). These 11 loci provide a considerable power of discrimination, but much less than that of the battery of autosomal STR loci (Roewer *et al*., 1996; Kayser *et al*., 1997b, 2000, 2002; Budowle *et al*., 2003b). Other candidate loci that may increase haplotype gene diversity values are GATA A7.1, GATA A7.2, GATA A10, GATA C4, and GATA H4 (White *et al*., 1999; Ayub *et al*., 2000; Beleza *et al*., 2003; Quintans *et al*., 2003).

26.3 METHODOLOGY

Forensic biological specimens often are limited in quality and quantity. Sometimes only one analysis may be possible. Undue consumption of evidence may leave nothing for retesting. To be effective, techniques that are developed for crime laboratories should be sufficiently robust so practitioners can use them readily. Protocols should minimize unnecessary steps and improve, if possible, performance time, reliability, and detection sensitivity. Moreover, the

procedures used in forensic applications are usually subjected to extensive validation. Validation is a process used to assess the ability of defined procedures to reliably obtain desired results, to define conditions that are required to obtain the result, to determine the limitations of the analytical procedure, and to identify aspects that must be monitored and controlled (Budowle *et al*., 2000). Validated methods are an essential requirement for routine work in the crime laboratory.

Although not discussed in detail in this chapter, quality assurance and quality control practices are requisite for ensuring high-quality results. Adherence to using tested reagents, calibrated equipment, and known control samples is necessary to obtain reliable results with confidence from standard operating protocols. Often when problems arise while analyzing good quality samples, it is likely that quality assurance and quality control practices were not followed. A quality assurance program should be implemented to demonstrate personnel, equipment, and reagents perform as expected (Budowle *et al*., 2000; see also Chapter 40).

The following discussion addresses a subset of potential issues to consider when typing Y-STR loci in forensic samples. Crime scene samples can be degraded and contaminated with materials that inhibit analytical processes, particularly the PCR. The practices employed and issues considered in forensic analyses could be useful for scientists in other disciplines where high-quality, as well as challenging, samples may be encountered.

26.3.1 Extraction

The success of DNA typing relies on the isolation of DNA of sufficient quantity, quality, and purity. DNA can be extracted from most types of biological material found at a crime scene. Depending on the typing procedure, the required quantity (i.e. the amount of retrievable DNA) and quality (i.e. the length of the fragmented molecules) of DNA can vary widely. There will be forensic samples where the quantity is too low or the sample has been subjected to environmental insults such that they are too degraded to analyze; no extraction procedure will improve the ability to type such samples. However, with the sensitivity afforded by the PCR, many small-sized, degraded samples can be typed. Purity of the DNA extract refers to a quality of cleanliness, such that subsequent analytical assays can be carried out effectively. Generally, DNA extraction protocols that overcome, remove, or dilute PCR inhibitors are sought. Although many extraction protocols are similar, tremendous variation in the specific details are possible, and yet comparably effective results can be obtained.

Ideally, extraction protocols should be simple and inexpensive to perform. Procedures should be suitable for extraction of DNA from small liquid blood samples, bloodstains, and other fluids and tissues. Because of space, very effective procedures such as salting out (Grimberg *et al*., 1989), organic extraction (Sambrook *et al*., 1989), or use of chaotropic reagent (Scherczinger *et al*., 1997; Shedlock *et al*., 1997; Sinclair and McKechnie, 2000; Iudica *et al*., 2001) will not be described here. These well-known procedures are compatible with both RFLP- and PCR-based procedures, because the isolated DNA is double stranded. Most molecular biologists are familiar with these procedures. Instead, two procedures used more commonly in forensic applications are presented. These are the use of Chelex-100 (Singer-Sam *et al*., 1989; Walsh *et al*., 1991) and the use of FTA paper (Del Rio *et al*., 1996; Burgoyne and Rogers, 1997).

Chelex-100 (BioRad, Hercules, CA) is an ion-exchange resin that binds cations that can inhibit the PCR. The Chelex protocol entails placing a small sample in 5% (w/v) Chelex-100 and incubating the sample at 56°C for 30 minutes. After incubation, the sample is boiled and centrifuged. A portion of the supernatant is used (Budowle *et al*., 2000). Boiling denatures the DNA, rendering it unsuitable for restriction enzyme digestion (i.e. RFLP typing). However, chelex-extracted DNA is compatible with current forensic DNA quantitation assays and with the PCR. The DNA extract used for the PCR should not contain residual Chelex, because the material will chelate magnesium ion (essential for polymerase activity). Chelex-100 can be used on a variety of tissues sources.

The other method of DNA purification, particularly useful for reference samples, is the washing away of cellular debris from DNA immobilized in FTA paper (Fitzco, Minneapolis, MN; Whatman, Clifton, NJ). Blood is spotted on FTA paper, the cells are automatically lysed, the DNA released from the cells is immobilized in the FTA matrix, and the sample is dried. Heme, or other cellular debris that may inhibit enzymatic activity in subsequent assays, is simply washed away without loss of DNA. The washed, immobilized DNA, housed in a circular pouch (approximately 1.2 mm in diameter), can be used directly as a DNA template source.

Blood deposited on FTA paper is an excellent source of genomic DNA. The paper facilitates handling, prevents degradation of the DNA, and enables storage of the material at ambient temperature. In addition to blood, epithelial cells from buccal swabs may be transferred and stored within the FTA matrix, facilitating the processing of reference samples. Buccal cells may be applied directly by a sponge applicator swab to the FTA paper during the sample collection phase, or the cells may be eluted from swabs and then placed on the FTA paper. Other cell-containing materials, such as urine, may be stored and processed with the FTA system.

The FTA system also provides several safety and quality assurance benefits. The matrix is impregnated with several reagents that are deleterious to bacteria, viruses, and fungi, as well as serving as a free radical trap to prevent DNA damage. Therefore handling and transport are simplified. The materials stored in FTA paper present fewer biohazard concerns. Moreover, the materials can be stored at ambient

temperature without concern regarding degradation; thus, refrigerated storage is unnecessary. The paper matrix is available in several formats from single spot cards to 96-positions per sheet for both manual and robotic processing. The DNA is immobilized in the matrix and experiments have demonstrated no cross-contamination during multiple sample sheet processing.

The evidence from rape cases, such as vaginal swabs and stained clothing, most often contains nucleated cells from the male contributor (i.e. predominantly sperm) and the female victim (i.e. epithelial cells). Sperm cells can be separated from other cells during extraction, because their cell membranes contain thiol-rich proteins and are resistant to cell lysis in the absence of a reducing agent (Gill *et al*., 1985; Giusti *et al*., 1986). The absence of a reducing agent results in a preferential lysis of non-sperm cells. The non-sperm DNA is released into the supernatant, whereas the sperm cells remain intact and are pelleted. After removal of the supernatant, the intact sperm can be lysed in the presence of a reducing agent. Many different extraction procedures have been modified to enable this differential extraction of sperm and non-sperm DNA.

26.3.2 Typing

The PCR process is well known and will not be described here. Commercial kits are available that contain the reagents necessary to amplify at least the 11 core SWGDAM Y-STR loci.

After the quantity of recovered DNA is determined (Waye *et al*., 1989; Walsh *et al*., 1992; Budowle *et al*., 1995; Giusti and Budowle, 1995), typically 150 pg to 1 ng of male template DNA are placed into the PCR. Because Y-STR loci are only being typed, female DNA can be present in much greater quantities and not impact on the male-specific results. For example, samples containing up to and beyond 1,000 ng of female DNA, along with the male DNA, have provided successful Y-STR results (Sinha *et al*., 2003a, b). Until recently, forensic DNA quantitation methods were human (and higher primate) specific, but not gender specific. Recently, a real-time PCR method was developed that can estimate the amount of male DNA in a sample. The Quantifiler™ Y Human DNA Quantification Kit (Applied Biosystems, Foster City, CA) has a dynamic range of at least 23 pg to 50 ng, is human (and higher primate) specific, and is commercially available.

The loci for analysis have been determined by the SWGDAM, giving direction to commercial manufacturers to develop requisite loci containing kits. The PowerPlex® Y System kit (Promega, Madison, WI) and the Y-PLEX™ 12 kit (Reliagene, New Orleans, LA) contain reagents needed for amplification, to include primer sets, which are specific for the various loci, PCR buffer, allelic ladders, and an internal lane standard. AmpliTaq Gold® DNA polymerase (Applied Biosystems, Foster City, CA) is purchased separately. Allelic ladders contain the common alleles in the general population for each locus and are used to normalize electrophoretic migration for the PCR products analyzed (see Fig. 26.1). The basic criteria for the 11 core Y-STR loci are displayed in Table 26.1. The primers are designed to amplify each locus specifically and tested for cross reactivity with other regions of the human genome, as well as with the other primers in the multiplex. The primer sets contained within each kit consist of both unlabeled and those labeled with one of three distinctive fluorescent dyes. One primer of each primer pair per Y-STR locus is labeled with a fluorescent dye so that PCR products can be detected during electrophoretic separation. For Y-STR typing, fluorescent tags are covalently bound to the 5′ end of one of the primers of a primer pair per locus. Thus, a fluorescent molecule is incorporated into PCR products and detected using laser-induced fluorescence. The internal lane standard is labeled with another distinctive fluorescent dye. The use of multicolor dyes permits the analysis of loci with overlapping size ranges.

The polymorphism of STR loci is based on different numbers of repeats contained within the locus (or amplified product) among individuals in the population. Thus, alleles are determined by the different size of amplified products, differing by repeat increments or portions of repeats (i.e. micro-variants). The amplified fragments are separated according to size by CE using, for example, the ABI Prism™ 310 or 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA). CE, rather than polyacrylamide slab gel electrophoresis, is used to fractionate the amplified products for allele designation, primarily because of increased automation and reduction of the need for manual skills in gel preparation by the analyst (see also Chapter 5). With CE, manual gel pouring and sample loading are eliminated; loading a sieving medium into the capillary and sample injection is achieved by automatic means. Because the charge-to-mass ratio of different size DNA molecules is the same, mobility in free solution is independent of fragment size. Therefore, a sieving medium is required to separate DNA fragments. Use of a soluble medium is advantageous, because the buffer and the polymer can be treated in a similar fashion to a free solution separation and readily pumped into the capillary. Therefore, the injection of the sieving medium can be automated, and fresh sieving medium can be used with each CE run. Refilling capillaries also reduces the chance of contamination from previous samples that were analyzed in the same capillary.

The high-surface area-to-volume ratio of a capillary enables electrophoretic separations to be carried out at very high field strengths by increasing heat dissipation. Faster separation times are thereby achieved, and resolution may be improved compared with some slab gel electrophoresis methods. Furthermore, real time detection is performed with CE. No post-electrophoresis detection is required. Results are stored directly in the computer, thus facilitating subsequent data manipulations and analyses.

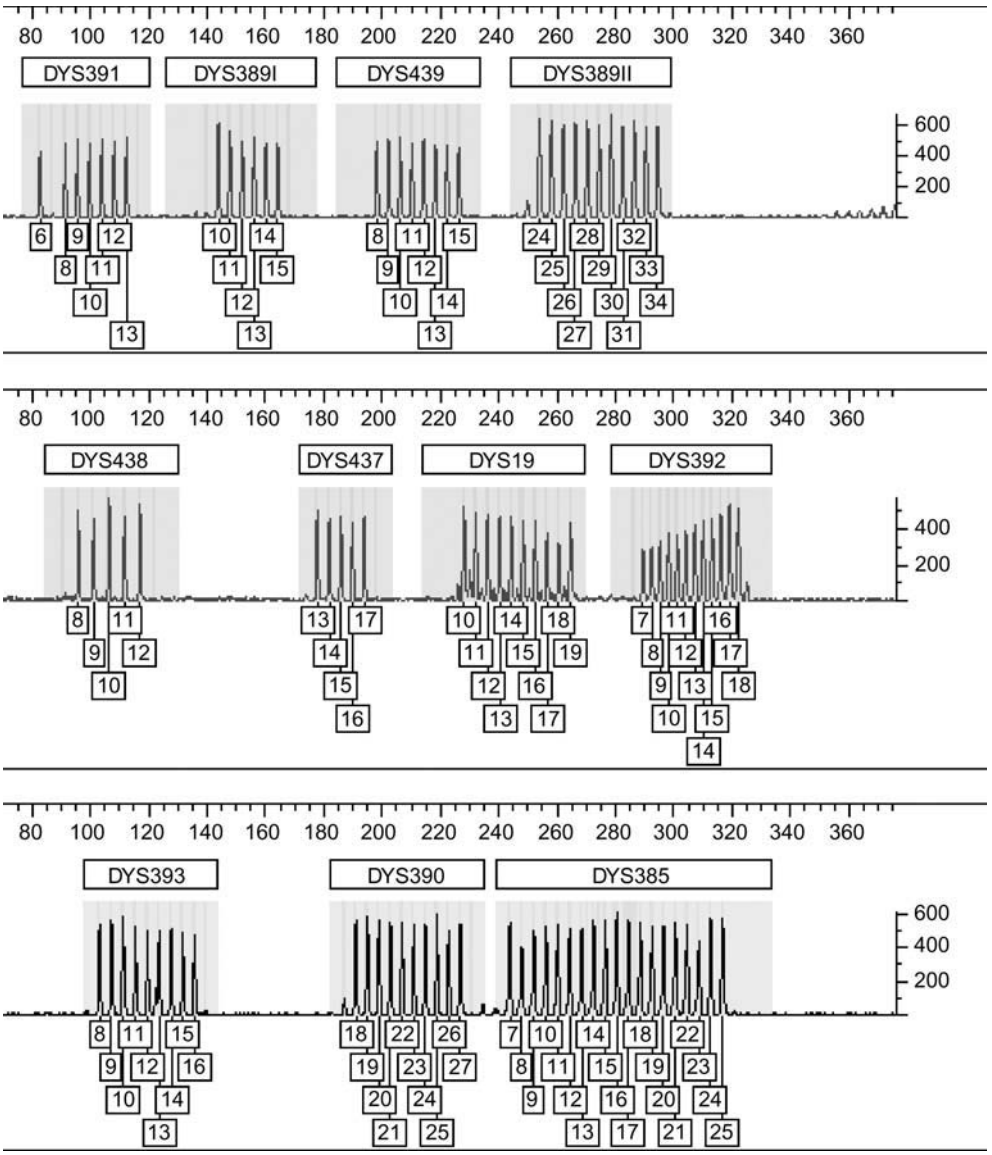


FIGURE 26.1 Profile of allelic ladder provided with the Power-Plex® Y System kit (Promega, Madison, WI). The common alleles at each locus are included. The values 80–360 above the peaks are scan units that relate to elution time. The numbers on the Y-axis (200, 400, 600) are relative fluorescent units and are correlated to relative yield of amplified product. Kindly provided by B. Krenke of the Promega Corporation.

Injection of DNA samples is performed electrokinetically. The amount of the DNA in the sample is dependent on the mobility and concentration of the other ions in the sample. The preferential injection of higher charge-to-mass molecules (e.g. Cl-1 ions) with an electrokinetic injection affects, to a degree, the quantity of samples that can be injected into the capillary. Dialysis of samples can remove such competing ions. For most forensic protocols, the samples do not need removal of competing ions.

26.3.3 Interpretation of Results

For the application of Y-STR loci, interpretation issues need to be considered. Many are the same as those for autosomal STR loci. These include: a quality evaluation of the data, whether the sample is composed of multiple donors or a single source sample, and the significance of matching data.

Figure 26.2 shows a Y-STR profile from a single male donor. Each locus displays one allele, except the DYS385 locus, which can present two alleles, due to a tandem duplication of the locus (Kayser *et al*., 1997b). A minimum amount of DNA is required to evaluate the profile. This threshold interpretation level is determined in-house during validation and is set at a minimum relative fluorescent units (rfus) level. When two or more alleles are observed at multiple loci, then the sample is likely a mixture. In addition to the alleles, there are four artefact classes to consider before rendering an interpretation of a mixture. The first

TABLE 26.1 General criteria of Y-STR loci in the two commercial multiplex kits*

Locus	Dye (Promega)	PCR Product Size (bases) (Promega)	Dye (Reliagene)	PCR Product Size (bases) (Reliagene)	Allele Range	Repeat Motif	GenBank Accession #
DYS392	JOE	294–327	FAM	103–139	37789	TAT	G09867
DYS390	TMR	191–227	FAM	163–207	17–28	TCTA / TCTG	G09611 AC011289
DYS385a/b	TMR	243–315	FAM	220–288	37826	GAAA	Z93950
DYS393	TMR	104–136	JOE	100–136	37849	AGAT	G09601
DYS389I	FL	148–168	JOE	179–207	37910	TCTG / TCTA	G09600 AF140635
DYS391	FL	90–118	JOE	230–262	37785	TCTA	G09613
DYS389II	FL	256–296	JOE	292–332	24–34	TCTG / TCTA	G09600 AF140635
DYS437	JOE	183–199	–	–	38002	TCTA / TCTG	AC002992
DYS19	JOE	232–268	NED	174–210	37912	TAGA	X77751
DYS439	FL	203–231	NED	230–258	37847	GATA	AC002992
DYS438	JOE	101–121	NED	292–327	37785	TTTTC	AC002531
Amelogenin	–	–	NED	104–110	X,Y	–	M55418 and M55419

Data are from PowerPlex® Y System Technical Manual No. D018 (Promega Corporation, Madison, WI) and Shewale and colleagues (2004).

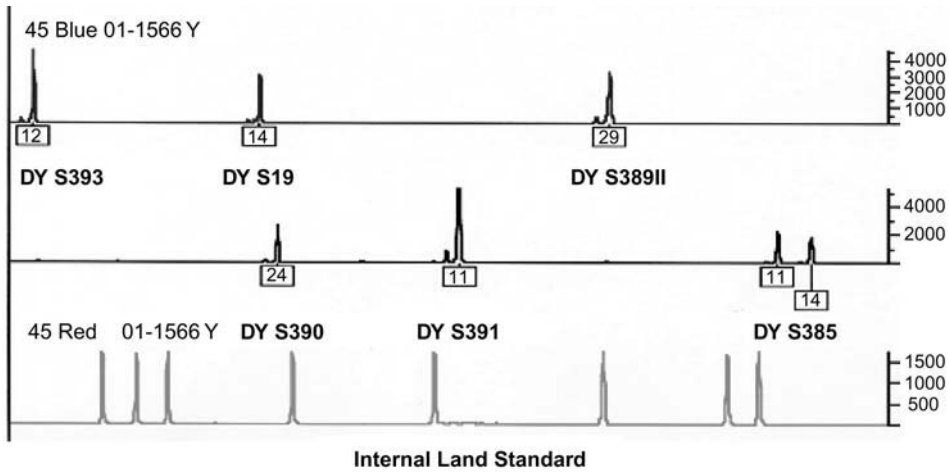


FIGURE 26.2 A six-locus Y-STR profile from a single male donor, generated using the Y-PLEXTM 6 kit (Reliagene, New Orleans, LA). The internal lane standard is a concoction of non-human DNA fragments of known mass; the standard is placed in every sample after PCR and prior to electrophoresis so DNA fragment sizes can be estimated and normalized. Kindly provided by S. Sinha of Reliagene.

are stutter peaks. These are amplification products typically one repeat smaller in size than the true allele. Stutter is inherent in the PCR and most likely due to repeat slippage (Levinson and Gutman, 1987; Schlotterer and Tautz, 1992). Most stutter products range from not detectable to about 15% of the associated allele (Sinha *et al*., 2003a, b). The specified YSTR loci were selected because stutter products are nominal. STR loci with smaller size repeats, particularly dinucleotide repeats and homopolymeric stretches, tend to have high levels of stutter and would substantially

complicate interpreting mixtures. The stutter peaks are considered more so, when interpreting mixtures samples where there is a minor contributor (displaying allelic products comparable in yield to that of the stutter peaks).

Terminal nucleotide addition is another artefact generated during the PCR (Smith *et al*., 1995; Brownstein *et al*., 1996; Magnuson *et al*., 1996), but is more controllable than generation of stutter. The enzyme *Taq* polymerase has a tendency to add a nucleotide, generally adenine, to the 3′ end of the amplicon without the presence of a template. Addition is termed +A and the true amplicon size is termed −A. The addition of the terminal nucleotide is dependent on temperature and the sequence adjacent to the extension site. There are two approaches to address addition of the nucleotide: either design the assay to minimize addition or design the assay to promote addition. It is easier to drive the reaction toward adenine addition. By adding an extension step at 60°C for 30 minutes to one hour after the PCR cycling, almost all of the amplified products achieve the +A state (Moretti *et al*., 2001). Allelic ladders provided with commercial kits contain alleles in the +A state as well. In some casework analyses, sample profiles may display significant −A products. These are indicative of too much template DNA in the PCR or the presence of PCR inhibitors in the sample. To eliminate the presence of –A products, one can dilute the sample and subject it to another PCR. Alternatively, after the PCR, fresh *Taq* polymerase can be added to the PCR and only the extension step at 60°C is carried out. AmpliTaq Gold® DNA polymerase should not be used in the latter approach. To activate AmpliTaq Gold® DNA polymerase, a heat activation step is required, which will denature the DNA duplex. Terminal nucleotide addition requires double-stranded DNA.

Another potential artefact is due to fluorophore synthesis by-products (or by leaching of bound tags). These products are not DNA derived and can appear in negative control samples. Often these artefacts are not observed until very small quantity samples are analyzed. The fluorescent products are present but cannot be seen in good quantity samples due to the scale. Because these by-products are present in negative controls, they can be readily considered when interpreting results.

The last artefact to consider is cross-reactivity of Y-STR primers with female DNA. The PCR primers are designed not to generate products with female DNA (and particularly with loci on the X-chromosome). However, when excessive amounts of female DNA are in the PCR, non-specific products may be observed. Sinha and colleagues (2003a) observed two TAMRA-labeled amplified products of sizes 255 bp and 448 bp when the quantity of female DNA in the PCR was 10 ng or greater. Thus, when an evidence sample dictates that large amounts of female DNA are placed in a PCR, it is imperative to run a positive control from, for example, the female victim with similar amounts of template DNA.

If the negative control samples show no product, the positive control sample types correctly, and the evidence and reference profiles are determined to be of sufficient quality, DNA profiles from evidentiary sample(s) are compared with profiles from known exemplar(s) to determine whether or not they are similar. There are three general interpretations that can be rendered:

1. Inclusion, or match – the Y haplotypes from the two samples are sufficiently similar and potentially could have originated from the same source.
2. Exclusion – the Y-haplotype profiles are dissimilar and could not have originated from the same source.
3. Inconclusive – there are insufficient data to render an interpretation (Budowle *et al*., 2003b).

Exclusions and inconclusive results require no further analysis. However, when a Y-haplotype obtained from a forensic specimen matches that of a suspect or victim (or cannot be excluded as arising from a biological relative), it is common practice to place some significance upon the likelihood of such an occurrence. Accordingly, statistics derived from population data are applied with the intent to provide an estimate of how common or rare a DNA profile is in the relevant population(s). The mode of Y-STR inheritance must be considered to decide the statistical approach to use when placing weight on an observed Y-STR haplotype. As described earlier, the Y-STR loci reside on the non-recombinant portion of the Y-chromosome. Thus, the rarity of a multilocus Y-STR profile cannot be estimated as the combined product of the allele frequencies at each locus, as is done for the autosomal STR loci (Roewer *et al*., 1996, 2000, 2001; de Knijff *et al*., 1997; Jobling *et al*., 1997; Kayser *et al*., 1997b, 2002; Gill *et al*., 2000; Budowle *et al*., 2003b). They must be evaluated as a haplotype of loci (i.e. inherited as a block). This can be achieved by evaluating the rarity of the Y-STR haplotypes in reference to which the observation of match is made (Budowle *et al*., 2003b).

The counting method is one method employed (and the most likely to be used in the USA) to convey an estimate of the rarity of the Y-haplotype (Budowle *et al*., 2003b). Basically, the number of times a particular haplotype is observed in a reference database(s) is counted and divided by the number of profiles in the data set. Then, a correction for sampling error is applied. It is a very conservative approach, because although all loci are physically linked, they are not all in disequilibrium to the same extent (Budowle *et al*., 2003b), likely due to the relatively high mutation rate of STR loci (Kayser *et al*., 2000).

The reference population database(s) used in forensics is composed of convenience samples (NRC II Report 1996);

that is, they are collected from such sources as paternity test laboratories and blood banks. The databases generally are divided into major population groups of the potential contributors of the evidence. For forensic applications the relevance and representativeness of these databases are considered and for autosomal loci the populations are relatively homogeneous within a major population group, such as Caucasians, African Americans, and Hispanics. Because of the lack of recombination and smaller effective population size of Y-markers compared with autosomal markers, the degree of heterogeneity is expected to be greater among populations within a major population group. Y-haplotype data demonstrate population substructure more so than the autosomal loci (de Knijff *et al*., 1997; Jobling *et al*., 1997; Kayser *et al*., 1997a; Kayser, 2003; Budowle *et al*., 2003b). Therefore, correcting for effects of substructure when estimating the rarity of the profile needs to be instituted. Such practices are already applied to calculations of the rarity of autosomal genotypes. The inbreeding coefficients will be larger for Y-STR loci. Data support that within Europe, interpopulation variability for Y-STR loci tends to be low within a geographic region. Common haplotypes appear throughout Europe. The fst values, similar to Wright's Fst values, are typically <0.01, such as those observed in central Europe. However, across Europe different regional affiliations are evident. The fst values can be as high as 0.10. Not surprisingly, distances for European to non-European range from 0.2 to 0.5 (Roewer *et al*., 1996, 2000). The USA is somewhat more amalgamated than Europe and people tend to migrate more so. Kayser and colleagues (2003) analyzed African American, US Caucasian, and Hispanic Y-STR data (on European minimal haplotype) and did not find significant geographic heterogeneity for Y-STR haplotypes within the major US population groups. For the major populations in the USA the effects of substructuring for the Y-chromosome are shown to be less than those observed across Europe.

26.4 CASE EXAMPLES

There are many cases that have been analyzed and myriad examples that could be presented to gain an appreciation of the application of Y-STR analysis in identity testing. Two cases are described generally: a paternity case and an assault case.

In a paternity test, a woman accused a man of being the father of her child and the man denied the allegation. A paternity test was undertaken using 15 autosomal STR markers. After accounting for the maternally inherited alleles, there were two loci in the child that had alleles (paternal in origin) that were not observed in the alleged biological father. Because of the relatively high mutation rate for STR loci (approximately at 10^{-3}), one or two exculpatory loci are not sufficient for excluding the man as the biological father. However, the paternity index is substantially reduced because of these two loci. Further testing is required to attain a sufficient paternity index. The 12-locus Y-haplotype for the child and the alleged father were identical. Thus, the genetic data could not exclude the alleged father as the biological father and because of the Y-STR typing a paternity index greater than 1,000 was obtained.

A woman was attacked at night and her identification of the assailant was questioned, but the alibi of the suspect was weak. However, she did manage to scratch the arm of her assailant. The material under her fingernails was removed by scraping. There was very little tissue from her male assailant. In fact, most of the tissue obtained derived from the victim. An attempt was made to type the extracted DNA with the 13 core STR loci and the result yielded a single source profile consistent with a reference sample from the victim. Subsequently, the total recovered DNA (about 200 ng) was typed for six Y-STR loci. A profile was generated equivalent to about 750 pg of male template DNA, and the Y-haplotype was different than that derived from the reference sample of the accused. If not for the special qualities afforded by Y-specific DNA testing, the suspect would not have been excluded as a possible perpetrator.

26.5 MITOCHONDRIAL DNA MARKERS ANALYSIS IN FORENSIC SCIENCE

Although the focus of this chapter has been on male lineage Y-STRs, it is worth mentioning briefly the forensic value and application of the female lineage equivalent; that is, mitochondrial (mt) DNA. Mitochondria are subcellular organelles that contain an extrachromosomal genome, separate and distinct from the nuclear genome. Human mtDNA differs from nuclear DNA in the following ways:

- It exists as a closed circular, rather than linear, genome.
- The genome is substantially smaller, consisting of approximately 16.5 kilobases.
- The molecule predominantly consists of coding sequences (for two ribosomal RNAs, 22 transfer RNAs, and 13 proteins), and only one notable non-coding region approximately 1,100 bp long, called the displacement loop (D-loop) or control region.
- All mtDNA molecules are inherited maternally.
- There are many copies per cell.
- Similar to the Y chromosome, mtDNA does not undergo recombination (for review see Budowle *et al*., 2003a).

The main advantage of analyzing mtDNA, compared with nuclear DNA, is the high copy number of mtDNA (Copenhagen and Clayton, 1974). Although each set of nuclear chromosomes is present in only two copies per cell, hundreds to thousands of mtDNA molecules reside within

a cell. For forensic cases, where the evidence is particularly challenging, such as old bones, severely decomposed or charred remains, or single hair shafts, there may be insufficient quantity and/or quality nuclear DNA for analysis. However, because of sheer copy-number, the probability of obtaining a successful typing result from mtDNA is higher than that from nuclear DNA. One can also hypothesize that the circular structure of the mtDNA may make the molecule more resistant to nuclease activity than nuclear DNA. Thus, mtDNA is likely to persist longer than nuclear DNA under environmentally difficult conditions.

Instead of variation in the number of repeats being exploited for forensic identification, the variation among individuals is at the sequence of the nucleotides in the mtDNA. The greatest concentration of genetic variation resides in two hyper variable regions in the control region designated HV1 and HV2 (Greenberg *et al*., 1983). These are sequenced using the Sanger method (Sanger *et al*., 1977), where four different fluorescently labeled dideoxyribonucleoside triphosphate analogs (ddATP, ddCTP, ddGTP, and ddTTP) are placed in the chain synthesis reaction. The ddNTP in the sequencing reaction is incorporated into the growing chain by complementary base pairing to the template, as are dNTPs, and therefore competes with its dNTP analog for incorporation. However, chain elongation is terminated at the point where the ddNTP is incorporated since, unlike dNTPs, ddNTPs do not have a 3′ hydroxyl group, which is necessary for chain elongation. Thus, the extended fragments are labeled by incorporation of a ddNTP with a fluorescent dye attached. The labeled fragments are separated and detected by CE as is done with Y-STRs.

Because the mtDNA molecule does not undergo recombination, the mtDNA markers, barring mutation, are identical for all maternal relatives, which includes siblings. Thus, mtDNA typing can be useful in cases, such as those involving the analysis of the remains of a missing person, where known maternal relatives, separated by several generations, can provide meaningful reference samples for direct comparison to the questioned mtDNA type (Ginther *et al*., 1992; Sullivan *et al*., 1992; Holland *et al*., 1993). For example, if one were attempting to identify the remains of a child using mtDNA analysis, reference samples could be obtained from the alleged mother, siblings, maternal grandmother, or children if the missing person is female. One of the first and most noted cases applying mtDNA typing was the identification of the remains of Tsar Nicholas II. The maternal reference samples were obtained from Countess Xenia Cheremeteff-Sfiri and the Duke of Fife (Gill *et al*., 1994; Ivanov *et al*., 1996).

Although bones and teeth are typical sources for identification in such cases as Tsar Nicholas II and in the identification of human remains in general, an important forensic application of mtDNA sequencing is in the analysis of hair shafts. Individual hairs contain very limited quantities of nuclear DNA, such that no molecular analysis of nuclear DNA markers is possible. However, mtDNA sequences can be obtained from as little as 1–2 cm of a single hair shaft (Wilson *et al*., 1995a, b; see Fig. 26.3). Thus, hair, a biological material often found at crime scenes, can be genetically characterized routinely to assist in resolving criminal cases.

26.6 LEGAL ADMISSIBILITY

DNA evidence, in itself, does not prove guilt or innocence. It conveys information about who may have contributed the sample and those who could not be the source of the evidence sample. Other evidence about the crime is considered by the fact finder to determine whether or not an accused individual is guilty. It is important to appreciate that the forensic field has a particular constraint not routinely encountered in other scientific disciplines, and that is, the results are often presented in a non-science forum, which is the courtroom. Although many genetic typing technologies have been demonstrated to be robust and reliable, the law determines the admissibility of the technology. The court carries out a process, called an admissibility hearing, to ensure that there is some level of confidence in the evidence before it may be presented in a court proceeding. The federal standard for admissibility, Daubert versus Merrell Dow Pharmaceuticals, Inc. (1993), places the trial judge as a gatekeeper for determining that an expert's testimony both

FIGURE 26.3 A sequence of a portion of the HV2 region of the human mtDNA genome. The top panel is from mtDNA derived from a hair shaft. The bottom panel is from mtDNA derived from a whole blood sample.

rests on a reliable foundation and is relevant to the issue(s). Factors to consider in evaluating subsequent expert testimony include:

- The testability of the science
- Peer review
- The error rate of the method
- Existence of standards
- General acceptance by those who are familiar with it

The court′s role is to attempt to screen out inappropriate and misleading (i.e. termed ''junk'') science.

Science and the law do not obtain the truth in the same manner. In science, beliefs and findings are questioned continuously using the scientific method. Hypotheses are proposed and experiments are carried out to test hypotheses. If the data do not refute a hypothesis, the hypothesis gains more support and through constructive incremental steps the hypothesis becomes grounded and accepted as reasonable and reliable. In the legal setting, an adversarial approach is used. Admissibility challenges to scientific evidence are classic examples of bias in ascertainment. Rarely is exculpatory evidence challenged. Inculpatory evidence is what is typically challenged for admissibility. Therefore, results that support the prosecution's hypothesis of guilt are the likely ones to be challenged and adversarially criticized. Even if the result is objectively obtained, it will support one side's beliefs over that of the other. Typically, a defense attorney is placed in the position or has the responsibility to vigorously create doubt, regardless of his/her personal beliefs of the client's guilt. However, at times, a prosecuting attorney may want to create doubt. In the courtroom, one may exploit the standard practice of science ''to question'' as lack of consensus, even if most, if not all, agree that the approach is reliable. Because the challenges are adversarial in the courtroom, the scientific arena and testability have been and are a better forum for determining the validity of forensic analytical methods.

26.7 CONCLUSIONS

The advent of molecular biology has been a boon to human forensic identification cases. Small, environmentally insulted samples can be genetically characterized and, in many cases, to such resolution that only one or a few individuals could be the source of the samples. In fact, DNA typing is one of the best exculpatory tools available when analyzing biological evidentiary materials. The technology and applications have become well established and routinely used. Now methods are being developed to address more challenging samples. One set of markers, those residing on the Y-chromosome, enable the forensic scientist to analyze samples that were not possible previously, such as mixtures composed of minute amounts of male DNA amidst a large background of female DNA. The Y-STRs were presented to exemplify some of the issues to consider when implementing DNA typing methods into the forensic laboratory. Core sets of Y-STR loci have been defined and commercial kits have been developed. For methods to be used routinely, they must be validated, quality assurance measures must be implemented, interpretation guidelines instituted, and methods should be developed for placing statistical weight on an evidence profile that matches a reference sample profile.

ACKNOWLEDGEMENTS

This is publication number 04-03 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

REFERENCES

Ayub, Q., Mohyuddin, A., Qamar, R., Mazhar, K., Zerjal, T., Mehdi, S.Q., and Tyler-Smith, C. (2000). Identification and characterization of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res.* 28, e8.

Ballantyne, J., and Hall, A. (2003). Strategies for the design and assessment of Y-short tandem repeat multiplexes for forensic uses. *Forensic Sci. Rev.* 15, 137–149.

Beleza, S., Alves, C., Gonzalez-Neira, A., Lareu, M., Amorim, A., Carracedo, A., and Gusmao, L. (2003). Extending STR markers in Y chromosome haplotypes. *Int. J. Legal Med.* 117, 27–33.

Brownstein, M.J., Carpten, J.D., and Smith, J.R. (1996). Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* 20, 1004–1010.

Budowle, B., and Baechtel, F.S. (1990). Modifications to improve the effectiveness of restriction fragment length polymorphism typing. *Appl. Theor. Electrophoresis* 1, 181–187.

Budowle, B., Chakraborty, R., Giusti, A.M., Eisenberg, A.J., and Allen, R.C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *Am. J. Hum. Genet.* 48, 137–144.

Budowle, B., Baechtel, F.S., Comey, C.T., Giusti, A.M., and Klevan, L. (1995). Simple protocols for typing forensic biological evidence: chemiluminescent detection for human DNA quantitation and RFLP analyses and manual typing of PCR amplified polymorphisms. *Electrophoresis* 16, 1559–1567.

Budowle, B., Koons, B.W., and Errera, J.D. (1996a). Multiplex amplification and typing procedure for the loci D1S80 and amelogenin. *J. Forensic. Sci.* 41, 660–663.

Budowle, B., Koons, B.W., Keys, K.M., and Smerick, J.B. (1996b). Methods for typing the STR triplex CSF1PO, TPOX, and HUMTHO1 that enable compatibility among DNA typing laboratories. In *Advances in Forensic Haemogenetics 6* (Carracedo, A., Brinkmann, B., and Bar, W., eds), pp. 107–114. Springer-Verlag, Berlin.

Budowle, B., Moretti, T.R., Niezgoda, S.J., and Brown, B.L. (1998). CODIS and PCR-based short tandem repeat loci: law enforcement tools. In *Second European Symposium on Human Identification 1998*, pp. 73–88. Promega Corp., Madison, Wisconsin.

Budowle, B., Smith, J.A., Moretti, T., and DiZinno, J. (2000). *DNA Typing Protocols: Molecular Biology and Forensic Analysis*. BioTechniques Books, BioForensic Sciences Series, Eaton Publishing, Natick, MA.

Budowle, B., Collins, P.J., Dimsoski, P., Ganong, C.K., Hennessy, L.K., Leibelt, C.S., Rao-Coticone, S., Shadravan, F., and Reeder, D.J. (July 2001a). Population data on the STR loci D2S1338 and D19S433. *Forensic Sci. Commun.* 3(2). Available: http://www.fbi.gov/hq/lab/fsc/backissu/july2001/index.htm.

Budowle, B., Hobson, D.L., Smerick, J.B., and Smith, J.A.L. (2001b). Low copy number –consideration and caution. In *Twelfth International Symposium on Human Identification 2001*. Promega Corp., Madison, WI. http://www.promega.com/ussymp12proc/default.htm.

Budowle, B., Masibay, A., Anderson, S.J., Barna, C., Biega, L., Brenneke, S., Brown, B.L., Cramer, J., De Groot, G.A., Douglas, D., Duceman, B., Eastman, A., Giles, R., Hamill, J., Haase, D.J., Janssen, D.W., Kupferschmid, T.D., Lawton, T., Lemire, C., Llewellyn, B., Moretti, T., Neves, J., Palaski, C., Schueler, S., Sgueglia, J., Sprecher, C., Tomsey, C., and Yet, D. (2001c). STR primer concordance study. *Forensic. Sci. Int.* 124, 47–54.

Budowle, B., and Brown, B.L. (2001). The use of DNA typing for forensic identification. *Forensica* 1, 9–37.

Budowle, B., Allard, M.W., Wilson, M.R., and Chakraborty, R. (2003a). Forensics and mitochondrial DNA: applications, debates, and foundations. *Annu. Rev. Genomics Hum. Genetics* 4, 119–141.

Budowle, B., Sinha, S.K., Lee, H.S., and Chakraborty, R. (2003b). Utility of Y chromosome STR haplotypes in forensic applications. *Forensic Sci. Rev.* 15, 153–164.

Bureau of Justice Statistics (2001). Criminal Victimization in the United States, Table 38, Library of Congress, NCJ 197064, available at http://www.ojp.usdoj.gov/bjs/pub/pdf/cvus01.pdf.

Burgoyne, L., and Rogers, C. (1997). Bacterial typing: storage and processing of stabilized reference bacteria for polymerase chain reaction without preparing DNA – an example of an automatable procedure. *Anal. Biochem.* 247, 223–227.

Chakraborty, R. (1985). Paternity testing with genetic markers: are Y-linked genes more efficient than autosomal ones? *Am. J. Med. Genet.* 21, 297–305.

Chakraborty, R., Stivers, D.N., Su, B., Zhong, Y., and Budowle, B. (1999). The utility of STR loci beyond human identification: implications for the development of new DNA typing systems. *Electrophoresis* 20, 1682–1696.

Comey, C.T., and Budowle, B. (1991). Validation studies on the analysis of the HLA-DQ alpha locus using the polymerase chain reaction. *J. Forens. Sci.* 36, 1633–1648.

Copenhagen, D., and Clayton, D.A. (1974). The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. *J. Biol. Chem.* 249, 7791–7795.

Corach, D., Risso, L.F., Marino, F., Penancino, G., and Sala, A. (2001). Routine Y-STR typing in forensic casework. *Forensic Sci. Int.* 118, 131–135.

Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., Coia, V., Wallace, D.C., Oefner, P.J., Torroni, A., Cavalli-Sforza, L.L., Scozzari, R., and Underhill, P.A. (2002). A back migration from Asia to sub-Saharan Africa is supported by high resolution analysis of human Y-chromosome haplotypes. *Am. J. Hum. Genet.* 70, 1197–1214.

Daubert versus Merrell Dow Pharmaceuticals, Inc. 509 US 579, 1993.

De Knijff, P., Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J., Weichhold, G., and Roewer, L. (1997). Chromosome Y microsatellite: population genetic and evolutionary aspects. *Int. J. Legal Med.* 110, 134–149.

Dekairelle, A.F., and Hoste, B. (2001). Application of Y-STR pentaplex PCR (DYS19, DYS389I and II, DYS390, and DYS393) to sexual assault cases. *Forensic Sci. Int.* 118, 122–125.

Del Rio, S., Marino, M.A., and Belgrader, P. (1996). Reusing the same bloodstained punch for sequential DNA amplifications and typing. *BioTechniques* 20, 970–974.

Edwards, A., Civitello, A., Hammond, H.A., and Caskey, C.T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* 49, 746–756.

Foster, E.A., Jobling, M.A., Taylor, P.G., Donnelly, P., de Knijff, P., Mieremet, R., Zerjal, T., and Tyler-Smith, C. (1998). Jefferson fathered slave's last child. *Nature* 396, 27–28.

Gill, P., Jeffreys, A.J., and Werrett, D.J. (1985). Forensic application of DNA fingerprints. *Nature* 318, 577–579.

Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., and Sullivan, K. (1994). Identification of the remains of the Romanov family by DNA analysis. *Nat. Genet.* 6, 130–135.

Gill, P., Brenner, C., Brinkman, B., Budowle, B., Carracedo, A., Jobling, M.A., de Knijff, P., Kayser, M., Krawczak, M., Mayr, W.R., Morling, N., Olaisen, B., Pascali, V., Prinz, M., Roewer, L., Schneider, P.M., Sajantila, A., and Tyler-Smith, C. (2000). DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y chromosome STRs. *Int. J. Legal Med.* 114, 305–309.

Ginther, C., Issel-Tarver, L., and King, M.C. (1992). Identifying individuals by sequencing mitochondrial DNA from teeth. *Nat. Genet.* 2, 135–138.

Giusti, A.M., Baird, M., Pasquale, S., Balasz, I., and Glassberg, G. (1986). Application of deoxyribonucleic acid (DNA) polymorphisms to the analysis of DNA recovered from sperm. *J. Forensic. Sci.* 31, 409–417.

Giusti, A.M., and Budowle, B. (1995). Chemiluminescence-based detection system for human DNA quantitation and restriction fragment length polymorphism (RFLP) analysis. *Appl. Theor. Electrophor.* 5, 89–98.

Greenberg, B.D., Newbold, J.E., and Sugino, A. (1983). Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* 21, 33–49.

Grimberg, J., Nawoschik, S., Belluscio, L., McKee, R., Turck, A., and Eisenberg, A. (1989). A simple and efficient nonorganic procedure for the isolation of genomic DNA from blood. *Nucleic Acids Res.* 17, 8390.

Gusmão, L., Alves, C., and Amorium, A. (2001). Molecular characterisation of four human Y-specific microsatellites (DYS434, DYS437, DYS438, DYS439) for population and forensic studies. *Ann. Hum. Genet.* 65, 285–291.

Holland, M.M., Fisher, D.L., Mitchell, L.G., Rodriguez, W.C., Canik, J.J., Merril, C.R., and Weedn, V.W. (1993). Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. *J. Forensic Sci.* 38, 542–553.

Holt, C.L., Buoncristiani, M., Wallin, J.M., Nguyen, T., Lazaruk, K.D., and Walsh, P.S. (2002). TWGDAM validation of AmpFlSTR PCR amplification kits for forensic DNA casework. *J. Forensic Sci.* 47, 66–96.

Honda, K., Roewer, L., and de Knijff, P. (1999). Male DNA typing from 25-year-old vaginal swabs using Y chromosomal STR polymorphisms in retrial request case. *J. Forensic Sci.* 44, 868–872.

Iudica, C.A., Whitten, W.M., and Williams, N.H. (2001). Small bones from dried mammal museum specimens as a reliable source of DNA. *BioTechniques* 30, 732–734.

Ivanov, P.L., Wadhams, M.J., Roby, R.K., Holland, M.M., Weedn, V.W., and Parsons, T.J. (1996). Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat. Genet.* 12, 417–420.

Jeffreys, A.J., Wilson, V., and Thein, S.L. (1985a). Hyper variable minisatellite regions in human DNA. *Nature* 314, 67–73.

Jeffreys, A.J., Wilson, V., and Thein, S.L. (1985b). Individual specific fingerprints of human DNA. *Nature* 316, 76–79.

Jobling, M.A., Pandya, A., and Tayler-Smith, C. (1997). The Y chromosome in forensic and paternity testing. *Int. J. Legal Med.* 110, 118–124.

Kasai, K., Nakamura, Y., and White, R. (1990). Amplification of a variable number of tandem repeat (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science. *J. Forensic Sci.* 35, 1196–1200.

Kayser, M., de Knijff, P., Dieltjes, P., Krawczak, M., Nagy, M., Zerjal, T., Pandya, A., Tyler-Smith, C., and Roewer, L. (1997a). Applications of microsatellite-based Y chromosome haplotyping. *Electrophoresis* 18, 1602–1607.

Kayser, M., Cagliá, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J., Weichhold, G., de Knijff, P., and Roewer, L. (1997b). Evaluation of Y chromosomal STRs: a multicenter study. *Int. J. Legal Med.* 110, 125–133.

Kayser, M., Kruger, C., Nagy, M., Geserick, G., and Roewer, L. (1998). Y-chromosomal experiences and recommendations. In *Progress in Forensic Genetics 7* (Olaisen, B., Brinkmann, B., and Lincoln, P.J., eds), pp. 494–496. Elsevier Science, Amsterdam.

Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Kruger, C., Krawczak, M., Nagy, M., Dobosz, T., Szibor, R., de Knijff, P., Stoneking, M., and Sajantila, A. (2000). Characterization and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.* 66, 1580–1588.

Kayser, M., Brauer, S., Willuweit, S., Schadlich, H., Batzer, M., Zawacki, J., Prinz, M., Roewer, L., and Stoneking, M. (2002). Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for US populations. *J. Forensic Sci.* 47, 513–519.

Kayser, M. (2003). The Human Y-chromosome – introduction into genetics and applications. *Forensic Sci. Rev.* 15, 77–89.

Kayser, M., Brauer, S., Schadlich, H., Prinz, M., Batzer, M., Zimmerman, P.A., Boatin, B.A., and Stoneking, M. (2003). Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic Ancestry. *Genome Res.* 13, 624–634.

Krenke, B.E., Tereba, A., Anderson, S.J., Buel, E., Culhane, S., Finis, C.J., Tomsey, C.S., Zachetti, J.M., Masibay, A., Rabbach, D.R., Amiott, E.A., and Sprecher, C.J. (2002). Validation of a 16-locus fluorescent multiplex system. *J. Forensic Sci.* 47, 773–785.

Krenke, B. (2004). The PowerPlex® Y system. *Profiles in DNA* 6, 6–9.

Krenke, B., Viculis, L., Richard, M.L., Prinz, M., Milne, S.C., Ladd, C., Gross, A.M., Gornall, T., Frappier, R., Eisenberg, A., Barna, C., Aranda, X., Adamowicz, M.S., and Budowle, B. (2005). Validation of a male-specific, 12-locus fluorescent short tandem repeat (STR) multiplex. *Forensic Sci. Int.* 148, 1–14.

Lell, J.T., Sukernik, R.I., Starikovskaya, Y.B., Su, B., Jin, L., Schurr, T.G., Underhill, P.A., and Wallace, D.C. (2002). The dual origin and Siberian affinities of Native American Y chromosomes. *Am. J. Hum. Genet.* 70, 192–206.

Levinson, G., and Gutman, G.A. (1987). Slipped strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4, 203–221.

Lins, A.M., Micka, K.A., Sprecher, C.J., Taylor, J.A., Bacher, J.W., Rabbach, D.R., Bever, R.A., Creacy, S.D., and Schumm, J.W. (1998). Development and population study of an eight locus short tandem repeat (STR) multiplex system. *J. Forensic Sci.* 43, 1168–1180.

Magnuson, V.L., Ally, P.S., Nylund, S.J., Karanjuwala, Z.E., Rayman, J.B., Knapp, J.I., Lowe, A.L., Ghosh, S., and Collins, F.S. (1996). Substrate nucleotide-determined non-templated addition of adenine by Taq DNA polymerase: implications for PCR-based genotyping. *BioTechniques* 21, 700–709.

Micka, K.A., Amiott, E.A., Hockenberry, T.L., Sprecher, C.J., Lins, A.M., Rabbach, D.R., Taylor, J.A., Bacher, J.W., Glidewell, D.E., Gibson, S.D., Crouse, C.A., and Schumm, J.W. (1999). TWGDAM validation of a nine-locus and a four-locus fluorescent STR multiplex system. *J. Forensic Sci.* 44, 1243–1257.

Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A., Keys, K.M., and Budowle, B. (2001). Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *J. Forensic Sci.* 46, 647–660.

Mountain, J.L., Knight, A., Jobin, M., Gignoux, C., Miller, A., Lin, A.A., and Underhill, P.A. (2002). SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Res.* 12, 1766–1772.

Murch, R., and Budowle, B. (1986). Applications of isoelectric focusing in forensic serology. *J. Forensic Sci.* 31, 869–880.

National Research Council (1996). *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington, DC.

Prinz, M., Boll, K., Baum, H., and Shaler, B. (1997). Multiplexing of Y chromosome specific STRs and performance of mixed samples. *Forensic Sci. Int.* 85, 209–218.

Quintans, B., Beleza, S., Brion, M., Sanchez-Diz, P., Lareu, M., and Carracedo, A. (2003). Population data of Galicia (NW Spain) on the new Y-STRs DYS437, DYS438, DYS439, GATA A10, GATAA7.1, GATAA7.2, GATAC4 and GATAH4. *Forensic Sci. Int.* 131, 220–224.

Roewer, L., Kayser, M., Dieltjes, P., Nagy, M., Bakker, E., Krawczak, M., and de Knijff, P. (1996). Analysis of molecular variance (AMOVA) of Y chromosome specific microsatellites in two closely related human populations. *Hum. Mol. Genet.* 7, 1029–1033.

Roewer, L., Kayser, M., de Knijff, P., Anslinger, K., Cagliá, A., Corach, D., Füredi, S., Henke, L., Hidding, M., Kärgel, H.J., Lessig, R., Nagy, M., Pascali, V.L., Parson, W., Rolf, B., Schmitt, C., Szibor, R., Teifel-Greding, J., and Krawczak, M. (2000). A new method for the evaluation of matches in nonrecombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. *Forensic Sci. Int.* 114, 31–43.

Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C., Amorim, A., Anslinger, K., Augustin, C., Betz, A., Bosch, E., Cagliá, A., Carracedo, A., Corach, D., Dekairelle, A.F., Dobosz, T., Dupuy, B.M., Füredi, S., Gehrig, C., Gusmaõ, L., Henke, J., Henke, L., Hidding, M., Hohoff, C., Hoste, B., Jobling, M.A., Kärgel, H.J., de Knijff, P., Lessig, R., Liebeherr, E., Lorente, M., Martínez-Jarreta, B., Nievas, P., Nowak, M., Parson, W., Pascali, V.L., Penacino, G., Polski, R., Rolf, B., Sala, A., Schmidt, U., Schmitt, C., Schneider, P.M., Szibor, R.,

Teifel-Greding, J., and Kayser, M. (2001). Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci. Int.* 118, 106–113.

Saferstein, R., ed. (1995). *Criminalistics – An Introduction to Forensic Science*, 5th edition. Prentice Hall Education, New Jersey.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia. *Science* 230, 1350–1354.

Saiki, R.K., Walsh, P.S., Levenson, C.H., and Erlich, H.A. (1989). Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86, 6230–6234.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edition. Cold-Spring Harbor Laboratory, Cold-Spring Harbor Laboratory Press, New York.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5468.

Scherczinger, C.A., Bourke, M.T., Ladd, C., and Lee, H.C. (1997). DNA extraction from liquid blood using QIAamp. *J. Forensic Sci.* 42, 893–896.

Schlotterer, C., and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20, 211–215.

Shedlock, A.M., Haygood, M.G., Pietsch, T.W., and Bentzen, P. (1997). Enhanced DNA extraction and PCR amplification of mitochondrial genes from formalin-fixed museum specimens. *BioTechniques* 22, 394–400.

Shewale, J.G., Nasir, H., Schneida, E., Gross, A.M., Budowle, B., and Sinha, S.K. (2004). Y-chromosome STR system, YPLEXTM 12, for forensic casework: development and validation. *J. Forensic Sci.* 49, 1278–1290.

Sinclair, K., and McKechnie, V.M. (2000). DNA extraction from stamps and envelope flaps using QIAamp and QIAshredder. *J. Forensic Sci.* 45, 229–230.

Singer-Sam, J., Tanguay, R.L., and Riggs, A. (1989). Use of Chelex to improve the PCR signal from a small number of cells. *Amplifications: A Forum for PCR Users* 3, 11.

Sinha, S.K., Budowle, B., Arcot, S.S., Richey, S.L., Chakraborty, R., Jones, M.D., Wojtkiewicz, P.W., Schoenbauer, D.A., Gross, A.M., Sinha, S.K., and Shewale, J.G. (2003a). Development and validation of a multiplexed Y-chromosome STR genotyping system, Y-PLEXTM6, for forensic casework. *J. Forensic Sci.* 48, 93–103.

Sinha, S.K., Nasir, H., Gross, A.M., Budowle, B., and Shewale, J.G. (2003b). Development and validation of the Y-PLEXTM5, a Y-chromosome STR genotyping system for forensic casework. *J. Forensic Sci.* 48, 985–1000.

Smith, J.R., Carpten, J.D., Brownstein, M.J., Ghosh, S., Magnuson, V.L., Gilbert, D.A., Trent, J.M., and Collins, F.S. (1995). Approach to genotyping errors caused by nontemplated nucleotide addition by Taq DNA polymerase. *Genome Res.* 5, 312–317.

Sullivan, K.M., Hopgood, R., and Gill, P. (1991a). Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *Int. J. Legal Med.* 105, 83–86.

Sullivan, K.M., Hopgood, R., Lang, B., and Gill, P. (1991b). Automated amplification and sequencing of human mitochondrial DNA. *Electrophoresis* 12, 17–21.

Sullivan, K.M., Hopgood, R., and Gill, P. (1992). Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *Int. J. Legal Med.* 105, 83–86.

Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazon Lahr, M., Foley, R.A., Oefner, P.J., and Cavalli-Sforza, L.L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 65, 43–62.

Walsh, S., Metzger, D.A., and Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* 10, 506–513.

Walsh, P.S., Varlaro, J., and Reynolds, R. (1992). A rapid chemiluminescent method for quantitation of human DNA. *Nucleic Acids Res.* 20, 5061–5065.

Waye, J., Presley, L., Budowle, B., Shutler, G.G., and Fourney, R.M. (1989). A simple and sensitive method for quantifying human genomic DNA in forensic specimen extracts. *Biotechniques* 7, 852–855.

Wells, R.S., Yuldasheva, N., Ruzibakiev, R., Underhill, P.A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., Balakrishnan, K., Read, M., Pearson, N.M., Zerjal, T., Webster, M.T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zalloua, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A., and Bodmer, W.F. (2001). The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* 98, 10244–10249.

White, P., Tatum, O., Deaven, L., and Longmire, J. (1999). New, male-specific microsatellite markers from the Y-chromosome. *Genomics* 57, 433–437.

Wilson, M.R., DiZinno, J.A., Polanskey, D., Replogle, J., and Budowle, B. (1995a). Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int. J. Legal Med.* 108, 68–74.

Wilson, M.R., Polanskey, D., Butler, J., DiZinno, J.A., Replogle, J., and Budowle, B. (1995b). Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. *Biotechniques* 18, 662–669.

Wyman, A.R., and White, R. (1980). A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA* 77, 6754–6758.

Chapter 27

Mass Disaster Victim Identification Assisted by DNA Typing

Daniel Corach

Servicio de Huellas Digitales Genéticas y Cátedra de Genética y Biología Molecular, School of Pharmacy and Biochemistry, University of Buenos Aires, Argentina

27.1 INTRODUCTION

At the beginning of the twentieth century, a relevant attribute of blood was detected by Karl Landsteiner (1901). His observations allowed the explanation of the underlying basis of "transfusion disease". Precipitin detection shed light on the antigenic effect of blood transfusion disorders. Accordingly, Landsteiner's contribution explained an induced immunological impairment and identified a monogenic polymorphism that was employed to complement kinship rejections.

The ABO blood group, coded in the DNA, became the first molecular tool to be used in human identification (see also Chapter 40). This discovery was followed by many others that included diverse serological polymorphic markers disclosed by Landsteiner and others (Owen, 2000). Over five decades elapsed between the detection of the first informative molecular monogenic marker and the establishment of the molecular structure of the double helix (Watson and Crick, 1953), and over 80 years for the detection of DNA polymorphisms (Wyman and White, 1980).

Therefore, for over 80 years, blood groups and other serological polymorphisms remained as the molecular diagnostic tools for inferring biological kinship. Besides the limitation imposed by their relatively reduced discriminatory power, these genetic markers required cell integrity and were not informative about the origin of the samples to be investigated – whether they came from evidentiary material or decomposed corpses or remains.

Soon after, a true revolution came about in forensic science with two groundbreaking findings: Alec Jeffreys's DNA fingerprinting (Jeffreys *et al.*, 1985) and Kary Mullis's development of the polymerase chain reaction (PCR; Mullis *et al.*, 1986). The major impact was brought forth in the field of human identification in cases involving living or deceased people, since even a certain degree of DNA degradation was likely to be neglected by the analysis of small polymorphic regions, termed short tandem repeats or STRs, widely distributed all around the genome (Weber and Mayr, 1989; Edwards *et al.*, 1991, 1992).

The availability of this new analytical approach efficiently contributed to the development of strategies that permitted the identification of decomposed corpses and fragmentary human remains, often "by-products" of mass disasters, by comparing their genetic profiles with those of their putative relatives. This chapter will focus on the characterization of victims and human remains from mass disasters by DNA-based methods.

27.2 CLASSIFICATION OF MASS FATALITIES AND DIVERSE SCENARIOS FOR HUMAN REMAINS RETRIEVAL

Hazards and mass disasters are extremely complex scenarios whose origin might be mainly determined by two major forces: natural and human-induced catastrophes. Some authors such as Wisner (2004) argue that all disasters are human induced, since human actions before the hazard can prevent their development into a mass disaster. Human failure in developing adequate disaster management preventive policies might result in human responsibility in most of them. Different kinds of hazards, which indicate the acting forces and their effect on the environment and human beings, are summarized in Table 27.1. It is important to keep in mind that some hazards are far from simple in terms of their effects, as they might produce a chain reaction with extremely severe effects. An example

TABLE 27.1 Summary of the different hazard types and their effect on the environment and humans.

Promoting force	Type of hazard	Description	Type of remains	DNA-based ID required
Natural	Geological	Earthquake	Buried/Decomposed	+/−
	Climatological	Volcanic eruption	Burnt	+/−
		Tsunami	Drowned/Decomposed	+
		Hurricanes	Drowned/Decomposed	+
		Landslide	Buried/Decomposed	+
	Biological	Epidemics	–	−
Human-induced	Technological	Environmental	–	−
		Engineering failures	Variable	+/−
		Transport accidents	Fragmentary/Decomposed	+
	Sociological	Terrorist attack	Fragmentary/Decomposed	+
		Wars	Skeletonized	+

of such a situation recently occurred in East Asia, in late December 2004, when an earthquake produced a tsunami and subsequent extensive flooding of flat coastal areas, with over 100,000 fatalities.

Many factors can produce these catastrophes and developing countries seem to be prone to their effects. In fact, in these countries human vulnerability is increased due to the lack of appropriate emergency management and planning. Accordingly, the unexpectedness of the event may intensify the adverse effects of the hazard, which may, in turn, result in extremely serious human, structural, and financial losses. However, generalizations are misleading. A clear example of such an event and its impact on an industrialized country is Hurricane Katrina which had a catastrophic effect on New Orleans in the USA, on August 29, 2005.

Both natural and human-induced fatalities might cause a variable number of victims. Nevertheless, natural disasters usually affect a great number of people and in most cases DNA-based identification (ID) is mandatory. Some of the most relevant circumstances surrounding natural disasters are: the time elapsed since the moment the hazard hits a vulnerable area, determining the disaster's occurrence, and the moment of retrieval of corpses and remains. In addition, the identification process might face serious restrictions such as the lack of surviving relatives that might otherwise have provided reference samples for comparison.

On the other hand, human-induced fatalities might also widely vary in terms of the number of resulting victims. In some terrorist attacks, these may range from a few (tens/hundreds), as in the case of suicide bombings, to a more considerable number (thousands), as in the event of the integrated terrorist attack that destroyed the World Trade Center in New York City on September 11, 2001. The retrieval of remains is a time-consuming process, strongly influenced by the magnitude of the disaster and the degree of environmental modifications produced by the hazard. Accordingly, corpses/fragmentary remains collection is usually inversely proportional to the dimension of the case, the number of victims, and the degree of environmental destruction. Hence, retrieval, cataloging, and storage represent the slowest steps in the analysis. Instead, DNA typing is considerably faster. DNA extractions can be carried out in a few hours with the help of automated platforms that may ensure high throughput of molecular typing results. In most cases, living relatives can provide reference samples for comparison. Nevertheless, in special cases where no relatives are available, victims' personal belongings are employed. In those terrorist attack cases where explosives are used, body fragmentation is usually produced. Under these conditions, it is important to analyze every tissue fragment in order to reconstruct the body for further comparison with reference samples. In some cases, this strategy may provide a clue concerning the number of victims resulting from the explosion and, potentially, to identify the suicide terrorist, if that is the case.

In some special cases, such as those produced by transport accidents in which the number and identity of the passengers are unknown but not all of them deceased, DNA-based ID is also mandatory.

27.3 CONVENTIONAL IDENTIFICATION CRITERIA ROUTINELY USED FOR HUMAN IDENTIFICATION

The ability to recognize and identify human individuals represents a complex cognitive process. Although identification can be considered an automatic recognizing reflex experienced either by babies during their early months of life or even by non-human mammals such as pets, its underlying mechanisms are far from simple. The basic process requires an information storage device, a sensorial system that might allow the information to be captured and stored and an adequate program that could establish coincidences between the previously stored information and the incoming data detected at a time after the first capture. We daily practice this process unconsciously. This is true for the immediate recognition of a person from his/her surrounding environment. This process is mostly subjective and is affected by memory loss or alterations in the cognitive abilities.

Beyond the individual process, personal identification requires objective criteria to be used for developing robust and reliable identification systems that can contribute to personal identification at a nationwide environment. What kind of traits can provide useful information for an unbiased identification? Basically, they should be highly polymorphic and remain unchanged throughout life. Fingerprints are the only morphological attributes that may fulfill these requirements. Their use in identification can be traced back to the end of the nineteenth century when Francis Galton (1892) described the identification potential of finger ridges or dermatogliphs.

During the twentieth century, the creation of dactyloscopic databases, which currently constitute part of the ID documentation systems in most countries, was spread worldwide. Their power of identification is remarkable since they can distinguish monozygous twins, a task that fails if carried out by DNA analysis. Other criteria include the signature, since calligraphic traces respond to an individual-specific pattern. Their analysis may allow correlating a piece of writing with a signature reference. Moreover, physiognomic traits are included as ID tools. However, it is evident that these features widely vary along a person's life. Nevertheless, in a given life span they might provide useful ID information.

The above-mentioned criteria can be put together and as a result, countrywide databases created. These criteria represent the conventional identification approaches stored in "institutional memories" and depicted in our ID cards and passports, depending on the countries' laws. These ID systems focus on establishing the identity of a given individual by comparing an evidence, constituted by a person's photographic image or a piece of writing or even a latent fingerprint, and its corresponding reference, stored in a database. These criteria, which might be efficient for identifying an individual, are not suited for establishing biological relatedness.

In some special cases, such as mass disasters, aircraft crashes, terrorist bombings, mass floods, or earth slides, in addition to the dactyloscopic and photographic data available from databases, other information that can be provided by living relatives seeking missing family members might be relevant – for example: personal belongings, ante mortem information including dental records, radiological studies revealing personal features, and detailed description of personal marks, such as tattoos, piercings, or scars.

The systematic use of these kinds of ID parameters may be useful but should be used with maximum care, since the emotional pressure of the relatives might lead to a biased recognition that could generate serious conflicts if the misidentification is detected at a later stage.

Currently, in addition to the so-called conventional criteria, mass disaster victim identification has been highly improved by the DNA-based identification approach. The first terrorist attack in which a wide set of DNA markers was used occurred in Argentina in 1992 (Corach *et al.*, 1994), when the Israeli Embassy was hit by an explosive charge supposed to be driven in a light truck by a suicide terrorist, whose remains were never found. The inclusion of this new ID approach required the development of remains handling strategies in order to ensure the optimization of DNA retrieval from samples, when possible.

27.4 CRITERIA FOR PRESERVATION OF REMAINS

The irruption of DNA-based ID methodology in the forensic field brought about dramatic changes concerning sample preservation. Since its first use, both RFLP and PCR-based approaches have been applied to catastrophe-emerging remains. The first kind of markers mentioned above required preventing sample degradation. This is because polymorphic changes are detected within a wide range of DNA fragment sizes, strongly influenced by the single locus probes used, and the restriction enzyme employed for DNA cleavage (mostly HaeIII with DNA fragment sizes ranging from 1 to10 kb). The second set of markers, known as short tandem repeats (STRs), were introduced by Edwards and coworkers in 1991 and are widely spread in the forensic field at present. These markers can be informative even with partially degraded DNA samples, but are highly sensitive to inter-individual contamination. Similar constraints are also valid for mitochondrial DNA sequencing of hypervariable regions I and II (HVRI and II) (Anderson *et al.*, 1981), also used in these cases. On the other hand, interspecies contamination has no effect on PCR-driven polymorphism detection, since species-specificity is provided by the primer sequence used and is mostly human specific, either for STRs or mitochondrial (mt) DNA HVR-I and II amplifications.

In close relation to the described characteristics of the molecular approaches used for disclosing polymorphic

attributes of the remains under study, practical procedures have been developed in order to optimize the material for preservation prior to analysis.

Mass fatalities that occurred in Argentina were investigated by our group. During the Embassy case, the collection of remains was far from ideal, no previous experience had been available, and therefore involuntary errors were committed. Inadequate handling of fragmentary body parts and long exposure to room temperature had led to a certain degree of sample degradation and in some cases to inter-individual contamination. These observations became apparent after DNA typing. The lesson we learnt from this dramatic experience led us to develop guidelines aimed at sample preservation procedures, whose implementation highly improved the results of other tragedies that were subsequently investigated.

These preservation procedures included the recovery of fragmentary remains in individual plastic bags, and abundant tap and sterile water washing and rinsing, previous to tissue sample collection. The tissues, mostly muscles, were placed in labeled 50 ml sterile plastic tubes, and kept at −20°C till the moment of DNA extraction. These simple and basic rules allowed for the improved recovery of DNA samples obtained from emerging victims' generally highly fragmented human remains.

The second mass fatality occurred two years later in Buenos Aires City, with more severe effects than in the previous case and producing a higher number of victims (Corach *et al.*, 1995, 1996a). A terrorist attack with explosives completely destroyed a building and buried some of the victims. Ten months after this fatality, when the building debris was removed completely, dehydrated muscle tissues were recovered. DNA was efficiently extracted and its quality was found suitable either for STR typing or for mtDNA sequencing. On the basis of the results obtained from the dehydrated soft tissue samples, a set of experiments was performed in order to develop a procedure that may warrant soft tissue preservation at room temperature. Since then, it has become our procedure of choice for fragmentary tissue preservation (Corach *et al.*, 1996b). At the moment of sample collection and for molecular identification purposes, the importance of complete sampling of all fragmentary remains must be stressed, even when some traits may lead to their visual identification. This practice will efficiently contribute to determine the number of victims and to reduce the number of unidentified remains at the morgue.

27.5 DNA POLYMORPHISMS USED FOR TRACING KINSHIP BETWEEN FRAGMENTARY HUMAN REMAINS AND THEIR DEMANDING RELATIVES

In the mid-1980s, two scientific breakthroughs revolutionized the existing human identification approaches. On the one hand, Alec Jeffrey's discovery of the hypervariable DNA mini-satellites became the landmark that set the bases for DNA identification technology (Jeffreys *et al.*, 1985). On the other hand, Kary Mullis's development of the PCR (Mullis *et al.*, 1986) permitted to broaden Jeffrey's contribution by optimizing it for forensic identification purposes.

During a short period of time between 1985 and 1992, a variety of diverse polymorphisms were identified, validated, and used in forensic casework investigation. Restriction fragment length polymorphism (RFLP) analysis allowed detection of multilocus (Jeffreys *et al.*, 1985) and single locus (Nakamura *et al.*, 1987) mini-satellites. PCR-based approaches included HLA DQ-A1 (Erlich and Bugawan, 1989) and Polymarker (Herrin *et al.*, 1994) allele-specific oligotyping detected by reverse-dot-blot hybridization, STR typing (Edwards *et al.*, 1991), and mtDNA sequence analysis (Ginther *et al.*, 1992). Additionally, the inclusion of the markers harbored in the Y-chromosome efficiently contributed to gender determination, which represented a relevant contribution to the identification of fragmentary remains.

With the exception of RFLP-based mini-satellite DNA typing and the DQ-A1/Polymarker systems, all other molecular approaches are currently in use with relevant improvement, such as the automation of analytical platforms, either for sequence or fragment analysis. Additional markers were included for investigating the fatalities of one of the most sadly conspicuous terrorist attacks: the explosions perpetrated on the World Trade Center on September 11, 2001 in New York City. The inclusion of single nucleotide polymorphisms (SNP) and the use of "mini" STRs highly improved the ability to retrieve informative genetic profiles and genotypes from severely degraded samples.

27.6 CHALLENGES CONCERNING DNA DEGRADATION AND CONTAMINATION

Between the occurrence of a given disaster and the collection of remains, a highly variable period of time might elapse. Accordingly, the decomposition of corpses and human body fragments characterizes the materials to be analyzed. In addition, the decaying process is strongly influenced by environmental conditions, such as temperature, humidity, soil, surrounding debris, etc. Decomposing is associated with macromolecular degradation, mostly hydrolysis and deamination of proteins and reduction of the molecular size of DNA fragments. The initially developed genetic polymorphic markers for human ID, although highly informative, proved to be of limited use when degraded samples needed to be analyzed. DNA integrity is also required for the evaluation of these markers, whose average fragment size should be as high as 10 kb. Furthermore, considerable amounts of genetic material are needed, due to the requirements imposed by the Southern blot hybridization approach. Efficient detection of

hybridization signals needs at least 50 nanograms of DNA. Another limitation, determined by the nature of multilocus probes (MLPs) (Jeffreys *et al.*, 1985) that are able to uncover highly polymorphic DNA mini-satellites, is related to its specificity. These genome-wide distributed polymorphisms are present in almost every organism, including animals, plants, and fungi. Hence, the lack of human specificity may obscure the hybridization patterns since heterologous contaminant DNA may produce overlapping multi-band results. The use of MLPs was abandoned a few years after their discovery. However, their use as probes for screening human genome libraries permitted the detection and isolation of a wide variety of human-specific clones, many of which denoted variable degrees of polymorphisms (Nakamura *et al.*, 1987). These single locus – SLP – markers were used as efficient probes that were sensitive to DNA integrity and quantity, but unaffected by heterologous contamination. Their use in forensic ID was widespread and some cases of mass fatalities were investigated with the help of these markers (Ludes *et al.*, 1994; Corach *et al.*, 1994). SLPs were also used to confirm STR results obtained after the analysis of fragmentary remains.

Perhaps some of the most informative genetic markers ever developed for human ID were derived from SLP markers, in particular the MS31s, on which the repeat units denote detectable variation. These markers were termed Mini-satellite Variant Repeats (MVR; Jeffreys *et al.*, 1990). The combination of long length PCR using primer specific reactions, followed by Southern blotting and hybridization with a specific SLP, allowed highly informative genetic "bar codes" to be obtained. Their use in highly degraded samples was extremely efficient, combining SLP informativeness and PCR sensitivity. Their use permitted the efficient identification of remains emerging from the terrorist attack on a Jewish Mutual Association (AMIA) in Buenos Aires (Corach *et al.*, 1996a). Unfortunately, these markers were not included in the routine casework marker battery.

Another efficient approach successfully employed for obtaining valuable ID information from decomposed or skeletonized remains was provided by mtDNA sequence analysis. The most relevant attributes of this genetic marker reside in its size, structure, and multiplicity of copies present per cell. The small size of the mitochondrial genome (16,569 bp long) and its circular structure make it resistant to physical rupture and exonuclease degradation, and the high copy number (with an average of five copies per organelle and over 1,000 organelles/cells depending on their metabolic condition) increases the possibility of DNA recovery. The inheritance mechanism of the mitochondria limits their informative potential. Maternal transmission of mtDNA only permits the matrilineage tracing of a given individual (see also Chapter 26). The control region or D-loop harbors a variety of polymorphic base substitutions and insertion/deletion polymorphisms within this non-coding region of about 1.2 kb. Three regions are recognized within it, denoted as HVR-I (16024–16365), HVR-II (73–349), and HVRIII (438–574). In some cases in which highly degraded material is analyzed, "mini-sequencing" strategies can be used in order to improve the results (Gabriel *et al.*, 2001). This strategy is based on the amplification of small amplicons included in the hypervariable regions for further sequencing. The amplification of overlapping fragments can "reconstruct" badly degraded DNA fragments, since smaller fragments are more abundant in degraded samples. This approach has been successfully used in analyzing skeletonized human remains after long burial time periods (Corach *et al.*, 1997).

As previously mentioned, forensic ID capabilities were boosted by the irruption of STR technology, which provided a powerful tool for routine use in establishing matching between suspects and crime scene evidence and investigating biological kinship relationships. Their main feature was to allow the typing of degraded samples containing minute DNA amounts. Moreover, the genome-wide distribution of the STRs offers a chance to diversify their application field. The first set of STRs that appeared to be effective for forensic application was located onto the autosomal and X-chromosomes (Edwards *et al.*, 1991) and soon after, those located onto the Y-chromosome were described (Roewer *et al.*, 1992). Autosomal STRs are fundamental for forensic evidence and individual identification, are highly informative, can undergo recombination, and are inherited from both parents. Instead, those located onto the sex chromosomes have clearly different attributes. Those harbored onto the X-chromosome behave as autosomal markers in females with normal recombination whereas in males they fail to recombine with their counterpart, the Y-chromosome. Accordingly, the Y-chromosome is transmitted unrecombined from fathers to sons, as a block of genetic information, and the sole variation they can incorporate is the product of mutation events. This characteristic restricts their ID potential to patrilineage level but provides a sensitive tool for detecting traces of male-specific material such as semen in rape cases and in the analysis of fragmentary remains. On the other hand, the X-STRs have proved to be a relevant application in father–daughter cases (Schmidtke *et al.*, 2004). The combined use of Y- and X-STRs has efficiently sped up the ID tasks performed in mass disaster case identification (Corach *et al.*, 1995, 1997).

Another additional improvement was introduced with the development of the "mini-STR". In these systems, the same loci are used. However, new primer sites were defined much closer to the variable region in which repeat units are tandemly organized. Accordingly, the amplification products are much more reduced in size, 100 bp shorter than their regular STR counterparts (Wiegand and Kleiber, 2001).

The most recent inclusion within the ID polymorphic systems has been SNP profiling. Since the number of alleles per system is restricted to two (bi-allele markers), the use of a complete set of SNPs may provide acceptable

discrimination and exclusion powers (Gill, 2001). Since polymorphism is restricted to a single nucleotide, it offers the possibility of analyzing highly degraded DNA samples. Its usefulness was demonstrated in the September 11 terrorist attacks (Vastag, 2002).

27.7 CRITERIA EVOLUTION AND TECHNICAL APPROACHES APPLIED TO DNA-BASED VICTIM IDENTIFICATION IN MASS DISASTERS FROM THE EARLY 1990s TO DATE

As in many other fields of science, the continuous improvement in analytical processes, programing techniques, and software development have allowed the speeding up of data acquisition and processing. In addition, software design has contributed to optimize the data gathering systems included in automated sequencers and can deal with extremely high numbers of genotypes for rapid comparison/matching purposes. This adequately led to the establishment of the National Database in Great Britain in 1995 and the Combined DNA Index System (CODIS) in the USA in 1997. Besides their relevant use in forensic caseworks, the latter proved to be of extraordinary help for the identification of victims emerging from the September 11 terrorist attacks (DNA President's initiative, 2004; http://www.dna.gov). However, many previous cases were investigated and positive identification resulted from these analyses. The successful use of manual platforms was attained, despite being more laborious and time consuming. Those labor-intensive approaches provided highly reproducible and robust results but their efficiency for handling massive sample sets was limited.

In the early 1990s, the Southern blot hybridization approach allowed SLPs typing, but few polymorphisms that could be detected by PCR-based methods were available. Most of these were at a status of experimental development; however, in some special cases their use was widely justified. Additionally, genotype comparison, originally carried out by simple counting (Corach *et al.*, 1995), was later replaced by simple programs that sped up the analytical process. This new approach made it possible to find samples that depicted genotype identities leading to body reconstruction or to establish possible kinship relationships between reconstructed bodies and putative relatives by comparing genotypes of remains/corpses and reference samples. Experimental strategies clearly varied due to technical feasibilities available at the time of each investigated case.

27.8 DESCRIPTION OF ANALYZED CASES

Different mass disaster situations are exemplified by five cases in which our laboratory was involved. These five mass disaster cases can be classified as human-induced disasters and their genesis can be defined as: (a) two open road accidents, one involving an airplane crash and the other two buses that crashed and caught fire and (b) three terrorist attack cases, two of them probably with religious motifs and one mass murder case conducted by state-terrorism during the last military dictatorship in Argentina. A description of these cases is discussed in the next paragraphs.

27.8.1 Terrorist Attack Cases

27.8.1.1 Terrorist Attack on the Embassy of Israel in Buenos Aires, Argentina

On March 17, 1992, a terrorist bomb completely destroyed the Israeli Embassy in Buenos Aires. Approximately 300 people were wounded and the number of fatalities was estimated to be 28, among them four Israeli members of the embassy staff, four local embassy employees, elderly residents at a nursing home, and school children on a passing bus. This attack marked the introduction of the Middle East (http://www.jewishvirtuallibrary.org/jsource/Terrorism/terrortoc.html) into South America and was the first mass disaster whose emerging remains were investigated by means of DNA-based technology in its early development stages.

Ten human remains were analyzed some time after the explosion. The material was partially putrid and intense bacterial contamination was denoted by the presence of 16S and 23S bacterial rRNA overimposed on some sample lanes. The analyses were carried out by using single locus mini-satellite YNH-24 (Nakamura *et al.*, 1987), six STRs, two of them located on the X-chromosome (HUMARA and HUMHPRTB) and four on autosomal chromosomes (HUMTHO-1, FABP, RENA-4 and CD-4). Gender on the fragmentary remains was determined by means of amplifying Yq13-ter sequences included within the heterochromatin region of the Y-chromosome, described by Kogan and coworkers (1987). MtDNA sequencing of HVR-I and II was carried out by asymmetrical amplification.

Most samples were completely typed and the data compared with putative relatives; no missing victims were officially recognized. This was the first DNA requested in Argentina by the Supreme Court of Justice in a criminal case.

27.8.1.2 Bombing to the Argentine Israeli Mutual Association (AMIA) in 1994

On July 18, 1994, the building of the Argentine Israeli Mutual Association was the target of a terrorist bomb attack in the city of Buenos Aires. In this attack, 85 people were killed whereas hundreds were injured. It was Argentina's deadliest terrorist bombing aggression. The overall investigation required the analysis of over 350 samples, including human remains and reference samples. A rapid molecular

screening was performed by employing five autosomal STRs (HUMTHO-1, FES/FPS, FABP, vWA, HUMRENA4, and two sex chromosome STRs: HUMHPRTB on the X-chromosome and Y27H39 on the Y-chromosome, also known as DYS19). STR profiling results were confirmed by SLP typing in those cases in which molecular integrity allowed the retrieval of quality-suited DNA for the analysis. Four mini-satellite probes (YNH-24, PH-30, MS1, and LH-1) were used in subsequent hybridization/dehybridization cycles employing a single Southern blotted membrane for each set of samples. In those cases with highly degraded DNA, confirmation was attained by the PCR-MVR approach (Jeffreys *et al.*, 1990) or mtDNA HVR-I and II sequencing.

27.8.1.3 Mass Murder Case Conducted by State Terrorism during the Last Military Dictatorship in Argentina

Argentina was ruled by military forces during the period between 1976 and 1983. During this time, the bloodiest repression ever was perpetrated by state terrorism. Over 30,000 people are supposed to have been killed or missing, of which over 11,000 disappearances have been well documented. Diverse kidnapping/torture/killing modalities were used, which resulted in some victims whose corpses were never retrieved. However, anonymous skeletal remains were discovered in public cemeteries, known as "NN burial sites". In order to investigate the efficiency of the molecular typing approaches, sets of bones and teeth were analyzed with several typing approaches including the nested-PCR reaction using external primer sequences to some STRs, mainly HUMTHO-1 and HUMFABP, and mtDNA sequencing. Although the first approach produced reproducible results in some samples, mtDNA sequencing proved to be successful in all samples tested. Unfortunately, at that time no sequence or genetic profile databases were available for comparison leading to missing person identification (Corach *et al.*, 1997).

27.8.2 Open Road Accidents

27.8.2.1 Bus Crash Accident

In January 9, 1993, two buses crashed and one of them caught fire. The accident produced 56 fatal victims and a total of 80 injured passengers. The accident occurred on national route 14 near Santo Tome in the Province of Corrientes, Argentina. Two years later, several people were still missing, and had probably been victims of the multiple bus crash. A DNA investigation was requested, for which a combination of genetic markers including autosomal short tandem repeats (STRs), Y-chromosome STRs (Y-STRs), and mtDNA HVR-I and II were characterized. A set of 25 human fragmentary remains including badly burned bones and teeth were analyzed, together with the reference samples obtained from the demanding relatives. Taking into account the quality of the analyzed remains, the DNA extraction and typing proved successful in 65% of cases (16 out of 25), including the child of a demanding couple. Complete genotypes were obtained with four autosomal, one X-chromosome STR, and nine Y-STRs. The analytical strategy was focused on Y-STR haplotype detection of the putative son with the haplotypes obtained from the fragmentary remains. A premolar revealed the same haplotype and also shared at least one autosomal allele with both descendants and with the putative mother. HVR-I and II confirmed the matrilineage origin since the putative mother presented an identical mitochondrial haplotype.

27.8.2.2 Open Road Airplane Crash

On August 31, 1999, LAPA Flight 3142 from Buenos Aires to the capital city of Cordoba, Argentina, crashed near Jorge Newbery airport, shortly after take-off. The crash resulted in 65 fatalities, 40 survivors and 17 others who were seriously injured, making it one of the worst accidents in the history of Argentine aviation. Due to human errors and technical failure, the aircraft retracted flaps during take-off and even though the aircraft had achieved minimum take-off speed it continued beyond the ramp, broke through the airport's fence, crossed a road, dragged with it an automobile and finally collided with road-construction machinery. Gasoline spilling over the hot engines caused incineration and total destruction of the aircraft.

A total of 38 human remains were received for DNA typing and the members of 13 family groups provided reference samples. DNA was extracted from corpse material in triplicate and analyzed blindly also in triplicate. The analytical strategy focused on autosomal STR typing and gender determination by using amelogenin. All detection analyses were carried out with automated sequencers and 13 autosomal STRs were typed with commercial kits, while Y-STRs were analyzed with custom-made multiplexes using the minimal haplotype marker system (Kayser *et al.*, 1997). After identifying male samples by amelogenin gender-associated dimorphism, patrilineage tracking was conducted by Y-STR haplotyping of suitable reference samples.

Nine remains were identified as belonging to persons that were biological relatives of the reference samples' donors. Four of them did not correspond to any family group. Unfortunately, bodies were prematurely returned to relatives and hence errors occurred, which forced a complete retyping of the exhumed bodies. This dramatic situation underscored the remarkable value of DNA typing not only for identifying remains and individuals, but for ensuring the proper use of other forensic identification practices conventionally employed, such as visual recognition, personal belongings, odontology, radiology, and others.

27.9 FUTURE PROSPECTS

Soon after the beginning of the new century, complex and multitudinous terrorist-produced mass disasters occurred. These challenging situations forced the forensic scientist to optimize and update the identification approaches based on DNA typing (see also Chapter 26). The most dramatic terrorist attacks were the September 11 bombing of the Twin Towers at the World Trade Center in New York City in 2001, which resulted in 2,974 fatal victims, the terrorist attack perpetrated against Atocha and other Madrid City Railway Stations on March 11, 2003, and the bombing attack on London's Public Transport system on July 7, 2005, which killed 55 people. In addition to these human-induced disasters are those that are produced naturally, such as the tsunami in South-East Asia in December 2004, Hurricane Katrina in the USA in August 2005, Cyclone Nargis which affected Myanmar on May 3, 2008, an extreme catastrophe that might have produced over 80,000 fatalities, and the destructive earthquake in China on May 15, 2008, whose final death toll was 69,136 victims.

In concordance with the magnitude of these catastrophes that might require DNA-based identification approaches, multi-laboratory facilities are needed. Additionally, international organizations such as Interpol have established detailed protocols for Mass Disaster Victim Identification, and the International Commission on Missing Persons may also provide advisory and technical help for mass fatalities worldwide. Currently, high-throughput platforms are efficiently powering a modern investigation of such disasters. Molecular diagnosis based on DNA-typing approaches has undergone an intense evolutionary process that might efficiently respond to the intensely demanding victim identification needs, imposed by extreme mass catastrophes that unfortunately characterize this age.

REFERENCES

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., and Young, I.G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290, 457–465.

Corach, D., Penacino, G., and Sotelo, A. (1994). Dealing with human remains in disaster areas. The case of the Israeli Embassy explosion occurred in Buenos Aires. *Adv. Forensic Haemogenet.* 5, 259–261.

Corach, D., Sala, A., Penacino, G., and Sotelo, A. (1995). Mass disasters: rapid molecular screening of human remains by means of STRs typing. *Electrophoresis* 16, 1617–1623.

Corach, D., Sala, A., Penacino, G., and Sotelo, A. (1996a). Combined STR, VNTR and MVR typing and mtDNA sequencing led to identification of human remains emergent from the AMIA explosion in Buenos Aires. *Adv. Forensic Haemogenet.* 6, 337–339.

Corach, D., Penacino, G., and Sala, A. (1996b). Room temperature preservation of putrid cadaveric soft tissues by means of inorganic compounds allows efficient recovery of DNA. *Proc. 7th International Symposium on Human Identification*, 203pp.

Corach, D., Penacino, G., Sala, A., Iannucci, N., Bernardi, P., Doretti, M., Fondebrider, L., Ginarte, A., Inchaurregui, A., Somigliana, C., Turner, S., and Hagelberg, E. (1997). Additional approaches to DNA typing of skeletal remains: the search for missing persons killed during the last dictatorship in Argentina. *Electrophoresis* 18, 1608–1612.

Edwards, A., Civitello, A., Hammond, H.A., and Caskey, C.T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* 49, 746–756.

Edwards, A., Hammond, H.A., Jin, L., Caskey, C.T., and Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12, 241–253.

Erlich, H.A., and Bugawan, T. (1989). HLA Class II gene polymorphism: DNA typing, evolution, and relationship to disease susceptibility. In *PCR Technology. Principles and Applications for DNA Amplification*, pp. 193–208. Stockton, New York.

Gabriel, M.N., Huffine, E.F., Ryan, J.H., Holland, M.M., and Parsons, T. J. (2001). Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy. *J. Forensic Sci.* 46, 247–253.

Galton, F. (1892). *Hereditary Genius: An Inquiry into its Laws and Consequences*. Macmillan/Fontana, London.

Gill, P. (2001). An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int. J. Legal Med.* 114, 204–210.

Ginther, C., Issel-Tarver, L., and King, M.C. (1992). Identifying individuals by sequencing mitochondrial DNA from teeth. *Nat. Genet.* 2, 135–138.

Herrin, G., Fildes, N., and Reynolds, R. (1994). Evaluation of the AmpliType® PM DNA test system on forensic case samples. *J. Forensic Sci.* 39, 1247–1253.

Jeffreys, A.J., Wilson, V., and Thein, S.L. (1985). Hypervariable "minisatellite" regions in human DNA. *Nature* 314, 37–73.

Jeffreys, A.J., Neumann, R., and Wilson, V. (1990). Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* 60, 473–485.

Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J., Weichhold, G., de Knijff, P., and Roewer, L. (1997). Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Legal Med.* 110, 125–133.

Kogan, S.C., Doherty, M., and Gitschier, J. (1987). An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. Application to hemophilia A. *N. Engl. J. Med.* 317, 985–990.

Landsteiner, K. (1901). Ueber Agglutinationserscheinungen normalen menschlichen Blutes. *Wien. Klin. Wochenschr.* 14, 1132–1134. [Translation: On agglutination phenomena of normal human blood, in S.H. Boyer (ed.), 1963, *Papers on Human Genetics*, pp. 27–31.]

Ludes, B., Tracqui, A., Pfitzinger, H., Kintz, P., Levy, F., Disteldorf, M., Hutt, J.M., Kaess, B., Haag, R., Memheld, B., Kaempf, C., Friederich, F., Evento, E., and Mangin, P. (1994). Medico-legal investigations of the Airbus, A320 crash upon Mount Ste-Odile, France. *J. Forensic Sci.* 5, 1147–1152.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* 51, 263–273.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235, 1616–1622.

Owen, R. (2000). Karl Landsteiner and the First Human Marker Locus. *Genetics* 155, 995–998.

Roewer, L., Amemann, J., Spurr, N.K., Grzeschik, K., and Epplen, J.T. (1992). Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum. Genet.* 89, 389–394.

Schmidtke, J., Kühnau, W., Wand, D., Edelmann, J., Szibor, R., and Krawczak, M. (2004). Prenatal exclusion without involving the putative fathers of an incestuous father-daughter parenthood. *Prenat. Diagn.* 24, 662–664.

Vastag, B. (2002). Out of tragedy, identification innovation. *JAMA* 288, 1221–1223.

Watson, J.D., and Crick, F.H. (1953). A structure for deoxyribose nucleic acid. *Nature* 171, 737–738.

Weber, J.L., and Mayr, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44, 388–396.

Wiegand, P., and Kleiber, M. (2001). Less is more length reduction of STR amplicons using redesigned primers. *Int. J. Legal Med.* 114, 285–287.

Wisner, B. (2004). Assessment of capability and vulnerability. In *Mapping Vulnerability Disasters, Development and People* (Bankoff, G., Frerks, G., and Hilhorst, D., eds), pp. 183–193. Earthscan, London.

Wyman, A.R., and White, R. (1980). A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA* 77, 6754–6758.

This page intentionally left blank

Chapter 28

Detection of Highly Pathogenic Viral Agents: Implications for Therapeutics, Vaccines and Biodefense

Kylene Kehn-Hall[1] and Sina Bavari[2]

[1] *The George Washington University, Department of Microbiology, Immunology, and Tropical Medicine, Washington DC, United States of America;*

[2] *United States Army Medical Research Institute of Infectious Diseases, Bacteriology Division, Ft. Detrick, MD, United States of America*

28.1 INTRODUCTION

Hemorrhagic fever viruses (HFV) are a group of diverse viruses that have a common denominator of causing viral hemorrhagic fever. These viruses are responsible for high levels of both mortality and morbidity in infected patients. The detection of HFV has been generally limited to serological-based assays, electron microscopy, or plaque assays all of which have serious limitations. However, the new standard in HFV detection is the use of quantitative reverse transcription PCR (q-RT-PCR) assays (see also Chapter 7). Herein we give a brief overview of q-RT-PCR assays, describe the published q-RT-PCR assays for HFV to date, and discuss the implications of these assays for use in expanding upon the limited therapeutic arsenal against HFV.

28.2 OVERVIEW OF HEMORRHAGIC FEVER VIRUSES

Viral hemorrhagic fever (VHF) is a group of severe diseases characterized mainly by fever and hemorrhage. Hemorrhage, from which the diseases are named, is only found in a fraction of patients and is rarely the cause of death. VHF has initial symptoms of fever, headache, myalgia, gastrointestinal problems, followed by organ failure. VHF is caused by multiple viral families, including filoviruses (Marburg and Ebola), arenaviruses (Lassa, Junin, Machupo, Guanarito, and Sabia), bunyaviruses (Crimean-Congo hemorrhagic fever virus (CCHFV), Rift Valley fever virus (RVFV), and hantaviruses), and flaviviruses (yellow fever, dengue, Omsk hemorrhagic fever, Kyasanur Forest disease) (Borio *et al.*, 2002). These particular viruses are thus termed hemorrhagic fever viruses (HFV). The majority of HFV are considered to pose serious risk as biological weapons and have therefore been classified as Category A agents (Borio *et al.*, 2002).

HFV are enveloped single-stranded RNA viruses with variable genomes (Fig. 28.1). With the exception of flaviviruses, HFV are negative-stranded RNA viruses. Both arenaviruses and bunyaviruses are segmented, whereas filoviruses and flaviviruses are non-segmented. The morphology of HFV is also very different with arenaviruses and bunyaviruses being spherical or pleomorphic, flaviviruses being isometric, and filoviruses being unusual filamentous particles.

The epidemiology of HFV varies depending on the virus in question (Table 28.1). Humans are not the natural reservoir for HFV; rather they become infected by coming into contact with the animal reservoir or through arthropod vectors. However, once humans are infected they can spread the disease from person to person, mainly by coming into contact with infectious blood and bodily fluids. Notable exceptions to this are RVFV and the flaviviruses, where person to person spread has not been observed (Borio *et al.*, 2002). Airborne transmission can also occur, but is not the primary means of spread.

The natural animal reservoir for the arenaviruses and hantaviruses are rodents. Humans generally become infected by coming into contact with an infected animal or through inhalation of rodent urine or feces particles. It is thought that the type of rodent reservoir localizes the virus to a particular geographic location. Lassa virus, the causative agent of Lassa fever, is endemic in Africa and

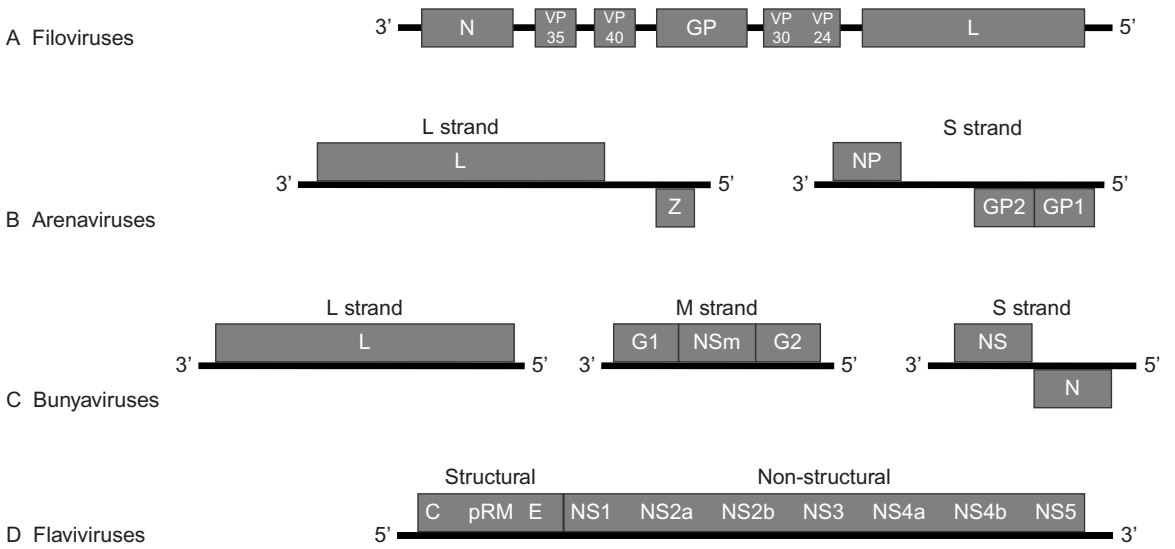


FIGURE 28.1 Genomes of the four HFV families. The general structures of the HFV genomes are depicted. For each viral family, only one example is shown to give a general overview of the genomic structure and allow for comparison between the viral families. For filoviruses the Ebola genome is depicted (A), for arenaviruses the Lassa genome (B), for bunyaviruses the RVFV genome (C), and the dengue genome for the flaviviruses (D).

TABLE 28.1 Characteristics of hemorrhagic fever viruses.

Family	Virus	Vector/Reservoir	Disease	Endemic region
Filovirus	Ebola	Unknown	Ebola hemorrhagic fever (EHF)	Sub-Saharan Africa
	Marburg	Unknown	Marburg hemorrhagic fever (MHF)	Sub-Saharan Africa
Arenavirus	Lassa	Rodents	Lassa fever	West Africa
	Junin	Rodents	Argentine hemorrhagic fever (AHF)	South America
	Machupo	Rodents	Bolivian hemorrhagic fever (BHF)	South America
	Guanarito	Rodents	Venezuelan hemorrhagic fever (VHF)	South America
	Sabia	Rodents		South America
Bunyavirus	Crimean-Congo hemorrhagic fever	Ticks	Crimean-Congo hemorrhagic fever (CCHF)	Africa, China, Middle East
	Hantavirus	Rodents	Hemorrhagic fever with renal syndrome (HFRS), human pulmonary syndrome (HPS)	Americas, Asia, and Europe
	Rift Valley fever	Mosquitoes	Rift Valley fever	Sub-Saharan Africa, Egypt, Saudi Arabia, Yemen
Flavivirus	Dengue	Mosquitoes	Dengue fever, dengue hemorrhagic fever, dengue shock syndrome	Asia, Africa, Pacific, and the Americas
	Yellow fever	Mosquitoes	Yellow fever	Sub-Saharan Africa, tropical regions of South America
	Omsk hemorrhagic fever	Ticks	Omsk hemorrhagic fever	Central Asia
	Kyasanur Forest disease	Ticks	Kyasanur Forest disease	Karnataka State, India

is hence classified as an Old World arenavirus. Junin, Machupo, Guanarito, and Sabia are all New World arenaviruses endemic to South America (Drosten *et al.*, 2002). Junin causes Argentina hemorrhagic fever and Machupo is the causative agent of Bolivian hemorrhagic fever, both with a case fatality rate of about 20% (Charrel and de Lamballerie, 2003). Venezuelan hemorrhagic fever is caused by Guanarito. Sabia virus was just recently isolated in 1994, from a patient in Brazil who died from an HF disease. Hantaviruses can be divided into Old and New World hantaviruses, which tend to cause distinct diseases. Old World hantaviruses are the causative agent of hemorrhagic fever with renal syndrome (HFRS) and New World hantaviruses cause human pulmonary syndrome (HPS) (Muranyi *et al.*, 2005). Hantaviral infections are found in the Americas, Asia, and Europe.

RVFV, yellow fever, and dengue all utilize mosquitoes as their vector, whereas CCHFV, Omsk hemorrhagic fever (OHF), and Kyasanur Forest disease (KFD) use ticks (Borio *et al.*, 2002; Monath, 2001). Both RVFV and CCHFV are zoonotic viral diseases that infect a number of wild and domestic animals, especially sheep and cattle. Humans can become infected by a bite of a mosquito or from coming into contact with blood or tissue of infected animals (Monath, 2001). RVFV is primarily found in sub-Saharan and North Africa. More recently there has also been outbreaks in Yemen and Saudi Arabia, which are thought to be due to infected sheep being imported from Africa (Madani *et al.*, 2003). The name Crimean-Congo originates from two different outbreaks of the virus. The first occurred in 1944 in Crimea and later a virus was isolated from the Congo, hence the name Crimean-Congo Hemorrhagic fever virus (Monath, 2001). CCHFV has a large geographic distribution being detected in over 30 countries in Africa, south-east Europe, Asia, and the Middle East (Hoogstraal, 1979). Yellow fever is endemic in tropical regions of South America and sub-Saharan Africa. Dengue infection results in a number of diseases including dengue fever, dengue hemorrhagic fever, and dengue shock syndrome and is prevalent in Asia, Africa, Pacific, and the Americas (Gubler, 1998a). There are approximately 450,000 cases of dengue hemorrhagic fever/dengue shock syndrome reported every year (Gubler, 1998a, b). Both OHF and KFD are rare diseases that have only recently been recognized. KFD is found in southwestern India, while OHF has been recognized in central Asia (Gritsun *et al.*, 2003, Pattnaik, 2006).

Filoviral infections have generally been contained to sub-Saharan Africa. The natural reservoir of filoviruses is unknown to date. There have been several reports suggesting that bats have an asymptomatic filoviral infection and are a potential reservoir. This hypothesis is based on the detection of both Ebola or Marburg nucleic acid as well as antibody to the virus (Leroy *et al.*, 2005; Swanepoel *et al.*, 2007; Towner *et al.*, 2007a) . However, live virus was unable to be isolated in any case. Therefore, the implication of these findings has yet to be fully demonstrated.

28.3 DETECTION METHODS

The standard virological assay to measure productive virus reproduction is the plaque assay (Moe *et al.*, 1981). Plaque assays for HFV are performed using African green monkey kidney (Vero) cells. Briefly, Vero cells are grown in a monolayer and media removed. Various dilutions of samples are applied to the monolayer. The samples are incubated to allow infection to occur, media removed, and fresh media added. Cell death due to the infection will produce an area of clearing, termed "plaques", which can be counted. The numbers of plaques are then correlated back to the dilution to determine the quantity of virus presented, expressed in plaque forming units per milliliter (pfu/ml). Although this is a widely used method, there are a number of major drawbacks to this procedure. First, plates need to be incubated for up to two weeks before results can be determined. Second, the entire procedure must be performed in either a biosafety level (BSL)-3 or -4 laboratory (depending on the virus of interest). Third, the quantitation determined is not necessarily reflective of the actual number of viruses present. This could be due to a number of reasons including inaccurate dilutions, defective viruses, the state of the cell, or the inherent low infectivity rate of the virus. Collectively, plaque assays are time consuming and questionably quantitative. Therefore, there is a need for a highly quantitative alternative that can be performed in a less restrictive laboratory environment (i.e. BSL-2). A number of molecular-based assays are currently being used as alternatives to plaque assays. These include standard PCR, nested PCR, and quantitative PCR.

28.3.1 RNA Preparation

The initial step in almost any PCR-based assay is RNA extraction. Reliable RNA sample preparation is critical for downstream analysis. The most commonly used RNA extraction methods employ an acidic phenol/guanididium isothiocynate mixture to inactivate RNases and allow the separation of nucleic acids into the aqueous phase. An added benefit of RNA extraction reagents is that they render the viruses non-infectious. Blow and coworkers (2004) demonstrated that various HFV were non-infectious after treatment with either Trizol LS reagent (Invitrogen) or AVL buffer (Qiagen) as demonstrated by plaque assays. As mentioned above, HFV must be handled in BSL-3 or -4 laboratories at all times. However, after treatment with Trizol LS, they can be safely removed from BSL-3/4 laboratories and enter a BSL-2 laboratory for further processing.

Phenol-based RNA extraction is a common and reliable method. However, there are two common mistakes that can occur: (1) the loss of RNA pellet upon removal of ethanol, resulting in loss of precious sample and (2) failure to dry the RNA pellet completely, resulting in residual ethanol, which often inhibits PCR. To avoid these problems, many laboratories have moved toward commercially available kits that bind either DNA or RNA to a silica matrix, followed by wash steps to remove contaminating substances, and elution. These kits provide reliable and reproducible results for both the beginner and the expert.

Phenol-based extraction is difficult with large sample numbers and not amenable to high-throughput screening. There are a number of commercially available kits that streamline RNA extraction in a 96-well format. However, as mentioned above, due to safety considerations, HFV must be treated with Trizol LS reagent (Invitrogen) or AVL buffer (Qiagen) to inactivate the virus before removal from the BLS-4 laboratories for further processing (Blow *et al.*, 2004). Therefore, the authors adapted and optimized Ambion's MagMax-96 for microarrays kit to handle their sample constraints. The MagMax-96 kit is a magnetic bead-based kit able to process 96 samples at a time. The protocol was modified slightly to accommodate larger volumes and the Trizol LS reagent. Mock extraction experiments were performed to test the reproducibility of the extraction procedure. Various amounts of purified Lassa RNA were resuspended in Trizol LS and extracted according to the modified MagMax-96 for microarray extraction procedure ($n = 6$ per genomic copy concentration). Lassa RNA was quantitated using an optimized Lassa q-RT-PCR assay. Mean Ct values were plotted versus the genomic copy-number input (Fig. 28.2). The high-throughput RNA extraction method provided reproducible extraction results for all quantities of input tested.

In 2007, two separate studies from the Center for Disease controls demonstrate the high-throughput RNA extraction of HFV (Bird *et al.*, 2007; Towner *et al.*, 2007b). Towner and coworkers (2007b) systematically demonstrated the safety and feasibility of a high-throughput RNA extraction utilizing NC lysis buffer (Applied Biosystems) and extraction with the 6100 Nucleic Acid Preparation Station (Applied Biosystems). The NC lysis buffer is not phenol based, but does contain guanididium isothiocynate. The optimal ratio of lysis buffer to induce complete viral inactivation was established by testing the infectivity of samples treated with various ratios of NC lysis buffer by plaque assays. A 3:1 ratio showed complete viral inactivation; however, a 6:1 ratio (NC lysis buffer:sample) was adopted to allow for a margin of error. Deep 96-well plates were heat sealed before being dunked into decontamination solution and transferred from the BSL-4 to the BSL-2 laboratory. This method proved to completely inactivate the virus without any cross-contamination between wells. This was demonstrated through alternating EBOV spiked sample and no template samples in the 96-well plate and subsequently monitoring by q-RT-PCR. No false positive or negative results were observed. This high-throughput RNA extraction procedure coupled with a Marburg q-RT-PCR assay successfully analyzed over 500 samples at a field laboratory that was established during the 2005 Marburg outbreak in Angola. Bird and coworkers (2007) adapted this high-throughput extraction protocol to analyze RVFV samples from Saudi Arabia outbreaks in 2000.

28.3.2 RT-PCR

All HFV are RNA viruses and therefore require reverse transcription-PCR (RT-PCR). RT-PCR involves the processing of RNA into cDNA prior to PCR. This can be done in either a separate reaction (two-step RT-PCR) or in a single reaction (one-step RT-PCR). One-step RT-PCR is possible due to the use of specialized enzymes. For example, one method employs Tth polymerase, a DNA-dependent DNA polymerase, which also contains reverse transcriptase activity (Gibb *et al.*, 2001a; Sanchez *et al.*, 1999). Another method utilizes both an RT enzyme and a DNA polymerase. The cDNA synthesis typically occurs between 50°C and 60°C, at which the DNA polymerase is inactive. The

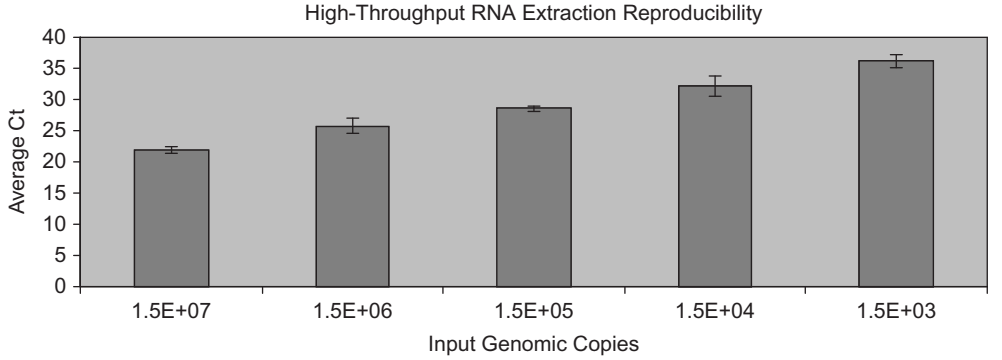


FIGURE 28.2 High-throughput RNA extraction reproducibility. Various amounts of purified LASV RNA (1.5e7 to 1.5e3) were resuspended in Trizol LS and extracted according to the modified MagMax-96 for microarray extraction procedure ($n = 6$ per genomic copy concentration). LASV RNA was quantitated using the optimized Lassa q-RT-PCR assay. Mean Ct values are plotted versus the genomic copy-number input.

DNA amplification reaction is then initiated at 95°C or "hot started", at which point the DNA polymerase now becomes active. The hot start of Taq polymerase increases specificity of the RT-PCR reaction. The decision to use a two-step versus a one-step RT-PCR system is based on a number of factors. One-step RT-PCR is generally favored because it is quicker, less expensive to use, and minimizes the number of steps. This last factor is critical because less handling of samples translates to fewer possibilities for errors and contamination. Quantitative one-step RT-PCR has also been shown to be more sensitive for specific genes (Peters *et al.*, 2004; Wacker and Godard, 2005). In the one-step method, gene-specific primers are used and both the RT and PCR occur in one reaction tube. Therefore, one potential drawback to the one-step method is that once the RNA sample is consumed, no additional reaction can be performed. If followed-up testing is needed, RNA samples should be aliquoted and stored in a RNA favorable solution. RNA secure (Ambion) is a commonly used RNA storage solution that is an enzymatically free buffer to prevent RNase degradation.

One advantage to two-step RT-PCR is that a cDNA sample is created from the RNA sample, which can be used for multiple PCR reactions. For two-step RT-PCR, random hexamers or oligo dT primers are typically used in an RT reaction in a separate tube. It is generally thought that cDNA is more stable than RNA, which is an advantage to the two-step RT-PCR method. A caveat to this is that mRNA may be more stable than the cDNA created in a two-step RT-PCR method (Wilkening and Bader, 2004).

Numerous standard RT-PCR assays have been developed for the detection and diagnosis of HFV (Drosten *et al.*, 2003). Nested RT-PCR, which has increased sensitivity and specificity over traditional RT-PCR, has also been described for multiple HFV, including CCHFV, RVFV, Lassa, and dengue (Drosten *et al.*, 2002; Lanciotti, 2003). Nested RT-PCR involves the use of two different primer sets and two PCR reaction steps. The second set of primers is within the region amplified by the first primer set and thus termed "nested". An important limitation of nested RT-PCR is that the additional handling and manipulation steps can lead to contamination. The advent of quantitative PCR and quantitative RT-PCR (q-RT-PCR) has highlighted multiple limitations of standard PCR as compared to q-PCR. Some important limitations are lower sensitivity, short dynamic range, non-automated, post-PCR processing, and non-quantitative. Therefore, q-RT-PCR is now the standard for molecular-based detection of HFV.

28.3.3 q-RT-PCR

q-RT-PCR technology allows the detection of amplified products as the reaction is occurring, through the use of either a specific fluorescently labeled probe or a non-specific double-stranded DNA binding fluorescent dye. This approach offers the benefits of high specificity, sensitivity, and high-throughput capabilities. In q-RT-PCR, the number of cycles needed for fluorescence to reach a specific detection threshold is termed the Ct (cycle threshold) value. The Ct value is inversely related to the amount of DNA in the sample. That is, the more starting template you have, the lower the Ct will be. Quantitation can be performed in either a relative or absolute manner. Relative quantitation compares the Ct values from the target RNA to the Ct values of an internal reference gene (see also Chapter 7). The results are expressed as a ratio of the target to the internal reference. This method is often referred to as the comparative Ct method. Absolute quantitation requires the use of a serially diluted standard template to generate a standard curve. The most commonly used template for standard curves are *in vitro* T7 transcribed RNA or serially diluted viral stocks. Absolute quantitation is used almost exclusively for HFV q-RT-PCR assays.

28.3.3.1 Primer Design

A critical aspect of PCR-based assays is the primer design. Without proper primer design the assay will fail to provide reliable and reproducible data. RNA viruses tend to have high mutation rates due to the error-prone nature of their RNA polymerase (Holland and Domingo, 1998). This high mutation rate results in a great amount of genetic variability, making primer design difficult for assays which seek to detect multiple viruses and/or viral strains (i.e. diagnostic tests). For example, when analyzing 13 different viral strains of CCHFV it was determined that the nucleotide divergence was 20% for the S (nucleopcapsid) gene, 31% for the M (glycoprotein) gene, and 22% for the L (polymerase) gene (Deyde *et al.*, 2006). Junin strains are also relatively diverse with up to 13% diversity at the nucleotide level for the GP1 gene and up to 9% for the NP gene (Garcia *et al.*, 2000). Primers can tolerate up to four internal mismatches between the template and primer without effecting PCR efficiency. In contrast, mismatches at the 3′ end are detrimental to the product yield (Christopherson *et al.*, 1997; Kwok *et al.*, 1990). With these restrictions primer design for multiple viruses and/or strains can be difficult. The development of high-quality primer and probe design software has helped to alleviate some of these concerns. One mechanism to deal with the genetic diversity is to utilize degenerate primers. Degenerate primers are primers that have more than one base possibility at certain positions, and thus can recognize more than one nucleotide and multiple sequences. A degenerate primer-based q-RT-PCR assay was designed to detect both Ebola Sudan and Ebola Reston (Gibb *et al.*, 2001a). An added benefit of this assay was that differentially labeled probes were specific to each viral strain and thus enabled the detection of both strains with one primer set, but also allowing probe specific distinction.

Due to the genetic variability, primer sets are often designed to recognize the most conserved genes. For example, all published CCHFV q-RT-PCR assays are designed to detect the S segment due to lower levels of genetic variability. Some other top candidates are the polymerase gene, L, GP, and N genes (Fig. 28.1). It should be noted that there is limited sequence availability for certain HFV (Huggins, 1989) and therefore primers may have to be redesigned as additional strains are identified. In fact, a review in 2003 analyzed published primer sets for various HFV and found that several primers would have five or more mismatches with newly available viral sequences and therefore would theoretically be unable to detect those viral strains (Drosten *et al.*, 2003). As an example, the Angola Marburg strain was recognized as a result of the 2005 outbreak in Angola (Towner *et al.*, 2006). The published q-RT-PCR assays at the time were not optimal for this strain and primer and probes were redesigned to accurately recognize and quantitate Angola Marburg (Weidmann *et al.*, 2007).

28.3.3.2 Detection Chemistries

In q-RT-PCR, detection is accomplished through the use of either a specific fluorescently labeled probe or a non-specific double-stranded DNA binding fluorescent dye. One of the most commonly used dyes is *SYBR Green*, which binds to the minor groove of double-stranded DNA (dsDNA). Prior to binding DNA, *SYBR Green* exhibits very little fluorescence, but upon binding to the DNA its fluorescence is greatly enhanced. The major drawback to using *SYBR Green* or other DNA binding dyes is the lack of specificity. *SYBR Green* binds to all dsDNA sequences present in the PCR reaction. Therefore, primer specificity is extremely important when *SYBR Green* is utilized. For these reasons, *SYBR Green* is generally used when money is limiting and/or as a starting point in assay development. The detection of non-specific products can be evaluated through melting curve analysis at the end of the assay run. However, this becomes problematic when there are low starting template concentrations, resulting in reduced sensitivity and specificity of DNA binding dyes (Mackay *et al.*, 2002). *SYBR Green*-based assays have been developed for multiple HFV due to large genetic variability limiting probe design.

For more specific product detection, fluorescently labeled probes are used. There are various varieties of probes available. Wong and Medrano (2005) compiled an excellent review of the detection chemistries. The main types of probes that have been utilized in HFV q-RT-PCR are Taqman probes and fluorescence resonance energy transfer (FRET) probes (see also Chapter 7). Taqman probes are often used due to their ease of use, reliability, and commercially available software for probe design (i.e. Primer Express). Hairpin probes, such as molecular beacons, have not been used for HFV, but are a useful technology to consider. It is important to note that the issues related to primer design (see section 28.3.3.1) also apply to probe design.

Taqman Probes

Taqman probes are a specific type of hydrolysis probe that use FRET technology combined with the Taq polymerase 5′ exonuclease activity. Taqman probes contain a reporter fluorescent molecular at the 5′ end of the probe and a quencher molecule at the 3′ end of the probe. The quencher molecule is able to suppress the fluorescence of the reporter molecule due to its close proximity. As the Taq polymerase replicates the DNA, the 5′ exonuclease activity degrades the probe, resulting in release of the quencher molecule and fluorescence of the reporter (Walker, 2002). In traditional Taqman probes, the quencher is fluorescent, which may demonstrate some background fluorescence in the absence of amplified product. Minor groove binding (MGB) probes, which contain a non-fluorescent quencher, are also available. MGB, such as dihydrocyclopyrroloindole tripeptide (DPI_3) can be used to increase the T_m and therefore allow the design of shorter probes (Wong and Medrano, 2005). Shorter probes may be useful when a conserved region of RNA is hard to find, such as in HFV.

FRET Hybridization Probes

FRET hybridization probes can rely on either the three or the four oligonucleotide method (Wong and Medrano, 2005). In the four oligonucleotide method a set of primers and two probes are utilized. The upstream probe contains an acceptor dye at the 3′ end and the downstream probe a donor dye at the 5′ end. The probes are sequence specific and bind to the template in a head-to-tail orientation, allowing an increase in FRET when both are bound. The three oligonucleotide method relies on a similar principle, but only has one probe with a donor dye. The upstream primer contains the acceptor dye at its 3′ end, functionally replacing the second probe used in the four oligonucleotide method. It is important to note that both Taqman and hybridization probes are based on FRET, but in opposite manners. Taqman probes rely on a decrease of FRET from the close proximity of the quencher and the reporter, whereas hybridization probes rely on an increase of FRET due to the close proximity of the acceptor and donor dye.

Molecular Beacons

Molecular beacons are single-stranded short hairpin probes that unfold in the presence of the target sequence. The molecular beacon consists of four main parts. The first part is an 18–30 bp loop region that is complementary to the target sequence. The second part contains a complementary stem portion that consists of 5–7 bp on both sides of the loop portion. These two structural portions allow the formation of a stem–loop structure, with the loop being

exposed and available for complementary binding to the target sequence. The third part (at the 5′ end) contains a fluorescent molecule that becomes active when the stem loop structure is disassociated. Finally, a non-fluorescent quencher is located at the 3′ end, which prevents the molecular beacon from emitting light when it is in the closed loop structure.

Due to the nature of complementary sequence binding, molecular beacons are well suited for the discrimination of single base pair differences. Assays utilizing molecular beacons would be very sensitive to HFV strain differences and may prove to be useful by providing strain-specific information.

28.3.3.3 Controls and PCR Inhibition

Controls are essential to ensure the integrity of each q-RT-PCR reaction. q-RT-PCR assays are particularly sensitive to PCR inhibition. PCR inhibition can be caused by improper sample preparation (residual ethanol, and/or salts) or it may originate from the sample itself, such as heme from blood samples. To avoid false negatives caused by PCR inhibition either an internal positive control or exogenous positive control can be used. Internal positive controls are typically artificial control RNA that is added either at the time of extraction or amplification. They are amplified in the same reaction as the sample under study, thus termed internal. Internal positive controls require the use of an additional probe and/or primers to specifically detect the target template and thus can be more complicated to implement. Internal positive controls have been successfully applied for HFV (Panning *et al.*, 2007; Wolfel *et al.*, 2007). There are also commercially available internal positive control kits, such as from the IPC kit from Applied Biosystems, which provides the template, primers, and probes in an all-in-one kit. Alternatively, exogenous controls can be spiked into a duplicate sample prior to RNA extraction to control for extraction efficiencies as well as PCR inhibition (Drosten *et al.*, 2003; Roth *et al.*, 1999; Towner *et al.*, 2007b). Samples should then be processed in parallel. Exogenous controls are usually *in vitro* transcribed RNA or virus. A negative result from samples containing either internal positive or exogenous positive controls indicates PCR inhibition and the sample should be diluted and reanalyzed. Finally, to control for differences in master mix volume a passive reference dye (such as ROX) is usually included in the q-RT-PCR reaction. This technology has been implemented in most HFV assays.

28.3.3.4 Published Methods

Over the past five years, there has been an abundance of q-RT-PCR assays developed for HFV. That being said, there is still a lack of available assays for certain viruses. For example, for the arenaviruses there are multiple assays available for Lassa, but none for Machupo or Sabia. The major differences in the assays lie in their use of detection chemistries, internal and exogenous positive controls, and the number and nature of primer and probes. In addition, the published assays vary widely based on the ability to identify multiple strains and/or multiple viruses.

Filoviruses

Filovirus q-RT-PCR assays were one of the first assays published. Gibb and coworkers (2001) described a one-step q-RT-PCR assay that utilized degenerate primers to amplify both Ebola Zaire and Sudan viruses. Taqman probes that were specific for either Zaire or Sudan allowed for the differentiation between the two strains. Another one-step q-RT-PCR assay from the same group was also published in 2001 for the detection of Marburg. This assay was designed to detect Musoke and Ravn with a sensitivity of 2–5 pfu/PCR reaction (Gibb *et al.*, 2001b). In 2002, Drosten and coworkers published a study describing q-RT-PCR assays for the detection of Ebola, Marburg, Lassa, CCHFV, RVFV, dengue and yellow fever (Drosten *et al*., 2002; Towner *et al.*, 2007b). One benefit of their design was that there were two sets of RT-PCR conditions applicable to all these assays, allowing several assays to be run in parallel enabling a quick turnaround time for diagnosis (Drosten *et al*., 2002). Filovirus primers for the L gene were used to detect both Ebola and Marburg. The q-RT-PCR assay for Ebola/Marburg was *SYBR Green* based and thus run in parallel with the Lassa and CCHF *SYBR Green*-based assays. After the Angola Marburg outbreak of 2005, a q-RT-PCR assay was developed that was able to recognize Angola strains as well (Weidmann *et al.*, 2007).

In 2007, a network of European BSL-4 laboratories in collaboration with Qiagen published the first industry-standard q-RT-PCR assay for filoviruses (Panning *et al.*, 2007). Their assay utilizes five primer sets and three probes targeted against the L gene and was designed to detect all known filoviral L sequences. This multiplexing strategy was necessary to detect all Ebola and Marburg strains in one assay, but did not decrease the overall assay sensitivity, which was approximately ten copies/assay. A copy of target RNA that has a mutated probe binding site and therefore must be detected by an alternative probe was included as an internal positive control that prevents false negative results due to poor sample preparation. To our knowledge, this is the first HFV q-RT-PCR assay to include an internal positive control and multiple primer/probes. This assay system is a major breakthrough that will enable the consistent and reliable detection of filoviruses.

Arenaviruses

There are a number of q-RT-PCR assays currently available for Lassa, but not for the other arenaviruses. The first

Lassa q-RT-PCR assay was developed by Drosten and coworkers (2002) and utilized *SYBR Green* detection chemistry targeting the GPC gene. *SYBR Green* was chosen due to the high-level genetic variability of Lassa (Bowen *et al.*, 2000), which prevented probe design that allowed detection of multiple Lassa strains. This one-step assay was capable of detecting 2,445 virus genome equivalents (geq)/ml at a 95% confidence interval. The primers used in this assay were originally designed for a standard RT-PCR (Demby *et al.*, 1994). Recently, a multiplex q-RT-PCR assay that targets the L gene has been developed (Vieth *et al.*, 2007). The L gene was chosen based on its higher level of conservation among not only Lassa, but other Old World arenaviruses. This assay is capable of detecting Lassa, Mobala, Ippy, Mopeia, Morogoro, and LCMV, all Old World arenaviruses. A mixture of three forward primers and two reverse primers is used to control for the genetic variability among the viruses, at the same time retaining a sensitivity of 4290 copies of Lassa/ml.

For the New World arenaviruses there are only q-RT-PCR assays published for the detection of Junin and Guanarito (Vieth *et al.*, 2005). These assays, which were designed to detect the NP gene, used FRET probe technology, requiring the use of two probes and two primers for each assay. While the assays specifically detected both Junin and Guanarito, the sensitivity of the Guanarito assay was only 1,000 copies per reaction. The lack of sensitivity observed here is possibly due to probe and primer restraints that were placed to try to recognize all strains of these viruses. In addition, the authors speculated that the lack of sensitivity was a result of secondary structure affecting the RT step (Vieth *et al.*, 2005).

Bunyaviruses

The first q-RT-PCR for bunyaviruses described by Drosten and coworkers (2002) (see filoviruses) was designed to detect the NP gene of CCHFV and the G2 gene of RVFV (Drosten *et al*., 2002). The CCHFV assay was *SYBR Green* based, due to the large genomic variability limiting probe designed, whereas the RVFV assay was Taqman probe-based. The sensitivity of these assays was 2,779 geq/ml for CCHFV and 2,835 gep/ml for RVFV. In the past three years, five additional q-RT-PCR assays all targeting the S segment of CCHFV have been published. To overcome the constraints of strain variation on primer and probe design Yapar and coworkers (2005) developed a one-step q-RT-PCR Taqman-based assay utilizing degenerate primers and probes to allow the detection of CCHFV infected samples (Yapar *et al.*, 2005). Eighteen serum samples from patients from East Anatolia were tested. However, it was not clear how many different strains of CCHFV this group of patients represented. A q-RT-PCR assay specific for CCHFV strains in the Balkan region and the Drosdov strain was established (Duh *et al.*, 2006). A FRET probe which uses the endonuclease activity of Taq polymerase was used. Detection of CCHFV strain Kosovo Hoti, clinical serum samples and ticks were used to verify the specificity and sensitivity of the assay. Importantly, this assay did not cross-react with the Dobrava or Puumla hantavirus RNA, which are prevalent in the Balkan region. In a follow-up study that evaluated the ability of q-RT-PCR to accurately measure viral load in 24 patients from Kosovo, the above described q-RT-PCR assay was modified slightly (Duh *et al.*, 2007). Multiple controls were incorporated into the assay, including a synthetic RNA to be used as a quantitative calibrator and a competitive internal control which recognizes a binding site present in the synthetic RNA, but not in the template RNA, to allow detection of PCR inhibitors. Wolfel and coworkers (2007) were able to develop a q-RT-PCR assay that detects a global spectrum of clinically relevant viral strains. They adopted an alternative strategy to deal with the genetic diversity of CCHFV through incorporating a multiple probe Taqman-based assay system, which used a primer set and three probes. This assay was tested against 63 serum samples from 31 patients with CCHFV infection. When compared to a standardized nested RT-PCR assay, the q-RT-PCR assay demonstrated sensitivity at least as high. Finally, a Taqman MGB probe-based assay was developed that was able to detect 18 strains of CCHF (Garrison *et al.*, 2007). The sensitivity was demonstrated to be ten genomic copies and it did not cross-react with RNA from 28 viruses or DNA from 78 organisms.

In addition to the RVFV q-RT-PCR assay described by Drosten and coworkers (2002), there are two other assays that have been published in the literature. The first is a Taqman-based assay that targeted the highly conserved NS region. Sera from mice infected with RVFV strain ZH548 were used successfully to evaluate the assay. In addition, the authors demonstrated the feasibility of using q-RT-PCR for antiviral drug screening through testing four compounds including ribavirin as a positive control. A study published in 2007 indicated that the primers and/or probes designed in both the Drosten and coworkers (2002) and Garcia and coworkers (2001) reports contained mismatches with recently acquired RVFV genomes that may hinder the ability to detect these newly recognized RVFV. Therefore, new primer and probes were designed that recognized 40 strains to establish a "pan-RVFV" q-RT-PCR assay (Bird *et al.*, 2007). Evaluation of this assay indicated that it had high sensitivity (~5 RNA copies of RNA/reaction) and high efficiency. Furthermore, this group coupled the q-RT-PCR assay with a high-throughput RNA extraction method that allowed the evaluation of 96 samples in 4.5 to 5 hours.

A number of q-RT-PCR assays for the detection of multiple hantaviruses (Puumala, Seoul, Dobrava, and Hantaan) in cell culture have been described (Aitichou *et al.*, 2005; Garin *et al.*, 2001; H *et al*., 2008). In particular one group established their q-RT-PCR assay for the evaluation of anti-viral

compounds (H *et al*., 2008). Kramski and coworkers (2007) developed five q-RT-PCR assays, three that specifically detected Dobrava, Puumala, or Tula hantaviruses (which cause HFRS) and two that recognize a combination of hantaviruses (which cause HPS). These assays were based on degenerate primers and MGB probes that recognize the S segment of hantaviruses. The Puumala and Dobrava hantavirus q-RT-PCR assays were able to be multiplexed with no adverse effects. Pyrosequencing (see also Chapter 8) was used to confirm the q-RT-PCR results. In a case study of the first Dobrava hantavirus patient with HFRS, the hantavirus infection was confirmed with an *SYBR Green*-based q-RT-PCR assay (Jakab *et al.*, 2007). A Puumala-specific q-RT-PCR assay utilizing Taqman MGB probes was also developed (Evander *et al.*, 2007). The sensitivity of this assay was particularly impressive, being able to detect down to two copies of RNA per reaction. One importance aspect of these studies is that HFRS displays similar clinical manifestations and there is therefore a need to determine the specific hantavirus causing disease. Another Taqman-based q-RT-PCR assay designed to specifically detect Dobrava hantavirus was tested against 46 patients with confirmed infection (Saksida *et al.*, 2008). This assay has the added benefit of an internal control probe and synthetic RNA transcripts, which allows for quantitation calibrators and PCR inhibitor controls. Viral RNA was detected in 26 of the patients, 13 with severe disease and 13 with mild to moderate disease. The authors speculated that the lack of detection of viral RNA in other patients could be due to low levels of virus present even in severe disease cases and that severity of disease does not depend solely on virus concentration. The results of this study are surprising in light of the sensitivity demonstrated by other q-RT-PCR assays.

Flaviviruses

Most standard flavivirus PCR assays utilize primers that target the highly conserved non-structural NS5 region (Drosten *et al.*, 2003). Taqman-based q-RT-PCR assays have been described that specifically target either the envelope or the 5′ non-coding region of the virus (Drosten *et al*., 2002). Sensitivity was determined to be 1,545 geq/ml (Drosten *et al*., 2002). q-RT-PCR assays have also been successfully implemented in the testing of yellow fever patients samples (Bae *et al.*, 2005). A multiplex q-RT-PCR assay designed to detect medically important flaviviruses, including yellow fever, has been described (Chao *et al*., 2007). This multiplex utilizes four primers and four probes, each probe containing four unique reporter dyes, allowing specific detection of viruses. Even with the multiplex platform, the assay sensitivity for yellow fever was 3.5 pfu/ml and the assay was specific to the virus of interest.

There are numerous published q-RT-PCR assays for the detection of dengue (Callahan *et al.*, 2001; Chutinimitkul *et al.*, 2005; Drosten *et al*., 2002; Houng *et al.*, 2001). One such assay has the ability to detect all dengue virus subtypes (Drosten *et al*., 2002). Other assays are able to distinguish between all four subtypes of dengue (Callahan *et al.*, 2001; Drosten *et al*., 2002; Houng *et al.*, 2001). Recently, two *SYBR Green*-based one-step q-RT-PCR assays were described. The first is a multiplex assay consisting of one conserved forward primer and four reverse primers targeting areas of the M and C genes of dengue 1, 2, 3, and 4 (Yong *et al.*, 2007). When examining dengue patient samples, the assay detected dengue in 178 out of 210 samples with a specificity of 100% being observed. In 2008, an *SYBR Green*-based one-step q-RT-PCR assay was developed against the 5′ region of the dengue genome (Dos Santos *et al.*, 2008). This assay was very specific for dengue, not detecting other flaviviruses and able to detect dengue virus 1, 2, and 3. The advantage of this assay is the use of one primer pair set versus the use of multiple primers.

To date there are no q-RT-PCR assays published for Omsk hemorrhagic fever or Kyasanur Forest disease. It would be interesting to determine if any of the published flavivirus q-RT-PCR assays would be capable of detecting Omsk hemorrhagic fever or Kyasanur Forest disease.

28.4 IMPLICATIONS FOR THERAPEUTICS, VACCINES, AND BIODEFENSE

HFV are a class of the most deadly viruses known to humans, with mortality rates approaching 90% in some Ebola outbreaks (Mason, 2008), and yet there are currently few effective therapies against them. This may be in part due to the fact that HFV are endemic in tropical and/or subtropical regions of the world, which have limited resources available for therapeutic development. In addition, outbreaks have been relatively infrequent, with a limited number of people becoming infected and quarantine of sick patients being effective in controlling epidemics. However, with the frequency of travel, the threat of infection in returning travelers is increasing (Hanna *et al.*, 2006; Schmitz *et al.*, 1996; Schwarz, 1996). In addition, since the terrorist attacks of September 11 in the USA, the public as well as the scientific community have an enhanced awareness of potential bioterrorism uses of HFV. In combination, this has led to the increased study of HFV and the search for effective therapeutics.

The only currently available vaccine for HFV is the live attenuated 17D vaccine for yellow fever virus, which has been available since the 1930s (Barnett *et al.*, 2008). Individuals traveling to endemic areas of Africa and South America are required to receive this vaccination (Barnett *et al.*, 2008). A number of vaccines have been approved as investigational new drugs (IND). Candid No. 1 is a live attenuated vaccine against Junin that was developed at USAMRIID (McKee *et al.*, 1992). This vaccine has been

proven to be safe and effective in protecting agricultural workers during a phase 3 clinical trial in South America (Maiztegui *et al.*, 1998). There are two IND vaccines against Rift Valley fever. One of which is a formalin-inactivated vaccine that is made available to at-risk laboratory workers and troops (Niklasson *et al.*, 1985; Pittman *et al.*, 1999). For Kyasanur Forest disease virus, there is also a formalin-inactivated vaccine. A field study performed in Shimoga, Uttar Kannada, and Chikmangalur districts during 1990–92 KFD epidemic seasons demonstrated a protective effect of vaccination among over 100,000 individuals vaccinated (Dandawate *et al.*, 1994).

There are several vaccine candidates in development for Lassa and filoviruses which have shown promising results in non-human primates. Virus-like particle (VLP) vaccines have shown especially effective protection in rhesus monkeys against lethal challenge with either Ebola or Marburg viruses (Bosio *et al.*, 2004; Swenson *et al.*, 2005; Warfield *et al.*, 2003, 2007a, b, 2005, 2004; Yang *et al.*, 2008; Ye *et al.*, 2006). The use of VLP is promising from a safety standpoint as they are non-infectious. A Lassa vaccine based on an attenuated recombinant vesicular stomatitis virus vector, which expresses Lassa GP, showed 100% protectivity against a lethal challenge of Lassa in cynomolgus monkeys (Geisbert *et al.*, 2005).

Currently, there are no licensed drug treatments for HFV, leaving supportive treatment as the main option for HFV diseases. Some studies have indicated that convalescent human plasma is effective in the treatment of arenaviral infections (Kilgore *et al.*, 1997). Ribavirin is a guanosine analog that displays very broad antiviral activities. It is effective both *in vitro* and *in vivo* against the arenaviruses, Hantaan virus, and possibly CCHF virus, but not against filoviruses or flaviviruses (Connor, 1990; Fisher-Hoch *et al.*, 1995; Huggins, 1989; Huggins *et al.*, 1991; McCormick *et al.*, 1986; McKee *et al.*, 1988; Rodriguez *et al.*, 1986; Weissenbacher *et al.*, 1986). However, ribavirin is only partially effective and is associated with negative side-effects (McKee *et al.*, 1988). Clearly, new therapeutic options for HFV are needed.

q-RT-PCR assays are an excellent option for large-scale drug/compound screening projects. There are real-time PCR instruments available that can process up to 384 samples simultaneously, with most instruments accommodating 96 well plates. To further aid RNA sample processing, one-step RT-PCR kits are available, which allow both the RT and PCR reactions to be performed in the same tube, eliminating the needed to produce cDNA before RNA samples can be processed by the real-time PCR instrument. Furthermore, RNA extraction reagents denature and inactivate viral RNA, allowing both extraction and q-RT-PCR analysis to be performed in a BSL-2 laboratory. This setup, as shown in Fig. 28.3, allows the efficient processing of samples in a high-throughput manner in a relatively short amount of time. This type of large-scale screening could never be performed accurately and in a timely fashion with plaque assays. In fact, two recent studies by the Centers for Disease Control have designed such a high-throughput q-RT-PCR detection method for Ebola and RVFV in clinical specimens for the application to outbreak settings (Bird *et al.*, 2007; Towner *et al.*, 2007b). However, there are currently no published studies on high-throughput compound screening, which would take advantage of this technology and hopefully lead to new therapeutic options.

It should be kept in mind that q-RT-PCR assays do not measure the infectivity of the virus, but rather only indicate the presence of a particular gene or genes of the virus. Assays that are designed to detect multiple viral genes, including the polymerase gene, can be suggestive of viral infectivity, but can never completely replace plaque assays. Plaque assays are an important assay to be used as a confirmation of results. This is reflected in the fact that viral titers measured through q-RT-PCR assays are 2 to 3 logs higher than viral titers measured through plaque assays (Garcia *et al.*, 2001). This is probably due to quantitation of non-infectious viral RNA by the q-RT-PCR assay. Therefore, in

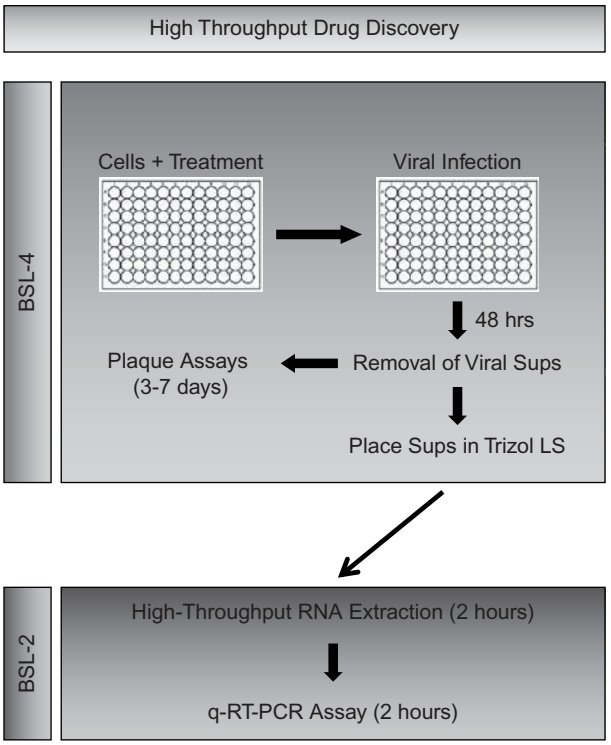


FIGURE 28.3 High-throughput drug discovery screening method. To achieve a high-throughput drug discovery screening method for HFV the depicted method can be utilized. *In vitro* cell culture samples can be pretreated with the compound of interest, infected with the appropriate virus, followed by post-treatment with the compound. After 48 hours, cellular supernatants containing released virus are collected and treated with RNA extraction buffer. Following RNA extraction buffer treatment, samples can be removed from BSL-4 laboratories and transferred to BSL-2 laboratories, where RNA extraction and q-RT-PCR assays will be completed. The entire turnaround time will be approximately three days versus up to two weeks if the effects of drug treatments were determined by plaque assays.

the case of large sample numbers, limited resources, and time constraints, q-RT-PCR assays have become standard, with plaque assays being used for confirmation of results when necessary.

28.5 CONCLUSIONS

HFV are deadly re-emerging pathogenic viruses with limited treatment options. To help identify new therapeutic options, compound screening should be implemented. Compound screening to identify therapeutics generally begins with *in vitro* assays where cells grown in cell culture are pretreated with the compound of interest, infected with the appropriate virus, followed by post-treated with the compound. Cellular supernatants containing released virus are then assayed to determine if the compound has decreased the viral titer as compared to an untreated control (Fig. 28.3). The standard virological assay to measure productive virus reproduction, the plaque assay, is both tedious and time consuming and must be performed in either BSL-3 or BSL-4 laboratories for HFV. Therefore, there is a need for an alternative to plaque assays that can be performed in a BSL-2 laboratory. q-RT-PCR technology is a molecular technique that has the ability to quantitate viral titers in lieu of traditional plaque assays. q-RT-PCR assays have been developed for almost all HFV and have proven to be sensitive, specific, and reliable. In addition, a high-throughput RNA extraction method for HFV has recently been described (Bird *et al.*, 2007; Towner *et al.*, 2007b) which when coupled with q-RT-PCR assays will enable high-throughput capabilities. This high-throughput screening method will streamline compound screening procedures against the notoriously difficult to work with VHF agents.

REFERENCES

Aitichou, M., Saleh, S.S., McElroy, A.K., Schmaljohn, C., and Ibrahim, M.S. (2005). Identification of Dobrava, Hantaan, Seoul, and Puumala viruses by one-step real-time RT-PCR. *J. Virol. Methods* 124, 21–26.

Bae, H.G., Drosten, C., Emmerich, P., Colebunders, R., Hantson, P., Pest, S., Parent, M., Schmitz, H., Warnat, M.A., and Niedrig, M. (2005). Analysis of two imported cases of yellow fever infection from Ivory Coast and The Gambia to Germany and Belgium. *J. Clin. Virol.* 33, 274–280.

Barnett, E.D., Wilder-Smith, A., and Wilson, M.E. (2008). Yellow fever vaccines and international travelers. *Expert Rev. Vaccines* 7, 579–587.

Bird, B.H., Bawiec, D.A., Ksiazek, T.G., Shoemaker, T.R., and Nichol, S. T. (2007). Highly sensitive and broadly reactive quantitative reverse transcription-PCR assay for high-throughput detection of Rift Valley fever virus. *J. Clin. Microbiol.* 45, 3506–3513.

Blow, J.A., Dohm, D.J., Negley, D.L., and Mores, C.N. (2004). Virus inactivation by nucleic acid extraction reagents. *J. Virol. Methods* 119, 195–198.

Borio, L., Inglesby, T., Peters, C.J., Schmaljohn, A.L., Hughes, J.M., Jahrling, P.B., Ksiazek, T., Johnson, K.M., Meyerhoff, A., O'Toole, T., Ascher, M.S., Bartlett, J., Breman, J.G., Eitzen, E.M., Jr., Hamburg, M., Hauer, J., Henderson, D.A., Johnson, R.T., Kwik, G., Layton, M., Lillibridge, S., Nabel, G.J., Osterholm, M.T., Perl, T.M., Russell, P., and Tonat, K. (2002). Hemorrhagic fever viruses as biological weapons: medical and public health management. *JAMA* 287, 2391–2405.

Bosio, C.M., Moore, B.D., Warfield, K.L., Ruthel, G., Mohamadzadeh, M., Aman, M.J., and Bavari, S. (2004). Ebola and Marburg virus-like particles activate human myeloid dendritic cells. *Virology* 326, 280–287.

Bowen, M.D., Rollin, P.E., Ksiazek, T.G., Hustad, H.L., Bausch, D.G., Demby, A.H., Bajani, M.D., Peters, C.J., and Nichol, S.T. (2000). Genetic diversity among Lassa virus strains. *J. Virol.* 74, 6992–7004.

Callahan, J.D., Wu, S.J., Dion-Schultz, A., Mangold, B.E., Peruski, L.F., Watts, D.M., Porter, K.R., Murphy, G.R., Suharyono, W., King, C.C., Hayes, C.G., and Temenak, J.J. (2001). Development and evaluation of serotype- and group-specific fluorogenic reverse transcriptase PCR (TaqMan) assays for dengue virus. *J. Clin. Microbiol.* 39, 4119–4124.

Chao, D.Y., Davis, B.S., and Chang, G.J. (2007). Development of multiplex real-time reverse transcriptase PCR assays for detecting eight medically important flaviviruses in mosquitoes. *J. Clin. Microbiol.* 45, 584–589.

Charrel, R.N., and de Lamballerie, X. (2003). Arenaviruses other than Lassa virus. *Antiviral Res.* 57, 89–100.

Christopherson, C., Sninsky, J., and Kwok, S. (1997). The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies. *Nucleic Acids Res.* 25, 654–658.

Chutinimitkul, S., Payungporn, S., Theamboonlers, A., and Poovorawan, Y. (2005). Dengue typing assay based on real-time PCR using I. *J. Virol. Methods* 129, 8–15.

Connor, J. (1990). Ribavirin pharmacokinetics. *Pediatr. Infect. Dis. J.* 9, S91–S92.

Dandawate, C.N., Desai, G.B., Achar, T.R., and Banerjee, K. (1994). Field evaluation of formalin inactivated Kyasanur forest disease virus tissue culture vaccine in three districts of Karnataka state. *Indian J. Med. Res.* 99, 152–158.

Demby, A.H., Chamberlain, J., Brown, D.W., and Clegg, C.S. (1994). Early diagnosis of Lassa fever by reverse transcription-PCR. *J. Clin. Microbiol.* 32, 2898–2903.

Deyde, V.M., Khristova, M.L., Rollin, P.E., Ksiazek, T.G., and Nichol, S. T. (2006). Crimean-Congo hemorrhagic fever virus genomics and global diversity. *J. Virol.* 80, 8834–8842.

Dos Santos, H.W., Poloni, T.R., Souza, K.P., Muller, V.D., Tremeschin, F., Nali, L.C., Fantinatti, L.R., Amarilla, A.A., Castro, H.L., Nunes, M. R., Casseb, S.M., Vasconcelos, P.F., Badra, S.J., Figueiredo, L.T., and Aquino, V.H. (2008). A simple one-step real-time RT-PCR for diagnosis of dengue virus infection. *J. Med. Virol.* 80, 1426–1433.

Drosten, C., Gottig, S., Schilling, S., Asper, M., Panning, M., Schmitz, H., and Gunther, S. (2002). Rapid detection and quantification of RNA of Ebola and Marburg viruses, Lassa virus, Crimean-Congo hemorrhagic fever virus, Rift Valley fever virus, dengue virus, and yellow fever virus by real-time reverse transcription-PCR. *J. Clin. Microbiol.* 40, 2323–2330.

Drosten, C., Kummerer, B.M., Schmitz, H., and Gunther, S. (2003). Molecular diagnostics of viral hemorrhagic fevers. *Antiviral Res.* 57, 61–87.

Duh, D., Saksida, A., Petrovec, M., Ahmeti, S., Dedushaj, I., Panning, M., Drosten, C., and Avsic-Zupanc, T. (2007). Viral load as predictor of Crimean-Congo hemorrhagic fever outcome. *Emerg. Infect. Dis.* 13, 1769–1772.

Duh, D., Saksida, A., Petrovec, M., Dedushaj, I., and Avsic-Zupanc, T. (2006). Novel one-step real-time RT-PCR assay for rapid and specific diagnosis of Crimean-Congo hemorrhagic fever encountered in the Balkans. *J. Virol. Methods* 133, 175–179.

Evander, M., Eriksson, I., Pettersson, L., Juto, P., Ahlm, C., Olsson, G.E., Bucht, G., and Allard, A. (2007). Puumala hantavirus viremia

diagnosed by real-time reverse transcriptase PCR using samples from patients with hemorrhagic fever and renal syndrome. *J. Clin. Microbiol.* 45, 2491–2497.

Fisher-Hoch, S.P., Khan, J.A., Rehman, S., Mirza, S., Khurshid, M., and McCormick, J.B. (1995). Crimean Congo-haemorrhagic fever treated with oral ribavirin. *Lancet* 346, 472–475.

Garcia, J.B., Morzunov, S.P., Levis, S., Rowe, J., Calderon, G., Enria, D., Sabattini, M., Buchmeier, M.J., Bowen, M.D., and St Jeor, S.C. (2000). Genetic diversity of the Junin virus in Argentina: geographic and temporal patterns. *Virology* 272, 127–136.

Garcia, S., Crance, J.M., Billecocq, A., Peinnequin, A., Jouan, A., Bouloy, M., and Garin, D. (2001). Quantitative real-time PCR detection of Rift Valley fever virus and its application to evaluation of antiviral compounds. *J. Clin. Microbiol.* 39, 4456–4461.

Garin, D., Peyrefitte, C., Crance, J.M., Le Faou, A., Jouan, A., and Bouloy, M. (2001). Highly sensitive Taqman PCR detection of Puumala hantavirus. *Microbes Infect* 3, 739–745.

Garrison, A.R., Alakbarova, S., Kulesh, D.A., Shezmukhamedova, D., Khodjaev, S., Endy, T.P., and Paragas, J. (2007). Development of a TaqMan minor groove binding protein assay for the detection and quantification of Crimean-Congo hemorrhagic fever virus. *Am. J. Trop. Med. Hyg.* 77, 514–520.

Geisbert, T.W., Jones, S., Fritz, E.A., Shurtleff, A.C., Geisbert, J.B., Liebscher, R., Grolla, A., Stroher, U., Fernando, L., Daddario, K.M., Guttieri, M.C., Mothe, B.R., Larsen, T., Hensley, L.E., Jahrling, P.B., and Feldmann, H. (2005). Development of a new vaccine for the prevention of Lassa fever. *PLoS Med.* 2, e183.

Gibb, T.R., Norwood, D.A., Jr., Woollen, N., and Henchal, E.A. (2001a). Development and evaluation of a fluorogenic 5′ nuclease assay to detect and differentiate between Ebola virus subtypes Zaire and Sudan. *J. Clin. Microbiol.* 39, 4125–4130.

Gibb, T.R., Norwood, D.A., Jr., Woollen, N., and Henchal, E.A. (2001b). Development and evaluation of a fluorogenic 5′-nuclease assay to identify Marburg virus. *Mol. Cell Probes,* 15, 259–266.

Gritsun, T.S., Nuttall, P.A., and Gould, E.A. (2003). Tick-borne flaviviruses. *Adv. Virus Res.* 61, 317–371.

Gubler, D.J. (1998a). Dengue and dengue hemorrhagic fever. *Clin. Microbiol. Rev.* 11, 480–496.

Gubler, D.J. (1998b). The global pandemic of dengue/dengue haemorrhagic fever: current status and prospects for the future. *Ann. Acad. Med. Singapore* 27, 227–234.

H, J., H, C., A, K., and R, Y. (2008). Real-time rt-PCR of hantaan virus RNA used for the detection of virus response to antiviral drugs. *Acta Virol.* 52, 67–70.

Hanna, J.N., Ritchie, S.A., Richards, A.R., Taylor, C.T., Pyke, A.T., Montgomery, B.L., Piispanen, J.P., Morgan, A.K., and Humphreys, J.L. (2006). Multiple outbreaks of dengue serotype 2 in north Queensland, 2003/04. *Aust. NZ J. Public Health* 30, 220–225.

Holland, J., and Domingo, E. (1998). Origin and evolution of viruses. *Virus Genes* 16, 13–21.

Hoogstraal, H. (1979). The epidemiology of tick-borne Crimean-Congo hemorrhagic fever in Asia, Europe, and Africa. *J. Med. Entomol.* 15, 307–417.

Houng, H.S., Chung-Ming Chen, R., Vaughn, D.W., and Kanesa-thasan, N. (2001). Development of a fluorogenic RT-PCR system for quantitative identification of dengue virus serotypes 1–4 using conserved and serotype-specific 3′ noncoding sequences. *J. Virol. Methods,* 95, 19–32.

Huggins, J.W. (1989). Prospects for treatment of viral hemorrhagic fevers with ribavirin, a broad-spectrum antiviral drug. *Rev. Infect. Dis.* 11(4), S750–S761.

Huggins, J.W., Hsiang, C.M., Cosgriff, T.M., Guang, M.Y., Smith, J.I., Wu, Z.O., LeDuc, J.W., Zheng, Z.M., Meegan, J.M., Wang, Q.N. *et al.* (1991). Prospective, double-blind, concurrent, placebo-controlled clinical trial of intravenous ribavirin therapy of hemorrhagic fever with renal syndrome. *J. Infect. Dis.* 164, 1119–1127.

Jakab, F., Sebok, J., Ferenczi, E., Horvath, G., and Szucs, G. (2007). First detection of Dobrava hantavirus from a patient with severe haemorrhagic fever with renal syndrome by SYBR Green-based real time RT-PCR. *Scand. J. Infect. Dis.* 39, 902–906.

Kilgore, P.E., Ksiazek, T.G., Rollin, P.E., Mills, J.N., Villagra, M.R., Montenegro, M.J., Costales, M.A., Paredes, L.C., and Peters, C.J. (1997). Treatment of Bolivian hemorrhagic fever with intravenous ribavirin. *Clin. Infect. Dis.* 24, 718–722.

Kramski, M., Meisel, H., Klempa, B., Kruger, D.H., Pauli, G., and Nitsche, A. (2007). Detection and typing of human pathogenic hantaviruses by real-time reverse transcription-PCR and pyrosequencing. *Clin. Chem.* 53, 1899–1905.

Kwok, S., Kellogg, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C., and Sninsky, J.J. (1990). Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.* 18, 999–1005.

Lanciotti, R.S. (2003). Molecular amplification assays for the detection of flaviviruses. *Adv. Virus Res.* 61, 67–99.

Leroy, E.M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassanin, A., Yaba, P., Delicat, A., Paweska, J.T., Gonzalez, J.P., and Swanepoel, R. (2005). Fruit bats as reservoirs of Ebola virus. *Nature* 438, 575–576.

Mackay, I.M., Arde, K.E., and Nitsche, A. (2002). Real-time PCR in virology. *Nucleic Acids Res.* 30, 1292–1305.

Madani, T.A., Al-Mazrou, Y.Y., Al-Jeffri, M.H., Mishkhas, A.A., Al-Rabeah, A.M., Turkistani, A.M., Al-Sayed, M.O., Abodahish, A.A., Khan, A.S., Ksiazek, T.G., and Shobokshi, O. (2003). Rift Valley fever epidemic in Saudi Arabia: epidemiological, clinical, and laboratory characteristics. *Clin. Infect. Dis.* 37, 1084–1092.

Maiztegui, J.I., McKee, K.T., Jr., Barrera Oro, J.G., Harrison, L.H., Gibbs, P.H., Feuillade, M.R., Enria, D.A., Briggiler, A.M., Levis, S. C., Ambrosio, A.M., Halsey, N.A., and Peters, C.J. (1998). Protective efficacy of a live attenuated vaccine against Argentine hemorrhagic fever. AHF Study Group. *J. Infect. Dis.* 177, 277–283.

Mardani, M., and Keshtkar-Jahromi, M. (2007). Crimean-Congo hemorrhagic fever Arch. *Iran. Med.* 10, 204–214.

Mason, C. (2008). The strains of Ebola. *CMAJ* 178, 1266–1267.

McCormick, J.B., King, I.J., Webb, P.A., Scribner, C.L., Craven, R.B., Johnson, K.M., Elliott, L.H., and Belmont-Williams, R. (1986). Lassa fever. Effective therapy with ribavirin. *N. Engl. J. Med.* 314, 20–26.

McKee, K.T., Jr., Huggins, J.W., Trahan, C.J., and Mahlandt, B.G. (1988). Ribavirin prophylaxis and therapy for experimental argentine hemorrhagic fever. *Antimicrob. Agents Chemother.* 32, 1304–1309.

McKee, K.T., Jr., Oro, J.G., Kuehne, A.I., Spisso, J.A., and Mahlandt, B. G. (1992). Candid No. 1 Argentine hemorrhagic fever vaccine protects against lethal Junin virus challenge in rhesus macaques. *Intervirology* 34, 154–163.

Moe, J.B., Lambert, R.D., and Lupton, H.W. (1981). Plaque assay for Ebola virus. *J. Clin. Microbiol.* 13, 791–793.

Monath, T.P. (2001). Yellow fever: an update. *Lancet Infect. Dis.* 1, 11–20.

Muranyi, W., Bahr, U., Zeier, M., and van der Woude, F.J. (2005). Hantavirus infection. *J. Am. Soc. Nephrol.* 16, 3669–3679.

Niklasson, B., Peters, C.J., Bengtsson, E., and Norrby, E. (1985). Rift Valley fever virus vaccine trial: study of neutralizing antibody response in humans. *Vaccine* 3, 123–127.

Panning, M., Laue, T., Olschlager, S., Eickmann, M., Becker, S., Raith, S., Courbot, M.C., Nilsson, M., Gopal, R., Lundkvist, A., Caro, A.,

Brown, D., Meyer, H., Lloyd, G., Kummerer, B.M., Gunther, S., and Drosten, C. (2007). Diagnostic reverse-transcription polymerase chain reaction kit for filoviruses based on the strain collections of all European biosafety level 4 laboratories. *J. Infect. Dis.* 196(2), S199–S204.

Pattnaik, P. (2006). Kyasanur forest disease: an epidemiological view in India. *Rev. Med. Virol.* 16, 151–165.

Peters, I.R., Helps, C.R., Hall, E.J., and Day, M.J. (2004). Real-time RT-PCR: considerations for efficient and sensitive assay design. *J. Immunol. Methods* 286, 203–217.

Pittman, P.R., Liu, C.T., Cannon, T.L., Makuch, R.S., Mangiafico, J.A., Gibbs, P.H., and Peters, C.J. (1999). Immunogenicity of an inactivated Rift Valley fever vaccine in humans: a 12-year experience. *Vaccine* 18, 181–189.

Rodriguez, M., McCormick, J.B., and Weissenbacher, M.C. (1986). Antiviral effect of ribavirin on Junin virus replication in vitro. *Rev. Argent. Microbiol.* 18, 69–74.

Roth, W.K., Weber, M., and Seifried, E. (1999). Feasibility and efficacy of routine PCR screening of blood donations for hepatitis C virus, hepatitis B virus, and HIV-1 in a blood-bank setting. *Lancet* 353, 359–363.

Saksida, A., Duh, D., Korva, M., and Avsic-Zupanc, T. (2008). Dobrava virus RNA load in patients who have hemorrhagic fever with renal syndrome. *J. Infect. Dis.* 197, 681–685.

Sanchez, A., Ksiazek, T.G., Rollin, P.E., Miranda, M.E., Trappier, S. G., Khan, A.S., Peters, C.J., and Nichol, S.T. (1999). Detection and molecular characterization of Ebola viruses causing disease in human and nonhuman primates. *J. Infect. Dis.* 179(1), S164–S169.

Schmitz, H., Emmerich, P., and ter Meulen, J. (1996). Imported tropical virus infections in Germany. *Arch. Virol.*(11), 67–74.

Schwarz, T.F. (1996). Imported vector- and rodent-borne virus infections – an introduction. *Arch. Virol. Suppl.* 11, 3–11.

Swanepoel, R., Smit, S.B., Rollin, P.E., Formenty, P., Leman, P.A., Kemp, A., Burt, F.J., Grobbelaar, A.A., Croft, J., Bausch, D.G., Zeller, H., Leirs, H., Braack, L.E., Libande, M.L., Zaki, S., Nichol, S.T., Ksiazek, T.G., and Paweska, J.T. (2007). Studies of reservoir hosts for Marburg virus. *Emerg. Infect. Dis.* 13, 1847–1851.

Swenson, D.L., Warfield, K.L., Negley, D.L., Schmaljohn, A., Aman, M.J., and Bavari, S. (2005). Virus-like particles exhibit potential as a pan-filovirus vaccine for both Ebola and Marburg viral infections. *Vaccine* 23, 3033–3042.

Towner, J.S., Khristova, M.L., Sealy, T.K., Vincent, M.J., Erickson, B. R., Bawiec, D.A., Hartman, A.L., Comer, J.A., Zaki, S.R., Stroher, U., Gomes da Silva, F., del Castillo, F., Rollin, P.E., Ksiazek, T.G., and Nichol, S.T. (2006). Marburgvirus genomics and association with a large hemorrhagic fever outbreak in Angola. *J. Virol.* 80, 6497–6516.

Towner, J.S., Pourrut, X., Albarino, C.G., Nkogue, C.N., Bird, B.H., Grard, G., Ksiazek, T.G., Gonzalez, J.P., Nichol, S.T., and Leroy, E. M. (2007a). Marburg virus infection detected in a common African bat. *PLoS One* 2, e764.

Towner, J.S., Sealy, T.K., Ksiazek, T.G., and Nichol, S.T. (2007b). High-throughput molecular detection of hemorrhagic fever virus threats with applications for outbreak settings. *J. Infect. Dis.* 196(2), S205–S212.

Vieth, S., Drosten, C., Charrel, R., Feldmann, H., and Gunther, S. (2005). Establishment of conventional and fluorescence resonance energy transfer-based real-time PCR assays for detection of pathogenic New World arenaviruses. *J. Clin. Virol.* 32, 229–235.

Vieth, S., Drosten, C., Lenz, O., Vincent, M., Omilabu, S., Hass, M., Becker-Ziaja, B., ter Meulen, J., Nichol, S.T., Schmitz, H., and Gunther, S. (2007). RT-PCR assay for detection of Lassa virus and related Old World arenaviruses targeting the L gene. *Trans. R. Soc. Trop. Med. Hyg.* 101, 1253–1264.

Wacker, M.J., and Godard, M.P. (2005). Analysis of one-step and two-step real-time RT-PCR using SuperScript III. *J. Biomol. Tech.* 16, 266–271.

Walker, N.J. (2002). Tech.Sight. A technique whose time has come. *Science* 296, 557–559.

Warfield, K.L., Bosio, C.M., Welcher, B.C., Deal, E.M., Mohamadzadeh, M., Schmaljohn, A., Aman, M.J., and Bavari, S. (2003). Ebola virus-like particles protect from lethal Ebola virus infection. *Proc. Natl. Acad. Sci. USA* 100, 15889–15894.

Warfield, K.L., Posten, N.A., Swenson, D.L., Olinger, G.G., Esposito, D., Gillette, W.K., Hopkins, R.F., Costantino, J., Panchal, R.G., Hartley, J.L., Aman, M.J., and Bavari, S. (2007a). Filovirus-like particles produced in insect cells: immunogenicity and protection in rodents. *J. Infect. Dis.* 196(2), S421–S429.

Warfield, K.L., Swenson, D.L., Demmin, G., and Bavari, S. (2005). Filovirus-like particles as vaccines and discovery tools. *Expert Rev. Vaccines* 4, 429–440.

Warfield, K.L., Swenson, D.L., Negley, D.L., Schmaljohn, A.L., Aman, M. J., and Bavari, S. (2004). Marburg virus-like particles protect guinea pigs from lethal Marburg virus infection. *Vaccine* 22, 3495–3502.

Warfield, K.L., Swenson, D.L., Olinger, G.G., Kalina, W.V., Aman, M. J., and Bavari, S. (2007b). Ebola virus-like particle-based vaccine protects nonhuman primates against lethal Ebola virus challenge. *J. Infect. Dis.* 196(2), S430–S437.

Weidmann, M., Hufert, F.T., and Sall, A.A. (2007). Viral load among patients infected with Marburgvirus in Angola. *J. Clin. Virol.* 39, 65–66.

Weissenbacher, M.C., Avila, M.M., Calello, M.A., Merani, M.S., McCormick, J.B., and Rodriguez, M. (1986). Effect of ribavirin and immune serum on Junin virus-infected primates. *Med. Microbiol. Immunol.* 175, 183–186.

Wilkening, S., and Bader, A. (2004). Quantitative real-time polymerase chain reaction: methodical analysis and mathematical model. *J. Biomol. Tech.* 15, 107–111.

Wolfel, R., Paweska, J.T., Petersen, N., Grobbelaar, A.A., Leman, P. A., Hewson, R., Georges-Courbot, M.C., Papa, A., Gunther, S., and Drosten, C. (2007). Virus detection and monitoring of viral load in Crimean-Congo hemorrhagic fever virus patients. *Emerg. Infect. Dis.* 13, 1097–1100.

Wong, M.L., and Medrano, J.F. (2005). Real-time PCR for mRNA quantitation. *Biotechniques* 39, 75–85.

Yang, C., Ye, L., and Compans, R.W. (2008). Protection against filovirus infection: virus-like particle vaccines. *Expert Rev. Vaccines* 7, 333–344.

Yapar, M., Aydogan, H., Pahsa, A., Besirbellioglu, B.A., Bodur, H., Basustaoglu, A.C., Guney, C., Avci, I.Y., Sener, K., Setteh, M.H., and Kubar, A. (2005). Rapid and quantitative detection of Crimean-Congo hemorrhagic fever virus by one-step real-time reverse transcriptase-PCR. *Jpn. J. Infect. Dis.* 58, 358–362.

Ye, L., Lin, J., Sun, Y., Bennouna, S., Lo, M., Wu, Q., Bu, Z., Pulendran, B., Compans, R.W., and Yang, C. (2006). Ebola virus-like particles produced in insect cells exhibit dendritic cell stimulating activity and induce neutralizing antibodies. *Virology* 351, 260–270.

Yong, Y.K., Thayan, R., Chong, H.T., Tan, C.T., and Sekaran, S.D. (2007). Rapid detection and serotyping of dengue virus by multiplex RT-PCR and real-time SYBR green RT-PCR. *Singapore Med. J.* 48, 628–662.

This page intentionally left blank

Chapter 29

Identification of Genetically Modified Organisms

Farid E. Ahmed
Department of Radiation Oncology, Leo W. Jenkins Cancer Center;
The Brody School of Medicine, LSB 014, East Carolina University, Greenville, NC, USA

29.1 INTRODUCTION AND HISTORICAL PERSPECTIVE

Adopting modern biotechnology, including genetic transformation, has developed new plant varieties. Genetically modified (GM) products contain an additional trait encoded by an introduced gene(s), which generally produces a protein(s) that confers the trait of interest. Raw material (e.g. grains) and processed products (e.g. foods) derived from GM crops might thus be identified by testing for the presence of introduced DNA, or by detection of expressed novel protein(s) encoded by the genetic material. Both qualitative (i.e. those that give a yes/no answer) and quantitative diagnostic methods are available (Ahmed, 2002).

Most developed countries over the last few years have established mechanisms for adjudicating on the safety of novel food before it is marketed (Moseley, 1999). GM foods have not gained worldwide acceptance because of unmollified consumer suspicion resulting from earlier food and environmental concerns, transparent regulatory oversight, and mistrust in government bureaucracies, all factors that fueled debates about the environmental and public health safety issues of introduced genes; for example, potential gene flow to other organisms, the destruction of agricultural diversity, allergenicity, antibiotic resistance, gastrointestinal problems (Gaskell *et al*., 1999; Hasslberger, 2000). Social acceptance of such novel foods or ingredients is not uniform in developed countries. Consumer concerns can be based on ethical considerations (i.e. scientists playing God) or safety worries (i.e. more testing needs to be done; Moseley, 1999). Other economical and ethical issues pertaining to intellectual property rights came into play (Serageldin, 1999) with the realization that inadvertent contamination of non-GM seeds with genetically modified organisms (GMOs) is likely. These factors induced countries, exemplified by the European Union (EU), either to restrict the import of bioengineered foods or to introduce legislation requiring mandatory labeling of GMO foods or food ingredients containing additives and flavorings that have been genetically modified or have been produced from GMOs (Commission Regulation, 2000a, b). EU regulations mandate labeling of food containing GMOs (Council Regulation, 1997, 1998). Norway and Switzerland, which are not members of the EU, demand the labeling of GMOs in their food (Hardegger *et al*., 1999). Moreover, in 1998, the EU introduced a *de facto* moratorium on the import and production of GM foods. In March 2003, the European Commission upheld the moratorium and is standing firm on its decision that any food containing more than 0.9% of a GM product would carry a label (Hellemans, 2003). In the USA, recent legislation did not stipulate mandatory labeling of GM foods but has instead recommended a voluntary labeling of bioengineered foods and requested that companies must notify the US Food and Drug Administration (FDA) of their intent to market GM foods at least 120 days in advance of launch (Editorial, 2000).

It appears that insufficient attention has been given to the following issues:

- The introduction of the same gene into different types of cells can produce distinct proteins.
- The introduction of any gene (either from the same or different species) can significantly change overall gene expression, and thus the phenotype of the recipient cells.
- Enzymatic pathways introduced to synthesize micronutrients may interact with endogenous pathways leading to production of novel metabolically active molecules.

Consequently, as with secondary modifications, it is possible that any or all of these perturbations may result in unpredictable outcome (Shubert, 2002). Although short- and

long-term toxicity and metabolic studies could be carried out to address concerns, it is probably not feasible that they will detect relevant changes unless extensive safety testing is carried out on GM crops, and more importantly provided the effect(s) in question is known so that it can be targeted and carefully studied.

Assessing the risk that a GM food poses to human health has been based on comparing its chemical composition to that of unmodified food; if the composition is similar, it is considered safe for human consumption. This concept, called substantial equivalence, has been based on relatively easy and inexpensive tests (Kuiper *et al*., 2001). However, some scientists in Europe criticize the methodologies currently used and contend that such tests do not include all biological, toxicological, and immunological aspects of GM food (Hellemans, 2003).

The first generation of ''input traits'' of GM crops (e.g. traits with purely agronomic benefits) is entering its ninth year, with the large majority derived from these leading crops in North America, namely canola (rapeseed), corn, and soybean. These products harbor traits that serve an agronomic purpose (i.e. benefiting farmers, but not necessarily the consumers). They entered the North American market with minimum regulations and without segregation, and have been judged by regulators as substantially equivalent to existing varieties. Second generation crops, which involve output modifications (traits with health and nutritional benefits; Agius *et al*., 2003), most likely will not be cleared unless their purity is assured; this is a problematic prospect given the current difficulties of attaining gene containment.

Third generation crops with new industrial, nutriceutical, or pharmaceutical properties most likely will require an effective gene control system so that they may be allowed to enter the markets (Kleter *et al*., 2001). Regardless of how effective regulations are, some producers (either deliberately or inadvertently) will misappropriate these technologies, creating risks and liabilities. Moreover, many plant species are sexually promiscuous, creating natural gene flow to related species, and leading to the following liability issues:

- The potential of volunteer seeds inadvertently left in the field to germinate the following year(s)
- The potential for pollen flown from GM crops to non-GM crops
- The potential of co-mingling of GM and non-GM crops, which could jeopardize the value of both crops and product lines if transgenes remain undetected before processing
- The potential for environmental risks associated with uncontrolled gene flow from GM varieties into related plants, which will impede export of GM varieties to countries not willing to adopt the new technologies

These liability issues have resulted in some disastrous consequences and have imposed significant cost on the food industry (Smyth *et al*., 2002). About 36 countries have legislation making it imperative that governments, the food industry, testing laboratories, and crop producers develop ways to accurately quantitate GMOs in crops, foods, and food ingredients to assure compliance with threshold levels of GM products (Butler and Reichardt, 1999; Kuiper, 1999; US National Academy of Sciences, 2000). The aim of this chapter is to detail diagnostic test methods, their potential, and their limitations.

29.2 SAMPLING PLANS

Both sample size and sampling procedures are important issues for testing GMOs in raw material and food ingredients if one is to avoid problems of non-homogeneity. The sampling plan should be performed in a manner that ensures that the sample is statistically representative, and the sample size must be sufficient to allow adequate sensitivity, because the statistical significance achievable with a small sample size is weak (Gilbert, 1999). When a sample is used to represent the content of a lot, its content is likely to deviate from the actual content of the lot. Sampling errors create risks for both the buyer and the seller in a transaction. The buyer may get a lot with a higher concentration than desired. The seller may also have a lot that meets a contract rejected. These two types of risks are referred to as seller's and buyer's risk, respectively, and the relationship between them is seen in Fig. 29.1. Probability can be used to estimate a likely range that a sample may deviate from the true lot content, as long as the sample is properly taken (Whitaker *et al*., 2001).

Results from theoretical and simulation research have shown that GMO distribution in bulk commodities is more likely to be heterogeneous than homogeneous, which poses a serious limit to the unconditional acceptance of the assumption of random distribution of GM material and to the use of binomial distribution to estimate producer and consumer risks (Paoletti *et al*., 2003). Due to the heterogenous distribution of biotech seeds (or beans or kernels)

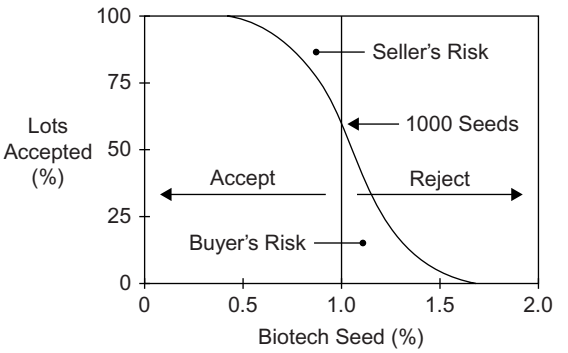


FIGURE 29.1 Operating characteristic curve illustrating buyer's and seller's risk. From Whitaker and colleagues (2001); with permission.

within a load, there is a probability that a sample from the load will contain a concentration that is higher or lower than the true concentration. Using a sample size of 1,000 seeds, the probability of accepting a load containing various percentages of biotechs is defined by the operating characteristic (OC) curve shown in Fig. 29.1. The buyer's and seller's risk for a load containing a true concentration of 1% biotech units, randomly distributed in a load, is indicated by the area of the figure between the QC curve and the line representing 1%. Using these statistic principles to manage marketing risks, different sampling and testing strategies have been developed to meet the demand of the market.

In situations where biotech grains have been approved by government authorities, appreciable concentration of biotech grains in a load may be tolerated. In this case, it is possible to imagine a transaction where the buyer wants high confidence that the beans being purchased contain less than 1% Roundup Ready™ (RR) soybeans to avoid labeling. At the same time, the seller wants reasonable confidence that if the load contains less than 1% RR, there is a high probability that the load will be accepted. Figure 29.2 illustrates an example of a sampling and testing protocol, called threshold testing, that uses lateral flow strips for qualitative determination of novel protein, which could be useful under such a scenario to provide acceptable assurances to both parties (Remund *et al*., 2001). A limitation of the use of this technology is the difficulty of testing all available biotech events using a single sample. To use the threshold testing protocol, it is necessary to limit the maximum number of kernels in a sample so that the presence of only a single biotech kernel will always give a positive response. When performing threshold testing for multiple biotech events in a single sample, it is necessary to limit the number of kernels so that the lowest expressing event will always be detected, and such a constraint significantly limits the sensitivity of the methods for high expressing events. This constraint on sensitivity, and the fact that there is a large and ever-changing list of biotech events, has curtailed efforts to develop methods to detect multiple biotech events in a single sample of corn grain (Stave, 2004).

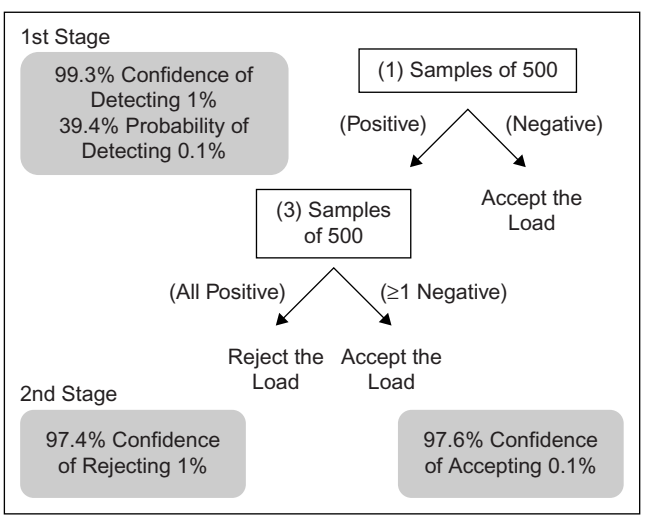


FIGURE 29.2 Threshold testing protocol providing high confidence that 1% will be rejected and 0.1% will be accepted. From Remund and colleagues (2001); with permission.

As raw material often comes from different suppliers, and given that industrial activities are structured in space and time, one can expect a portion of the original chronological order always to be present in the spatial structure of any lot. Under this assumption, a systematic sampling approach is recommended over a random sampling approach. Systematic sampling should take place when lots are being unloaded (i.e. when there is the option of continuous sampling of the entire consignment). This is preferable to the situation of large batches in silos or trucks, where it is difficult to access remote parts even when employing sampling probes (Paoletti *et al*., 2003).

It is difficult to make clear recommendations on the number of increments used to produce the bulk sample, because the number of increments required to minimize the sampling error will depend upon the heterogeneity of the lot under investigation. The lack of data on the expected distributions of real lots makes it impossible to establish objective criteria to address this problem. When defining the number of increments to be sampled, it should be taken into account that even modest levels of heterogeneity will compromise sampling reliability when 30 to 50 increments are used to produce the bulk sample (Paoletti *et al*., 2003).

All the sampling steps necessary to produce final samples of suitable working size from large bulk samples are needed. However, assuming random distribution to estimate the errors associated with each of these secondary sampling steps is not going to pose a problem as long as grinding and mixing of the material is properly carried out. Sampling size reduction should be attempted only when all the sampled material is reduced to the same particle size as the smaller, and the more uniform this is after milling, the more successful the mixing will be in ensuring homogeneity in the population, and ultimately in minimizing sampling error. Sampling size is selected to best meet the needs of the buyer and seller, and often involves a compromise between precision and cost (Paoletti *et al*., 2003).

29.3 CERTIFIED REFERENCE MATERIAL

Appropriate reference materials for positive and negative controls provide the basis for the validation of analytical procedures and for assessing the performance of methods and laboratories. Reference material should be independent of the analytical methods and should be focused on raw material or base ingredients rather than on finished foods (Ahmed, 2002). Various types of material can be used as a certified reference material (CRM) for the detection of

GMOs: matrix-based material (produced from seeds), pure DNA, or protein standards. Pure DNA standards can consist of either genomic DNA extracted from plant leaves or plasmid DNA in which a short sequence of a few hundred base pairs (bp) has been cloned and amplified in a suitable vector. Pure protein can be extracted from ground seed, or produced by recombinant DNA technology. Different requirements have to be met by CRM depending on their use for identification and/or quantitation of GMOs, and each has advantages and disadvantages (see Table 29.1). In contrast to protein detection methods, in which a single standard can be settled on relatively easy, DNA-based methods are better served through combinations of several positive controls. The availability of reference materials is currently limited owing to concerns over intellectual property rights and costs (Serageldin, 1999). The Institute of Reference Materials and Measurements at the Joint Research Center in Geel, Belgium, offer through Fluka (Buchs, Switzerland) a limited number of reference materials for modified soya, corn, and maximizer maize (MM) (Ahmed, 2002).

29.4 PROTEIN-BASED TESTING METHODS

Immunoassay technologies using antibodies are ideal for qualitative and quantitative detection of many types of proteins in complex matrices when the target analyte is known (Brett *et al*., 1999). Both monoclonal (highly specific) and polyclonal (often more sensitive) antibodies can be used depending on the amounts needed and the specificity of the detection system (e.g. antibodies to whole protein, or to specific peptide sequences) depending on the particular application, time allotted for testing, and cost. On the basis of typical concentrations of transgenic material in plant tissue (>10 mg/tissue), the detection limits of protein immunoassays can predict the presence of modified proteins in the range of 1% GMOs (Stave, 1999). Immunoassays in which antibodies are attached to a solid phase have been used in two formats:

- A competitive assay in which the detector and analyte compete to bind with capture antibodies
- A two-site (double antibody sandwich) assay in which target analyte sandwich between the capture antibody and the detector antibody; this assay is deemed preferable

29.4.1 Western Blot

The Western blot is a highly specific method that provides qualitative results suitable for determining whether a sample contains the target protein below or above a

TABLE 29.1 Advantages and disadvantages of various types of GMO CRMs. Adapted from Trapmann and colleagues (2004); with permission.

Type of CRM	Advantages	Disadvantages
Matrix GMO CRMs	• suitable for protein- and DNA-based methods	• different extractability (?)
	• extraction covered	• large production needed
	• commutability	• low availability of raw material due to restricted use of seeds
		• degradation
		• variation of the genetic background
Genomic DNA CRMs	• good calibrant	• large production needed
	• fewer seeds needed	• low availability of raw material due to restricted use of seeds
	• commutability	• variation of the genetic background
		• long-term stability (?)
Pure protein DNA CRMs	• fewer seeds needed	• commutability (?)
Plasmidic DNA CRMs	• easy to produce in large quantities	• plasmid topology
	• broad dynamic range	• discrepancies
		• commutability (?)

predetermined threshold level (Lipton *et al*., 2000), and is particularly useful for the analysis of insoluble protein (Brett *et al*., 1999). Further, because electrophoretic separation of protein is carried out under denaturing conditions, any problems of solubilization, aggregation, and coprecipitation of the target protein with adventitious proteins are eliminated (Rogan *et al*., 1999; Sambrook and Russel, 2000). The detection limits of Western blots vary between 0.25% for seeds and 1% for toasted meal (Smyth *et al*., 2002). This method, even though it can provide quantitative results and is sensitive, is considered more suited to research applications than to routine testing because it is not amenable to automation. Western blotting generally takes about two days and costs about $150/sample (Ahmed, 2002).

29.4.2 ELISA

ELISA assumes more than one format: a microwell plate (or strip) format, and a coated tube format. The antibody-coated microwell strips, with removable strips of 8 to 12 wells, are quantitative, highly sensitive, economical, provide high throughput, and are ideal for quantitative high-volume laboratory analysis, provided that the protein is not denatured. The typical run time for a plate assay is 90 min, and an optical plate reader determines concentration levels in the samples. Detection limits for CP4 EPSPS soybean protein was 0.25% for seeds and 1.4% for toasted meal. The antibody-coated tube format is suited for field testing, with typical run times ranging from 15 to 30 min, and tubes can be read either visually or by an optical tube reader; results are qualitative. Because there is no quantitative internal standard within the assay, no extra information can be obtained concerning the presence of GMO at the ingredient level in food. The ELISA test generally takes about 1 to 2 h, and costs about $5.0/sample on the average (Ahmed, 2002).

29.4.3 Lateral Flow Strip

A lateral flow strip is a single unit device that allows for manual testing of individual samples. Each nitrocellulose strip consists of three components: a reservoir pad on which an antibody coupled to a colored particle such as colloidal gold or latex is deposited, a result window, and a filter cover (see Fig. 29.3). An analyte-specific capture antibody (Ab) is also immobilized on the strip. Inserting the strip in an eppendorf vial containing an extract from plant tissue solution harboring a transgenic protein, leads to the solution moving toward the reservoir pad solubilizing the reported Ab, which binds to target analyte and forms an analyte–Ab complex that flows with the liquid sample laterally along the surface of the strip. When the complex passes over the zone where the capture Ab has been immobilized, it binds to the Ab and produces a colored band. The presence of two bands indicates a positive test for the protein of interest. A single band indicates that the test was performed correctly, but that there was no protein of interest present in the sample. This test provides a yes/no or threshold (semi-quantitative) determination of the target protein in about 10 min at an average cost of about $2/sample, is appropriate for field or on-site applications, and does not require specialized equipment (Ahmed, 2002). Commercially available lateral flow strips are currently limited to few biotechnology derived protein-producing GMO products, but strips that can detect multiple proteins simultaneously are being developed (Smyth *et al*., 2002).

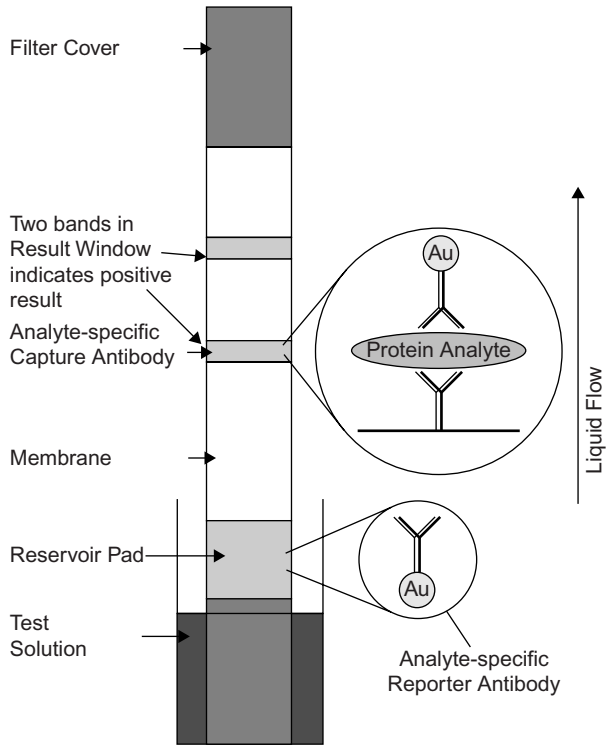


FIGURE 29.3 A schematic view of lateral flow strip assay format illustrating the principles of the assay. From Lipton and colleagues (2000); with permission.

29.4.4 Other Immunoassay Formats

In addition to microplate ELISA and later flow devices, other immunoassay formats use magnetic particles as the solid support surface. The magnetic particles can be coated with the capture antibody and the reaction carried out in a test tube. The particles with bound reactants are separated from unbound reactants in solution using a magnet. Advantages of this format are superior kinetics because the particles are free to move in reaction solutions, and increased precision owing to uniformity of the particles. Other less commonly used formats utilize nesting or a combination of two steps in one (Brett *et al*., 1999; Stave, 1999; Lipton *et al*., 2000; Ahmed, 2002). In the near future, improvements in immunoassays are expected to occur via

advances in antibody technology and improved instruments (Smyth *et al*., 2002). Recent advances in proteomics have implications for GMO diagnosis. However, their sensitivity has to be increased, and instrument cost has to be brought down (Ahmed, 2004).

29.5 DNA-Based Testing Methods

Molecular diagnostic methods for GMOs detect and quantify those DNA sequences that have been introduced into the organism during the process of gene modification. The DNA that has been engineered into a crop consists of several elements that govern its functioning. They are typically a promoter sequence, structural gene, and a stop sequence for the gene. Although several techniques are available, two are commonly used: Southern blot, and, particularly, PCR analyses. Microarray-based and real-time biosensor technologies have recently also been applied for diagnosing GMOs.

29.5.1 Southern Blot

The method involves fixing isolated sample DNA onto nitrocellulose or nylon membranes, probing with double-stranded (ds)-labeled nucleic acid probe(s) specific to the organisms whose diagnosis is desired, and detecting hybridization radiographically, fluoremetrically, or by chemiluminescence. Earlier probes were labeled with 32P. However, non-radioactive probe labeling methods employing the indirect or the direct approach have been developed to allow convenient diagnosis and avoid use of radioactive isotopes. Indirect methods relying on fluorescence and chemiluminescence have been developed (Ross *et al*., 1999). Direct methods utilizing fluorophores are faster than indirect methods because the antibody conjugates, incubation, and associated blocking and washing steps have been eliminated (Osborne, 2000).

A comparison of the performance of PCR, ELISA, and DNA hybridization for the diagnosis of the causal agent of bacterial ring rot *Calvibacter michiganensis* subsp. *sepedonicus* in crude extracts of field grown potatoes was carried out. Results showed that PCR was slightly better than ELISA, and both PCR and ELISA were superior to DNA hybridization in detection sensitivity. On the other hand, the two DNA-based assays (PCR and DNA hybridization) have the advantage of not relying on an arbitrary positive threshold, and had greater specificity (Drennan *et al*., 1993; Slack *et al*., 1996).

Recently, an alternative Southern blot technology has been attempted using near infrared (IR) fluorescent dyes (emissions at approximately 700 and 800 nm) coupled to a carbodiimide reactive group and attached directly to DNA in a 5 min reaction. The signals for both dyes are detected simultaneously (limit in the low zeptomolar range) by two detectors of an IR imager, something not yet possible with conventional radioactive or chemiluminescent detection techniques (Stull, 2001).

Although Southern hybridization can be quantitative, it is mainly used as a research tool for GMO detection, and is not suitable for routine day-to-day diagnosis because of its low throughput, as it takes on average two days to complete at a cost of $150/sample, and it is unsuitable for automation (Ahmed, 2002).

29.5.2 Polymerase Chain Reaction (PCR)

Because not all GM foods contain expressed protein(s) or have antibodies available to detect them, and the rather low expression levels of transgenic products in tissue used for human consumption (Lipton *et al*., 2000), more sensitive PCR methods are used for diagnosis. A positive result, however, has to be conformed by a specific assay determining the unique modification (Schreiber, 1999). Two essential prerequisites for the application of PCR-based diagnostic methods are complete knowledge of the foreign gene construct within the GMO to be detected, and the ability to extract significant amounts of amplifiable DNA from the samples to be diagnosed (Ahmed, 1995); whereas the availability of CRM and criteria for standardization are the limiting factors for PCR diagnosis (Wurz *et al*., 1999).

PCR exploits the specificity of DNA polymerase to allow the selective amplification of specific DNA segments occurring at low frequency in a complex mixture of other DNA sequences. In a standard PCR test, two pairs of primers are used: forward, sense or 5′A 3′A; and reverse, antisense or 3′A 5′A. These primers are designed to hybridize on opposite strands of the sequence of interest, and through a series of repetitive cycles (of 2–3 thermal steps) amplify the sequence between the primers millions of times (see also Fig. 29.2). Amplified fragments can be subjected to agarose gel electrophoresis to separate amplified DNA according to size, although other separation methods such as high-performance liquid chromatography (HPLC) and capillary electrophoresis (CE) are used (Ahmed, 1995; De Palma, 2001).

PCR diagnosis of GMOs involves four critical steps:

1. Sampling and sample preparation
2. DNA purification and aliquot size
3. PCR amplification and detection of reaction products
4. Interpretation of results

Sampling issues have been discussed in a previous section. For routine purposes, a sample size in the 2.5 to 3 kg range is recommended (Fagan, 2004). The field samples are ground and homogenized, and duplicate subsamples (between 1 and 2 g) are taken in a manner that ensures that the analytical sample is representative of field samples in accordance with recommendations of the ISO/CEN working group on GMO testing (Whitaker *et al*., 2001). DNA is then isolated independently from the duplicate analytical

subsample for later PCR analysis. Purification procedures not producing DNA that is free from PCR inhibitors or DNA degradation must be minimized, and DNA yield must be sufficient for reliable analysis. There are two basic procedures commonly used for isolating DNA from food and ingredient products for GMO diagnosis: the CTAB method, based on incubating food sample in the presence of the detergent cetyltrimethylammonium bromide; and the Wizard method, employing DNA-binding silica resins (Promega Corp., Madison, Wisconsin). Both methods produce satisfactory DNA isolation without unacceptable DNA degradation and are cost effective. Extraction methods for GMO diagnosis employing several commercial kits have recently been reviewed in detail (Terry *et al*., 2002).

Food and agricultural products contain numerous compounds that are inhibitory to PCR such as polysaccharides, caramelized sugar, proteins, fats, cocoa extracts, phenols, Ca ++, Fe ++, and other secondary metabolites and trace compounds. Although not easy to separate these inhibitors, an extraction kit (e.g. QI Amp DNA Stool Kit™ (QIAGEN Inc., Valencia, CA)) was reported to effectively remove PCR inhibitory substances such as cocoa from highly processed foods containing GMOs (Tengel *et al*., 2001).

Factors such as excessive heat, nuclease activity, and low pH (quite common in food processing) also contribute to DNA degradation. This is most likely for products with long shelf lives, such as prepared meatballs in tomato sauce and beef burgers. Data suggest that the critical minimum average DNA size for successful PCR analysis is approximately 400 bp (Meyer, 1999). Remedies such as increasing the amount of sample DNA added, and selecting primer sets whose recognition sites within the fragment DNA are very close together resulting in amplicons approximately 100 bp long, have been recommended (Fagan, 2004).

The most currently available GMOs in the EU contain any of three genetic elements: the cauliflower mosaic virus (CaMV) 35S promoter, the nopalin synthase (NOS) terminator, or the kanamycin-resistance marker gene (nptII), and others (Ahmed, 2002). These elements also occur naturally in some plants and soil microorganisms, and can thus be detected using PCR giving false positive results. If the PCR assay gives a positive result, product-specific PCR methods that have been developed for a range of different GM foods, can be carried out. These product-specific PCR methods are based on the use of a primer pair set that spans the boundary of two adjacent genetic elements (e.g. promoters, target genes, and terminators), or that are specific for detection of the altered target gene sequence. Detection limits are in the range of 20 pg to 10 ng of target DNA and 0.0001–1% of the mass fraction of GMOs (Smyth *et al*., 2002).

Different methods can be used to confirm the PCR results:

- Specific cleavage of the amplified product by restriction endonuclease digestion (Meyer, 1999)
- Hybridization with a DNA probe specific for a target sequence
- Direct sequencing of the PCR product (Sambrook and Russel, 2000)
- Nested PCR (Bouw *et al*., 1998), in which two sets of primer pairs bind specifically to the amplified target sequence

29.5.2.1 Qualitative PCR

The first method for GMO identification in foodstuffs developed to identify the Flavr Savr™ tomatoes was a qualitative PCR application assay because the genetic modification did not produce a protein in the plant (Meyer, 1995). This tomato contains – in addition to the polygalacturonase (PG) gene, which degrades pectin in the cell wall – the Kanr gene, conferring resistance to kanamycin and the cauliflower mosaic virus promoter CaMV 35S. PCR detection was achieved by designing two pairs of primers: one pair amplified a 173 bp fragment for Kanr, and the second pair amplified a 427 bp fragment that contains part of the promoter sequence (Lüthy, 1999). Other methods for detection of RR soybean, containing the genetic element from the crown gall causing bacterium *Agrobacterium tumifaciens* producing the enzyme 3-enolpyruvyl-shikimate-5-phosphate-synthase (EPSPS) that makes the plant resistant to the herbicide glyphosate, and of maximizer maize™ (MM) containing the synthetic endotoxin cryIA (b) gene employing qualitative PCR, were developed (Lüthy, 1999).

Later on, a nested PCR method was applied to the detection of EPSPS gene in soya meal pellets and flour, as well as a number of processed complex products such as infant formulas, tofu, tempeh, soy-based desserts, bakery products, and meal-replacing products (Bouw *et al*., 1998). In this two-step method, an outer primer was used to amplify a 352 bp fragment, followed by an inner primer set to amplify a 156 bp. This resulted in improved selectivity and sensitivity of the PCR reaction. RR bean DNA could be detected at 0.02%, but processed products (e.g. candy, biscuits, lecithins, cocoa-drink powder, and vegetarian paste) were undetectable by PCR due to DNA breakdown as a result of heating, and low pH, which resulted in increased nuclease activity leading to depurination and hydrolysis (Meyer, 1999; Wurz *et al*., 1999). The presence of inhibitory components and low amounts of DNA in some material (e.g. lecithin, starch derivatives, and refined soya oil) makes it difficult to develop a single reliable method for detection of all products (Lüthy, 1999).

29.5.2.2 Limiting Dilution PCR

A semiquantitative method for RR detection based on the limited dilution method has been the official method for detection of GM foods in Germany (German Food Act

LMBG § 35, Jankiewciz *et al*., 1999). This method is based on optimization of the PCR so that amplification of an endogenous control gene will take place in an all-or-none fashion occurring at the terminal plateau phase of the PCR, and the premise that one or more targets in the reaction mixture (e.g. GMO) will give rise to a positive result. Accurate quantitation is achieved by performing multiple replicates at serial dilutions of the material(s) to be assayed.

At the limit of dilution, where some endpoints are positive and some are negative, the number of targets present can be calculated from the proportion of negative endpoints by using Poisson statistics (Sykes *et al*., 1998). In this method, two measurements are used for setting limits for the GMO content of foods: a theoretical detection limit (L_{Theoret}), defined as the lowest detectable amplification determined from the serial dilution of target DNA with/without background DNA; and a practical detection limit (L_{Pract}), defined as the lowest detectable amplicon determined by examining certified reference material CRM containing different mass fraction of GM and non-GM organisms. The L_{Theoret} for both RR and MM (0.0005%) is generally two or more orders of magnitude lower than L_{Pract} (0.1%, Sykes *et al*., 1998). An advantage of this method is that it does not require coamplification of added reporter DNA. However, caution should be exercised when using this technique because of potential contamination of PCR reactions due to various dilutions and manipulations (Hupfer *et al*., 1998).

29.5.2.3 Quantitative Endpoint PCR

An important aspect in GMO food analysis is quantitation, since maximum limits of GMO in food are the basis for labeling in countries like the EU, and the fast increasing number of GM foods on the market demands the development of more advanced multidetection systems (Schreiber, 1999). Therefore, adequate quantitative PCR detection methods were developed. Quantitative competitive (QC)-PCR was first applied for the determination of the 35S promoter in RR and MM in Switzerland as described by Studer and colleagues (1998). In this method, an internal DNA standard was coamplified with target DNA (Hübner *et al*., 1999). The standard was constructed using linearized plasmids containing a modified PCR amplicon, which was an internal insert of 21 or a 22 bp deletion in the case of RR and MM DNA, respectively. QC-PCR consists of four steps:

1. Coamplification of standard and target DNA in the same reaction tube
2. Separation of the products by an appropriate method such as agarose gel electrophoresis and staining the gel by ethidium bromide
3. Analysis of the gel densitometrically
4. Determining the relative amounts of target and standard DNA by regression analysis

At the equivalent point, the starting concentration of internal standard and target are equal. In the QC-PCR, the competition between the amplification of internal standard DNA and target DNA generally leads to loss of detection sensitivity. Nevertheless, the method allows for the detection of as little as 0.1% GMO DNA (Hübner *et al*., 1999).

29.5.2.4 Quantitative Real-Time PCR

To overcome some of the limitations of conventional quantitative endpoint PCR, a real-time PCR was introduced that provided a large dynamic range of amplified molecule, allowing for higher sample throughput, decreased labor, and increased fluorescence (Ahmed, 2002).

Several commercially available real-time PCR thermal cyclers, although different in design and operation, all automate the analytical procedure and allow cycle-by-cycle monitoring of reaction kinetics; this mode of monitoring permits calculation of the concentration of target sequence. Several formats are used to estimate the amount of PCR product:

- The ds-DNA-binding dye SYBR Green I
- Hybridization probes, or fluorescence resonance energy transfer (FRET) probes
- Hydrolysis probes (TaqMan® technology)
- Molecular beacons (see also Chapter 7)

These systems also permit differentiation between specific- and non specific-PCR products (such as primer-dimer) by the probe hybridization or by using melt curve analysis of PCR products, as non-specific products tend to melt at a much lower temperature than the longer specific products (Ahmed, 2000).

29.5.2.5 Quality Assurance Issues

There are generally two sources of sample-to-sample variability in PCR experiments: differences caused by variation in the quantity or quality of the samples (e.g. partial degradation or the presence of contaminants), and random sample-to-sample variation (which includes user-induced ones). Unfortunately, random variability is a fact in PCR; the best way to minimize it is to run duplicate samples and average the data. Variability caused by operator error can be minimized by making a cocktail of reagents or master mix (Fagan, 2004).

PCR is a very sensitive amplification process. Therefore, extreme care must be taken to prevent contamination of primers and PCR reagents with cDNAs, cRNAs, RNase, or DNase. PCR reactions should be set up in a laminar flow hood, employing the same precautions used in aseptic procedures, in a laboratory separate from the laboratory where the reaction will be run. Examples of these procedures include cleaning all work surfaces with DNA Away (Molecular Bio-Products, San Diego, CA), wearing

gloves, using aerosol-resistant pipette tips, using dedicated pipettors, routinely cleaning the pipettes with DNA Away, using DEPC-treated water for all dilutions, and adding primers and/or DNA to the master mixes after all component tubes have been closed to prevent cross contamination (Ahmed, 1995).

Changes made in reagents by manufacturers often produce different results, or may change the level of sensitivity of the assays. Collecting and storage procedures were found to influence detectable levels of soluble markers; even using the same kits, studies demonstrated differences among results obtained by various laboratories suggesting that factors relating to assay performance may be responsible. It is important that a laboratory be attuned to quality control and good laboratory practice issues in testing and sample storage (see also Chapter 40); assessing analytical sensitivity for minimum detection level of markers under consideration; intra- and interassay variability; problems inherent in use of different suppliers – and sometimes different batches – for the same reagents; and the need to establish median, mean SD, and 5th and 95th percentiles for references' range (Ahmed, 2000).

A method for preamplification inactivation of amplified DNA that allows for the selective destruction of previously amplified DNA can be used, in which deoxynucleotide 5′A-uridine triphosphate (dUTP) is substituted for deoxynucleotide 5′A-thymine triphosphate (dTTP) in the master mix. All amplified DNA therefore will contain U instead of T. When a new amplification reaction is set up, the master mix is supplemented with the enzyme uracil-DNA-glycosylase (UDG) prior to the start of the temperature cycling. This enzyme catalyzes the excision of uracil from ss- and ds-DNA (but not RNA). During the first denaturation cycle of the PCR program, there is strand scission of the aglycosidic linkage, destroying contaminating templates. UDG is inactive at most temperatures used during PCR cycling (i.e. above 55°C), so newly synthesized amplicons will not be degraded prior to detection. The level of effective sterilization approaches 106 copies of input DNA, which is much more than should be necessary to control aerosol carryover contamination. Moreover, the presence of UDG at the start of the amplification appears to enhance the specificity of the amplification reaction. Because residual UDG activity can potentially degrade dU-PCR products under certain conditions (i.e. prolonged incubation of the product at either 4 or 25°C) following thermal cycling, the samples are kept at 72°C to protect amplified dU products. The UDG inhibitor protein Ugi is used to inactivate residual UDG activity and protect dU-containing products during benchtop manipulations (Ahmed, 2003).

Optimization experiments for PCR can be carried out by attempting various cycles (from 25 to 40); changing annealing temperatures (in 3°C increments above or below calculated annealing temperature); employing hot start Taq DNA polymerase; and varying $MgCl_2$ and dNTPs concentrations. In the case of amplifying regions of high GC content, denaturants such as formamide and DMSO may be tried. Both positive and negative controls should be employed. To detect inhibitors of Taq DNA polymerase, additional validation procedures, or a method referred to as quality control PCR, need to be employed (Ahmed, 2003).

Maintaining uniformly high standards of performance in the laboratory carrying out PCR diagnosis includes adherence to a thorough system of standardized operating procedures, assessment of laboratory performance from within and through external performance assessment schemes, accreditation, and availability of international standardized and validated testing methods (Fagan, 2004).

29.5.3 DNA Microarray Technology

The study of gene expression by microarray technology is important because changes in the physiology of an organism or a cell are accompanied by changes in patterns of gene expression (Jordan, 1998). The technology is based on immobilization of cDNA or oligonucleotides on solid support. The major advances of DNA microarray technology results from the small size of the array, permitting more information to be packed into the chip, thereby allowing for higher sensitivity, enabling the parallel screening of a large number of genes, and providing the opportunity to use smaller amounts of starting material (see also Chapter 16). Mainly, the introduction of differently labeled fluorescent probes for control and test samples made possible miniaturization of arrays, and because microarrays can be produced in series, this facilitates comparative analysis of a number of samples. Microarray technology has been employed for diagnosing GMOs in plants and foods (Van Hal *et al*., 2000). The principles of the technique are illustrated in Fig. 29.4.

Like any analytical technique, quality assurance issues (e.g. sensitivity, specificity, ruggedness, test performance) must be rigorously tested and evaluated. The sensitivity of microarray detection is an area where research has been carried out to improve its performance. Recent reports have indicated sensitivities to be in the order of fg of purified DNA in complex samples containing potential PCR inhibitors and competing target DNAs (Wilson *et al*., 2002). If carried out correctly, high-density oligonucleotide arrays were shown in some cases to confirm the identity of predicted sequence changes within the gene with about 99% accuracy covering a minimum of 97% of the sequence under study (Hacia *et al*., 1998). The hindrances to utilization of the microarray technique on a wide scale in diagnosis include lack of standardization, and the initial expense of the technology. A typical basic experiment involving 50 or so pairwise hybridizations would cost between $5,000 and $10,000 in reagents. An investment of about one million dollars will bring into existence a DNA microarray

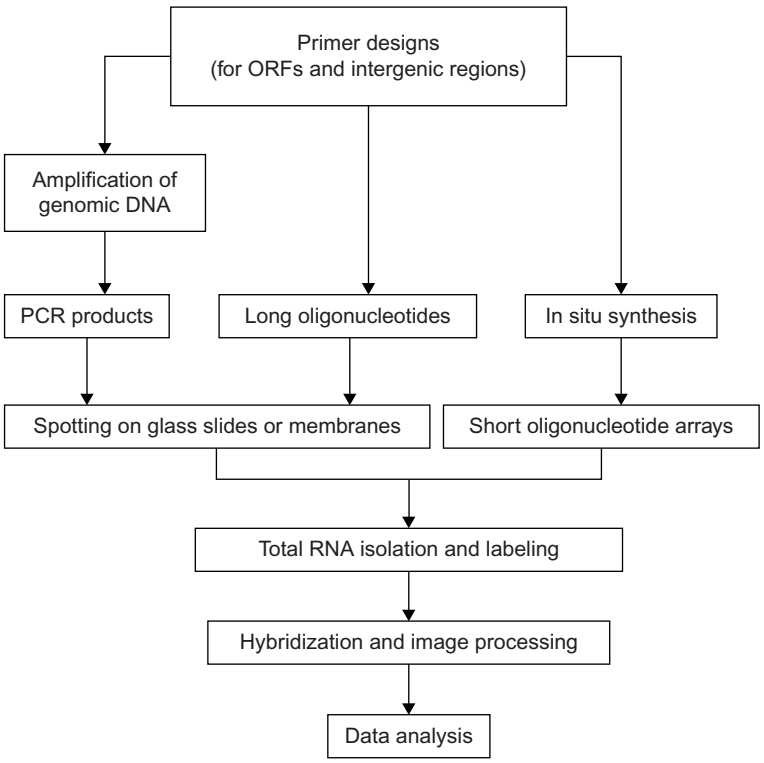


FIGURE 29.4 Construction of a DNA microarray showing principles of gene expression analysis. From Ye and colleagues (2001); with permission.

facility (Ahmed, 2000). Attempts have been made to reduce cost by employing a visualization system based on an enzyme-linked system to generate a dye visible to the naked eye, and increasing the size of spots on the arrays where probes are attached forms a few microns of approximately 1 to 2 mm in diameter, thus allowing signal visualization by the naked eye, instead of using sophisticated confocal microscopy (Grohmann, 2002).

29.5.4 DNA-Based Biosensors

DNA biosensors are based on the immobilization of a DNA probe, and on monitoring and recording the variation of a transducer signal when the complementary target in solution interacts with the probe, forming a stable complex. There are currently two types of real-time biosensing transducers for GMO diagnosis: piezoelectric, in the form of quartz crystal microbalance (QCM), and surface plasmon resonance (SPR); both the QCM and SPR biosensors have been used by immobilizing specific 25-mer oligonucleotide probes using different sensing surfaces (e.g. a screen-printed electrode (or chip), the piezoelectric crystal, and the Biacore™ sensor (M5) dextran-coated chip). These sequences are complementary to the sequences of the CaMV ^{35}S-promoter and the NOS terminator (Minunni *et al.*, 2001).

The piezoelectric and SPR sensors require biotinylation of the probes, but not for the electrochemical one. The PCR probe must first be denatured to produce a single-stranded (ss) DNA capable of hybridization. For QCM sensing, denaturation at 95°C for 5 min followed by incubation on ice was reported to be adequate. For SPR sensing using the Biacore X instrument (Upsala, Sweden), the thermal treatment did not allow for an adequate amount of ss-DNA to reach the sensor surface and react with the probe. However, when magnetic beads were used, the problem was overcome. Figure 29.5 shows QCM signals for three different denaturing methods: thermal, enzymatic, and magnetic using CaMV ^{35}S PCR-amplified plasmid DNA. Signals obtained from GMO samples for different foods (e.g. dietetic snacks and drinks, certified reference material (CRM)) were treated only thermally. The system has been optimized using synthetic complementary oligos (25-mer), and the specificity of the system (which relies on the immobilized probe sequence) was tested with a non-complementary oligo (23-mer). The hybridization study was performed also with DNA samples from CRM soybean powder containing 2% GMO and amplified by PCR. Non-amplified genomic or plasmid DNA was also used. The amplified CaMV ^{35}S resulted in a fragment 195 bp long (Minunni, 2003).

These affinity systems are attractive for DNA sensing since their versatility is often associated with probability

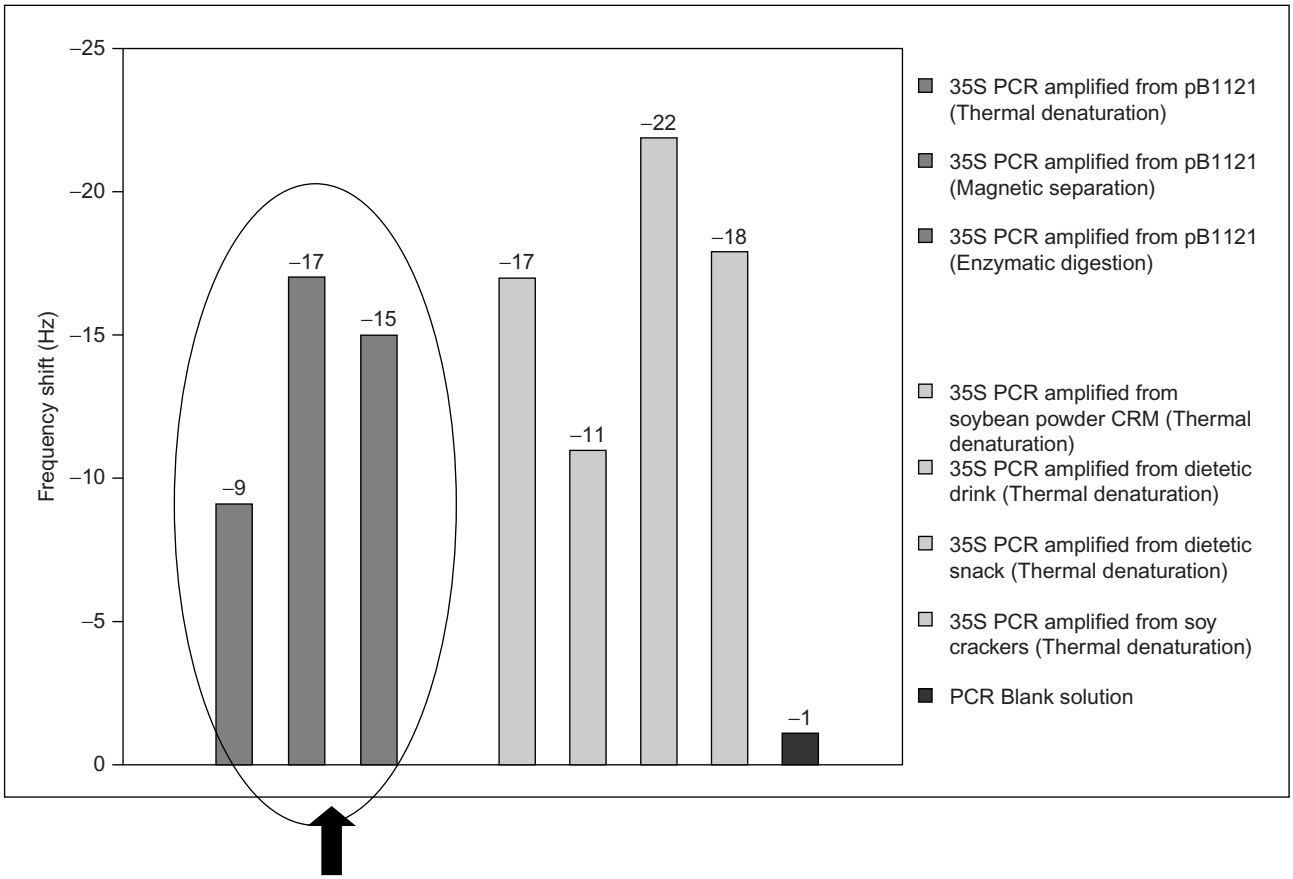


FIGURE 29.5 QCM sensor. Results obtained with different PCR amplified samples: plasmid DNA from pBI121 (containing CaMV ^{35}S) treated with three different denaturation methods (thermal, enzymatic, and magnetic particles); transgenic CRM, and dietetic snacks and light drinks (also containing CaMV 35S). From Minunni and colleagues (2003); with permission.

and with absence of any labeling. Additionally, many analyses can be performed on the same sensor surface with the possibility of reuse of the QCM and SPR devices up to 30 times and more than 100 times, respectively. Future prospects in this rapidly developing area will result – within few years – in new equipment and formats that will enhance the detection sensitivity of these methods.

The BIACORE X instrument is priced at $112,000. The chip employed for SPR analysis (e.g. dextran modified) costs about $200 per chip, with one chip capable of performing 100 analyses of PCR-amplified samples. A system based on piezo sensing, however, is much cheaper, averaging about $13,000 (including software). The sensing element (i.e. piezoelectric crystal) costs approximately $25, allowing up to 25 analyses per surface, and doubling the number of reactions with both surfaces used. Reagents for both methods cost nearly the same (about $70 for biotinylated probes immobilized on the transducer surface). Usually about 35 probes are needed for the analysis, making a probe cost per chip in the range of $1. Synthetic oligonucleotides are needed to generate a calibration curve, at a cost of $3 per chip.

Based on affinity biosensing, GeneScan Europe has recently introduced a test kit for diagnosing of GMOs in food products, which allows a multiplex PCR for the specific detection of DNA sequences from plant species and GM traits using a biosensor chip and a biochip reader. The detection limit for the GMO chip kit is in the range of 250 copies of each of the target DNA sequences in the PCR (GeneScan Europe, 2001).

29.6 NEAR-INFRARED (NIR) TECHNOLOGY

NIR transmittance spectroscopy has been widely used by grain handlers in elevators in the USA and in most of the world for non-destructive analysis of whole grains for the prediction of parameters including moisture, protein, oil, fiber, and starch. The technique has recently been used to distinguish RR from conventional soybean, but unfortunately, the sensitivity has been quite low (Rousel *et al.*, 2001). Attempts are being made to increase the sensitivity of NIR systems by accumulating an adequate database from which to glean a common NIR ‘‘signature’’ for a

given GMO, and linking the systems to sophisticated computers to increase signal detection, but the capacity has not yet been realized (Roussel and Cogdill, 2004).

29.7 CONCLUSIONS

The commonly used methods for diagnosing GMOs in foods are presented in Table 29.2. To respond to global regulations requiring food labeling, a tiered approach may be employed using first qualitative PCR for GMO detection. If no GMOs were detected using a validated qualitative method, the product(s) would be evaluated for the presence of protein. If no protein is detected, the product is presumed not detectable. If the qualitative PCR showed a positive result, the product is considered as ''non-approved GMO'', and a validated real-time PCR is used to detect the level of GMO. If the level is above an established threshold, the product is also considered non-approved GMO, but if below the threshold, the product need not be labeled (Ahmed, 2002). The high sensitivity and specificity of PCR methods and their ability to be applied to different food matrices make them suitable for diagnosing GMOs at low thresholds in various foods. Developments in DNA microarrays, biosensor devices, and proteomics promise to increase diagnostic approaches for GMO detection dramatically (Ahmed, 2004).

The greatest uncertainty of using DNA-based assays, as for protein-based methods, is that not all products derived from GM foods (e.g. refined oil) contain enough DNA. In addition, heating and other processes associated with finished food production can degrade DNA. Similarly, if GMO is expressed on a relative basis (i.e. % GMO), it is important to know whether the estimate is to be based on total DNA from all sources, or on the basis of analyzed product DNA. This approach, known as genetic equivalence, which is sound and pragmatic, was correlated with results of studies where GMO content was expressed on a percent weight basis. Quantitative PCR might best be applied at the early stages in the food production chain. Using the genome equivalent approach to assess the GMO content of food ingredients and tracking the ingredients used should allow for accurate estimation of GMOs. This approach is also consistent with current EU food labeling

TABLE 29.2 Summary of methods that specifically detect recombinant deoxyribonucleic acid (rDNA) products produced by GM foods. Modified from Ahmed (2002); with permission.

Parameter	Protein-based				DNA-based				
	Western blot	ELISA	Lateral flow strip	Southern blot	Qualitative PCR[1]	QC-PCR and limiting dilution	Real-time PCR	DNA microarrays	DNA sensors
	Difficult	Moderate	Simple	Difficult	Difficult	Difficult	Difficult	Difficult	Difficult
Needs special equipment	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Sensitivity	High	High	High	Moderate	Very high	High	High	High	Low
Duration[2]	2 d	30–90 min	10 min	6 h[3]	1.5 d	2 d	1 d	2 d	2 d
Cost/sample	US$150	US$5	US$2	US$150	US$250	US$350	US$450	US$600	US$200
Provides quantitative results	No	Yes[4]	No	No	No	Yes	Yes[5]	Yes	No
Suitable for field test	No	Yes[4]	Yes	No	No	No	No	Yes	Yes
Employed mainly in	Academic labs	Test facility	Field testing	Academic labs	Test facility	Test facility	Test facility	Academic[6] labs	Academic[6] labs

NIR detects structural changes (not DNA or protein), is fast (<1 min) and inexpensive (~$1).
[1]Including nested PCR and GMO chip.
[2]Excluding time allotted for sample preparation.
[3]When non-radioactive probes are used; otherwise 30 h with 32P-labeled probes.
[4]As in the antibody-coated tube format.
[5]With high precision.
[6]In development.

regulations that focus on ingredients, and is also applicable to finished products containing more than one GMO-derived ingredient (Ahmed, 2002).

REFERENCES

Agius, F., Gonzalez-Lamothe, R., Cabarello, J.L., Manoz-Blanco, J., Botella, M.A., and Valpuesta, V. (2003). Engineering increased vitamin C levels in plants by overexpression of a D-galacturonic acid reductase. *Nat. Biotechnol.* 21, 177–181.

Ahmed, F.E. (1995). Applications of molecular biology to biomedicine and toxicology. *J. Env. Sci. Health,* C11, 1–51.

Ahmed, F.E. (2000). Molecular markers for early cancer detection. *J. Env. Sci. Health,* C18, 75–125.

Ahmed, F.E. (2002). Detection of genetically modified organisms in foods. *Trends Biotechnol.* 20, 215–223.

Ahmed, F.E. (2003). Colon cancer: prevalence, screening, gene expression and mutation, and risk factors and assessment. *J. Env. Sci. Health,* C21, 65–131.

Ahmed, F.E. (2004). Other methods for GMOs detection and overall assessment of the risks. In *Testing of Genetically Modified Organisms in Food* (Ahmed, F.E., ed.), pp. 285–313. Haworth Press, Binghamton, NY.

Bouw, E., Keijer, J., Kok, E.J., Kuiper, H.A., and Van Hoef, A.M.A. (1998). Development and application of a selective detection method for genetically modified soy and soy-derived products. *Food Addit. Contam.* 15, 767–774.

Brett, G.M., Chambers, S.T., Huang, L., and Morgan, M.R.A. (1999). Design and development of immunoassays for detection of proteins. *Food Control* 10, 401–406.

Butler, D., and Reichardt, T. (1999). Long-term effect of GM crops serves up food for thought. *Nature* 398, 651–656.

Commission Regulation (EC) No 50/2000, 10 January 2000a, on the labeling of foodstuffs and food ingredients containing additives and flavorings that have been genetically modified or have been produced from genetically modified organisms. *Official J. Eur. Communities: Legislation,* 6, 15–17.

Commission Regulation (EC) No 49/2000, 10 January 2000b, amending Council Regulation (EC) No. 1139/98 concerning the compulsory indication on the labeling of certain foodstuffs produced from genetically modified organisms of particulars other than those provided for in Directive 79/112/EEC, *Official J. Eur. Communities: Legislation* 6, 13–14.

Council Regulation (EC) No 285/97, 27 January 1997, concerning Novel Foods and Novel Food Ingredients. *Official J. Eur. Communities*, No. L43, 1–5.

Council Regulation (EC) No. 1139/98, 26 May 1998, concerning the compulsory indication of the labelling of certain foodstuffs produced from genetically modified organisms, *Official J. Eur. Communities: Legislation* 159, 4–7.

De Palma, A. (2001). Capillary electrophoresis. *Gen. Eng. News,* 21, 21–22, 81.

Drennan, J.L., Westra, A.A.G., Slack, S.A., Delserone, L.M., and Collmer, A. (1993). Comparison of a DNA hybridization probe and ELISA for the detection of *Calvibacter michiganensis* subsp. *Sepodonicus* in field-grown potatoes. *Plant Dis.* 77, 1243–1247.

Editorial (2000). Overseeing biotech foods. *Genet. Engineer.* 20, 1–37, 83.

Fagan, J. (2004). DNA-based methods for detection and quantification of GMOs. In *Testing of Genetically Modified Organisms in Food* (Ahmed, F.E., ed.), pp. 163–220. Haworth Press, Binghamton, NY.

Gaskell, G., Bauer, M.W., Durant, J., and Allum, N.C. (1999). Worlds apart? The reception of genetically modified food in Europe and the U.S. *Science* 285, 384–386.

GeneScan Europe (2001). GMO Chip: Test Kit for the Detection of GMOs in Food Products, Cat. No. 5321300105, Bremen, Germany.

Gilbert, J. (1999). Sampling of raw materials and processed foods for the presence of GMOs. *Food Control* 10, 363–365.

Grohmann, L. (2002). GMO chip internal validation results. Symposium – GMO Analytik Heute, Frankfurt. Ch. 19.qxd 4/9/05 3:55 PM Page 246.

Hacia, J.G., Makalowski, W., Edgemon, K., Erdos, M.R., Robbins, C.M., Fodor, S.P.A., Brody, L.C., and Collins, F.S. (1998). Evolutionary sequence comparison using high-density oligonucleotide array. *Nat. Genet.* 18, 155–158.

Hardegger, M., Brodmann, P., and Hermann, A. (1999). Quantitative detection of the 35S promoter and the NOS terminator using quantitative competitive PCR. *Eur. Food Res. Technol.* 209, 83–87.

Haslberger, A.G. (2000). Monitoring and labeling for genetically modified products. *Science* 287, 431–432.

Hellemans, A. (2003). Consumers fear cancels European GM research. *The Scientist* 17, 52–54.

Hübner, P., Studer, E., and Lüthy, J. (1999). Quantitative competitive PCR for the detection of genetically modified organisms in food. *Food Control* 10, 353–358.

Hupfer, C., Hotzel, H., Sachse, K., and Engel, K-H. (1998). Detection of genetic modification in heat treated products of Bt maize by polymerase chain reaction. *Z. Lebensum. Unters. Forsch.* 206, 203–207.

Jankiewciz, A., Broll, H., and Zagon, J. (1999). The official method for the detection of genetically modified soybeans (German Food Act LMBG 35): a semi-quantitative study of sensitivity limits with glyphosate-tolerant soybeans (Roundup Ready) and insect resistant Bt maize (Maximizer). *Eur. Food. Res. Technol.* 209, 77–82.

Jordan, B.R. (1998). Large-scale expression measurement by hybridization methods: from high-density membranes to ''DNA chips''. *J. Biochem.* 124, 251–258.

Kleter, G.A., van der Krieken, W.M., Kok, E.J., Bosch, D., Jordi, W., and Gilissen, L.J.W.J. (2001). Regulation and exploitation of genetically modified crops. *Nat. Biotechnol.* 19, 1105–1110.

Kuiper, H.A. (1999). Summary report of the ILSI Europe workshop on detection methods for novel foods derived from genetically modified organisms. *Food Control* 10, 339–349.

Kuiper, H.A., Kleter, G., Noteborn, H.P.J.M., and Kok, E. (2001). Assessment of the food safety issues related to genetically modified foods. *Plant J.* 27, 503–528.

Lipton, C.R., Dautlick, X., Grothaus, G.D., Hunst, P.L., Magin, K.M., Mihaliak, C.A., Ubio, F.M., and Stave, J.W. (2000). Guidelines for the validation and use of immunoassays for determining of introduced proteins in biotechnology enhanced crops and derived food ingredients. *Food Agric. Immunol.* 12, 153–164.

Lüthy, J. (1999). Detection strategies for food authenticity and genetically modified foods. *Food Control* 10, 359–361.

Meyer, R. (1995). Nachweis genetechnisch der Flavr Savr™ Tomate. *Z. Lebensum. Unters. Forsch.* 201, 583–586.

Meyer, R. (1999). Development and application of DNA analytical methods for the detection of GMOs in food. *Food Control* 10, 391–399.

Minunni, E. (2003). Biosensors based on nucleic acid interaction. *Spectroscopy* 17, 613–625.

Minunni, M., Tombelli, S., Mariotti, E., Mascini, M., and Mascini, M. (2001). Biosensors as a new analytical tool for detection of genetically modified organisms (GMOs). *Fres. J. Anal. Chem.* 369, 589–593.

Moseley, B.E.B. (1999). The safety and social acceptance of novel foods. *Int. J. Food Microbiol.* 50, 25–31.

Osborne, J. (2000). A review of radioactive and non-radioactive based techniques used in life science applications – Part I: Blotting techniques. *Life Sci. News (Amersham Biosciences)* 6, 1–4.

Paoletti, C., Donatelli, M., Kay, S., and Van der Eede, G. (2003). Simulating kernel lot sampling: the effect of heterogeneity on the detection of GMO contamination. *Seed Sci. Technol.* 31, 629–638.

Remund, K., Dixon, D.A., Wright, D.L., and Holden, L.R. (2001). Statistical considerations in seed purity testing for transgenic traits. *Seed Sci. Res.* 11, 101–119.

Rogan, G.J., Dudin, Y.A., Lee, T.C., Magin, K.M., Astwood, J.D., Bhakta, N. S., Leach, J.N., Sanders, P.R., and Fuchs, R.L. (1999). Immunodiagnostic methods for selection of 5- enolpyruvyl shikimate-3-phosphate synthase in Roundup Ready® soybeans. *Food Control* 10, 407–414.

Ross, R., Ross, X-L., Rueger, R.B., Laengin, T., and Reske-Kunz, A.B. (1999). Nonradioactive detection of differentially expressed genes using complex RNA or DNA hybridization probes. *BioTechniques* 26, 150–155.

Roussel, S., and Cogdill, R.P. (2004). Near infrared spectroscopic methods. In *Testing of Genetically Modified Organisms in Food* (Ahmed, F.E., ed.), pp. 255–284. Haworth Press, Binghamton, NY.

Roussel, S.A., Hardy, C.L., Hurburgh, C.R., Jr, and Rippke, G.R. (2001). Detection of Roundup Ready™ soybean by near-infrared spectroscopy. *Applied Spectroscopy* 55, 1425–1430.

Sambrook, J., and Russel, D. (2000). *Molecular Cloning: A Laboratory Manual*, 3rd edition. Cold Spring Harbor Laboratory Press.

Schreiber, G.A. (1999). Challenges for methods to detect genetically modified DNA in foods. *Food Control* 10, 351–352.

Serageldin, I. (1999). Biotechnology and food security. *Science* 285, 387–389.

Shubert, D. (2002). A different perspective on GM food. *Nat. Biotechnol.* 20, 696.

Slack, S.A., Drennan, J.L., and Westra, A.A.G. (1996). Comparison of ELISA and hybridization for the detection of *Calvibacter michiganensis* subsp. *Sepodonicus* in field-grown potatoes. *Plant Dis.* 80, 519–524.

Smyth, S., Khachatourians, G.G., and Phillips, P.W.B. (2002). Liabilities and economics of transgenic crops. *Nat. Biotechnol.* 20, 537–541.

Stave, J.W. (1999). Detection of new or modified proteins in novel foods derived from GMO-future needs. *Food Control* 10, 367–374.

Stave, J.W. (2004). Protein-based methods: case studies for GMO detection. In *Testing of Genetically Modified Organisms in Food* (Ahmed, F.E., ed.), pp. 147–161. Haworth Press, Binghamton, NY.

Studer, E., Rhyner, C., Lüthy, J., and Hübner, P. (1998). Quantitative competitive PCR for the detection of genetically modified soybean and maize. *Z. Lebensum. Unters. Forsch.* 207, 207–213.

Stull, D. (2001). A feat of fluorescence. *The Scientist* 15, 20–21.

Sykes, P.J., Neoh, S.H., Brisco, M.J., Hughes, E., Condon, J., and Morley, A.A. (1998). Quantitation of targets for PCR by use of limiting dilution. In *The PCR Technique: Quantitative PCR* (Larrick, J.W., ed.), pp. 81–93. Eaton Publishing Co, Natick, Massachusetts.

Tengel, C., Scuber, P., Setzke, E., Balles, J., and Sprenger-HauBels, M. (2001). PCR-based detection of genetically modified soybean and maize in raw and highly processed food stuff. *BioTechniques* 31, 426–429.

Terry, C.F., Harris, N., and Parkes, H.C. (2002). Detection of genetically modified crops and their derivatives: critical steps in sample preparation and extraction. *J. AOAC Inter.* 85, 768–774.

Trapmann, S., Corbisier, H., and Schimmel, H. (2004). Reference materials and standards. In *Testing of Genetically Modified Organisms in Food* (Ahmed, F.E., ed.), pp. 101–115. Haworth Press, Binghamton, NY.

US National Academy of Sciences (2000). *Genetically Modified Pest-Protected Plants: Science and Regulations*. National Academy Press.

Van Hal, N.L.W., Vorst, O., van Houwelingen, A.M.M.L., Kok, E.J., Peijumenburg, A., Aharoni, A., van Tunen, A.J., and Keijer, J. (2000). The application of DNA microarrays in gene expression analysis. *J. Biotechnol.* 78, 271–280.

Whitaker, T.B., Freese, L.F., Giesbrecht, F.G., and Slate, A.B. (2001). Sampling grain shipments to detect genetically modified seed. *J. AOAC Inter.* 84, 1941–1946.

Wilson, W.J., Strout, C.L., DeSantis, T.Z., Stilwell, J.L., Carrano, A.V., and Andersen, G.L. (2002). Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol. Cell. Probes,* 16, 119–127.

Wurz, A., Bluth, A., Zeltz, P., Pfeifer, C., and Willmund, R. (1999). Quantitative analysis of genetically modified organisms (GMO) in processed food by PCR-based methods. *Food Control* 10, 385–389.

Ye, R.W., Wang, T., Bedzyk, L., and Croker, K.M. (2001). Application of DNA microarrays in microbial systems. *J. Microbiol.* 47, 257–272.

Chapter 30

Molecular Diagnostics and Comparative Genomics in Clinical Microbiology

Alex van Belkum

Erasmus MC, Department of Medical Microbiology and Infectious Diseases, Unit Research and Development, Rotterdam, The Netherlands

30.1 INTRODUCTION

Invasive infections in humans can be caused by a wide variety of protozoan, fungal, microbial, and viral pathogens. This renders microbiological and viral diagnosis a complicated task, especially since many of these pathogens generate similar and sometimes even identical clinical syndromes. As a further complicating factor, clinical samples submitted for culture-based diagnostic procedures are frequently contaminated by microorganisms naturally colonizing the epithelial lining of the human body. Moreover, the specimens also may contain substances that limit the chances for pathogen's survival and, thereby, cultivation success. On the whole, microbial diagnostics is usually performed on a wide variety of clinical samples and the microbial agents requiring identification may differ in prevalence with age, sex, habits, demographics, health status, and several other personal characteristics of the individual patient. The prevalence of infectious agents may also be subject to strong seasonal and geographic variation and is currently considered to be strongly influenced by travel habits as well. In all, culture-oriented clinical microbiology seems to be an art rather than a precise scientific approach. With respect to correctness of a diagnosis, much depends on the expertise of the clinical microbiologist or infectious disease (ID) specialist: during the diagnostic procedures several important decisions must be taken and these may be biased due to personal preferences rather than scientific appropriateness.

The choice of microbiological laboratory procedure is usually dictated by personal experiences: doctors ask for a specific type of diagnostic assay, but this may not always be the most obvious one. After clinical parameters have been assessed at bedside, the first step in microbial diagnosis normally consists of a straightforward (gram-) staining procedure. This segregates gram-positive from gram-negative bacteria and already, in part, may lead to direct species identification of the pathogen involved. Malaria parasites, for instance, show very distinct morphological features upon simple Giemsa staining of a blood smear (Iqbal *et al*., 2003). Subsequently, the clinical specimen may be further analyzed by cultivation-based tests. In addition, concomitant host serum samples may be analyzed for antibodies raised by the host's immune system against the pathogen involved. Microbial antigens rather than the complete and viable organism itself can be searched for as well. Finally, when investigations are finished and a pathogen has been identified beyond any reasonable doubt, rational therapy can be implemented, if available. Therapy is usually preceded by assessment of the antimicrobial susceptibility profile and/or virulence characteristics of the microorganism involved; that is, in case of successful cultivation, of course. It goes without saying that speed and quality of the diagnostic procedures in the end determine the clinical impact and efficacy of antibiotic therapies.

Since some pathogens may be transmitted easily between humans, not only diagnosis is important. It is also relevant to keep track of the spreading of certain pathogens, both in the hospital and in the open population. Classic microbiology uses methods such as phagetyping, serotyping, or antibiogram comparison to assess epidemiological relatedness among strains of a given pathogenic species. The diagnostic and epidemiological methods just mentioned have been in place for already quite some time. Seeing the continuing impact of infectious diseases on modern society (i.e. the AIDS pandemic, biological warfare, emergence of multidrug resistant bacteria, the recent SARS outbreak), there is a persisting need for continuous improvement of diagnostic microbiology and virology (Demain, 1999). Initially, serological tests, covering both

host antibody responses assessment and bacterial serotyping, helped improve adequate diagnosis of several infectious diseases. More recently, molecular microbiology has provided the laboratory with additional tools that will help significantly to improve the quality of microbial detection and (subspecies) identification (see Fig. 30.1).

Molecular microbiological techniques initially were developed during the 1960s and 1970s in fundamental research laboratories. As a direct result, over the past one or two decades, the identification of the organisms that are infectious to humans has been greatly facilitated by the development and application of specific molecular hybridization (probe) tests (Ksiazek *et al*., 2003). In addition, the availability of (real-time) nucleic acid amplification methods has been instrumental in the development of another, even more recent category of direct, highly sensitive diagnostic assays. The polymerase chain reaction (PCR) is among the most popular of these methods (Wolk *et al*., 2001). The currently available molecular assays enable both the direct detection of antigens (DNA and RNA) of putative pathogens in clinical material and the genetic identification (also known as DNA fingerprinting or comparative genomics) of microorganisms obtained by culture. Fingerprinting also allows for species identification and, more importantly, facilitates the identification of molecular markers that are associated with the disease invoking potential of a given strain of a potentially pathogenic microorganism. So-called pathotyping is an important recent development.

The availability of molecular diagnostics initially was considered a panacea, but replacement of conventional tests for detection and identification of microorganisms by molecular procedures turned out to be a slow and cautious process. However, several of the innovative nucleic acid identification tests are currently just beyond or slightly ahead of their breakthrough. For all the clinically relevant microorganisms one or more molecular tests for detection and identification, even below the species level, have now become available, although commercially developed tests are for sale for only the most prevalent infectious disease agents. High-throughput use of these tests, however, is becoming more common although still somewhat restricted to the larger laboratories possessing adequate technical equipment and analytical expertise.

In this chapter, the current state-of-the-art of molecular diagnostics and comparative genomics in medical microbiology will be described, together with the technological advances that have been made recently. In addition, some of the problems remaining to be solved prior to general acceptation of nucleic acid-mediated detection and identification of microbial pathogens will also be reviewed. This chapter highlights the success of the novel applications by providing examples of modern molecular diagnostic approaches in the field of bacterial infections. Viral diagnosis will not be discussed in depth, but a short section describing the current state of affairs within virology will be provided later in this chapter.

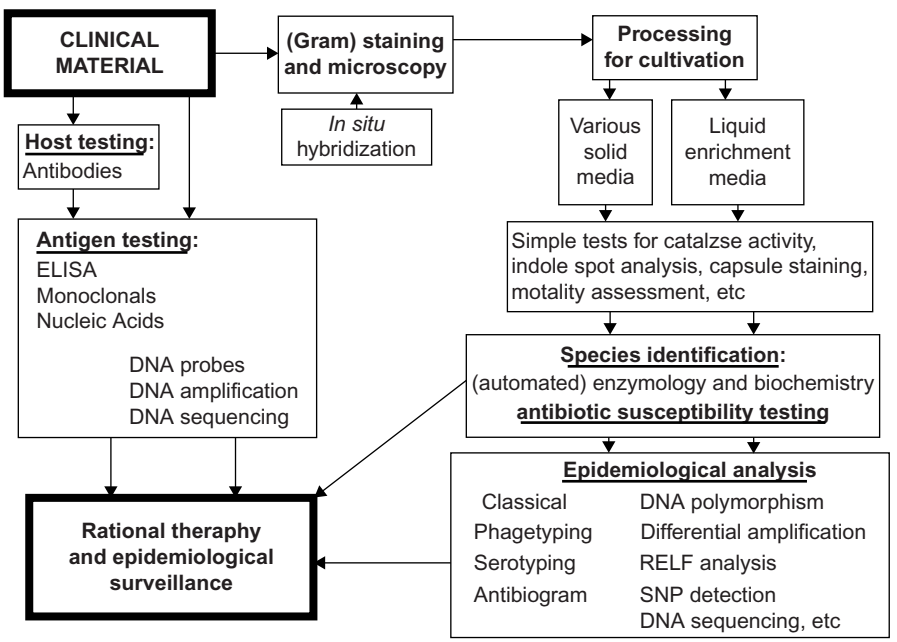


FIGURE 30.1 Schematic representation of sample routing in clinical microbiology. The classic methods are depicted in standard letter type, whereas the novel molecular diagnostic tools are depicted in italic and bold lettering. The arrows indicate the flow of activities; note that epidemiological investigations can be guided by data from both the antigen detection and culture-based test systems (ELISA: enzyme-linked immunosorbent assay; RFLP: restriction fragment length polymorphism; SNP: single nucleotide polymorphism).

30.2 TECHNOLOGICAL IMPROVEMENTS

Essentially, molecular diagnostics of infectious diseases is simple and straightforward. A clinical sample (or a not yet identified organism derived from a microbiological culture experiment) is provided and the use of a broad-spectrum nucleic acid isolation procedure generates material to be included as a template in nucleic acid hybridization or amplification assays. These reactions proceed and positive or negative results are produced, interpreted, and translated into a diagnostic result. However, this simple scenario has been compromised by a variety of obvious and emerging problems. Most of these have been solved, but some remain and are continuously investigated in search of elegant solutions. Some examples of recent technological successes follow.

30.2.1 Optimization of the Isolation of Template Nucleic Acid Molecules

During the past decade, the guanidinium isothiocyanate/Celite affinity procedure was accepted as the gold standard in many molecular diagnostic laboratories (Boom *et al*., 1990). Its main downside was the fact that the manual version of this procedure was and still is quite time-consuming and laborious. Fortunately, alternative systems have become available. Extraction simplification and automation have progressed well and after the commercial availability of handy spin column assays (e.g. the Qiagen kits; McOrist *et al*., 2002), fully automated nucleic acid isolation systems have been developed. bioMérieux has automated the Celite affinity procedure through the Nuclisense Extractor machine. Roche Molecular Systems is marketing the MagnaPure DNA isolation robot, which allows for parallel DNA extractions for hundreds of samples per working day, invoking limited hands-on time only (Van Doornum *et al*., 2003). In a recent study, three methods (NucliSens miniMAG (bioMérieux), MagNA Pure DNA Isolation Kit III Bacteria/Fungi (Roche), and a silica-guanidiniumthiocyanate (Si-GuSCN-F) procedure for extracting DNA from stool specimens) were compared. The manual Si-GuSCN-F procedure showed the highest analytical performance at the lowest associated costs per extraction (euro 4.28). However, this method did require the longest hands-on time. Both miniMAG and MagNA Pure extraction showed similar performances. With regard to amplification inhibition all methods showed relatively low rates (<4%). Costs, finally, were euro 4.28 for Si-GuSCN, euro 6.69 for MagNA Pure, and euro 9.57 for miniMAG (Schuurman *et al*., 2007). Further improvements in the nucleic acid yield and sample throughput are anticipated. The currently separate DNA extraction and amplification machines will soon be combined in a single apparatus (Cockerill and Smith, 2002).

To monitor the extraction process for loss of target or contamination by inhibitory compounds, internal process controls should be introduced (see also section 30.4 for more details). This methodology identifies deleterious effects during extraction (loss of sample) and amplification (inhibition), and enables one to be more confident on both positive and negative results generated. It has to be stated that molecular microbiologists have been putting in enormous efforts to demonstrate the efficacy of nucleic acid-mediated testing over the past years. Far more than classic microbiology, DNA/RNA testing has been subjected to extremely detailed and stringent intercenter quality control studies. Interestingly, most of the new generations of tests withstood the comparisons with classic testing quite easily, thereby confirming their superiority over the widely accepted classic test systems (Harmsen *et al*., 2001; Niesters, 2002). Recently, the molecular diagnostics of methicillin resistant *Staphylococcus aureus* MRSA bacteria has also been the subject of such a multicenter validation study (van Belkum *et al*., 2007a).

30.2.2 Prevention of Carry-Over Contamination

The implementation of PCR and other amplification tests in routine laboratories requires logistic adaptation and training programs for the personnel involved. For one, the various stages in PCR diagnostics (i.e. nucleic acid purification, preparation of enzyme master mixes, collating the complete PCR samples, amplification, and, finally, the analysis of the amplified material) should preferably be performed in physically separate laboratories. It is advisable even to plan personnel involvement in a downstream manner: individuals that have been in contact with amplimers should not be allowed to enter the clean laboratory anymore that day. The most important adaptations are those required to prevent contamination and subsequent false positive results. The current generation of commercial assays uses intelligent systems such as the UNG glycosylase approach (Meier *et al*., 1993). This results in pre-PCR destruction of amplimers generated in prior tests; this system has proven to be quite robust (Rys and Persing, 1993). There are several other measures and precautions to be taken that essentially prevent carry-over contamination from occurring. Replacement of DNA-contaminated disposable articles by clean ones, avoiding pipette abuse (aerosol formation!), adequate routing of patient materials, and the use of molecular-grade PCR ingredients are a few of these alternative options. A comprehensive discussion of most of the important factors was recently published by Millar and colleagues (2002).

30.2.3 Real-Time PCR Systems

The initial molecular tests for infectious diseases were strictly qualitative in nature: they provided a straightforward

yes-or-no answer to questions relating to the absence or presence of certain pathogens. With recent improvements of the nucleic acid-mediated technology, quantitative aspects also became addressable. Current methodologies allow for the assessment of the relative quantity of a given nucleic acid target molecule in a clinical sample (Barken *et al*., 2007). A variety of real-time PCR machines have been developed. These include, among others, the GeneAmp 5700 and Prism 7700 by Applied Biosystems, the BioRad iCycler, the Roche LightCycler, the Cepheid Smartcycler and GeneXpert, the MX400 by Stratagene and the Rotor Gene by Corbette Research (see also Chapter 7). All these machines share a high assay speed (between 20 minutes and 2 hours per run) and sufficient sample capacity (from 16 to 384 samples per run). Real-time technologies combining PCR and TaqMan hydrolysis probes (Lunge *et al*., 2002), NASBA and molecular beacons (Weusten *et al*., 2002), or hybridization probes (Bidet *et al*., 2003) helped to improve the tedious process of real-time detection and identification. It has been well established that the new technologies have the reproducible ability to detect (small amounts of) an infectious agent in all sorts of clinical specimens.

The question whether nucleic acid quantification may have an added value for clinical infectious disease management in distinct groups of patients should and must be answered in the near future. The main issue here is whether larger amounts of pathogens present may be an indication of the severity of (underlying) disease. If this is the case (as suggested by some virology studies; Schutten and Niesters, 2001) this implies that quantitative tests could also be used successfully to monitor the efficacy of a treatment protocol.

30.2.4 Broad Spectrum PCR and PCR Multiplexing

In some cases, species of microbial pathogens capable of invoking an infectious syndrome may be large in number. Infection of the lower and upper airways, for instance, can be caused by a variety of bacterial species. These include, in random order, *Streptococcus pneumoniae*, *Moraxella catarrhalis*, *Haemophilus influenzae*, *Bordetella pertussis*, *Mycoplasma pneumoniae*, *Legionella pneumophila*, *Chlamydia pneumoniae*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, and others covering a wide variety of viruses, fungi, and parasites as well. It would be very cost effective if these pathogens, involved in similar, clinically manifest diseases, could be detected by a single assay. Broad spectrum PCR coupled to species-specific nucleic acid probing assays could provide such an opportunity (Roth *et al*., 2000), whereas multiplex PCRs have been described that are capable of achieving a similar feat (Ko *et al*., 2003). This type of assay will become increasingly important over the coming years, especially with the introduction of quantitative variants of these tests. The downside of multiplexing is that the multiplex PCR is usually considered to be less sensitive than the simplex approach. This problem could be circumvented by applying a new generation of PCR-mediated tests. For instance, the multiplex ligation probe assay is supposed to be relatively insensitive to the downside of classical PCR multiplexing (Reijans *et al*., 2008). A recent test was described which supposedly detects 15 respiratory viruses in one reaction. The MLPA reaction is preceded by a pre-amplification step which ensures the detection of both RNA and DNA viruses with the same specificity and sensitivity as individual monoplex real-time RT-PCRs. The use of this test showed an increase of the diagnostic yield from 35.5 up to 60% compared to cell culture. The test provided a user-friendly and high-throughput tool for the simultaneous detection of 15 respiratory viruses with excellent overall performance statistics.

The detection of ill-defined or even previously unrecognized bacterial pathogens can be pursued by broad spectrum PCR, a method already alluded to in the previous section (Relman, 1998). The method usually involves PCR targeting of universally conserved, eubacterial PCR priming sites. A well-known example of such targets are the ribosomal genes, although successful broad spectrum amplification of genes encoding DNA polymerase, DNA gyrase, or protein elongation factors has been described as well. Broad spectrum PCR has recently been used to try to resolve a variety of clinical syndromes. These include, for instance, periodontal diseases (Kumar *et al*., 2003) and non-chlamydial, non-gonococcal urethritis in men (Riemersma *et al*., 2003). The latter study elegantly demonstrated that in infectious diseases not only the presence of a pathogen may be important, but also the absence of certain apparently healthy components of the resident flora may be helpful in diagnosing disease. There may be, however, fundamental obstructions to the diagnostic implementation of broad spectrum PCR. For instance, there is an intense discussion still ongoing on the presence of native, non-pathogenic commensal bacteria in human blood (Nikkari *et al*., 2001; McLaughlin *et al*., 2002). Such phenomena may in the end completely frustrate the application of broad spectrum PCR in nucleic acid extracts of these matrices.

30.3 PERSISTING PROBLEMS WITH MOLECULAR DIAGNOSTICS

In clinical microbiology, companies focus on a limited number of economically attractive microbes for the development of well-standardized, commercially available assays. From this perspective, it is deemed unlikely that the number and scope of actively marketed infectious disease tests will increase rapidly over the coming years: all of the economically interesting diagnostic areas now seem to be covered and primarily the competition between diagnostic

companies may increase significantly when different tests aiming at the same agent become more widely available. However, it should be emphasized that there is a large and still growing list of less prevalent microbial pathogens for which the commercial availability of nucleic acid test systems would be much appreciated by medical microbiologists and infectious disease specialists. For most of these minor pathogens, molecular diagnosticians have to rely on tests that have been developed in-house; a major drawback is the fact that these tests generally are not accepted and widely used. These tests usually suffer from restricted quality control only (see also Chapter 40). Hopefully, the small niche markets will raise the interest of start-up or other small companies.

Besides the current lack of specific tests availability, another bottleneck is due to the lack of convenient and reproducible nucleic acid isolation systems used in combination with the diversity of clinical specimens received by the clinical microbiology laboratory. Although several open systems are available, there still is a clear need for versatile, highly reliable automated systems. When available, these machines may significantly enhance full implementation of nucleic acid detection in the day-to-day laboratory setting. Furthermore, this will enable the laboratory to generate results within a short turnaround time, which is essential for infectious disease management. In addition, only limited amounts of information are available on the (dys)functioning of the present generation of robots and other instruments in use for diagnostic analyses. This scant availability has thus far frustrated detailed intercenter comparisons on the machines' performance. It is interesting to note that integrated platforms for both nucleic acid extraction and amplification are now available. It is particularly noteworthy that some of these machines can even be considered portable due to their small size (Seme *et al*., 2008).

Lack of standardization and quality control programmes is another persisting problem area in molecular diagnostics (see also Chapter 40). It has been known since the early 1990s that one of the most significant hurdles to be overcome by molecular technologies is false positivity due to contamination and false negativity due to the large differences in sensitivity between various test systems (Schuurman *et al*., 1996). European intercenter quality control research programs have indeed shown the clear need for standardized reference materials and the general desire to participate in quality control programs, for instance the ones organized by the Quality Control for Molecular Diagnostics (QCMD) organization (Van Vliet *et al*., 2001; van Belkum *et al*., 2007b). This latter organization, initially funded by the European Union, provides quality control schemes for an increasing number of viral and bacterial targets.

Besides the need for well-defined qualitative quality control programs, there is also a lack of standardization of (semi) quantitative assays. External standards should be developed in a format that shows maximum identity to the samples of interest. In the end this integration of molecular standards and maybe even unit definition would facilitate the intercenter communication of matching data. The conclusion still is that at present absolute quantification is very hard to achieve, a problem that may persist for years to come. Also, in the field of subspecies identification of clinical bacterial isolates, big problems have been noted. Many of the molecular technologies used for fingerprinting the genomes of microbes failed to be reproducibly implemented even in different but closely collaborating laboratories (van Belkum *et al*., 1995; van Belkum *et al*., 1998a). Although incidental successes have been reported, there is no current gold standard procedure for microbial typing (De Lencastre *et al*., 1996; Murchan *et al*., 2003). The technology most likely to survive in the end and already now providing useful tools for detailed international microbial dissemination studies probably will be multilocus sequence typing (MLST), a technology that can now be applied using Affymetrix DNA chip technology as well (Maiden *et al*., 1998; Van Leeuwen *et al*., 2003). In addition, the application of bacterial whole genome sequencing and other forms of ''deep sequencing'' are now becoming more accessible. These technologies will be discussed in a later section of this chapter.

30.4 MOLECULAR VIRUS DETECTION

One has to realize that the translation from experimental research tools into routine molecular diagnostics is still not completed in many cases. In virology this is mainly hampered by the fact that many of the new testing systems are really new: there are no historic alternatives for the sensitive detection procedure by non-nucleic acid methodologies, with the detection of the hepatitis virus, the human metapneumovirus, or the coronavirus causing SARS as recent examples (Su *et al*., 2002; Poon *et al*., 2003; Van den Hoogen *et al*., 2001). This renders comparative quality assessment difficult if not impossible and both laboratory scientists and clinicians need to put their trust in a test that stands largely unverified. However, the powerful versatility and reliability of nucleic acid testing has already convinced most investigators with clinical responsibilities.

The availability of unprecedented tests in virology has accelerated the development of process and quality controls (Oberste *et al*., 2000; Van Elden *et al*., 2001; Zaaijer *et al*., 1993). One of the means to measure and validate the performance of diagnostic testing is the use of external quality assessment (EQA) control programs. The QCMD network again supplies standardized reference sample panels to be used for monitoring of development and implementation of nucleic acid detection technologies, both in a qualitative and a quantitative manner (Quint *et al*., 1995; Valentine-Thon *et al*., 2001; Wallace, 2003). The use of EQA has

until recent years been focusing almost exclusively to blood-borne viruses, for which standardized tests are now commercially available. However, the need for EQA relating to the large panel of clinically relevant but commercially less interesting targets is growing (Espy *et al*., 2000; Savolainen *et al*., 2002).

Besides external quality control, virology has also excelled in the development of internal control procedures. One of these elegant control systems concerns a complete non-human seal herpes DNA virus (Van Doornum *et al*., 2003). A real-time and quantitative TaqMan assay was developed for this phocine herpes virus (PhHV). The virus can be grown relatively easily in cell culture and, hence, is available in sheer unlimited amounts. Also, a universal RNA virus was introduced as an internal control; the phocine distemper virus (PDV) could serve the same purpose as PhHV, for example for reverse transcriptase (RT) PCR applications (Niesters, 2004). The assumption here is that an intact virus, when used as universal internal control, behaves more similarly in the extraction procedure as compared to target viruses of interest, in contrast to using, for example, a plasmid as internal control. A low and fixed amount of this virus (equal to an amount giving a cut-off (Ct) value in the real-time assay of approximately 30–33 cycles of amplification) needs to be added to each clinical sample before the extraction procedure starts. In virology this mostly involves serum or plasma, although analysis of sputum and throat samples has also been successful. The internal control virus genomes are co-extracted and subsequently amplified in a quantitative manner in (currently) a separate tube.

30.5 EXAMPLES FROM BACTERIOLOGY

In all protocols, molecular diagnostic bacteriology starts with the purification of DNA or RNA from the bacteria involved. In other words, procedures aimed at the detection of a certain specific pathogen can be adapted simply to the detection of other pathogens in general. The commercially available molecular procedures suited for the detection of the sexually transmitted, intracellular pathogen *Chlamydia trachomatis* will be highlighted in the following pages. In addition, methods suited for the detection and subspecific identification of the gram-positive microorganism *Staphylococcus aureus*, often based on in-house testing, will be discussed in detail.

30.5.1 Commercial Test Systems for *Chlamydia trachomatis* Diagnosis

C. trachomatis is a microbe capable of causing two major clinical syndromes. When infecting the eye it can cause trachoma, which ultimately leads to blindness. When infecting the fallopian tubes it can cause pelvic inflammatory disease (PID), a major cause of infertility in women (for a review see Schachter, 1985). It is clear that diagnosis of *C. trachomatis* infection, in Western countries usually disseminated by sexual contacts or through vertical transmission during birth, requires some priority. Classical *C. trachomatis* diagnosis depended on the organism's capability to infect certain receptor cell-lines. Cervical swabs were taken, transported to the laboratory in a specialized conservation medium and the suspected material was inoculated into the cell culture system. The appearance of a cytopathogenic effect was indicative of the presence of *C. trachomatis* (Suchland *et al*., 2003). This test system was insensitive, laborious, expensive, and required high levels of analytical expertise. Hence, the availability of an initial DNA hybridization-based culture confirmation test spurred enthusiasm from the diagnostic community (Tenover, 1993).

In addition to the extended culture diagnosis, the first direct nucleic acid tests became available in the early 1980s of the last century. These commercial tests manufactured by GenProbe (San Diego, CA) were initially aimed at direct ribosomal RNA detection, the second generation of tests was also able to generate limited ribosomal RNA amplification (Verkooyen *et al*., 2003). Ultimately, various amplification mediated tests were developed based on technologies such as bacteriophage Qb RNA polymerase mediated amplification (Stefano *et al*., 1997), the ligase chain reaction or LCR (Blocker *et al*., 2002), and, of course, the PCR (see Verkooyen *et al*., 2003 and references therein). These tests all underwent extensive comparisons with the culture-based assay and among each other, and ultimately the PCR tests prevailed, although the LCR tests demonstrated adequate sensitivity and specificity (Pannekoek *et al*., 2003).

With the apparent preference of many researchers for the PCR, commercial developments took various quantum leaps. Roche Molecular Systems is now selling an integral PCR-based detection system called COBAS AMPLICOR. This system facilitates automated *C. trachomatis* diagnostics, not only based on the use of cervical swabs but also on less invasive, more patient-friendly urine samples (Leslie *et al*., 2003). Similarly, commercial diagnostic systems developed by Becton Dickinson and based on the Strand Displacement Amplification principle were introduced and found to be versatile and reliable as well (Bang *et al*., 2003; Cosentino *et al.*, 2004). In conclusion, in a mere 15 years a drastic diagnostic change has been observed concerning the detection of *C. trachomatis*. From the cumbersome cell culture systems, to be combined with cervical swabs, we now have a new gold standard technology: the detection of *C. trachomatis* in urine can be performed with a machine generating an adequate result within half a working day. So, added to the increased sensitivity and specificity of the test system, speed and throughput also have improved significantly over the past decade (for a concise meta-analysis of some of the available data, see Table 30.1). Although detailed cost-effectiveness

TABLE 30.1 Meta-analysis of *Chlamydia trachomatis* molecular diagnostics versus culture. Various diagnostic procedures are listed in the column on the left, some information on each and every system is indicated in the other columns. This is not a complete survey; this is meant for highlighting some of the more common technologies and their respective performance in clinical diagnostics.

Procedure	Target sequence	Specificity	Sensitivity	Number of samples	References
Radioactive probe hybridization	Cryptic 7 kbp plasmid	91–94%	87–90%	1,214 conjunctival samples	Dean *et al.* (1989)
PCR	16S rRNA	No cross-reactivity with DNA from 13 different bacterial species	1 bacterial cell in 100,000 host cells	DNA extracts from cell lines and various bacterial species	Pollard *et al.* (1989)
PCR and radioactive RNA hybridization	Conserved plasmid	100%	100%	200 vaginal swabs	Griffais and Thibon (1989)
Probe hybridization	16S rRNA	Not discussed	10–100 picograms	Variable numbers from different sources, mainly conjunctival swabs	Cheema *et al.* (1991)
PCR-RFLP analysis	Plasmid and outer membrane protein 1 (omp1)	Not relevant	30 culture positive, 38 PCR positive	209 cervical scrapes	Lar *et al.* (1993)
LCR	Major outer membrane protein (MOMP)	All serovars positive, no cross-reaction not even with C. psittaci and C. pneumoniae	3 elementary bodies	Dilution series of serovar L2 DNA	Dille *et al.* (1993)
LCR	Plasmid and MOMP	99.8–100% (40–85% when culture efficacy was assessed)	93–98% for plasmid PCR, 68% for the MOMP PCR	1,500 urines and urethral swabs	Chernesky *et al.* (1994)
Qβ replicase	16S rRNA	85 out of 88 culture negatives were also PCR negative	1,000 molecules, five out of six culture positive ones	94 urogenital samples	Shah *et al.* (1994)
Qβ replicase and PCR	Both 16S rRNA	Both methods added three positives to the culture positives	Five elementary bodies for both methods	94 endocervical samples	An *et al.* (1995)
PCR	16S rRNA and MOMP for discordance analysis	PCR can give positive signals for two weeks after antibiotic treatment	92.7 for the PCR, 79.1% for culture after discordance analysis	1,110 cervical swabs	Goessens *et al.* (1995)
Capture PCR	C1Q capture and endogenous plasmid	Not discussed	90–95% respectively for low and high positive cultures; capture helps concentrate bacteria	71 cervical swabs including samples with high and low positivity in the culture test	Herbrink *et al.* (1995)
Transcription-mediated amplification TMA, LCR and COBAS AMPLICOR PCR	TMA: 16S rRNA LCR: plasmid Amplicor: plasmid	TMA: 98% LCR: 99% Amplicor: 99% versus culture	TMA: 90% LCR: 91% Amplicor: 96% versus culture	First void urine and urethral/cervical scrapes for 544 males and 456 females, respectively	Goessens *et al.* (1997)
TMA and COBAS AMPLICOR PCR	TMA: 16S rRNA Amplicor: plasmid	TMA: 100% Amplicor: 99%	TMA: 85% Amplicor: 97%	First void urine of 320 males and 338 females	Pasternack *et al.* (1997)

studies have not been published thus far, preliminary explorations already indicated that in populations with a prevalence of *C. trachomatis* infections over 3.9% the direct costs of PCR mediated screening programs are low, as compared to symptom-driven analyses (Paavonen *et al*., 1998). Recently, in an evaluation of *C. trachomatis* testing with pregnant women, first-void urine specimens from 750 consecutive asymptomatic pregnant women were collected (Rours *et al*., 2005). The COBAS AMPLICOR test was compared to the COBAS AMPLICOR test with prior DNA isolation by use of the MagNA Pure. Next, using all 750 urines, the COBAS AMPLICOR performance for individual testing was compared to pooled testing with the standard COBAS AMPLICOR procedure and subsequently to pooled testing with COBAS AMPLICOR in combination with the MagNA Pure bacterial DNA isolation kit. The sensitivity of COBAS AMPLICOR was 65% on individual and 42% on pooled urines but improved to 92% on pooled urines with the MagNA Pure bacterial DNA isolation kit, making this combination the best screening method. The *C. trachomatis* prevalence in this population appeared to be 6.4%. Additionally, the cost of the combined MagNA Pure bacterial DNA isolation kit and COBAS AMPLICOR method on pooled urines was only 56% of the cost of the standard COBAS AMPLICOR test applied to individual urines. Costs per positive case detected in the combined method were 39% of standard costs. It is noteworthy that the molecular testing for *C. trachomatis* continues to surprise the scientific community. Cases of conjunctivitis can be directly linked to maternal carriage (Rours *et al*., 2008) whereas there are also ongoing discussions on the relatedness between premature delivery and *C. trachomatis* status (Caan, 2006).

30.5.2 Molecular Detection and Identification of *Staphylococcus aureus*

Classical *S. aureus* culture is still amenable to improvement. Novel selective growth media providing excellent yield and specificity are still strong competitors to diagnostic DNA testing. In the case of *S. aureus*, especially the detection of the methicillin-resistant version of *S. aureus* (MRSA), it is an important driving factor behind diagnostic improvement. Colonization of patients or medical personnel with this particularly antibiotic-multiresistant bug predisposes to dangerous, difficult-to-treat infections. Colonization and subsequent infection, or spread of MRSA in the hospital setting, must therefore be prevented, hence the continuous need for methods for MRSA detection and genetic identification (see Table 30.2 for a short list of currently available methods). MRSA, like methicillin-susceptible *S. aureus* (MSSA), has its prime ecological niche in the vestibulum nasi, the foremost compartment of the nose.

Diagnostic procedures are often focused on the analysis of nasal swabs. An example of a recently described, new medium is the Oxacillin Resistance Screening Agar developed by Blanc and colleagues (2003). This medium provides adequate results within 48 hours after inoculation. The most extensive comparison of culture-based assays was recently presented by Safdar and colleagues (2003), who compared 32 different media. Their conclusion was that optimal samples are taken with standard rayon swabs, the material should be enriched overnight in salt-containing trypticase soy broth, and the pre-enrichment culture should subsequently be inoculated on lipovitellin-containing mannitol salt agar containing oxacillin. Addition of an oxacillin disk on the agar surface further increased the specificity.

It has been shown that the quality of the clinical specimen is one of the prime determinants of diagnostic efficacy. For instance, nasal swabs are to be preferred over perianal or throat swabs for assessing the colonization status of a person (Singh *et al*., 2003). However, PCR is still considered to be the future gold standard diagnostic tool, especially for the detection and identification of MRSA. Many of the *S. aureus* PCR tests in one way or another include the detection of the methicillin resistance gene encoding the penicillin binding protein PBB2′ (Miyamoto *et al*., 2003). The first mecA-specific test was published as early as 1991 (Murakami *et al*., 1991) and many followed over the past decade.

Testing for the presence of MRSA is clinically important. For instance, Hallin and colleagues (2003) demonstrated that PCR diagnosis was instrumental in the modification and optimization of antibiotic therapy in 7/28 (25%) of patients included in their clinical study. Also, among cardiac surgery patients the prevention of post-surgical *S. aureus* wound infections was better guided by PCR than by culture (Shrestha *et al*., 2003). Besides PCR, other molecular amplification methodologies are also applied to the detection of *S. aureus*. A new method, isothermal signal amplification (ISA), allows for the detection of at least 2×10^5 MRSA cells (Levi *et al*., 2003). Although the method may have its advantages, the fact that in general a time-consuming enrichment culture is required prior to ISA may ultimately be a significant obstacle. The latest development is that of real-time quantitative PCR (Borg *et al*., 2003; Fang and Hedin, 2003; Gueudin *et al*., 2003). Real-time procedures have been successfully applied for infection control purposes and high-throughput application has been facilitated by spectacular technological developments (Boyce *et al*., 2008; Gilpin *et al*., 2007; Kluytmans, 2007).

Many other nucleic acid-controlled technologies are applied in order to reach further refinement of molecular diagnostics. This includes genome sequencing and computerized or experimental comparative genomics in order to increase the number of adequate target sequences (Baba *et al*., 2002). Recently, it was shown that whole genome sequencing could help to identify the development of mutation in the staphylococcal genome that contributes to the development of vancomycin resistance (Mwangi *et al*.,

TABLE 30.2 Examples of MRSA-specific test systems. Four categories of test systems are included. Again, this is not a complete survey; this is meant for highlighting some of the more common technologies and their respective performance in clinical diagnostics.

Test principle	Data output	References
Culture-based tests		
The **BBL crystal MRSA ID test** assesses viability by means of biochemical detection of oxygen consumption through a fluorescent compound	Fluorescence, commercially available multi-well system for assessing growth in the presence of various antibiotics	Qadri *et al*. (1994); Zambardi *et al*. (1996); Kubina *et al*. (1999)
The **soft salt mannitol agar cloxacillin test** consists of tubes containing this medium; can be used at bedside for direct swab inoculation	An indicator substance induces a color change upon the presence of MRSA	Mir *et al*. (1998)
CHROMagar Staph aureus and oxa resistance screen agar are chromogenic plate media; addition of 4 μg/ml renders the media suitable for MRSA identification	*S. aureus* colonies on the agar show a distinct purple color on Chromagar; some problems have been noted with test sensitivity	Merlino *et al*. (2000); Apfalter *et al*. (2002)
Lipovitellin salt mannitol agar is a selective medium; lipovitellin is a glycosylated, lipid binding protein present in the yolk of egg laying animals, showing homology to human apolipoprotein B	Growth or not	Verghese *et al*. (1999)
Vitek technology is developed by bioMerieux and represents automated biochemical and anti-microbial screening of bacterial isolates	Biochemical and antimicrobial profiles are produced with limited hands-on time	Knapp *et al*. (1994)
Oxacillin resistant screen agar is another chromogenic, selective culture medium	Blue stained colonies, does require follow-up identification reactions	Blanc *et al*. (2003)
Protein detection systems		
230 kilodalton *S. aureus* surface protein was found to be useful in the agglutination of non-agglutinable strains of MRSA	Experimental set-up using purified antibodies reacting with the protein	Kuusela *et al*. (1994)
MRSA screen facilitates the detection of PBP-2a in crude cell preparations	Latex agglutination system	Van Leeuwen *et al*. (1999)
Nucleic acid detection systems		
EVIGENE probe hybridization can be used for non-PCR-mediated detection of MRSA	Staining and spectrophotometry at 405 nm	Levi and Towner (2003)
Velogene rapid MRSA identification assay is a probe-mediated procedure, which uses a chimeric probe for cycling mediated recognition of MecA gene	Fluorescence value based on fluorescein, this procedure is intermediate between probe mediated and nucleic acid amplification mediated testing	Louie *et al*. (2000); Van Leeuwen *et al*. (2001); Arbique *et al*. (2001)
Nucleic acid amplification		
Multiplex PCR for nuc, Tnase and MecA facilitates detection of MRSA	Home-brew PCR, analysis of products by gel electrophoresis	Brakstadt *et al*. (1993)
Multiplex PCR for femA and mecA facilitates detection of MRSA	Home-brew PCR, analysis of products by gel electrophoresis. Has also been combined with immunoassay for product identification, including mupirocin resistance	Vannuffel *et al*. (1998); Towner *et al*. (1998); Perez-Roth *et al*. (2001); Jonas *et al*. (2002)
Combined **immunomagnetic enrichment of *S. aureus*** followed by **TaqMan PCR using the SmartCycler**	Fluorescence measurements using the SmartCycler hard- and software	Francois *et al*. (2003)
TaqMan PCR for the *S. aureus* nuc gene in combination with selective culture-based pre-enrichment	Fluorescence measurement	Fang and Hedin (2003)

2007). Such studies demonstrate that sequencing is useful for both diagnostic but also epidemiological analyses, not forgetting the linkage between genome flexibility and phenotypes. Transcription profiling using DNA chips is useful for enhancing our understanding of stage- or stimulus-specific gene expression in *S. aureus* (Mongodin *et al*., 2003). In addition, exploring the human genome for polymorphisms that can be associated with staphylococcal infection or carriage could in the end generate alternative strategies for the identification of people at elevated risk of acquiring staphylococcal infections (van Belkum *et al.*, 2007a, b; van den Akker *et al*., 2006; Wertheim *et al*., 2008).

A very important new feature of the diagnosis of *S. aureus* infection is the possibility of simultaneously detecting the presence of certain virulence genes. For instance, French studies convincingly showed that the presence of the gene for the Panton-Valentine leukocidin (PVL) confers strong disease invoking capacities upon the staphylococcus. PVL-positive strains have been implicated as causal agents in very severe cases of necrotizing pneumonia, a disease that is leading to significant mortality (Gillet *et al*., 2002). In addition, it has been shown by a Dutch research group that the presence of one or more virulence genes may lead to enhancement of impetigo in children. It was shown that the presence of the exfoliative toxin B (ETB) gene and also the PVL gene led to an increase in the number and the overall size of the impetigo lesions (Koning *et al*., 2003).

With the availability of complete inventories of putative virulence genes, as based on whole genome comparisons, the possibilities for predictive diagnosis will increase in the future: the virulence gene repertoire of a colonizing *S. aureus* strain can be assessed by molecular diagnostics, and depending on its virulence gene profile it may be decided that the strain needs to be eliminated prior to a patient undergoing surgical treatment, which reduces the risk of post-surgical wound infection.

Finally, the detection of subspecies genetic polymorphism is important for fingerprinting staphylococcal isolates and, hence, facilitates epidemiologic studies into the dissemination of clones of MSSA and MRSA. In the past, epidemiologic studies essentially were based on a strain's antimicrobial resistance profile, biochemical characteristics, or its phage type (Weiss and Nitzkin, 1971). Phage typing determines a strain's susceptibility toward infection by a large panel of lytic *S. aureus*-specific bacteriophages. Essentially, a binary code is developed consisting of alternating sensitivity or resistance toward given phages (Weller, 2000). Forty years ago, phage typing was the epidemiological gold standard and used even for successful and informative nationwide analysis of the dissemination of identical *S. aureus* phagetypes. This clonal dissemination revealed interesting staphylococcal colonization and infection dynamics and set the stage for the development of more stable and reproducible typing systems. Again, molecular microbiology provided most of the alternative possibilities. Phage typing has, for instance, been compared in detail with random amplification of polymorphic DNA, a PCR method generating DNA fingerprints consisting of non-identified DNA molecules (van Belkum *et al*., 1993). This showed that the DNA-mediated procedure was better in the sense that it appeared to be more reproducible and that its resolving power was clearly enhanced.

Over the past years a relatively large set of alternative strategies for typing *S. aureus* has become available. This includes the assessment of mutation in a variable number of tandem repeat (VNTR) loci (Sabat *et al*., 2003), a method aiming at the detection of unit number variation in certain genetic loci of fast-evolving repetitive DNA. The mechanism behind this form of DNA polymorphism is slipped strand mispairing: during replication the DNA polymerase skips or adds a repeat unit as a consequence of the complex tertiary structure of the repetitive DNA domain (van Belkum *et al*., 1998b). Multilocus analysis, so-called MLVA approaches have now been developed for all medically relevant species and for *S. aureus* multiple systems have been described (Tenover *et al*., 2007; Francois *et al*., 2005). Specific for MRSA, various methods for finetyping of the PBP-encoding gene and its neighboring sequences have been developed (Ito *et al*., 2001; Oliveira and De Lencastre, 2002); these methods usually depend on the selective PCR-mediated amplification of locus-specific sequence elements (see Fig. 30.2).

Although nearly all DNA typing methods appear to be useful for epidemiologic analysis of MRSA and MSSA (van Belkum, 2000), pulsed field gel electrophoresis (Stranden *et al*., 2003) and multilocus sequence typing (MLST; Feil *et al*., 2003) are the two methods that are currently best appreciated. Hundreds of papers have been published describing the use of PFGE for epidemiological comparisons of sets of MRSA and MSSA strains (van Belkum, 2000; see Fig. 30.2 for examples). These research efforts all strongly contributed to the current awareness of the nature of international MRSA dissemination (Aires de Sousa and De Lencastre, 2003). MLST is relatively new and defines single nucleotide polymorphisms (SNPs) in housekeeping genes. Each gene sequence is translated into an allele code and the accumulation of all (generally 7 or 8) alleles lead to an allelic profile. This can be redefined into a single digit sequence type (ST). Its resolving power is not as strong as that of PFGE but its biggest advantage is that sequencing data are extremely portable. Data can be put in a single large database and each and every individual researcher can compare his strains according to MLST type with the database entries. This facilitates worldwide comparisons on staphylococcal genotypes to be made and a number of really seminal studies on the population structure of MRSA and MSSA have recently been published (Day *et al*., 2001; Fitzgerald *et al*., 2001; Enright *et al*., 2002; Feil *et al*., 2003). Overall, five or six major MRSA clones have traveled the world and many more minor types, restricted to certain locations, have been identified.

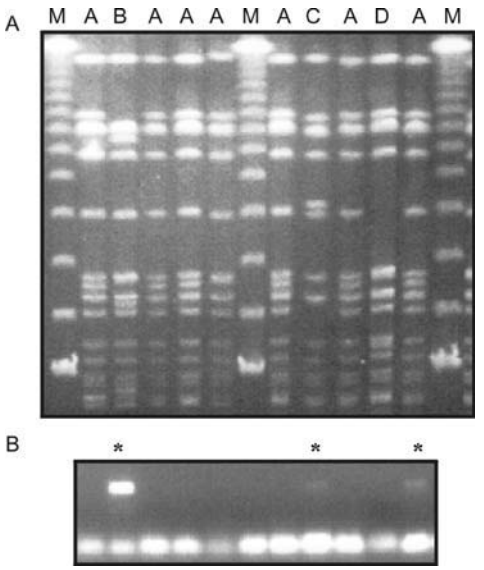


FIGURE 30.2 Gel electrophoretic analysis of genetic diversity and the presence of a methicillin resistance gene in strains of *Staphylococcus aureus*. **A.** Fingerprints generated by pulsed field gel electrophoresis (PFGE) of DNA macrorestriction fragments. Four different characteristic banding patterns are observed (indicated by the letters A to D, respectively, above each lane); these serve the purpose of unequivocal determination of relatedness between strains. Lanes marked M contain concatemeric phage lambda genomes, the smallest one being 50,000 bp in size. **B.** PCR amplification of part of the mecA gene, the gene encoding the penicillin binding protein 2 A, the product of which shows diminished affinity toward the antibiotic; hence its causal involvement in resistance. Three positive reactions are observed (marked with an asterisk).

Presently, also the allelic variation in a single staphylococcal gene can be used for epidemiological tracing of isolates within the species. The protein A-encoding *spa* genes contain a repeat region where the number of repeats is not the single variant, also the repeat sequence itself varies in sequences. So, on the basis of the number of repeats and their sequence identity, a binary code can be attributed to a strain (Strommenger *et al*., 2008). This method is highly reproducible (Aires de Sousa *et al.*, 2006) and has already generated massive amounts of well-communicable data. A dedicated website exists and the *spa* typing is widely advocated as a tool that should bring national reference centers closer to their customers and, also, to their colleague centers. Even more importantly, *spa* typing is now suggested to be the first system that could possibly facilitate real-time typing in clinical settings (Mellmann *et al*., 2006). Whether this is true or false needs to be defined in well-controlled studies still.

30.6 FUTURE PERSPECTIVES

Nucleic acid-based tests are now being introduced with increasing speed into routine clinical microbiology laboratories (Check, 2001). Part of the initial and obvious delay was due to the inadequately perceived and prejudiced shortcomings of nucleic acid testing; it was supposed to be laborious, expensive, and requiring high levels of laboratory expertise (Vaneechoutte and Van Eldere, 1997). However, these historic objections are slowly being taken apart, and accelerated introduction of molecular diagnostics should be pursued in many cases. PCR testing for *Legionalla pneumophila*, the agent of Legionnaires' disease, for instance (sensitivity 80–100%, specificity more than 90%), improves significantly over culture (sensitivity 10–80%, specificity 100%) and urine antigen testing (sensitivity 70–90%, specificity 99%, Murdoch, 2003). The same is true for *Trichomonas vaginalis*, an example of a sexually transmitted parasite. Several PCR tests showing increased sensitivity and excellent specificity have been described (Van Der Schee *et al*., 1999), even using urine instead of an invasive swab as clinical specimen. Why have these tests not been implemented immediately and massively in clinical microbiology? The most frequently perceived comments in this respect still concern the supposed levels of complexity and elevated costs of the molecular tests. It is becoming clear that there is cost effectiveness in the molecular technology. In addition, it is regularly shown that clear improvement in the sensitivity and specificity of clinical testing is achieved by introducing molecular tests, so swift introduction of such tests into clinical practice will remain important to pursue.

Another important future application of molecular testing will be in the realm of host susceptibility toward infections (Relman, 2002). When genetic profiling tests for humans become more widely available (and their implications better understood) all infectious disease specialists will ultimately profit from these new molecular services. Many examples of genes important in host defense against infections have been described already (Santos *et al*., 2002; Leveque *et al*., 2003; and many, many more!!). It will be interesting to see where these exciting developments will lead over the coming decade. Whether or not this type of genetic data ultimately should be collected for each and every individual is already now a matter of intense, ethical debate. It is anticipated that for certain diseases the identification of host susceptibility factors will eventually replace microbial diagnostics.

Genomics and proteomics, not to forget metabolomics, transcriptomics, and all of the other ''omics'' sciences, have started to dominate the microbiology field over the past five years. The availability of complete genome sequences for both bacteria and the larger viruses has opened new avenues of fundamental and applied research (Cummings *et al*., 2002). From the diagnostic perspective, genomics has facilitated the characterization of novel diagnostic and epidemiologically intriguing target sequences, whereas proteomics has facilitated the identification of complete protein profiles expressed by pathogens under different environmental conditions. The use of high-throughput ''omics'' methods has enhanced our understanding of infectious diseases

significantly and the emphasis for the coming years should be on functional ''omics'' approaches, where microbiological ''omics'' data will be linked to microbial phenotypes or features of the host pathogen interaction.

30.7 CONCLUDING REMARKS

It needs to be emphasized that in principle the issues covered in this chapter for some very specific pathogens can be extrapolated to species and isolates of each and every other microbial infectious disease agent. This exemplifies the beauty of molecular diagnostics: the focused, nearly universally applicable technology is not only suited for the detection of pathogens; the subsequent genetic profiling of pathogens and assessment of their virulence potential is enabled by the same technology. Moreover, homologous technology also can be used to measure host response and eventually predict host susceptibility toward infectious diseases. It is good to witness that an obvious systems biology approach emerges upon the introduction of all of these new tools in the microbiology laboratory. The outcome of molecular diagnostic studies will continue to improve our understanding of infectious disease over the coming years. Ours are exciting times!!

REFERENCES

Aires de Sousa, M., and De Lencastre, H. (2003). Evolution of sporadic isolates of methicillin resistant *Staphylococcus aureus* (MRSA) in hospitals and their similarities to isolates of community acquired MRSA. *J. Clin. Microbiol.* 41, 3806–3815.

Aires-de-Sousa, M., Boye, K., de Lencastre, H., Deplano, A., Enright, M. C., Etienne, J., Friedrich, A., Harmsen, D., Holmes, A., Huijsdens, X.W., Kearns, A.M., Mellmann, A., Meugnier, H., Rasheed, J. K., Spalburg, E., Strommenger, B., Struelens, M.J., Tenover, F.C., Thomas, J., Vogel, U., Westh, H., Xu, J., and Witte, W. (2006). High interlaboratory reproducibility of DNA sequence-based typing of bacteria in a multicenter study. *J. Clin. Microbiol.* 44, 619–621.

An, Q., Liu, J., O'Brien, W., Radcliffe, G., Buxton, D., Popoff, S., King, W., Vera-Garcia, M., Lu, L., and Shah, J. (1995). Comparison of characteristics of Q beta replicase-amplified assay with competitive PCR assay for *Chlamydia trachomatis*. *J. Clin. Microbiol.* 33, 58–63.

Apfalter, P., Assadian, O., Kalczyk, A., Lindenmann, V., Makristathis, A., Mustafa, S., Rotter, M., and Hirschl, A.M. (2002). Performance of a new chromogenic oxacillin resistance screen medium (Oxoid) in the detection and presumptive identification of methicillin-resistant *Staphylococcus aureus*. *Diagn. Microbiol. Infect. Dis.* 44, 209–211.

Arbique, J., Forward, K., Haldane, D., and Davidson, R. (2001). Comparison of the Velogene Rapid MRSA identification assay, Denka MRSA Screen assay and BBL Crystal MRSA ID system for rapid identification of methicillin-resistant *Staphylococcus aureus*. *Diagn. Microbiol. Infect. Dis.* 40, 5–10.

Baba, T., Tajeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K., and Hiramatsu, K. (2002). Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* 359, 1819–1827.

Bang, D., Angelsø, L., Schirakow, B., and Westh, H. (2003). Comparison of the Becton Dickinson strand displacement amplification and Cobas Amplicor Roche PCR for the detection of *Chlamydia trachomatis*: pooling versus individual tests. *Clin. Microbiol. Infect.* 9, 1020–1023.

Barken, K.B., Haagensen, J.A., and Tolker-Nielsen, T. (2007). Advances in nucleic acid-based diagnostics of bacterial infections. *Clin. Chim. Acta.* 384, 1–11.

Bidet, P., Brahimi, N., Chalas, C., Aujard, Y., and Bingen, E. (2003). Molecular characterisation of serotype III Group B Streptococcal isolates causing neonatal meningitis. *J. Infect. Dis.* 188, 1132–1137.

Blanc, D.S., Wenger, A., and Bille, J. (2003). Evaluation of a novel medium for screening specimens from hospitalised patients to detect methicillin-resistant *Staphylococcus aureus*. *J. Clin. Microbiol.* 41, 3499–3502.

Blocker, M.E., Krysiak, R.G., Behets, F., Cohen, M.S., and Hobbs, M.M. (2002). Quantification of *Chlamydia trachomatis* elementary bodies in urine by ligase chain reaction. *J. Clin. Microbiol.* 40, 3631–3634.

Boom, R., Sol, C.J.A., Salimans, M.M.M., Jansen, C.L., Wertheim-van Dillen, P.M.E., and Van der Noordaa, J. (1990). Rapid and simple method for the purification of nucleic acids. *J. Clin. Microbiol.* 28, 495–503.

Borg, I., Rohde, G., Loseke, S., Bittscheidt, J., Schultze-Werninghaus, G., Stephan, V., and Bufe, A. (2003). Evaluation of a quantitative real-time PCR for the detection of respiratory syncytial virus in pulmonary diseases. *Eur. Respir. J.* 21, 944–951.

Boyce, J.M., and Havill, N.L. (2008). Comparison of BD GeneOhm methicillin-resistant Staphylococcus aureus (MRSA) PCR versus the CHROMagar MRSA assay for screening patients for the presence of MRSA strains. *J. Clin. Microbiol.* 46, 350–351.

Brackstad, O.G., Maeland, J.A., and Tveten, Y. (1993). Multiplex PCR for detection of genes for *Staphylococcus aureus* thermonuclease and methicillin resistance and correlation with oxacillin resistance. *APMIS* 101, 681–688.

Caan, W. (2006). Preterm delivery in primiparous women at low risk: could epidemic chlamydia contribute to rise in preterm births? *BMJ* 332, 1094.

Check, W. (2001). Nucleic acid-based tests move slowly into clinical labs. *ASM News* 67, 560–565.

Cheema, M.A., Schumacher, H.R., and Hudson, A.P. (1991). RNA-directed molecular hybridization screening: evidence for inapparent chlamydial infection. *Am. J. Med. Sci.* 302, 261–268.

Chernesky, M.A., Lee, H., Schachter, J., Burczak, J.D., Stamm, W.E., McCormack, W.M., and Quinn, T.C. (1994). Diagnosis of *Chlamydia trachomatis* urethral infection in symptomatic and asymptomatic men by testing first-void urine in a ligase chain reaction assay. *J. Infect. Dis.* 170, 1308–1311.

Cockerill, F.R., III, and Smith, T.F. (2002). Rapid real-time PCR: a revolution for clinical microbiology. *ASM News* 68, 77–83.

Cosentino, L.A., Landers, D.V., and Hillier, S.L. (2003). Detection of Chlamydia trachomatis and Neisseria gonorrhoeae by strand displacement amplification and relevance of the amplification control for use with vaginal swab specimens. *J. Clin. Microbiol.* 41, 3592–3596.

Cummings, L., Riley, L., Black, L., Souvorov, A., Reschenchuk, S., Dondoshka, A., and Tatusova, T. (2002). Genomic BLAST: custom defined virtual databases for complete and unfinished genomes. *FEMS Microbiol. Lett.* 216, 133–138.

Day, N.P., Moore, C.E., Enright, M.C., Berendt, A.R., Smith, J.M., Murphy, M., Peacock, S.J., Spratt, B.G., and Feil, E.J. (2001). A link between virulence and ecological abundance in natural populations of *Staphylococcus aureus*. *Science* 292, 114–116.

Dean, D., Pant, C.R., and O'Hanley, P. (1989). Improved sensitivity of a modified polymerase chain reaction amplified DNA probe in comparison with serial tissue culture passage for detection of Chlamydia trachomatis in conjunctival specimens from nepal. *Diagn. Microbiol. Infect. Dis.* 12, 133–137.

De Lencastre, H., Severina, E.P., Roberts, R.B., Kreiswirth, B.N., and Tomasz, A. (1996). Testing the efficacy of a molecular surveillance network: MRSA and VREF genotyping in six hospitals in the metropolitan New York area. The Emerging Infectious Diseases Initiative Pilot Group. Bacterial Antibiotic Resistance Surveillance Group. *Microb. Drug Resist.* 2, 343–351.

Demain, A.L. (1999). Achievements in microbial technology. *Biotechnol. Adv.* 8, 291–301.

Dille, B.J., Butzen, C.C., and Birkenmeyer, L.G. (1993). Amplification of *Chlamydia trachomatis* DNA by ligase chain reaction. *J. Clin. Microbiol.* 31, 729–731.

Enright, M.C., Robinson, D.A., Randle, G.E., Feil, E.J., Grundmann, H., and Spratt, B.G. (2002). The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proc. Natl. Acad. Sci. USA,* 99, 7687–7692.

Espy, M.J., Uhl, J.R., Mitchell, P.S., Thorvilson, J.N., Svien, K.A., Wold, A.D., and Smith, T.F. (2000). Diagnosis of herpes simplex virus infections in the clinical laboratory by LightCycler PCR. *J. Clin. Microbiol.* 38, 795–799.

Fang, H., and Hedin, G. (2003). Rapid screening and identification of methicillin-resistant *Staphylococcus aureus* from clinical samples by selective-broth and real-time PCR. *J. Clin. Microbiol.* 41, 2894–2899.

Feil, E.J., Cooper, J.E., Grundmann, H., Robinson, D.A., Enright, M. C., Berendt, T., Peacock, S.J., Maynard Smith, J., Murphy, M., Spratt, B.G., Moore, C.E., and Day, N.P.J. (2003). How clonal is Staphylococcus aureus? J. *Bacteriol.* 185, 3307–3316.

Fitzgerald, J.R., Sturdevant, D.E., Mackie, S.M., Gill, S.R., and Musser, J. M. (2001). Evolutionary genomics of *Staphylococcus aureus*: insight into the origin of MRSA and the toxic shock syndrome epidemic. *Proc. Natl. Acad. Sci. USA,* 98, 8821–8826.

Francois, P., Huyghe, A., Charbonnier, Y., Bento, M., Herzig, S., Topolski, I., Fleury, B., Lew, D., Vaudaux, P., Harbarth, S., van Leeuwen, W., van Belkum, A., Blanc, D.S., Pittet, D., and Schrenzel, J. (2005). Use of an automated multiple-locus, variable-number tandem repeat-based method for rapid and high-throughput genotyping of Staphylococcus aureus isolates. *J. Clin. Microbiol.* 43, 3346–3355.

Francois, P., Pittet, D., Bento, M., Pepey, B., Vaudaux, P., Lew, D., and Schrenzel, J. (2003). Rapid detection of methicillin resistant *Staphylococcus aureus* directly from sterile or non-sterile clinical samples by a new molecular assay. *J. Clin. Microbiol.* 41, 254–260.

Gillet, Y., Issartel, B., Vanhems, P., Fournet, J.C., Lina, G., Bes, M., Vandenesch, J., Piemont, Y., Brousse, N., Floret, D., and Etienne, J. (2002). Association between *Staphylococcus aureus* strains carrying the genes for Panton-Valentine leukocidin and highly lethal necrotising pneumoniae in young immunocompetent patients. *Lancet* 359, 753–759.

Gilpin, D.F., Tunney, M.M., Funston, C., Savage, K., Gardiner, A., and Kearney, M.P. (2007). Rapid detection of MRSA in a routine diagnostic laboratory using a real-time PCR assay. *J. Hosp. Infect.* 67, 97–99.

Goessens, W.H.F., Kluytmans, J.A., Den Toom, N., Van Rijsoort-Vos, T., Niesters, B.G., Stolz, E., Verbrugh, H.A., and Quint, W.G.V. (1995). Influence of volume of sample processed on detection of *Chlamydia trachomatis* in urogenital samples by PCR. *J. Clin. Microbiol.* 33, 251–253.

Goessens, W.H.F., Mouton, J.W., Van der Meijden, W.I., Deelen, S., Van Rijsoort-Vos, T., Lemmens-den Toom, N., Verbrugh, H.A., and Verkooijen, R.P. (1997). Comparison of three commercially available amplification assays, AMP CT, LCx, and COBAS AMPLICOR for detection of *Chlamydia trachomatis* in first-void urine. *J. Clin. Microbiol.* 35, 2628–2633.

Griffais, R., and Thibon, M. (1989). Detection of *Chlamydia trachomatis* by the PCR. *Res. Microbiol.* 140, 139–141.

Gueudin, M., Vabret, A., Petitjean, J., Gouarin, S., Brouard, J., and Freymuth, F. (2003). Quantitation of respiratory syncytial virus RNA in nasal aspirates of children by real-time RT-PCR assay. *J. Virol. Methods,* 109, 39–45.

Hallin, M., Maes, N., Byl, B., Jacobs, F., De Gheldre, Y., and Struelens, M.J. (2003). Clinical impact of a PCR assay for identification of *Staphylococcus aureus* and determination of methicillin resistance directly from blood cultures. *J. Clin. Microbiol.* 41, 3942–3944.

Harmsen, D., Singer, C., Rothganger, J., Tonjum, T., De Hoog, G.S., Shah, A.J., and Frosch, M. (2001). Diagnostics of neisseriaeae and moraxellaceae by ribosomal DNA sequencing: ribosomal differentiation of medical microorganisms. *J. Clin. Microbiol.* 39, 936–942.

Herbrink, P., Van den Munckhof, H.A., Niesters, H.G., Goessens, W.H.F., Stolz, E., and Quint, W.G.V. (1995). Solid phase C1q-directed bacterial capture followed by PCR for detection of *Chlamydia trachomatis* in clinical specimens. *J. Clin. Microbiol.* 33, 283–286.

Iqbal, J., Hira, P.R., Al-Ali, F., Khalid, N., and Sher, A. (2003). Modified Giemsa staining for rapid diagnosis of malaria infections. *Med. Princ. Pract.* 12, 156–159.

Ito, T., Katayama, Y., Asada, K., Mori, N., Tsutsumimoto, K., Tiensasitorn, C., and Hiramatsu, K. (2001). Structural comparison of three types of staphylococcal cassette chromosome *mec* integrated in the chromosome of methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 45, 1323–1336.

Jonas, D., Speck, M., Daschner, F.D., and Grundmann, H. (2002). Rapid PCR based identification of methicillin resistant *Staphylococcus aureus* from screening swabs. *J. Clin. Microbiol.* 40, 1821–1823.

Kluytmans, J. (2007). Control of methicillin-resistant Staphylococcus aureus (MRSA) and the value of rapid tests. *J. Hosp. Infect.* 65 Suppl. 2, 100–104.

Knapp, C.C., Ludwig, M.D., and Washington, D.A. (1994). Evaluation of differential inoculum disk diffusion method and Vitek GPS-SA card for detection of oxacillin resistant staphylococci. *J. Clin. Microbiol.* 32, 433–436.

Ko, K.S., Kim, J.M., Kim, J.W., Jung, B.Y., Kim, W., Kim, I.J., and Kook, Y.H. (2003). Identification of *Bacillus anthracis* by rpoB sequence analysis and multiplex PCR. *J. Clin. Microbiol.* 41, 2908–2914.

Koning, S., Van Belkum, A., Snijders, S., Van Leeuwen, W., Verbrugh, H., Nouwen, J., Op 't Veld, M., Suijlekom-Smit, L.W., Van der Wouden, J., and Verduin, C. (2003). Severity of nonbullous *Staphylococcus aureus* impetigo in children is associated with strains harboring genetic markers for exfoliative toxin B, Panton Valentine leukocidin and the multidrug resistance plasmid pSK41. *J. Clin. Microbiol.* 41, 3017–3021.

Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Urbani, C., Comer, J.A., Lim, W., Rollin, P.E., Dowell, F.S., Ling, A.F., Humar, C.D., Shieh, W.J., Guarner, J., Paddock, C.D., Rota, P., Fields, B., DeRisi, J., Cox, N., Hughes, J.M., LeDuc, J.W., Bellini, W.J., Anderson, L.J., and the SARS Working Group (2003). A novel coronavirus associated with severe acute respiratory syndrome. *New Engl. J. Med.,* 348, 1953–1966.

Kubina, M., Jaulhac, B., Delabranche, X., Lindenmann, C., Piemont, Y., and Monteil, H. (1999). Oxacillin susceptibility testing of staphylococci directly from Bactec Plus blood cultures by the BBL Crystal MRSA ID system. *J. Clin. Microbiol.* 37, 2034–2036.

Kumar, P.S., Griffen, A.L., Barton, J.A., Paster, B.J., and Moeschberger, M.L. (2003). New bacterial species associated with chronic periodontitis. *J. Dent. Res.* 82, 338–344.

Kuusela, P., Hilden, P., Savolainen, K., Vuento, M., Lyytikainen, O., and Vuopio-Varkila, J. (1994). Rapid detection of methicillin-resistant *Staphylococcus aureus* strains not identified by slide agglutination tests. *J. Clin. Microbiol.* 32, 143–147.

Lar, J., Walboomers, J.M., Roosendaal, R., Van Doornum, G.J., MacLaren, D.M., Meijer, C.J., and Van den Brule, A.J. (1993). Direct detection and genotyping of *Chlamydia trachomatis* in cervical scrapes by using PCR and restriction fragment length polymorphism analysis. *J. Clin. Microbiol.* 31, 1060–1065.

Leslie, D.E., Azzato, F., Ryan, N., and Fyfe, J. (2003). An assessment of the Roche Amplicor *Chlamydia trachomatis/Neisseria gonorrhoeae* multiplex PCR assay in diagnostic use on a variety of specimen types. *Commun. Dis. Intell.* 27, 373–379.

Leveque, G., Forgetta, V., Morroll, S., Smith, A.L., Bumstead, N., Barrow, P., Loredo-Osti, J.C., Morgan, K., and Malo, D. (2003). Allelic variation in TLR4 is linked to susceptibility to *Salmonella enterica* serovan Typhimurium infection in chickens. *Infect Immun.* 71, 1116–1124.

Levi, K., Bailey, C., Bennett, A., Marsh, P., Cardy, D.L.N., and Towner, K. J. (2003). Evaluation of an isothermal signal amplification method for rapid detection of methicillin-resistant *Staphylococcus aureus* from patient-screening swabs. *J. Clin. Microbiol.* 41, 3187–3191.

Levi, K., and Towner, K.J. (2003). Detection of methicillin resistant *Staphylococcus aureus* (MRSA) in blood with the EVIGENE MRSA detection kit. *J. Clin. Microbiol.* 41, 3890–3892.

Louie, L., Matsumara, S.O., Choi, E., Louie, M., and Simor, A.E. (2000). Evaluation of three rapid methods for detection of methicillin resistance in *Staphylococcus aureus*. *J. Clin. Microbiol.* 38, 2170–2173.

Lunge, V.R., Miller, B.J., Livak, K.J., and Batt, C.A. (2002). Factors affecting the performance of the 5¢ nuclease PCR assay for *Listeria monocytogenes* detection. *J. Microbiol. Methods,* 51, 361–368.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., and Spratt, B.G. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA,* 95, 3140–3145.

McLaughlin, R.W., Vali, H., Lau, P.C.K., Palfree, R.G.E., De Ciccio, A., Sirois, M., Ahmad, D., Villemur, R., Desrosiers, M., and Chan, E. C.S. (2002). Are there naturally occurring pleomorphic bacteria in the blood of healthy humans? *J. Clin. Microbiol.* 40, 4771–4775.

McOrist, A.L., Jackson, M., and Bird, A.R. (2002). A comparison of five methods for extraction of bacterial DNA from human faecal samples. *J. Microbiol. Methods,* 50, 131–139.

Meier, A., Persing, D.H., Finken, M., and Bottger, E.C. (1993). Elimination of contaminating DNA within polymerase chain reaction reagents: implications for a general approach to detection of uncultured pathogens. *J. Clin. Microbiol.* 31, 646–652.

Mellmann, A., Friedrich, A.W., Rosenkötter, N., Rothgänger, J., Karch, H., Reintjes, R., and Harmsen, D. (2006). Automated DNA sequence-based early warning system for the detection of methicillin-resistant Staphylococcus aureus outbreaks. *PLoS Med.* 3, e33.

Merlino, J., Leroi, M., Bradbury, R., Veal, D., and Harbour, C. (2000). New chromogenic identification and detection of *Staphylococcus aureus* and methicillin resistant *S. aureus. J. Clin. Microbiol.* 38, 2378–2380.

Millar, B.C., Xu, J., and Moore, J.E. (2002). Risk assessment models and contamination management: implications for broad-range ribosomal DNA PCR as a diagnostic tool in medical bacteriology. *J. Clin. Microbiol.* 40, 1575–1580.

Mir, N., Sanchez, M., Baquero, F., Lopez, B., Canlderon, C., and Canton, R. (1998). Soft salt-mannitol agar cloxacillin test: a highly specific bedside screening test for detection of colonization with methicillin resistant *Staphylococcus aureus*. *J. Clin. Microbiol.* 36, 986–989.

Miyamoto, H., Imamura, K., Kojima, A., Takenaka, H., Hara, N., Ikenouchi, A., Tanabe, T., and Taniguchi, H. (2003). Survey of nasal colonisation by, and assessment of novel multiplex PCR method for detection of biofilm-forming methicillin-resistant staphylococci in healthy medical students. *J. Hosp. Infect.* 53, 215–223.

Mongodin, E., Finan, J., Climo, M.W., Rosato, A., Gill, S.R., and Archer, G.L. (2003). Microarray transcription analysis of clinical *Staphylococcus aureus* isolates resistant to vancomycin. *J. Bacteriol.* 185, 4638–4643.

Murakami, K., Minamide, W., Wada, K., Nakamura, E., Teraoka, A., and Watanabe, S. (1991). Identification of methicillin resistant strains of *Staphylococcus aureus* by polymerase chain reaction. *J. Clin. Microbiol.* 29, 2240–2244.

Murchan, S., Kaufmann, M.E., Deplano, A., De Ryck, R., Struelens, M., Fussing, V., Salmenlinna, S., Vuopio-Varkila, J., El Solh, N., Cuny, C., Witte, W., Tassios, P.T., Legakis, N., Van Leeuwen, W., Van Belkum, A., Vindel, A., Garaizar, J., Haeggman, S., Olsson Lilljequist, B., Ransjo, U., Coombs, J., and Cookson, B.D. (2003). Harmonization of PFGE protocols for epidemiological typing of strains of methicillin resistant *Staphylococcus aureus*: a single approach developed by consensus in ten European laboratories and its application for tracing the spread of related strains. *J. Clin. Microbiol.* 41, 1574–1585.

Murdoch, D.R. (2003). Diagnosis of *Legionella* infection. *Clin. Infect. Dis.* 36, 64–69.

Mwangi, M.M., Wu, S.W., Zhou, Y., Sieradzki, K., de Lencastre, H., Richardson, P., Bruce, D., Rubin, E., Myers, E., Siggia, E.D., and Tomasz, A. (2007). Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing. *Proc. Natl. Acad. Sci. USA,* 104, 9451–9456.

Niesters, H.G. (2002). Clinical virology in real time. *J. Clin. Virol.* 125, 3–12.

Niesters, H.G. (2004). Molecular and diagnostic clinical virology in real time. *Clin. Microbiol. Infect.* 10, 5–11.

Nikkari, S., McLaughlin, I.J., Bi, W., Dodge, D.E., and Relman, D.A. (2001). Does blood from healthy human subjects contain bacterial DNA?. *J. Clin. Microbiol.* 39, 1956–1959.

Oberste, M.S., Maher, K., Flemister, M.R., Marchetti, G., Kilpatrick, D. R., and Pallansch, M.A. (2000). Comparison of classic and molecular approaches for the identification of untypeable enteroviruses. *J. Clin. Microbiol.* 38, 1170–1174.

Oliveira, D.C., and De Lencastre, H. (2002). Multiplex PCR strategy for rapid identification of structural types and variants of the *mec* element in methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 46, 2155–2161.

Paavonen, J., Puolakkainen, M., Paukku, M., and Sintonen, H. (1998). Cost benefit analysis of first void urine *Chlamydia trachomatis* during a screening program. *Obstet. Gynecol.* 92, 292–298.

Pannekoek, Y., Westenberg, S.M., Eijk, P.P., Repping, S., Van der Veen, F., Van der Ende, A., and Dankert, J. (2003). Assessment of *Chlamydia trachomatis* infection of semen specimens by ligase chain reaction. *J. Med. Microbiol.* 52, 777–779.

Pasternack, R., Vuorinen, P., and Miettinen, A. (1999). Comparison of a transcription-mediated amplification assay and PCR for the detection of *Chlamydia trachomatis* in first void urine. *Eur. J. Clin. Microbiol. Infect. Dis.* 18, 142–144.

Perez-Roth, E., Claverie-Martin, F., Villar, J., and Mendez- Alvarez, S. (2001). Multiplex PCR for simultaneous identification of *Staphylococcus aureus* and detection of methicillin and mupirocin resistance. *J. Clin. Microbiol.* 39, 4037–4041.

Pollard, D.R., Tyler, S.D., Ng, C.W., and Kozee, K.R. (1989). A PCR protocol for the specific detection of *Chlamydia* spp. *Mol. Cell. Probes,* 3, 383–389.

Qadri, S.M., Ueno, Y., Imambaccus, H., and Almodovar, E. (1994). Rapid detection of MRSA by Crystal MRSA ID system. *J. Clin. Microbiol.* 32, 1830–1832.

Quint, W.G.V., Heijtink, R.A., Schirm, J., Gerlich, W.H., and Niesters, H. G.M. (1995). Reliability of methods for hepatitis B virus DNA detection. *J. Clin. Microbiol.* 33, 225–228.

Reijans, M., Dingemans, G., Klaassen, C.H., Meis, J.F., Keijdener, J., Mulders, B., Eadie, K., van Leeuwen, W., van Belkum, A., Horrevorts, A.M., and Simons, G. (2008). RespiFinder: a new multiparameter test to differentially identify 15 respiratory viruses. *J. Clin. Microbiol.* (in press).

Relman, D.A. (1998). Detection and identification of previously unrecognised microbial pathogens. *Emerg. Infect. Dis.* 4, 382–389.

Relman, D.A. (2002). New technologies, human-microbe interactions, and the search for previously unrecognized pathogens. *J. Infect. Dis.* 186, S254–S258.

Riemersma, W.A., Van der Schee, C.J., Van der Meijden, W., Verbrugh, H.A., and Van Belkum, A. (2003). Microbial population diversity in the urethras of healthy men and males suffering from nonchlamydial, nongonococcal urethritis. *J. Clin. Microbiol.* 41, 1977–1986.

Roth, A., Reischl, U., Streubel, A., Naumann, L., Kroppenstedt, R.M., Haber, Z., Fischer, M., and Mauch, H. (2000). Novel diagnostic algorithm for identification of mycobacteria by genus specific amplification of the 16S-23S rRNA gene spacer and application of restriction endonuclease. *J. Clin. Microbiol.* 38, 1094–1104.

Rours, I.G., Hammerschlag, M.R., Ott, A., De Faber, T.J., Verbrugh, H.A., de Groot, R., and Verkooyen, R.P. (2008). Chlamydia trachomatis as a cause of neonatal conjunctivitis in Dutch infants. *Pediatrics* 121, e321–e326.

Rours, G.I., Verkooyen, R.P., Willemse, H.F., van der Zwaan, E.A., van Belkum, A., de Groot, R., Verbrugh, H.A., and Ossewaarde, J.M. (2005). Use of pooled urine samples and automated DNA isolation to achieve improved sensitivity and cost-effectiveness of large-scale testing for Chlamydia trachomatis in pregnant women. *J. Clin. Microbiol.* 43, 4684–4690.

Rys, P.N., and Persing, D.H. (1993). Preventing false positives: quantitative evaluation of three protocols for inactivation of PCR amplification products. *J. Clin. Microbiol.* 31, 2356–2360.

Sabat, A., Krzyszto-Russjan, J., Strzalka, W., Filipek, R., Kusowska, K., Hryniewicz, W., Travis, J., and Potemka, J. (2003). New method for typing *Staphylococcus aureus* strains: multiple locus variable number of tandem repeat analysis of polymorphism and genetic relatedness of clinical isolates. *J. Clin. Microbiol.* 41, 1801–1804.

Safdar, N., Narans, L., Gordon, B., and Maki, D.G. (2003). Comparison of culture screening methods for detection of nasal carriage of methicillin-resistant *Staphylococcus aureus*: a prospective study comparing 32 methods. *J. Clin. Microbiol.* 41, 3163–3166.

Santos, A.R., Suffys, P.N., Vanderborght, P.R., Moraes, M.O., Vieira, L. M.M., Cabello, P.H., Bakker, A.M., Matos, H.J., Huizinga, T.W.J., Ottenhoff, T.H.M., Sampaio, E.P., and Sarno, E.N. (2002). Role of tumour necrosis factor-alpha and interleukin 10 promoter gene polymorphisms in leprosy. *J. Infect. Dis.* 186, 1687–1691.

Savolainen, C., Blomqvist, S., Mulders, M.N., and Hovi, T. (2002). Genetic clustering of all 102 human rhinovirus prototype strains: serotype 87 is close to human enterovirus 70. *J. Gen. Virol.* 83, 333–340.

Schachter, J. (1985). Overview of *Chalamydia trachomatis* infection and the requirement for a vaccine. *Rev. Infect. Dis.* 7, 713–716.

Schutten, M., and Niesters, H.G.M. (2001). Clinical utility of viral quantification as a tool for disease monitoring. *Expert. Rev. Mol. Diagn.* 2, 153–162.

Schuurman, R., Descamps, D., Weverling, G.J., Kaye, S., Tijnagel, J., Williams, I., Van Leeuwen, R., Tedder, R., Boucher, C.A., Brun-Vezinet, F., and Loveday, C. (1996). Multicenter comparison of three commercial methods for quantification of human immunodeficiency virus type 1 RNA in plasma. *J. Clin. Microbiol.* 34, 3016–3022.

Schuurman, T., de Boer, R., Patty, R., Kooistra-Smid, M., and van Zwet, A. (2007). Comparative evaluation of in-house manual, and commercial semi-automated and automated DNA extraction platforms in the sample preparation of human stool specimens for a Salmonella enterica 5′-nuclease assay. *J. Microbiol. Methods.* 71, 238–245.

Seme, K., Mocilnik, T., Komlos, K.F., Doplihar, A., Persing, D.H., and Poljak, M. (2008). GeneXpert Enterovirus assay: one-year experience in a routine laboratory setting and evaluation on three proficiency panels. *J. Clin. Microbiol.* (in press).

Shah, J.S., Liu, J., Smith, J., Popoff, S., Radcliffe, G., O'Brien, W. J., Serpe, G., Olive, D.M., and King, W. (1994). Novel ultrasensitive Q beta-replicase amplified hybridization assay for detection of *Chlamydia trachomatis*. *J. Clin. Microbiol.* 32, 2718–2724.

Shrestha, N.K., Shermock, K.M., Gordon, S.M., Tuohy, M.J., Wilson, D.A., Cwynar, R.E., Banbury, M.K., Longworth, D.L., Isada, C.M., Mawhorter, S.D., and Procop, G.W. (2003). Predictive value and cost-effectiveness analysis of rapid polymerase chain reaction for preoperative detection of nasal carriage of *Staphylococcus aureus*. *Infect. Control Hosp. Epidemiol.* 24, 327–333.

Singh, K., Gavin, P.J., Vescio, T., Thomson, R.B., Deddish, R.B., Fisher, A., Noskin, G.A., and Peterson, L.R. (2003). Microbiologic surveillance using nasal cultures alone is sufficient for detection of methicillin-resistant *Staphylococcus aureus* in neonates. *J. Clin. Microbiol.* 41, 2755–2757.

Stefano, J.E., Genovese, L., An, Q., Lu, L., McCarty, J., Du, Y., Stefano, K., Burg, J.L., King, W., and Lane, D.J. (1997). Rapid and sensitive detection of *Chlamydia trachomatis* using a ligatable binary RNA probe and Q beta replicase. *Mol. Cell. Probes,* 11, 407–426.

Stranden, A., Frei, R., and Widmer, A.F. (2003). Molecular typing of methicillin-resistant *Staphylococcus aureus*: can PCR replace pulsed-field gel electrophoresis? *J. Clin. Microbiol.* 41, 3181–3186.

Strommenger, B., Braulke, C., Heuck, D., Schmidt, C., Pasemann, B., Nübel, U., and Witte, W. (2008). Spa typing of Staphylococcus aureus as a frontline tool in epidemiological typing. *J. Clin. Microbiol.* 46, 574–581.

Suchland, R.J., Geisler, W.M., and Stamm, W.E. (2003). Methodologies and cell lines used for antimicrobial susceptibility testing of *Chlamydia* spp. *Antimicrob. Agents Chemother.* 47, 636–642.

Tenover, F.C. (1993). DNA hybridisation techniques and their application to the diagnosis of infectious diseases. *Infect. Dis. Clin. North Am.* 7, 171–181.

Tenover, F.C., Vaughn, R.R., McDougal, L.K., Fosheim, G.E., and McGowan, J.E., Jr (2007). Multiple-locus variable-number tandem-repeat assay analysis of methicillin-resistant Staphylococcus aureus strain. *J. Clin. Microbiol.* 45, 2215–2219.

Towner, K.J., Talbot, J.C., Curran, R., Webster, C.A., and Humphreys, H. (1998). Development and validation of a PCR-based immunoassay for the rapid detection of methicillin resistant *Staphylococcus aureus*. *J. Med. Microbiol.* 47, 607–613.

Valentine-Thon, E., Van Loon, A.M., Schirm, J., Reid, J., Klapper, P.E., and Cleator, G.M. (2001). European Proficiency Testing Program for

Molecular Detection and Quantitation of Hepatitis B Virus DNA. *J. Clin. Microbiol.* 39, 4407–4412.

van Belkum, A., Bax, R., Peerbooms, P., Goessens, W., Van Leeuwen, N., and Quint, W.G.V. (1993). Comparison of phage typing and DNA fingerprinting by polymerase chain reaction for discrimination of methicillin-resistant *Staphylococcus aureus*. *J. Clin. Microbiol.* 31, 798–803.

van Belkum, A., Kluytmans, J., Van Leeuwen, W., Bax, R., Quint, W., Peters, E., Fluit, A., Vandenbroucke Grauls, C., Van den Brule, A., Koeleman, H., Elaichoune, A., Vaneechoutte, M., Tenover, F., and Verbrugh, H. (1995). Multicenter evaluation of arbitrarily primed PCR for typing of *Staphylococcus aureus* strains. *J. Clin. Microbiol.* 33, 1537–1547.

van Belkum, A., Scherer, S., Van Alphen, L., and Verbrugh, H. (1998a). Short sequence repeats in prokaryotic genomes. *Microbiol. Mol. Biol. Rev.* 62, 275–293.

van Belkum, A., Van Leeuwen, W., Kaufmann, M.E., Cookson, B.D., Forey, F., Etienne, J., Goering, R., Tenover, F., Steward, C., O'Brien, F., Grubb, W., Legakis, N., Morvan, A., El Solh, N., De Ryck, R., Struelens, M., Salmenlinna, S., Vuopio-Varkila, J., Koositra, M., Witte, W., and Verbrugh, H. (1998b). Assessment of resolution and intercenter reproducibility of genotyping of *Staphylococcus aureus* by pulsed field gel electrophoresis of SmaI macrorestriction fragments: a multi-center study. *J. Clin. Microbiol.* 36, 1653–1659.

van Belkum, A. (2000). Molecular epidemiology of MRSA strains: state of affairs and tomorrows possibilities. *Microb. Drug Resist.* 6, 173–188.

van Belkum, A., Niesters, H.G., MacKay, W.G., and van Leeuwen, W. B. (2007a). Quality control of direct molecular diagnostics for methicillin-resistant Staphylococcus aureus. *J. Clin. Microbiol.* 45, 2698–2700.

van Belkum, A., Emonts, M., Wertheim, H., de Jongh, C., Nouwen, J., Bartels, H., Cole, A., Hermans, P., Boelens, H., Toom, N.L., Snijders, S., Verbrugh, H., and van Leeuwen, W. (2007b). The role of human innate immune factors in nasal colonization by Staphylococcus aureus. *Microbes Infect.* 9, 1417–1471.

van den Akker, E.L., Nouwen, J.L., Melles, D.C., van Rossum, E.F., Koper, J.W., Uitterlinden, A.G., Hofman, A., Verbrugh, H.A., Pols, H.A., Lamberts, S.W., and van Belkum, A. (2006). Staphylococcus aureus nasal carriage is associated with glucocorticoid receptor gene polymorphisms. *J. Infect. Dis.* 194, 814–818.

Van den Hoogen, B.G., De Jong, J.C., Groen, J., Kuiken, T., De Groot, R., Fouchier, R.A., and Osterhaus, A.D. (2001). A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat. Med.* 7, 719–724.

van der Schee, C.J., van Belkum, A., Zwijgers, L., Van der Brugge, E., O'Neill, E., Luijendijk, A., Van Rijsoort Vos, T., Van der Meijden, W. I., and Verbrugh, H.A. (1999). Improved diagnosis of *Trichomonas vaginalis* infection by PCR using vaginal swabs and urine specimens compared to diagnosis by wet mount microscopy, culture and fluorescent staining. *J. Clin. Microbiol.* 37, 4127–4130.

Van Doornum, G.J., Guldemeester, J., Osterhaus, A.D., and Niesters, H. G.M. (2003). Diagnosing herpesvirus infections by real time amplification and culture. *J. Clin. Microbiol.* 41, 576–580.

Vaneechoutte, M., and Van Eldere, J. (1997). The possibilities and limitations of nucleic acid amplification technology in diagnostic microbiology. *J. Med. Microbiol.* 46, 188–194.

Van Elden, L.J., Nijhuis, J., Schipper, P., Schuurman, R., and Van Loon, A.M. (2001). Simultaneous detection of influenza viruses A and B using real-time quantitative PCR. *J. Clin. Microbiol.* 39, 196–200.

Van Leeuwen, W.B., Van Pelt, C., Luijendijk, A., Verbrugh, H.A., and Goessens, W.H.F. (1999). Rapid detection of methicillin resistance in *Staphylococcus aureus* isolates by the MRSA screen latex agglutination test. *J. Clin. Microbiol.* 37, 3029–3030.

Van Leeuwen, W.B., Kreft, D.E., and Verbrugh, H.A. (2002). Validation of rapid screening tests for the identification of methicillin resistance in staphylococci. *Microb. Drug Resist.* 8, 55–59.

Van Leeuwen, W., Jay, C., Snijders, S., Durin, N., Lacroix, B., Verbrugh, H.A., Enright, M.C., Troesch, A., and van Belkum, A. (2003). Multilocus sequence typing of *Staphylococcus aureus* with DNA array technology. *J. Clin. Microbiol.* 41, 3323–3326.

Vannuffel, P., Laterre, P.F., Bouyer, M., Gigi, J., Vandercam, B., Reynaert, M., and Gala, J.L. (1998). Rapid and specific molecular identification of methicillin-resistant *Staphylococcus aureus* in endotracheal aspirates from mechanically ventilated patients. *J. Clin. Microbiol.* 36, 2366–2368.

Van Vliet, K.E., Muir, P., Echevarria, J.M., Klapper, P.E., Cleator, G.M., and Van Loon, A.M. (2001). Multicenter proficiency testing of nucleic acid amplification methods for the detection of enteroviruses. *J. Clin. Microbiol.* 39, 3390–3392.

Verghese, S., Padmaja, P., Sudha, P., Vanitha, V., and Mathew, T. (1999). Nasal carriage of MRSA in a cardiovascular tertiary care centre and its detection by lipovitellin salt mannitol agar. *Indian J. Pathol. Microbiol.* 42, 441–446.

Verkooyen, R., Noordhoek, G.T., Klapper, P.E., Reid, J., Schirm, J., Clears, A., Ieven, M., and Hoddevik, G. (2003). Reliability of nucleic amplification methods for detection of *Chlamydia trachomatis* in urine: results of the first international collaborative quality control study among 96 laboratories. *J. Clin. Microbiol.* 41, 3013–3016.

Wallace, P.S. (2003). Linkage between the journal and Quality Control Molecular Diagnostics (QCMD). *J. Clin. Virol.* 27, 211–212.

Weiss, D.L., and Nitzkin, P. (1971). Relationship between lysogeny and lysis in phage *S. aureus* systems. *Am. J. Clin. Pathol.* 56, 593–596.

Weller, T.M. (2000). Methicillin resistant *Staphylococcus aureus* typing methods should be the international standard. *J. Hosp. Infect.* 44, 160–172.

Wertheim, H.F., Walsh, E., Choudhurry, R., Melles, D.C., Boelens, H.A., Miajlovic, H., Verbrugh, H.A., Foster, T., and van Belkum, A. (2008). Key role for clumping factor B in Staphylococcus aureus nasal colonization of humans. *PLoS Med.* 5, e17.

Weusten, J.J., Carpay, W.M., Oosterlaken, T.A., Van Zuijlen, M.C., and Van der Paar, P.A. (2002). Principles of quantitation of viral loads using nucleic acid-sequence based amplification in combination with homogeneous detection using molecular beacons. *Nucleic Acids Res.* 30, e26.

Wolk, D., Mitchell, S., and Patel, R. (2001). Principles of molecular microbiology testing methods. *Infect. Dis. Clin. North Am.* 15, 1157–1204.

Zaaijer, H.L., Cuypers, H.T., Reesink, H.W., Winkel, I.N., Gerken, G., and Lelie, P.N. (1993). Reliability of polymerase chain reaction for detection of hepatitis C virus. *Lancet* 341, 722–724.

Zambardi, G., Fleurette, J., Schito, G.C., Auckenthaler, R., Bergogne-Berezin, E., Hone, R., King, A., Lenz, W., Lohner, C., Makrishtatis, A., Marco, F., Muller-Serieys, C., Nonhoff, C., Phillips, I., Rohner, P., Rotter, M., Schaal, K.P., Struelens, M., and Viebahn, A. (1996). European multicentre evaluation of a commercial system for identification of MRSA. *Eur. J. Clin. Microbiol. Infect. Dis.* 15, 747–749.

Chapter 31

Genetic Monitoring of Laboratory Rodents

Jean-Louis Guénet[1] and Fernando J. Benavides[2]

[1] *Département de Biologie du Développement – Institut Pasteur – Paris – France;*

[2] *The University of Texas M.D. Anderson Cancer Center, Science Park-Research Division, 1808 Park Road 1C, PO Box 389, Smithville, Texas 78957, USA*

31.1 INTRODUCTION

The use of laboratory rodents in biological research raises a number of issues. Some of these are related to the ethical aspects of experimenting with live animals, and others are related to the cost of the experiments or the type of facilities required to perform these experiments. A way to alleviate these problems is to design experiments in a way such that they each yield the largest possible amount of novel and reliable information, thus reducing to a minimum assays leading to inconclusive results (Festing and Altman, 2002). To reach this goal, experiments should be performed with carefully designed protocols and genetically defined animals. In this chapter, different aspects related to the genetic quality of laboratory rodent strains and its control will be discussed. This is a very important aspect of experimental medicine that should be considered as a frontline priority.

31.2 THE GENETIC STRUCTURE OF LABORATORY STRAINS AND STOCKS OF RODENTS

31.2.1 The Different Kinds of Laboratory Strains

Strains of laboratory rodents are of two kinds. One kind consists of genetically uniform populations, each of these populations representing only a sample among the many allelic forms that exist in the species. The other kind consists of genetically heterogeneous populations, segregating for a variety of alleles. *Inbred strains* belong to the first kind. They are artificial populations, analogous (but not identical) to a clone of mammals with, however, the two sexes. They are very different from human populations but they have the enormous advantage of being extremely uniform from the genetic point of view and they are supposed to remain so, generation after generation, provided they are bred with an appropriate protocol (Festing, 1979, 1996; Hedrich, 1990). According to Grüneberg, the introduction of inbred strains into biology was "comparable in importance with that of the analytical balance in chemistry" (Morse, 1978a). Inbred strains have derivatives such as *recombinant inbred strains* (RIS; Taylor, 1996), *consomic strains* (CS; Nadeau *et al.*, 2000), *congenic strains* and *recombinant congenic strains* (RCS; Groot *et al.*, 1996).

Opposed to inbred strains, random-bred and outbred stocks are genetically heterogeneous, and in this respect, they are more similar to human populations. *Random-bred stocks*, as their name indicates, are bred with no specific protocol: in other words, progenitors at generation G are mated randomly to produce generation G + 1. The genetic constitution of these stocks is unknown and may change with time. *Outbred stocks*, on the contrary, are bred with a specific protocol and, although they segregate for a variety of alleles at a number of loci, the frequency of these alleles in the population fluctuates within limits that are determined by the breeding system and monitored (Silver, 1995; Hartl, 2001; Chia *et al.*, 2005).

The decision to use an inbred strain rather than an outbred stock in an experimental protocol depends on the experiment and, in particular, on the biological question that is addressed. In fact, both populations are interesting and the reasons for choosing one or the other will become obvious in the following sections, once their genetic constitution and main characteristics are explained.

31.2.2 What Exactly is an Inbred Strain?

For the International Committee on Genetic Nomenclature a strain can be regarded as inbred if it has been propagated by mating systematically brothers to sisters for 20 or more consecutive generations, and individuals of the strain can be traced to a single ancestral pair at the twentieth or subsequent generation. At this point the genome of each animal within the strain, on the average, will have no more than 1 to 2% residual heterozygosity (excluding any genetic drift) and members of the same inbred strain can then be regarded, for most purposes, as genetically identical (Davisson, 1996). In practice, most of the mouse and rat inbred strains that are commonly used in research have undergone several tens of generations of inbreeding and some have even been bred with this system since the beginning of the last century, which means for over 200 generations. This definition of an inbred strain is important and calls for a few comments.

As it has already been mentioned, mice and rats of the same inbred strain are genetically identical. They are also homozygous at all loci of their genome and, in particular, at all loci that were polymorphic in the founder ancestors. After a few tens of generations, one of the alleles that were segregating at a given locus becomes fixed and the others (i.e. up to three) are lost. This process of allele loss (or fixation) is easy to understand considering that, when by chance, an allele that was present at generation F is not represented in at least one of the two mice (one brother and its sister) that are mated to produce generation F + 1, it is then permanently lost. In other words, alleles are lost during the inbreeding process but, except in the rare case of *de novo* mutations, none are ever introduced from an external source. The sorting of the alleles that are retained in the strain and those that are lost depends (in most cases) on chance and, if the experiment could be remade from scratch, with the same founder mice, it would lead to a strain with a different genetic constitution after the same 20 generations. This means that an inbred strain represents a unique, although fortuitous, assortment of alleles. To give an idea of the genetic profile of an inbred strain one could imagine a totally artificial scenario where an X-bearing spermatozoa, taken from wild mice, fertilizes an oocyte. Then the female pronucleus is removed from the oocyte in question (before it fuses with the male pronucleus) and finally the chromosomes that were brought into the oocyte by the spermatozoa are duplicated. This new ($n + n = 2n$) conceptus would be a female with the two chromosomes of each pair absolutely identical. This is precisely how the genome of all members of an inbred strain looks, with the exception, of course, of the sex chromosomes.

During the process of inbreeding, the progression toward homozygosity is not constant. It is relatively fast during the first few generations, then it slows down and, after 20 generations of unrelaxed inbreeding, it is no more than about 1 to 2% of the loci that are still segregating. A mathematical series based on Fibonacci's numbers is traditionally used to model the progress toward homozygosity while inbreeding progresses, but it is only an approximation (Fig. 31.1).

Since complete homozygosity is virtually reached, at all loci, after a few tens of brother-to-sister matings, it may then come to mind that it is no longer necessary to use such a stringent breeding system to propagate the strain. In fact, this would be hazardous because mutations constantly occur and the interruption of inbreeding would allow an increasing load of new polymorphisms to accumulate. Even if the spontaneous mutation rate is very low, there are so many genes in a mammalian genome that this source of polymorphism is not negligible. Accordingly, inbreeding must not be relaxed.

All animals of an inbred strain are isogenic. This means that they all are genetically identical. This is an enormous advantage because scientists working with the same inbred strain, but in different laboratories or at different periods in time, can perform experiments where fluctuations in the experimental results, by definition, will not be the consequence of possible differences in the genetic constitution. Being isogenic also means that one can define in detail the phenotypic characteristics of an inbred strain by accumulating experimental data concerning this strain from several sources. For example, The Jackson Laboratory has developed a program to establish a collection of baseline phenotypic data on the 48 most popular inbred strains of mouse, through a coordinated international effort. Information collected in this program (The Mouse Phenome Database)

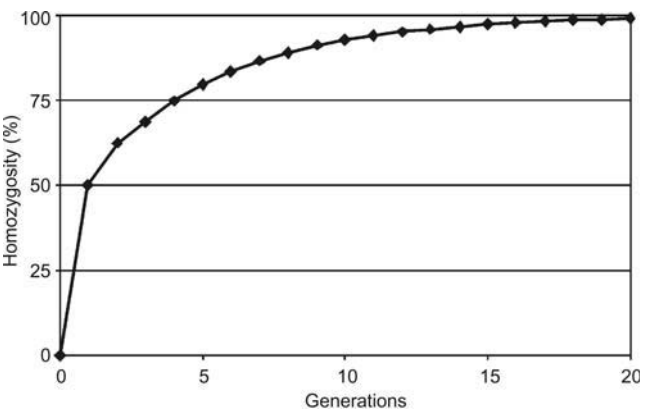


FIGURE 31.1 When a population of laboratory animals is propagated exclusively by mating brothers to sisters, the percentage of loci that are heterozygous decreases regularly. This progression towards homozygosity is modeled quite faithfully by the successive terms of a mathematical series where the numerators are Fibonacci's numbers (1, 1, 2, 3, 5, 8, 13, etc., the term ranked N being the sum of the terms ranked N-l plus N-2) and denominators the successive terms of the series 2^N. From the fifth generation (F5) onward, the level of heterozygosity falls off roughly by 19% at each generation so that, at F + 40, 99.98% of the genome is expected to be homozygous. In practice, considering that most of the loci were already homozygous in the first progenitors (as they are in most mammals, irrespective of the species) and keeping in mind that loci in a mammalian genome are not assorting independently but on the contrary are inherited as chromosomal blocks of various size, one can consider that a strain is totally inbred after 60 generations of inbreeding, not taking into account the continual appearance of spontaneous mutations.

(Paigen and Eppig, 2000) is freely available to the research community through the internet: http://phenome.jax.org/pub-cgi/phenome/mpdcgi?rtn=docs/home. The development of this database, which is regularly updated, is possible only because mice of inbred strains are isogenic. Around 400 different mice and 200 rat inbred strains are available worldwide and a dozen among these strains, in both species, have become extremely popular.

To finish with the unique characteristics of inbred strains, one could mention that they are fundamental tools for all experiments where a high level of standardization is required. They can be used as such, and they can be used for the production of interstrain FI hybrids, which also represent genetically homogeneous populations and which, when taken by pair, are absolutely comparable to monozygotic twins. Inbred strains can also be used to generate genetically heterogeneous populations when, for example, F1 hybrids between strain A and strain B (abbreviated ABFl) are crossed with similar Fl hybrids between strain C and strain D (CDFl) to generate a four-way heterogeneous stock. In this case, although genetically heterogeneous, the basic components of such a stock, the original inbred strains, are perfectly identified, and such genetically heterozygous stocks can be produced over and over, at will, when necessary. Genetically heterogeneous stocks with an even more complex structure (for example eight-way crosses stemming from eight different and totally unrelated inbred strains) are now being bred on a large scale for research in quantitative genetics (The Complex Trait Consortium, 2004).

Describing the genetic characteristics of inbred strains, we said that these strains were artificial since their genetic constitution (isogenicity, homozygosity) has no natural equivalent in any species, including man. We now know, from historical records and molecular information collected at the DNA level, that these classical laboratory inbred strains of mice are also artificial because they do not derive from a single subspecies of the *Mus* genus but from at least two subspecies: *Mus musculus domesticus* and *Mus musculus musculus* (Frazer *et al.*, 2007). This means that the genome of these mice is a patchwork of chromosomal segments stemming in these two subspecies. This characteristic, which is probably unique to the laboratory mouse, is particularly interesting in the sense that it contributes to an increase in terms of genetic polymorphisms, making each strain different from the next. Studies on the genetic determinism of complex traits will benefit from this situation (Bonhomme *et al.*, 1987; Wade *et al.*, 2002).

31.2.3 The Difficulties of Maintaining Strain Purity

People in charge of a breeding colony should ask themselves these simple questions: Are the animals really from an inbred strain? Has this strain been previously used for similar studies? Is this strain used for other investigators in the same type of experiments? The real challenge for these professionals is to keep inbred strains genetically pure. Three factors can lead to changes in the genetic constitution of inbred strains: the *genetic drift* due to residual heterozygosity, the occurrence of *mutations*, and, the most important of all, the contamination by accidental outcross with another strain.

Genetic drift due to residual heterozygosity is more a theoretical than an actual threat because most of the strains that are used nowadays stem from a rather small number of primitive strains that were already highly inbred when they began to be distributed. However, we have examples of this type of genetic drift: most of the existing differences among the major substrains of the C3H family (C3H/An from Andervont, C3H/He from Heston, C3H/Bi from Bittner, and C3H/Fg from Fuchs, etc.) are, for the most part, due to incomplete inbreeding before distribution by L.C. Strong of the original progenitors (Morse, 1978b). Accordingly, if genetic drift can now be disregarded as a source of genetic divergence this was not the case at the beginning of last century and must be kept in mind, for a rigorous designation of the various substrains. Neglecting this may lead to serious difficulties, or even inconsistencies, in the interpretation of the experimental results.

Mutations are important to consider for two reasons: first, because their occurrence is beyond the control of the colony manager, and second, because they are very insidious and frequently impossible to detect by simple, superficial phenotypic observation. Based on extensive statistics, the mutation rates have been estimated to be in the range of 10^{-7} to 0.5×10^{-6} per locus per gamete for mutations towards a dominant allele in the range of $0.6-0.8 \times 10^{-6}$ per locus per gamete for mutations towards a recessive allele (Schlager and Dickie, 1967). Because these mutation rates are not negligible, the International Committee on Genetic Nomenclature has decided that two strains with the same origin but separated in different colonies by 100 or more generations (for example, 47 in laboratory A and 53 in laboratory B) should be considered as two different substrains and designated differently, in compliance with the rules for nomenclature (Davisson, 1996). Examples are common in the literature where two substrains of the same original inbred strain behave differently because a mutation occurred in one of the two substrains and altered the standard phenotype of the strain (Sultzer, 1968; Moisset, 1978; Morse, 1978a; Bulfield *et al.*, 1984).

Data recently collected, concerning single nucleotide polymorphisms (SNPs) in different C57BL/6 substrains, kept independently for a few years at The Jackson Laboratory indicates that the mutation rate generating SNPs is very low (Wade, 2002). In addition, assuming that only one SNP out of ten is translated into a functional polymorphism (Beier, 2000), this would suggest that the occurrence

of new mutations is not a serious issue in the generation of subline divergence. The main problem, however, is that the consequences of a new mutation are totally unpredictable. Mice of the C57BL/6JOlaHsd substrain, for example, are homozygous for a deletion of the *Snca* locus (encoding for α-synuclein) on chromosome 6 (Specht and Schoepfer, 2001). This deletion has modest phenotypic effects (Chen *et al.*, 2002) but might interfere in an unpredictable manner with other mutations if, for example, the C57BL/6JOlaHsd substrain is used as a background strain for making knockout congenics (Wotjak *et al.*, 2003). Similarly, if mice of substrain C3H/HeJ are to be experimentally infected with Gram-negative bacteria they may react very differently from mice of substrain C3H/OuJ. This is explained by the occurrence of a mutation at the *Tlr4* locus (encoding for a *Toll*-like receptor) that occurred in the substrain C3H/HeJ where all mice are homozygous for the defective allele (Poltorak *et al.*, 1998). A very similar comment could be made for mice of the CBA/N substrain which, unlike mice of all other CBA substrains, are homozygous for an X-linked mutation (Btk^{xid}) producing a syndrome of immunodeficiency homologous to the Bruton's disease in man (Rawlings *et al.*, 1993).

The accidental matings of individuals from one inbred strain with another strain is by far the most important source of alteration of the genetic profile of inbred strains. Genetic contaminations of this type, which always result in a sudden and massive exchange of alleles, are generally observed between strains that have similar coat color (i.e. albino (Tyr^c/Tyr^c), agouti (*A/A*), or non-agouti (*a/a*)). Even more frequently, these accidental crosses occur between an interstrain Fl and one of the parental strains. Mouse strains A2G and C57BL/Ks are two well-known examples where massive genetic contamination occurred in the past. A2G was considered a substrain of strain A until it was discovered that they became contaminated after an "illegitimate" mating with an unknown partner (probably a wild mouse?). C57BL/Ks (now C57BLKS) derives from strain C57BL/6 but was contaminated with up to 30% of the DBA/2 genome. This was suspected because C57BL/Ks mice have an *H2* haplotype which is not like the haplotype normally found in C57BL/6 mice (they are $H2^d$ instead of $H2^b$) and also because congenic mice for the same obese (Lep^{ob}) mutation on these two backgrounds (C57BL/6J and C57BL/Ks) exhibited a totally different phenotype (Herberg and Coleman, 1977). The suspicion of genetic contamination has now been molecularly documented.

All three sources of changes in the genetic constitution of inbred strains must be taken into account by animal breeders. If the genetic drift due to residual heterozygosity is no longer a serious problem and the occurrence of mutation is insidious and unavoidable, the genetic contamination after an accidental cross with another strain is, by far, the most serious threat.

31.3 MONITORING THE GENETIC QUALITY OF INBRED STRAINS

A variety of techniques have been described in the past to assay the genetic quality of inbred strains. All these techniques were based on the postulates that each inbred strain, as previously mentioned, is *a priori* expected to be homozygous at all loci of its genome and that all animals of the same strain have exactly the same genetic make-up. These techniques, summarized in Table 31.1, were designed following the progress of the genetic tools available for the species and consisted in the analysis of a few traits, controlled by a set of specific alleles, and defining a specific pattern for each strain.

Reciprocal skin grafting, for example, was extensively used in the 1960s because histocompatibility (the complete and permanent acceptance of transplants between any two mice or rats) is controlled by many genes and requires complete genetic identity between the donor and the recipient (Bailey and Usama, 1960). Skin grafting was a relatively inexpensive procedure and fast (in skilled hands!) but, unfortunately, it was often influenced in both directions by environmental factors yielding false positive and false negative results. It is no longer used for the control of strain purity although some refinements in the procedure, using some modern *in vitro* assays, might still be interesting as supplementary tests. Analysis of the electric charge of enzymatic proteins (isozymes) by electrophoresis in gels became popular in the mid-1970s because the technique was highly reliable and relatively easy to handle (Groen, 1977). However, this technique had the major drawback of being expensive to apply because each test required the use of specific and costly reagents.

TABLE 31.1 Summary of the most representative techniques to assay the genetic quality of inbred strains.

Phenotype-based
External characteristics
Reproductive performances
Skin grafting
Protein analysis by gel electrophoresis
DNA-based
RFLP analysis
PCR amplification of microsatellite markers (SSLP)
SSCP analysis of 100/250 bp long DNA stretches
DNA sequencing (detection of SNPs)

Currently, most of the genetic monitoring techniques applied to inbred strains are based on DNA analysis and are extremely powerful. However, it must be kept in mind that the control of genetic purity should be undertaken in a broader context, considering also several simple parameters of very different nature, and not only by applying very sophisticated molecular techniques. Among these parameters, a careful observation of individuals from the same inbred strain, even if it may appear rather subjective, is always a very important source of information.

31.3.1 Observation of Characteristic Phenotypes

Being genetically identical, all members of an inbred strain must have exactly the same phenotype and, in pigmented strains, the observation of different coat colors must be considered highly suspicious. A white spot on the belly of C3H mice is, however, common and should even be regarded as a criterion of purity as well as occasional tail kinks or eye abnormalities in C57BL/6. With experience, a sagacious observer should even be able to recognize subtle characteristics discriminating the five major albino strains of mice (A, AKR, BALB/c, FVB, and SJL) or rats only considering the hair texture and tail shape. Similarly, several traits such as aggressiveness, open field behavior, life span, spontaneous diseases (tumors in particular), as well as a whole range of biochemical, immunological, and physiological characteristics are very much strain specific and are important to consider as criteria of genetic purity. For example, the observation of testicular teratomas in some substrains of 129 mice (129S2/SvPas in particular) is a criterion of great value because that sort of tumor is under polygenic control. In short, no other strains than those of the 129 group develop testicular teratocarcinomas.

Another change that is almost always the direct consequence of a genetic contamination is a dramatic increase in the breeding performance. With a few exceptions, inbred strains are not very prolific and the age of puberty, age of first pregnancy, and the average number of pups weaned per week or per female are highly heritable characteristics. An abrupt change in these performances must be considered suspicious.

31.3.2 Observation of Structural Variations at the DNA Level

Apart from these zootechnical parameters, which are easy to monitor by the mere analysis of the breeding records and a careful analysis of the phenotypes of the animals, other tests are available for the monitoring of genetic purity. Most of these tests are based on the analysis of strain-specific DNA sequences, revealed by the routine techniques of DNA structural analysis such as endonuclease-generated restriction fragment length polymorphisms (RFLP), minisatellite-based DNA fingerprinting (Benavides *et al.*, 1998), polymerase chain reaction (PCR) amplification of microsatellites (also known as simple sequence length polymorphism, SSLP), single strand conformation polymorphism (SSCP, see also Chapter 4) or DNA sequencing to identify various SNPs.

31.3.2.1 Genetic Monitoring with Microsatellite Markers

Microsatellite markers are very popular because they are extremely easy to type at a very low cost (Montagutelli and Guénet, 1994; Benavides, 1999; Mashimo *et al.*, 2006). The technique consists in the amplification of short repeated sequences, in general dinucleotides of the type $(CA)_n$ or $(TA)_n$, with flanking primers using genomic DNA. There are an enormous number of microsatellite loci in the mouse and rat genomes (probably around 10^5) and it is generally not a problem to find a set of such molecular markers whose amplification products define a strain-specific pattern (Fig. 31.2). This strain-specific pattern may be assayed on a sample of animals of the strain and compared to a reference pattern that is archived in the laboratory. Routine analysis of DNA samples with microsatellite markers will confirm isogenicity and, provided the markers have been carefully selected, it could also guarantee that the strain whose DNA is assayed indeed corresponds to its designation. The use of fluorescently labeled primers for microsatellite loci combined with capillary electrophoresis represent a new, fast, automated system for genetic monitoring (Bothe *et al.*, 2004; Mashimo *et al.*, 2006) (Fig. 31.3). With this method, the resulting PCR products can be distinguished from one another by both their size and the fluorescent dye associated with them. The availability of different dyes allows the possibility of developing multiplex PCR (i.e. the combination of primers for multiple loci in one reaction) and pooling several PCR products in one capillary.

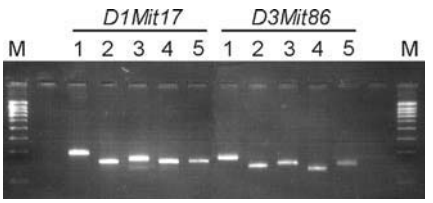


FIGURE 31.2 Segment of an agarose gel showing the characteristic bands obtained after PCR amplification of SSLP markers. The amplification products were obtained using genomic DNA from five mouse inbred strains: SEG/Pas (1), C57BL/6Pas (2), C3H/HePas (3), DBA/2Pas (4), and BALB/cPas (5). The first five bands represent microsatellite marker *D1Mit17*, and the other five, markers *D3Mit86*. Note that each inbred strain exhibits only one band (one allele) per SSLP locus (indicative of their homozygous state), and that the bands have a standard size for each strain. Using a set of carefully selected microsatellite markers it is possible to define a pattern characteristic of a given strain. M: DNA size marker.

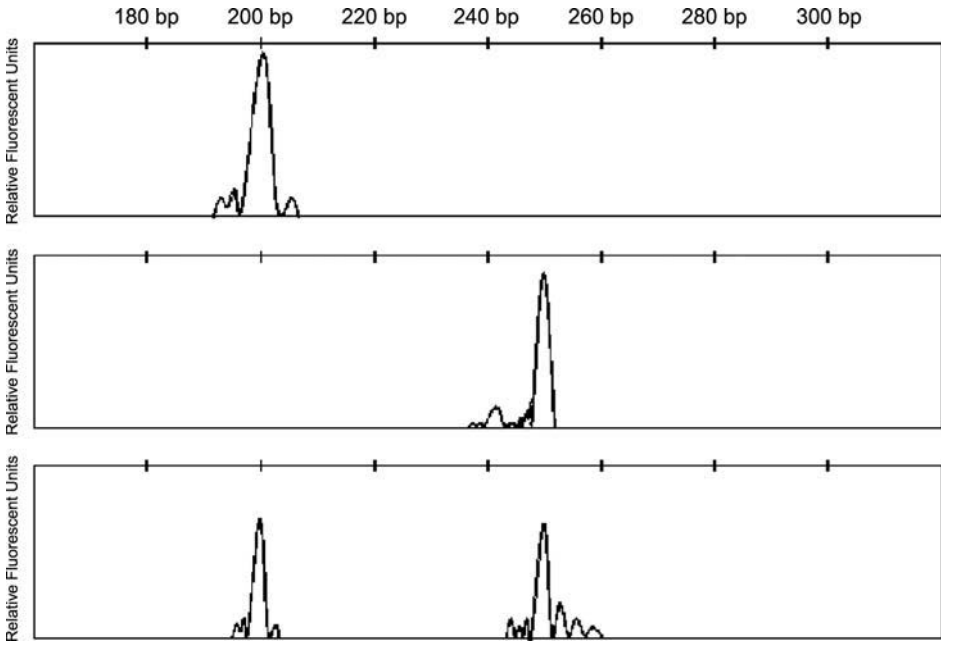


FIGURE 31.3 Representative data output for analysis of mouse DNA with a microsatellite marker using fluorescent primers (6-FAM). This marker was amplified in a single PCR reaction and separated by capillary electrophoresis (see also Chapter 5). The upper panel shows a mouse homozygous for a 200 bp allele (represented by the pick of fluorescence). The middle panel shows a mouse homozygous for a 250 bp allele. The lower panel shows a mouse heterozygous for the alleles depicted above. The sizes in base pares (bp) are indicated on the top of the *x*-axis (with smaller alleles to the left and larger alleles to the right). The values on the *y*-axis indicate fluorescent signal intensity (relative fluorescent units).

Using this system, Bryda and Riley (2008) have developed a genotyping panel containing 87 microsatellite markers that are polymorphic among commonly used inbred rat strains.

Microsatellite markers have been a true revolution in the genetic monitoring of laboratory inbred strains because the test is extremely simple, affordable, and highly reliable. It can then be repeated more frequently and on a large sample of animals. A drawback of this test, maybe the only one, is a consequence of the occasional occurrence of variations in size of the microsatellite sequences. Typing DNA samples from the BXD set of recombinant inbred strains, between the parental strains C57BL/6 and DBA/2, Dallas and coworkers (1992) found that several amplification products had a size different from the parental types and were then considered as "mutants". This occurred in the frequency of 10^{-2} to 10^{-4}, a frequency that is not so low. When such a mutation is discovered, it is then recommended to type a dozen more microsatellites, some of them flanking the mutant allele. In general, this is sufficient to clarify the situation. It would also be advisable to type all the breeders in the foundation colony and to eliminate the new mutant allele (or to select it in the homozygous state) but not to keep it segregating in the inbred strain.

31.3.2.2 Genetic Monitoring Based on the Use of Single Nucleotide Polymorphisms

Even though microsatellites are excellent markers for genetic monitoring of inbred strains, they have limitations in the sense that their mutation frequency is rather high and that they do not allow for the easy identification of the origin of a genetic contamination when such an accident is detected in a given strain. Sequencing a few short DNA stretches looking for SNPs is an alternative approach that is now popular. SNP genotyping is inexpensive and can be performed in most research institutions (see Fig. 31.4). Petkov and coworkers (2004), from The Jackson Laboratory (Maine, USA), have described the allelic distribution of 235 SNPs in 48 mouse strains and selected a panel of 28 such SNPs, enough to characterize most of the almost 300 inbred, wild-derived, congenic, consomic, and recombinant inbred strains maintained at the Jackson Laboratory. This set of markers encompassing all mouse chromosomes, is an excellent tool for detecting genetic contaminations in mouse facilities by way of automated PCR systems. SNP genotyping assays are currently based on allele-specific PCR, real-time PCR (TaqMan®), direct sequencing, or DNA arrays (Moran *et al.*, 2006; Shifman *et al.*, 2006).

Finally, it is worth recalling that, regardless the technique used, the occurrence of a new mutation in a coding sequence cannot be detected, except by chance, or if it results in an obvious phenotypic change. The only possibility for detecting such changes would be to re-sequence the genome of the strain, something that would be prohibitively expensive and probably not worth doing in the end. Techniques are now being developed for the detection of DNA mismatches, which may become useful for the detection of new mutations, but their cost is still high (Yamana *et al.*, 2005).

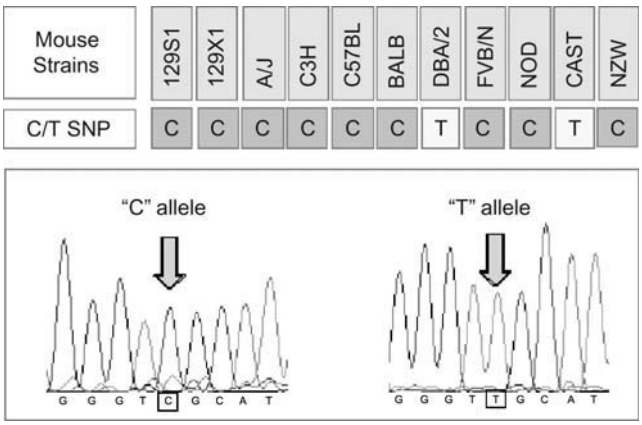


FIGURE 31.4 SNPs are discrete DNA sequence variations occurring when a single nucleotide in the genome differs between members of the same species. These SNPs are common and they are scattered throughout the genome of all species. They result from random point mutations occurring at a constant rate during evolution, either in the coding regions or in between genes and they are inherited like a Mendelian trait. In the mouse genome, they are very unevenly distributed along the chromosomes with "SNP-rich" and "SNP-poor" regions depending on the phylogenetic origin of the chromosomal segment. This allows the determination of an SNP pattern, which is unique to a given strain and accordingly can be used for assessing strain purity. The upper panel represents a C/T SNP that is polymorphic between strains DBA/2 and CAST (homozygous for the "T" allele), and other common inbred strains (homozygous for the "C" allele). The lower panel presents DNA sequencing electropherograms showing the SNP (arrow).

31.4 PRESERVING THE GENETIC PURITY OF INBRED STRAINS

As it has been already discussed in the previous paragraphs, there are very efficient techniques to monitor the genetic quality of inbred strains. However, once a strain is recognized as contaminated, the situation is irreversible. All individuals of the strain must be discarded, and another colony must be developed from a new set of breeders. The dramatic consequences of a genetic contamination imply that steps should be taken to prevent it from occurring again. Roughly, one can imagine that there are two efficient ways to preserve an inbred nucleus from genetic contamination: embryo freezing and complete physical isolation.

Embryo freezing is, theoretically, the most efficient way of preservation because, once frozen, genomes are insensitive to mutations (DNA does not replicate) and of course contamination cannot occur. Experiments performed several years ago indicated that cosmic radiation does not represent a serious risk for deep-frozen embryos (Whittingham *et al.*, 1977). This means that choosing this procedure for genome preservation is very secure and relatively inexpensive. However, the technique has two serious drawbacks. First, it can be achieved safely only in laboratories where the technical staff is trained for embryo handling and, second, some strains are very difficult to preserve and the percentage of living embryos recovered after thawing is extremely low or null. This drawback, which is inherent to particular strains, cannot be bypassed. Sperm freezing may, in some circumstances, be used as an alternative to embryo freezing, but is of no use when a diploid genome must be preserved (Glenister and Thornton, 2000).

Complete isolation of a breeding nucleus, for example, into a plastic isolator is a very efficient way to preserve genetic integrity and it is also an elegant way to preserve, at the same time, the health status of a rodent colony. Most of the commercial breeders of laboratory rats and mice have chosen this strategy, which combines several advantages at a relatively low cost. Finally, and as an anecdotal comment, it is interesting to note that any recessive coat color mutation occurring in an inbred nucleus, at least theoretically, can be used as a natural seal of genetic purity since any illegitimate mating will result in the immediate disappearance of the trait. At several occasions, new alleles of the mutations leaden ($MLph^{ln}$-chromosome 1), brown ($Tyrp^{b}$-chromosome 4), dilute ($Myo5a^{d}$-chromosome 9), or pink-eyed dilution ($Oca2^{p}$-chromosome 7) were found segregating in the author's colonies of C3H or C57BL/6, which could have been used to tag those strains with an institutional marker. However, it seems that scientists, when offered these naturally tagged mice are always reluctant to use them; a C57BL/6 inbred mouse must be solid black and any other color is suspicious!

31.5 ASSESSING THE GENETIC STANDARD OF OUTBRED STOCKS

Outbred stocks are genetically heterogeneous and segregate for several alleles (in general two, sometimes three) at a few loci. They are very useful tools in many areas of experimental biology because the response of this type of stock depends upon a large pool of genes and not only upon a few randomly selected alleles as in the case of inbred strains. In this respect, they appear complementary to the inbred strains but obviously they are neither homozygous nor isogenic and accordingly cannot be genetically monitored with the same protocols. What is important in this case is to make sure that the percentage of the different alleles in the population does not change much from one generation to the next and the strategies used for the monitoring of these strains are based on sampling and statistics. Here again, microsatellite markers or SNPs can give an indication and hence this is a safe strategy to test a large sample of presumptive breeders, retaining for the next generation a pool of mice or rats in which the frequency of each allelic variant is not statistically different from the frequency observed in generation G-1. Of course, the larger the sample the better it is to type an even greater number of microsatellites. Finally, it must also be kept in mind that, only in the case of inbred strains, one brother and one sister

are required to breed generation F + 1, while in the case of outbred stocks the larger the pool of breeders the better. Breeding programs have been developed to minimize the fluctuation of allele frequency at each generation (Hartl, 2001) but it is safe, from time to time, to perform a very large survey of the structure of the population and to keep the results as a reference.

Four- and eight-way genetically heterogeneous stocks, which are produced by intercrossing Fl hybrids bred from four or eight independent inbred strains, are an interesting alternative to the outbred stocks. In addition, they have the enormous advantage that genetic monitoring is not required. Like the best champagne wines, which have a constant taste because they are a blend of several vintages, genetically heterogeneous stocks are produced by crossing a small number of highly standardized inbred strains.

31.6 THE CONTROL OF HEALTH STATUS

Although the control of the health status is out of the scope of this chapter, it is worth making a brief comment about it because the health status is a major source of variation in the biological response of laboratory animals and accordingly it has become a serious concern (Whary and Fox, 2006). In fact, although very efficient strategies exist for the control of the genetic quality of laboratory animals, it is very difficult (and also very expensive) to perform a thorough control of their microbiological status and even more difficult to keep it constant. Lists of micro-organisms that are pathogenic for rats and mice have been previously published and a battery of tests exists for the diagnosis of these pathogens that are reliable provided they are applied routinely and carefully performed. Some of these tests are PCR-based and allow the detection of some specific pathogens (*Helicobacter* spp., mouse hepatitis virus (MHV), mouse parvovirus (MPV), etc.; see also Chapter 30). One can then guarantee with a high level of confidence that a particular rodent colony is free of specific pathogens (specific pathogen free or SPF). Animals of this type have been available for several decades from most laboratory animal vendors. Since these mice are also evaluated at regular intervals for their genetic quality one may consider that the quality of laboratory animals is now totally under control. However, scientists have reported that variations in the biological response of some transgenic mice could be seriously influenced by differences in the population of enteric bacteria even if these bacteria are not registered as pathogenic for the mouse (Sellon *et al.*, 1998). In other words, it looks as if "clean" mice may not be the best material for some investigations in immunology. These observations indicate that, in the future, a standardized microflora for laboratory rodents may have to be defined and with it the procedures to monitor it.

31.7 CONCLUSIONS

In this chapter, the problems associated with the maintenance of standardized strains of laboratory rodents and the need to control their genetic quality and health status regularly have been reviewed. These controls, provided they are carefully and regularly performed, will contribute to the optimization of the experimental protocols, which in turn result in reducing the cost of the experiments by minimizing the number of animals needed. It is also of cardinal importance to use and follow with care, in all published works, the rules of nomenclature for inbred strains in order to avoid possible inconsistencies that may result from subtle genetic differences between the many substrains available.

REFERENCES

Bailey, D.W., and Usama, B. (1960). A rapid method of grafting skin on tails of mice. *Plast. Reconstr. Transplant. Bull.* 25, 424–425.

Beier, D.R. (2000). Sequence-based analysis of mutagenized mice. *Mamm. Genome* 11, 594–597.

Benavides, F., Cazalla, D., and Pereira, C. (1998). Evidence of genetic heterogeneity in a BALB/c mouse colony as determined by DNA fingerprinting. *Lab. Anim.* 32, 80–85.

Benavides, F. (1999). Genetic contamination of an SJL/J mouse colony: rapid detection by PCR-based micro satellite analysis. *Contemp. Top. Lab. Anim. Sci.* 38, 54–55.

Bonhomme, F., Guénet, J.-L., Dod, B., Moriwaki, K., and Bulfield, G. (1987). The polyphyletic origin of laboratory inbred mice and their rate of evolution. *Biol. J. Linnean Society* 30, 51–58.

Bothe, G.W., Bolivar, V.J., Vedder, M.J., and Geistfeld, J.G. (2004). Genetic and behavioral differences among five inbred mouse strains commonly used in the production of transgenic and knockout mice. *Genes Brain Behav.* 3, 149–157.

Bryda, E.C., and Riley, L.K. (2008). Multiplex microsatellite marker panels for genetic monitoring of common rat strains. *J. Am. Assoc. Lab. Anim. Sci.* 47, 37–41.

Bulfield, G., Siller, W.G., Wight, P.A., and Moore, K.J. (1984). X chromosome-linked muscular dystrophy (mdx) in the mouse. *Proc. Natl. Acad. Sci. USA* 81, 1189–1192.

Chen, P.E., Specht, C.G., Morris, R.G., and Schoepfer, R. (2002). Spatial learning is unimpaired in mice containing a deletion of the alpha-synuclein locus. *Eur. J. Neurosci.* 16, 154–158.

Chia, R., Achilli, F., Festing, M.F., and Fisher, E.M.C. (2005). The origins and uses of mouse outbred stocks. *Nat. Genet.* 37, 1181–1186.

Dallas, J.E., Wotjak, C.T., and Holter, S.M. (1992). Estimation of micro satellite mutation rates in recombinant inbred strains of mouse. *Mamm. Genome* 3, 452–456.

Davisson, M.T. (1996). Rules for nomenclature of inbred strains. In *Genetic Variants and Strains of the Laboratory Mouse* (Lyon, M.E., Rastan, S., and Brown, S.D.M., eds), pp. 1532–1536. Oxford University Press, Oxford.

Festing, M.F. (1979). *Inbred Strains in Biomedical Research*. The MacMillan Press Ltd.

Festing, M.F. (1996). Origins and characteristics of inbred strains of mice. In *Genetic Variants and Strains of the Laboratory Mouse* (Lyon, M.E.,

Rastan, S., and Brown, S.D.M., eds), pp. 1537–1576. Oxford University Press, Oxford.

Festing, M.F., and Altman, D.G. (2002). Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR J.* 43, 244–258.

Frazer, K.A., Eskin, E., Kang, H.M., Bogue, M.A., Hinds, D.A., Beilharz, E.J., Gupta, R.V., Montgomery, J., Morenzoni, M.M., Nilsen, G.B., Pethiyagoda, C.L., Stuve, L.L., Johnson, F.M., Daly, M.J., Wade, C.M., and Cox, D.R. (2007). A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* 448, 1050–1053.

Glenister, P.H., and Thornton, C.E. (2000). Cryoconservation archiving for the future. *Mamm. Genome* 11, 565–571.

Groen, A. (1977). Identification and genetic monitoring of mouse inbred strains using biochemical polymorphisms. *Lab. Anim.* 11, 209–214.

Groot, P.C., Moen, C.J., Hart, A.A., Snoek, M., and Demant, P. (1996). Recombinant congenic strains – genetic composition. In *Genetic Variants and Strains of the Laboratory Mouse* (Lyon, M.E., Rastan, S., and Brown, S.D.M., eds), pp. 1660–1670. Oxford University Press, Oxford.

Hartl, D.L. (2001). Genetic management of outbred laboratory rodent populations. Charles River Genetic Literature, http://www.criver.com/techdocs/index.html.

Hedrich, H.J. (1990). Inbred strains in biomedical research. In *Genetic Monitoring of Inbred Strains of Rats* (Hedrich, H.J., ed.), pp. 1–7. Gustav Fischer Verlag, Stuttgart.

Herberg, L., and Coleman, D.L. (1977). Laboratory animals exhibiting obesity and diabetes syndromes. *Metabolism* 26, 59–99.

Mashimo, T., Voigt, B., Tsurumi, T., Naoi, K., Nakanishi, S., Yamasaki, K., Kuramoto, T., and Serikawa, T. (2006). A set of highly informative rat simple sequence length polymorphism (SSLP) markers and genetically defined rat strains. *BMC Genet.* 7, 19.

Moisset, B. (1978). Subline differences in behavioral responses to pharmacological agents. In *Origins of Inbred Mice* (Morse, H.C., 3rd ed.), pp. 483–484. Academic Press, New York.

Montagutelli, X., and Guénet, J.L. (1994). Genetic monitoring of inbred strains by analysis of micro satellite polymorphisms. Fifth Symposium of the Federation of European Laboratory Animal Science Associations. Royal Society of Medicine Press, London.

Moran, J.L., Bolton, A.D., Tran, P.V., Brown, A., Dwyer, N.D., Manning, D.K., Bjork, B.C., Li, C., Montgomery, K., Siepka, S.M., Vitaterna, M.H., Takahashi, J.S., Wiltshire, T., Kwiatkowski, D.J., Kucherlapati, R., and Beier, D.R. (2006). Utilization of a whole genome SNP panel for efficient genetic mapping in the mouse. *Genome Res.* 16, 436–440.

Morse, H.C., 3rd (1978a). Introduction (also available via the internet at http://www.informatics.jax.org). In *Origins of Inbred Mice* (Morse, H.C., 3rd ed.), pp. 3–22. Academic Press, New York.

Morse, H.C., 3rd (1978b). Differences among sub lines of inbred mouse strains (also available via the internet at http://www.informatics.jax.org). In *Origins of Inbred Mice* (Morse, H.C., 3rd ed.), pp. 441–444. Academic Press, New York.

Nadeau, J., Singer, J., Matin, A., and Lander, E. (2000). Analyzing complex genetic traits with chromosome substitution strains. *Nat. Genet.* 24, 221–225.

Paigen, K., and Eppig, J.T. (2000). A mouse phenome project. *Mamm. Genome* 11, 715–717.

Petkov, P.M., Cassell, M.A., Sargent, E.E., Donnelly, C.J., Robinson, P., Crew, V., Asquith, S., Vonder Haar, R., and Wiles, M.V. (2004). Development of a SNP genotyping panel for genetic monitoring of the laboratory mouse. *Genomics* 83, 902–911.

Poltorak, A., He, X., Smimova, I., Liu, M.Y., Van Huffel, C., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C., Freudenberg, M., Ricciardi-Castagnoli, P., Layton, B., and Beutler, B. (1998). Defective LPS signaling in C3H/HeJ and C57BL/lOScCr mice: mutations in Tlr4 gene. *Science* 282, 2085–2088.

Rawlings, D.J., Saffran, D.C., Tsukada, S., Largaespada, D.A., Grimaldi, J.C., Cohen, L., Mohr, R.N., Bazan, J.F., Howard, M., Copeland, N.G. (1993). Mutation of unique region of Bruton's tyrosine kinase in immunodeficient XID mice. *Science* 261, 358–361.

Schlager, G., and Dickie, M.M. (1967). Spontaneous mutations and mutation rates in the house mouse. *Genetics* 57, 319–330.

Sellon, R.K., Tonkonogy, S., Schultz, M., Dieleman, L.A., Grenther, W., Balish, E., Rennick, D.M., and Sartor, R.B. (1998). Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-IO-deficient mice. *Infect. Immun.* 66, 5224–5231.

Shifman, S., Bell, J.T., Copley, R.R., Taylor, M.S., Williams, R.W., Mott, R., and Flint, J. (2006). A high-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biol.* 4, e395.

Silver, L.M. (1995). *Mouse Genetics: Concepts and Applications* (also available via the internet at http://www.informatics.jax.org). Oxford University Press, Oxford.

Specht, C.G., and Schoepfer, R. (2001). Deletion of the alphasynuclein locus in a subpopulation of C57BL/6J inbred mice. *BMC Neurosci.* 2, 11.

Sultzer, B.M. (1968). Genetic control of leucocyte responses to endotoxin. *Nature* 219, 1253–1254.

Taylor, B.A. (1996). Recombinant inbred strains. In *Genetic Variants and Strains of the Laboratory Mouse* (Lyon, M.E., Rastan, S., and Brown, S.D.M., eds), pp. 1597–1599. Oxford University Press, Oxford.

The Complex Trait Consortium (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* 36, 1133–1137.

Wade, C.M., 3rd, Kulbokas, E.J., Kirby, A.W., Zody, M.C., Mullikin, J.C., Lander, E.S., Lindblad-Toh, K., and Daly, M.J. (2002). The mosaic structure of variation in the laboratory mouse genome. *Nature* 420, 574–578.

Whary, M.T., and Fox, J.G. (2006). Detection, eradication, and research implications of Helicobacter infections in laboratory rodents. *Lab. Anim.* 35, 25–36.

Whittingham, D.G., Lyon, M.E., Glenister, P.H., Wotjak, C.T., and Holter, S.M. (1977). Long-term storage of mouse embryos at 196 degrees C: the effect of background radiation. *Genet. Res.* 29, 171–181.

Wotjak, C.T., Henniger, M.S., and Holter, S.M. (2003). C57BLack/BOX? The importance of exact mouse strain nomenclature. *Trends Genet.* 19, 183–184.

Yamana, K., Fukunaga, Y., Ohtani, Y., Sato, S., Nakamura, M., Kim, W.J., Akaike, T., and Maruyama, A. (2005). DNA mismatch detection using a pyrene–excimer-forming probe. *Chem. Commun. (Camb.)* 19, 2509–2511.

This page intentionally left blank

Chapter 32

Safety Analysis in Retroviral Gene Therapy: Identifying Virus Integration Sites in Gene-Modified Cells

Stephanie Laufs[1], Luisa Schubert[2], Patrick Maier[3], W. Jens Zeller[2] and Stefan Fruehauf[4]

[1] Department of Experimental Surgery, Medical Faculty Mannheim, University of Heidelberg, and Department of Molecular Oncology of Solid Tumors Unit, German Cancer Research Center (DKFZ), Heidelberg, Germany;

[2] Research Group Pharmacology of Cancer Treatment;

[3] Department of Radiation Oncology, Mannheim Medical Centre, University of Heidelberg;

[4] Center for Tumor Diagnostics and Therapy, Paracelsus Hospital, Osnabrück, Germany

32.1 INTRODUCTION

Retroviruses are a class of enveloped viruses containing two single-stranded RNA molecules as genome. Following infection, the viral genome is reverse-transcribed into double-stranded DNA, which integrates into the host genome and is referred to as a provirus. Sequences, known as long terminal repeats (LTRs), are located at each end of the viral genome, which include promoter/enhancer regions and sequences involved in integration.

In retroviral vectors the viral genes (*gag*, *pol*, and *env*) are replaced with the transgene of interest (Fig. 32.1). Due to their intrinsic ability to integrate into the host genome, retroviral vectors were the first and most commonly used gene transfer systems for transgene expression in hematopoietic cells (Edelstein *et al.*, 2007).

Low transduction efficiency of vectors into hematopoietic stem cells (HSC) made gene transfer via retroviral vectors inoperative for therapeutic application for a long time. In the early 1990s the transduction efficiency of those vectors hardly reached 1% in different clinical trials. Meanwhile, this situation has changed due to:

- the application of adequate hematopoietic growth factors (Drexler, 1996; Petzer *et al.*, 1996; Shah *et al.*, 1996),
- the use of fibronectin fragments in the transduction process (Moritz *et al.*, 1996; Hanenberg *et al.*, 1996),
- application of pseudo-typed viral particles with increased binding specificity for HSCs (Porter *et al.*, 1996; Movassagh *et al.*, 1998; Kelly *et al.*, 2000; Gatlin *et al.*, 2001; Kiem *et al.*, 1998, Sanders, 2002),
- the optimization of transgene expression of retroviral vector constructs (Halene and Kohn, 2000; Moritz and Williams, 1998).

These advanced conditions today facilitate an *in vivo* transduction efficiency of 5 to 20% (Abonour *et al.*, 2000) and contributed to the resumption of several clinical trials on gene therapy of monogenetic diseases (Bordignon and Roncarolo, 2002; Grez *et al.*, 2000; Kohn, 2002). Successful correction of gene defects and significant therapeutic benefits using retroviral gene transfer have been achieved for X-chromosomal severe combined immunodeficiency (X-SCID) (Cavazzana-Calvo *et al.*, 2000; Cavazzana-Calvo and Fischer, 2007), adenosine deaminase (ADA)-linked SCID (Gaspar *et al.*, 2004, 2006), or chronic granulomatous disease (CGD) (Ott *et al.*, 2006).

However, occurrence of retroviral vector-integration related leukemia in humans following the X-SCID trial (Bonetta, 2002; Hacein-Bey-Abina *et al.*, 2003; Marshall, 2003; Fischer and Cavazzana-Calvo, 2005; Cavazzana-Calvo and Fischer, 2007), as well as in non-human primates (Modlich *et al.*, 2005; Seggewiss *et al.*, 2006) and mice (Li *et al.*, 2002) display the risks of retroviral gene therapy. Furthermore, observation that insertion-mediated up-regulated gene expression could confer selective growth advantage to affected cells in non-human primates (Calmels *et al.*, 2005) and humans (Ott *et al.*, 2006; Schwarzwaelder *et al.*, 2007) highlights the need for a better understanding of the mechanism of retroviral vector integration into the host genomes. Therefore, methods that recognize insertion sites of retroviral vectors defined as junctions where

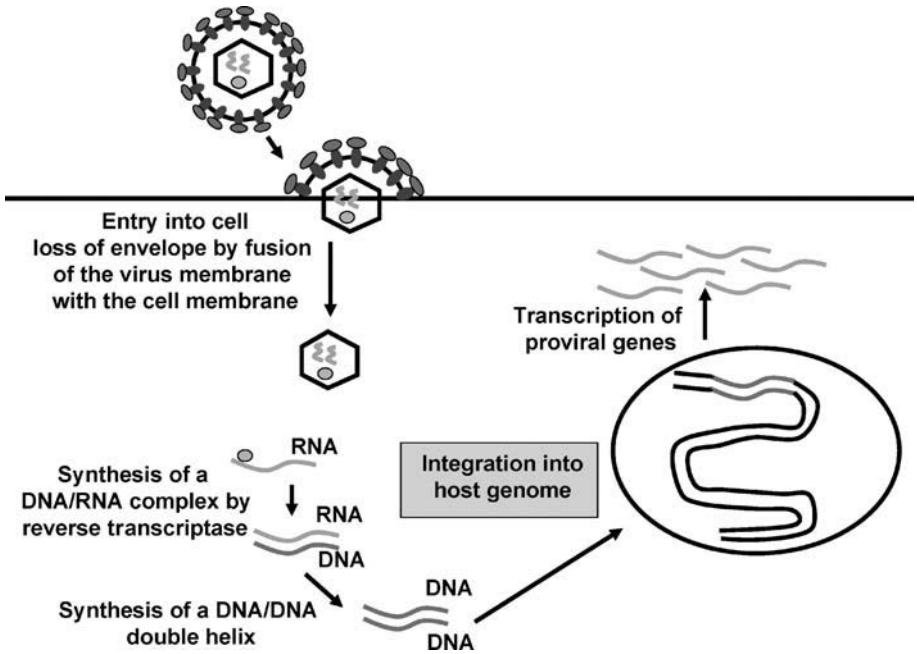


FIGURE 32.1 Retroviral gene transduction. The retroviral vector is packaged into a retroviral particle, which consists of envelope and nucleocapsid. The retrovirus interacts with the host cell surface. After fusion of the virus membrane with the cell membrane, the nucleocapsid enters the cytoplasm of the host cell. A DNA/RNA complex is synthesized by the reverse transcriptase and a double-stranded (ds) DNA is produced. After transport into the host cell nucleus, the viral DNA integrates into the host cell genome. The integrated viral DNA is referred to as a provirus. Transgene expression is mediated by the host cell transcription and translation repertoire.

genomic DNA flanks the provirus have become indispensable for the safety assessment in clinical trials and experimental studies involving retroviral-mediated gene therapy (Fruehauf *et al.*, 2002; Baum *et al.*, 2003).

32.2 METHODS USED TO DETECT RETROVIRAL INTEGRATION SITES

The experimental analysis of proviral integration sites in human stem cells (HSCs) is challenging. Stem cell tracking techniques based on detecting the retroviral integration sites and using this as a unique tag traditionally were used in isologous mouse studies (Dick *et al.*, 1985; Jordan and Lemischka, 1990). However, these techniques are not easily applied to xenogenic transplantation models due to lower contents of engrafted cells carrying the proviral tag in this latter setting. Retroviral integration patterns in transduced human cord blood cells transplanted into immune-deficient mice were detected by Southern blotting if engraftment and transduction efficiency were high (Barquinero *et al.*, 2000; Guenechea *et al.*, 2001). These early methods to isolate and characterize vector integration sites were – due to the inability to reliably detect integrations in samples with high background of untransduced cells (Laufs *et al.*, 2003) – superseded by more sensitive PCR-based methods such as inverse polymerase chain reaction (PCR; Nolta *et al.*, 1996) (Fig. 32.2a) or arbitrary primer PCR (Fig. 32.2b) (Gentner *et al.*, 2002).

Sensitivity and specificity of methods to identify and characterize vector integration sites were further increased by developing linker-mediated PCR methods that enable the simultaneous detection of different integration sites present in a sample. The oligo-cassette-mediated PCR technique described by Rosenthal and Jones (1990) was modified by Schmidt and coworkers (2001). The latter group developed linear amplification-mediated PCR (LAM-PCR) (Fig. 32.2c). LM-PCR actually represents an optimized LAM-PCR method for analysis of peripheral blood progenitor cells (PBPCs; Laufs *et al.*, 2003) (Fig. 32.2d). Exponential Fragment Marking PCR (EFRAM-PCR) (Fig 32.2e) in turn was developed as optimization of LM-PCR, reducing the required amount of starting DNA and increasing sensitivity and specificity significantly (Bozorgmehr *et al.*, 2007). Fluorescence *in situ* hybridization (FISH; see also Chapter 10) has been used successfully to confirm integration sites detected with LM-PCR (Laufs *et al.*, 2003). However, for routine application to identify retroviral integration sites this method is too complex.

32.3 IDENTIFYING VIRUS INTEGRATION SITES BY FLUORESCENCE *IN SITU* HYBRIDIZATION

Metaphase spreads from vector-transduced cells are hybridized by using a retroviral vector probe (Lichter and Cremer, 1992) followed by whole chromosome painting probes for

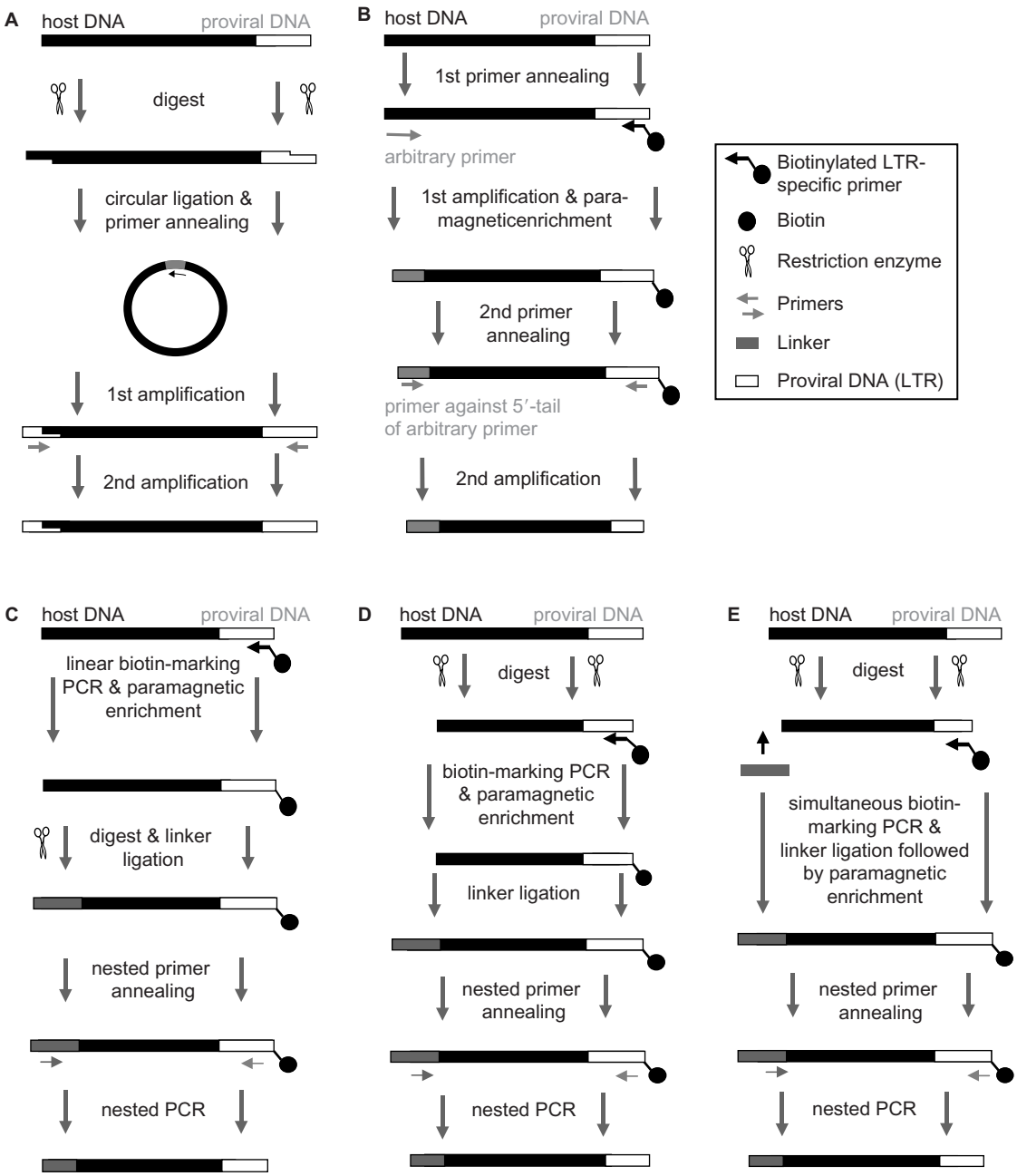


FIGURE 32.2 Schematic representation of Inverse PCR (A), Arbitrary Primer PCR (B), LAM-PCR (C), LM-PCR (D), and EFRAM-PCR (E) to identify viral vector integration sites. While Inverse PCR and Arbitrary Primer PCR are simpler PCR-based methods, linker-based methods such as LAM-, LM-, and EFRAM-PCR are more sensitive methods able to detect multiple integration sites in one reaction.

24-color FISH (Speicher *et al.*, 1996). For transduced cells, at least 10 to 15 metaphase spreads should be acquired. Subsequently, the Re-FISH protocol (Müller *et al.*, 2002) can be performed, and metaphase chromosomes are hybridized by using the multicolor FISH protocol (Speicher *et al.*, 1996). At least five pools of whole chromosome painting probes are amplified and labeled by degenerate oligonucleotide primed-PCR (DOP-PCR; Telenius *et al.*, 1992) with the use of five spectrally distinguishable fluorochromes (fluorescein isothiocyanate (FITC), Cy3, Cy3.5, Cy5, Cy5.5). Each probe is hybridized in the presence of Cot-1 DNA for 48 hours. For evaluation, metaphase spreads are acquired by using highly specific filter sets and images are processed using the appropriate software, such as the Leica Multicolor Karyotyping (MCK) software (Fig. 32.3).

32.4 IDENTIFYING VIRUS INTEGRATION SITES USING PCR-BASED METHODS

PCR-based methods describe an integration site in detail by amplifying the fusion fragment where the proviral LTR

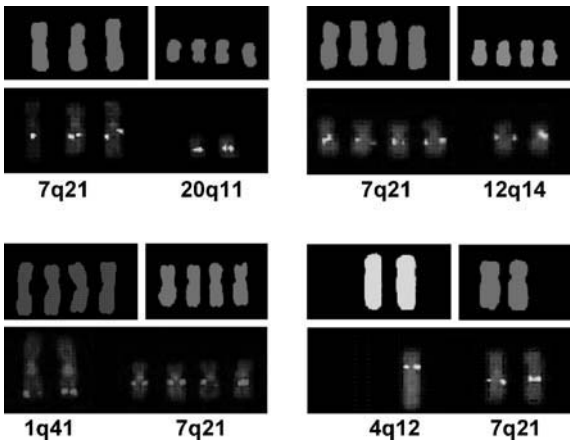


FIGURE 32.3 Chromosomal mapping of proviral sequences by FISH in cell-lines. FISH analysis using SF1m vector (gammaretroviral vector with *MDR*1 as transgene) plasmid DNA as probe and subsequently performed multicolor FISH was established to detect proviral inserts in SF1m-transduced HT1080 cell-line clones. Chromosomes were identified by different colors. Chromosomes with a hybridization signal are highlighted: the *MDR*1 transgene could be detected on chromosomes 1q41, 4q12, 12q14, and 20q11. The human wild-type *MDR*1 locus (7q21) could also be detected. The highlighted chromosomes (top panel) show no translocations, so that the specific vector signal (bottom panel) can be assigned clearly to the depicted chromosomes. (Adapted from Laufs *et al.*, 2003; with permission.)

joins the genomic host DNA. Usually, PCR amplification requires knowledge about both sequences flanking the DNA of interest since specific primers need to be designed. Nonetheless, several methods such as inverse PCR or arbitrary primer PCR have been established to overcome this limitation and allow PCR to be used for amplification of DNA outside a region of known sequence.

32.4.1 Inverse PCR and Arbitrary Primer PCR

In inverse PCR (Nolta *et al.*, 1996), a "restriction-digested" DNA template containing the junction between vector LTR and genomic host DNA, is circularized by self-ligation. The circular-ligated DNA serves as template for inverse PCR since it contains a fragment of known vector LTR sequence. This fragment enables the design of two primers. Starting from both primers, replication progresses in the opposite direction with direct amplification of the sequence of interest (Fig. 32.2a). However, sensitivity of this method has been questioned as a detection rate of only 30 to 40% of integration sites was reported (Kim *et al.*, 2000). Limited sensitivity of inverse PCR might be contributed to inefficient amplification subsequently to numerous DNA preparation and dilution steps. Additionally, this approach cannot reliably detect multiple integration sites in one reaction (Kim *et al.*, 2000).

Arbitrary primer PCR (Gentner *et al.*, 2003), also referred to as two-step PCR, in contrast is a sensitive and rapid method to identify retroviral integration sites in small clonal cell samples such as hematopoietic colonies. It does not require DNA manipulations like DNA restriction enzyme digest, adapter ligation, or circularization of DNA fragments prior to amplification. Therefore, arbitrary primer PCR can potentially be performed with very small amounts of DNA. Methods using arbitrary primers (Silver and Keerikatte, 1989; Sørensen *et al.*, 1993) or primers against repetitive genomic DNA sequences (Butler *et al.*, 2001) that hybridize to the unknown flanking DNA regions have been described for the analysis and identification of retroviral integration sites.

To detect proviral integrants in human genomic DNA, a PCR reaction is performed using biotinylated LTR-specific and arbitrary primers. Arbitrary primers consist of a long 5′ tail of known sequence, seven random nucleotides, and five fixed nucleotides at the 3′ end. Due to the fact that these primers are partly degenerate (random nucleotides), they enable amplification of proviral flanking human DNA fragments, without requiring any information on sequence. The biotinylated fragments are enriched by streptavidin-coated paramagnetic beads. A nested PCR is performed using a nested LTR-specific primer and a primer that binds to the 5′ tail sequence of the arbitrary primer. This second PCR enhances the specificity of the amplification. The obtained PCR products are excised from agarose gel and directly sequenced following DNA extraction (Fig. 32.2b).

The arbitrary primer PCR technique (Sørensen *et al.*, 1993) was adapted by Gentner and coworkers (2002), and the results were validated with LM-PCR (described below) on retrovirally transduced cell-line clones. The applicability of arbitrary primer PCR was shown to analyze retroviral integration sites in colony-forming human $CD34^+$ PBPCs, a cell population that is an indicator for transduction of HSCs. Seven different arbitrary primers were constructed according to a model described by Sørensen and coworkers (1993), which has previously been used for integration analysis of wild-type retroviruses (Sørensen *et al.*, 1993; Sørensen *et al.*, 1996) (Fig. 32.4).

To identify the most suitable primers for the application, arbitrary primer PCRs were performed with each of those primers in combination with a specific primer for the known proviral LTR segment on two human fibrosarcoma cell-line clones (HT1080 clone N2 and N3) which were transduced with the MDR1 gene. Three arbitrary primers (FP2, FP4, and FP5; Fig. 32.4) turned out to be most useful, as they did not avidly bind to proviral sequences close to the 3′ LTR.

FP2, FP4, and FP5 generated PCR products of the LTR-genomic DNA junctions (5′ LTR external fragment) in both of the HT1080 cell-line clones, which were subsequently sequenced. In clone N3, two different junctions were detected. The first one was detected by FP2 and FP5, and the second junction was detected by FP4. Importantly, both

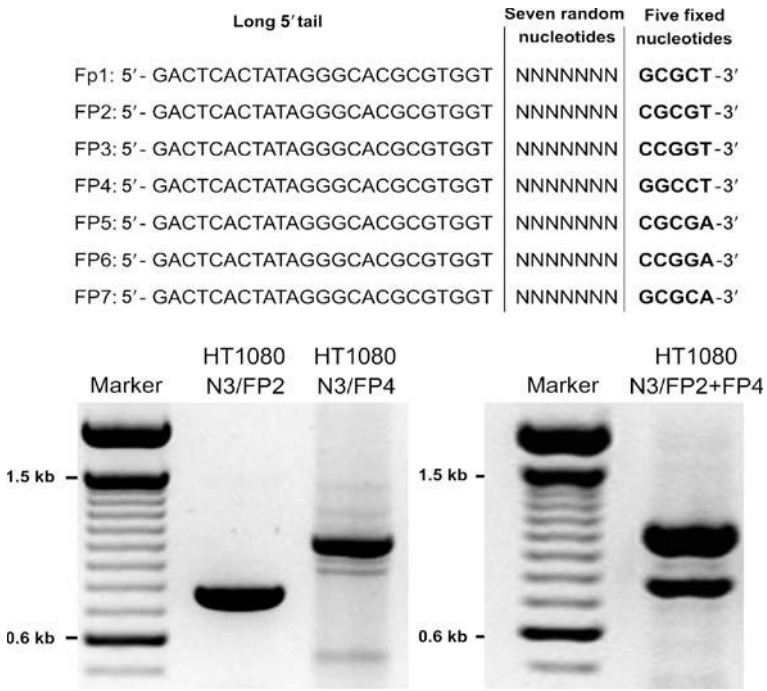


FIGURE 32.4 Combination of arbitrary primers allows simultaneous detection of retroviral integration sites. Sequences of arbitrary primers with partly degenerate tail ("N", top panel). Individual SF1m-transduced HT1080 cells were selected by single cell deposition. Arbitrary primer PCR with a primer cocktail allowed simultaneous detection of different retroviral integration sites in HT1080 cell-line clone N3. Arbitrary primer PCR was performed on HT1080 cell-line clone N3 using either arbitrary primers FP2 and FP4 alone or FP2/FP4 in a cocktail.

integrations could be detected in one reaction using a cocktail of the primers FP2 and FP4. Furthermore, these HT1080 cell-line clones were analyzed with the LM-PCR method in order to validate the results. After extensive studies with three different restriction enzymes to create a large variety of amplification permissive restriction fragment length polymorphisms, a total of three different retroviral integrations in the two HT1080 cell-line clones were found. It appeared that vector integrations detected with arbitrary primer PCR were identical to those detected by LM-PCR, thus confirming the reliability and sensitivity of both methods.

Next, the optimized arbitrary primer PCR technique was applied in order to detect retroviral vector (a Friend mink cell focusforming/murine embryonic stem-cell virus-type vector) integration sites in colony-forming human PBPCs. Mobilized $CD34^+$ PBPCs were transduced with the SF1m retroviral vector and cultured in semisolid medium in the presence or absence of vincristine. Vincristine was added to the plates in order to select for MDR1-expressing (vector positive) colonies. As expected, the percentage of MDR1-positive colonies was up to five fold increased in the presence of vincristine.

A total of 182 individual colonies from five different SF1m transduced PBPC donor samples were analyzed for the presence of the MDR1 gene. In total, 72 colonies (40%) were positive for the MDR1 transgene and thus qualified for integration site analysis. The DNA amount of a MDR1-positive colony corresponds to a sensitivity of 75 or more cells per reaction. Of the 72 colonies, 61 yielded a specific PCR product; that is, either an informative band (25 colonies) or an internal 3′ LTR fragment. In most cases, one to five clear, robust bands per colony were amplified. In 10 out of 25 colonies containing external bands, the informative fragment was the only amplicon seen in the gel.

Arbitrary primer PCR has been used successfully to study wild-type retroviral integrations in cell-lines (Sørensen *et al.*, 1993), retroviral integrations in tumor DNA from mice (Sørensen *et al.*, 1996), and hepatitis B virus integrations in peripheral blood mononuclear cells (Laskus *et al.*, 1999). It was demonstrated that this simple and fast approach also can be used to study retroviral vector integrations in day 14 colonies of $CD34^+$ PBPCs, a setting in which DNA amount is very limited. High-quality sequences of the LTR-genomic DNA junctions could be obtained after only five standard steps (1st PCR, enrichment, 2nd PCR, agarose gel electrophoresis, direct sequencing).

32.5 IDENTIFYING VIRUS INTEGRATION SITES USING LINKER-MEDIATED PCR METHODS

Linker-mediated PCRs such as linear amplification-mediated (LAM-) PCR (Schmidt *et al.*, 2001, 2002), ligation-mediated (LM-) PCR (Laufs *et al.*, 2003), or exponential

fragment marking (EFRAM-) PCR (Bozorgmehr *et al.*, 2007) have been designed to simultaneously detect different integration sites in a sample. All linker-mediated PCRs share the general principle to detect the junction between the integrated provirus and the flanking genomic sequence: while the sequence of the LTR is known, the flanking genomic sequence is unknown. Therefore, following restriction enzyme digest, a linker is ligated to the blunt-end cleavage site. Integration sites are then amplified using one primer that binds to the viral DNA end and another primer binding to the DNA linker. After performance of a second amplification with nested primers, PCR products containing host-virus DNA junctions are cloned, sequenced, and mapped to the corresponding genome sequence. Protocols, however, differ in terms of initial DNA amounts, digestion steps, restriction enzymes, linkers, polymerases, template types, and PCR product cloning.

32.5.1 Linear Amplification-Mediated PCR

In LAM-PCR (Schmidt *et al.*, 2001, 2002), isolated DNA is linearly preamplified with biotinylated LTR-specific primers. Biotinylated fragments are enriched by the use of paramagnetic beads coated with streptavidin, thus binding the target DNA to a solid phase. Using random hexanucleotide primers, second strand of the target fragment is synthesized. Resulting double-stranded DNA is digested, then an adapter oligonucleotide is ligated to the LTR-distant portion of digested fragments. Using linker-specific and LTR-specific primers, nested PCR is performed and resulting products are cloned and sequenced (Fig. 32.2c). LAM-PCR sensitivity and specificity have been questioned by a recent study reporting failure of detection of 30 to 40% of the clones in artificially generated mixtures of retrovirally marked clones (Harkey *et al.*, 2007).

32.5.2 Ligation-Mediated PCR

In LM-PCR (Laufs *et al.*, 2003), genomic DNA is digested with a restriction enzyme, which cuts upstream from the LTR in the genomic sequence and in the proviral sequence, but not within the LTR region. Fragments containing the LTR-genomic DNA junctions are tagged with a biotinylated LTR specific primer. Since the 5′ and 3′ LTR have identical sequence, two types of fragments are tagged: the external band, which contains the unknown flanking sequence; and the internal band, which contains only vector sequence. The biotin-tagged fragments are enriched using streptavidin-coated paramagnetic beads. Thus, other DNA fragments are removed.

In the next step, an adapter oligonucleotide cassette of known sequence is ligated on the LTR-distant end of fragments allowing PCR amplification. To increase the specificity and the sensitivity of PCR, a second, nested PCR is performed. The resulting PCR products are excised from agarose gel, separately cloned following DNA extraction, and sequenced (Fig. 32.2d). Sequences are confirmed as integration sites when the following criteria are met:

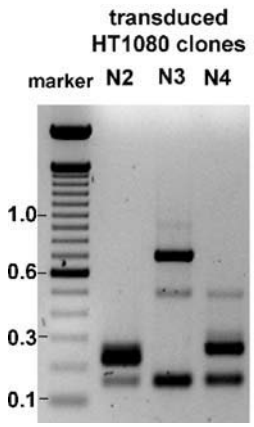
- Sequence is flanked by LTR and adapter sequences
- Sequence is matched uniquely (>90% identity) to human genome

The LM-PCR method first has been established and optimized on transduced HT1080 cell-line clones (Laufs *et al.*, 2003) obtained by single cell sorting. Therefore, two cell-line clones with one integration site each and one cell-line clone with two integration sites have been analyzed. These cell-line clones were mixed together to test for the ability to detect multiple clones in one reaction. Four different integration sites and one internal band were readily co-amplified (Fig. 32.5). These integration sites obtained by LM-PCR were confirmed by FISH analysis (Table 32.1).

The addition of genomic mouse DNA (as background DNA) to the sample did not change the results, and negative controls with mock-transduced mouse DNA produced no PCR bands on agarose gel. An additional optimization step was the excision of the bands from agarose gel, which increased the number of identified integrations.

The LM-PCR method has been performed on transduced human $CD34^+$ PBPCs, repopulating the bone marrow of NOD/SCID mice (Laufs *et al.*, 2003). The human multidrug resistance 1 (*MDR1*) gene was transferred into human mobilized $CD34^+$ PBPCs by a Friend mink cell focusforming/murine embryonic stem-cell virus-type vector, the SF1m retroviral vector, carrying the MDR1 gene (Baum *et al.*, 1995). This vector has been shown to mediate high P-glycoprotein (*MDR1* gene product) expression, conferring a drug-resistant phenotype in early hematopoietic cells *in vitro* (Baum *et al.*, 1995; Eckert *et al.*, 1996; Hildinger *et al.*, 1998, 1999) and permitting sustained transgene expression *in vivo* (Baum *et al.*, 1996; Schiedlmeier *et al.*, 2000). Chimeric mouse bone marrow DNA was digested and LM-PCR was performed. An integration site library was constructed by cloning the resulting PCR products. The cloning procedure allowed sensitive detection of fragments. Subsequently, the cloned PCR products were sequenced to prove the presence of LTR-genomic DNA junctions, and up to 32 different clones were identified from one chimeric bone marrow. Repeated LM-PCR using the same chimeric bone marrow DNA resulted in less than 10% overlap between two separate LM-PCR analyses, pointing to the polyclonality of human hematopoiesis following experimental transplantation.

To investigate the specificity of the method, the LM-PCR was performed on DNA, isolated from bone marrow of three mice that had received transplants of non-transduced human $CD34_+$ cells (mock transduction). The LM-PCR product was concentrated and cloned as a whole, omitting



cell line clone	band-length (kb)	vector sequence	human flanking DNA sequence	GenBank accession number	chromosomal localization
N2	0.22	..gggctttca	CTTCCCATTCT..	AC121337	12q13
N3	0.33	..gggctttca	TCAGACCCTTC..	AC092015	1q41
N3	0.67	..gggctttca	TATGGAAGCTT..	AC083798	3q21
N4	0.25	..gggctttca	AACCCATTCAC..	AL109824	20q11
IB	0.15	..gggctttca	ttccccccttttct		

FIGURE 32.5 Retroviral integrations detected by LM-PCR. Three SF1m-transduced HT1080 cell-line clones were analyzed. DNA from each clone was used in separate reactions (N2, N3, and N4) and mixed together in one reaction (N2/3/4). The internal band (IB) originates from the 3′ LTR and is identical for all SF1m vector-transduced cells. The length and the sequence of each band are shown. The mapping result of the detected human flanking sequences is shown as the matched Genbank accession number and as chromosomal localization. (Adapted from Laufs *et al.*, 2003; with permission.)

TABLE 32.1 Correlation between flanking DNA sequences mapped to human chromosomes and localization of FISH signals.

Cell-line clone	Integration site sequences obtained by LM-PCR	FISH signals
N2	**12q13**	**12q14**
N3	**1q41**	**1q41**
N4	**20q11**	**20q11**

procedures in which PCR fragments could get lost. Not a single LTR-flanking DNA junction could be identified in three mock mice that had received untransduced human PBPCs from different transplantation experiments (Laufs *et al.*, 2003).

32.5.3 Exponential Fragment Marking PCR

In order to reduce the starting amount of DNA required to perform an LAM-PCR or LM-PCR reaction, highly sensitive EFRAM-PCR was developed. While in LAM-PCR and LM-PCR biotin marking is performed prior to adapter ligation, in EFRAM-PCR a linker is ligated prior to biotin marking. Furthermore, in contrast to LAM-PCR which uses linear biotin marking of DNA fragments, and LM-PCR where fragments are biotin marked in a one-step amplification reaction, target fragments in EFRAM-PCR are simultaneously biotin marked and amplified (Fig. 32.2e). Thus, both sensitivity and specificity increase significantly. As a result, 100 ng of starting DNA amount containing 1 ng vector transduced DNA and 99 ng untransduced background DNA have shown to be sufficient for detection of viral integration sites (Fig. 32.6) (Bozorgmehr *et al.*, 2007).

Sensitivity and specificity of LM-PCR and EFRAM-PCR were compared in a study about biosafety of retroviral MDR1 gene transfer in a non-human primate model with direct relevance to human biology. Mobilized $CD34^+$ PBPCs from two rhesus macaques were transduced either with the Harvey-based retroviral vector containing the *MDR1* gene (HaMDR1) or the neomycin-phosphotransferase (NEO_R) gene as a control (G1Na). Following transduction, cells were autologously transplanted into the monkeys (Sellers *et al.*, 2001). No evidence for expansion of HaMDR1 vector-transduced cells or a myeloproliferative disorder had been observed in a one year follow-up (Sellers *et al.*, 2001). The observation period of one year was prolonged up to four years after transplantation. The analysis identifying and characterizing specific vector integration sites in the genome of mononuclear cells and granulocytes of both monkeys using EFRAM-PCR and LM-PCR followed by sequencing of gained products was extended. This is the longest study period of *MDR1* transduced cells in rhesus macaques or other non-human primates reported so far (Fig. 32.7; Bozorgmehr *et al.*, 2007).

Integrations could be assigned to be either G1Na or HaMDR1 vector integrations due to the fact that both vectors differ in their LTR sequence at position 68 which is a thymidine in the HaMDR1 LTR and a cytosine in the G1Na LTR. Besides characterizing the integration patterns of both the G1Na and HaMDR1 vectors, the risk of insertional mutagenesis was assessed by determining the occurrence of integrations in oncogenes.

Overall, 122 integrations were detected (Bozorgmehr *et al.*, 2007). All 122 detected integration sequences were aligned with the human genome using the IntegrationMap

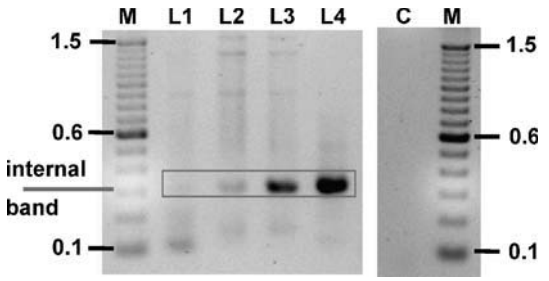


FIGURE 32.6 Agarose gel electrophoresis of PCR products obtained with EFRAM-PCR. EFRAM-PCR was performed on serial dilutions of the cell-line HT1080 containing 0.1–100% retrovirally transduced DNA in a background of untransduced DNA. Lane 1 (L1): 0.1 ng vector-transduced DNA and 99.9 ng untransduced background DNA. Lane 2 (L2): 1 ng vector-transduced DNA and 99 ng untransduced background DNA. Lane 3 (L3): 10 ng vector-transduced DNA and 90 ng untransduced background DNA. Lane 4 (L4): 100 ng vector-transduced DNA. Internal band representing internal control of the EFRAM-PCR method and expected at 300 bp is visible in L2, L3, and L4. EFRAM-PCR allows detection of 1 ng transduced DNA in a background of 99 ng untransduced DNA (L2). To confirm specificity of the obtained bands, they were cloned and sequenced. Mock controls were negative. M: size marker (numbers denote fragment size in kb). C: negative control.

tool (Giordano *et al.*, 2007). This bioinformatic tool can be used on http://www.gtsg.org. With IntegrationMap, valid integration sequences get automatically assigned to their genomic location (Giordano *et al.*, 2007). Of 122 integration sequences, 87 were successfully mapped: 72 integration sequences contained the G1Na vector, while the remaining 15 integration sequences contained the HaMDR1 vector. Of 87 mapped sequences, 46 integrations occurred into RefSeq genes, of which five integrations were located in gene loci involved in malignant transformation of cells. Of those five integrations, three occurred with the G1Na vector and two with the HaMDR1 vector. The integrations hit the *TIAM1* (T-cell lymphoma invasion and metastasis 1) oncogene, the *SKI* (v-ski sarcoma viral oncogene homolog) gene, the *MECT1* (mucoepidermoid carcinoma translocated 1 isoform) gene, as well as the *SIN3A* (transcriptional co-repressor SIN3A) transcription regulating gene. In line with reported recurrent integrations into the *EVI1* (ecotropic virus integration site 1) and related *MDS/EVI1* gene locus in animal models and humans (Buonamici *et al.*, 2004; Calmels *et al.*, 2005; Kustikova *et al.*, 2005; Du *et al.*, 2005; Nienhius *et al.*, 2006; Ott *et al.*, 2006; Deichmann *et al.*, 2007), integration into the

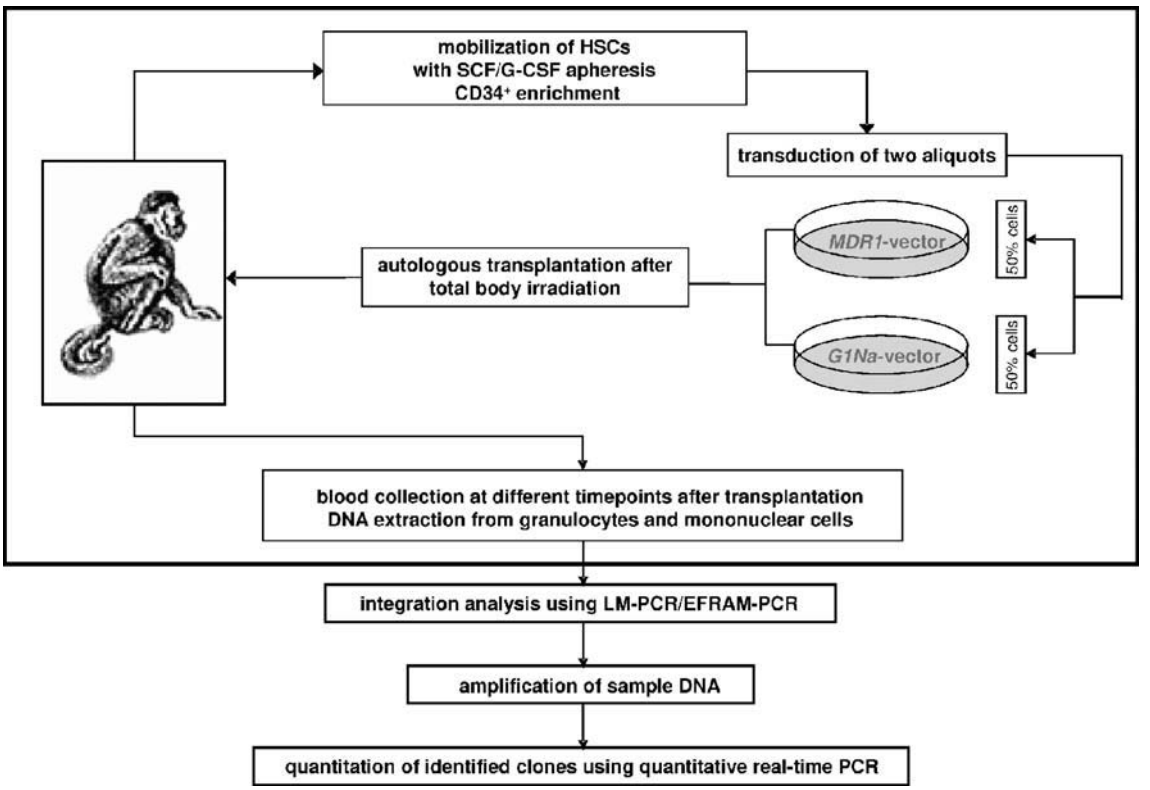


FIGURE 32.7 Peripheral blood cells from two rhesus macaques were mobilized and $CD34^+$ cells were enriched. Obtained cells from both macaques, respectively, were split into equal aliquots and transduced either with an *HaMDR1* vector expressing the *MDR1* gene or with a control *NEO* expressing vector *G1Na*. Before autologous transplantation, animals were conditioned with 5 Gy of total body irradiation. Blood samples were taken at different time points, from eight weeks up to four years after transplantation and separated into granulocytes and mononuclear cells. DNA was isolated from both cell types. Integration analysis was performed applying LM-PCR or EFRAM-PCR. After integration site identification, cell clones were quantified with quantitative real-time PCR.

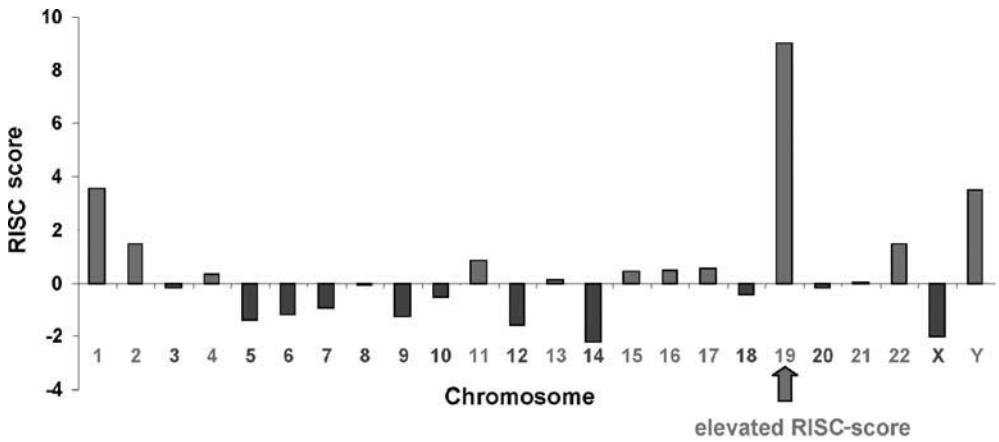
EVI1 proto-oncogene was also found. While Li and coworkers (2002) observed leukemia development in mice after integration of a retroviral vector into EVI1, here no evidence for retroviral vector or *MDR1*-associated leukemia in both monkeys was observed.

To test whether the number of integrations was equally distributed along the chromosomes, a RISC-score (retroviral insertion site into chromosomes) was calculated (Laufs *et al.*, 2004). Taking into account the size of the chromosomes, the RISC-score compares the observed number of integrations (oi) with the statistically expected number of integrations (ei) in each chromosome. Expected integration counts are computed assuming a discrete uniform distribution of integrations but considering differences in chromosomal size. Considering the total number of mapped integration site sequences of 72 for the G1Na vector, 1.4 integrations (ei) were expected, whereas 5 integrations (oi) were observed for chromosome 19. Thus, a significantly elevated RISC-score value of 9.0 was determined for G1Na integrations into chromosome 19 (Fig. 32.8).

To monitor behavior and kinetics of transduced cells after transplantation, the contribution of transduced cells to hematopoiesis in one monkey was assessed over the period of four years applying a fluorescence-based real-time quantitative PCR (QRT-PCR) method subsequent to total DNA amplification. Single transduced repopulating cells as well as HaMDR1 and G1Na vector amount were followed and monitored in a long-term follow-up of four years. Compared to earlier studies that analyzed single clones in samples taken six or eight weeks after transplantation (Nagy *et al.*, 2004) or the analysis of exclusively one integration site (Calmels *et al.*, 2005), this is the first time that single transduced repopulating cells have been followed by quantification in a time frame of eight weeks up to four years following transplantation. Observed overall contribution of HaMDR1 or G1Na transduced cells was stable (Fig. 32.9a). Furthermore, contribution of single clones to transduced cell fraction and hematopoiesis was low, ranging between 0 and 2.4% to overall transgene-modified cell population (Fig. 32.9b). Thus, no evidence for clonal dominance or proliferative disorder associated to retroviral or MDR1 transgene vector gene transfer was observed. Moreover, no indication of correlation between site of integration, clone size, or vector type was determined.

32.6 CONCLUSIONS

For a long time, retroviral vector integration was supposed to be random. However, MLV-based vector studies reported preferential vector integrations into genes, near transcription start sites of active genes, SINEs, and LINEs



$$\text{RISC-score} = \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$\text{expected integrations} = \frac{\text{determined bases of chromosome}}{\text{determined bases of genome}} \times \text{total number of mapped integrations}$$

chromosome	19
expected integrations	1.4
observed integrations	5
RISC-score	9

FIGURE 32.8 Schematic distribution of retroviral vector integrations into chromosomes of retrovirally transduced rhesus macaque mononuclear cells and granulocytes. RISC-score (**r**etroviral **i**ntegration e**s**timate into **c**hromosome) is defined from observed integrations (oi) and expected integrations (ei). For calculating the expected integrations, the chromosome size and completed human genome sequences (EMBL genome monitoring table, http://www.ebi.ac.uk/genomes/mot) are taken into account. In case of non-preferential integration, an RISC-score of 0 would be expected. For chromosomes 19 a much higher RISC-score is obtained than for the other chromosomes. Results based on the human genome sequencing status from 10/24/2002.

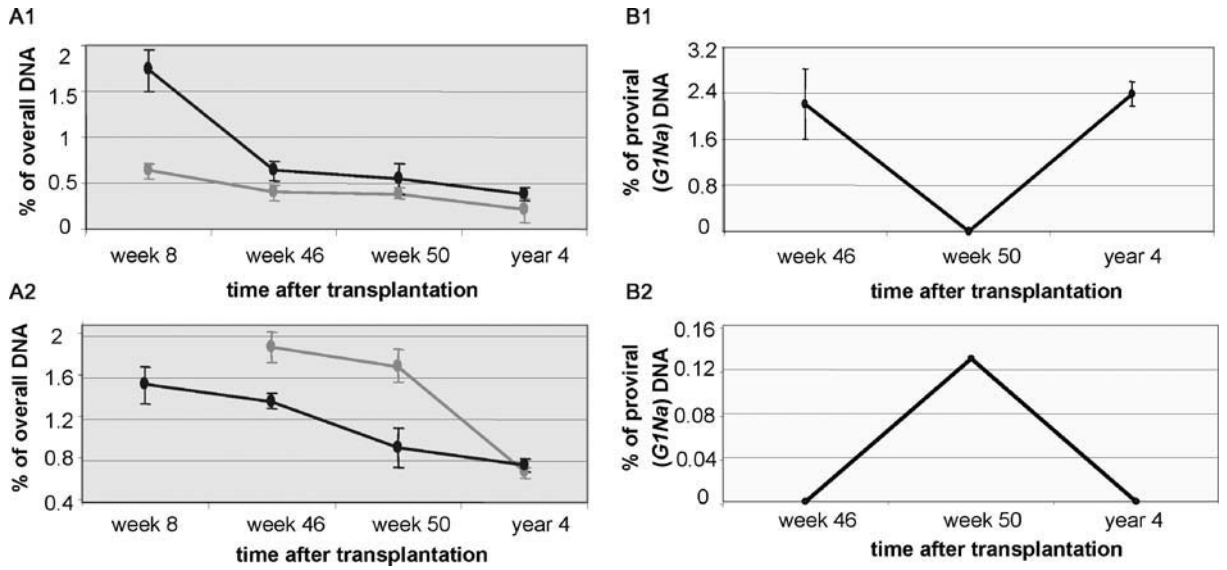


FIGURE 32.9 **A.** Contribution of total *HaMDR1* (A1) and *G1Na* (A2) vector-transduced cells in granulocytes (gray) and mononuclear cells (black) of a rhesus macaque measured via quantitative real-time PCR in samples eight weeks up to four years after transplantation (*G1Na* clone quantification at time point week 8 after transplantation in granulocytes not shown). **B.** Contribution of single clones to transgene modified DNA. Two *G1Na* clones (B1 and B2) were quantified via QRT-PCR in DNA samples of granulocytes collected 46 weeks, 50 weeks and four years after transplantation of a rhesus macaque. Size of single clones referring the transgene-modified population displayed in percent.

(Wu *et al.*, 2003; Laufs *et al.*, 2003, 2004; Hematti *et al.*, 2004; Mitchell *et al.*, 2004; Schroder *et al.*, 2002; Holmes-Son *et al.*, 2001; Kitamura *et al.*, 1992; Weidhaas *et al.*, 2000; Deichmann *et al.*, 2007). Furthermore, the presence of DNA binding proteins, chromatin structure of DNA and cellular targeting proteins that affect DNA accessibility (Bushman *et al.*, 2005; Wu *et al.*, 2005) have shown to influence integration target site of retroviral vectors. In fact, besides increased frequency of integrations into chromosome 19 in transplanted rhesus macaques, significantly increased frequency of vector integrations into chromosomes 19 and 17 (Fig. 32.8) and into specific regions of chromosomes 6, 13, and 16 were also observed, following integration analysis in immune-deficient mice. Here, animals were transplanted with multiple transduced human PBPCs. After LM-PCR and sequencing of retroviral vector integration sites, 141 proviral inserts were unambiguously mapped to the human genome. These findings, reported by Laufs and coworkers (2003), revealed for the first time that retroviral vector integration into human marrow repopulating cells can be non-random ($P = 0.00037$). On chromosome 17 and chromosome 19, ten and nine integrations resulting from three donors were found, respectively, and three integrations were expected for both chromosomes when the chromosome size and completed human genome sequences (EMBL genome monitoring table, http://www.ebi.ac.uk/genomes/mot/) were taken into account.

Besides non-random retroviral vector integration into host genome, the risk of retroviral vector-induced mutagenesis has been underestimated. Initially, the risk of insertional mutagenesis caused when genes involved in critical cellular functions were disrupted or activated by insertion of the vector into the genome was supposed to be low ranging between 10^{-6} and 10^{-8} per insertion event (Stocking *et al.*, 1993). However, this risk has recently been considered to be much higher: transforming insertion event frequency in a 10 kb diameter of a potential proto-oncogene is now estimated to range between 10^{-2} and 10^{-3} (Baum *et al.*, 2003). Activation of proto-oncogenes or disruption of tumor-suppressor genes by a retroviral vector and subsequent leukemia development are actually known complications of retrovirus-mediated gene transfer. However, no evidence of abnormal events and insertional mutagenesis was observed in several retroviral gene transfer studies including humans, mice, dogs, or non-human primates (Kohn *et al.*, 2003; Bunting *et al.*, 1999; Kiem *et al.*, 1998, 2004), although in one animal model insertion of a retroviral vector contributed to malignancy (Li *et al.*, 2002). Following sequential transplantation of retroviral-marked HSCs leukemic transformation and hematopoietic disorder development was observed. Molecular analysis showed that in all diseased mice the retroviral vector had integrated in the first exon of the *EVI1* oncogene (ecotropic viral integration site1) encoding for a transcription factor, probably leading to its activation (Li *et al.*, 2002; Baum *et al.*, 2003; Nienhuis *et al.*, 2006). In normal hematopoietic cells, EVI1 expression is not detected but its activation is associated with the development of aggressive myeloid leukemia (Buonamici *et al.*, 2004).

Also in humans biosafety and applicability of retroviral gene transfer have been dramatically questioned. Insertional mutagenesis occurred in the first clinical gene

therapy trial in 1999 where retroviral vectors were used to treat patients suffering from SCID (Cavazzana-Calvo *et al.*, 2000). Treatment was successful in many patients, but three children developed a T-cell lymphoproliferative disorder; insertional mutagenesis due to integration in or near the LMO2 oncogene was blamed for this (Hacein-Bey-Abina *et al.*, 2003; Fischer and Cavazzana-Calvo, 2005). Furthermore, a fatal myeloid malignancy in one of seven rhesus macaques that had been transplanted with retrovirally transduced autologous $CD34^+$ cells has been reported five years after transplantation (Seggewiss *et al.*, 2006). Dominance of single HSC clones following retroviral gene transfer without development of leukemia has been observed in several studies: in mice enhanced survival and clonal dominance of HSCs due to insertions following retroviral gene transfer was observed (Du *et al.*, 2005; Kustikova *et al.*, 2005). In non-human primates, growth or survival advantages were determined for retrovirally transduced $CD34^+$ cells that displayed insertion into the MDS/EVI1 gene region (Calmels *et al.*, 2005), an alternative isoform of the *EVI1* gene locus (Buonamici *et al.*, 2004). With 14 MDS/EVI1 integrations in granulocytes of nine animals, integration into this locus was overrepresented (Calmels *et al.*, 2005). In addition, two patients with CGD successfully treated by gene therapy have shown clonal dominance with integrations in the MDS/EVI1, and related PRDM16 gene which is also termed MEL1 (MDS1/EVI1-like gene) (Ott *et al.*, 2006).

All these events associated with retroviral gene transfer highlight the need for the above described methods to characterize vector integration sites in individual stem cells and their progeny in order to ensure biosafety of retroviral vectors. Therefore, highly sensitive and specific integration analyzing methods such as LM-PCR or EFRAM-PCR and sequencing of retroviral vector integration sites are indispensable tools.

Following integration analysis, bioinformatic tools such as IntegrationMap and IntegrationSeq (also accessible on http://www.gtsg.org) are highly valuable since they enable fast large-scale mapping of integration site sequences and standardized comparison of generated data. While IntegrationSeq trims and processes sequences yielded via molecular biology methods such as LM-PCR or EFRAM-PCR, IntegrationMap enables automatic genomic mapping since it first runs a BLAST search against the RefSeq_mRNA database to identify exon integrations. Mapping to specific chromosomes is yielded by a MEGABLAST analysis. The output of IntegrationMap contains information about chromosomal localization of integrations, the exact localization about the transcriptional unit (exon/intron number), and information about several parameters like the distance from adjacent transcription start sites, next repetitive elements (e.g. short interspersed nucleotide elements (SINE), long interspersed elements (LINE), simple repeats, LTR elements), or next CpG-Islands. Both tasks have been validated and demonstrated reliability and efficiency. Two hundred and fifty-nine raw sequences obtained from integration-site analysis were processed via IntegrationSeq, further processed with the IntegrationMap task, and compared to conventional analysis. Use of IntegrationSeq and IntegrationMap revealed higher sensitivity and specificity, and results were calculated 15 times faster than using conventional manual analysis. IntegrationMap output files displayed 99.8% identity with results gained by much slower conventional mapping methods using the ENSEMBL alignment tool (Giordano *et al.*, 2007; Appelt *et al.*, 2009).

Moreover, RISC-score calculation is highly helpful to assess safety of current preclinical or clinical stem cell gene therapy protocols. A data bank project "retroviral insertion estimate of chromosomal integration" (RISC) has been set up to recognize critical genomic regions and genes involved with possible transforming capacity. Such data collections will allow further safety-type studies and monitoring that are needed to restore confidence in clinical trials involving integrating vectors.

ACKNOWLEDGEMENTS

This work was supported in part by Grant 10-1294-Ze3 of the Deutsche Krebshilfe/Dr. Mildred-Scheel-Stiftung, by Grant FR1732/3-1, FR1732/3-2 of the Priority Program SPP1230 of the German Research Foundation (Bonn, Germany) and the H.W. & J. Hector Foundation (Mannheim, Germany).

REFERENCES

Abonour, R., Williams, D.A., Einhorn, L., Hall, K.M., Chen, J., Coffman, J., Traycoff, C.M., Bank, A., Kato, I., Ward, M., Williams, S.D., Hromas, R., Robertson, M.J., Smith, F.O., Woo, D., Mills, B., Srour, E.F., and Cornetta, K. (2000). Efficient retrovirus-mediated transfer of the multidrug resistance 1 gene into autologous human long-term repopulating hematopoietic stem cells. *Nat. Med.* 6, 652–658.

Appelt, J.U., Giordano, F.A., Ecker, M., Roeder, I., Grund, N., Hotz-Wagenblatt, A., Opelz, G., Zeller, W.J., Allgayer, H., Fruehauf, S., Laufs, S. (2009). HYPERLINK "http://www.ncbi.nlm.nih.gov/pubmed/19387483?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum" QuickMap: a public tool for large-scale gene therapy vector insertion site mapping and analysis. *Gene Ther.* 2009, Apr 23. [Epub ahead of print].

Barquinero, J., Segovia, J.C., Ramirez, M., Limon, A., Guenechea, G., Puig, T., Briones, J., Garcia, J., and Bueren, J.A. (2000). Efficient transduction of human hematopoietic repopulating cells generating stable engraftment of transgene-expressing cells in NOD/SCID mice. *Blood* 95, 3085–3093.

Baum, C., Hegewisch-Becker, S., Eckert, H.G., Stocking, C., and Ostertag, W. (1995). Novel retroviral vectors for efficient expression of the multidrug resistance (mdr-1) gene in early hematopoietic cells. *J. Virol.* 69, 7541–7547.

Baum, C., Eckert, H.G., Stockschlader, M., Just, U., Hegewisch- Becker, S., Hildinger, M., Uhde, A., John, J., and Ostertag, W. (1996). Improved vectors for hematopoietic stem cell protection and in vivo selection. *J. Hematother.* 5, 323–329.

Baum, C., Dullmann, J., Li, Z., Fehse, B., Meyer, J., Williams, D.A., and von Kalle, C. (2003). Side effects of retroviral gene transfer into hematopoietic stem cells. *Blood* 101, 2099–2114.

Bonetta, L. (2002). Leukemia case triggers tighter gene-therapy controls. *Nat. Med.* 8, 1189.

Bordignon, C., and Roncarolo, M.G. (2002). Therapeutic applications for hematopoietic stem cell gene transfer. *Nat. Immunol.* 3, 318–321.

Bozorgmehr, F., Laufs, S., Sellers, S.E., Roeder, I., Zeller, W.J., Dunbar, C.E., and Fruehauf, S. (2007). No evidence of clonal dominance in primates up to 4 years following transplantation of multidrug resistance 1 retrovirally transduced long-term repopulating cells. *Stem Cells* 25, 2610–2618.

Bunting, K.D., Galipeau, J., Topham, D., Benaim, E., and Sorrentino, B.P. (1999). Effects of retroviral-mediated MDR1 expression on hematopoietic stem cell self-renewal and differentiation in culture. *Ann. NY Acad. Sci.* 872, 125–140.

Buonamici, S., Li, D., Chi, Y., Zhao, R., Wang, X., Brace, L., Ni, H., Saunthararajah, Y., and Nucifora, G. (2004). EVI1 induces myelodysplastic syndrome in mice. *J. Clin. Invest.* 114, 713–719.

Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S., and Hoffmann, C. (2005). Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol.* 3, 848–858.

Butler, S.L., Hansen, M.S., and Bushman, F.D. (2001). A quantitative assay for HIV DNA integration in vivo. *Nat. Med.* 7, 631–634.

Calmels, B., Ferguson, C., Laukkanen, M.O., Adler, R., Faulhaber, M., Kim, H.J., Sellers, S., Hematti, P., Schmidt, M., von Kalle, C., Akagi, K., Donahue, R.E., and Dunbar, C.E. (2005). Recurrent retroviral vector integration at the Mds1/Evi1 locus in nonhuman primate hematopoietic cells. *Blood* 106, 2530–2533.

Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L., Bousso, P., Deist, F.L., and Fischer, A. (2000). Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* 288, 669–672.

Cavazzana-Calvo, M., and Fischer, A. (2007). Gene therapy for severe combined immunodeficiency: are we there yet? *J. Clin. Invest.* 117, 1456–1465.

Deichmann, A., Hacein-Bey-Abina, S., Schmidt, M., Garrigue, A., Brugman, M.H., Hu, J., Glimm, H., Gyapay, G., Prum, B., Fraser, C.C., Fischer, N., Schwarzwaelder, K., Siegler, M.L., de Ridder, D., Pike-Overzet, K., Howe, S.J., Thrasher, A.J., Wagemaker, G., Abel, U., Staal, F.J., Delabesse, E., Villeval, J.L., Aronow, B., Hue, C., Prinz, C., Wissler, M., Klanke, C., Weissenbach, J., Alexander, I., Fischer, A., von Kalle, C., and Cavazzana-Calvo, M. (2007). Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. *J. Clin. Invest.* 117, 2225–2232.

Dick, J.E., Magli, M.C., Huszar, D., Phillips, R.A., and Bernstein, A. (1985). Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of W/Wv mice. *Cell* 42, 71–79.

Drexler, H.G. (1996). Expression of FLT3 receptor and response to FLT3 ligand by leukemic cells. *Leukemia* 10, 588–599.

Du, Y., Jenkins, N.A., and Copeland, N.G. (2005). Insertional mutagenesis identifies genes that promote the immortalization of primary bone marrow progenitor cells. *Blood* 106, 3932–3939.

Eckert, H.G., Stockschlader, M., Just, U., Hegewisch-Becker, S., Grez, M., Uhde, A., Zander, A., Ostertag, W., and Baum, C. (1996). High-dose multidrug resistance in primary human hematopoietic progenitor cells transduced with optimized retroviral vectors. *Blood* 88, 3407–3415.

Edelstein, M.L., Abedi, M.R., and Wixon, J. (2007). Gene therapy clinical trials worldwide to 2007 – an update. *J. Gene Med.* 9, 833–842.

Fischer, A., and Cavazzano-Calvo, M. (2005). Integration of retroviruses: a fine balance between efficiency and danger. *PLoS Med.* 2, 22–24.

Fruehauf, S., Veldwijk, M.R., Berlinghoff, S., Basara, N., Baum, C., Flasshove, M., Hegewisch-Becker, S., Kroger, N., Licht, T., Moritz, T., Hengge, U.R., Zeller, W.J., and Laufs, S. (2002). Gene therapy for sarcoma. *Cells Tissues Organs* 172, 133–144.

Gaspar, H.B., Bjorkegren, E., Parsley, K., Gilmour, K.C., King, D., Sinclair, J., Zhang, F., Giannakopoulos, A., Adams, S., Fairbanks, L.D., Gaspar, J., Henderson, L., Xu-Bayford, J.H., Davies, E.G., Veys, P.A., Kinnon, C., and Thrasher, A.J. (2006). Successful reconstitution of immunity in ADA-SCID by stem cell gene therapy following cessation of PEG-ADA and use of mild preconditioning. *Mol. Ther.* 14, 505–513.

Gaspar, H.B., Parsley, K.L., Howe, S., King, D., Gilmour, K.C., Sinclair, J., Brouns, G., Schmidt, M., Von Kalle, C., Barington, T., Jakobsen, M.A., Christensen, H.O., Al Ghonaium, A., White, H.N., Smith, J.L., Levinsky, R.J., Ali, R.R., Kinnon, C., and Thrasher, A.J. (2004). Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet* 364, 2181–2187.

Gatlin, J., Melkus, M.W., Padgett, A., Kelly, P.F., and Garcia, J.V. (2001). Engraftment of NOD/SCID mice with human CD34(+) cells transduced by concentrated oncoretroviral vector particles pseudotyped with the feline endogenous retrovirus (RD114) envelope protein. *J. Virol.* 75, 9995–9999.

Gentner, B., Laufs, S., Nagy, K.Z., Zeller, W.J., and Fruehauf, S. (2002). Rapid detection of retroviral vector integration sites in colony-forming human peripheral blood progenitor cells using PCR with arbitrary primers. *Gene Ther.* 10, 789–794.

Giordano, F.A., Hotz-Wagenblatt, A., Lauterborn, D., Appelt, J.U., Fellenberg, K., Nagy, K.Z., Zeller, J.W., Suhai, S., Fruehauf, S., and Laufs, S. (2007). New bioinformatic strategies to rapidly characterize retroviral integration sites of gene therapy. *Methods Inf. Med.* 46, 542–547.

Grez, M., Becker, S., Saulnier, S., Knoss, H., Ott, M.G., Maurer, A., Dinauer, M.C., Hoelzer, D., Seger, R., and Hossle, J.P. (2000). Gene therapy of chronic granulomatous disease. *Bone Marrow Transplant.* 25(Suppl. 2), S99–S104.

Guenechea, G., Gan, O.I., Dorrell, C., and Dick, J.E. (2001). Distinct classes of human stem cells that differ in proliferative and self-renewal potential. *Nat. Immunol.* 2, 75–82.

Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.L., Fraser, C.C., Cavazzana-Calvo, M., and Fischer, A. (2003). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* 348, 255–256.

Halene, S., and Kohn, D. (2000). Gene therapy using hematopoietic stem cells: Sisyphus approaches the crest. *Hum. Gene Ther.* 11, 1259–1267.

Hanenberg, H., Xiao, X.L., Dilloo, D., Hashino, K., Kato, I., and Williams, D.A. (1996). Colocalization of retrovirus and target cells on specific fibronectin fragments increases genetic transduction of mammalian cells. *Nat. Med.* 2, 876–882.

Harkey, M.A., Kaul, R., Jacobs, M.A., Kurre, P., Bovee, D., Levy, R., and Blau, C.A. (2007). Multiarm high-throughput integration site detection: limitations of LAM-PCR technology and optimization for clonal analysis. *Stem Cells Dev.* 16, 381–392.

Hematti, P., Hong, B.K., Ferguson, C., Adler, R., Hanawa, H., Sellers, S., Holt, I.E., Eckfeldt, C.E., Sharma, Y., Schmidt, M., von Kalle, C.,

Persons, D.A., Billings, E.M., Verfaillie, C.M., Nienhuis, A.W., Wolfsberg, T.G., Dunbar, C.E., and Calmels, B. (2004). Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol.* 2, e423.

Hildinger, M., Eckert, H.G., Schilz, A.J., John, J., Ostertag, W., and Baum, C. (1998). FMEV vectors: both retroviral long terminal repeat and leader are important for high expression in transduced hematopoietic cells. *Hum. Gene Ther.* 5, 1575–1579.

Hildinger, M., Schilz, A., Eckert, H.G., Bohn, W., Fehse, B., Zander, A., Ostertag, W., and Baum, C. (1999). Bicistronic retroviral vectors for combining myeloprotection with cell-surface marking. *Gene Ther.* 6, 1222–1230.

Holmes-Son, M.L., Appa, R.S., and Chow, S.A. (2001). Molecular genetics and target site specificity of retroviral integration. *Adv. Genet.* 43, 33–69.

Jordan, C.T., and Lemischka, I.R. (1990). Clonal and systemic analysis of hematopoiesis in the mouse. *Genes Dev.* 4, 220–232.

Kelly, P.F., Vandergriff, J., Nathwani, A., Nienhuis, A.W., and Vanin, E.F. (2000). Highly efficient gene transfer into cord blood nonobese diabetic/severe combined immunodeficiency repopulating cells by oncoretroviral vector particles pseudotyped with the feline endogenous retrovirus (RD114) envelope protein. *Blood* 96, 1206–1214.

Kiem, H.P., Andrews, R.G., Morris, J., Peterson, L., Heyward, S., Allen, J.M., Rasko, J.E., Potter, J., and Miller, A.D. (1998). Improved gene transfer into baboon marrow repopulating cells using recombinant human fibronectin fragment CH-296 in combination with interleukin-6, stem cell factor, FLT-3 ligand, and megakaryocyte growth and development factor. *Blood* 92, 1878–1886.

Kiem, H.P., Sellers, S., Thomasson, B., Morris, J.C., Tisdale, J.F., Horn, P.A., Hematti, P., Adler, R., Kuramoto, K., Calmels, B., Bonifacino, A., Hu, J., von Kalle, C., Schmidt, M., Sorrentino, B., Nienhuis, A., Blau, C.A., Andrews, R.G., Donahue, R.E., and Dunbar, C.E. (2004). Long-term clinical and molecular follow-up of large animals receiving retrovirally transduced stem and progenitor cells: no progression to clonal hematopoiesis or leukemia. *Mol. Ther.* 9, 389–395.

Kim, H.J., Tisdale, J.F., Wu, T., Takatoku, M., Sellers, S.E., Zickler, P., Metzger, M.E., Agricola, B.A., Malley, J.D., Kato, I., Donahue, R.E., Brown, K.E., and Dunbar, C.E. (2000). Many multipotential gene-marked progenitor or stem cell clones contribute to hematopoiesis in nonhuman primates. *Blood* 96, 1–8.

Kitamura, Y., Lee, Y.M., and Coffin, J.M. (1992). Nonrandom integration of retroviral DNA in vitro: effect of CpG methylation. *Proc. Natl. Acad. Sci. USA* 89, 5532–5536.

Kohn, D.B. (2002). Adenosine deaminase gene therapy protocol revisited. *Mol. Ther.* 5, 96–97.

Kohn, D.B., Sadelain, M., Dunbar, C.E., Bodine, D., Kiem, H.P., Candotti, F., Tisdale, J., Riviere, I., Blau, C.A., Richard, R.E., Sorrentino, B., Nolta, J., Malech, H., Brenner, M., Cornetta, K., Cavagnaro, J., High, K., and Glorioso, J. (2003). American Society of Gene Therapy (ASGT) ad hoc subcommittee on retroviral-mediated gene transfer to hematopoietic stem cells. *Mol. Ther.* 8, 180–187.

Kustikova, O., Fehse, B., Modlich, U., Yang, M., Dullmann, J., Kamino, K., von Neuhoff, N., Schlegelberger, B., Li, Z., and Baum, C. (2005). Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. *Science* 308, 1171–1174.

Laskus, T., Radkowski, M., Wang, L.F., Nowicki, M., and Rakela, J. (1999). Detection and sequence analysis of hepatitis B virus integration in peripheral blood mononuclear cells. *J. Virol.* 73, 1235–1238.

Laufs, S., Gentner, B., Nagy, K.Z., Jauch, A., Benner, A., Naundorf, S., Kuehlcke, K., Schiedlmeier, B., Ho, A.D., Zeller, W.J., and Fruehauf, S. (2003). Retroviral vector integration occurs into preferred genomic targets of human hematopoietic stem cells with long-term bone marrow repopulating ability. *Blood* 101, 2191–2198.

Laufs, S., Nagy, K.Z., Giordano, F.A., Hotz-Wagenblatt, A., Zeller, W.J., and Fruehauf, S. (2004). Insertion of retroviral vectors in NOD/SCID repopulating human peripheral blood progenitor cells occurs preferentially in the vicinity of transcription start regions and in introns. *Mol. Ther.* 10, 874–881.

Li, Z., Dullmann, J., Schiedlmeier, B., Schmidt, M., von Kalle, C., Meyer, J., Forster, M., Stocking, C., Wahlers, A., Frank, O., Ostertag, W., Kuhlcke, K., Eckert, H.G., Fehse, B., and Baum, C. (2002). Murine leukemia induced by retroviral gene marking. *Science* 296, 497.

Lichter, P., and Cremer, T. (1992). Chromosome analysis by non-isotopic in situ hybridization. In *Human Genetics – A Practical Approach* (Rooney, D.E., and Czepulkowski, B.H., eds), pp. 157–192. IRL Press, Oxford, United Kingdom.

Marshall, E. (2003). Second child in French trial is found to have leukemia. *Science* 299, 320.

Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* 2, E234.

Modlich, U., Kustikova, O.S., Schmidt, M., Rudolph, C., Meyer, J., Li, Z., Kamino, K., von Neuhoff, N., Schlegelberger, B., Kuehlcke, K., Bunting, K.D., Schmidt, S., Deichmann, A., von Kalle, C., Fehse, B., and Baum, C. (2005). Leukemias following retroviral transfer of multidrug resistance 1 (MDR1) are driven by combinatorial insertional mutagenesis. *Blood* 105, 4235–4246.

Moritz, T., Dutt, P., Xiao, X., Carstanjen, D., Vik, T., Hanenberg, H., and Williams, D.A. (1996). Fibronectin improves transduction of reconstituting hematopoietic stem cells by retroviral vectors: evidence of direct viral binding to chymotryptic carboxy-terminal fragments. *Blood* 88, 855–862.

Moritz, T., and Williams, D.A. (1998). Methods of gene transfer: genetic manipulation of hematopoietic stem cells. In *Hematopoietic Cell Transplantation* (Thomas, E.D., Blume, K., and Forman, S.J., eds), pp. 79–86. Blackwell Science.

Movassagh, M., Desmyter, C., Baillou, C., Chapel-Fernandes, S., Guigon, M., Klatzmann, D., and Lemoine, F.M. (1998). High-level gene transfer to cord blood progenitors using gibbon ape leukemia virus pseudotype retroviral vectors and an improved clinically applicable protocol. *Hum. Gene Ther.* 9, 225–234.

Müller, S., Neusser, M., and Wienberg, J. (2002). Towards unlimited colors for fluorescence in situ hybridization (FISH). *Chromosome Res.* 10, 223–232.

Nagy, K.Z., Laufs, S., Gentner, B., Naundorf, S., Kuehlcke, K., Topaly, J., Buss, E.C., Zeller, W.J., and Fruehauf, S. (2004). Clonal analysis of individual marrow-repopulating cells after experimental peripheral blood progenitor cell transplantation. *Stem Cells* 22, 570–579.

Nienhuis, A.W., Dunbar, C.E., and Sorrentino, B.P. (2006). Genotoxicity of retroviral integration in hematopoietic cells. *Mol. Ther.* 13, 1031–1349.

Nolta, J.A., Dao, M.A., Wells, S., Smogorzewska, E.M., and Kohn, D.B. (1996). Transduction of pluripotent human hematopoietic stem cells demonstrated by clonal analysis after engraftment in immune-deficient mice. *Proc. Natl. Acad. Sci. USA* 93, 2414–2419.

Ott, M.G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U., Glimm, H., Kuhlcke, K., Schilz, A., Kunkel, H., Naundorf, S., Brinkmann, A., Deichmann, A., Fischer, M., Ball, C., Pilz, I., Dunbar, C., Du, Y., Jenkins, N.A., Copeland, N.G., Luthi, U., Hassan, M., Thrasher, A.J., Hoelzer, D., von Kalle, C., Seger, R., and Grez, M. (2006). Correction of X-linked chronic granulomatous disease by

gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat. Med.* 12, 401–409.

Petzer, A.L., Hogge, D.E., Landsdorp, P.M., Reid, D.S., and Eaves, C.J. (1996). Self-renewal of primitive human hematopoietic cells (long-term-culture-initiating cells) in vitro and their expansion in defined medium. *Proc. Natl. Acad. Sci. USA* 93, 1470–1474.

Porter, C.D., Collins, M.K., Tailor, C.S., Parkar, M.H., Cosset, F.L., Weiss, R.A., and Takeuchi, Y. (1996). Comparison of efficiency of infection of human gene therapy target cells via four different retroviral receptors. *Hum. Gene Ther.* 7, 913–919.

Rosenthal, A., and Jones, D.S. (1990). Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction. *Nucleic Acids Res.* 18, 3095–3096.

Sanders, D.A. (2002). No false start for novel pseudotyped vectors. *Curr. Opin. Biotechnol.* 13, 437–442.

Schiedlmeier, B., Kuehlcke, K., Eckert, H.G., Baum, C., Zeller, W.J., and Fruehauf, S. (2000). Quantitative assessment of retroviral transfer of the human multidrug resistance 1 gene to human mobilized peripheral blood progenitor cells engrafted in nonobese diabetic/severe combined immunodeficient mice. *Blood* 95, 1237–1248.

Schmidt, M., Hoffmann, G., Wissler, M., Lemke, N., Muessig, A., Glimm, H., Williams, D.A., Ragg, S., Hesemann, C.U., and von Kalle, C. (2001). Detection and direct genomic sequencing of multiple rare unknown flanking DNA in highly complex samples. *Hum. Gene Ther.* 12, 743–749.

Schmidt, M., Zickler, P., Hoffmann, G., Haas, S., Wissler, M., Muessig, A., Tisdale, J.F., Kuramoto, K., Andrews, R.G., Wu, T., Kiem, H.P., Dunbar, C.E., and von Kalle, C. (2002). Polyclonal long-term repopulating stem cell clones in a primate model. *Blood* 100, 2737–2743.

Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110, 521–529.

Schwarzwaelder, K., Howe, S.J., Schmidt, M., Brugman, M.H., Deichmann, A., Glimm, H., Schmidt, S., Prinz, C., Wissler, M., King, D.J., Zhang, F., Parsley, K.L., Gilmour, K.C., Sinclair, J., Bayford, J., Peraj, R., Pike-Overzet, K., Staal, F.J., de Ridder, D., Kinnon, C., Abel, U., Wagemaker, G., Gaspar, H.B., Thrasher, A.J., and von Kalle, C. (2007). Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo. *J. Clin. Invest.* 117, 2241–2249.

Seggewiss, R., Pittaluga, S., Adler, R.L., Guenaga, F.J., Ferguson, C., Pilz, I.H., Ryu, B., Sorrentino, B.P., Young, W.S., 3rd, Donahue, R.E., von Kalle, C., Nienhuis, A.W., and Dunbar, C.E. (2006). Acute myeloid leukemia is associated with retroviral gene transfer to hematopoietic progenitor cells in a rhesus macaque. *Blood* 107, 3865–3867.

Sellers, S.E., Tisdale, J.F., Agricola, B.A., Metzger, M.E., Donahue, R.E., Dunbar, C.E., and Sorrentino, B.P. (2001). The effect of multidrug-resistance 1 gene versus neo transduction on ex vivo and in vivo expansion of rhesus macaque hematopoietic repopulating cells. *Blood* 97, 1888–1891.

Shah, A.J., Smogorzewska, E.M., Hannum, C., and Crooks, G.M. (1996). Flt3 ligand induces proliferation of quiescent human bone marrow CD34+CD38-cells and maintains progenitor cells in vitro. *Blood* 87, 3563–3570.

Silver, J., and Keerikatte, V. (1989). Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. *J. Virol.* 63, 1924–1928.

Sørensen, A.B., Duch, M., Jorgensen, P., and Pedersen, F.S. (1993). Amplification and sequence analysis of DNA flanking integrated proviruses by a simple two-step polymerase chain reaction method. *J. Virol.* 67, 7118–7124.

Sørensen, A.B., Duch, M., Amtoft, H.W., Jorgensen, P., and Pedersen, F.S. (1996). Sequence tags of provirus integration sites in DNAs of tumors induced by the murine retrovirus SL3-3. *J. Virol.* 70, 4063–4070.

Speicher, M.R., Gwyn Ballard, S., and Ward, D.C. (1996). Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat. Genet.* 12, 368–375.

Stocking, C., Bergholz, U., Friel, J., Klingler, K., Wagener, T., Starke, C., Kitamura, T., Miyajima, A., and Ostertag, W. (1993). Distinct classes of factor-independent mutants can be isolated after retroviral mutagenesis of a human myeloid stem cell line. *Growth Factors* 8, 197–209.

Telenius, H., Carter, N.P., Bebb, C.E., Nordenskjold, M., Ponder, B.A., and Tunnacliffe, A. (1992). Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13, 718–725.

Weidhaas, J.B., Angelichio, E.L., Fenner, S., and Coffin, J.M. (2000). Relationship between retroviral DNA integration and gene expression. *J. Virol.* 74, 8382–8389.

Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300, 1749–1751.

Wu, X., Li, Y., Crise, B., Burgess, S.M., and Munroe, D.J. (2005). Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.* 79, 5211–5214.

Chapter 33

Preimplantation Genetic Diagnosis

Martine de Rycke and Karen Sermon
Centre for Medical Genetics, UniversitairZiekenhuis Brussel, Department of Embryology and Genetics, Vrije Universiteit Brussel, Laarbeeklaan 101, 1090 Brussels, Belgium

33.1 WHAT IS PREIMPLANTATION GENETIC DIAGNOSIS?

Preimplantation genetic diagnosis (PGD) is an alternative to prenatal diagnosis. Oocytes or preimplantation embryos are obtained *in vitro* and are genetically analyzed, after which only those embryos that are judged to be free of the genetic defect under consideration are transferred. To this end, polar bodies are removed from oocytes, blastomeres (at the cleavage stage), or trophectoderm cells (at the blastocyst stage) are removed from preimplantation embryos, and these cells are used for genetic diagnosis. The main advantage is the circumvention of therapeutic abortion for a genetic disease; the main disadvantage is the requirement of *in vitro* fertilization. The first report on PGD was by Handyside and coworkers (1990) and described the sexing of embryos in families who were at risk of X-linked disease. The sex of the embryos was established using polymerase chain reaction (PCR) of repetitive Y-sequences, and resulted in two pregnancies and the birth of healthy female babies.

Initially, PGD was developed to help couples at risk for monogenic diseases. The only technique available to analyze single cells (polar bodies or blastomeres) was PCR, which was used either to sex the embryos (Handyside *et al.*, 1990) or to detect specific mutations such as the p.F508del mutation in cystic fibrosis (Handyside *et al.*, 1992; Verlinsky *et al.*, 1992). Later, fluorescent *in situ* hybridization (FISH; see also Chapter 10) was downscaled to the single-cell level (Griffin *et al.*, 1992) and this opened possibilities to analyze embryos at the chromosomal level: again, either for sexing (Harper *et al.*, 1994) or for chromosomal aberrations, such as Robertsonian or reciprocal translocations (Conn *et al.*, 1999), or, finally, for aneuploidy screening (Verlinsky and Kuliev, 1996).

Guidelines for best practice have been drawn up by the European Society for Human Reproduction and Embryology (ESHRE) PGD Consortium and by the Preimplantation Genetic Diagnosis International Society (PGDIS) (Thornhill *et al.*, 2005; PGDIS, 2004, 2008).

33.2 INDICATIONS FOR PGD

33.2.1 Monogenic Diseases

Not surprisingly, most PGD cycles for monogenic diseases have been performed for the most frequent monogenic diseases. These are also more or less the monogenic diseases for which most prenatal diagnoses are performed. Cystic fibrosis (CF) was the first monogenic disease for which a specific diagnosis at the single-cell level was performed, analyzing the most frequent p.F508del mutation (Handyside *et al.*, 1992), and several reports using slightly different approaches have been published since (Verlinsky *et al.*, 1992; Liu *et al.*, 1994; Moutou and Viville, 1999; Goossens *et al.*, 2000). More recently, multiplex approaches designed for analysis of single cells, and analyzing a CF mutation with one or more linked markers (Strom *et al.*, 1998; Moutou *et al.*, 2002; Goossens *et al.*, 2003), or a panel of linked markers, have been described (Dreesen *et al.*, 2000; Eftedal *et al.*, 2001; Vrettou *et al.*, 2002). Similarly, strategies have been developed for PGD for spinal muscular atrophy, both based on analysis of the most prevalent mutation (a large deletion encompassing several exons, Dreesen *et al.*, 1998; Fallon *et al.*, 1999; Moutou *et al.*, 2001) and on the multiplex analysis of the mutation and linked markers (Moutou *et al.*, 2003; Girardet *et al.*, 2007).

Mutations in the β-globin gene, both leading to sickle cell disease (SCD) and β-thalassemia, have received considerable attention from the PGD community, because of their high frequency in certain populations (Patrinos *et al.*, 2004). Because so many different mutations can be found in the β-globin gene (only one of them leading to SCD, the others leading to various types of β-thalassemia; Hardison *et al.*, 2002), the strategies developed for PGD are different, for example, from spinal muscular atrophy and aim to detect

as many different mutations as possible with the same assay. Examples of this are given by Xu and coworkers (1999), on PGD for SCD; Vrettou and coworkers (1999), using DGGE for PGD for β-thalassemia; Kuliev and coworkers (1999), on polar body biopsy for PGD for β-thalassemia; De Rycke and coworkers (2001), on PGD for SCD and β-thalassemia using fluorescent PCR; Jiao and coworkers (2003), using primer-extension preamplification (PEP), followed by nested PCR and reverse dot-blot for PGD for β-thalassemia; Vrettou and coworkers (2004) on real-time PCR for PGD for SCD and β-thalassemia; and the more recent multiple displacement amplification technique by Hellani and coworkers (2004).

Patients at risk of transmitting an autosomal dominant disease have always been particularly interested in PGD because of the high recurrence risk. Examples of this abound in the literature, and again the more frequent diseases have received the most attention. The authors (Sermon *et al.*, 1999a; Spits *et al.*, 2006a) and others (Harton *et al.*, 1996; Blaszczyk *et al.*, 1998; Lledó *et al.*, 2006) have described several different approaches for Marfan's disease. PGD for diseases caused by dynamic mutations have also been reported. Noteworthy here are myotonic dystrophy type 1 (DM1; Sermon *et al.*, 1998a, 2001; Piyamongkol *et al*., 2001; Dean *et al.*, 2001; Kakourou *et al.*, 2008) and Huntington's disease (Sermon *et al.*, 1998a, 2001; Stern *et al.*, 2002; Moutou *et al.*, 2004).

Sexing with FISH has been most frequently used for X-linked diseases such as Duchenne's muscular dystrophy, hemophilia A and B, retinitis pigmentosa, and others (Staessen *et al.*, 1999), but more and more specific DNA diagnoses have been developed and used. The advantages of a specific DNA diagnosis are important: first, healthy male embryos are not discarded and second, female carriers can be identified and, according to the patient's wishes and the center's policy, are then not selected for transfer. Examples of tests developed for the specific diagnosis of an X-linked disease are given by Hussey and coworkers (1999), Ray and coworkers (2001), and Girardet and coworkers (2003) for Duchenne's muscular dystrophy. Sermon and coworkers (1999b) and Apessos and coworkers (2001) described protocols for PGD for Fragile-X syndrome. Because girls who are carriers of Fragile-X can be affected, only a specific DNA analysis can be used for PGD.

PGD is currently available for a large number of monogenic diseases and has lately also been applied for indications for which prenatal diagnosis is ethically difficult, such as late-onset diseases and (cancer) predisposition syndromes (Verlinsky *et al.*, 2001; Spits *et al.*, 2007). Another more recent indication involves human leukocyte antigen (HLA) typing of preimplantation embryos to select an embryo that is HLA compatible with an affected sibling. At birth, hematopoietic stem cells, collected from the umbilical cord blood, can be used to transplant to the sick child. HLA typing is carried out in cases of leukemias and anemias or it is combined with the detection of mutations underlying immunodeficiencies and hemoglobin disorders (Kuliev *et al.*, 2005; Verlinsky *et al.*, 2007).

33.2.2 Chromosomal Aberrations

Reciprocal translocations are characterized by the exchange of fragments between chromosomes, whereas in Robertsonian translocations a whole acrocentric chromosome is translocated to another one through centromeric fusion. Carriers of these balanced translocations have a normal phenotype but are at risk of having children with congenital anomalies and mental retardation due to chromosomal imbalances or more frequently suffer from recurrent miscarriages or infertility (especially if the male is a carrier). This explains the large interest this group of patients has shown for PGD. The first reports (Munné *et al.*, 1998; Iwarsson *et al.*, 1998) described the use of probes that were custom designed for one specific translocation usually occurring in only one family, which is why the application of this approach remained limited. However, it was only since the widespread availability of fluorescent probes in different colors (Conn *et al.*, 1999; Van Assche *et al.*, 1999; Coonen *et al.*, 2000) that it has been possible to propose PGD to these patients in more than a handful of highly specialized centers. Examples of reports on larger series are those by Munné and coworkers (2000), Durban and coworkers (2001), and Pickering and coworkers (2003).

33.2.3 Aneuploidy Screening

It has been a well-established fact that human embryos carry cytogenetic abnormalities in high proportions: using classical karyotyping to investigate embryos, between 23% and 80% of embryos were found to be aneuploid, the latter number found in embryos of poor quality (Zenzes and Casper, 1992; Pellestor *et al.*, 1994). More detailed information has become available after the advent of FISH. Earlier reports by Delhanty and coworkers (1993), Munné and coworkers (1993), and Harper and coworkers (1995) showed a whole range of abnormalities, such as monosomies, trisomies, triploidies, and combined abnormalities. These authors also reported that 70% of the embryos analyzed were developing abnormally, even if only five chromosomes (X, Y, 13, 18, and 21) were analyzed.

In another report on a large number of non-viable cleavage stage embryos, Marquez and coworkers (2000) could show that aneuploidy (from 1.4% in patients between 20 and 34 years to 52.4% in patients between 40 and 47) increases with maternal age, and polyploidy and mosaicism is related to poor embryo morphology. Several authors suggested that, considering the high rate of abnormalities in preimplantation embryos, together with the higher risk for fetal aneuploidy at an advanced maternal age and the fact

that 50% to 60% of all spontaneous abortions from clinically recognized pregnancies carry an abnormal karyotype (Boué *et al.*, 1985), embryo selection based on chromosome complement would improve *in vitro* fertilization (IVF) results in groups of patients with poor outcome, as well as avoid the birth of babies with chromosomal defects. The obvious patient groups or cases for whom PGD-aneuploidy screening (PGD-AS) could be beneficial are those with advanced maternal age (AMA) (Verlinsky *et al.*, 1999; Munné *et al.*, 2003a), repetitive implantation failure after IVF (Gianaroli *et al.*, 1999), and recurrent miscarriage not due to translocations (Wilton, 2002; Rubio *et al.*, 2003).

Comparative genomic hybridization (CGH, section 33.3.4.2; see also Chapter 12) (Voullaire *et al.*, 2000; Wells and Delhanty, 2000; Wilton *et al.*, 2001; Wells *et al.*, 2002) has been used in PGD-AS: the important advantage of CGH over FISH is that a whole karyotype is obtained. In this way, abnormalities are found in embryos, which would have been missed by FISH. Wilton and coworkers (2003) estimate that FISH for five or nine chromosomes would have missed 38% and 25% of the abnormal blastomeres, respectively. However, the complexity of the CGH, as well as the time currently needed to obtain a karyotype (five days), explains why for the time being CGH is not as widely applied as FISH. PGD, combined with aneuploidy screening (PGD-AS) using FISH, has now become widely applied in the patient groups mentioned earlier, because of the relative ease of the technique and the large potential patient group. However, the evaluation of the benefit of these treatments awaits the results of large prospective randomized controlled trials (RCT), because the PGD-AS data in these studies are not compared to a suitable control group (Wilton, 2002). Three RCTs have been performed for AMA and found no differences or even lower live birth rates when comparing PGS-treated patients with control IVF patients (Staessen *et al.*, 2004; Stevens *et al.*, 2004; Mastenbroek *et al.*, 2007). But then again, a flaw in design or execution could be brought forward for each of these studies. Another recent RCT of blastocyst biopsy followed by PGS using 5-color FISH versus no biopsy in young and infertile women showed no difference in live birth rate (Jansen *et al.*, 2008). Large multicenter, randomized well-designed and well-executed studies are currently set up and will allow future evaluation of the efficiency of PGD-AS and delineation of patient groups who will most benefit from PGD-AS (Harper *et al.*, 2008b).

33.3 TECHNOLOGIES USED IN PGD

33.3.1 Assisted Reproductive Technology

The assisted reproductive technology (ART) used to obtain embryos for PGD will only briefly be outlined, as it lies mostly outside the scope of this chapter.

The first step in a PGD cycle, as in many ART, is controlled ovarian hyperstimulation (COH), aimed at obtaining a large cohort of mature oocytes. Although there is a tendency in regular IVF to apply milder stimulation protocols to reduce the number of oocytes in order to obtain better quality oocytes and to prevent multiple pregnancies (Macklon and Fauser, 2000; Baart *et al.*, 2007), in PGD a large number of oocytes are still required because many embryos will be deemed untransferable because they are affected or carry important aneuploidies (Vandervorst *et al.*, 1998). COH is carried out in an agonist protocol, using gonadotrophin releasing hormone analogs combined with human menopausal gonadotrophins (hMG) or recombinant folicle stimulating hormone (FSH), or an antagonist protocol using recombinant FSH combined with GnRHa according to the patient's profile (age, body mass index (BMI)) and/or previous response to ovarian stimulation. When blood hormone and ultrasound monitoring show that a sufficient number of mature oocytes are present, an injection of human chorionic gonadotrophin (hCG) is given to induce ovulation (Kolibianakis *et al.*, 2004). Follicles can then be punctured transvaginally and under ultrasound guidance, and oocytes are aspirated.

Intracytoplasmic sperm injection (ICSI) is usually preferred over regular IVF, regardless of the sperm quality, because the presence of sperm stuck to the zona pellucida (ZP) after IVF represents an important source of contamination when PCR is used for diagnosis (Liebaers *et al.*, 1998). The aspirated oocytes are first denuded of the surrounding granulosa cells, either through enzymatic (hyaluronidase) or mechanical (aspiration in and out of a pipette with a small diameter) means, or a combination of both. Concurrently, sperm is washed and prepared for ICSI. During ICSI, the oocyte is held in a holding pipette at nine o'clock, while the sperm is injected with a fine injection pipette at three o'clock (Joris *et al.*, 1998).

After about 16 hours in culture, normal fertilization can be observed if two pronuclei and two polar bodies are present. Nowadays, embryo culture media are complex systems, with special formulations adapted for each step in the embryo's development. Fertilization, early development (up to day 3), and development to blastocyst are performed in different formulations. Several of these so-called sequential media are now commercially available (Schoolcraft and Gardner, 2001).

33.3.2 Biopsy Techniques: Pros and Cons

33.3.2.1 Polar Body Biopsy

Just before fertilization, a normal oocyte is at the metaphase II stage of the meiosis; that is, it has extruded the first polar body (PBI). The second polar body (PBII) is then extruded after normal fertilization. These two PBs have no further function in the embryonic development, and can

thus be retrieved for analysis. The genetic content of the oocyte is the mirror image of the genetic content of the polar bodies and can thus indirectly be deduced (Verlinsky *et al.*, 1992, 1999). The advantages of PB biopsy are self-evident: the embryo proper is not disturbed and there are thus no detrimental effects of the decrease of embryonic mass as in blastomere biopsy. Ethics is another important argument. Indeed, it is possible to biopsy and analyze the two PBs before syngamy of the male and female pronucleus, before what is legally regarded in several countries (e.g. Germany) as the beginning of life. A third advantage advanced by the proponents of PB biopsy, but which has failed to convince the authors, is that the difficulties raised by mosaicism in the embryo are avoided. The validity of this argument depends on the importance imparted on the presence of mosaicism. The important disadvantages have led most centers to prefer cleavage stage embryo biopsy. First, only the maternal contribution is analyzed, thus the technique is not applicable in autosomal dominant diseases or translocations where the father is the carrier. It is also not applicable for sexing. Second, if the analysis is to be finished before syngamy, it leaves very little time to complete the diagnosis. Conversely, if ethics are not an issue, more time is available for analysis than in cleavage stage biopsy (see section 33.3.2.2).

Technically, PB biopsy is quite straightforward. The ZP is breached using either mechanical slitting with a fine needle or laser technology. The chemical breaching of the ZP is not a valid alternative as the Acidic Tyrode's solution used for this purpose damages the oocyte. After hatching, a small diameter pipette is introduced into the hole and the two polar bodies are removed (Verlinsky *et al.*, 1999). If a FISH analysis is to be performed (for translocations or PGD-AS), then the two PBs can be removed in one operation because FISH analysis will reveal which cell is PBI and which is PBII. However, if analysis at the DNA level is carried out, then it is mandatory to biopsy the PBI before fertilization and the PBII after fertilization, to be able to follow the segregation of the respective alleles in oocyte and PBs (Rechitsky *et al.*, 1999).

33.3.2.2 Cleavage Stage Biopsy

Cleavage stage biopsy is the most widely spread technique for obtaining embryonic material for PGD (Sermon *et al.*, 2005). One or two blastomeres are biopsied on the morning of day 3 when the embryo is normally at the eight-cell stage. Since human embryos have a quite erratic cleavage pattern, some embryos will be arrested at the four-cell stage, whereas others will be retarded and will have reached only the six-cell stage, and still others will have a lead and will have reached the ten-cell stage. Compaction has not occurred yet in most embryos and it is assumed that all cells are still totipotent. It is still not known whether and to what extent the random removal of one or two nucleated cells harms embryonic development and implantation potential. Some studies seem to indicate that good quality embryos easily recover from the removal of one quarter of their cell mass (Van de Velde *et al.*, 2000; Parriego *et al.*, 2003). Goossens and coworkers (2008) carried out a prospective randomized study assessing the influence of one-cell biopsy versus two-cell biopsy on diagnostic efficiency and accuracy as well as on embryonic development and live birth rate. The diagnostic efficiency for FISH PGD cycles was similar in the one-cell and the two-cell biopsy group, but for PCR PGD cycles, a significant lower diagnostic efficiency was demonstrated after one cell removal. However, this did not influence the transfer rates; an important reason being that the majority of embryos remaining without diagnosis were embryos of inferior quality (embryo quality as based on morphology), yielding failure in PCR amplification and with limited contribution to a good quality embryo transfer (Goossens *et al.*, 2008). The removal of one blastomere at day 3 was less invasive than two-cell removal with respect to day 5 embryonic developmental competence. However, the day 3 developmental stage was a stronger predictor for further *in vitro* development. In other words, a good quality embryo on day 3 has a higher chance of reaching the blastocyst stage on day 5, while a six-cell embryo on day 3 mainly arrests or becomes a moderate or poor quality embryo on day 5. Despite significant differences for post-biopsy *in vitro* development, the data show that the live birth rate per started cycle is not significantly different after two-cell biopsy than after one-cell biopsy. The data reveal an absolute difference in live birth rate of 3%. Taken together with their findings that for PCR the false positive rate was not higher in the case of one-cell biopsy and that for the FISH cycles the false positive rate was not significantly different if the diagnosis was based on one-cell versus two-cell analysis (Michiels *et al.*, 2006), the authors recommended removal of one cell whenever an accurate diagnosis method is in place (Goossens *et al.*, 2008).

Several techniques have been used to breach the ZP and to make the blastomeres accessible. The ZP can be opened mechanically, as for PB biopsy. Usually, two perpendicular slits are made with a needle, giving rise to flaps in the ZP that can be lifted to allow the introduction of the biopsy pipette (Cieslak *et al.*, 1999). The chemical opening is another method. A thin stream of Acidic Tyrode's (AT) solution (pH 2.2) is applied with a drilling pipette of about 10–12 mm to the ZP to dissolve it. This technique requires some skill, as the AT can lyse the cells immediately under the ZP. Nowadays, the opening of the ZP using a non-contact diode infrared laser is the most widespread method. Two or exceptionally three pulses of 5–8 msec and with a wavelength of 1.48 mm are applied at a safe distance (more than 8 mm from the nearest blastomere; De Vos and Van Steirteghem, 2001). Joris and coworkers (2003) have found that the number of cells lysed after ZP hatching is reduced significantly using the laser as compared to AT, and that

the time needed to biopsy the cell(s) is also reduced significantly. Moreover, the laser does not have a detrimental effect on the further development of the embryos.

At about the eight-cell stage, embryos start compaction; that is, the cell-to-cell adhesions increase and the cell boundaries become indistinguishable, a process that is completed at the morula stage. Especially with the introduction of the use of sequential media, more embryos are at the compacted stage at the moment of biopsy and this can lead to important difficulties to retrieve one or two cells without cell lysis. It was shown by Dumoulin and coworkers (1998) that the use of Ca^{++} and Mg^{++}-free medium to decompact the embryos did not have a detrimental effect on the development of the embryo. This is why most centers performing PGD at the cleavage stage incubate the embryos in Ca^{++} and Mg^{++}-free medium, either only before biopsy, or during the whole biopsy procedure (ESHRE PGD Consortium Steering Committee, 2002).

For the biopsy itself, several techniques have been developed and applied with changing success. Biopsy by flow displacement (where a flow of fluid is applied through one opening in the ZP to release the blastomere through another hole) or extrusion (where pressure is applied to the embryo to push a blastomere through the opening in the ZP) is just mentioned for completeness. The most widely used method is described extensively in Joris and coworkers (2003). The embryo is held with a holding pipette (100 mm outer, 25–30 mm inner diameter) through the application of a slight negative pressure. A blunt biopsy pipette (inner diameter of 35–40 mm) is introduced through the hole in the ZP and the blastomere is gently aspirated. The blastomere is usually not aspirated completely into the pipette, but is "grabbed" partially and pulled through the hole in the ZP (see Fig. 33.1).

33.3.2.3 Blastocyst Biopsy

Trophectoderm (TE) biopsy is an emerging approach in PGD. The feasibility of trophectoderm biopsy was first shown by Dokras and coworkers (1990). The use of non-contact laser facilitated TE biopsy (Veiga *et al.*, 1997) and successful clinical applications followed, showing that this approach is compatible with implantation and normal development (de Boer *et al.*, 2004; Kokkali *et al.*, 2005). Preliminary comparisons even suggest that TE biopsy may yield substantially higher implantation and live birth rates than cleavage stage biopsy (McArthur *et al.*, 2005; Kokkali *et al.*, 2007). Many of the PGD cycles with TE biopsy have been published by the Sydney IVF center; their latest report further confirms that TE biopsy may be more advantageous than cleavage stage biopsy (McArthur *et al.*, 2008).

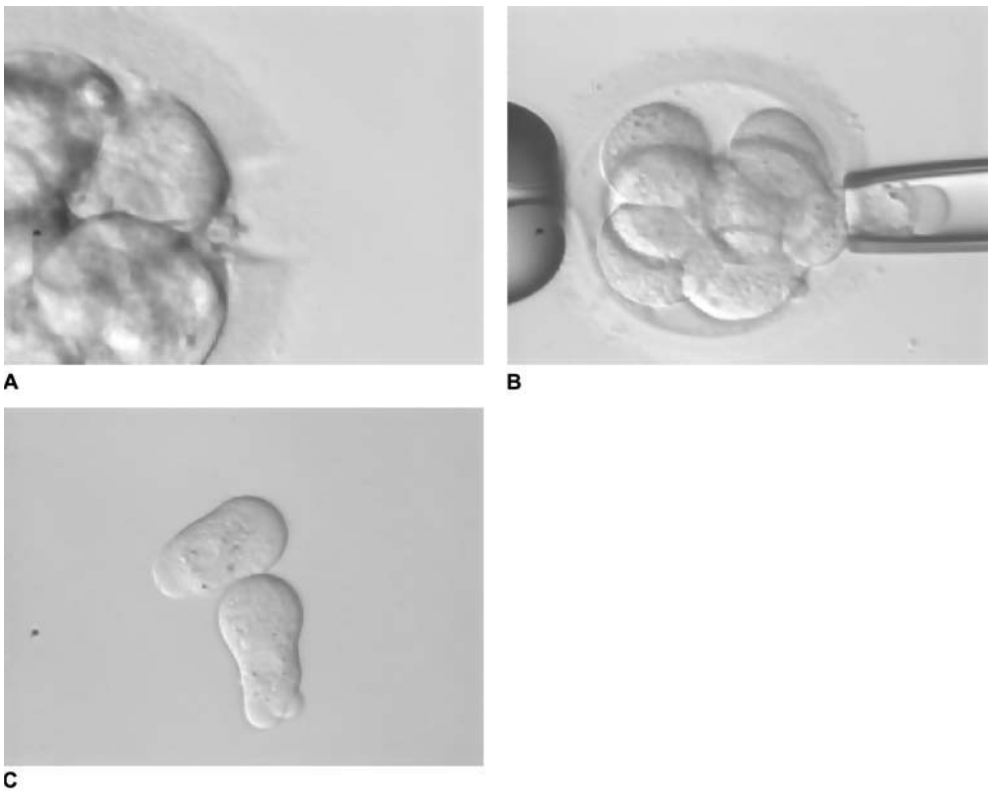


FIGURE 33.1 Example of an embryo biopsy. Panel **A.** Shows the hole in the ZP after application of the laser for ZP hatching. Panel **B.** Shows how one blastomere is gently aspirated outside the embryo. Panel **C.** Shows two biopsied blastomeres with a clear nucleus.

Technically, TE biopsy involves two steps: on day 3 or 4 an opening is made in the ZP by laser, directly opposite the ICM, allowing the TE to herniate from the zona at blastulation. At day 5 or 6, 4–20 herniating TE cells are aspirated in a biopsy pipette and dissected from the embryo by laser, leaving the ICM from which the fetus develops intact and removing only 10–20% of total blastocyst mass. This is an advantage compared to cleavage stage biopsy; another advantage of blastocyst biopsy compared to PB and cleavage stage biopsy is that multiple cells are available for genetic testing, which may improve diagnostic accuracy. The main disadvantage is the limited time left for genetic analysis and the fact that fewer blastocysts (one-half to one-third of cleavage stage embryos) can be biopsied, which may restrict the success rate of PGD for some couples.

33.3.3 PGD for Monogenic Diseases

33.3.3.1 Single-Cell PCR: Pitfalls and Their Solution

PCR, conceived by Kary Mullis in 1983, allows for the exponential amplification of a particular sequence from a DNA sample. PCR applications are widely used in molecular biology, including PGD where it can generate millions of copies of the region of interest starting from one or just two DNA molecules. Since the first report on PGD, the methods of analysis of the PCR products have evolved from electrophoresis on simple agarose gels (Handyside *et al.*, 1990), over fragment analysis on automated sequencers (Sermon *et al.*, 1998b), to minisequencing (Fiorentino *et al.*, 2003), and real-time PCR (Rice *et al.*, 2002). Although the refinement of the analysis methods has increased the efficiency and accuracy tremendously, there are still a number of pitfalls that are inherent to single-cell PCR and that still cause incorrect diagnosis. These are:

- The (lack of) specificity of the PCR,
- Contamination with DNA extraneous to the analyzed cell, and
- Allele drop-out

33.3.3.1.1 Specificity of the PCR

From one or two DNA template molecules, it takes about 50 to 60 PCR cycles to visualize the PCR product on simple ethidium bromide-stained agarose gels. However, a-specific products and smears appear long before that, because Taq DNA polymerase incorporates mistakes during the elongation step. This problem was solved in the early 1990s with the introduction of nested PCR (Holding and Monk, 1989). In this method, a first PCR round is performed, followed by a second PCR round in which a small amount of the first round PCR product serves as a template, and primers are used to amplify a fragment inside the first fragment. A-specific fragments are thus not amplified in the second PCR round, and a clear, pure PCR product is obtained.

Later, the number of PCR cycles necessary to obtain sufficient PCR product for analysis was significantly reduced with the introduction of fluorescent PCR. Fluorescent PCR products are obtained using one fluorescently labeled primer, after which the fragments are analyzed on an automated sequencer. Not only is this method much more sensitive than the ethidium bromide staining (approximately 1,000-fold), but the resolution of fragments, up to 1 base, is also much better (Lissens and Sermon, 1997).

The introduction of DNA polymerases, with a high proofreading activity and/or highly suited for multiplex PCR, has reduced the number of cycles necessary for a sufficient yield even more (Sermon *et al.*, 1998b).

33.3.3.1.2 Contamination with Extraneous DNA

The introduction of one molecule of foreign DNA in the PCR tube along with the cell that is to be analyzed can lead to a wrong diagnosis. Different types of contamination with extraneous DNA can be distinguished: contamination with DNA from cells (e.g. from the operator, sperm cells, or granulosa cells stuck to the embryo), from extracted DNA (genomic DNA), or from carry-over with PCR fragments from previous DNA amplifications.

A number of measures need to be taken to avoid and/or detect contamination (Lissens and Sermon, 1997):

1. Granulosa cells are meticulously removed from the oocytes and fertilization is achieved through ICSI to avoid contamination with sperm.
2. The pre-PCR area, where all the reactions are set up, and the post-PCR area, where the PCR products are analyzed, need to be strictly separated. Each area has its dedicated materials and equipment. No DNA, except primers and single-cell samples or highly diluted DNA, should be brought into the pre-PCR area.
3. The PCR reactions are set up in a laminar flow (UV light to be turned on overnight and regularly cleaning of working areas with DNA degrading detergents is required). Pre-PCR products and materials are kept from any DNA source; pre-PCR reagents are autoclaved whenever possible, and aliquots are stored to reduce the number of reactions from a single aliquot.
4. The manipulation of the small PCR tubes requires a certain amount of care, and even skill, in order to prevent contamination. Especially, ill-fitting gloves can be a source of carry-over contamination.
5. For each sample, a blank containing all PCR components except DNA should be run. However, a contaminated blank is only an indication of a more general problem and does not mean that the corresponding sample is contaminated. A more efficient way to detect (genomic) contamination is the use of linked or

unlinked polymorphic markers, amplified in duplex with the locus of interest.

33.3.3.1.3 Allele Drop-Out and Preferential Amplification

Allele drop-out (ADO) is defined as the non-amplification of one of two alleles when starting from a single cell. It can thus be detected only in a heterozygous cell. In preferential amplification, one allele is less amplified than the other. Both are related problems because the allele that is less amplified, but still present, can go undetected if detection methods with a low sensitivity are used (e.g. ethidium bromide-stained agarose gels). ADO can lead to serious misdiagnoses, for example in the detection of mutations in autosomal dominant diseases. When the affected allele drops out, the cell appears to be unaffected and the corresponding embryo could be transferred. This has been called an unacceptable error (Navidi and Arnheim, 1991). Conversely, in an autosomal recessive situation where both parents carry the same mutation (e.g. p.F508del in CF) ADO would not lead to the transfer of affected embryos: if ADO occurs in a carrier cell, drop-out of the healthy allele would lead to the diagnosis of an affected embryo, and the embryo would not be transferred, whereas if the affected allele drops out, the embryo would be diagnosed as homozygous healthy instead of carrier, which is still a healthy embryo, and could be transferred. This was termed an acceptable misdiagnosis.

Supposedly, ADO has led to two misdiagnoses in CF (Harper and Handyside, 1994; Verlinsky, 1996). The parents were both carrying different mutations and the embryos in which a misdiagnosis occurred were compound heterozygotes.

A first efficient way to significantly reduce ADO is to use fluorescent PCR. As mentioned earlier, fluorescent PCR is 1,000-fold more sensitive than detection of PCR products on agarose gels, requires much fewer cycles to obtain an analyzable amount of PCR product, and allows for a clear distinction between preferential amplification and real ADO. Together with the use of more efficient DNA polymerases, fluorescent PCR has been the most important breakthrough so far to reduce the risk of ADO (Sermon *et al.*, 1998b).

ADO cannot be reduced to zero, so detection methods have been devised. The most important is the use of markers in multiplex PCR. Multiplex PCR in which amplification of the mutation is combined with linked markers or in which just a set of linked markers is amplified has now become the gold standard in PGD for monogenic diseases (Lewis *et al.*, 2001). Usually, microsatellites are used as linked markers because they are highly polymorphic and thus can be used in several families (Rechitsky *et al.*, 1999; Dreesen *et al.*, 2000; Apessos *et al.*, 2001; Dean *et al.*, 2001; Pyiamongkol *et al.*, 2001; Goossens *et al.*, 2003; Moutou *et al.*, 2003). Figure 33.2 shows how multiplex PCR can detect ADO. Simply stated, the results in one locus should concord with the results in another locus. The main advantage of linked markers is that they not only ensure an indirect diagnosis of the mutation but simultaneously allow detection of ADO, contamination and recombination events, thereby increasing the accuracy of the test.

33.3.3.1.4 Adaptation of PCR to the Single-Cell Level

First, it has to be decided whether a mutation-specific protocol will be developed or whether only linked markers will be used. For monogenic diseases such as spinal muscular atrophy or achondroplasia where most of the couples present with the same mutation, the strategy may be to develop a mutation-specific protocol in which at least one closely linked marker is included (Altarescu *et al.*, 2008).

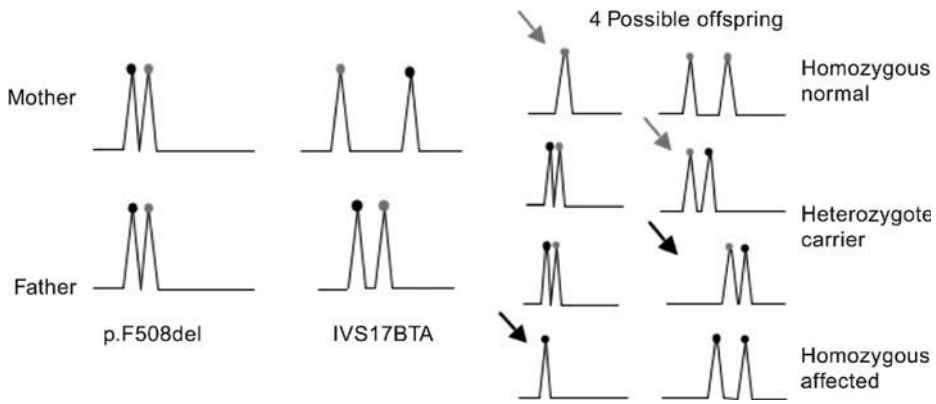


FIGURE 33.2 The principle of a duplex PCR for cystic fibrosis (mutation p.F508del and IVS17BTA intragenic marker). Both parents are carriers of the p.F508del mutation (black dot). In the marker, the black dot segregates with the mutation and the gray dot with the normal allele. Four different offspring are possible. If a result is obtained with the combination of the gray arrows (i.e. normal for the mutation and heterozygous for the marker), it can be assumed that ADO of the affected allele occurred in the mutation. This embryo could be transferred because both cells give the same result (transferable embryo). If the two black arrows are found, an ADO of the normal allele can be assumed in the mutation. Nevertheless, this embryo will not be transferred because both loci give opposite results (transferable/not transferable).

For monogenic diseases, such as Marfan's syndrome or neurofibromatosis type I for which a large number of private mutations have been described, the development of linked-marker-based PCR protocols may be more practical and efficient (Spits *et al.*, 2005, 2006a, b). Such protocols are useful for several couples regardless of the mutation, on condition that the couples are informative for the markers and segregation analysis using family samples to establish the segregation phase is possible.

The design of primers for the locus of the mutation and microsatellites should aim at similar annealing temperatures and their specific annealing should be theoretically checked in the Blast database. Primer sets should first be tested and multiplexed on genomic DNA. Primer concentrations and reaction components may be adapted to reach optimal amplification of each amplicon; further fine-tuning may be required once the reaction is carried out at the single-cell level. Finally, the single-cell protocol is validated on series of single cells (generally 50 cells and corresponding blanks) to assess the amplification efficiency as well as contamination and ADO rates.

33.3.3.2 Whole Genome Amplification

Despite all significant improvements, PCR-PGD still has limitations and these are linked to the minute quantity of start DNA. The fine-tuning of the PCR conditions during the adaptation of the multiplex-PCR to the single-cell level and the validation may take several weeks. This may be overcome when relying on whole genome amplification techniques. In this way, the PGD protocols would consist of a first universal step, the whole genome amplification of the single cell carried out in a dedicated area, followed by regular PCR methods that use standard conditions and do not require long optimization. Moreover, the yield from a single cell would be sufficient not solely to serve as input DNA for multiple, regular PCR analyses but also for HLA testing, SNP genotyping as well as (array) CGH (see section 33.3.4.2) for molecular karyotyping.

Different approaches aiming at the reliable amplification of the whole genome have been explored for many years. Ideally, a whole genome amplification (WGA) procedure should have high yield (a few μg of DNA) and should cover the whole genome with faithful representation. Moreover, for PGD applications, the procedure should be straightforward and quick. Primer extension preamplification (PEP) and degenerated oligonucleotide primed PCR (DOP-PCR) are well-established WGA methods which have been applied in PGD (Zhang *et al.*, 1992; Telenius *et al.*, 1992; reviewed by Coskun and Alsmadi, 2007). Nevertheless, PEP and DOP-PCR, both PCR based, are associated with a representation bias and high ADO when applied to single cells. The more recent multiple displacement amplification (MDA) method is non-PCR based and seems a better WGA method. In MDA, DNA is amplified isothermally using random hexamer primers and the bacteriophage ϕ29 DNA polymerase (Lizardi *et al.*, 1998; Dean *et al.*, 2002). Several commercial kits have been introduced and proved highly efficient with very good amplification rates and a lower amplification bias than PEP and DOP-PCR (Lovmar *et al.*, 2003). The first single-cell MDA reports mentioned relatively long amplification steps (6–16 h), but protocols were further optimized and shortened (2–4 h) (Handyside *et al.*, 2004; Hellani *et al.*, 2004; Spits *et al.*, 2006b; Kumar *et al.*, 2008). Single-cell MDA in clinical PGD has been applied already for CF, β-thalassemia, and non-ketotic hyperglycinaemia (Hellani *et al.*, 2005, 2008), for Zellweger syndrome (Al-Sayed *et al.*, 2007), for Marfan's syndrome, X-linked adrenoleukodystrophy, and X-linked retinoschisis (Lledó *et al.*, 2006, 2007, 2008), for Fragile X (Burlet *et al.*, 2006), and for CF and Duchenne's muscular dystrophy (Renwick *et al.*, 2006). The major problem with MDA so far is the relatively high preferential amplification and ADO rate (on average 25% which is about 3–5 higher than with PCR-based protocols). Analysis of multiple loci near the gene of interest can overcome this problem and allow reliable reconstitution of the parental haplotypes in the embryos, ensuring a diagnostic efficiency similar to PCR-based protocols. Yet considering that segregation analysis of affected family members may not always be possible, or even finding several informative, closely linked markers may be difficult, it will be necessary to further improve the ADO rate to reduce the number of markers required during analysis to make this new MDA approach the gold standard.

33.3.4 Single-Cell Cytogenetics

33.3.4.1 Fluorescence in Situ Hybridization

In FISH, fluorescent probes carrying distinct fluorochromes for different chromosomes are hybridized to cell nuclei spread either in metaphase or in interphase. When embryos are analyzed, the cells are usually in interphase. Careful choosing of the type and location of the probes allows not only enumeration of chromosomes, as in sexing and PGD-AS, but also the diagnosis of chromosome imbalances in structural abnormalities.

Two methods are currently in use for the fixation of the blastomeres. Briefly, the first method (derived from the Tarkowski method to fix embryos) consists of the following steps. First, the blastomere is lysed in hypotonic solution (1% sodium citrate in water) to release the nucleus. Second, the fixative (3:1 acetic acid:methanol) is used to dissolve away the cytoplasm and to fix the nucleus to the glass slide. The second method was described by Coonen and coworkers (2000): the blastomere is brought in spreading solution (0.1 N hydrochloric acid, 0.01% Tween 20), which is then removed and the process repeated until the nucleus is free of cytoplasm and attached to the poly-L-lysine-coated glass

slide. Some authors claim that the modified Tarkowski method yields better spread nuclei, so that the different fluorescent dots are better separated and more easily enumerated (Velilla *et al*., 2002). The Coonen method is easier to use, and gives reasonable spreading results (Staessen *et al.*, 1999), even when five different chromosomes are analyzed concurrently, in the laboratories that are more experienced with this method (Staessen *et al.*, 2003).

Fluorescent probes can be made in-house or, as is more common nowadays, are commercially available. For more widespread uses, such as sexing and PGD-AS, kits are available, containing directly labeled probes for chromosomes X, Y, 13, 16, 18, 21, and 22. These chromosomes were chosen either because they are present in live born trisomies (trisomies 13, 18, and 21) or because they are frequently present in miscarriages (16 and 22). Because the number of fluorochromes available is limited, the number of chromosomes that are analyzed can be increased either through ratio labeling and computerized analysis, or through the application of two and three hybridization rounds. Munné and coworkers (2003a) showed that in this way, up to nine different chromosomes (X, Y, 13, 15, 16, 17, 18, 21, and 21) could be analyzed for aneuploidy screening, and these authors reported an increase in implantation rate, as compared to the implantation rate obtained after analysis with five probes.

For structural abnormalities, a judicious choice of probes is made depending on the chromosomes involved and the breakpoints present. The first cases were performed using probes that specifically delineate the translocation breakpoints (Munné *et al.*, 1998, 2000). The disadvantage was that appropriate probes first had to be cloned and then shown to be useful in FISH, but the advantage was that balanced translocated karyotypes could be distinguished from normal karyotypes. A more general approach was described by Conn and coworkers (1999), taking advantage of the availability of centromeric and subtelomeric probes for each chromosome. For Robertsonian translocations, a combination of one locus-specific probe centromeric to the breakpoint of one of the chromosomes involved and one on the telomere of the other chromosome involved is applied. Three FISH probes are used for reciprocal translocations: two on the centromeric side of the breakpoints of the chromosomes involved, and a third one on one of the telomeres on the other side of the breakpoint (see Fig. 33.3). The two or three probes used must each carry a different fluorochrome, and it must be possible to use them together in one assay (Scriven *et al.*, 2000; Scriven and Ogilvie, 2007).

33.3.4.2 Comparative Genomic Hybridization

CGH was developed to characterize the often complex chromosomal rearrangements present in cancer tissue (see also Chapter 12). Total genomic DNA from the cells to be analyzed (e.g. the cancer cells) is labeled with a green fluorochrome, and a normal DNA sample is labeled with another, for example red, fluorochrome. Both labeled DNA samples are mixed and hybridized to a normal metaphase. A computerized system allows analysis of the ratios between both fluorochromes for each chromosome and detection of imbalances. If the test sample contains more

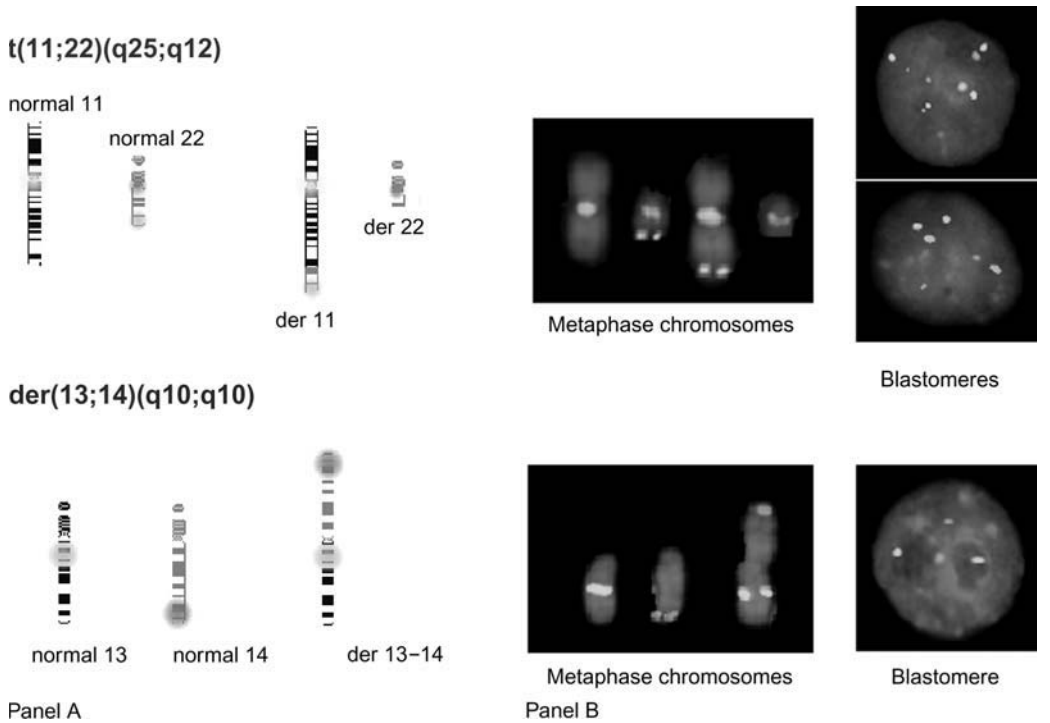


FIGURE 33.3 Example of the probes chosen for a reciprocal translocation (t(11;22), panel **A.** And for a Robertsonian translocation (t(13;14), panel **B.** The results are shown schematically, on lymphocyte metaphase spreads, and on interphase spreads of blastomeres.

of a certain sequence, for example in a duplication, then that sequence will show up more green after computerized analysis of the metaphase. Conversely, if the test sample contains less of a certain sequence (e.g. a deletion), this sequence will be more red on analysis. CGH has been down-scaled to the single cell level by introducing a first step of whole genome amplification of the DNA in the single blastomere (Voullaire *et al.*, 2000; Wells and Delhanty, 2000; Wilton *et al.*, 2001; Wells *et al.*, 2002).

The most widely used WGA method is degenerate oligonucleotide primed PCR (or DOP-PCR). More recently, multiple displacement amplification has been applied as WGA. As CGH at the single cell level required several days for analysis, groups that have presented clinical application of CGH in PGD either had to recur to polar body analysis (Wells *et al.*, 2002), or to cryopreservation of the embryos (Voullaire *et al.*, 2000; Wilton *et al.*, 2001). Wilton (2005) reviewed single-cell CGH in PGD.

Array-CGH, in which the metaphase spread is replaced by DNA sequences specific to human chromosomes spotted on an array, usually a glass slide, has a lower hybridization time, reducing the time necessary for the CGH analysis, and bringing it back well within time to transfer embryos on day 5. Other advantages of array-CGH are the higher resolution (100–200 kb instead of 2–10 Mb), and the important computerization of the whole procedure (Wells and Brynn, 2003). DOP-PCR has been used in array-CGH of single lymphocytes, fibroblasts, and single oocytes as reported by Hu and coworkers (2004) and Fragouli and coworkers (2006), respectively. Array-CGH following MDA was used for the detection of chromosomal imbalances in a single lymphoblast, fibroblast, and blastomere (Le Caignec *et al.*, 2006). The high price of the arrays is currently still a drawback and may explain why no reports on clinical PGD applications have been published yet. Nevertheless, it is regarded as the most promising future strategy to study the full chromosome set in a single cell within the timeframe of PGD.

33.4 OUTCOME OF PGD

33.4.1 Accuracy of the Diagnosis

One misdiagnosis for sexing and two for CF are mentioned in early reports (Harper and Handyside, 1994; Lissens and Sermon, 1997), and these were mainly due to the low efficiency of single-cell PCR. Further technical developments – that is, FISH for sexing and multiplex fluorescent PCR – have ruled out the recurrence of this type of error. Munné and coworkers (1999) reported one misdiagnosis (trisomy 21 after aneuploidy screening) on a total of 57 pregnancies. These authors estimate that the misdiagnosis rate after biopsy of one blastomere was around 7% in a large series; nearly 6% of the embryos were misdiagnosed due to mosaicism in the embryo (Munné *et al.*, 2003b). Pickering and coworkers (2003) reported one misdiagnosis for spinal muscular atrophy out of 18 ongoing pregnancies. The most recent report of the ESHRE PGD Consortium (2008) has mentioned 18 misdiagnoses since the start of data collection in 1999: nine after PCR and nine after FISH. The identification of these misdiagnoses and the reasons why they occurred has led to the initiation of an ESHRE working group that will develop guidelines to minimize risks (Harper *et al.*, 2008a, b).

33.4.2 Pregnancy Outcome

The ESHRE PGD consortium has reported on PGD cycles and their outcome through seven data collections so far (Harper *et al.*, 2008a, b). The latest data collection, which was quite similar to previous data collections, showed a 15% overall live birth rate per cycle to oocyte retrieval. This number is lower than can be expected in a regular IVF cycle, but it must be taken into account that a large cohort of embryos is diagnosed as affected or abnormal. Especially in patients carrying reciprocal translocations where as many as 80% of the embryos carry an unbalanced karyotype (Munné *et al.*, 1998; Van Assche *et al.*, 1999; Pickering *et al.*, 2003). The International Working Group reported that more than 3,000 clinical PGD cycles had been applied by mid-2001, resulting in around a 24% pregnancy rate (Report of the 11th Annual Meeting of International Working Group on Preimplantation Genetics (2001). Close to 700 children have been born following these pregnancies and 4.9% of these were reported to show abnormalities. The ESHRE PGD Consortium reports indicate that pregnancies and babies born after PGD are comparable to ICSI pregnancies and ICSI babies in many aspects such as birth weight and congenital malformations (Bonduelle *et al.*, 2002; Harper *et al.*, 2008a, b).

The most important reason for morbidity and mortality in the pregnancies conceived after PGD is multiplicity. The ESHRE data collection VII for 2004 reports on a high number of multiple pregnancies in PGD (36%). It is hoped that the single embryo transfer policy of regular IVF will also be introduced in PGD and substantially reduce the multiple births the coming years.

Follow-up studies of the children have been reassuring. A neurodevelopmental screening of 49 children born after PGD indicated no adverse effects of the procedure on the children's health (Banerjee *et al.*, 2008). A study of 2-year-old children born after PGD/PGS showed that the socio-emotional and language as well as the mental and psychomotor developmental outcomes were very similar to those of children born after ICSI and children naturally conceived (Nekkebroeck *et al.*, 2008a, b). Another large cohort of children, not included in the ESHRE PGD Consortium report, was reported by Strom and coworkers (2000). A total of 109 infants were described and here, too, the conclusions were that children born after PGD are very comparable with

children born after ICSI, and that PGD is a safe method to avoid the birth of children with genetic defects.

33.5 CONCLUSIONS

In more than a decade, PGD has gone from an experimental procedure to a widely accepted alternative for prenatal diagnosis. Although it is still labour intensive and time consuming, introduction of new technologies has made it safer and more reliable. Newer technologies will not only increase the reliability further, but will also make it easier and more accessible to our patients. Nevertheless, it is the authors' firm conviction that a close follow-up of the results obtained with PGD, not only in terms of technical safety and prevention of misdiagnosis, but more importantly in the follow-up of the pregnancies and babies born, is mandatory to bring PGD at the same level with prenatal diagnosis. Thus, PGD will make it possible to offer patients a real choice in how they make decisions in their reproductive lives.

REFERENCES

Al-Sayed, M., Al-Hassan, S., Rashed, M., Qeba, M., and Coskun, S. (2007). Preimplantation genetic diagnosis for Zellweger syndrome. *Fertil. Steril.* 87, 1468.e1–1468.e3.

Altarescu, G., Renbaum, P., Brooks, P.B., Margalioth, E.J., Ben Chetrit, A., Munter, G., Levy-Lahad, E., and Eldar-Geva, T. (2008). Successful polar body-based preimplantation genetic diagnosis for achondroplasia. *Reprod. Biomed. Online* 16, 276–282.

Apessos, A., Abou-Sleiman, P., Harper, J., and Delhanty, J. (2001). Preimplantation genetic diagnosis of the fragile X syndrome by use of polymorphic linked markers. *Prenat. Diagn.* 21, 504–511.

Baart, E.B., Martini, E., Eijkemans, M.J., Van Opstal, D., Beckers, N.G., Verhoeff, A., Macklon, N.S., and Fauser, B.C. (2007). Milder ovarian stimulation for in-vitro fertilization reduces aneuploidy in the human preimplantation embryo: a randomized controlled trial. *Hum. Reprod.* 22, 980–988.

Banerjee, I., Shevlin, M., Taranissi, M., Thornhill, A., Abdalla, H., Ozturk, O., Barnes, J., and Sutcliffe, A. (2008). Health of children conceived after preimplantation genetic diagnosis: a preliminary outcome study. *Reprod. Biomed. Online* 16, 376–381.

Blaszczyk, A., Tang, Y., Dietz, H., Adler, A., Berkeley, A., Krey, L., and Grifo, J. (1998). Preimplantation genetic diagnosis of human embryos for Marfan's syndrome. *J. Assist. Reprod. Genet.* 15, 281–284.

Bonduelle, M., Liebaers, I., Deketelaere, V., Derde, M.-P., Camus, M., Devroey, P., and Van Steirteghem, A. (2002). Neonatal data on a cohort of 2889 infants born after ICSI (1991–1999) and of 2995 infants born after IVF (1983–1999). *Hum. Reprod.* 17, 671–694.

Boué, A., Boué, J., and Gropp, A. (1985). Cytogenetics of pregnancy wastage. *Adv. Hum Genet.* 14, 1–57.

Burlet, P., Frydman, N., Gigarel, N., Kerbrat, V., Tachdjian, G., Feyereisen, E., Bonnefont, J.P., Frydman, R., Munnich, A., and Steffann, J. (2006). Multiple displacement amplification improves PGD for fragile X syndrome. *Mol. Hum. Reprod.* 12, 647–652.

Cieslak, J., Ivakhnenko, V., Wolf, G., Sheleg, S., and Verlinsky, Y. (1999). Three-dimensional partial zona dissection for preimplantation genetic diagnosis and assisted hatching. *Fertil. Steril.* 71, 308–313.

Conn, C., Cozzi, J., Harper, J., Winston, R., and Delhanty, J. (1999). Preimplantation genetic diagnosis for couples at high risk of Down syndrome pregnancy owing to parental translocation or mosaicism. *J. Med. Genet.* 36, 45–50.

Coonen, E., Martini, E., Dumoulin, J., Hollanders-Crombach, H., de Die-Smulders, C., Geraedts, J., Hopman, A., and Evers, J. (2000). Preimplantation genetic diagnosis of a reciprocal translocation t(13;11)(q27.3;q24.3) in siblings. *Mol. Hum. Reprod.* 6, 199–206.

Coskun, S., and Alsmadi, O. (2007). Whole genome amplification from a single cell: a new era for preimplantation genetic diagnosis. *Prenat. Diagn.* 27, 297–302.

De Boer, K.A., Catt, J.W., Jansen, R.P., Leigh, D., and McArthur, S. (2004). Moving to blastocyst biopsy for preimplantation genetic diagnosis and single embryo transfer at Sydney IVF. *Fertil. Steril.* 82, 295–298.

De Rycke, M., Van de Velde, H., Sermon, K., Lissens, W., De Vos, A., Vandervorst, M., Vanderfaeillie, A., Van Steirteghem, A., and Liebaers, I. (2001). Preimplantation genetic diagnosis for sicklecell anemia and for beta-thalassemia. *Prenat. Diagn.* 21, 214–222.

De Vos, A., and Van Steirteghem, A. (2001). Aspects of biopsy procedures prior to preimplantation genetic diagnosis. *Prenat. Diagn.* 21, 767–780.

Dean, N., Tan, S., and Ao, A. (2001). The development of preimplantation genetic diagnosis for myotonic dystrophy using multiplex fluorescent polymerase chain reaction and its clinical application. *Mol. Hum. Reprod.* 7, 895–901.

Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S.F., Egholm, M., and Lasken, R.S. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* 99, 5261–5266.

Delhanty, J., Griffin, D., Handyside, A., Harper, J., Atkinson, G., Pieters, M., and Winston, R. (1993). Detection of aneuploidy and chromosomal mosaicism in human embryos during preimplantation sex determination by fluorescent in situ hybridisation, (FISH). *Hum. Mol. Genet.* 2, 1183–1185.

Dokras, A., Sargent, I.L., Ross, C., Gardner, R.L., and Barlow, D.H. (1990). Trophectoderm biopsy in human blastocysts. *Hum. Reprod.* 5, 821–825.

Dreesen, J., Bras, M., de Die-Smulders, C., Dumoulin, J., Cobben, J., Evers, J., Smeets, H., and Geraedts, J. (1998). Preimplantation genetic diagnosis of spinal muscular atrophy. *Mol. Hum. Reprod.* 4, 881–885.

Dreesen, J., Jacobs, L., Bras, M., Herbergs, J., Dumoulin, J., Geraedts, J., Evers, J., and Smeets, H. (2000). Multiplex PCR of polymorphic markers flanking the CFTR gene: a general approach for preimplantation genetic diagnosis of cystic fibrosis. *Mol. Hum. Reprod.* 6, 391–396.

Dumoulin, J.C., Bras, M., Coonen, E., Dreesen, J., Geraedts, J.P., and Evers, J.L. (1998). Effect of Ca2+/Mg2+-free medium on the biopsy procedure for preimplantation genetic diagnosis and further development of human embryos. *Hum Reprod.* 13, 2880–2883.

Durban, M., Benet, J., Boada, M., Fernandez, E., Calafell, J.M., Lailla, J.M., Sanchez-Garcia, J.F., Pujol, A., Egozcue, J., and Navarro, J. (2001). PGD in female carriers of balanced Robertsonian and reciprocal translocations by first polar body analysis. *Hum. Reprod. Update* 7, 591–602.

Eftedal, I., Schwartz, M., Bendtsen, H., Andersen, A., and Zieve, S. (2001). Single intragenic microsatellite preimplantation genetic diagnosis for cystic fibrosis provides positive allele identification of all CFTR genotypes for informative couples. *Mol. Hum. Reprod.* 7, 307–312.

ESHRE PGD Consortium Steering Committee (2002). ESHRE Preimplantation Genetic Diagnosis Consortium: data collection III (May 2001). *Hum. Reprod.* 17, 233–246.

Fallon, L., Harton, G.L., Sisson, M.E., Rodriguez, E., Field, L.K., Fugger, E.F., Geltinger, M., Sun, Y., Dorfmann, A., Schoener, C., Bick, D., Schulman, J., Levinson, G., and Black, S.H. (1999). Preimplantation genetic diagnosis for spinal muscular atrophy I. *Neurology.* 53, 1087–1090.

Fiorentino, F., Magli, M.C., Podini, D., Ferraretti, A.P., Nuccitelli, A., Vitale, N., Baldi, M., and Gianaroli, L. (2003). The minisequencing method: an alternative strategy for preimplantation genetic diagnosis of single gene disorders. *Mol. Hum. Reprod.* 9, 399–410.

Fragouli, E., Wells, D., Thornhill, A., Serhal, P., Faed, M.J., Harper, J.C., and Delhanty, J.D. (2006). Comparative genomic hybridization analysis of human oocytes and polar bodies. *Hum. Reprod.* 21, 2319–2328.

Gianaroli, L., Magli, C., Ferraretti, A., and Munné, S. (1999). Preimplantation diagnosis for aneuploidies in patients undergoing in vitro fertilization with a poor prognosis: identification of the categories for which it should be proposed. *Fertil. Steril.* 72, 837–844.

Girardet, A., Hamamah, S., Déchaud, H., Anahory, T., Coubes, C., Hédon, B., Demaille, J., and Claustres, M. (2003). Specific detection of deleted and non-deleted dystrophin exons together with gender assignment in preimplantation genetic diagnosis of Duchenne muscular dystrophy. *Mol. Hum. Reprod.* 9, 421–427.

Girardet, A., Fernandez, C., and Claustres, M. (2007). Efficient strategies for preimplantation genetic diagnosis of spinal muscular atrophy. *Fertil. Steril.* 90, 443.e7–443.e12.

Goossens, V., Sermon, K., Lissens, W., Vandervorst, M., Vanderfaeillie, A., De Rycke, M., De Vos, A., Henderix, P., Van de Velde, H., Van Steirteghem, A., and Liebaers, I. (2000). Clinical application of preimplantation genetic diagnosis for cystic fibrosis. *Prenat. Diagn.* 20, 571–581.

Goossens, V., Sermon, K., Lissens, W., De Rycke, M., Saerens, B., De Vos, A., Henderix, P., Van de Velde, H., Platteau, P., Van Steirteghem, A., Devroey, P., and Liebaers, I. (2003). Improving clinical preimplantation genetic diagnosis for cystic fibrosis by duplex-PCR using two polymorphic markers or one polymorphic marker in combination with the detection of the DF508 mutation. *Mol. Hum. Reprod.* 9, 559–567.

Goossens, V., De Rycke, M., De Vos, A., Staessen, C., Michiels, A., Verpoest, W., Van Steirteghem, A., Bertrand, C., Liebaers, I., Devroey, P., and Sermon, K. (2008). Diagnostic efficiency, embryonic development and clinical outcome after the biopsy of one or two blastomeres for preimplantation genetic diagnosis. *Hum. Reprod.* 23, 481–492.

Griffin, D., Wilton, L., and Handyside, A. (1992). Dual fluorescent in situ hybridisation for simultaneous detection of X and Y chromosome-specific probes for sexing of human preimplantation embryonic nuclei. *Hum. Genet.* 89, 18–22.

Handyside, A., Kontogianni, E., Hardy, K., and Winston, R. (1990). Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature* 344, 768–770.

Handyside, A., Lesko, J., Tarin, J., Winston, R., and Hughes, M. (1992). Birth of a normal girl after in vitro fertilization and preimplantation diagnostic testing for cystic fibrosis. *N. Eng. J. Med.* 327, 905–909.

Handyside, A.H., Robinson, M.D., Simpson, R.J., Omar, M.B., Shaw, M.A., Grudzinskas, J.G., and Rutherford, A. (2004). Isothermal whole genome amplification from single and small numbers of cells: a new era for preimplantation genetic diagnosis of inherited disease. *Mol. Hum. Reprod.* 10, 767–772.

Hardison, R.C., Chui, D.H., Giardine, B., Riemer, C., Patrinos, G.P., Anagnou, N., Miller, W., and Wajcman, H. (2002). HbVar: a relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. *Hum. Mutat.* 19, 225–233.

Harper, J., and Handyside, A. (1994). The current status of preimplantation diagnosis. *Curr. Obst. Gynecol.* 4, 143–149.

Harper, J.C., Coonen, E., Ramaekers, F.C., Delhanty, J.D., Handyside, A.H., Winston, R.M., and Hopman, A.H. (1994). Identification of the sex of human preimplantation embryos in two hours using an improved spreading method and fluorescent in-situ hybridization (FISH) using directly labelled probes. *Hum. Reprod.* 9, 721–724.

Harper, J., Coonen, E., Handyside, A., Winston, R., Hopman, A., and Delhanty, J. (1995). Mosaicism of autosomes and sex chromosomes in morphologically normal, monospermic preimplantation human embryos. *Prenat. Diagn.* 15, 41–49.

Harper, J.C., de Die-Smulders, C., Goossens, V., Harton, G., Moutou, C., Repping, S., Scriven, P.N., SenGupta, S., Traeger-Synodinos, J., Van Rij, M.C., Viville, S., Wilton, L., and Sermon, K.D. (2008a). ESHRE PGD consortium data collection VII: cycles from January to December 2004 with pregnancy follow-up to October 2005. *Hum. Reprod.* 23, 741–755.

Harper, J., Sermon, K., Geraedts, J., Vesela, K., Harton, G., Thornhill, A., Pehlivan, T., Fiorentino, F., SenGupta, S., de Die-Smulders, C., Magli, C., Moutou, C., and Wilton, L. (2008b). What next for preimplantation genetic screening?. *Hum. Reprod.* 23, 478–480.

Harton, G., Tsipouras, P., Sisson, M., Starr, K., Mahoney, B., Fugger, E., Schulman, J., Kilpatrick, M., Levinson, G., and Black, S. (1996). Preimplantation genetic testing for Marfan syndrome. *Mol. Hum. Reprod.* 2, 713–715.

Hellani, A., Coskun, S., Benkhalifa, M., Tbakhi, A., Sakati, N., Al-Odaib, A., and Ozand, P. (2004). Multiple displacement amplification on single cell and possible PGD applications. *Mol. Hum. Reprod.* 11, 847–852.

Hellani, A., Coskun, S., Tbakhi, A., and Al-Hassan, S. (2005). Clinical application of multiple displacement amplification in preimplantation genetic diagnosis. *Reprod. Biomed. Online* 10, 376–380.

Hellani, A., Sammour, A., Johansson, L., and El-Sheikh, A. (2008). Delivery of a normal baby after preimplantation genetic diagnosis for non-ketotic hyperglycinaemia. *Reprod. Biomed. Online* 16, 893–897.

Holding, C., and Monk, M. (1989). Diagnosis of beta-thalassaemia by DNA amplification in single blastomeres from mouse preimplantation embryos. *Lancet* 2, 532–535.

Hu, D.G., Webb, G., and Hussey, N. (2004). Aneuploidy detection in single cells using DNA array-based comparative genomic hybridization. *Mol. Hum. Reprod.* 10, 283–289.

Hussey, N., Donggui, H., Froiland, D., Hussey, D., Haan, E., Matthews, C., and Craig, J. (1999). Analysis of five Duchenne muscular dystrophy exons and gender determination using conventional duplex polymerase chain reaction on single cells. *Mol. Hum. Reprod.* 5, 1089–1094.

Iwarsson, E., Richter-Ährlund, L., Inzunza, J., Rosenlund, B., Fridström, M., Hillensjö, T., Sjöblom Nordenskjöld, M., and Blennow, E. (1998). Preimplantation genetic diagnosis of a large pericentric inversion of chromosome 5. *Mol. Hum. Reprod.* 7, 719–723.

Jansen, R.P., Bowman, M.C., de Boer, K.A., Leigh, D.A., Lieberman, D.B., and McArthur, S.J. (2008). What next for preimplantation genetic screening (PGS)? Experience with blastocyst biopsy and testing for aneuploidy. *Hum. Reprod.* 23, 1476–1478.

Jiao, Z., Zhou, C., Li, J., Shu, Y., Liang, X., Zhang, M., and Zhuang, G. (2003). Birth of healthy children after preimplantation diagnosis of b-thalassemia by whole-genome amplification. *Prenat. Diagn.* 23, 646–651.

Joris, H., Nagy, P., Van de Velde, H., De Vos, A., and Van Steirteghem, A. (1998). Intracytoplasmic sperm injection: laboratory set-up and injection procedure. *Hum. Reprod.* 13(Suppl. 1), 76–86.

Joris, H., De Vos, A., Janssens, R., Devroey, P., Liebaers, I., and Van Steirteghem, A. (2003). Comparison of the results of human embryo biopsy and outcome of PGD after zona drilling using acid Tyrode medium or a laser. *Hum. Reprod.* 18, 1896–1902.

Kakourou, G., Dhanjal, S., Mamas, T., Gotts, S., Doshi, A., Fordham, K., Serhal, P., Ranieri, D.M., Delhanty, J.D., Harper, J.C., and SenGupta, S.B. (2008). Preimplantation genetic diagnosis for myotonic dystrophy type 1 in the UK. *Neuromuscul. Disord.* 18, 131–136.

Kokkali, G., Vrettou, C., Traeger-Synodinos, J., Jones, G.M., Cram, D.S., Stavrou, D., Trounson, A.O., Kanavakis, E., and Pantos, K. (2005). Birth of a healthy infant following trophectoderm biopsy from blastocysts for PGD of beta-thalassaemia major. *Hum. Reprod.* 20, 1855–1859.

Kokkali, G., Traeger-Synodinos, J., Vrettou, C., Stavrou, D., Jones, G.M., Cram, D.S., Makrakis, E., Trounson, A.O., Kanavakis, E., and Pantos, K. (2007). Blastocyst biopsy versus cleavage stage biopsy and blastocyst transfer for preimplantation genetic diagnosis of beta-thalassaemia: a pilot study. *Hum. Reprod.* 22, 1443–1449.

Kolibianakis, E.M., Zikopoulos, K., Verpoest, W., Camus, M., Joris, H., Van Steirteghem, A.C., and Devroey, P. (2004). Should we advise patients undergoing IVF to start a cycle leading to a day 3 or a day 5 transfer? *Hum. Reprod.* 19, 2550–2554.

Kuliev, A., Rechistky, S., Verlinsky, O., Ivakhnenko, V., Cieslak, J., Evsikov, S., Wolf, G., Angostiniotis, M., Kalakoutis, G., Strom, C., and Verlinsky, Y. (1999). Birth of healthy children after preimplantation diagnosis of thalassemias. *J. Assist. Reprod. Genet.* 16, 207–211.

Kuliev, A., Rechitsky, S., Verlinsky, O., Tur-Kaspa, I., Kalakoutis, G., Angastiniotis, M., and Verlinsky, Y. (2005). Preimplantation diagnosis and HLA typing for haemoglobin disorders. *Reprod. Biomed. Online* 11, 362–370.

Kumar, G., Garnova, E., Reagin, M., and Vidali, A. (2008). Improved multiple displacement amplification with phi29 DNA polymerase for genotyping of single human cells. *Biotechniques* 44, 879–890.

Le Caignec, C., Spits, C., Sermon, K., De Rycke, M., Thienpont, B., Debrock, S., Staessen, C., Moreau, Y., Fryns, J.P., Van Steirteghem, A., Liebaers, I., and Vermeesch, J.R. (2006). Single-cell chromosomal imbalances detection by array CGH. *Nucleic Acids Res.* 34, e68.

Lewis, C., Pinel, T., Whittaker, J., and Handyside, A. (2001). Controlling misdiagnosis errors in preimplantation genetic diagnosis: a comprehensive model encompassing extrinsic and intrinsic sources of error. *Hum. Reprod.* 16, 43–50.

Liebaers, I., Sermon, K., Staessen, C., Joris, H., Lissens, W., Van Assche, E., Nagy, P., Bonduelle, M., Vandervorst, M., Devroey, P., and Van Steirteghem, A. (1998). Clinical experience with preimplantation genetic diagnosis and intracytoplasmic sperm injection. *Hum. Reprod.* 13(Suppl. 1), 186–195.

Lissens, W., and Sermon, K. (1997). Preimplantation genetic diagnosis: current status and new developments. *Hum. Reprod.* 12, 1756–1761.

Liu, J., Lissens, W., Silber, S., Devroey, P., Liebaers, I., and Van Steirteghem, A. (1994). Birth after preimplantation diagnosis of the cystic fibrosis DF508 mutation by polymerase chain reaction in human embryos resulting from intracytoplasmic sperm injection with epididymal sperm. *JAMA* 272, 1858–1860.

Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C., and Ward, D.C. (1998). Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* 19, 225–232.

Lledó, B., Ten, J., Galán, F.M., and Bernabeu, R. (2006). Preimplantation genetic diagnosis of Marfan syndrome using multiple displacement amplification. *Fertil. Steril.* 86, 949–955.

Lledó, B., Bernabeu, R., Ten, J., Galán, F.M., and Cioffi, L. (2007). Preimplantation genetic diagnosis of X-linked adrenoleukodystrophy with gender determination using multiple displacement amplification. *Fertil. Steril.*, 1327–1333.

Lledó, B., Ten, J., Rodriguez-Arnedo, D., Llácer, J., and Bernabeu, R. (2008). Preimplantation genetic diagnosis of X-linked retinoschisis. *Reprod. Biomed. Online* 16, 886–892.

Lovmar, L., Fredriksson, M., Liljedahl, U., Sigurdsson, S., and Syvänen, A.C. (2003). Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. *Nucleic Acids Res.* 31, e129.

Macklon, N., and Fauser, B. (2000). Regulation of follicle development and novel approaches to ovarian stimulation for IVF. *Hum. Reprod. Update* 6, 307–312.

Marquez, C., Sandalinas, M., Bahçe, M., Alikani, M., and Munné, S. (2000). Chromosome abnormalities in 1255 cleavage-stage human embryos. *Reprod. Biomed. Online* 1, 17–26.

Mastenbroek, S., Twisk, M., van Echten-Arends, J., Sikkema-Raddatz, B., Korevaar, J.C., Verhoeve, H.R., Vogel, N.E., Arts, E.G., de Vries, J.W., Bossuyt, P.M., Buys, C.H., Heineman, M.J., Repping, S., and van der Veen, F. (2007). In vitro fertilization with preimplantation genetic screening. *N. Engl. J. Med.* 357, 9–17.

McArthur, S.J., Leigh, D., Marshall, J.T., de Boer, K.A., and Jansen, R.P. (2005). Pregnancies and live births after trophectoderm biopsy and preimplantation genetic testing of human blastocysts. *Fertil. Steril.* 84, 1628–1636.

McArthur, S.J., Leigh, D., Marshall, J.T., Gee, A.J., De Boer, K.A., and Jansen, R.P. (2008). Blastocyst trophectoderm biopsy and preimplantation genetic diagnosis for familial monogenic disorders and chromosomal translocations. *Prenat. Diagn.* 28, 434–442.

Michiels, A., Van Assche, E., Liebaers, I., Van Steirteghem, A., and Staessen, C. (2006). The analysis of one or two blastomeres for PGD using fluorescence in-situ hybridization. *Hum. Reprod.* 21, 2396–2402.

Moutou, C., and Viville, S. (1999). Improvement of preimplantation genetic diagnosis (PGD) for cystic fibrosis mutation DF508 by fluorescent polymerase chain reaction. *Prenat. Diagn.* 19, 1248–1250.

Moutou, C., Gardes, N., Rongières, C., Ohl, J., Bettahar-Lebugle, K., Wittemer, C., Gerlinger, P., and Viville, S. (2001). Allele-specific amplification for preimplantation genetic diagnosis (PGD) of spinal muscular atrophy. *Prenat. Diagn.* 21, 498–503.

Moutou, C., Gardes, N., and Viville, S. (2002). Multiplex PCR combining delta F508 mutation and intragenic microsatellites of the CFTR gene for pre-implantation genetic diagnosis (PGD) of cystic fibrosis. *Eur. J. Hum. Genet.* 10, 231–238.

Moutou, C., Gardes, N., and Viville, S. (2003). Duplex PCR for preimplantation genetic diagnosis (PGD) of spinal muscular atrophy. *Prenat. Diagn.* 23, 685–689.

Moutou, C., Gardes, N., and Viville, S. (2004). New tools for preimplantation genetic diagnosis of Huntington's disease and their clinical applications. *Eur. J. Hum. Genet.* 12, 1007–1014.

Munné, S., Lee, A., Rosenwaks, Z., Grifo, J., and Cohen, J. (1993). Diagnosis of major chromosome aneuploidies in human preimplantation embryos. *Hum. Reprod.* 8, 2185–2191.

Munné, S., Fung, J., Cassel, M., Marquez, C., and Weier, H. (1998). Preimplantation genetic diagnosis of translocations: case specific probes for interphase cell analysis. *Hum. Genet.* 102, 663–674.

Munné, S., Magli, C., Cohen, J., Morton, P., Sadowy, S., Gianaroli, L., Tucker, M., Marquez, C., Sable, D., Ferraretti, A., Massey, J., and Scott, R. (1999). Positive outcome after preimplantation diagnosis of aneuploidy in human embryos. *Hum. Reprod.* 1, 2191–2199.

Munné, S., Sandalinas, M., Escudero, T., Fung, J., Gianaroli, L., and Cohen, J. (2000). Outcome of preimplantation genetic diagnosis of translocations. *Fertil. Steril.* 73, 1209–1218.

Munné, S., Sandalinas, M., Escudero, T., Velilla, E., Walmsley, R., Sadowy, S., Cohen, J., and Sable, D. (2003a). Improved implantation after preimplantation genetic diagnosis of aneuploidy. *Reprod. Biomed. Online* 7, 91–97.

Munné, S., Sandalinas, M., Escudero, T., Marquez, C., and Cohen, J. (2003b). Chromosome mosaicism in cleavage-stage human embryos: evidence of a maternal age effect. *Reprod. Biomed. Online* 4, 223–232.

Navidi, W., and Arnheim, N. (1991). Using PCR in preimplantation genetic disease diagnosis. *Hum. Reprod.* 6, 836–849.

Nekkebroeck, J., Bonduelle, M., Desmyttere, S., Van den Broeck, W., and Ponjaert-Kristoffersen, I. (2008a). Mental and psychomotor development of 2-year-old children born after preimplantation genetic diagnosis/screening. *Hum. Reprod.* 23, 1560–1566.

Nekkebroeck, J., Bonduelle, M., Desmyttere, S., Van den Broeck, W., and Ponjaert-Kristoffersen, I. (2008b). Socio-emotional and language development of 2-year-old children born after PGD/PGS, and parental well-being. *Hum. Reprod.* 23, 1849–1857.

Parriego, M., Solé, M., Vidal, F., Boada, M., Santalo, J., Egozcue, J., Veiga, A., and Barri, P.N. (2003). One- or two-cell biopsy: does it affect implantation potential in PGD? *Reprod. Biomed. Online* 7(Suppl. 1), 31.

Patrinos, G.P., Giardine, B., Riemer, C., Miller, W., Chui, D.H., Anagnou, N.P., Wajcman, H., and Hardison, R.C. (2004). Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. *Nucleic Acids Res.* 32, D537–D541.

Pellestor, F., Girardet, A., Andréo, B., Arnal, F., and Humeau, C. (1994). Relationship between morphology and chromosomal constitution in human preimplantation embryo. *Mol. Reprod. Dev.* 39, 141–146.

Pickering, S., Polidoropoulos, N., Caller, J., Scriven, P., Mackie Ogilvie, C., and Braude, P. The PGD Study Group (2003). Strategies and outcomes of the first 100 cycles of preimplantation genetic diagnosis at the Guy's and St. Thomas' Center. *Fertil. Steril.* 79, 81–90.

Piyamongkol, W., Harper, J., Sherlock, J., Doshi, A., Serhal, P., Delhanty, J., and Wells, D. (2001). A successful strategy for preimplantation genetic diagnosis of myotonic dystrophy using multiplex fluorescent PCR. *Prenat. Diagn.* 21, 223–232.

Preimplantation Genetic Diagnosis International Society (2004). The Preimplantation Genetic Diagnosis International Society (PGDIS): guidelines for good practice in PGD. *Reprod. Biomed. Online* 9, 430–434.

Preimplantation Genetic Diagnosis International Society (PGDIS) (2008). Guidelines for good practice in PGD: programme requirements and laboratory quality assurance. *Reprod. Biomed. Online* 16, 134–147.

Ray, P., Vekemans, M., and Munnich, A. (2001). Single cell multiplex PCR amplification of five dystrophin gene exons combined with gender determination. *Mol. Hum. Reprod.* 7, 489–494.

Rechitsky, S., Strom, C., Verlinsky, O., Amet, T., Ivakhnenko, V., Kukharenko, V., Kuliev, A., and Verlinsky, Y. (1999). Accuracy of preimplantation diagnosis of single-gene disorders by polar body analysis of oocytes. *J. Assist. Reprod. Genet.* 16, 192–198.

Renwick, P.J., Trussler, J., Ostad-Saffari, E., Fassihi, H., Black, C., Braude, P., Ogilvie, C.M., and Abbs, S. (2006). Proof of principle and first cases using preimplantation genetic haplotyping – a paradigm shift for embryo diagnosis. *Reprod. Biomed. Online* 13, 110–119.

Report of the 11th Annual Meeting of International Working Group on Preimplantation Genetics (2001). Preimplantation genetic diagnosis: experience of 3000 clinical cycles. *Reprod. Biomed. Online,* 3, 49–53.

Rice, J., Aquiles Sanchez, J., Pierce, K., and Wangh, L. (2002). Real-time PCR with molecular beacons provides a highly accurate assay for detection of Tay-Sachs alleles in single cells. *Prenat. Diagn.* 2, 1130–1134.

Rubio, C., Simon, C., Vidal, F., Rodrigo, L., Pehlivan, T., Remohi, J., and Pellicer, A. (2003). Chromosomal abnormalities and embryo development in recurrent miscarriage couples. *Hum. Reprod.* 18, 182–188.

Schoolcraft, W., and Gardner, D. (2001). Blastocyst versus D2 or D3 transfer. *Sem. Reprod. Med.* 19, 259–268.

Scriven, P., O'Mahony, F., Bickerstaff, H., Yeong, C.-T., Braude, P., and Mackie Ogilvie, C. (2000). Clinical pregnancy following blastomere biopsy and PGD for a reciprocal translocation carrier: analysis of meiotic outcomes and embryo quality in two IVF cycles. *Prenat. Diagn.* 20, 587–592.

Scriven, P.N., and Ogilvie, C.M. (2007). Fluorescence in situ hybridization on single cells. (Sex determination and chromosome rearrangements). *Methods Mol. Med.* 132, 19–30.

Sermon, K., Goossens, V., Seneca, S., Lissens, W., De Vos, A., Vandervorst, M., Van Steirteghem, A., and Liebaers, I. (1998a). Preimplantation diagnosis for Huntington's disease (HD): clinical application and analysis of the HD expansion in affected embryos. *Prenat. Diagn.* 18, 1427–1436.

Sermon, K., De Vos, A., Van de Velde, H., Seneca, S., Lissens, W., Joris, H., Vandervorst, M., Van Steirteghem, A., and Liebaers, I. (1998b). Fluorescent PCR and automated fragment analysis for the clinical application of preimplantation genetic diagnosis of myotonic dystrophy (Steinert's disease). *Mol. Hum. Reprod.* 4, 791–795.

Sermon, K., Lissens, W., Messiaen, L., Bonduelle, M., Vandervorst, M., Van Steirteghem, A., and Liebaers, I. (1999a). Preimplantation genetic diagnosis of Marfan syndrome with the use of fluorescent polymerase chain reaction and the Automated Laser Fluorescence DNA Sequencer. *Fertil. Steril.* 71, 163–165.

Sermon, K., Seneca, S., Vanderfaeillie, A., Lissens, W., Joris, H., Vandervorst, M., Van Steirteghem, A., and Liebaers, I. (1999b). Preimplantation diagnosis for Fragile X syndrome based on the detection of the non-expanded paternal and maternal CGG. *Prenat. Diagn.* 19, 1223–1230.

Sermon, K., Seneca, S., De Rycke, M., Goossens, V., Van de Velde, H., De Vos, A., Platteau, P., Lissens, W., Van Steirteghem, A., and Liebaers, I. (2001). PGD in the lab for triplet repeat diseases – myotonic dystrophy, Huntington's disease and Fragile-X syndrome. *Mol. Cell. Endocrinol.* 183(Suppl. 1), 77–85.

Sermon, K., Moutou, C., Harper, J., Geraedts, J., Scriven, P., Wilton, L., Magli, M.C., Michiels, A., Viville, S., and De Die, C. (2005). ESHRE PGD Consortium data collection IV: May–December 2001. *Hum. Reprod.* 20, 19–34.

Spits, C., De Rycke, M., Van Ranst, N., Joris, H., Verpoest, W., Lissens, W., Devroey, P., Van Steirteghem, A., Liebaers, I., and Sermon, K. (2005). Preimplantation genetic diagnosis for neurofibromatosis type 1. *Mol. Hum. Reprod.* 11, 381–387.

Spits, C., De Rycke, M., Verpoest, W., Lissens, W., Van Steirteghem, A., Liebaers, I., and Sermon, K. (2006a). Preimplantation genetic diagnosis for Marfan syndrome. *Fertil. Steril.* 86, 310–320.

Spits, C., Le Caignec, C., De Rycke, M., Van Haute, L., Van Steirteghem, A., Liebaers, I., and Sermon, K. (2006b). Whole-genome multiple displacement amplification from single cells. *Nat. Protoc.* 1, 1965–1970.

Spits, C., De Rycke, M., Van Ranst, N., Verpoest, W., Lissens, W., Van Steirteghem, A., Liebaers, I., and Sermon, K. (2007). Preimplantation genetic diagnosis for cancer predisposition syndromes. *Prenat. Diagn.* 27, 447–456.

Staessen, C., Van Assche, E., Joris, H., Bonduelle, M., Vandervorst, M., Liebaers, I., and Van Steirteghem, A. (1999). Clinical experience of sex determination by fluorescent in situ hybridisation for preimplantation genetic diagnosis. *Mol. Hum. Reprod.* 5, 392–399.

Staessen, C., Tournaye, H., Van Assche, E., Michiels, A., Van Landuyt, L., Devroey, P., Liebaers, I., and Van Steirteghem, A. (2003). PGD in 47, XXY Klinefelter's syndrome patients. *Hum. Reprod. Update* 9, 319–330.

Staessen, C., Platteau, P., Van Assche, E., Michiels, A., Tournaye, H., Camus, M., Devroey, P., Liebaers, I., and Van Steirteghem, A. (2004). Comparison of blastocyst transfer with or without preimplantation genetic diagnosis for aneuploidy screening in couples with advanced maternal age: a prospective randomized controlled trial. *Hum. Reprod.* 19, 2849–2858.

Stern, H., Harton, G., Sisson, M., Jones, S., Fallon, L., Thorsell, L., Getlinger, M., Black, S., and Schulman, J. (2002). Nondisclosing preimplantation genetic diagnosis for Huntington disease. *Prenat. Diagn.* 22, 303–307.

Stevens, J., Wale, P., Surrey, E.S., and Schoolcraft, W.B. (2004). Is aneuploidy screening for patients aged 35 or over beneficial? A prospective randomized trial. *Fertil. Steril.* 82, 249.

Strom, C., Ginsberg, N., Rechitsky, S., Cieslak, J., Ivakhenko, V., Wolf, G., Lifchez, A., Moise, J., Valle, J., Kaplan, B., White, M., Barton, J., Kuliev, A., and Verlinsky, Y. (1998). Three births after preimplantation genetic diagnosis for cystic fibrosis with sequential first and second polar body analysis. *Am. J. Obstet. Gynecol.* 178, 1298–1306.

Strom, C., Levin, R., Strom, S., Masciangelo, C., Kuliev, A., and Verlinsky, Y. (2000). Neonatal outcome of preimplantation genetic diagnosis by polar body removal: the first 109 infants. *Pediatrics* 106, 650–653.

Telenius, H., Carter, N.P., Bebb, C.E., Nordenskjöld, M., Ponder, B.A., and Tunnacliffe, A. (1992). Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13, 718–725.

Thornhill, A.R., De Die-Smulders, C.E., Geraedts, J.P., Harper, J.C., Harton, G.L., Lavery, S.A., Moutou, C., Robinson, M.D., Schmutzler, A.G., Scriven, P.N., Sermon, K.D., and Wilton, L. (2005). ESHRE PGD Consortium ''Best practice guidelines for clinical preimplantation genetic diagnosis (PGD) and preimplantation genetic screening (PGS)''. *Hum. Reprod.* 20, 35–48.

Van Assche, E., Staessen, C., Vegetti, W., Bonduelle, M., Vandervorst, M., Van Steirteghem, A., and Liebaers, I. (1999). Preimplantation genetic diagnosis and sperm analysis by fluorescence in situ hybridization for the most common reciprocal translocation t(11;22). *Mol. Hum. Reprod.* 5, 682–690.

Van de Velde, H., De Vos, A., Sermon, K., Staessen, C., De Rycke, M., Van Assche, E., Lissens, W., Vandervorst, M., Van Ranst, H., Liebaers, I., and Van Steirteghem, A. (2000). Embryo implantation after biopsy of one or two cells from cleavage stage embryos with a view to preimplantation genetic diagnosis. *Prenat. Diagn.* 20, 1030–1037.

Vandervorst, M., Liebaers, I., Sermon, K., Staessen, C., De Vos, A., Van de Velde, H., Van Assche, E., Joris, H., Van Steirteghem, A., and Devroey, P. (1998). Successful preimplantation genetic diagnosis is related to the number of available cumulus-oocyte-complexes. *Hum. Reprod.* 13, 3169–3176.

Veiga, A., Sandalinas, M., Ben Khalifa, M., Boada, M., Carrera, M., Santaló, J., Barri, P.N., and Ménézo, Y. (1997). Laser blastocyst biopsy for preimplantation diagnosis in the human. *Zygote* 5, 351–354.

Velilla, E., Escudero, T., and Munné, S. (2002). Blastomere fixation techniques and risk of misdiagnosis for preimplantation genetic diagnosis of aneuploidy. *Reprod. Biomed. Online* 4, 210–217.

Verlinsky, Y., Rechitsky, S., Evsikov, S., White, M., Cieslak, J., Lifchez, A., Valle, J., Moise, J., and Strom, C. (1992). Preconception and preimplantation diagnosis for cystic fibrosis. *Prenat. Diagn.* 12, 103–110.

Verlinsky, Y. (1996). Preimplantation genetic diagnosis. *J. Assist. Reprod. Genet.* 13, 87–89. (Editorial).

Verlinsky, Y., and Kuliev, A. (1996). Preimplantation diagnosis of common aneuploidies in fertile couples of advanced maternal age. *Hum. Reprod.* 11, 2076–2077.

Verlinsky, Y., Cieslak, J., Ivakhnenko, V., Evsikov, S., Wolf, G., White, M., Lifchez, A., Kaplan, B., Moise, J., Valle, J., Ginsberg, N., Strom, C., and Kuliev, A. (1999). Prevention of age-related aneupoidies by polar body testing of oocytes. *J. Assist. Reprod. Genet.* 16, 165–169.

Verlinsky, Y., Rechitsky, S., Verlinsky, O., Xu, K., Schattman, G., Masciangelo, C., Ginberg, N., Strom, C., Rosenwaks, Z., and Kuliev, A. (2001). Preimplantation diagnosis for p53 tumour suppressor gene mutations. *Reprod Biomed Online* 2, 102–105.

Verlinsky, Y., Rechitsky, S., Sharapova, T., Laziuk, K., Barsky, I., Verlinsky, O., Tur-Kaspa, I., and Kuliev, A. (2007). Preimplantation diagnosis for immunodeficiencies. *Reprod. Biomed. Online* 14, 214–223.

Voullaire, L., Slater, H., Williamson, R., and Wilton, L. (2000). Chromosome analysis of blastomeres from human embryos by using CGH. *Hum. Genet.* 105, 210–217.

Vrettou, C., Palmer, G., Kanavakis, E., Tretis, M., Antoniadi, T., Mastrominas, M., and Traeger-Synodinos, J. (1999). A widely applicable strategy for single cell genotyping of b-thalassemia mutations using DGGE analysis: application to preimplantation genetic diagnosis. *Prenat. Diagn.* 19, 1209–1216.

Vrettou, C., Tzetis, M., Traeger-Synodinos, J., Palmer, G., and Kanavakis, E. (2002). Multiplex sequence variation detection throughout the CFTR gene appropriate for preimplantation genetic diagnosis in populations with heterogeneity of cystic fibrosis mutations. *Mol. Hum. Reprod.* 8, 880–886.

Vrettou, C., Traeger-Synodinos, J., Tzetis, M., Palmer, G., Sofocleous, C., and Kanavakis, E. (2004). Real-time. PCR for single-cell genotyping in sickle cell and thalassemia syndromes as a rapid, accurate, reliable, and widely applicable protocol for preimplantation genetic diagnosis. *Hum. Mutat.* 23, 513–521.

Wells, D., and Delhanty, J. (2000). Comprehensive chromosomal analysis of human preimplantation embryos using whole genome amplification and single cell comparative gnomic hybridisation. *Mol. Hum. Reprod.* 6, 1055–1062.

Wells, D., Escudero, T., Levy, B., Hirschhorn, K., Delhanty, J., and Munné, S. (2002). First clinical application of comparative genomic hybridisation and polar body testing for preimplantation genetic diagnosis of aneuploidy. *Fertil. Steril.* 78, 543–549.

Wells, D., and Brynn, L. (2003). Cytogenetics in reproductive medicine: the contribution of comparative genomic hybridisation (CGH). *BioEssays* 25, 289–300.

Wilton, L., Williamson, R., McBain, J., Edgar, D., and Voullaire, L. (2001). Birth of a healthy infant after preimlantation confirmation of euploidy by comparative genomic hybridisation. *N. Engl. J. Med.* 34, 1537–1541.

Wilton, L. (2002). Preimplantation genetic diagnosis for aneuploidy screening in early human embryos: a review. *Prenat. Diagn.* 22, 312–318.

Wilton, L. (2005). Preimplantation genetic diagnosis and chromosome analysis of blastomeres using comparative genomic hybridization. *Hum. Reprod. Update* 11, 33–41.

Wilton, L., Voullaire, L., Sargeant, P., Williamson, R., and McBain, J. (2003). Preimplantation aneuploidy screening using comparative genomic hybridization or fluorescence in situ hybridization of embryos from patients with recurrent implantation failure. *Fertil. Steril.* 80, 860–868.

Xu, K., Shi, Z., Veeck, L., Hughes, M., and Rosenwaks, Z. (1999). First unaffected pregnancy using preimplantation genetic diagnosis for sickle cell anemia. *JAMA* 281, 1701–1706.

Zenzes, M., and Casper, R. (1992). Cytogenetics of human oocytes, zygotes, and embryos after in vitro fertilization. *Hum. Genet.* 88, 367–375.

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., and Arnheim, N. (1992). Whole genome amplification from a single cell: implications for genetic analysis. *Proc. Natl. Acad. Sci. USA* 89, 5847–5851.

Chapter 34

Automated DNA Hybridization and Detection

Bert Gold
Human Genetics Section, Cancer and Inflammation Program, Center for Cancer Research, National Cancer Institute at Frederick, Frederick, MD, USA

34.1 INTRODUCTION

The accuracy and precision of molecular identification has considerably improved as a consequence of the application of robots to the tasks of nucleic acid hybridization. Automation has enabled higher-throughput molecular assays and diagnostics. Inspiration for automating these processes has come from the successful effort to map and sequence the human genome in the past two decades. Specific tasks that have been automated have been manually chained together so that nucleic acid extraction, amplification, hybridization, and detection are dovetailed. In a clinical molecular diagnostics laboratory, these processes are monitored, interpreted and reported by skilled technologists and diagnosticians. In a research environment, chaining these processes has permitted high-throughput single nucleotide polymorphism (SNP) characterization and voluminous data production. Many instrument and reagent manufacturers have participated in the automation initiative. The short-term future should witness increasing integration and a diminished manual component for dovetailing automated, high-throughput subtasks. As laboratory throughput has grown, the need for biological information handling has grown too. As a consequence, database management and result reporting and analysis systems are poised for a period of rapid growth (see also Chapter 25). Since a broad review of this topic (Beck, 1995), the human genome initiative has permitted great strides in automating molecular hybridization and detection technology. While non-radioactive detection was then a novelty, it has now become commonplace. And, while clinical molecular diagnosis was then an exception, it is now the accepted standard for cystic fibrosis (CF) carrier screening in the USA. In fact, American College of Medical Genetics/American College of Obstretrics and Gynecology guidelines promulgate prenatal carrier screening of every Caucasian couple of childbearing age (Grody *et al*., 2001).

Automated hybridization now encompasses nucleic acid extraction, microarray spotting, colony and plaque hybridization, picking, and many parts of DNA sequencing, primer extension, and SNP detection. The aforementioned earlier review also subsumed fluorescent *in situ* hybridization (FISH) applications under its rubric, but that topic, comparative genomic hybridization (CGH), array-CGH, and spectral karyotyping (SKY), is beyond the scope of this chapter. More details on these techniques can be found in Chapters 10 and 12.

34.2 DNA HYBRIDIZATION

34.2.1 Origins of DNA Hybridization

DNA hybridization was anticipated by Chargaff's rule and enumerated in Watson and Crick's 1953 landmark publications. Spiegelman's contribution to the field was recognizing that it could be used to gauge phage similarity and quantify viruses (Gillespie and Spiegelman, 1965). During the same period, Ron Davis, a graduate student in Norman Davidson's laboratory at the time, developed the heteroduplex method for visualizing deletions in phage genomes with the electron microscope (Davis and Davidson, 1968). Roy Britten, Barbara Hough-Evans, and Eric Davidson began experiments aimed at investigating sequence diversity through examination of hybridization kinetics (Davidson and Hough, 1969). Of course, all of these were wet bench experiments carried out with manually purified and quantitated (often sheared) nucleic acids in laboratory glassware or quartz cuvettes.

34.2.2 Southern Blot Hybridization

Ed Southern invented his namesake Southern blot in 1975 (Southern, 1975). The original method required the purchase of a qualified batch of nitrocellulose membrane; an explosive, brittle solid support that occasionally dehydrates and loses wettability characteristics required for nucleic acid transfer. The blotting method, as originally described, required osmotic wicking to provide a flow of high-salt transfer solution through the gel and directly onto a properly wet membrane. After transfer, blots were washed in a lower salt buffer and then dried in a vacuum oven. Radioactive probing in the first 10 years of Southern blotting was standard operating procedure and only in the past ten years has given way to non-isotopic techniques (Gold *et al*., 2000). The Southern blot provided a cumbersome yet reliable method for molecular diagnosis. It permitted direct interrogation of human genomic sequence through cleavage pattern identification using an appropriate restriction endonuclease. Unfortunately, it also often required the use of radioactivity and copious amounts of human genomic DNA.

From 1986, when Lander and Botstein first suggested that restriction fragment length polymorphism (RFLP) could be used for gene mapping, through the early 1990s, Southern blot characterization of RFLP became a central method for gene mapping.

DNA from families affected with a genetic disease was collected, digested with various inexpensive restriction endonucleases, and sized by electrophoresis. The Oncor ProbeTech, shown in Fig. 34.1a, was helpful for this procedure, and was purchased by a number of laboratories to help automate Southern blotting. The device, which is no longer in production, consisted of an electrophoresis chamber with pumps for recirculating buffer, and a safety lock to prevent accidental shock to the operator. Recirculation is a requirement to minimize joule heating and provide high-resolution Southern blotting. Without buffer recirculation, ion polarization during electrophoresis distorts the band pattern made by near saturating, high-molecular weight DNA as it runs through agarose pores (Dracopoli *et al.*, 2000). In addition to the recirculation pumps, the device, shown in Fig. 34.1, contained convenient gel trays that permitted denaturation and neutralization steps, both pre-blotting requirements of the Southern blot protocol to be accomplished efficiently. The company sold a line of pre-tested buffers that could be used off the shelf, or diluted for electrophoresis, denaturation, and neutralization. Also, a charged nylon membrane supplied by the company was precut to fit the gel chamber and a vacuum pump built into the device provided the osmotic force to transfer sized, genomic DNA from the gel to the membrane.

Although production of these devices ceased in the middle 1990s, they were workhorses in a number of molecular diagnostics and gene mapping laboratories, and provide an exemplary first effort at automation.

FIGURE 34.1 Oncor Probetech and Tecan ProfiBlot. **A.** The Oncor Probetech was an automated Southern blot device that is no longer in production. It was routinely used to produce Southern blots for RFLP studies intended for gene mapping through the 1990s. It was also extensively used for Southern blot hybridization required for characterization of Fragile X syndrome diagnostic confirmation and carrier testing. **B.** The Tecan ProfiBlot is an automated RDB developing device that provides more reproducible LiPA results. Latter photo courtesy of Roche Molecular Systems.

34.2.3 From Southern Blot to PCR

During the 1990s, as molecular genetic diagnosis came of age, polymerase chain reaction-RFLP (PCR-RFLP) analysis became the ''gold standard'' for molecular genetic testing. PCR-RFLP was believed by some clinicians to provide greater analytic specificity than spot hybridization techniques (Gold, 2001, 2003). This is because two levels of specificity were required for its proper function. In Southern blotting, both hybridization and sizing form the basis for unambiguous identification of a molecular fragment. Similarly, in PCR-RFLP, both accurate amplification and proper sizing of an amplified fragment are required for unambiguous molecular identification. However, PCR-RFLP is not practical for genetically heterogeneous disorders (hereditary disorders with hundreds of known disease-bearing alleles, such as CF or β-thalassemia). Direct DNA sequencing would be the most thorough mutation detection approach, and is being tried in some laboratories

(Groman *et al*., 2002; Strom *et al*., 2003), but it is arduous, capital intensive, requires relatively expensive supplies, and is time consuming to analyze. Rapid reliable molecular diagnosis of genetically heterogeneous disorders requires a unique approach. Because the majority of these hereditary illnesses occur in families with no prior history, a high-sensitivity screening test to identify at-risk carriers is required. Molecular carrier screening tests must be cost effective to be practical. Commercialization of molecular diagnostics for genetically heterogeneous disorders involves special considerations. Among these are low cost, high throughput, and rapid interpretation. Automated printing and processing of standardized reverse dot-blots (Fig. 34.1b) in conjunction with robust, optimized multiplex PCR reagents provides a practical commercial mutation detection method. Technical factors, such as methods of attachment of oligonucleotide probes to solid supports, the nature of the solid supports themselves and PCR product visualization techniques play an important role in formatting high-throughput reverse dot-blot diagnostics. Reverse dot-blots (RDB) have become popular for clinical molecular disease diagnosis recently (see also Chapter 2); however, newer chip-based techniques discussed here and homogeneous assay systems are likely to supplant the reverse dot-blot in the not-too-distant future.

34.3 DNA EXTRACTION

In the past ten years, four automated platforms for DNA extraction have emerged as field leaders. These are the Roche MagnaPure system, the Qiagen BioRobot systems, and the Qiagen Autopure LS system (pictured in Fig. 34.2, panels A to C). In addition, a new entrant to the field has begun to make inroads in the marketplace: this is a paramagnetic system based on technology developed at the Massachusetts Institute of Technology (MIT) Whitehead Institute. Paramagnetic methods were originally developed as a partnership between Dynal and Beckman, which has since morphed into a subsidiary of Beckman-Coulter named Agencourt Bioscience Corporation (pictured in Fig. 34.2, panel D). Unique specifications and requirements describe each of these systems. And yet, each automatic method provides sufficient pure genomic DNA product from peripheral blood lymphocytes to perform conventional PCR and restriction endonuclease cleavage for Southern blotting.

The extraction principle underlying the Roche MagnaPure system (Fig. 34.2a) is that glass beads are coated with a magnetic layer held inside reaction tips. Bead bound nucleic acids are cycled in reaction tips through a series of processing steps. Once the sample is loaded, there is no operator intervention until extraction is complete. Roche sells two MagnaPure Systems, of these the larger takes 90 minutes to purify 32 samples with a per sample expendable cost of approximately US$2. Typical yield is 10 μg from 200 μL of peripheral blood collected in EDTA anticoagulant. High-salt nucleic acid adherent hydrated silica gel tubes provide the basis for Qiagen's nucleic acid extraction technology. Subsequent to lysis, gel adherent cell nucleic acids are extracted by high salt and alcohol buffer washed, dried, and eluted in a low-salt buffer. Unfortunately, a final centrifugation step requires removing the samples from the robot, as the centrifuge has not been interfaced. Qiagen's BioRobot MDX system (Fig. 34.2b) takes approximately two hours to extract 96 samples, with yields in the range of 4–6 μg from 200 μL of whole blood. The MDX is a gantry robot that uses a vacuum manifold to enable rapid purification and standardize processing. During 2006, Qiagen acquired Gentra Systems. This acquisition permitted Qiagen to sell Puregene kits, based on a salting out procedure, which have long provided nucleic acid substrate for molecular diagnostic assays. It also made Qiagen distributor of the Autopure LS instrument (Fig. 34.2c). This fully automated system provides 35 μg per mL of whole human peripheral blood DNA, with more than 70–90% between 100 and 200 kb. The DNA is also consistent with long-term storage characteristics (up to ten years) which have been discussed in the literature (Farkas *et al*., 1996). Sample sizes up to 10 mL of whole blood are accommodated by the instrument, with a throughput of eight samples in less than one hour, 16 samples in 80 minutes, and 96 samples per eight hour shift. The instrument accommodates 16 samples at a time.

A new entrant to the high-throughput nucleic acid extraction marketplace is the Agencourt Bioscience Corporation, a subsidiary of Beckman-Coulter. Beckman's gantry robots have been used for automated hybridization (see below and pictured in Fig. 34.2d), but the new NX and FX series have been programmed for paramagnetic bead addition protocols that provide high-quality nucleic acids through Solid Phase Reversible Immobilization (Hawkins *et al*., 1995).

In a recent publication, performance of several different extraction methods was compared using quantitative PCR of a housekeeping gene as a quality assurance endpoint (Cler *et al*., 2006). It was found that each tested commercial kit provided better quality DNA than phenol-chloroform extraction or single-step proteinase K digestion, and of the methods discussed here, Puregene and Qiagen column purification produced amplifiable DNA more frequently than paramagnetic beads for starting cell numbers $<$50,000.

34.4 QUANTIFYING DNA

Diverging views characterize methods for genomic DNA quantification in preparation for PCR or restriction endonuclease cleavage assays. On the one hand, the optical density

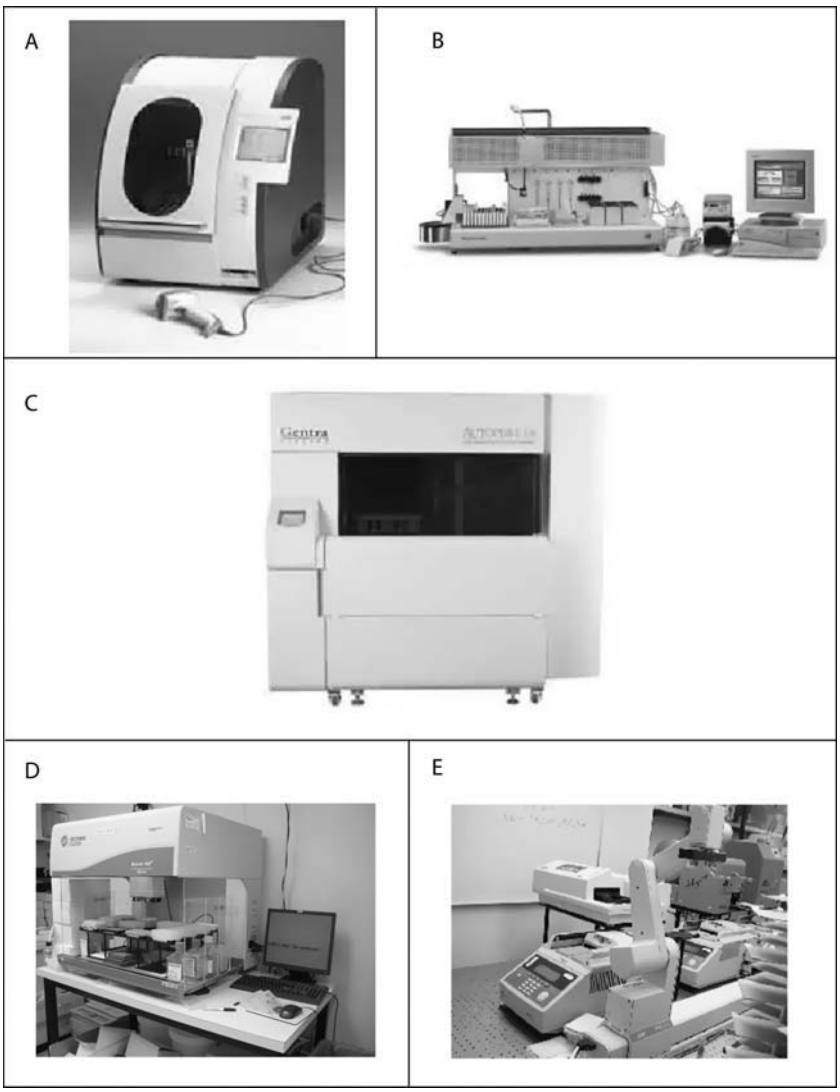


FIGURE 34.2 DNA extraction and one spectrophotometric quantitation system. **A.** The MagNA Pure Compact System is a small robot intended to automate nucleic acid extraction in low-throughput laboratories. The system is sold in two different formats; the smaller can simultaneously extract up to eight samples using the same principles as the larger Roche system which is capable of 32 simultaneous nucleic acid isolations. The system uses silica coated magnetic core beads to extract nucleic acid in disposable reaction tips. Operating costs and details are discussed in the text. **B.** Qiagen's BioRobots MDX uses adherence to a silica bead in high-salt conditions to provide a basis for one of its robotic extraction methods. An on-board vacuum system provides the basis for separation of silica adherent nucleic acids from wash solutions. **C.** Qiagen's Autopure LS was originally designed by Gentra Systems. It uses a well-characterized ''salting out'' method to provide high DNA yield and purity on 16 simultaneous peripheral blood samples with ''walk-away'' operation. **D.** Beckman-Coulter's NX gantry robot set-up for use with the paramagnetic DNA extraction method first developed with Dynal, and later made into the Agincourt subsidiary of Beckman. Fresh peripheral blood can be processed into amplifiable DNA using this system; however, low leukocyte counts (below 500,000) have shown substandard performance. **E.** A Molecular Devices SpectraMax Plus 384 micro-well plate scanning spectrophotometer can use specially engineered 384-well quartz cuvette plates to provide accurate optical densities at 230, 260, and 280 nanometers. Here, the spectrophotometer is shown integrated into a Beckman FX-based sequencing pipeline constructed by David Munroe and his collaborators at the National Cancer Institute in Frederick, Maryland.

at wavelength maxima for the macromolecule (OD 260) compared with the wavelength maxima for phenylalanine and tyrosine, representing protein presence (OD 280) have been customarily compared in a ratio to estimate nucleic acid purity (OD 260/OD 280). Early workers charted ratios to provide rigor to purity judgments. In the current milieu, this has meant using high-throughput spectrophotometers, such as Molecular Devices Spectra Max micro-well plate reader to provide wavelength scans for 96 to 384 samples simultaneously (Hilbertm *et al*., 2000). This device rigged into a Beckman FX liquid handling robot is shown in Fig. 34.2e. On the other hand, numerous labs (CIDR, Myriad Genetic Laboratories) report more accurate genomic DNA measures using fluorescent quantitation with the intercalating dye Pico Green or one of the Hoechst stains (Hoechst 33258 and its relatives; Hopwood *et al*., 1997). These are read on

a fluorimeter (such as the Molecular Devices Gemini series or Tecan Genios) in 96-well or 384-well format and compared with a standard dilution series of DNAs, compiled on the same instrument at the E_{max} of the intercalating dye.

34.5 ROBOTICS

In his chapter, Stephan Beck (1995) points to several breadboards as providing initial hybridization robotics. Among these were Beck's own DNA sequencing device; his hybridization device (Alderton *et al*., 1994), and George Church's genomic sequencing machine (Church and Gilbert, 1984, 1985). However, the first efforts at using off-the-shelf liquid handling technology to successfully automate an ongoing laboratory operation were the application of the Beckman Biomek 1000 to RDB in the Department of Human Genetics of the Baylor College of Medicine (DeMarchi *et al*., 1994). In this research institute, a team began robotically spotting successive nylon membranes with amplification products and then hybridizing these with a variety of probes. It was through this effort that robotization was first attempted for forward dot-blots. It was not until 2001 that a robot procedure was published for spotting reverse dot-blots (Lappin *et al*., 2001). Commercialization and the desire to retain competitive advantage means most automated procedures remain unpublished in detail. In fact, no central repository for automated laboratory robotic codes or procedures has yet been established.

34.5.1 A Revolution in Liquid Handling Robotics

One important force propelling molecular hybridization technology forward is the revolution witnessed in liquid handling technology. Beckman-Coulter's development of a series of disposable tip liquid handling robots (Multimek, Biomek 1000 and 2000, and Beckman FX) has permitted the automation of a variety of recalcitrant laboratory tasks. Nearly simultaneously, companies like Tecan and Packard developed a series of gantry robots using continuous pipetting technology to allow relatively inexpensive liquids (because the robots have significant void volumes) that have long shelf-lives to be accurately dispensed. Hamilton Corporation, a micro-syringe manufacturer, built a series of robots that provide small volume accuracy based on positive pressure displacement, and can be programmed for bleach or distilled water wash between solution additions. This non-comprehensive description of current liquid handling technology has made possible a variety of automated solutions to blotting tasks; however, there is a hidden cost: there has been little standardization of the software required to operate these robots even though some computer engineers have advocated the use of Tcl and Tk to the task (Ousterhout, 1994). Unfortunately, the use of proprietary robot controller software has flourished, and therefore a true automated laboratory requires an engineering team's support. As a consequence, engineering costs may cause automation to be too expensive for laboratories that would otherwise use it.

34.6 REVERSE DOT-BLOT

Details of the use of a liquid handling laboratory robot for printing of RDBs were published by Lappin and coworkers (2001). RDBs are most often arrayed so that the normal oligonucleotide probes sit on a line adjacent to one another while variant probes are not on the same line. This permits rapid interpretation of heterozygotes where one normal and one variant allele are indicated (see also Chapter 2). However, where two mutations are close to each other and fall within the sequence of the oligonucleotide probe (a situation often observed with respect to hemoglobin Hb A, Hb S, and Hb C), a different pattern is obtained. DNA from individuals homozygous for Hb S that have sickle cell disease will not hybridize to the normal probe at either the S or C position of the blot. Instead, such DNA hybridizes to both the mutant S and mutant C probes. Similar RDB results are observed in the case of variants neighboring the p.F508del, such as p.I507del or p.I506V. Confounding neighboring variants are also present in β-thalassemia diagnosis where IVS I-1 G > T (c.92 + 1G > T) and IVS I-6 T > C (c.92 + 6T > C) mutations, five nucleotides apart, can each contribute to disease. Nevertheless, even though RDBs possess limitations, they provide a widely used and versatile mutation detection platform.

In early 2000, the American College of Medical Genetics and the American College of Obstetrics and Gynecology recommended CF carrier testing for all US Caucasian couples that anticipate children (Grody *et al*., 2001). Since that time, two CF line blot assays that provide mutation detection have become popular methods in US molecular diagnostics laboratories. The CF LAp, offered by Roche Diagnostics (Indianapolis, IN, USA) and Innogenetics (Gent, Belgium) line probe assay (LiPA), have been well received by molecular diagnostics laboratories. Recently, a Johns Hopkins group has evaluated a 58 allele line probe assay for CF, and discussed its analytic utility in a variety of diagnostic situations (Wang *et al*., 2002). Limitations, such as individuals who are compound heterozygote at closely spaced loci, occasionally fail to signal the presence of one or the other allele because of interference. As mentioned above, it is one characteristic of all sequence-specific assays that nucleotide variants within the probed region (usually approximately 17 nucleotides, although some new chemistries permit probes approximately 13 nucleotides and still maintain specificity) affect test accuracy. An example of this kind of interaction

is detection of a p.S549/p.R553X heterozygote by Wang and coworkers (2002). This failed to hybridize with the p.G551D wild-type probe encompassed by the probe for p.R553X. Another example is a p.F508del mutation/p.I506V polymorphism heterozygote that failed to hybridize against the normal sequence in the region. Confounding results should notify an alert diagnostician that a given test uncertainty bears further investigation.

As mentioned above, RDBs are versatile and might be printed for the detection of a variety of genetic lesions. Efforts to commercialize RDB hybridization for CF has motivated many laboratories to engage in home-brew testing via reverse allele-specific oligonucleotides for a variety of hereditary illness and infectious disease typing. In spite of certain limitations, this technology is useful for the reproducible detection of any point mutation or small deletion or insertion in any amplified genomic sample. The general utility of RDBs might allow mechanized production of screening strips for aldolase B mutations, causing hereditary fructose intolerance (Lau and Tolan, 1999), nondeletion α-thalassemia (Chan *et al*., 1999), or adult onset mitochondrial disorders, such as Leber hereditary optic neuropathy (Schollen *et al*., 1997) or even hepatitis A contamination in food (Jean *et al*., 2001). A recently published RDB assay for congenital adrenal hyperplasia has high clinical utility in identifying sexually ambiguous newborns (Yang *et al*., 2001). As discussed above, the RDB is flexible, inexpensive to implement and uses off-the-shelf commercially available hardware, reagents and software.

34.6.1 Commercialization of the Reverse Dot-Blot

Commercialization of RDB strips required significant effort in ensuring their reproducibility and uniformity. During the period of scale-up at Roche Molecular Systems, it was found that dot-blots themselves often gave equivocal results in the form of a halo around the dots. This had more to do with surface tension considerations when applying amino-conjugated or BSA-conjugated oligonucleotides (printing) than it had to do with any other single consideration. As a consequence, a decision was made to convert the dots to lines of uniform width. Automation for printing the line blots was first developed at Ismeca USA, Inc. (Carlsbad, CA, USA), which never found an adequate market to maintain production of their Bio-Line Dispenser. Commercial RDB printing equipment is currently available from Bio-Dot, Inc. (Irvine, CA, USA). Roche developed a Linter Striper and a MultidispenseR2000 Controller (IVEK, N. Springfield, VT, USA) to coat nylon backed sheets. Stephen Lappin, Jeffrey Cahlik, and the author described the techniques used for RDB automated printing at Quest diagnostics during 2001 (Lappin *et al*., 2001). Extensive improvements in reproducibility and manufacture accompanied the commercialization process at Roche Molecular Systems and production of strips at Quest. Incubation temperature, temperature equilibration and uniformity in conjugate and substrate distribution are known to be critical variables for color or luminescence detection. Both Tecan (Maennedorf, Switzerland) and Dynal Biotech (Oslo, Norway) manufacture specially engineered incubators and chemical dispenser devices that automate part of the RDB hybridization and developing process (Tecan's ProfiBlot is picture in Fig. 34.1b). While RDB assays are generally limited by an inability to detect large or quantitative deletions and an inability to characterize all but modestly expanded repeat sequences (ascertainment of the exact size of an expansion is often required for accurate molecular diagnosis or binning), these strips can provide a means of accurate and reproducible genotype assignments. Automated spotting or line blotting of RDB strips allows the printing of large numbers of these with a minimum of operator intervention. Automation also permits higher density arraying. Optimum conditions for nucleotide hybridization and development of the RDB strips were devised and automated through the use of Tecan's ProfiBlot (Fig. 34.1b) and Dynal's AutoReli. These shaking baths give uniformity to the strip development that is not easily achieved in any other fashion.

34.7 5′ NUCLEOTIDASE (TAQMAN) ASSAYS

David Gelfand and his coworkers invented the 5′ nucleotidase assay on the basis of his observation that Taq polymerase retained some 5′ endonuclease activity even when its $3' > 5'$ proofreading activity was blocked, as in Amplitaq Gold (see also Chapter 7). These researchers took advantage of this activity to modify PCR so that amplification, conducted in the presence of a probe labeled with a label on its 5′ end, permits visualization of a perfectly hybridized sequence. At first, Gelfand's group used a radioactive label, but colleagues at Applied Biosystems and Genentech modified a Fluorescent Energy Transfer (FRET) method, dual labeling oligonucleotide probes with a 3′ quencher and a 5′ fluorescent reporter to detect perfectly hybridized oligomers (Heid *et al*., 1996). On a FRET probe, the fluorescent moieties on the probe differ in their excitation and emission spectra, such that one quenches the other when in close proximity. The fluorescent molecule selected as the quencher remains attached to the oligonucleotide probe before, during, and after the assay. The reporter is liberated during the PCR extension and allows the measurement of increasing fluorescence throughout the assay. By conducting the PCR in optically clear plastic tubes irradiating these during the assay at the excitation wavelength of the reporter, the increasing fluorescence of a positive sample can be visualized during (or shortly after) the test. A small modification of the assay allows two alleles to be discerned

in the test sample. Instead of using only one probe, with one fluorescent and one quencher molecule, two differently labeled oligonucleotides – one with fluorescent moiety A and a one base pair difference with an oligonucleotide with an alternative fluorescent moiety B – can be used simultaneously in the same tube (Morin *et al*., 1999). One of these detects the mutation while the other detects the normal gene sequence. Through application of this strategy, it is possible to visualize the presence of normal, homozygous mutant, or heterozygous samples in a single tube. Drawbacks of the technique include the difficulty of optimizing assay conditions, the expense of the complicated synthesis of TaqMan probes, and the cost of readers required for visualizing the results. In spite of these disadvantages, this method has been adopted in several high-throughput genotyping facilities.

34.8 CAPILLARY THERMAL CYCLER

The Light-cycler was originally a circulating hot air thermal cycler equipped with 32 experimental positions that can be excited and monitored in real time, distributed by Idaho Technologies. During 2000, Roche Diagnostic Systems purchased this technology and distributed it for the following 5 years, while engaged in research and development for a newly formatted LightCycler™. Thermal cycling is very rapid, with a standard 30-cycle reaction completing in less than half an hour (Wittwer *et al*., 1989, 1990, 1997; Wittwer and Garling, 1991; Ririe *et al*., 1997). For homogeneous analysis using the 5′ nucleotidase technology, the dynamic range of the detection circuit in the capillary tube instrument often failed to detect reactions easily quantitated in Applied Biosystems TaqMan instruments (ABI 7900 HT). However, the LightCycler was unique in that it permitted production of a melting curve of hybridized fluorescent-labeled probe annealed to amplified genomic DNA. Thus, alleles not easily visualized using 5′ nucleotidase reactions could be resolved on the LightCycler™. In short order after Roche entered the TaqMan instrumentation marketplace, the 5′ nucleotidase platforms (ABI 7700; ABI 7900) were updated to permit visualization of melt curves using a 5′ GC-clamp strategy, with three probes, Stoffel Fragment and SybrGreen in the reaction mixes. This permitted visualization of alleles through synthesis of relatively inexpensive oligonucleotide primers and a melting strategy. Then, in 2005, Roche Diagnostic Systems released a new version of the LightCycler™ that involved a 96- or 384-micro-well plate format, was versatile enough to visualize 5′ nucleotidase reactions, and could also perform melting curve analysis (see also Chapter 15). This Roche LightCycler™ 480 instrument has improved optical visualization characteristics and provides an incremental advance in this field.

34.9 ELECTRONIC HYBRIDIZATION

Newer technologies began to make their way into the hybridization automation marketplace during the first few years of the new century. Nanogen (San Diego, CA) developed a CF assay on their proprietary, silicon chip-based platform (Sosnowski *et al*., 1997). The chip consisted of a silicon microchip with 100 electronically addressable sites. A hydrogel layer overlying the chip was embedded with streptavidin. Electronic addressing was used to concentrate DNA to a particular test site(s) on the microarray, enabling sequential addition of biotinylated oligonucleotides, in the capture down mode, or biotinylated samples, in the amplicon down mode, to different test sites. Interrogation and genotyping were accomplished by hybridization of fluorescently labeled reporter probes. Laser excitation and a charge-coupled device (CCD) camera were used to monitor hybridization. The Nanogen technology was a leap forward in automation of DNA hybridization because it permitted electronic alteration of hybridization specificity. However, the equipment required a sizable capital investment in specialized hardware and relatively expensive cartridges. Perhaps for this reason, the Nanogen ended production of the hardware in 2006 and will end consumable support in the next few years, which will render this platform of historical value only.

34.10 HYBRIDIZATION ARRAYS

Affymetrix and its HapMap partner, Perlegen, have been photolithographically attaching oligonucleotides to quartz enabling the production of dense gene interrogation microarrays for the past ten years (Pease *et al*., 1994). Quartz is used because of its natural hydroxylation properties. To begin the process, quartz is washed, to provide a uniform surface, and submerged into silane, which reacts with the hydroxyl groups to form a matrix of covalent linker molecules. The intramolecular distance between silanes determines the probe's packing density. Affymetrix and Perlegen have improvised proprietary methods to permit synthesis of arrays holding over 500,000 probe locations, also known as ''features'' within a 1.28 square ''coverslip''. Linker molecules, attached to the silane matrix, provide a surface that may be spatially activated by light. Probe synthesis occurs in parallel, permitting simultaneous nucleotide addition of A, C, T, or G nucleotide to the ''features''. Definition of which oligonucleotide chains will receive a nucleotide in each step is accomplished through the use of photolithographic masks, each with $18\mu^2$ to $20\mu^2$ windows corresponding to the dimensions of individual features. The use of ultraviolet light to irradiate the mask in the first step of synthesis permits exposed linkers to become de-protected and available for nucleotide coupling. Chrome marks on the wafer ensure accurate mask alignment during the

de-protection step. After activation, the wafer surface is flooded with a solution containing a single type of deoxyribonucleotide with a removable protection group. Nucleotides attach to exposed (unmasked) activated linkers.

Occasionally, activated molecules fail to attach the new nucleotide. To prevent these aberrant chains from compromising chip quality, a capping step is used to truncate them. In addition, branched oligonucleotide formation is prevented through specific side chain blocking chemistry. Iterative synthesis steps involve masking, de-protection, and coupling. The process is repeated until the probes reach approximately 25 nucleotides. Although each position in the sequence of an oligonucleotide is occupied by 1 of 4 nucleotides, resulting in an apparent need for $25 \times 4 = 100$ different masks per wafer, the synthesis process is optimized to minimize masking requirements. This is done through the use of algorithms that adjust probe growth kinetics and the informatic identification of probe products where the same mask can be used multiple times.

At synthesis conclusion, the wafers are de-protected, diced, and the resulting individual arrays are packaged in flow cell cartridges. Depending on the number of probe features per array, a single 5 inch square quartz wafer can yield between 49 and 400 arrays.

Subsequent to the manufacturing process, both Affymetrix and Perlegen do a comprehensive series of quality control tests including sampling arrays from every wafer to run control hybridizations. A quantitative test of hybridization is also performed using standardized control probes.

Application of Affymetrix arrays to the detection of SNPs is accomplished through a series of labeling steps (Fan *et al*., 2000). First, approximately 250 μg of total genomic DNA is digested with restriction enzyme XbaI for the 100K SNP array, or NspI and Sty1 for the larger 500K or Affy 6.0 array, and ligated to adapters provided by Affymetrix that recognize the cohesive 4 bp 5′ overhang. There is no purification between digestion and adapter ligation so both large and small restriction fragments are ligated with adapters. PCR primers that contain the adapter sequence are then used to amplify the fragments in the 250 to 1,000 bp size range. These molecules are then labeled in a reaction that consists of a proprietary biotin-labeled end labeling reagent and terminal deoxynuclotidyl transferase. This incubation produces the labeled molecules that are hybridized to the GeneChip array. Subsequent to hybridization and washing in the instrument that Affymetrix has engineered for this purpose, chips are flooded with phycoerythrin-conjugated streptavidin, which binds to the biotinylated amplicons attached to the matrix, and permits visualization in an elaborate scanner supplied by the company.

The Illumina BeadArray is composed of tens of millions of 3-micron silica beads assembled on a modified microscope slide (examples shown in Fig. 34.3c). Each bead is photolithographically printed with hundreds of thousands of specific oligonucleotide sequences that act as capture probes for a single oligonucleotide address moiety. Separately, patient DNA is subjected to whole genome amplification, PCR, and thermophilic ligase treatment in order to specifically label hundreds to millions of SNPs through the use of three primers: one primer that is at a fixed sequence in close proximity to the SNP, a second primer that provides base extension if and only if the SNP is represented as the major allele, and one primer that provides base extension if and only if the SNP is represented as the minor allele. As the SNP region is extended, a sequence homologous to the bead array address is co-amplified. When the beads are hybridized with the amplification products, they become specifically fluorescent. The BeadArray is subjected to post-hybridization washing and visualization in a proprietary BeadArray detector called an iScan system. This reader is controlled through proprietary Laboratory Integration Management System (LIMS) and interpretative software

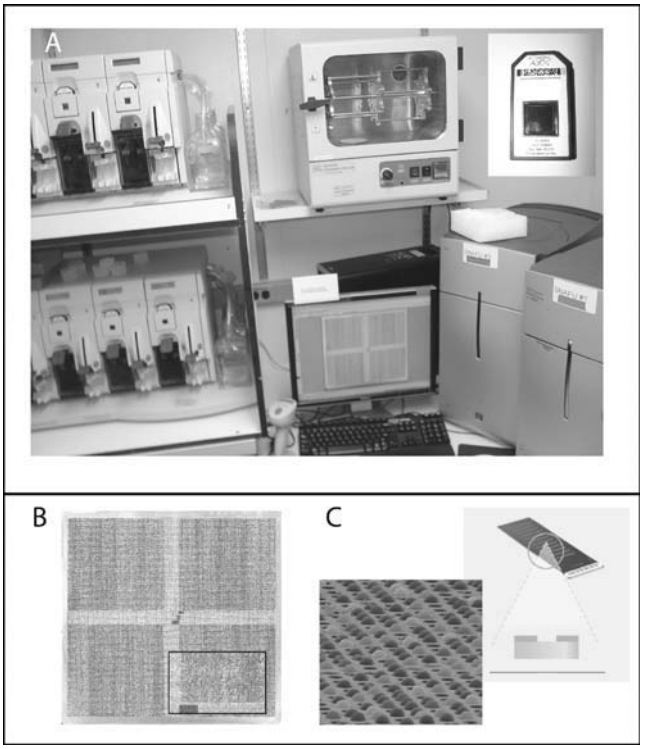


FIGURE 34.3 Affymetrix and Illumina Microarray Genotyping Systems. **A.** The entire Affymetrix system as it sits on a benchtop. From left to right, microarray washer, controller computer below and microarray hybridization oven above, and two microarray readers on the right. Upper inset, one Affymetrix version 6.0 SNP microarray cartridge. **B.** Visualization of the Affymetrix microarray 6.0 using the proprietary reader software. The cross pattern denotes the perfect match and mismatch QA probes included on the Affymetrix version 6.0 microarray. The inset provides a magnification of the visualization of the upper right-hand quadrant near the QA probe border. Internal signal dilution probes are features that are also part of microarray QA/QC. **C.** Upper right, a schematic view of an Illumina BeadArray containing genotyping slide. As described in the text, the BeadArrays are read in a proprietary reader that calibrates the fluorescence on particular tags co-amplified with SNPs during a single base pair extension PCR and ligation.

called BeadStudio. BeadStudio permits the interpretation of BeadArray address sequences into nucleotide calls.

The past five years have witnessed a revolution in the capability and resolution of the Affymetrix gene chip arrays. Several iterations of the platform have been marketed and supplanted in a short time frame. The most current configuration of an Affymetrix GeneChip array is the 6.0, which consists of 950,000 SNP features, of which approximately 720,000 are scoreable, and 950,000 features that are intended to assess copy-number variation. Several mathematical algorithms for interpretation of the fluorescent signal into base pair genotype calls have been proposed and tested. These include an early algorithm still used for judging genotyping quality, named the Dynamic Model (DM) which involves using both perfect match and mismatch probes, of which there are a limited number on the 6.0 chip; and later, the CRLMM and BRLMM algorithms, by which interpretation of base calls is rapid and mainly accurate. Similar algorithms have been implemented by Illumina in their BeadStudio software.

34.11 MICRO-WELL PLATE ARRAYS

Chemical attachment of amino-conjugated oligonucleotide probes to plastic micro-well plates is now a well-established methodology. This is the system used by Roche for popular micro-well plate assays for HIV and HCV viral load (Mulder *et al*., 1994). It has also been used by several groups for Factor V Leiden mutation detection (Zehnder *et al*., 1997; Kowalski *et al*., 2000). In brief, the method consists of attachment of amino-conjugated oligonucleotides to BSA-coated micro-well plates or plates coated with N-oxysuccinimide esters; such esters react with the nucleophilic primary amine conjugates. Once coupling is accomplished, plates are extensively washed before being used for hybridization capture of biotinylated gene-specific amplicons. After post-hybridization washing, subsequent incubation of the plate with a streptavidin conjugated with alkaline phosphatase permits visualization using tetramethylbenzidine or acid-tetramethylbenzidine chemistry. Finally, plates containing visualized amplicons are optically scanned using a micro-well plate reader and interpreted using appropriate software. These assays are efficient and low cost compared with PCR-RFLP or PCR-RFLP coupled with probe hybridization; however, they have two major drawbacks. These are the extensive manipulation of post-PCR products, because they are non-homogeneous and the requirement for preparation and validation of appropriate capture plates.

34.12 MICROARRAY PRINTING

Now that microarray printing and visualization has become routine in many laboratories (Schena, 1999), former constraints preventing indel or gene copy-number detection via microarray have diminished (Beheshti *et al*., 2002; Yu *et al*., 2003; Fiegler *et al*., 2003). As a consequence, the ability to gauge gene copy-number, genomic deletions, or the presence of amplified regions has become possible through the printing of oligonucleotide arrays. These may be printed on glass slides or coverslips through the use of microarray preparation robots. Competitive hybridization between normal and patient genomic DNA labeled with Cy3 and Cy5 through sample amplification permits visualization using a microarray scanner and consequent label ratio determination. Metrics for comparative genomic hybridization (CGH; see also Chapters 10 and 12), developed by Pinkel and coworkers (1998), can then be used to determine genomic regions that have undergone amplification or deletion. In addition, careful control of hybridization conditions, probe synthesis, and attachment should permit SNPs scoring at large numbers of sites using this technology. This kind of ''home-brew'' microarray manufacture for genetic disease research and diagnosis is still in its infancy, in part because of the expense of oligonucleotide synthesis; however, these methods should provide some healthy competition for the suppliers of still relatively expensive photolithographically manufactured genomic chips discussed earlier.

34.13 SUMMARY

Genotyping assay efficiency and throughput has considerably improved as a consequence of automation. Automation of many of the subtasks of nucleic acid extraction, quantitation, aliquotting, amplification, and hybridization along with electrophoresis, and homogeneous signal detection has enabled the establishment of higher-throughput molecular assays and diagnostics. Process automation has been motivated in part by the successful effort to linkage map, sequence, and then provide a haplotype map (HapMap) for the human genome. Specific subtask automation has been accompanied by manual chaining together so that each process, for example extraction, amplification, hybridization, and detection, are dovetailed. There are incremental gains in clinical laboratory efficiency through automated process monitoring, result interpretation, and reporting afforded by coupling this automated task management with laboratory management information systems. In a research environment, chaining manual processes with automated subtasks, such as robotic liquid handling, has permitted high-throughput genomic sequencing, genotyping, and voluminous data production. Many instrument and reagent manufacturers have participated in the automation initiative. There will undoubtedly be many improvements in automation and detection of DNA hybridization in the short-term future.

ACKNOWLEDGEMENTS

The content of this publication does not necessarily reflect the views or policies of the Department of Health

and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. This chapter has been funded in whole or in part with federal funds from the Center for Cancer Research, National Cancer Institute, and the National Institutes of Health.

REFERENCES

Alderton, R.P., Kitau, J., and Beck, S. (1994). Automated DNA hybridization. *Anal. Biochem.* 218, 98–102.

Beck, S. (1995). Automated DNA hybridization and detection. In *Molecular Biology: Current Innovations and Future Trends* (Griffin, A.M., and Griffin, H.G., eds), Vol. 2, pp. 83–91. Horizon Scientific Press, Wymondham, UK.

Beheshti, B., Park, P.C., Braude, I., and Squire, J.A. (2002). Microarray CGH. *Methods Mol. Biol.* 204, 191–207.

Chan, V., Yam, I., Chen, F.E., and Chan, T.K. (1999). A reverse dot-blot method for rapid detection of non-deletion alpha thalassaemia. *Br. J. Haematol.* 104, 513–515.

Church, G.M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA,* 81, 1991–1995.

Church, G.M., and Gilbert, W. (1985). The genomic sequencing technique. *Prog. Clin. Biol. Res.* 177, 17–21.

Cler, L., Bu, D., Lewis, C., and Euhus, D. (2006). A comparison of five methods for extracting DNA from paucicellular clinical samples. *Mol. Cell. Probes,* 20, 191–196.

Davidson, E.H., and Hough, B.R. (1969). High sequence diversity in the RNA synthesized at the lampbrush stage of oogenesis. *Proc. Natl. Acad. Sci. USA,* 63, 342–349.

Davis, R.W., and Davidson, N. (1968). Electron-microscopic visualization of deletion mutations. *Proc. Natl. Acad. Sci. USA,* 60, 243–250.

DeMarchi, J.M., Richards, C.S., Fenwick, R.G., Pace, R., and Beaudet, A.L. (1994). A robotics-assisted procedure for large scale cystic fibrosis mutation analysis. *Hum. Mutat.* 4, 281–290.

Dracopoli, N.C., Haines, J.L., Korf, B.R., Morton, C.C., Seidman, C.E., Seidman, J.G., and Smith, D.R. (2000). *Current Protocols in Human Genetics*. Wiley, New York, NY.

Fan, J.B., Chen, X., Halushka, M.K., Berno, A., Huang, X., Ryder, T., Lipshutz, R.J., Lockhart, D.J., and Chakravarti, A. (2000). Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res.* 10, 853–860.

Farkas, D.H., Drevon, A.M., Kiechle, F.L., DiCarlo, R.G., Heath, E.M., and Crisan, D. (1996). Specimen stability for DNA-based diagnostic testing. *Diagn. Mol. Pathol.* 5, 227–235.

Fiegler, H., Carr, P., Douglas, E.J., Burford, D.C., Hunt, S., Scott, C.E., Smith, J., Vetrie, D., Gorman, P., Tomlinson, I.P., and Carter, N.P. (2003). DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer* 36, 361–374.

Gillespie, D., and Spiegelman, S. (1965). A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane. *J. Mol. Biol.* 12, 829–842.

Gold, B. (2001). DNA genotyping. *Adv. Clin. Chem.* 36, 171–234.

Gold, B. (2003). Origin and utility of the reverse dot-blot. *Expert Rev. Mol. Diagn.* 3, 143–152.

Gold, B., Radu, D., Balanko, A., and Chiang, C.S. (2000). Diagnosis of Fragile X syndrome by Southern blot hybridization using a chemiluminescent probe: a laboratory protocol. *Mol. Diagn.* 5, 169–178.

Grody, W.W., Cutting, G.R., Klinger, K.W., Richards, C.S., Watson, M. S., and Desnick, R.J. (2001). Laboratory standards and guidelines for population-based cystic fibrosis carrier screening. *Genet. Med.* 3, 149–154.

Groman, J.D., Meyer, M.E., Wilmott, R.W., Zeitlin, P.L., and Cutting, G.R. (2002). Variant cystic fibrosis phenotypes in the absence of CFTR mutations. *N. Engl. J. Med.* 347, 401–407.

Hawkins, M.E., Pfleiderer, W., Mazumder, A., Pommier, Y.G., and Balis, F.M. (1995). Incorporation of a fluorescent guanosine analog into oligonucleotides and its application to a real time assay for the HIV-1 integrase 3′-processing reaction. *Nucleic Acids Res.* 23, 2872–2880.

Heid, C.A., Stevens, J., Livak, K.J., and Williams, P.M. (1996). Real time quantitative PCR. *Genome Res.* 6, 986–994.

Hilbertm, H., Lauber, J., Lubenow, H., and Dusterhoft, A. (2000). Automated sample-preparation technologies in genome sequencing projects. *DNA Seq.* 11, 193–197.

Hopwood, A., Oldroyd, N., Fellows, S., Ward, R., Owen, S.A., and Sullivan, K. (1997). Rapid quantification of DNA samples extracted from buccal scrapes prior to DNA profiling. *Biotechniques* 23, 18–20.

Jean, J., Blais, B., Darveau, A., and Fliss, I. (2001). Detection of hepatitis A virus by the nucleic acid sequence-based amplification technique and comparison with reverse transcription-PCR. *Appl. Environ. Microbiol.* 67, 5593–5600.

Kowalski, A., Radu, D., and Gold, B. (2000). Colorimetric microwell plate detection of the factor V Leiden mutation. *Clin. Chem.* 46, 1195–1198.

Lander, E.S., and Botstein, D. (1986). Mapping complex genetic traits in humans: new methods using a complete RFLP linkage map. *Cold Spring Harb. Symp. Quant. Biol.* 51(Pt 1), 49–62.

Lappin, S., Cahlik, J., and Gold, B. (2001). Robot printing of reverse dot blot arrays for human mutation detection. *J. Mol. Diagn.* 3, 178–188.

Lau, J., and Tolan, D.R. (1999). Screening for hereditary fructose intolerance mutations by reverse dot-blot. *Mol. Cell. Probes* 13, 35–40.

Morin, P.A., Saiz, R., and Monjazeb, A. (1999). High-throughput single nucleotide polymorphism genotyping by fluorescent 5′ exonuclease assay. *Biotechniques* 27, 538–540. 542, 544 *passim*.

Mulder, J., McKinney, N., Christopherson, C., Sninsky, J., Greenfield, L., and Kwok, S. (1994). Rapid and simple PCR assay for quantitation of human immunodeficiency virus type 1 RNA in plasma: application to acute retroviral infection. *J. Clin. Microbiol.* 32, 292–300.

Ousterhout, J.K. (1994). *Tcl and the Tk Toolkit*. Addison-Wesley, Readingx Mass.

Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., and Fodor, S.P. (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. USA,* 91, 5022–5026.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B. M., Gray, J.W., and Albertson, D.G. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* 20, 207–211.

Ririe, K.M., Rasmussen, R.P., and Wittwer, C.T. (1997). Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal. Biochem.* 245, 154–160.

Schena, M. (1999). *DNA Microarrays: A Practical Approach*. Oxford University Press, Oxford; New York.

Schollen, E., Vandenberk, P., Cassiman, J.J., and Matthijs, G. (1997). Development of reverse dot-blot system for screening of mitochondrial DNA mutations associated with Leber hereditary optic atrophy. *Clin. Chem.* 43, 18–23.

Sosnowski, R.G., Tu, E., Butler, W.F., O'Connell, J.P., and Heller, M.J. (1997). Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control. *Proc. Natl. Acad. Sci. USA,* 94, 1119–1123.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503–517.

Strom, C.M., Huang, D., Chen, C., Buller, A., Peng, M., Quan, F., Redman, J., and Sun, W. (2003). Extensive sequencing of the cystic fibrosis transmembrane regulator gene: assay validation and unexpected benefits of developing a comprehensive test. *Genet. Med.* 5, 9–14.

Wang, X., Myers, A., Saiki, R.K., and Cutting, G.R. (2002). Development and evaluation of a PCR-based, line probe assay for the detection of 58 alleles in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. *Clin. Chem.* 48, 1121–1123.

Wittwer, C.T., and Garling, D.J. (1991). Rapid cycle DNA amplification: time and temperature optimization. *Biotechniques* 10, 76–83.

Wittwer, C.T., Fillmore, G.C., and Garling, D.J. (1990). Minimizing the time required for DNA amplification by efficient heat transfer to small samples. *Anal. Biochem.* 186, 328–331.

Wittwer, C.T., Fillmore, G.C., and Hillyard, D.R. (1989). Automated polymerase chain reaction in capillary tubes with hot air. *Nucleic Acids Res.* 17, 4353–4357.

Wittwer, C.T., Ririe, K.M., Andrew, R.V., David, D.A., Gundry, R.A., and Balis, U.J. (1997). The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques* 22, 176–181.

Yang, Y.P., Corley, N., and Garcia-Heras, J. (2001). Reverse dot-blot hybridization as an improved tool for the molecular diagnosis of point mutations in congenital adrenal hyperplasia caused by 21-hydroxylase deficiency. *Mol. Diagn.* 6, 193–199.

Yu, W., Ballif, B.C., Kashork, C.D., Heilstedt, H.A., Howard, L.A., Cai, W.W., White, L.D., Liu, W., Beaudet, A.L., Bejjani, B.A., Shaw, C.A., and Shaffer, L.G. (2003). Development of a comparative genomic hybridization microarray and demonstration of its utility with 25 well-characterized 1p36 deletions. *Hum. Mol. Genet.* 12, 2145–2152.

Zehnder, J.L., Benson, R.C., and Cheng, S. (1997). A microplate allele-specific oligonucleotide hybridization assay for detection of factor V Leiden. *Diagn. Mol. Pathol.* 6, 347–352.

This page intentionally left blank

Chapter 35

The Use of Microelectronic-Based Techniques in Molecular Diagnostic Assays

Piotr Grodzinski[1], Michael Ward[2], Robin Liu[3], Kathryn Scott[4], Saul Surrey[5] and Paolo Fortina[4,6]

[1] *Center for Strategic and Scientific Initiative, National Cancer Institute, Bethesda, MD;*

[2] *Invitrogene Corp, Eugene, OR, USA;*

[3] *Molecular Diagnostics, Osmetech Molecular Diagnostics, Pasadena, CA, USA;*

[4] *Department of Cancer Biology, Kimmel Cancer Center, Thomas Jefferson University, Jefferson Medical College, Philadelphia, PA, USA;*

[5] *Department of Medicine, Thomas Jefferson University, Jefferson Medical College, Philadelphia, PA, USA;*

[6] *Department of Experimental Medicine, University "La Sapienza", Rome Italy*

35.1 INTRODUCTION

Contemporary on-chip biological assays have benefited from the multitude of techniques developed during the microelectronic revolution of the last three decades. The ability of delivering electric and magnetic signals to the localized portions of the biochip permitted exploitation of fundamental dielectric and magnetic properties of biological molecules and cells. For example, most biological molecules exhibit an electric charge and thus they migrate when exposed to an electric field. The interaction between an external electric field and these molecules can be used for their manipulation and concentration. The differentiation in dielectric properties of cells led to the development of dielectrophoretic techniques allowing for sorting, separation, and isolation of cells from different organisms and tissues. The ability to monitor electron transfer between the molecule and the external measurement system formed the basic principles for electronic detection of DNA. Similarly, the vast arsenal of microfabrication capabilities developed in the microelectronic industry has been leveraged toward on-chip assay applications, where integration of conductive, heating, and magnetic elements resulted in development of micro-PCR devices, microarray DNA chips containing thermal gradients, and dielectrophoretic and magnetic devices for cell and molecule preconcentration. This chapter will include a discussion on microfabrication, cell sorting, molecule preconcentration, and sample preparation techniques followed by DNA/RNA amplification on chips, electronic-assisted hybridization assays, and examples of commercially available platforms.

35.2 MICROFABRICATION

The engagement of microfabrication techniques permits for realization of high-throughput, integrated analytical systems. Use of microfabrication techniques allows for the building of miniature networks of chambers and channels with a dimension resolution <1 mm. Performing assays within these devices reduces reagent volume and facilitates reduction of assay duration (Kricka, 1998; Freemantle, 1999). In addition, high parallelism of reactions is achieved by placement of many assay paths within the same device; thus, analysis throughput can be increased (Fodor *et al*., 1991; Simpson *et al*., 1998). Furthermore, a multitude of assay functions can be integrated into the same device leading to assay automation.

Individual components of a microfluidic device are required to perform a variety of functions including mixing, thermal cycling, valving, and detection, which eventually will need to be integrated into one chip. There are a number of material platforms under consideration for fabrication of such devices including glass (Manz *et al*., 1992; Harrison *et al*., 1993), silicon (Mastrangelo *et al*., 1998; Burns *et al*., 1998), plastic (Alonso-Amigo, 2000; Anderson *et al*., 2000; Grodzinski *et al*., 2001), and hybrid approaches (Burns *et al*., 1998). Devices made of glass and silicon both take advantage of lithographic techniques developed in the integrated circuit chip industry. They use photolithography, wet etching, metal sputtering, and anodic and thermal bonding. The glass-based platform is attractive due to its well-understood surface chemistry, and it possesses a surface charge, which allows for using electrokinetic pumping methods for

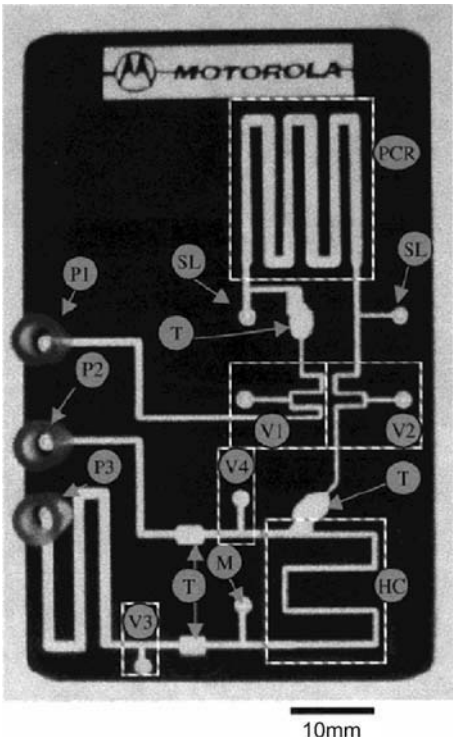


FIGURE 35.1 Monolithic, integrated DNA assay device produced using direct laser writing in 250 mm polycarbonate plates and boned through thermal bonding process. Serpentine PCR channel (PCR), hybridization channel (HC), pluronics valves (V1–4), pluronic traps (T), hydrophobic air permeable membrane (M), PCR reagent loading holes (SL), sample driving syringe pump P1, waste withdrawing syringe pump (P2), and wash syringe pump (P3) are indicated (adapted from Liu *et al*., 2002).

fluid transport. Glass is also transparent in the UV region of the spectrum, which simplifies detection. Silicon, on the other hand, permits for integration of complex complementary metal-oxide semiconductor (CMOS) electronics for control and data processing within a chip.

Plastics offer an attractive alternative for fabrication of disposable devices due to low cost (particularly important for disposable devices) and a wide range of advantageous physical and chemical properties, including very good biocompatibility. Bulk micromachining of plastics can be achieved through hot embossing, injection molding, and casting. Rapid prototyping techniques taking advantage of direct writing in polymer plates are also available and permit production of complex structures (see Fig. 35.1). Due to the distinct advantages of various chip fabrication techniques, hybrid approaches with integrated electronics on silicon substrates with glass or plastic channels are also being pursued (Burns *et al*., 1998).

35.3 CHIPS FOR SAMPLE PREPARATION

Sample preparation represents the most time-consuming and labor-intensive procedure in DNA analysis, and also introduces one of the largest variables in subsequent analyses due to its complexity. Due to these reasons, initial efforts in bringing the assay to the chip level were focused on development of back-end detection schemes resulting in demonstrations of DNA microarray biochips (Fodor *et al*., 1991; Fortina *et al*., 2000) and electrophoresis separation chips (Manz *et al*., 1992; Harrison *et al*., 1993; Woolley and Mathies, 1994). However, the quest for a complete, integrated analytical chip system requires the treatment of ''real'' samples of bodily or environmental fluids. Recent research attempts use methods of separating cells from complex samples. Two of these methods are discussed in the following sections, which in the authors' opinion are most promising. These methods utilize dielectrophoresis and immunomagnetic separation.

35.3.1 Use of Dielectrophoresis

Immersion of a biological cell into a medium of different dielectric property, and subsequent application of an external non-uniform electric field, causes cell polarization and induction of charges at the cell/medium interface (Wang and Cheng, 2001). The dipole moment develops due to separation of positive and negative charges within the cell. Dielectrophoretic force (DEP) originates from interaction of this dipole moment with an external non-uniform electric field. Depending on the relationship of dipole moment and electric field vectors (directions), the DEP force can be positive (parallel vectors) or negative (anti-parallel vectors), causing motion of the cell in different directions (Pethig *et al*., 1992). These vector relationships depend on the relative dielectric property of the cell and the medium. The former depends on the cell type and physiological state. Negative effects on development or behavior of cells were not observed either during or after application of electrical fields (Pethig *et al*., 1992).

The DEP technique does not require cell labeling, and uncomplicated devices containing only interdigitated electrodes can be built. When non-uniform AC electric field is used, cells migrate to locations of minimum dielectric potentials and get immobilized there. Furthermore, differences in dielectric properties between cells allow for tuning of electric field conditions so that only single cell types from a complex mixture are immobilized.

For example, DEP separation techniques, which exploit differential dielectric properties among different biological cells, have been developed. Cultured cervical carcinoma (HeLa) cells and *E. coli* cells were separated and isolated from normal blood cells (Cheng *et al*., 1998a,b) on a silicon chip (see Fig. 35.2). The NanoChip™, developed by Nanogen, contains a two-dimensional silicon-based array of the platinum electrode sites. These sites can be addressed individually, such that AC electric field can be directed to the site of choice. Cells, bacteria, or other molecules can be

FIGURE 35.2 NanoChip™ cartridge and electronic microarray. The latest generation NanoChip™ cartridge is used on the automated Molecular Biology Workstation developed by Nanogen. Each microarray has 100 test sites.

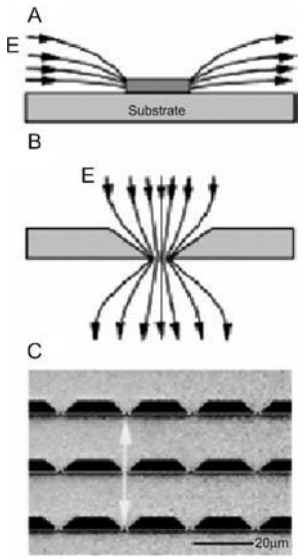


FIGURE 35.3 Schematic of a microfluidic DEP trap. **A.** A metallic DEP trap is made of microfabricated wire(s) on a substrate. The wire(s) may be either free floating or connected to a voltage source. **B.** An electrode-less DEP trap made of dielectric constrictions. The solid lines, indicated by E, are electric field lines. **C.** A scanning electron micrograph of an electrode-less DEP device consisted of a constriction array etched in quartz. The constrictions are 1 mm wide and 1.25 mm deep. The whole chip measures 1×1 cm. The applied electric field direction z is shown by the double-headed arrow (adapted from Chou *et al*., 2002).

moved rapidly and concentrated to the appropriate address in the NanoChip electronic microarray. A permeation layer, usually porous agarose or polyacrylamide, allows an electrical contact between the electrode and cells or molecules in solution, while preventing physical contact, which may lead to undesired oxidation or denaturing.

As described in Cheng and colleagues (1998b), separation of HeLa cells from blood cells took approximately 3 min after addition of the mixture. Furthermore, a chip also was designed to integrate subsequent sample processing and DNA/RNA hybridization (Cheng *et al*., 1998b). The DEP separation of bacteria containing a plasmid from blood cells, electronic lysis of isolated cells, and digestion of proteins were performed. Presence of plasmid sequence was detected by hybridization, and the entire process took approximately 4 min to complete after the mixture was added to the chip. Lysis was performed using a series of electronic pulses, and all positions had a proteinase K mixture added to digest proteins released during lysis. Plasmid and genomic RNA and DNA were released from the cells as a result of the lysis. Hybridization was performed on another chip, requiring removal of the nucleic acids. The hybridization chip contained a permeation layer of streptavidin agarose with biotinylated capture probes at specific addresses. Each target sequence was electronically moved to the separate addresses and bound as expected to the proper fluorescent capture probe.

A diversity of dielectrophoretic techniques resulted in modifications of simple cell arrest approaches and produced more sophisticated operation schemes, such as Field-Flow Fractionation (FFF) DEP and traveling-wave DEP (Gascoyne and Vykoukal, 2002). The former is used to separate cells in the vertical direction within a flow channel and combines variation of flow velocity due to parabolic distribution of the flow front in pressure-driven flow and DEP. The latter involves application of an AC electrical field in a sequential manner to the set of electrodes enabling simultaneous concentration and motion of the cells.

DEP devices also can be used for focusing (Edman *et al*., 1998), preconcentration, and selective entrapment of DNA fragments (Chou *et al*., 2002). The latter approach used an electrode-less DEP chip configuration where electrodes are removed to the end of fluidic channels, and DEP force is created due to crowding of electric field lines within the microfabricated channel constriction (see Fig. 35.3). The advantage of this approach lies in prevention of DNA hydrolysis when it is brought to the vicinity of the metal electrode.

35.3.2 Use of Magnetic Devices

Molecule and cell labeling with particles or beads, which could exhibit magnetic properties, are common in molecular biology. The magnetic beads can then be manipulated, transferred, and immobilized in a magnetic field. Resulting molecule or cell separation and sorting can then occur. Incorporating chip elements enabling selective production of such fields allows for localized formation of these fields. Traditional cell separation protocols rely predominantly on immunological methods utilizing labeling of cells with specific antibodies, which are usually covalently linked to particles, for example paramagnetic beads (Chalmers *et al*., 1998a, b). The bead particles can be trapped in an external magnetic field, and a subset of cells bound to the beads can

be isolated via an antibody. This method carries substantial promise for separation of rare cells occurring at low concentrations, due to its potential for high selectivity. Thus, developing chips capable of producing magnetic fields and magnetic gradients and enabling isolation of magnetically labeled cells within the chips at predetermined locations is of great importance. A method to effectively produce strong magnetic gradients is important while capturing small (<150 nm) beads, which possess too weak a momentum to be trapped in uniform fields.

Introduction of complex fabrication processes to fabricate on-chip integrated microcoils has been successful, and has facilitated an alternative approach instead of use of external magnets for bead arrest. Meander-type magnetic coils have been demonstrated using a planar magnetic conductor and two-level magnetic core (Ahn and Allen, 1993). Such coils can be used directly to drive magnetic actuators. Hybrid processes, where wire bonding combined with microfabrication techniques have been used, result in coils (Christenson *et al*., 1995). Microfluidic structures with manually wound coils used for bead trapping and coil-to-coil bead transport have also been evaluated (Joung *et al*., 2000). Shallow channels need to be used to maintain gradient strength in the bead flow field.

Methods to produce high magnetic gradients for cell separation have been discussed (Tibbe *et al*., 1999, 2002; Berger *et al*., 2001; Ward *et al*., 2002), and devices of approximately 1 mm wide and 100 nm tall sputtered ferromagnetic lines used to create very high local gradients at the edges of the lines. A cell-tracks device (Immunicon Corp., Hundington Valley, PA) uses nickel magnetized in an external field (Tibbe *et al*., 1999, 2002), whereas another uses permanently magnetized cobalt-chrome-tantalum materials (Berger *et al*., 2001). When labeled cells are brought in close proximity (50 mm) to the lines, they can be captured by the magnetic force resulting from the edge gradients. The Immunicon device (see Fig. 35.4) achieves this proximity using a specialized external gradient magnet while the other uses shallow flow channels. The Immunicon device is engineered for optical observation of trapped cells in a CD-like configuration capable of laser scanning; the second is designed for cell sorting and uses hydrodynamic forces in the channel to direct captured cells along the wires to collection points.

Ward and colleagues (2002) developed microfluidic channels with built-in gradients using submillimeter iron or nickel-iron matrix elements fabricated into acrylic and polycarbonate substrate plastic channels. Those metal features geometrically concentrate external magnetic flux to form the necessary magnetic gradients for capture of weakly susceptible particles. Targets are collected at the walls of the channel where they can be washed or otherwise processed in a flow stream. These first generation devices were used to concentrate *E. coli* cells from blood and employed double-sided iron embedded polycarbonate channels 100 mm in height. Plating experiments of *E. coli* inoculated blood demonstrated that large sample volumes could be concentrated in minutes with relatively high recovery rates. About 50% recovery was obtained at the maximum tested average flow rate of approximately 19 mm/sec, which corresponded to an approximately 350 ml/min flow rate. For capture of rare target mammalian cells from blood, alternative, more uniform and more easily manufactured designs using compression-molded, saw-toothed ridges in acrylic, which are coated with soft magnetic material, were also explored (see Fig. 35.5). A 90% cell capture efficiency using a 120 ml/min flow rate was achieved.

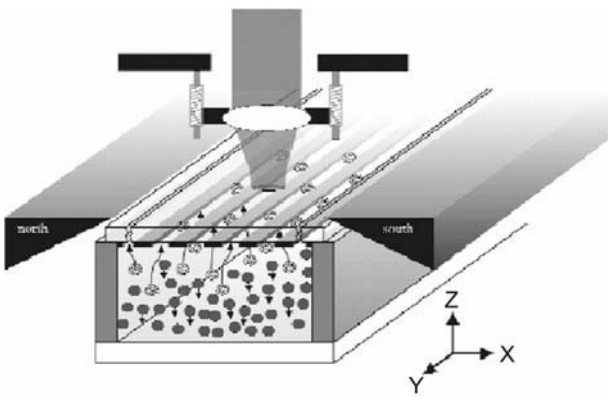


FIGURE 35.4 Schematic of Celltracks device and method. The chamber with microfabricated nickel lines is placed in the external magnet device. Labeled cells in the chamber are drawn upward where they align along the nickel. They can then be analyzed by the scanning optics (adapted from Tibbe *et al*., 1999).

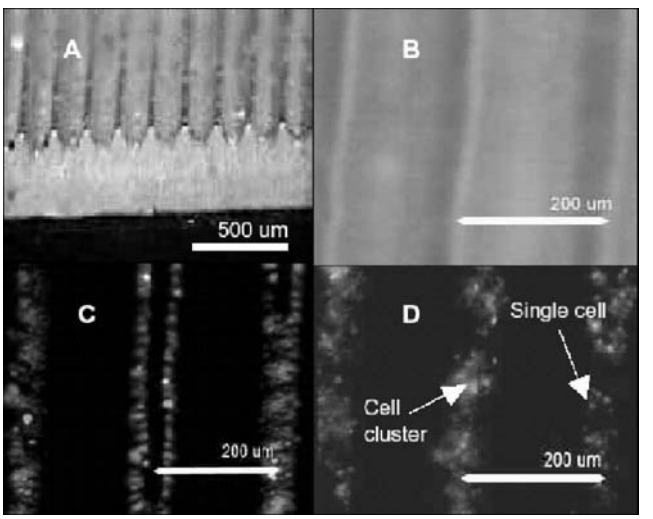


FIGURE 35.5 **A.** Laser-machined acrylic ridge structure prior to electroplating. **B.** Ni-Fe ridge device with fluorescein-labeled CD45 antibody coupled to IgG1-specific magnetic beads prior to separation. **C.** Same device after separation. Fluorescence is focused on the ridges. **D.** Similar device after separation of white blood cells using the fluorescent anti-CD45 antibodies (adapted from Ward *et al*., 2002).

35.3.3 Cell Separation by Size

Physical filters relying on separation of cells by size have been demonstrated (Wilding *et al*., 1998). White blood cells were isolated from whole blood in silicon-glass 4.5 ml microchips containing a series of 3.5 mm feature-sized weir-type filters, formed by an etched silicon dam spanning the flow chamber. Genomic DNA targets, for example, the dystrophin gene, then were directly amplified using PCR from cells isolated on the filters. This dual function microchip provides a means to simplify nucleic acid analyses by integrating in a single device two key steps in the analytical procedure: cell isolation and PCR (Cheng *et al*., 1996, 1998c; Wilding *et al*., 1998).

35.4 DNA AND RNA AMPLIFICATION IN MICROCHIP FORMAT

DNA amplification using polymerase chain reaction (PCR) is essential to most genetic analysis applications of integrated microchips. Rapid operation, small sample volume, and parallel amplification of different amplicons within the same chip are among the desired features of amplification in the microchip environment. Design and development of such chips is hindered by several challenges including loss of sample to chamber walls due to dramatic increases in surface-to-volume ratio, evaporation in small volumes, and effective heat dissipation in order to achieve rapid thermal cycling. Micro-PCR devices have been fabricated in glass (Kopp *et al*., 1998; Waters *et al*., 1998; Lagally *et al*., 2000, 2001), silicon (Wilding *et al*., 1994; Belgrader *et al*., 1998, 1999), and plastic (Boone *et al*., 1998; Yu *et al*., 2000; Kricka, *et al*., 2002). Silicon, due to its superior thermal conductivity (approximately 10× that of glass and 700× that of polymers), allows for very fast temperature ramp times and results in short on-chip protocols (Belgrader *et al*., 1999). Recently, with increasing emphasis on disposable devices, use of plastic and plastic fabrication methods have become popular in microreactor development. Such chips are inexpensive, optically transparent, and biocompatible (Alonso-Amigo, 2000; Grodzinski *et al*., 2001). Despite all the advantages, plastic possesses a major challenge to a designer of PCR microreactors due to its poor thermal conductivity and resulting difficulty in achieving rapid thermal cycling. Recently, successful DNA amplification in polycarbonate chips was reported (Yang *et al*., 2002). Thirty thermal cycles took 30 min, which, considering the poor thermal conductivity characteristics of polymers, is a significant achievement.

Low-volume operation and rapid thermal cycling have been early motivators for on-chip PCR development. An infrared heating scheme with glass PCR devices containing a 1.7 ml microchamber was described, in which amplification was achieved by 15 thermal cycles in 4 min (Giordano *et al*., 2001). In an integrated monolithic silicon-glass device, a submicroliter (280 nl) volume was thermally cycled as fast as 30 sec/cycle (Lagally *et al*., 2001). The low volume limit for micro-PCR was further reduced by using a picoliter microchamber array (Nagai *et al*., 2001). Using a real-time PCR device, PCR detection in 7 min of *Erwinia*, a vegetative bacterium, was reported (Belgrader *et al*., 1999). An integrated rapid PCR-detection system with amplification times in the 20 min range coupled with capillary electrophoresis analysis also was presented (Khandurina *et al*., 2000; Lagally *et al*., 2000).

Despite all these examples, sensitivity of micro-PCR chips has not been well studied. Most reports focused on achieving amplification *per se*, but the systematic evaluation of amplification yield and reaction sensitivity was not usually explored. The initial template concentration used to achieve fast and small volume amplification was usually high, ranging from 0.1 ng of phage DNA (Giordano *et al*., 2001) to 100 ng of human genomic DNA (Cheng *et al*., 1998c). The most sensitive micro-PCR assay demonstrated so far involved use of a single molecule DNA template, which was amplified in a glass-integrated microfluidic device (Lagally *et al*., 2001; Fig. 35.6). For ''real'' sample (containing target cells, rather than purified DNA) analysis, the most sensitive silicon microstructure, which could perform rapid real-time PCR analysis from a sample containing a low target concentration (*Erwinia*), was reported (Belgrader *et al*., 1999). A positive amplicon signal was detected in less than 35 cycles (17 sec/cycle) with the

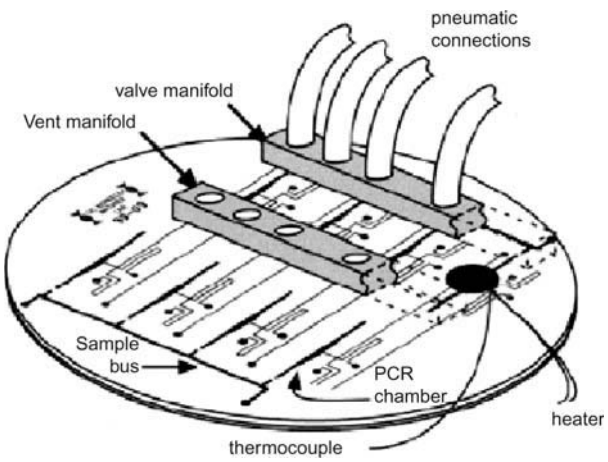


FIGURE 35.6 Microfluidic PCR-CE device. The PCR chambers are connected to a common sample bus through a set of valves. Hydrophobic vents at the other end of the PCR chambers are used to locate the sample and to eliminate gas. The PCR chambers are connected directly to the cross channel of the CE system for product injection and analysis. Two aluminum manifolds, one each for the vents and valves, are placed onto the respective ports and clamped in place using vacuum. The manifolds connect to external solenoid valves for pressure and vacuum actuation. Thermal cycling is accomplished using a resistive heater and a miniature thermocouple below the 280 nL chamber (adapted from Lagally *et al*., 2001).

starting template concentration as low as five cells. A systematic study on sensitivity of micro-PCR in plastic devices was reported for the first time (Yang *et al*., 2002), demonstrating feasibility of amplifying template concentrations as low as ten *E. coli* cells (50 fg of DNA) in the presence of blood. Similarly, multiplexing PCR was demonstrated within the same reactor chamber for four different bacteria species.

Investigators have worked on sample preparation using real-time PCR or hybridization in an array format as an assessment tool. An integrated, monolithic genetic assay device performing serial and parallel multistep molecular operations, including nucleic acid hybridization, was presented (Anderson *et al*., 2000). More recently, a system with online PCR detection, on-chip spore preconcentration, lysing, and DNA purification was demonstrated (Taylor *et al*., 2001). In addition, a prototype integrated semidisposable microchip analyzer for cell separation and isolation, PCR amplification, and amplicon detection was developed (Yuen *et al*., 2001). The analyzer includes three glass-silicon microchips (an integrated sample preparation and PCR microchip, a denaturation microchip, and a reaction microchip) and a detection microarray made from a standard glass slide. The microchips and microarray are connected through a microfluidic network. The individual chips are assembled into a working system using three platforms: an upper plexiglas, a disposable lower plexiglas, and an aluminum platform.

Plastic, disposable chips for pathogen detection, PCR amplification, DNA hybridization, and hybridization in a single device were demonstrated (Liu *et al*., 2002). DNA probes were immobilized in plastic channels to reduce target volume and to enable target motion to improve detection sensitivity. On-chip valving using phase-change Pluronics material was also implemented to facilitate separation of different stages of the assay. The chip was sandwiched between external Peltier heaters during PCR amplification, and sample transport was accomplished using external syringe pumps. The level of integration was expanded further to a complete self-contained biochip capable of magnetic, bead-based cell capture, cell preconcentration, purification, lysis, PCR amplification, DNA hybridization, and electrochemical detection of hybridization events (Liu *et al*., 2003). The device is completely self-contained and does not require external pressure sources, fluid storage, mechanical pumps, or valves (see Fig. 35.7). It uses phase-transition paraffin valves and electrochemical pumps for efficient sample transport and isolation between adjacent assay stages.

It should be noted that a wide range of flow rates needs to be used, since large sample volumes (approximately 2 ml) have to be processed in short periods of time (<10 min). Pathogenic bacteria preconcentration and subsequent amplification and detection of genomic DNA as well as single nucleotide polymorphism (SNP) analysis directly from blood was reported using this system.

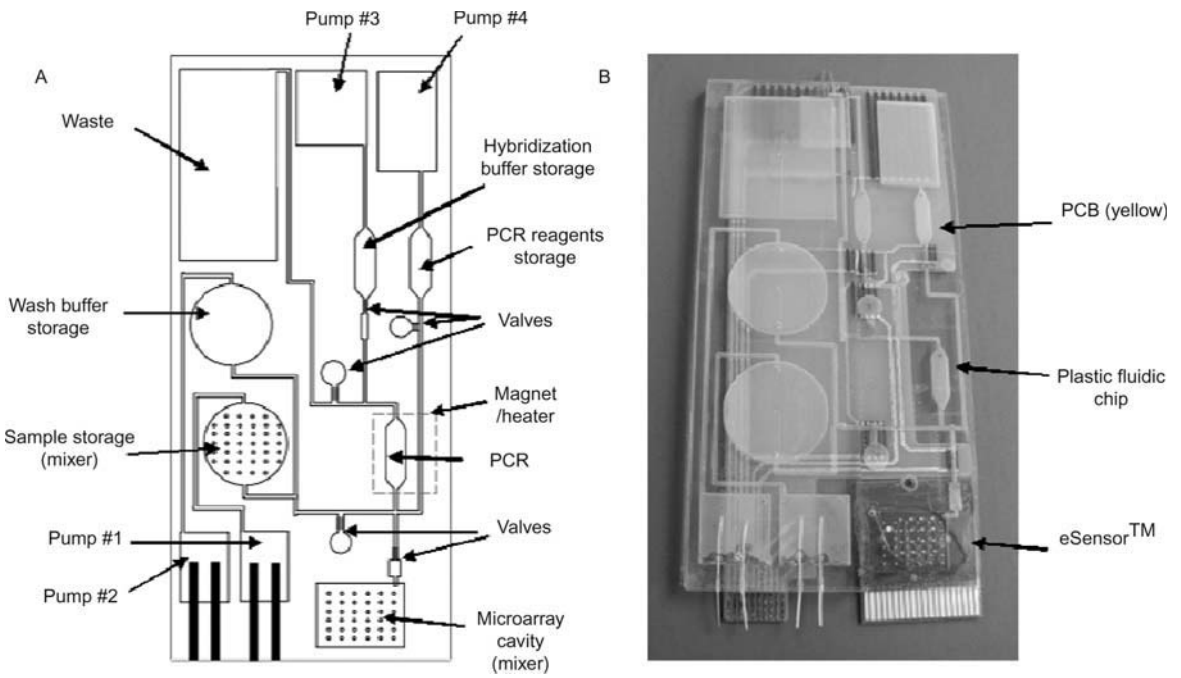


FIGURE 35.7 **A.** Schematic of the plastic fluidic chip. Pumps 1–3 are electrochemical pumps; pump 4 is a thermo-pneumatic pump. **B.** Photograph of the integrated device that consists of a plastic fluidic chip, a printed circuit board (PCB), and a Motorola eSensor™ microarray chip. The piezoelectric disks glued on the top of the sample storage chamber and microarray chamber that provide acoustic micromixing are not shown here (adapted from Liu *et al*., 2003).

Quantitation of mRNA abundance in a micro-PCR format remains an additional challenge. Single-step RT-PCR using silicon dioxide-coated chips has been reported with an efficiency of 70% compared to that from a tube control. Analyses were conducted using an entangled-solution capillary electrophoresis (ESCE) system and laser-induced fluorescence (LIF) detection (Cheng and Mitchelson, 2001). Similarly, Anderson and colleagues (2000) developed chips, which, among others, performed reverse transcription reactions. Recently, PCR and RT-PCR in one chip using a continuous flow scheme was demonstrated (Obeid, 2003).

35.5 COMMERCIAL IMPLEMENTATION OF MOLECULAR ASSAYS WITH THE USE OF MICROELECTRONICS

Many of the research approaches just discussed have matured and grown to be commercialized, taking advantage of microfabrication and microelectronic technologies. In the following sections, a few primary examples will be discussed, including:

- Hitachi high-technology microarray chips with thermal gradients
- Nanogen approach to electronic focusing of DNA in hybridization microarrays
- Motorola–Clinical Micro Sensors hybridization chips using electrochemical detection for SNP analysis
- CombiMatrix microarrays with *in situ* probe synthesis and electronic detection for immunoassay and DNA applications
- GeneOhm detection and direct DNA conductivity measurements
- Hybridization arrays with magnetic sensing
- Cepheid technology for real-time micro-PCR amplification and sample preparation
- Caliper capillary electrophoresis technology for DNA and protein separation

35.5.1 Thermal Gradient Chips

With the increasing complexity and size of the contemporary microarrays, the determination of optimum conditions for multiple sequences to be analyzed on the same chip is becoming increasingly difficult. Template/primer and temperature differences do not allow all sequences to have similar performances when being hybridized simultaneously. Several groups have tried to optimize primer length, melting temperature, GC content, buffer concentrations, wash conditions, and hybridization temperature in attempting to find one set of optimal hybridization conditions (Khrapko *et al*., 1991; Sosnowski *et al*., 1997; Gerry *et al*., 1999; Gilles *et al*., 1999; Lipshutz *et al*., 1999; Fan *et al*., 2000; Hirschhorn *et al*., 2000; Mei *et al*., 2000; Pastinen *et al*., 2000; Dong *et al*., 2001; Jobs *et al*., 2002). These multiple attempts, which did not lead to an ultimate success, triggered a new research direction where chip platforms allowing for the independent control of the temperature at each individual array site were constructed (Kajiyama *et al*., 2003). This independent temperature control of hybridization events allowed for multiple genes and sequences to be assayed together. Genotyping SNPs in two genes was demonstrated with this newly developed chip, and four clinically relevant loci were selected for testing: two in the Factor VII gene and two in the hemochromatosis gene (Kajiyama *et al*., 2003).

A standard lithography process has been employed to design and fabricate the chip containing 100 individually controlled sites. Each site was 500×500 microns in size and contained a heater and temperature sensor. A diagram of the gradient chip, an individual reaction site diagram, and images of the individual sites for reaction are shown in Fig. 35.8. Heat dissipation fins were added to the chip after initial testing to minimize any thermal effects one site may have on neighboring sites.

The thermal gradient chip has proven to be very effective in genotyping, as it allowed for multiple loci with different probe melting temperatures (T_m) to be typed on the same chip. The melting/hybridization temperature of a sequence is determined by several factors, including the GC content and size of the PCR product. A few degrees difference in T_m can determine success or failure of a successful analysis of the sequence of interest, providing that hybridization or wash conditions are not optimized for the individual locus. The GC content of the four loci studied varied from 43% to as high as 71%, and PCR product size ranged from 150 to 390 bp, and T_m from 51 to 65°C. A Cy5 label was added to each PCR product to detect hybridization of the chip-bound normal or variant oligonucleotides. As expected, heterozygous samples hybridized to both chip-bound normal and mutant probes. The ability to control the temperature for hybridization for each locus facilitated genotyping of multiple loci simultaneously on the same chip. Such thermal gradient chips help minimize the number of ambiguous calls, making them potentially attractive for clinical laboratory testing.

35.5.2 Use of Electric Field to Accelerate Hybridization

Although microarrays are making a large impact on genomics, disease diagnosis, and biology in general, a significant problem still exists – hybridization is slow. Just as with Southern blotting, the solution containing the molecules to be hybridized with those attached to the surface typically takes 16 to 24 hours. Furthermore, the sample must be sufficiently concentrated in these molecules to permit hybridization even at these extended times. For example,

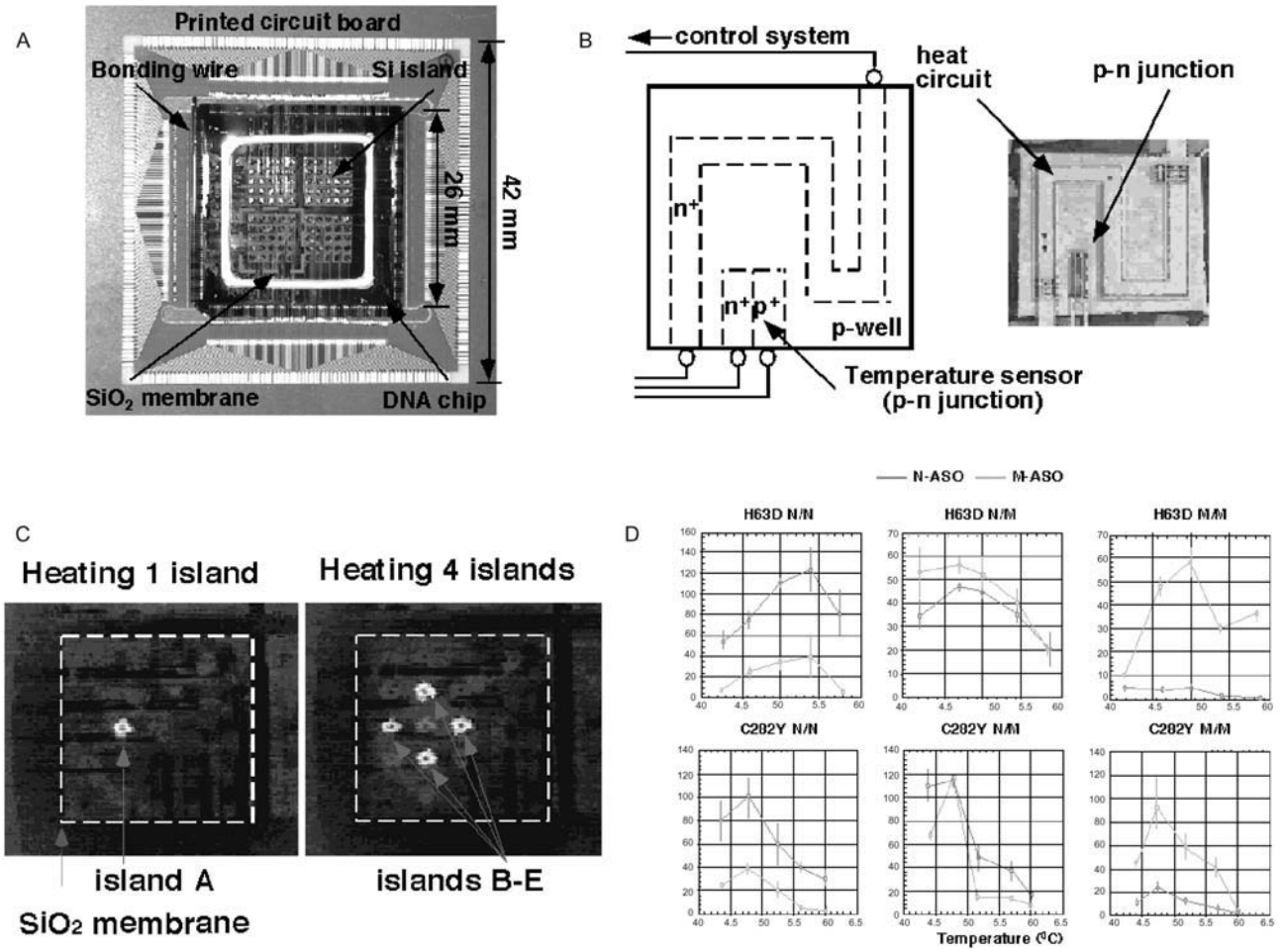


FIGURE 35.8 Thermal gradient chip images. **A.** Completed thermal gradient chip wire bonded to a small printed circuit board for testing. The silicon islands and supporting silicon dioxide support structure are visible in the center. **B.** Location and size of the heating and thermal sensing regions within a single island with photographic details. **C.** When four islands surrounding an unheated island are heated (right), their combined effect on the temperature of the central island is much greater than when a single island is heated (left). **D.** Genotype analyses for two SNPs in the gene for hemochromatosis: fluorescence units (y-axis) as a function of chip island temperature (x-axis) are shown for hybridization of Cy 5-labeled, denatured PCR, solution-phase targets to chip-bound normal (N-ASO, red) or mutant (M-ASO, green) oligonucleotide probes. Genomic DNA from either heterozygous (N/M, middle panels), wild-type (N/N, left panels), or homozygous mutant individuals (M/M, right panels), encompassing either the H63D (upper three panels) or C282Y (lower three panels) SNPs in the gene for hemochromatosis, were amplified by PCR. Samples then were heat denatured and annealed to chips containing different thermally isolated islands. Numbers in brackets refer to the ratio of intensities of annealing of labeled targets to normal (red)/mutant (green) probe spots on the array (adapted from Kajiyama *et al*., 2003). *Note*: The e-book for this title, including full-color images, is available for purchase at www.elsevierdirect.com.

concentration of about 4 nanomolar DNA is usually employed to achieve reasonably complete hybridization of 150 bp probes in 10 hours. Recommended protocols for mRNA use 2 microgram or (assuming size of 1,000 to 10,000 bp) roughly 19 to 190 nM mRNA. Lower concentrations simply do not produce a measurable signal in any reasonable period of time. The reason for the relatively large difference is that larger molecules diffuse much more slowly than smaller ones.

Due to these deficiencies, numerous attempts to accelerate the hybridization process have been undertaken. They include target stirring or mixing with the hybridization cavity, dynamic DNA hybridization using paramagnetic beads (Fan *et al*., 1999), rotation of the whole device (Chee *et al*., 1996), and the use of a micro porous three-dimensional biochip with the hybridization solution being pumped continuously through it (Cheek *et al*., 2001).

One of the most effective hybridization kinetics acceleration methods developed to date is to use an electric field to attract and concentrate DNA molecules at the chip surface.

NanoChip™ microelectronic arrays use electronic field control to drive the transport and concentration of nucleic acids to specific test sites within the array. The currently available array comprises 100 microelectrode test sites, each of which can be controlled individually and electronically to carry a positive, negative, or neutral charge. This flexibility differentiates NanoChip™ microelectronic arrays from other array platforms and enables analysis of multiple

patient samples and/or multiple SNPs on a single array (Foglieni *et al*., 2004).

Nanogen's (http://www.nanogen.com) electronic arrays are composed of silicon microchips coated with a synthetic hydrogel permeation layer containing streptavidin. By applying a positive charge to one or more test sites on the array, DNA, which is negatively charged, is attracted to those positions. Incorporation of a biotin moiety onto the DNA allows the sample to remain anchored at the test site after the electronic charge has been discontinued.

To perform a standard SNP genotyping assay, amplified, biotinylated genomic DNA is electronically addressed to one or more test sites. After all samples have been applied, fluorescently labeled reporter oligonucleotides that discriminate between the wild-type and polymorphic alleles are hybridized to the DNA, indicating the genotype of each sample. This entire process is highly automated; subsequent to sample preparation, the electronic addressing, hybridization, fluorescence scanning, and data analysis steps are performed on the NanoChip™ Molecular Biology Workstation.

The Nanogen approach has been effective in reducing the hybridization duration events to several minutes; however, the design of these chips remains complex and thus chip cost is still high.

Investigators at the University of Pennsylvania (Su *et al*., 2002) employed the electric field to attract DNA target to the chip surface, while drastically changing the chip design. They used standard glass slides coated commercially with a layer of indium and tin oxide (ITO) as the conductive material, which acts as an electrode. ITO is transparent and is used as the active material in liquid crystal displays (LCDs). It was found that standard methods for attaching oligonucleotides to a glass surface (silanization with an amino silane, reaction with a difunctional isothiocyanate, and reaction with an amino oligonucleotide) worked well with ITO surfaces. After oligonucleotide spots are attached to the conductive surface of one slide, a small amount of solution containing the complementary oligonucleotide is placed on the surface inside a thin gasket (typically around 25 microliters) and the second glass slide with electrode facing the solution is clamped on. A low voltage (<1 V) is applied between the electrodes such that the array side is at positive potential as compared to the second electrode. After a fixed time, the cell is disassembled and the hybridization is quantified with Cy5 dye labeling and laser scanning. Results are expressed as relative fluorescence intensity, defined as the ratio of (spot signal – background signal) to (reference spot signal – background signal).

A series of studies with PCR products of different sizes were performed to evaluate the improvement in hybridization times using this technique (also known as electric field-enhanced hybridization, or EFEH). The enhancements from about 40 to 200 times for DNA in the size ranges considered have been measured. In addition to improving speed of hybridization, EFEH showed an improved discrimination between perfectly matched and single-base mismatched hybrids. In this case, the field is applied in the reverse direction for a short time (sec) following the application in the forward direction to drive target to the surface. The bond strengths for imperfect matches are lower than for perfect ones and therefore may be removed more easily by the reverse field.

The main value of the proposed EFEH technique is its simplicity. It uses glass slides (versus the more expensive silicon ones) and it moves away from individual addressing of the hybridization sites. Thus, fabrication of these devices is much less complex and less expensive. By eliminating the individual addressing scheme, this method loses some flexibility.

35.5.3 Electronic Detection of DNA Hybridization: Clinical Microsensor Technology

Electronic detection of DNA hybridization has been developing rapidly, since it offers potential toward further miniaturization of DNA diagnostics (Wang, 1999, 2000). Use of electrochemical detection schemes obviates the need for large and expensive fluorescent scanners. The detection chips can be produced inexpensively using Si- or printed circuit board-based fabrication technologies. Most importantly, detection schemes not requiring external labels can be devised, further simplifying signal collection and enabling studies of reaction kinetics.

Clinical Micro Sensors (http://www.cms.com) has developed electrochemical detection (Farkas, 1999; Umek *et al*., 2001), low-density hybridization arrays for diagnostic applications, in particular, detection of SNPs. This so-called eSensor™ device is composed of a printed circuit board (PCB) chip and a plastic cartridge that is attached to the PCB chip using double-sided adhesive tape to form an 80 microliter reaction chamber. The PCB chip consists of an array of gold electrodes modified with a multicomponent, self-assembled monolayer (SAM) that includes presynthesized oligonucleotide (DNA) capture probes that are covalently attached to the electrode through an alkyl thiol linker. When a sample solution containing target DNA is introduced into the detection cartridge, specific capture probes on an electrode surface encounter complementary DNA from the sample and hybridization occurs. Two ferrocene, electronic labels (capture probe and signaling probe) bind the target in a sandwich configuration.

Binding of the target sequence to both the capture probe and the signaling probe connects the electronic labels to the surface through the chains of molecular wires built into SAMs (Creager *et al*., 1999). This adds a circuit element to the bioelectronic circuit on that electrode, and presence of hybridized (double-stranded) DNA can be detected using

alternative current voltammetry (ACV). Since the incoming target itself is not labeled, the washing step (to remove excessive, non-bound target) prior to the signal collection is not required. A continuous monitoring of the binding process with a quantitative measurement of the target accumulation is possible.

35.5.4 Hybridization Microarrays with *in situ* Probe Synthesis: CombiMatrix Technology

The use of on-chip *in situ* probe synthesis provides an attractive alternative to hybridization microarrays fabricated using presynthesized oligonucleotides and their physical deposition on the substrate. Chips with *in situ* probe synthesis are much more flexible to build, and valuable in the experiments where new sequences need to be introduced frequently. A method relying on photolithography and combinatorial chemistry allowing for parallel synthesis of probes nucleotide-by-nucleotide has been developed (Fodor *et al*., 1991). The use of photolabile chemistry allows for deprotection of the linker under the UV light exposure and attachment of the nucleotide. The fabrication costs of the photolithographic mask sets are, however, fairly high. Therefore, this technology has been modified by using Digital Light Processor (DLP) to develop the Maskless Array Synthesizer (MAS), which delivers focused light beams to selected portions of the substrate, thus eliminating need for the photomask (Nuwaysir *et al*., 2002).

Another step toward miniaturization of diagnostic assays can be undertaken when electrochemical synthesis of probes is implemented. CombiMatrix (http://www.combimatrix.com) introduced electrochemical *in situ* synthesis using electronic CMOS devices (Dill *et al*., 1999; Oleinikov *et al*., 2003). These chips contain high-density arrays of individually addressable microelectrodes. A typical chip has 1,024 microelectrodes that are each 100 mm in diameter (see Fig. 35.9). Each microelectrode in the array can be addressed independently. An activated electrode generates protons, resulting in a local change of pH. In turn, a pH-controlled chemical reaction can occur in the vicinity of a given electrode. The CombiMatrix biochips are coated with proprietary approximately 1 mm thick porous reaction layer material, which is used for immobilization and synthesis of biomolecules used for subsequent binding of target. Covalent linkage of the molecules within the porous layer is accomplished using reagents that are generated *in situ* by the microelectrodes.

In immunoassay format, biotin is immobilized in the porous layer. Subsequently, biotin-labeled sites then can conjugate with streptavidin, which in turn can capture biotin-labeled proteins or antibodies. Sandwich immunoassays can be used for larger entities and competitive immunoassays for smaller molecules. The technique was used for detection of saxitoxin, BG spores, murine IgG, and salmonella, among others (Dill *et al*., 2001).

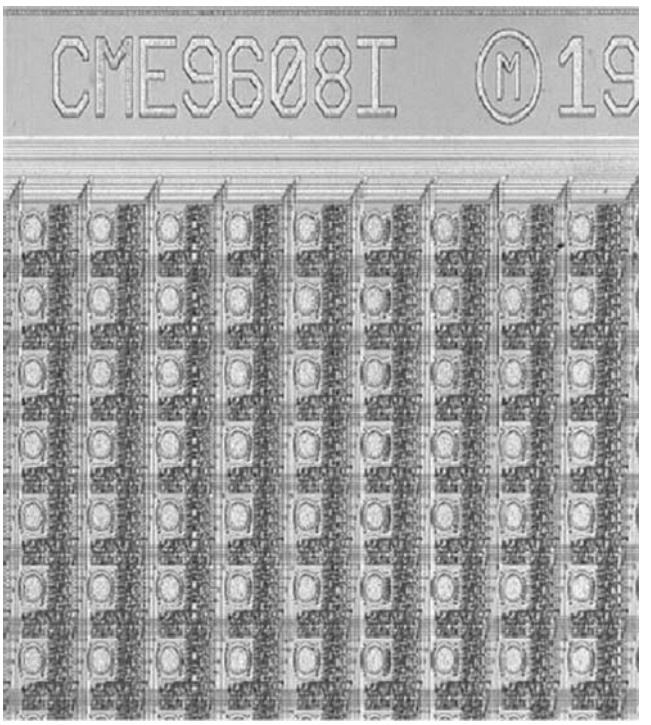


FIGURE 35.9 A section of the CombiMatrix CME9608 chip. Each of the small round disks is an electrode. Each electrode site is associated with CMOS circuitry for individually addressing and operating the electrode. CombiMatrix first generation CME9608 series chips have densities of over 1,000 sites per square centimeter. These chips are fabricated using a conventional 3-micron CMOS process. The next generation of chips will have over 6,000 sites per square centimeter. CombiMatrix also has special-purpose designs that achieve densities well over one million sites per square centimeter (adapted from Dill *et al*., 2001).

The use of phosphoramidite chemistry at the electrode sites, combined with sequential addressing of these electrodes in order to produce protons, results in the synthesis of oligonucleotides (Oleinikov *et al*., 2003). CombiMatrix demonstrated an *in situ* synthesis of 50-mers oligonucleotides. A complete chip with such probes can be prepared within 24 hours. Using this technique, CombiMatrix demonstrated SNP and gene expression assays. Furthermore, they also showed protein recognition using oligonucleotide arrays. Proteins under study were tagged with matching (to the one synthesized at the electrode site) oligonucleotide through streptavidin-biotin linker. Hybridization of oligonucleotides allowed immobilization of protein on the chip surface (Oleinikov *et al*., 2003).

At the initial stage of CombiMatrix development, detection was accomplished using fluorophore-tagged antibodies and epifluorescent microscopy. However, since the electrode sites on the chip are individually addressable, they can also be used to detect events occurring on the chip surface using electrochemistry (Oleinikov *et al*., 2003).

35.5.5 E-Chem Detection and Direct DNA Conductivity Measurement: GeneOhm

GeneOhm (http://www.geneohm.com) uses electron flow through a double helical DNA molecule to generate a current, which is dependent upon complementary base pairing within the helix. Briefly, an oligonucleotide spanning the region to be interrogated is deposited on a gold surface array. PCR target DNA from the patient is then heat denatured and annealed to the array. Perfect matches allow electrons to flow through the helix generating a measurable current, whereas single-base mismatches limit current flow. Multiple loci can be interrogated since each position on the array has an independent current sensor at each register, thereby enabling parallel analysis (Boon *et al*., 2000, 2002; Drummond *et al*., 2003).

35.5.6 Hybridization Arrays with Magnetic Sensing

The use of magnetic devices in biological assays is not limited to cell separation. They can also be used for detection of hybridization events, when the target DNA is labeled with a magnetic bead. Pioneering work by the Naval Research Labs on their Bead Array Counter (BARC) series demonstrated feasibility of this approach (Edelstein *et al*., 2000; Miller *et al*., 2001) using giant magneto-resistance (GMR) multilayer magnetic sensors. Departure from optical or electrochemical detection methods allowed detector size to be considerably reduced while maintaining high sensitivity. This size reduction can occur due to feasibility of integrating the magnetic sensor with electronics for signal acquisition on a single chip. Furthermore, this approach allows for improvement of reaction kinetics and control of hybridization stringency. The former is achieved through utilization of magnetic field for target preconcentration at the probe location, and the latter through use of an AC field to repel mismatched targets.

35.5.7 On-Chip Amplification and Sample Preparation: Cepheid Technology

The importance of microchip PCR developments was previously described, the major advantages being the speed of operation and reduced reagent volumes. Cepheid (http://www.cepheid.com) has built its technology around rapid recognition of bacterial and viral pathogens using DNA amplification in microchip format (Belgrader *et al*., 1998, 1999). Due to its market target applications within biodefense, food, and environmental testing, the company embarked on development of microdevices, which operate at rapid thermal cycling rates in order to reduce the PCR assay duration time to a minimum (Belgrader *et al*., 1999). The microchip format technology, originally developed at Lawrence Livermore National Laboratory, uses polypropylene, single-use reaction vials (volume of 25–100 microliters) equipped with a silicon ring heater for rapid thermal cycling. Silicon has a very high thermal conductivity and is a prime choice for a heater in such applications. The ramp rates achieved in solution are 10°C/sec for heating between 50 and 95°C and 2.5°C/sec for cooling at the same temperature range. In addition, to further reduce the duration of the assay, the chip is equipped with an optical window and online fluorescent detection to terminate thermal cycling as soon as the level of amplified signal reaches a detection threshold. This feature in combination with a modified two-temperature PCR protocol allowed for detection of *Erwinia* in 7 min (Belgrader *et al*., 1999).

A similar assay for *B. anthracis* starting with 100 pg of DNA can be completed in 13 min, and using 0.1 pg DNA required 19 min. The original chip solution just described has evolved into the Smart Cycler II and Smart Cycler II TD instruments (Belgrader *et al*., 2001), capable of amplification of 16 different samples at a time, with four-plex/well multiplexing capability and four-color fluorescence detection.

In order to facilitate sample preparation prior to DNA amplification, Cepheid developed a cell lysing unit. *B. anthracis* and *B. subtilis* spores are difficult to lyse using traditional thermal or chemical methods. They utilized sonication (Belgrader *et al*., 2000), which relies on the use of ultrasonic horn operating at approximately 40 kHz to deflect a flexible wall of the sample container at high impact. In order to improve lysis efficiency, purified spores are mixed with glass beads prior to sonication. There was also a need for the use of large volumes of starting sample (100 microliters to 1 microliters), since initial concentrations of pathogens under study may be very low. This need for analysis of large sample volumes led to development of sample preconcentration units.

The individual units for sample preparation and PCR-based DNA detection are designed in a modular format using an I-CORE platform. The unit subsequently is assembled into the analytical system under the GeneXpert product name. GeneXpert is designed to purify, concentrate, detect, and identify targeted DNA sequences, taking unprocessed sample to result in less than 30 min (Taylor *et al*., 2001).

35.5.8 LabChip: Caliper Technologies

Caliper Technologies Corporation (http://www.calipertech.com) designs and manufactures microfluidic chips for separating molecules. The chip incorporates electrokinetics and pressure to move molecules through microfluidic channels. For the analysis of cellular components, pressure is used for controlling movement of the cells through the chip. For DNA, RNA, and protein assays the charged molecules are moved through the channels using electrokinetics (Panaro *et al*., 2000).

In collaboration with Agilent Technologies, a platform was developed for the analysis of DNA, RNA, proteins, or cells. The Agilent 2100 BioAnalyzer allows researchers to use small amounts of their sample for analysis and to see data within 30 minutes of sample loading. The software provided with the instrument analyzes the results for each sample. Samples are added to wells and are fluorescently labeled with an intercalating dye. They move through the microfluidic channels and are injected into a separation chamber for analysis. The labeled fragments of DNA, RNA, or protein are separated by molecular sieving and detected by fluorescence. Samples are injected and analyzed individually. The entire time from sample loading to completion of the assay is approximately 30 minutes for 12 samples. The RNA 6000 NanoChip has become the standard method in laboratories for quantitating RNA used in gene expression analysis.

Cells are analyzed on the instrument using two-color fluorescence in combination with flow cytometry. The preparation of cells for flow cytometry analysis normally involves a lengthy staining and labeling step; however, cells are stained on the chip before analysis to allow for a faster workflow. The BioAnalyzer allows researchers to use fewer cells than in traditional flow cytometry assays. Developed assays and protocols for cell analysis are provided with the Cell Fluorescence LabChip Kit, but can be adjusted for researchers wishing to make changes to the protocol. Six samples will take approximately 25 minutes to be analyzed on the system (http://www.chem.agilent.com).

35.6 CONCLUSIONS

The convergence of molecular biology and high microelectronic technology has been unprecedented in the last decade. This bidirectional utilization and the cross-enrichment of these techniques are leading to the development of integrated multifunctional genetic analytical systems, which will have high sensitivity and specificity and will be cheaper and faster to operate as compared to their benchtop predecessors.

The ability to produce large arrays of identical and miniature elements using lithography led to the birth of high-density hybridization arrays and multichannel separation systems, and revolutionized high-throughput analysis in gene discovery, mutation studies, and SNP analysis. Furthermore, the microfabrication techniques allow for delivering and harvesting of electrical and magnetic signals at the distinct locations of the analytical chip with an ability for choosing individual addresses in large sensor arrays. These developments allow for a design of localized sensors analyzing single cells or molecules immobilized in distinct locations on the chip. It is expected that further progression in capability of these techniques will be achieved through the emerging field of nanotechnology.

REFERENCES

Ahn, C.H., and Allen, M.G. (1993). A planar micromachined spiral inductor for integrated magnetic microactuator applications. *J. Micromech. Microeng.* 3, 37–44.

Alonso-Amigo, G. (2000). Polymer microfabrication for microarrays, microreactors and microfluidics. *J. Assoc. Lab. Aut.* 5, 96–101.

Anderson, R.C., Su, X., Bogdan, G.J., and Fenton, J. (2000). A miniature integrated device for automated multistep genetic assays. *Nucleic Acids Res.* 28, E60.

Belgrader, P., Benett, W., Hadley, D., Long, G., Mariella, R., Jr, Milanovich, F., Nasarabadi, S., Nelson, W., Richards, J., and Stratton, P. (1998). Rapid pathogen detection using a microchip PCR array instrument. *Clin. Chem.* 44, 2191–2194.

Belgrader, P., Benett, W., Hadley, D., Richards, J., Stratton, P., Mariella, R., Jr, and Milanovich, F. (1999). PCR detection of bacteria in seven minutes. *Science* 284, 449–450.

Belgrader, P., Okuzumi, M., Pourahmadi, F., Borkholder, D.A., and Northrup, M.A. (2000). A microfluidic cartridge to prepare spores for PCR analysis. *Biosens. Bioelectron.* 14, 849–852.

Belgrader, P., Young, S., Yuan, B., Primeau, M., Christel, L.A., Pourahmadi, F., and Northrup, M.A. (2001). A battery-powered notebook thermal cycler for rapid multiplex real-time PCR analysis. *Anal. Chem.* 73, 286–289.

Berger, M., Castelino, J., Huang, R., Shah, M., and Austin, R.H. (2001). Design of a microfabricated magnetic cell separator. *Electrophoresis* 22, 3883–3892.

Boon, E.M., Ceres, D.M., Drummond, T.G., Hill, M.G., and Barton, J.K. (2000). Mutation detection by electrocatalysis at DNA-modified electrodes. *Nat. Biotechnol.* 18, 1096–1100.

Boon, E.M., Salas, J.E., and Barton, J.K. (2002). An electrical probe of protein-DNA interactions on DNA-modified surfaces. *Nat. Biotechnol.* 20, 282–286.

Boone, T., Hooper, H., and Soane, D. (1998). Integrated chemical analysis on plastic microfluidic devices. Hilton Head Island, South Carolina, June 8–11th.

Burns, M.A., Johnson, B.N., Brahmasandra, S.N., Handique, K., Webster, J.R., Krishnan, M., Sammarco, T.S., Man, P.M., Jones, D., Heldsinger, D., Mastrangelo, C.H., and Burke, D.T. (1998). An integrated nanoliter DNA analysis device. *Science* 282, 484–487.

Chalmers, J., Zborowski, M., Sun, L., and Moore, L. (1998a). Flow through immunomagnetic cell separation. *Biotechnol. Prog.* 14, 141–148.

Chalmers, J.J., Mandal, S., Fang, B., Sun, L., and Zborowski, M. (1998b). Theoretical analysis of cell separation based on cell surface marker density. *Biotechnol. Bioeng.* 59, 10–20.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S., and Fodor, S.P. (1996). Accessing genetic information with high-density DNA arrays. *Science* 274, 610–614.

Cheek, B., Matthew, A., Torres, P., Yu, Y., and Yang, H. (2001). Chemiluminescence detection for hybridization assays on the flow-thru chip, a three-dimensional microchannel biochip. *Anal. Chem.* 73, 5777–5783.

Cheng, J., Shoffner, M.A., Hvichia, G.E., Kricka, L.J., and Wilding, P. (1996). Chip PCR. II. Investigation of different PCR amplification systems in microfabricated silicon-glass chips. *Nucleic Acids Res.* 24, 380–385.

Cheng, J., Sheldon, E.L., Wu, L., Heller, M.J., and O'Connell, J.P. (1998a). Isolation of cultured cervical carcinoma cells mixed with

peripheral blood cells on a bioelectronic chip. *Anal. Chem.* 70, 2321–2326.

Cheng, J., Sheldon, E., Wu, L., Uribe, A., Gerrue, L., Carrino, J., Heller, M., and O'Connell, J. (1998b). Preparation and hybridization analysis of DNA/RNA from E. coli on microfabricated bioelectronic chips. *Nat. Biotechnol.* 16, 541–546.

Cheng, J., Waters, L.C., Fortina, P., Hvichia, G., Jacobson, S.C., Ramsey, J.M., Kricka, L.J., and Wilding, P. (1998c). Degenerate oligonucleotide primed-polymerase chain reaction and capillary electrophoretic analysis of human DNA on microchip based devices. *Anal. Biochem.* 257, 101–106.

Cheng, J., and Mitchelson, M. (2001). Capillary electrophoresis of nucleic acids. In *Practical Applications of Capillary Electrophoresis*, Vol. 2, pp. 211–219. Humana Press, Totowa, NJ.

Chou, C.-F., Tegenfeldt, J.O., Bakajin, O., Chan, S.S., Cox, E.C., Darnton, N., Duke, T., and Austin, R.H. (2002). Electrodeless dielectrophoresis of single- and double-stranded DNA. *Biophysical J.* 83, 2170–2175.

Christenson, T.R., Klein, J., and Guckel, H. (1995). An electromagnetic micro dynamometer. IEEE Conf., pp. 386–391. Amsterdam, The Netherlands.

Creager, S., Bamdad, C., MacLean, T., Lam, E., Chong, Y., Olsen, G.T., Luo, J., Gozin, M., and Kayyem, J.F. (1999). Electron transfer at electrodes through conjugated ''molecular wires'' bridges. *J. Am. Chem. Soc.* 121, 1059–1064.

Dill, K., Stanker, L.H., and Young, C.R. (1999). Detection of salmonella in poultry using a silicon chip-based biosensor. *J. Biochem. Biophys. Methods,* 41, 61–67.

Dill, K., Montgomery, D.M., Wang, W., and Tsai, J.C. (2001). Antigen detection using microelectrode array microchips. *Anal. Chim. Acta.* 444, 69–78.

Dong, S., Wang, E., Hsie, L., Cao, Y., Chen, X., and Gingeras, T.R. (2001). Flexible use of high-density oligonucleotide arrays for single-nucleotide polymorphism discovery and validation. *Genome Res.* 1, 1418–1424.

Drummond, T.G., Hill, M.G., and Barton, J.K. (2003). Electrochemical DNA sensors. *Nat. Biotechnol.* 21, 1192–1199.

Edelstein, R.L., Tamanaha, C.R., Sheehan, P.E., Miller, M.M., Baselt, D.R., Whitman, L.J., and Colton, R.J. (2000). The BARC biosensor applied to the detection of biological warfare agents. *Biosensors Bioelectr.* 14, 805–813.

Edman, C., Raymond, D., Wu, D., Tu, E., Sosnowski, R., Butler, W., Nerenberg, M., and Heller, M. (1998). Electric field directed nucleic acid hybridization on microchips. *Nucleic Acids Res.* 25, 4907–4914.

Fan, Z.H., Mangru, S., Granzow, R., Heaney, P., Ho, W., Dong, Q., and Kumar, R. (1999). Dynamic DNA hybridization on a chip using paramagnetic beads. *Anal. Chem.* 71, 4851–4859.

Fan, J.B., Chen, X., Halushka, M.K., Berno, A., Huang, X., Ryder, T., Lipshutz, R.J., Lockhart, D.J., and Chakravarti, A. (2000). Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res.* 10, 853–860.

Farkas, D.H. (1999). Bioelectronic detection of DNA and the automation of molecular diagnostics. *J. Assoc. Lab. Autom.* 4, 125–129.

Fodor, S.P., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., and Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* 251, 767–773.

Foglieni, B., Cremonesi, L., Travi, M., Ravani, A., Giambona, A., Rosatelli, M.C., Perra, C., Fortina, P., and Ferrari, M. (2004). Beta-thalassemia microelectronic chip: a fast and accurate method for mutation detection. *Clin. Chem.* 50, 73–79.

Fortina, P., Delgrosso, K., Sakazume, T., Santacroce, R., Moutereau, S., Su, H., Graves, D., Mckenzie, S., and Surrey, S. (2000). Simple two-color array-based approach for mutation detection. *Eur. J. Hum. Genet.* 8, 884–894.

Freemantle, M. (1999). Downsizing chemistry. *Chem. Eng. News.* 77, 27–36.

Gascoyne, P.R.C., and Vykoukal, J. (2002). Particle separation by dielectrophoresis. *Electrophoresis* 23, 1973–1976.

Gerry, N.P., Witowski, N.E., Day, J., Hammer, R.P., Barany, G., and Barany, F. (1999). Universal DNA microarray method for multiplex detection of low abundance point mutations. *J. Mol. Biol.* 292, 251–262.

Gilles, P.N., Wu, D.J., Foster, C.B., Dillon, P.J., and Chanock, S.J. (1999). Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. *Nat. Biotechnol.* 17, 365–370.

Giordano, B., Ferrance, J., Swedberg, S., Huhmer, A., and Landers, J. (2001). Polymerase chain reaction in polymeric microchips: DNA amplification in less than 240 seconds. *Anal. Biochem.* 291, 124–132.

Grodzinski, P., Liu, R.H., Chen, H., Blackwell, J., Liu, Y., Rhine, R., Smekal, T., Ganser, D., Romero, C., Yu, H., Chan, T., and Kroutchinina, N. (2001). Development of plastic microfluidic devices for sample preparation. *Biomed. Microdevices,* 3, 275.

Harrison, D.J., Fluri, K., Seiler, K., Fan, Z., Effenhauser, C.S., and Manz, A. (1993). Micromachining a miniaturized capillary electrophoresis-based chemical analysis system on a chip. *Science* 261, 895–897.

Hirschhorn, J.N., Sklar, P., Lindblad-Toh, K., Lim, Y.M., Ruiz-Gutierrez, M., Bolk, S., Langhorst, B., Schaffner, S., Winchester, E., and Lander, E.S. (2000). SBE-TAGS: an array-based method for efficient single-nucleotide polymorphism genotyping. *Proc. Natl. Acad. Sci. USA,* 97, 12164–12169.

Jobs, M., Fredriksson, S., Brookes, A.J., and Landegren, U. (2002). Effect of oligonucleotide truncation on single nucleotide distinction by solid-phase hybridization. *Anal. Chem.* 74, 199–202.

Joung, J., Shen, J., and Grodzinski, P. (2000). Micro pumps based on alternating high gradient magnetic fields. *IEEE Trans. Magnetics,* 36, 36. 2012.

Kajiyama, T., Miyahara, Y., Kricka, L.J., Wilding, P., Graves, D.J., Surrey, S., and Fortina, P. (2003). Genotyping on a thermal gradient DNA chip. *Genome Res.* 13, 467–475.

Khandurina, J., Meknight, T.E., Jacobson, S.C., Waters, L.C., Foote, R.S., and Ramsey, J.M. (2000). Integrated system for rapid PCR-based DNA analysis in microfluidic devices. *Anal. Chem.* 72, 2995–3000.

Khrapko, K.R., Lysov, Y.-P., Khorlin, A.A., Ivanov, I.B., Yershov, G.M., Vasilenko, S.K., Florentiev, V.L., and Mirzabekov, A.D.A. (1991). Method for DNA sequencing by hybridization with oligonucleotide matrix. *DNA Sequence* 1, 375–388.

Kopp, M., De Mello, A., and Manz, A. (1998). Chemical amplification: continuous-flow PCR on a chip. *Science* 280, 1046–1048.

Kricka, L. (1998). Miniaturization of analytical systems. *Clin. Chem.* 44, 2008–2014.

Kricka, L., Fortina, P., Panaro, N., Wilding, P., Alonso-Amigo, G., and Becker, H. (2002). Fabrication of plastic microchips by hot embossing. *Lab on a Chip* 2, 1–4.

Lagally, E.T., Simpson, P.C., and Mathies, R.A. (2000). Monolithic integrated microfluidic DNA amplification and capillary electrophoresis analysis system. *Sens. Actuators B* 63, 138–146.

Lagally, E.T., Medintz, I., and Mathies, R.A. (2001). Singlemolecule DNA amplification and analysis in an integrated microfluidic device. *Anal. Chem.* 73, 565–570.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. *Nat. Genet.* 21(Suppl.), 20–24.

Liu, Y., Rauch, C., Stevens, R., Lenigk, R., Yang, J., Rhine, D., and Grodzinski, P. (2002). DNA Amplification and hybridization assays in integrated plastic monolithic devices. *Anal. Chem.* 74, 3063–3070.

Liu, R.H., Yang, J., Lenigk, R., Bonanno, J., Grodzinski, P., and Zenhausern, F. (2003). Self-contained, integrated biochip system for sample-to-answer genetic assays. *Proc. mTAS 2003 Symp.*, p. 1319. Kluwer Academic Publishers.

Manz, A., Harrison, D.J., Verpoorte, E., Fettinger, J.C., Paulus, A., Ludi, H., and Widmer, H.M. (1992). Planar chips technology for miniaturization and integration of separation techniques into monitoring systems: capillary electrophoresis on a chip. *J. Chromatogr.* 593, 253–258.

Mastrangelo, C., Burns, M., and Burke, D. (1998). Microfabricated devices for genetic diagnostics. *Proc. IEEE,* 86, 1769–1787.

Mei, R., Galipeau, P.C., Prass, C., Berno, A., Ghandour, G., Patil, N., Wolff, R.K., Chee, M.S., Reid, B.J., and Lockhart, D.J. (2000). Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Res.* 10, 1126–1137.

Miller, M.M., Sheehan, P.E., Edelstein, R.L., Tamanaha, C.R., Zhong, L., Bounnak, S., Whitman, L.J., and Colton, R.J. (2001). A DNA array sensor utilizing magnetic microbeads and magnetoelectronic detection. *J. Magnetism Magnetic Mat.* 225, 276–281.

Nagai, H., Murakami, Y., Morita, Y., Yokoyama, K., and Tamiya, E. (2001). Development of a microchamber array for picoliter PCR. *Anal. Chem.* 73, 1043–1047.

Nuwaysir, E.F., Huang, W., and Albert, T.J. (2002). Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.* 12, 1749–1755.

Obeid, P., Christopoulos, T., Crabtree, H., and Backhouse, C. (2003). Microfabricated device for DNA and RNA amplification by continuous-flow polymerase chain reaction and reverse transcription-polymerase chain reaction with cycle number selection. *Anal. Chem.* 75, 288–295.

Oleinikov, A.V., Gray, M.D., Zhao, J., Montgomery, D.D., Ghindilis, L.L., and Dill, K. (2003). Self-assembling protein arrays using electronic semiconductor microchips and in vitro translation. *J. Proteome Res.* 2, 313–319.

Panaro, N.J., Yuen, P.K., Sakazume, T., Fortina, P., Kricka, L.J., and Wilding, P. (2000). Evaluation of DNA fragment sizing and quantification by the Agilent 2100 Bioanalyzer. *Clin. Chem.* 46, 1851–1852.

Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L., and Syvanen, A.C. (2000). A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res.* 10, 1031–1042.

Pethig, R., Huang, Y., Wang, X.-B., and Burt, J.P.H. (1992). Positive and negative dielectrophoretic collection of colloidal particles using interdigitated castellated microelectrodes. *J. Appl. Phys.* 24, 881–888.

Simpson, P.C., Roach, D., Wooley, A.T., Thorsen, T., Johnston, R., Sensabaugh, G.F., and Matthies, R.A. (1998). High throughput genetic analysis using microfabricated 96-sample capillary array electrophoresis microplates. *Proc. Natl. Acad. Sci. USA,* 95, 2256–2261.

Sosnowski, R.G., Tu, E., Butler, W.F., O'Connell, J.P., and Heller, M.J. (1997). Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control. *Proc. Natl. Acad. Sci. USA,* 94, 1119–1123.

Su, H.-J., Surrey, S., McKenzie, S.E., Fortina, P., and Graves, D.J. (2002). Kinetics of heterogeneous hybridization on indium tin oxide surfaces with and without an applied potential. *Electrophoresis* 23, 1551–1557.

Taylor, M.T., Belgrader, P., Joshi, R., Kintz, G.A., and Northrup, M.A. (2001). Fully automated sample preparation for pathogen detection performed in a microfluidic cassette. In *Micro Total Analysis Systems 2001* (Ramsey, J.M., and Van den Berg, A., eds), pp. 670–672. Kluwer Academic Publishers, Monterey, CA.

Tibbe, A.G., de Grooth, B.G., Greve, J., Liberti, P.A., Dolan, G.J., and Terstappen, L.W. (1999). Optical tracking and detection of immunomagnetically selected and aligned cells. *Nat. Biotechnol.* 17, 1210–1213.

Tibbe, A.G., de Grooth, B.G., Greve, J., Dolan, G.J., Rao, C., and Terstappen, L.W. (2002). Magnetic field design for selecting and aligning immunomagnetic labeled cells. *Cytometry* 47, 163–172.

Umek, R.M., Vielmetter, J., Terbrueggen, R.H., Irvine, B., Yu, C.J., Kayyem, J.F., Yowanto, H., Blackburn, G.F., Farkas, D.H., and Chen, Y.P. (2001). Electronic detection of nucleic acids: a versatile platform for molecular diagnostics. *J. Mol. Diagn.* 3, 74–84.

Wang, J. (1999). Electroanalysis and biosensors. *Anal. Chem.* 71, 328R–332R.

Wang, J. (2000). From DNA biosensors to gene chips. *Nucleic Acids Res.* 28, 3011–3016.

Wang, X.-B., and Cheng, J. (2001). Electronic manipulation of cells on microchip-based devices. In *Biochip Technology* (Cheng, J., and Kricka, L.J., eds), pp. 135–159. Harwood Academic Publishers.

Ward, M.D., Quan, J., and Grodzinski, P. (2002). Metal-polymer hybrid microchannels for microfluidic high gradient separations. *Eur. Cells Materials J.* 3, 123–126.

Waters, L.C., Jacobson, S.C., Kroutchinina, N., Khandurina, J., Foote, R. S., and Ramsey, J.M. (1998). Microchip device for cell lysis, multiplex PCR amplification, and electrophoretic sizing. *Anal. Chem.* 70, 158–162.

Wilding, P., Shoffner, M.A., and Kricka, L.J. (1994). PCR in a silicon microstructure. *Clin. Chem.* 40, 1815–1818.

Wilding, P., Kricka, L.J., Cheng, J., Hvichia, G., Shoffner, M.A., and Fortina, P. (1998). Integrated cell isolation and polymerase chain reaction analysis using silicon microfilter chambers. *Anal. Biochem.* 257, 95–100.

Woolley, A.T., and Mathies, R.A. (1994). Ultra-high speed DNA fragment separations using microfabricated capillary array electrophoresis. *Proc. Natl. Acad. Sci. USA,* 91, 11348–11351.

Yang, J., Liu, Y., Rauch, C.B., Stevens, R.L., Liu, R.H., Lenigk, R., and Grodzinski, P. (2002). High sensitivity PCR assay in plastic micro reactors. *Lab on a Chip* 2, 179–187.

Yu, H., Sethu, P., Chan, T., Kroutchinina, N., Blackwell, J., Mastrangelo, C., and Grodzinski, P. (2000). *A Miniaturized and Integrated Plastic Thermal Chemical Reactor for Genetic Analysis*. pp. 545–548. Kluwer Academic Publishers, The Netherlands.

Yuen, P., Kricka, L., Fortina, P., Panaro, N., Sakazume, T., and Wilding, P. (2001). Microchip module for blood sample preparation and nucleic acid amplification reactions. *Genome Res.* 11, 405–412.

Chapter 36

Human Gene Patents and Genetic Testing

Timothy Caulfield[1] and Yann Joly[2]
[1]*University of Alberta, Health Law Institute, Law Centre, Edmonton, Alberta, Canada;*
[2]*Université de Montréal, Centre de recherche en droit public, Montréal, Québec, Canada*

36.1 INTRODUCTION

Since the start of the Human Genome Project, the patenting of human genetic material has been the focus of a great deal of social controversy. Concerns have ranged from claims that such patents infringe human dignity, to a possible adverse impact on the research environment. Despite the ongoing controversies, gene patents continue to be granted in jurisdictions throughout the world and those with an eye toward economic growth insist that they are an important element in the innovation and technology transfer process.

This chapter provides an overview of the gene patenting controversy, particularly emphasizing how gene patents may impact, for better or worse, the development and provision of genetic testing technologies. The chapter begins with a discussion of benefits and social concerns often associated with human gene patents. This is followed by an analysis of gene patents in the health care context and the related patent reform recommendations that have been proposed by various policy-making entities.

36.2 BENEFITS FROM PATENTS

Patents are meant to benefit society by encouraging innovation through granting the inventor an exclusive monopoly over the invention. This monopoly, which generally lasts 20 years, gives the inventor the ability to control the use of the invention, thus putting the inventor in a position to profit. Specifically, the patent holder may prevent anyone else from making, constructing, emulating, using, or selling the patented invention, or any other invention that achieves the same result in substantially the same manner. However, patents are also meant to ensure that the details of the invention are disclosed to the public. Indeed, one of the criteria for obtaining a patent is that the invention is described in a sufficient fashion so that an individual with reasonable skill in the relevant art or technical field would have the ability to replicate the invention. This disclosure of information is meant to facilitate the advancement of further innovation and the avoidance of wasteful, duplicative research (ALRC, 2004).

Patents are viewed as an increasingly important part of the world economy. Many commentators have suggested that the world now lives in an era where ideas and innovation have come to dominate the new, knowledge-based economy. Just as the industrial revolution changed the way in which wealth was generated, in a similar fashion areas such as information technology, computers, biotechnology, and, perhaps one day, nanotechnology are opening new exciting market opportunities. In such fields, where ideas, as opposed to tangible goods, are often the primary commodity, intellectual property protection is perceived to be fundamentally important to economic growth. Some commentators have gone so far as to suggest that the ability of a nation to generate patents is a strong indicator of the health and long-term prospects of the region's economy. For example, Juan Enriquez (2001) has suggested that ''patents are a good window (although not the only window) on who might triumph and who might lose over the course of the next two decades''.

Few other industries rely on patent protection as much as the biotechnology sector (Straus, 1998; Cook-Deegan and McCormack, 2001; Scherer, 2002; Stott and Valentine, 2003). Often, patents are the only assets that biotechnology companies have, and because potential products can take years to bring to market, investors believe they need strong intellectual property protection in order to provide a small sense of long-term security. The perceived importance of patents to the biotechnology sector is well illustrated by the way in which investment seems to follow strong intellectual property protection frameworks. When a rough map of the human genome was completed in 2000, US

President Clinton and British Prime Minister Blair made a joint announcement suggesting that the ''gene map belongs to all'' (Evenson, 2000). The statement was not meant to implicate patent policy. Nevertheless, the fear that the two leaders would seek to diminish intellectual property protection sent stocks in biotechnology companies tumbling (Schehr and Fox, 2000). Similarly, when the Supreme Court of Canada held that a genetically modified mouse, known as the oncomouse, was not patentable (*Harvard College* vs. *Canada* (Commisssioner of Patents), 2002), a case that makes Canada the only developed nation that has explicit jurisprudence not allowing the patenting of ''higher life forms'', there were many predictions that the decision would have dire consequences on the Canadian biotechnology sector (Abraham, 2002). For example, one Canadian commentator noted that: ''While the technical impact of the decision will likely be minor, the negative effect on future biotechnology could be huge'' (Gervais, 2002).

Although most in the biotechnology industry have strong views about the importance of patenting to the strength of the biotechnology sector, some commentators have noted that there is, in fact, little systematic research to support the claim. For example, Richard Gold and colleagues have noted that: ''Despite the assumption within intellectual property systems that they are necessary to encourage research and development, there is only a modest body of empirical evidence to support this in the biotechnology industry'' (Gold *et al*., 2000, 2002a).

36.3 PATENTABILITY

Though the patenting of human genetic material has been the source of social controversy, there have never been any significant legal barriers to the patenting of human genetic material. Ever since the famous US Supreme Court case of *Diamond* v. *Chakrabarty* [447 US 303 (1980)], there have been few legal obstacles to the patenting of biologically based ''inventions''. In the *Chakrabarty* case, the plaintiff challenged the US Patent Office after a patent on an oil-eating bacterium had been refused. The Patent Office took the position that the bacterium was not an invention but a product of nature. The Supreme Court sided with *Chakrabarty*, noting that whether an organism was alive or not was deemed irrelevant for the application of the law, and as long as an element of human inventiveness was involved, the innovation could be patented.

Since the *Chakrabarty* case, the boundaries of patent law have been pushed even further. In the USA, for example, numerous patents have been awarded on plants and animals (Kevles and Berkowitz, 2001). The patenting of human genes is consistent with this trend in the jurisprudence. From the perspective of patent offices in Europe, Japan, the USA, and Canada, human gene inventions are patentable as long as they satisfy the basic criteria of national patent regimes: the invention must be new, non-obvious, and useful. Though many may view human gene inventions as the mere discovery of something that already exists in nature, if a degree of human inventiveness has been applied, such as isolating the gene to make it useful, the gene is potentially patentable. Just as a naturally occurring chemical can be patented if an inventor has devised a method of purifying and making the chemical useful, so too are human gene sequences. As summarized by Demaine and Fellmeth (2003), ''the [US] Patent and Trademark Office and federal courts now routinely hold discovered natural substances patentable if they are 'isolated and purified' or otherwise insubstantially modified. Naturally occurring DNA and protein biomolecules have, consequently, become the subject of patent applications.''

Similarly, the European Patent Office Guidelines state that to ''find a substance freely occurring in nature is also a mere discovery and therefore unpatentable. However, if a substance found in nature has first to be isolated from its surroundings and a process for obtaining it is developed, that process is patentable. Moreover, if the substance can be properly characterized either by its structure, by the process by which it is obtained or by other parameters and is 'new' in the absolute sense of having no previously recognized existence, then the substance per se may be patentable'' (Bostyn, 2003).

Gene patenting reached its peak in the USA around the mid-1990s. During this period, the National Institutes of Health filed patent applications for 6,800 expressed sequence tags. Although these applications would subsequently be withdrawn, the general trend continued. In 1996, Incyte Pharmaceuticals filed patent applications for as many as 400,000 expressed sequence tags (Holman and Munzer, 2000). Concurrently, a promising discipline amalgamating genomics and pharmaceutical science termed ''pharmacogenomics'' caught the eye of the biotechnology business community, which led to the first patent applications on single nucleotide polymorphisms (Joly, 2003). As of the year 2000, over 25,000 DNA-based patents have been issued in the USA alone (Cook-Deegan and McCormack, 2001). Gene patenting never reached the same level of effervescence elsewhere as it did in the USA. Thomas and colleagues (2002) found that between 1996 and 1999, out of 6,786 DNA patents filed, the majority were filed in the USA (62%). Given that the major part of the biotechnology industry remains concentrated in this country, these figures are hardly surprising (OECD, 2006).

In the case of patents on fundamental genetic research tools, there has been a change in outlook in recent years. These patents have recently become more difficult to obtain and to enforce in both North America and Europe. In North America, the enactment of *Utility Examination Guidelines*, validated by the US Court of Appeals for the Federal Circuit in the landmark *In Re Fisher* vs. *Lalgudi* case, introduced the requirement that in order to be eligible

for patent protection, DNA patents should have a substantial utility demonstrable through specific benefits (Joly, 2006). In Europe, a restrictive application of the patentability criteria made by the European Patent Office Board of Appeals has led to similar results in practice (Hopkins *et al*., 2007). Moreover, the value of these patents, as well as the surrounding commercial hype around biotechnology patenting, seems to have somewhat diminished (Gura, 2002).

36.4 GENERAL CONCERNS

Human gene patents have become the source of a great deal of social controversy. Indeed, the appropriateness and potential impact of patents on human genetic material has emerged as one of the dominant social and ethical concerns in the area of human genetics. Jeremy Rifkin (1998), one of the most vocal opponents of human gene patents, has stated, ''the debate over life patents is one of the most important issues ever to face the human family. Life patents strike at the core of our beliefs about the very nature of life and whether it is to be conceived of as having intrinsic or mere utility value.''

Many commentators have suggested that patenting human genes has the potential to facilitate the commodification of humans generally. The Australian Law Reform Commission recently summarized this concern stating: ''Another objection to patents on genetic materials is that they may engender a lack of respect for human life and dignity'' (ALRC, 2004). On this view, to grant a patent – a proprietary right – on something suggests that it is a fit subject for such rights. Consequently, patents on genetic materials are thought to commodify parts of human beings by treating them as objects, or as something to be placed in the stream of commerce for financial gain. Commercialisation of parts of human beings is seen as ethically problematic because it might affect how we value people'' (Resnik, 2004).

Likewise, it has been suggested that the human genome should be considered a natural resource, part of the common heritage of humankind and, as such, should not belong to any one individual or corporation (Rifkin, 1998). Similarly, it has been put forward that genetic information is a global public good whose benefit should accrue to everyone (Thorsteinsdóttir *et al*., 2003). These ideas have gained a degree of traction with some policy-making entities. For example, the Human Genome Organization's (HUGO) Ethics Committee has suggested that human genomic databases are global public goods whose benefit should be shared by all (HUGO, 2002). Additionally, UNESCO's 1997 Universal Declaration on the Human Genome and Human Rights states that ''the human genome in its natural state shall not give rise to financial gains'' (UNESCO, 1997). However, considering the human genome at the level of the species as a common heritage, or genetic information as a global public good, does not preclude granting of individual patents (Knoppers and Joly, 2007). Consequently, these policy declarations have had very little practical impact on domestic patent law. In other words, they have not affected how individual patent offices view the patentability of human genetic material.

The patenting of biological substances has also been associated with the notion of ''biopiracy'' (Rifkin, 1998; Sheremeta and Knoppers, 2007). In general, this concern relates to the exploitation of one region's genetic heritage or genetic resources by researchers or industry from another region. This may include, for example, researchers from a developed country using genetic samples from a developing nation without appropriate consent or authorization. Concern about this kind of phenomenon led the HUGO Ethics Committee to recommend that organizations or individuals who profit from genetic research should have an obligation to return a percentage (1–3%) of the gross profit to the region and/or community involved in the research (HUGO, 2000). In Newfoundland (Canada), a recent report released at the request of the provincial government has recommended the adoption of a mandatory approbation model for every commercial genetic research in human genetics. This model would require sponsors of any research project with commercial potential to submit a motivated benefit-sharing proposition along with the research protocol (Pullman and Latus, 2003).

Though the stimulation of innovation is one of the stated goals of the patent system, there has also been a long-standing concern that the patenting process may, paradoxically, hurt research by promoting a more secretive and less collaborative research environment. Research conducted by Blumenthal and colleagues (1997, 2002, 2006) suggests academic–industry collaborations are ''significantly associated with the tendency to withhold the results of research''. They also found that the most commonly cited reason for the delay in publication was the need to allow time for filing a patent application.

Similarly, many commentators have suggested that ''upstream'' patents, for example, on gene sequences of unclear function may hinder the development of research on ''downstream'', clinically useful inventions, such as genetic tests or therapies. This concern is closely related to what has been called the ''tragedy of the anti-commons''. First articulated by Heller and Eisenberg (1998), this theory suggests that patents may hurt innovation in the context of gene research because they may create a disincentive to research. Researchers may avoid investigating a particular gene or gene region for fear of infringing an existing patent, thus stifling what may be useful research (Andrews, 2002a, b). In addition, patents may increase the cost of research by compelling researchers to agree to pay licensing fees to use the patented invention. This latter concern is particularly problematic given the uncertainty surrounding the application of the research exemption in most relevant jurisdictions. For example, the recent US case of *Madey* vs. *Duke*

University in 2002 has severely limited the research exemption in that country. The court held that the research exemption is applicable only if the research is done ''solely for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry''. The exemption does not apply if the alleged offender has a legitimate business interest in the patented invention. The US Supreme Court has recently refused to review this decision.

Despite the frequent articulation of similar concerns, very little in the way of substantial patent reform has occurred in any jurisdiction. Indeed, many of these issues were identified over a decade ago by commentators like Hubbard and Wald (1993) who stated, ''It is important not to lose sight of the fact that the pieces of DNA being sequenced are part of our bodies; they are not being invented by these researchers. If the base sequences of the DNA can be patented, rather than remaining in the public domain, the rights to the commercial use of these sequences will belong to the NIH or to companies that buy them from the MRC. In the end, consumers will be the losers. They will first pay the costs of the research and patenting with their taxes, then pay prices inflated by monopolies.'' Nevertheless, the rules applying to the patenting of human genetic material have remained largely unchanged since the start of the Human Genome Project. The emerging evidence might justify this cautious approach to a biotechnology influenced patent reform (Caulfield *et al*., 2006). Recent studies (Walsh *et al*., 2003, 2005; Nicol and Nielsen, 2003; Nagaoka, 2006; Hansen *et al*., 2006; see also http://www.oecd.org/dataoecd/36/22/1817995.pdf) have demonstrated that the patent system has not, so far, created widespread problems of access to genetic research tools. However, the available evidence is still limited and some of these same studies also recommend careful monitoring of the situation for the next few years in order to avoid the formation of parasitic patent thickets capable of delaying the progress of science in broad research areas. Consequently, some authors still favor a legislative solution to the issue of gene patenting (Gold *et al*., 2007; Herder and Johnston, 2008).

In response to some of these concerns, in 2007, US congressman Becerra introduced a draft bill (Genomic Research and Accessibility Act 2007) that would prohibit patents on nucleotide sequences. However, it is unlikely that the bill will gain the necessary support from other congressmen to be enacted into law. This is not the first legislative attempt in this country to limit genetic patents and it likely won't be the last (Holman, 2007). After a transition period, the patent system might have adapted itself to the challenge posed by genetic research tools' patents. However, new challenges are emerging in the field of stem cell research, where broad patents, such as that of the Wisconsin Alumni Research Foundation (WARF), could potentially threaten the progress of science (Loring and Campbell, 2006). Issues, often linked to overbroad patents or inadequate licensing practices, have also surfaced in the area of genetic testing.

36.5 CONCERNS RELATED TO THE PROVISION OF HEALTH CARE

As more and more genetic technologies move from the laboratory toward clinical use, new patent controversies have started to emerge. Specifically, that patenting may drive up the cost of health care and impede availability of new medical innovations (Henry *et al*., 2002). Of course, concern about the potential impact of patents on health care systems is not unique to genetic technologies and services. In a recent report on the Canadian Health Care system, the problem was discussed in the context of pharmaceuticals: ''On the one hand, [a patent] protects the intellectual property of pharmaceutical companies and helps offset the considerable investment they make in researching and developing new drugs. On the other hand, it delays the introduction of low cost generic drugs'' (Romanow, 2002). However, several high-profile controversies have caused policy makers throughout the world to focus on the potential adverse implications of gene patenting in the context of health systems (Andrews, 2002a, b).

As previously noted, patents give the inventor a monopoly over the use and control of the invention. This allows the patent holder to charge a premium for the use of the invention. Indeed, the ability to reap substantial rewards for the use of the invention is the primary incentive that is meant to encourage innovation. However, this monopoly also means that there are few ways to control costs. Unless a country has some kind of formal price control schemes in place, patent holders are well within their rights to charge whatever they deem appropriate. The downside of this loss of control is most readily apparent in countries, such as Canada, that have a publicly funded health care system where global budgets may not be able to accommodate the demanded monopoly price. In such situations, the patent may result in a loss of public access to a necessary health care service. This can happen either because the administrators of the public system decide that they will not pay for the patented test, or because the patent holder simply refuses to allow access (Caulfield *et al*., 2000, 2002).

The practical impact of this situation is well illustrated by the controversy surrounding Myriad Genetics in Canada, though there have been, of course, a number of similar patenting controversies, including the patenting and exclusive licensing of the Canavan gene in Florida (Hahn, 2003), and the patenting of the SARS genome (Gold, 2003). Indeed, Williams-Jones (2002) characterized the case as follows: ''The Myriad case is a harbinger of an increasing number of instances where gene patents provide companies with monopolies on the development, marketing, and provision of genetic tests and therapeutics. Not surprisingly, this case

has become a focal point in Canada and Europe for debates about the social and ethical implications of DNA patenting and the commercialization of genetic tests.''

Myriad Genetics controls the patents for the *BRCA1/2* genes and for the related testing processes. These genes are associated with predisposition to breast and ovarian cancer. Though there is still an ongoing debate regarding the clinical utility of testing for *BRCA1/2* gene mutations (Healy, 1997), the provision of the test has become the standard of care in many regions throughout the world (Gold *et al*., 2002b). In Canada, various provinces provided the test as part of the public health care system for a relatively reasonable cost, though there was a degree of variation in testing and sequencing techniques.

In the summer of 2001, Myriad Genetics decided to take steps to enforce its patents over the *BRCA1/2* genes. Provincial health care ministries in Canada received a cease and desist letter from Myriad. They were told that all future genetic testing that utilizes the *BRCA1/2* genes must be done through Myriad's laboratories. The Myriad test is quite expensive as compared to the testing process already being done in Canada. The Myriad test would cost approximately 3,800 Canadian dollars per test. In some cases, this was more than four times the cost of the testing being done within the provincial system. As a result, a number of Canadian provincial health ministers stated that the public system could not afford the Myriad test. Some provinces decided to ignore the patent and continue testing; others, at least temporarily, simply stopped offering the test (Williams-Jones and Burgess, 2004). In Europe, the situation grew even more contentious. Myriad's patents were perceived by a considerable number of geneticists and by the general population as a threat not only to public health, but also to scientific development and personal privacy as well. Consequently, European-wide opposition procedures were engaged against all three of Myriad's *BRCA1* gene patents. These procedures eventually resulted in one of Myriad's patents being revoked and the two others limited in scope (Matthijs and Halley, 2002; Cassier and Stoppa-Lyonnet, 2005; Matthijs, 2006).

Following the conclusion of the Myriad saga, another patented test, using the *JAK2* gene as an indicator of myeloproliferative disorders, was licensed exclusively in Canada in 2007 and threatened to raise similar issues again. However, this time around, the parties were offered a forum at a governmental workshop to peacefully discuss their differences before the story got too litigious. Interestingly, during the event, both parties were able to quickly and collaboratively solve the problems of access (Gold, 2008). This encouraging resolution of the story seems to suggest that the main issues with respect to the Myriad Genetics case were likely related to the lack of communication and openness between the parties as well as to the impact of the overly negative portrayal of the debate in the media (Caulfield *et al*., 2007; Gold, 2008).

These controversies highlighted a number of interesting policy issues. First, the Myriad dilemma illustrated the potential conflict between two key governmental priorities, access to affordable health care and the promotion of innovation and the economy (Caulfield, 2003). Over the past few decades, containing the cost of publicly funded health care systems has been a key policy issue for most OECD countries (Flood, 2000). The Myriad story stood as an example of how patents, at least initially, might drive up costs. However, as noted previously, patents are also viewed as an essential component of the commercialization process. Thus, for those in industry and government interested in economic growth, diminishing the strength of patent protection is hardly an attractive option. On the contrary, within universities and in many sectors of the government, the obtainment of biotechnology patents is aggressively promoted. As such, a strange policy paradox has emerged where explicitly pro-biotechnology politicians, such as Ontario's former Premier Mike Harris, argued against gene patents. For example, shortly after the Myriad story broke, Premier Harris suggested: ''The benefits of a world-wide effort such as the Human Genome Project should not be the property of a handful of people or companies. Our genetic heritage belongs to everyone. We must share the benefits fairly and do what we can to make genetic tests and therapies affordable and accessible'' (Benzie, 2001). One wonders what his position would have been if Myriad Genetics was located in Ontario.

Second, there is also concern that gene patents might inhibit the development of cheaper or more efficient and effective ways of doing the same testing. For example, there are those who believe that the Myriad test may not be the most comprehensive procedure for the determination of *BRCA1/2* gene mutations, though this remains controversial (Gold *et al*., 2002a). Researchers at the Institute Curie in France, and the Erasmus University Medical Center in the Netherlands, claim that the Myriad technique misses 10% to 20% of expected mutations, seriously jeopardizing the quality of test results and usefulness of this information for patient care (Williams-Jones, 2002; Matthijs and Halley, 2002). If true, and there are more efficient and effective procedures available, the current Myriad patents would block their use. Again, this could result in increased costs to the health care system.

Recently, Cho and colleagues (2003) analyzed many of these concerns. In that study, researchers interviewed directors of genetic laboratories throughout the USA. They asked the directors various questions about the impact of gene patenting and licensing on clinical practice. The results showed that, in general, laboratory directors had a negative impression of gene patents. Indeed, the researchers noted that ''It was striking that virtually no respondents, including those from commercial laboratories, thought the effects of patents and licenses on the costs, access, and development of genetic tests have been positive''.

Moreover, their findings are also consistent with concerns that ''patents are inhibiting commercialization of genetic tests''. Given that one of the justifications for patents is to facilitate the commercialization process and the dissemination of useful technologies, these findings raise interesting questions about the utility and effectiveness of the current patent regime in the context of genetic innovations. Cho and colleagues (2003) summarize their findings thus: ''We conclude that patents and licenses have a significant negative effect on the ability of clinical laboratories to continue to perform already developed genetic tests, and that these effects have not changed substantially throughout the past 3 years. Furthermore, the development of new genetic tests for clinical use, based on published data on disease-gene associations, and information sharing between laboratories, seemed to be inhibited.''

Interestingly, there is some evidence that the public has similar concerns about the impact of gene patents on access to health care services. The public has the same divided view of patents as that represented in much of the literature. A survey of the Canadian public found that few of those surveyed had moral or religious objections to the patenting of human genes and a majority (63%) saw more benefits than risks associated with the patenting process (Pollara, 2000). However, in focus groups, it was found that there were major concerns based on issues of access and equity. In the context of health care, at least in Canada, access seems to be the dominant public consideration.

Another concern associated with patents, and the commercialization process in general, is a possible impact on utilization and uptake. The individuals and organizations that hold patents have a natural and understandable desire to see their invention utilized by as broad a market as possible. However, some have argued that this market pressure may cause commercial labs to offer their services to an inappropriately broad sector of the population (the broader the definition of at-risk population, the larger the market; Martone, 1998; Caulfield *et al*., 2001). Companies that control a patent for a particular disease gene may inappropriately stress the seriousness of a disorder in order to encourage people to get tested (Loeben *et al*., 1998; Biesecker and Marteau, 1999). Such an approach may conflict with a more dispassionate assessment of the clinical utility for a given genetic service. For example, if one compares Myriad Genetics' indications for BRCA tests with those recommended by an independent body, one finds that the latter excludes women without a family history of breast or ovarian cancer, whereas Myriad's guidelines include these lower risk women (Willison and MacLeod, 2002).

Naturally, this kind of market pressure could create inappropriate utilization and unnecessary costs for the health care system. Controlling the dissemination and uptake of new technologies has always been a challenge, particularly because few health care systems have the necessary framework to make system-wide implementation policies. Utilization and uptake are the result of a complex mix of professional decision making, market pressure, and patient preference and, in most countries, coverage by the public system. However, as research and health care become increasingly privatized, the market seems likely to have an increasingly important impact on the introduction of genetic tests into clinical practice. Blancquaert (2000) summarizes the problem as such: ''Not only will commercialization increase public demand and affect the availability of genetic tests and services, it will certainly have an impact on the financial and human resources of national health care systems as well.''

It should not be forgotten that if, as it is often predicted, pharmacogenetics and multiplex testing become a reality, then many of the previously noted concerns could be amplified (Evans and Relling, 1999). This is because the use of both pharmacogenetics and multiplex testing – that is, testing for many mutations at the same time – could implicate numerous, perhaps dozens, of patented gene sequences. As such, obtaining the necessary licenses to provide the genetic service could be a tremendous challenge. For example, pharmacogenetics is premised on the idea that we will be able to ''individualize'' pharmaceutical treatments to maximize effectiveness and minimize adverse events (Shah, 2003, see also Chapter 22). However, this will necessarily involve testing an individual for the presence or absence of a variety of gene sequences. If these sequences are owned by another company or, more likely, several companies, the development of pharmacogenetic clinical protocols could become a complicated and expensive venture. The use of pharmacogenetics by drug companies to revive old patents by marketing the ''rejuvenated'' drug, sometime along with a genetic test, to a specific genotype or racial category (such as was attempted in the BiDil case) is also a concern.

36.6 SUGGESTED REFORMS

Over the past few years, partly as a result of the previously noted controversies, a variety of governmental and policy entities have been seen to take on the gene patenting issues. Even though the UK Nuffield Council supported the use of patents on diagnostic tests as a means of ''rewarding the inventor'', they concluded that ''in the future, the granting of patents that assert rights over DNA sequences should become the exception rather than the norm'', and went on to make a number of specific reform suggestions, including applying existing patenting criteria more stringently (Nuffield Council, 2002). In addition, and perhaps most controversial, the Council also suggested that ''in the case of patents that have been granted for diagnostic tests based on genes, compulsory licensing may be required to ensure reasonable licensing terms are available to enable alternative tests to be developed'' (Nuffield Council, 2002).

Another interesting UK report, which explicitly endorses the Nuffield Council's recommendations, was issued in July of 2003 (Cornish *et al*., 2003). This report was commissioned by the UK Department of Health because of emerging ''serious concern about the impact of intellectual property upon research and the use of novel developments in genetics affecting health care''. The authors provide a series of recommendations including the suggestion that the Department of Health should consider ''a robust central policy for 'licensing in' designed to moderate excessive demands by licensors by considering, as possible options, the use of compulsory licensing, competition law and Crown use''.

In what appeared to be a direct response to the Myriad controversy (Caulfield *et al*., 2006), the Ontario government struck a policy group to examine the potential adverse social implications of gene patenting. The government's 2002 Report to the Premiers made a number of similar recommendations from those of the Nuffield Council. The Ontario report – which, on January 25, 2002 was adopted by the Premiers from all the Canadian provinces at the Premiers Conference on Healthcare in Vancouver – recommends *inter alia*, a clarification of patent criteria in relation to human genes, the exclusion of broad-based genetic patents covering multiple uses, a clarification of the experimental and non-commercial exceptions, and an expansion of the methods of medical treatment exclusion. The Ontario government has also decided to become an intervener in a plant patent case in hopes of raising issues associated with the impact of gene patents on health care policy. It was reported that the Ontario affidavit to the Supreme Court says the case ''has important implications for the development of public policy in Ontario including the delivery of health care to its residents'' (Bueckert, 2003). The Ontario report was followed in 2006 by a report from the Canadian Biotechnology Advisory Committee on Human Genetic Materials, Intellectual Property and the Health Sector (CBAC, 2006). The highlights of the proposed recommendations were: some clarification to the wording of the Patent Act, the elaboration of a statutory research exemption, as well as the development of licensing guidelines that could be adopted by national Canadian research granting agencies. As of 2008, the Canadian government has yet to implement the suggested substantive changes to the Canadian Patent Act.

The Australian Law Reform Commission (ALRC) has also conducted a major inquiry on the topic of gene patenting and human health. It released a final report containing its recommendations for improving the Australian patent system (ALRC, 2004) in 2004. The proposed avenues of reform are relatively narrow in nature, involving administrative changes and the adoption of technical guidelines, rather than substantial changes in the law. The few statutory changes that were discussed included the introduction of a research exemption for experimental use and the drafting of more refined crown use and compulsory licensing provisions.

There are a number of common themes that run through these and other policy documents. Tightening patent criteria, clarifying exemptions, and the use of compulsory licensing are all suggestions that consistently emerge. Another common suggestion is to introduce a mechanism that would allow social and ethical issues to be considered in the assessment of patents (Gold and Caulfield, 2002). This could, for example, be similar to Article 53(a) of the European Patent Convention, which allows the rejection of inventions, the publication or exploitation of which would be contrary to *ordre public* or morality. Some worry, however, that such an approach would introduce too much uncertainty into a system that was never meant to address ethical issues. Crespi (2003), for example, emphasizes: ''Most professional patent practitioners are somewhat dubious as to the wisdom or appropriateness of introducing ethical and moral questions into a law which is primarily based on the assessment of originality, innovativeness, and practical advances in science and technology.'' Given that the morality exception has been interpreted very broadly by the EPO to deny the patenting of embryonic stem cells at the regional level in Europe, there are certainly some grounds for concern with this approach from a commercial standpoint (Plomer, 2006).

But not all proposed solutions imply an in-depth reform of the patent system. A recent initiative led by the OECD was intended to promote the adoption of more informed and less restrictive licensing policies that would foster a broader use of patented genetic inventions. As a consequence, the OECD developed model-licensing guidelines to be adapted and enforced at the national level by member countries. According to these guidelines, ''It is important that licensors and licensees are encouraged to consider the possible impact of their license arrangements on the healthcare system and on patients. Licensors and licensees should, while meeting their economic needs, design their licensing arrangements so that patients have access to new health products and services and so that healthcare system administrators have reasonable flexibility to determine how best to implement new healthcare services and products'' (OECD, 2001, 2006). The American NIH took a similar stand in 2005 (NIH, 2005). Other less exacting solutions to the challenge raised by genetic patent include the creation of patent pools or patent clearinghouses as well as the use of open source schemes of peer productions to ensure greater access to both research tools and health care products (Sheremeta and Gold, 2003; Ebersole *et al*., 2005; Joly, 2007; Hope, 2008).

36.7 CONCLUSIONS

Despite the continued articulation of concerns regarding the impact of gene patents and the publication of numerous policy recommendations, there are many reasons why altering

the existing system will remain a challenge. International trade agreements, for example, have the potential to create technical barriers for those seeking patent reform. Trade-related Intellectual Property Rights (TRIPS) agreement of the World Trade Organization, which came into effect in 1995 and currently involves 151 nations, was envisioned as a mechanism for creating and maintaining strong patent protection within all member countries. Patent reform, which greatly alters or appears to erode patent protection, may be subject to challenge. At a minimum, existing international treaties add an element of uncertainty to the policy-making process (Caulfield and von Tigerstrom, 2006). Moreover, the momentum of the biotechnology sector, and the imbedded perception that patent protection is crucially important to economic development, rightly or not, will create political barriers to significant patent reform. Finally, only the best possible evidence should inform any reform process. Currently, there is still limited empirical data available on the actual benefits and risks of patents. Thus, further research is clearly warranted. As intellectual property becomes a more significant part of our social landscape, new challenges will undoubtedly emerge. Does the current patent system possess the flexibility needed to confront the challenges set by human genetics patents? Policy makers throughout the world must strive to make the necessary adjustments to create an intellectual property regime that balances a variety of laudable social goals, including the stimulation of the economy, the encouragement of innovation, and the facilitation of access to reasonably priced health care technologies.

ACKNOWLEDGEMENTS

We would like to thank our colleagues Lori Sheremeta and Richard Gold for their insight into the issues relevant to this area. Indeed, much of this chapter draws on work we have done together. We would also like to thank Suzanne Debow for her editorial help and Genome Canada, the Stem Cell Network, the Alberta Heritage Foundation for Medical Research, the Canadian Institutes of Health Research for RMEthnet and Genome Quebec for the continued funding support.

REFERENCES

Abraham, C. (December 6, 2002). Mouse ruling may stall research. *The Globe and Mail.*, A1.

Andrews, L. (2002). Genes and patent policy: rethinking intellectual property rights. *Nat. Rev. Genet.* 3, 803.

Andrews, L.B. (2002). The gene patent dilemma: balancing commercial incentives with health needs. *Houston J. Health Law Policy,* 65, 106.

Australian Law Reform Commission (2004). *Genes and Ingenuity: Gene Patenting and Human Health.* ALRC 99. Available online: http://www.austlii.edu.au/au/other/alrc/publications/reports/99/

Benzie, R. (September 20, 2001). Ontario to defy U.S. patents on cancer genes: province will pay for $800 test, not $3,850 version by Myriad Genetic Laboratories Inc. *The National Post*, A15.

Biesecker, B., and Marteau, T. (1999). The future of genetic counseling: an international perspective. *Nat. Genet.* 22, 133–137.

Blancquaert, I. (2000). Availability of genetic services: implementation and policy issues. *Community Genet.* 3, 179–183.

Blumenthal, D., Campbell, E.G., Anderson, M.S., Causino, N., and Louis, K.S. (1997). Withholding research results in academic life science. Evidence from a national survey of faculty. *JAMA* 277, 1224–1228.

Blumenthal, D., Campbell, E.G., Clarridge, B.R., Manjusha Gokhale, M., Birenbaum, L., Hilgartner, S., and Holtzman, N.A. (2002). Data withholding in academic genetics. *JAMA* 287, 473–480.

Blumenthal, D., Campbell, E.G., Gokhale, M., Yucel, R., Clarridge, B., Hilgartner, S., and Holtzman, N.A. (2006). Data withholding in genetics and other life sciences: prevalence and predictors. *Acad. Med.* 81, 137–145.

Bostyn, S.J.R. (2003). The prodigal son: the relationship between patent law and health care. *Med. Law Rev.* 11, 67–120.

Bueckert, D. (October 7, 2003). *Ontario government intervenes in high-profile gene patenting case*. The Canadian Press.

Canadian Biotechnology Advisory Committee (2006). *Human Genetic Materials, Intellectual Property and the Health Sector.* Available online: http://www.cbac.gc.ca/epic/site/cbac-cccb.nsf/en/ah00578e.html

Cassier, M., and Stoppa-Lyonnet, D. (2005). L'opposition contre les brevets de Myriad Genetics et leur révocation totale ou partielle en Europe. *Medicine/Sciences* 21, 658–662.

Caulfield, T. (2003). Sustainability and the balancing of the health care and innovation agendas: the commercialization of genetic research. *Saskatchewan Law Rev.* 66, 629–645.

Caulfield, T., and von Tigerstrom, B. (2006). Gene patents, health care policy and licensing schemes. *Trends Biotechnol.* 24, 251–254.

Caulfield, T., Gold, E.R., and Cho, M.K. (2000). Patenting human genetic material: refocusing the debate. *Nat. Rev. Genet.* 1, 227–231.

Caulfield, T., Burgess, M.M., Williams-Jones, B., Baily, M.-A., Chadwick, R., Cho, M., Deber, R., Fleising, U., Flood, C., Friedman, J., Lank, R., Owen, T., and Sproule, J. (2001). Providing genetic testing through the private sector: a view from Canada. *ISUMA. Canadian J. Policy Res.* 2, 72–81.

Caulfield, T.A., Knoppers, B.M., Gold, E.R., Sheremeta, L.E., and Bridge, R.J. (2002). Genetic technologies, health care policy and the patent bargain. *Clin. Genet.* 63, 15–18.

Caulfield, T., Bubela, T., and Murdoch, C.J. (2007). Myriad and the mass media: the covering of a gene patent controversy. *Genet. Med.* 9, 850.

Caulfield, T., Cook-Deegan, R.M., Kieff, F.S., and Walsh, J.P. (2006). Evidence and Anecdotes: an analysis of human gene patenting controversies. *Nat. Biotechnol.* 24, 1091–1094.

Cho, M.K., Illangasekare, S., Weaver, M.A., Leonard, D.G.B., and Merz, J.F. (2003). Effects of patents and licenses on the provision of clinical genetic testing services. *J. Mol. Diagn.* 5, 3–8.

Cook-Deegan, R.M., and McCormack, S.J. (2001). Intellectual property. Patents, secrecy and DNA. *Science* 293, 217.

Cornish, W.R., Llewelyn, M., and Adcock, M. (2003). Intellectual property rights and genetics: a study into the impact and management of intellectual property rights within the healthcare sector. *Cambridge Genetics Knowledge Park*, 1–134.

Crespi, R.S. (February 26, 2003). Patenting and ethics: a dubious connection. *Bio-Science Law Review*.

Demaine, L., and Fellmeth, A. (2003). Patent law: distinguishing natural substances from patentable inventions. *Science* 300, 1375–1376.

Ebersole, T., Guthrie, M.C., and Goldstein, J. (2005). Patent pools and standard setting in diagnostic genetics. *Nat. Biotechnol.* 23, 937–938.

Enriquez, J. (2001). *As the Future Catches You: How Genomics & Other Forces are Changing Your Life, Work Health & Wealth*. Crown Business, New York.

Evans, W., and Relling, M. (1999). Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* 286, 487–491.

Evenson, B. (March 15, 2000). Gene map belongs to all. *The National Post*, A1.

Flood, C. (2000). *International Health Care Reform: A Legal, Economic and Political Analysis*. Routledge, London.

Genomic Research and Accessibility Act (2007). H.R.977.IH.

Gervais, D. (December 14 2003). Eeek! the mouse confounds the house! *The Ottawa Citizen*, B7.

Gold, E.R. (2000). Finding common cause in the patent debate. *Nat. Biotechnol.* 18, 1217–1218.

Gold, E.R. (2003). SARS genome patent: symptom or disease? *Lancet* 361, 2002–2003.

Gold, E.R., and Caulfield, T.A. (2002). The ''moral tollbooth'': a method that makes use of the patent system to address ethical concerns in biotechnology. *Lancet* 359, 2268–2270.

Gold, E.R., Castle, D., Cloutier, L.M., Daar, A.S., and Smith, P.J. (2002a). Needed: models of biotechnology intellectual property. *Trends Biotechnol.* 20, 327–329.

Gold, E.R., Caulfield, T.A., and Ray, P. (2002b). Gene patents and the standard of care. *Canadian Med. Assoc. J.* 167, 256–257.

Gold, E.R., Bubela, T., Miller, F.A., Nicol, D., and Piper, T. (2007). Gene patents – more evidence needed, but policymakers must act. *Nat. Biotechnol.* 25, 388–389.

Gold, R. (2008). *Beyond Myriad: JAK2 and Compromise.* Centre for Intellectual Property Policy. Blog, March 30, 2008. Available online: http://www.cipp.mcgill.ca/blog/ (last visited April 7, 2008).

Gura, T. (2002). After the gold rush: gene firms reinvent themselves. *Science* 297, 1982.

Hahn, L. (May 2003). Owning a piece of Jonathan. *Chicago*.

Hansen, S., Brewster, A., Asher, J., and Kisielewski, M. (2006). The Effects of Patenting in the AAAS Scientific Community. American Association for the Advancement of Science. Washington. Available on line: http://sippi.aaas.org/Pubs/AAAS_IP_Survey_Report.pdf

Harvard College vs. Canada (Commissioner of Patents), [2002] SCC 76; 219 D.L.R. (4th) 577.

Healy, B. (1997). BRCA genes: bookmaking, fortune telling, and medical care. *N. Engl. J. Med.* 336, 1448–1449.

Heller, M.A., and Eisenberg, R.S. (1998). Can patents deter innovation? The anticommons in biomedical research. *Science* 280, 698–701.

Henry, M.R., Cho, M.K., Weaver, M.A., and Merz, J.F. (2002). Genetics. DNA patenting and licensing. *Science* 297, 1279.

Herder, M., and Johnston, J. (2008). *Licensing for Knowledge Transfer in Human Genetics Research: A Study of Business Models for Licensing and Technology Transfer in Human Genetics Patents*. Stanford University.

Holman, C.M. (2007). Patent border wars: defining the boundary between scientific discoveries and patentable inventions. *Trends Biotechnol.* 25, 539–543.

Holman, M., and Munzer, S. (2000). Intellectual property rights in genes and gene fragments: a registration solution for expressed sequence tags. *Iowa Law Review* 85, 3.

Hope, J. (2008). *BioBazaar: Biotechnology and the Open Source Revolution*. Harvard University Press, Cambridge, Massachusetts; London, England.

Hopkins, M., Mahdi, S., Patel, P., and Thomas, S.M. (2007). DNA patenting: the end of an era. *Nat. Biotechnol.* 25, 185.

Hubbard, R., and Wald, E. (1993). *Exploding the Gene Myth*. Beacon Press, Boston.

HUGO Ethics Committee (2000). Statement on Benefit Sharing. Genome Digest 6, 7–9, online, Human Genome Organization, http://www.hugo-international.org

HUGO Ethics Committee (2002). *Statement on Human Genomic Databases*. Available online: http://www.hugo-international.org/Statement_on_Human_Genomic_Databases.htm

In Re Fisher, 421 F.3d 1365 (Fed. Cir. 2005).

Joly, Y. (2003). Accès aux médicaments: le système international des brevets empêchera-t-il les pays du tiers monde de bénéficier des avantages de la pharmacogénomique. *Les Cahiers de Propriété Intellectuelle* 16, 131.

Joly, Y. (2006). Wind of change: *In re Fisher* and the evolution of the American biotechnology patent law. In *Patent Law and Biological Inventions* (Rimmer, M., ed.). The Federation Press, NSW.

Joly, Y. (2007). Open source approaches in biotechnology: Utopia revisited. *Maine Law Rev.* 59, 386.

Kevles, D., and Berkowitz, A. (2001). The gene patenting controversy: a convergence of law, economic interests, and ethics. *Brooklyn Law Rev.* 67, 233–248.

Knoppers, B.M., and Joly, Y. (2007). Our social genome? *Trends Biotechnol.* 25, 284–288.

Loeben, G.L., Marteau, T.M., and Wilfond, B.S. (1998). Mixed messages: presentation of information in cystic fibrosis screening pamphlets. *Am. J. Hum. Genet.* 63, 1181–1189.

Loring, J.F., and Campbell, C. (2006). Intellectual property and human embryonic stem cell research. *Science* 311, 1716.

Madey vs. Duke University, 307 F.3d 1351 (Fed. Cir. 2002).

Martone, M. (1998). The ethics of the economics of patenting the human genome. *J. Bus. Ethics* 17, 1679–1684.

Matthijs, G. (2006). The European opposition against the BRCA gene patents. *Fam. Cancer* 5, 95–102.

Matthijs, G., and Halley, D. (2002). European-wide opposition against the breast cancer gene patents. *Eur. J. Hum. Genet.* 10, 783–785.

Nagaoka, S. (2006). Presentation to OECD Conference on Research Use of Patented Inventions. Madrid, May 18–19, 2006. Available online: http://www.oecd.org/dataoecd/20/54/36816178.pdf

National Institute of Health. Office of Technology Transfer (2005). *Best Practices for the Licensing of Genomic Inventions: Final Notice*. April 2005. Available online: http://www.ott.nih.gov/policy/lic_gen.html

Nicol, D., and Nielsen, J. (2003). Patents and medical biotechnology: an empirical analysis of issues facing the Australian industry – Occasional Paper No. 6. Centre for Law & Genetics, Sandy Bay, Australia.

Nuffield Council on Bioethics (2002). *The Ethics of Patenting DNA*, Nuffield Council on Bioethics, London. nsf/c5ce8ffa41835d64c125685d005300b0/c1256985004c66e3c1256ac600350f21/$FILE/JT00112476.pdf.

Ontario (2002). *Genetics, Testing and Gene Patenting: Charting New Territory in Healthcare*. Ontario Ministry of Health and Long Term Care, http://www.health.gov.on.ca:80/english/public/pub/ministry_reports/geneticsrep02/report_e.pdf.

Organization for Economic Co-operation and Development (OECD). *Statistics in OECD Member Countries*, September 13, 2001. Available online: http://www.olis.oecd.org/olis/2001doc

Organization for Economic Co-Operation and Development (OECD). *OECD Biotechnology Statistics – 2006*, Paris 2006. Available online: http://www.oecd.org/dataoecd/51/59/36760212.pdf

Plomer, A. (2006). Stem Cell Patents: European Patent Law and Ethics Report (Nottingham, European Commission). Available online: http://www.nottingham.ac.uk/law/StemCellProject/reports.htm

Pollara Research and Earnscliffe Research and Communications (2000). Public opinion research into biotechnology issues, third wave. Biotechnology Assistant Deputy Minister Coordinating Committee, Ottawa. Available online: Canadian Biotechnology Strategy, http://www.biotech.gc.ca/docs/engdoc/3Wavexece.pdf

Pullman, D., and Latus, A. (2003). Benefit sharing in smaller markets: the case of Newfoundland and Labrador. *Community Genet.* 6, 178–181.

Resnik, D. (2004). *Owing the Genome: A Moral Analysis of DNA Patenting*. State University of New York Press, Albany.

Rifkin, J. (1998). *The Biotech Century*. Penguin Putnam, New York.

Romanow, R.J. (2002). Building on Values: the Future of Health Care in Canada. Commission on the Future of Health Care. 1–339. Commissioner, Saskatoon. Available online: Health Canada, http://www.hc-sc.gc.ca/english/pdf/care/romanow_e.pdf

Scherer, F.M. (2002). The economics of human gene patents. *Acad. Med.* 77, 1348–1365.

Schehr, R., and Fox, J. (2000). Human genome bombshell. *Nat. Biotechnol.* 18, 365.

Shah, J. (2003). Economic and regulatory considerations in pharmacogenomics for drug licensing and healthcare. *Nat. Biotechnol.* 21, 747–753.

Sheremeta, L., and Gold, E.R. (2003). Creating a patent clearinghouse in Canada: a solution to problems of equity and access. *Health Law Rev.* 11, 17.

Sheremeta, L., and Knoppers, B.M. (2007). Beyond the rhetoric: population genetics and benefit-sharing. In *Accessing and Sharing the Benefits of the Genomics Revolution* (Phillips, P.W.B., and Onwuekwe, C.B., eds). Springer, The Netherlands.

Stott, M., and Valentine, J. (2003). Impact of gene patenting on R&D and commerce. *Nat. Biotechnol.* 21, 729–731.

Straus, J. (1998). Bargaining around the TRIPS agreement: the case for ongoing public-private initiatives to facilitate worldwide intellectual property transactions. *Duke J. Comp. Intern. Law* 9, 91–107.

Thomas, S., Hopkins, M.M., and Brady, M. (2002). Shares in the human genome – the future of patenting DNA. *Nat. Biotechnol.* 20, 1185–1188.

Thorsteinsdóttir, H., Daar, A., Smith, R., and Singer, P. (2003). Genomics – a global public good? *Lancet* 361, 891–892.

United Nations Educational, Scientific and Cultural Organization (UNESCO) (1997). Universal Declaration on the Human Genome and Human Rights, 29th Sess., 29 C/Resolution 19.

Walsh, J.P., Cho, C., and Cohen, W.M. (2005). View from the bench: patents and material transfers. *Science* 309, 2002–2003.

Walsh, J.P., Cohen, W.M., and Arora, A. (2003). Patenting and licensing of research tools and biomedical innovation. In *Patents in the Knowledge-Based Economy* (Cohen, W.M., and Merrill, S., eds). National Academies Press, Washington, DC.

Williams-Jones, B. (2002). History of a gene patent: tracing the development and application of commercial BRCA testing. *Health Law J.* 10, 123–146.

Williams-Jones, B., and Burgess, M.M. (2004). Social contract theory and just decision making: lesson from genetic testing for the BRCA mutations. *Kennedy Institute of Ethics J.* 14, 115.

Willison, D.J., and MacLeod, S.M. (2002). Patenting of genetic material: are the benefits to society being realized? *CMAJ* 167, 259–262.

Chapter 37

Genetic Counseling and Ethics in Molecular Diagnostics

Mary Petrou

University College London, Institute of Women's Health, Haemoglobinopathy Genetics Centre (Perinatal Centre), University College London Hospitals NHS Foundation Trust, London UK

37.1 INTRODUCTION

Genetic counseling is inseparable from genetic diagnosis. It aims to replace misunderstandings about the causes of genetic disease with correct information, and to increase people's control of their own and their family's health by informing them of the resources available for diagnosis, treatment, and prevention. Though counseling has a role in many medical consultations, it is particularly important in medical genetics because of the often predictive nature of genetic information, the implications for other family members, the difficult choices that sometimes have to be made, and the important ethical problems that can be involved.

Genetic counseling has been defined as the process by which patients or relatives at risk of a disorder that may be hereditary are advised of the consequences of the disorder, and the probability of developing and transmitting it and the ways in which this may be prevented or ameliorated (Harper, 1988). This definition requires:

- A correct diagnosis in the presenting family member,
- Explanation of the nature and prognosis of the disorder, the treatment available, and where to find it,
- Estimation of genetic risk for parents and family members. This requires drawing a family tree, and it may also call for investigations on other family members,
- Communication of genetic risks, and the options for avoiding them, including the chances for parents and other family members of passing the disorder on to other children, and explanation of the risk. The options for avoiding further affected children are techniques of prenatal diagnosis, problems, risk of error, and complications,
- Supporting the individual or couple in making the decision that is right for them,
- Accessibility for long-term contact: people at risk often need counseling and support at several points in their life.

Therefore specialist genetic knowledge, training in counseling skills, time, ability to communicate, and back-up by a medical geneticist or trained genetic counselor all are required. More than 7,000 single gene defects have been described worldwide (http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM&itool=toolbar). Single gene defects fall into two main groups, the common and the rare single gene disorders. The common single gene disorders include the hemoglobin disorders (sickle cell anemia and thalassemia), oculocutaneous albinism, and cystic fibrosis (Christianson *et al*., 2006). Because the author's expertise is in hemoglobin disorders, the remainder of this chapter will draw on experiences focused on these disorders, particularly thalassemia.

Hemoglobin disorders constitute the most common lethal inherited disorders worldwide (reviewed in Weatherall and Clegg, 2001). They are common in populations in tropical Africa, Asia, and the Mediterranean region and have spread by migration throughout the world (Angastiniotis and Modell, 2003; Stuart and Nagel, 2004). It is estimated that 307,900 children are born annually with a severe hemoglobin disorder and 60–70% of births occur in sub-Saharan Africa (Christianson *et al*., 2006). Consequently, sickle cell disease (SCD) accounts for 70% of hemoglobin disorders worldwide because of the high frequency of the gene. In Africa it is estimated that 224,200 infants are born annually with a sickle cell disorder and

most die before they reach the age of five (Akinyanju 1989; Fleming *et al*., 1979). Thalassemia is prevalent in the Mediterranean area, the Middle East and South East Asia, and the Pacific. The carrier rates range from 2 to 19% in the different populations (van Baal *et al*., 2007). The birth prevalence of the hemoglobin disorders in countries affected by migration of populations varies according to the geographic location and the origin of the populations (Patrinos *et al*., 2004). It is estimated that 0.7 in 1,000 pregnancies are affected in the United Kingdom and 0.4 in 1,000 in North America (WHO, 1994; Angastiniotis and Modell, 2003).

The hemoglobin disorders are caused by mutations, which affect the genes that direct synthesis of the globin chains of hemoglobin, and may result in reduced synthesis (thalassemia syndromes) or structural changes (hemolytic anemia, polycythemia, or more rarely, cyanosis).

Thalassemia mutations and various abnormal hemoglobins interact to produce a wide range of disorders of varying degrees of severity. There are four main categories of interactions associated with severe disease states, for which genetic counseling and prenatal diagnosis are indicated (Old, 1996): thalassemia major (co-inheritance of β- and/or δβ-thalassemia mutations), SCD (and analogous interactions, e.g. Hb S/C, Hb S/β-thalassemia, Hb S/D Punjab, Hb S/O Arab, Hb S/Lepore), Hb E disease (co-inheritance of β-thalassemia mutations with Hb E or Hb O Arab), Hb Bart's Hydrops Fetalis syndrome (homozygous α^0-thalassemia), and (rarely) Hb H Hydrops Fetalis syndrome (α^0/α^T). These disorders are common in the United Kingdom, because of migration of ethnic minority populations.

37.2 PROBLEMS IN GENETIC COUNSELING

Medical training rarely equips doctors to provide adequate genetic counseling and discuss complex issues with their patients, in order to help patients reach their own decisions. The responsibility involved in genetic counseling should not be underestimated. There are ethical problems faced by the counselor as there is no universal model for genetic counseling; counseling is the understanding a set of facts according to the counselor's frame of reference, background in science and genetics and previous training and experience in effectively communicating with the individual/couple. In order to communicate effectively the counselor must take into account the educational and religious background of the individual/couple. The obstacles to effective genetic counseling are emotional conflicts and lack of knowledge of genetics (Clarke, 1990). Genetic counseling should be non-directive and the counselor's main role is to provide information in a non-biased manner and support decisions that are morally right for the individual, couple, or family. When the decision is genuinely their own, the individuals are more likely to be able to live with it. It is often indeed difficult for a counselor to impart this information in an unbiased manner because of his/her patients' family history including parental age, ethnic background, and reproductive history. A counselor may be faced with an at-risk couple, where the woman is 40 years of age, pregnant, with no living children. This couple has the option of prenatal diagnosis, with a 25% chance of an affected fetus and the option of selective termination for an affected fetus, but also a small risk of miscarriage following prenatal diagnosis (Jauniaux and Petrou, 2003). But, there is also the prospect that this may be the last chance for this couple to have a child. Situations such as this may lead the counselor to adopt a directive rather than a non-directive approach to genetic counseling (Clarke, 1991). The major difference between directive and non-directive counseling is whether or not the counselor helps the couple to make a decision that is right for them. Directive counseling can have a positive influence on the couple's decision; the non-directive approach involves presenting the facts in an unbiased manner and helping the couple come to the decision that is right for them and leaving the responsibility of the decision making to the couple. However, can the counselor completely disassociate the couple from their own values and present the information in such a way that they are completely free to make their own decision? Another example of this is the tone or manner that the counselor uses to provide the information to a couple, which is dependent on their level of personal involvement in a particular case. The counselor may change the manner of speech and body gestures, which may in turn influence how a couple interprets the information. For instance, in cases where the counselor feels it is better for the couple to continue a pregnancy without prenatal diagnosis, they could say to the couple "Your chances of a healthy baby are really quite high at 75%." However, for another couple the counselor may say "Your chances of an affected baby with thalassemia major are really quite high at 25%." Therefore the information being imparted to both couples is correct but presented in different ways.

There are other dilemmas, conflicts, and controversies of treatment strategies for thalassemia major (Constantinou, 2003; Savulescu, 2004; Constantinou *et al.*, 2004; Hoffbrand, 2004; Spino, 2004; Olivieri, 2004), which the counselor is faced with that may influence how the counselor transmits the information to an at-risk couple, which in turn may influence the information that the counselor chooses to share with the couple. These conflicts, in turn, cause serious anxieties among the affected patient population.

Genetic diagnosis is often difficult in view of the enormous diversity of the conditions involved, and misdiagnosis and misinformation can have disastrous consequences for individuals and their families.

For instance, when counseling at-risk couples for hemoglobin disorders the genetic counselor should have a good knowledge of the molecular genetics of thalassemia and therefore be able to comprehend the molecular mechanisms, and to communicate this information to the families. This is very important in view of the phenotypic heterogeneity of thalassemia. Generally, the inheritance of two β-thalassemia mutations results in a blood transfusion-dependent thalassemia. However, there are mild β-thalassemia mutations resulting in thalassemia intermedia. For instance, when both partners carry the mild β^+-88 (C > T) (*HBB*:g.-88C > T) thalassemia mutation, the homozygous state generally results in a very mild clinical phenotype. In this case, should the parents consider prenatal diagnosis?

This type of mild thalassemia intermedia poses an ethical dilemma as far as termination of pregnancy is concerned. Health professionals also are faced with this ethical dilemma when couples request termination of pregnancy for a condition such as this. Although they have to accept the couple's decision when termination of pregnancy is requested for a possible mild condition, they themselves are often uneasy about such decisions. The situation is made even more complex, as there is a wide clinical phenotypic heterogeneity in thalassemia intermedia syndromes and it is often not possible to predict how mild the condition will be and therefore impossible to reassure at-risk couples that their affected child will be mildly affected. Even the term ''mild'' raises controversies as its acceptance for one couple may differ with the acceptance for another couple.

However, there are conditions when the couple can be reassured, such as when there are silent β-thalassemia mutations; that is, the β^+ -101 (C > T) (*HBB*:g.-101C > T) that produces mild or very mild clinical phenotypes in the homozygous state and when interacting with severe β-thalassemia mutations, or when one partner carries β-thalassemia trait and the other carries triplicated α-globin genes (ααα) or hereditary persistence of fetal hemoglobin (HPFH). Prenatal diagnosis is not offered for these cases. However, there are other ameliorating factors when the clinical phenotype is not so consistent.

The counselor should have the experience and expertise to communicate this information and the likely outcome to the patients. Indeed, as the knowledge of the molecular mechanisms that cause thalassemia continues to accumulate this scientific area will become even more complex.

In the United Kingdom, sickle cell disease is also common because of the ethnic minority groups present. Counseling couples who are at-risk for sickle cell disorders is often perceived as relatively simple, but in fact it is quite complex because of the wide range of severity of sickle cell disorders, ranging from the very mild to the very severe. As a result, parents face considerable difficulty deciding whether or not to request prenatal diagnosis. A study performed in London showed decision making was greatly influenced by the experiences of sickle cell disease within the family: a couple with an affected child was more likely to undergo prenatal diagnosis and selective termination of an affected fetus than a couple with no such experience. The study also found that uptake of prenatal diagnosis was influenced by the gestation at referral: a woman referred in the second trimester was less likely to proceed with prenatal diagnosis (Petrou *et al*., 1992a).

In contrast, counseling couples at risk for α-thalassemia hydrops fetalis is more straightforward because of the usually hopeless prognosis for an affected fetus and the possibility of life-threatening obstetric risks for the mother (Petrou *et al*., 1992b). It is rare for such couples to decline prenatal diagnosis. However, even here new complexity is introduced by the possibility of initiating regular intra-uterine transfusion for an affected fetus, provided that diagnosis is made early enough (Ng *et al*., 1998). More information is needed about outcomes and the best way to provide this service, if it is requested. An international collaborative study should be conducted, where data on all cases undergoing intra-uterine transfusion including fetal treatment should be collected, and then guidelines on the most successful approaches must be issued. This information is essential before this treatment can be considered as an approach that could be offered widely. Until then, the approach should be applied on a research basis. The main problem with surviving babies of intra-uterine transfusion therapy has been severe mental and physical handicap, though some appear to be doing well. However, is it justified to provide this treatment when the life of an α-thalassemia major fetus is saved, but then all the problems associated with the treatment of β-thalassemia major, such as regular blood transfusions and iron chelation therapy are created?

New medical possibilities created by the rapid advance of genetic technology need to be included in genetic counseling, often before the ethical and moral dilemmas they involve have been adequately considered in some societies. This also concerns the acceptability of prenatal diagnosis and selective abortion of an affected fetus.

In practice, people's options are greatly influenced by the stage in life at which they learn of their risk, and by whether prenatal diagnosis is available or not. If the risk is found before marriage then the options are to remain single, not to marry another carrier, or marry as usual. If, however, the risk is found after marriage, then the options are to separate and find a non-carrier partner, have few or no children, take the chance and have children as usual, or use prenatal diagnosis and selective termination of pregnancy if it is acceptable and available, in case the fetus is affected (see also Chapter 38).

Unless there is an active carrier screening program, it is unusual for a couple to learn they are at risk of having children with a hemoglobin disorder before marriage or before starting a family. At present, in countries where premarital or antenatal screening is not offered, most couples learn of

their risk only after the diagnosis of an affected child, and this limits their choices.

Therefore, there are challenges involved in genetic counseling, because all the available choices involve difficult moral and social problems. In most cases, there appears to be no right answer, but on the other hand, once people understand their risk, they cannot escape from choosing an option. Even the decision not to choose constitutes a choice. In most Western countries, prenatal diagnosis for couples at risk of hemoglobin disorders is available, with the option of selective abortion. It is recognized that selective abortion is not an optimal or easy solution, that early prenatal diagnosis is preferable to later prenatal diagnosis, and that not all couples at risk of having children with thalassemia feel that prenatal diagnosis is the right choice for them. However, these attitudes have evolved only gradually. Prenatal diagnosis became possible in the early 1970s, and at that time its acceptance in different Western countries depended largely on the status of family planning, and the existing abortion law in each country.

It is difficult for people to deal with the idea of prenatal diagnosis unless they are already used to the concept of controlling their own reproduction, a social change that takes time. When genetic diagnosis began to be developed in the 1950s, the options available to couples at reproductive risk were to ignore the information and hope for the best, to remain unmarried or separate, or to limit their reproduction using family planning (i.e. the same choices as those now available in countries where prenatal diagnosis is not available). Family planning was already becoming widely accepted in the West at that time, and it soon became clear that when options were so limited, most people chose either to avoid or ignore information on risk, or to limit the size of their family (Carter, 1974; Modell *et al*., 1980).

37.3 OPTIONS AVAILABLE TO PEOPLE WITH A REPRODUCTIVE RISK

It is useful to examine the options available for people with a genetic reproductive risk in more detail. Because carriers of thalassemia can be detected and advised of reproductive risk either before or after marriage, this opens a wide range of choices for couples. There is extensive experience of the choices that people make in practice.

How can information be obtained on which approaches are acceptable, and which are unacceptable to a particular population? One possibility is through public meetings, discussions, and questionnaires addressed to interested professionals and the general public. However, this presupposes an informed public, which can be found in only a few countries worldwide at present, so this approach also requires a substantial component of public education. The most extensive of such consultation exercises was carried out in the early 1990s by the Canadian Royal Commission on New Reproductive Technologies (Final Report of the Canadian Royal Commission, 1993). The results, like many other surveys, showed that the majority of people, including many who might not use the service themselves for religious or other reasons, approved of the availability of prenatal diagnosis and termination of pregnancy for serious genetic disorders. Another possibility is to seek the views of informed people who are themselves at genetic risk. However, there can be a considerable difference between what people think they would do, and what they actually do when they have to make a choice. Therefore, a more objective approach is to observe and report the choices that people at risk actually make. Data obtained in all these ways is referred to in the following discussion.

Though community-based programs for the prevention of hemoglobin disorders in Europe now include the option of prenatal diagnosis and selective abortion, carrier screening and genetic counseling were introduced in some countries either before prenatal diagnosis was possible (Stamatoyannopoulos, 1974; Angastioniotis and Hadjiminas, 1981; Gamberini *et al*., 1991) or before it was legal (Gamberini *et al*., 1991). This early experience produced useful information on the acceptability of some alternative approaches in Europe.

The core ethical principles of genetic counseling are the autonomy of the individual or couple, their right to full information, and the highest standards of confidentiality (Fletcher *et al*., 1985; Zeesman *et al.*, 1994). This highlights the important role of information in genetics, and the need to develop systems for providing and accessing genetic information. Specialist genetic information must be accessible to health workers, patients and the community in order to permit informed choice. Additionally, informed choice should determine outcome at both the individual and the community level, so patient choices should be recorded, collected, reported, and used to guide service development (Modell *et al*., 2000a).

People at risk, especially those identified by population screening, represent a randomly selected sample of the population. They have to learn in depth about the risk, and often have to use the information in life-determining choices. Each choice represents a complex personal judgment, for example couples at genetic reproductive risk must balance the burden of the disease such as thalassemia against the burden of prenatal diagnosis and termination of a wanted but affected pregnancy. The statistical accumulation of individuals' informed choices represents a collective judgment on what services are available within the health care of the country (Petrou *et al*., 2000; Modell *et al*., 2000b, c; Modell and Petrou, 2003). Since such choices are or should be recorded, registers that aggregate this information and outcomes can be used to listen to the community verdict of their reproductive choices such as the acceptability of prenatal diagnosis.

The outcome of high uptake of prenatal diagnosis by couples at risk of thalassemia in the United Kingdom and Mediterranean countries reflects the at-risk community's view of the heavy burden of the disease and its treatment. However, things can change as treatment becomes more effective and acceptable and uptake of prenatal diagnosis may fall. If new approaches like preimplantation genetic diagnosis (Verlinsky and Kuliev, 2000; see also Chapter 33) become easier and more widely available, then the burden of prevention may decline and uptake may rise.

Life expectancy for thalassemia has improved significantly with modern medical treatment, therefore quality of life (QoL) should now be considered an important factor of effective health care. The burden of treatment and QoL should be discussed and considered within genetic counseling. An assessment of QoL differs from other forms of medical assessment in that it focuses on the individuals' own views of their well-being and assesses other aspects of life giving a more holistic view of well-being. Within countries that provide optimal treatment, thalassemia major patients survive into their 50s and beyond; however, some do not. How does QoL have an impact on survival? QoL is defined as "…individuals' perceptions of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns" (WHO, 1997 (http://www.who.int/mental_health/media/68.pdf)).

QoL is a concept affected in a complex way by the person's physical health, psychological state, personal beliefs, social relationships, and their relationship to salient features of their environment (Telfer *et al*., 2005).

However, at a conference organized by the United Kingdom Thalassemia Society, QoL was defined as: Graduating, having a suitable job, getting married, having children (Constantinou, 2002). What constitutes an acceptable QoL depends also on cultural context, age, and a complex interplay of the social, environmental, and spiritual background of the patient. To achieve these goals it is necessary for patients to be integrated within society and treated as equal members of society. QoL is dependent on factors such as making a success of: education, profession, relationships with their partner and family, and parenthood. However, these are all related to treatment and the creative ways treatment is offered that consequently has an impact on the burden of the disease. Life can be improved by introducing effective therapy, which is easy and convenient, such as evening and weekend transfusions, evening clinics, easy effective iron chelation treatment, psychological support, and very importantly a more tolerant, informed, and accepting society towards affected patients. However, do society's social and cultural beliefs allow patients to integrate into society? Is it possible for quality treatment encompassing an optimal QoL to be integrated into the health system to reduce the burden of the disease? By denying patients this holistic approach the burden of the disease remains high and consequently uptake of prenatal diagnosis remains high. Genetic counselors will always need to bear these issues in mind when counseling individuals and couples.

How does the option of prenatal diagnosis in, say, the United Kingdom vs. Pakistan differ? In Pakistan, it is estimated that 5,250 infants with beta thalassemia major are born annually. Although only a minority of these cases are diagnosed, charitable thalassemia centers sustain thousands of affected children with monthly blood transfusions. Such transfusions permit an excellent quality of life during childhood but lead to iron overload and death in adolescence. Therefore the comparison of QoL based on the definitions provided by the WHO and Constantinou (2002), renders the patient in Pakistan with a low QoL and high burden of the disease.

37.4 PREMARITAL SCREENING

It is often thought that affected births can be prevented if at-risk couples are identified prior to marriage, on the assumption that they will then decide to separate and each find another, non-carrier partner. However, it is not always easy to explain how most marriages come about to be able to make any valid assumptions about how choice of partner might be affected by genetic information. In many societies, marriage is a complex social phenomenon that involves many other family members besides the prospective couple, and marriage partners usually are selected either because of a strong personal preference, or for valid family or traditional reasons, or a mixture of all three.

If a planned marriage needs to be rearranged because both partners carry a genetic disorder, this causes social embarrassment or stigma to the young couple and their families, and there is a risk that the problem will recur if the new partners found are also carriers for the same disorder. For example, if population carrier frequency is 6%, the chance that one or both new partners will be a carrier is 12%. Therefore the recurrence risk for the couples is 12% (or even higher if the new potential partner is a relative).

A second possibility is to marry as planned, but avoid having children altogether. This is always a difficult choice, and is unrealistic in a strongly family-centered society. A more realistic option for married at-risk couples is to limit their family size. If they limit themselves to two healthy children, 56% of them will never have an affected child.

Other options for having children while avoiding a known genetic risk, such as artificial insemination by donor, egg donation, or adoption, so far have not proved popular in any society, and are unacceptable in many communities. Another possibility is to marry and wait to have a family until appropriate methods for prevention become available. For example, when Cypriot at risk couples were informed that prenatal diagnosis would become available in the

foreseeable future, many postponed conceiving until they could use the service (Angastiniotis and Hadjiminas, 1981).

A final possibility is to marry and have a family as usual, trusting in fate. This seems often to be a common choice in the absence of alternatives that are acceptable to families. Before prenatal diagnosis was feasible for hemoglobin disorders, Stamatoyannopoulos (1974) conducted a research study in the area of Arta (Greece), where 20% of the population are carriers of either thalassemia or SCD. All young people in that area were screened and counseled, and counseling contact was maintained for a two-year period. When the pattern of marriages was assessed at the end of this period, there was no measurable effect on choice of partner (Stamatoyanopoulos, 1974).

More recently, mandatory premarital screening for thalassemia and SCD has been conducted in Saudi Arabia with the objective of decreasing at-risk marriages. However, following counselling, almost 90% of couples married despite being aware of their risk (Al Hamdan *et al*., 2007). The option of prenatal diagnosis is usually unavailable in the current Saudi health system.

A similar approach was tried in Cyprus much earlier, when marriages between carriers were actively discouraged. This approach proved to be unacceptable to the population and was soon abandoned ''because of evasions'' (Angastiniotis *et al*., 1986). Once prenatal diagnosis became possible for thalassemia, it was made available within the Cypriot health service. Soon after, confidential premarital screening was made mandatory among Greek Cypriots by the Greek Orthodox Church, and among Turkish Cypriots by the civil authorities. It was then found that 98% of at-risk couples detected just prior to marriage proceed to marry, even though Cypriot parents often have considerable influence on their children's choice of partner. Nevertheless, the annual number of new births of children with thalassemia major has decreased almost to zero in Cyprus, because couples use the information on genetic risk in a variety of ways to obtain a healthy family. It has been reported that less than 5% of the decrease in thalassemia major births is due to separation of engaged couples, about 80% is due to prenatal diagnosis and selective abortion, and 20% is because at-risk couples have fewer children, on average, than couples not at risk (Angastiniotis *et al*., 1986). In both Cyprus and Sardinia, the population is now very well informed, and has gained confidence that the thalassemia control programs involve little if any coercion. There is now increasing popular demand for carrier testing in high schools, so that young people and their families can take information on carrier status into account at an earlier stage in the choice of marriage partner (WHO, 1993). It will be of great interest to continue to follow the results of these Mediterranean social experiments.

Each choice made by at-risk couples within these Mediterranean countries based on informed choice has reduced the birth rate to almost zero. Such uniform decisions by population are often questioned, as they may be led by the existence of bias within communities towards thalassemia. The other word for a thalassemia carrier in Greek is ''stigma'' which means ''a mark of social disgrace''. Does this ''stigma'' still exist in society, despite the many changes that have occurred in society's attitude to thalassemia and improved survival for thalassemia? In the author's practice it is quite common for Cypriot at-risk couples living in the United Kingdom to present in the clinic and immediately demand prenatal diagnosis. These couples carry with them society's preconceived ideas of thalassemia and make comments such as ''children with thalassemia die early, and are sickly''. For these couples to take on board the modern treatment and excellent prognosis is very difficult. This is not to say that modern-day treatment is not without its complications and there are no guarantees (UKTS, 2005; Telfer *et al*., 2006; Modell *et al*., 2000d; B. Modell, in preparation). Therefore it is not surprising that following counseling these couples still proceed with prenatal diagnosis. Therefore stigma is still present although certainly improved and not as deep rooted as in the Middle East and in India where the stigma of having the thalassemia trait is a deterrent to the disclosure of thalassemia status as well as to testing. Being a carrier in India may render an individual unfit as a suitable marriage partner and testing after marriage or prenatal counseling would be more acceptable to the majority. The stigma associated with being a carrier can only be reduced significantly through greater awareness by involving community leaders and people who are involved with arranging marriages (Chattopadhyay, 2006).

In Canada, a program of information and screening for carriers of Tay Sachs disease or thalassemia in high schools in Montreal has proved highly acceptable, and also highly informative about pitfalls in foreseeing how people will use genetic information (Zeesman *et al*., 1984). A fraction of the young people screened was followed up with a questionnaire that included the questions: ''Do you think that a couple planning to marry who found they were both carriers would change their marriage plans? Would you change your own marriage plans?'' Interestingly, almost 80% thought that other couples would change their marriage plans, but only 10% thought they would change their own plans. Clearly, it is all too easy to underestimate the importance of other people's inner lives. Practical experience is the only reliable guide to how people are prepared to use information on genetic risk.

Therefore, there seems to be a strong case for early carrier diagnosis and genetic counseling, so that couples at genetic risk and their families can make an informed choice whether to separate or stay together. However, knowledge of a genetic risk may alone be insufficient to change a lifetime partner.

Prenatal diagnosis is now available in all the countries whose experience has been drawn on earlier. In addition,

assessment of its desirability from the medical, religious, and social points of view has now begun in many other countries, including a large number of Muslim countries (Alwan and Modell, 2003) and indeed prenatal diagnosis is available in some of them, such as Pakistan (Ahmed *et al*., 2000) and Iran (Samavat and Modell, 2004).

37.5 CAN PREVENTION PROGRAMS BE CONSIDERED EUGENICS?

The word *eugenics*, coming from the Greek word *ευγνικλή*, meaning good birth, was coined by the British scientist Sir Francis Galton. Although it still carries its original meaning in some countries, usually it is associated with Nazi programs to eradicate so-called inferior groups.

The WHO has suggested another definition of eugenics: ''A coercive policy intended to further a reproductive goal against the rights, freedoms and choices of the individual'' (WHO, 2001). For this definition coercion includes laws, regulations, positive or negative incentives, including the lack of accessibility to affordable medical services, put forward by states or other social institutions.

Globally, there is little evidence for eugenics practice according to the new definition. It has been suggested that programs, available in countries such as Sardinia and Cyprus, of carrier screening and prenatal diagnosis are eugenics programs, because they have limited the birth of affected individuals. However, in Cyprus couples make their own decisions once mandatory screening takes place. Couples can still decide to marry and have children with thalassemia. Although most marry, very few choose to carry on a pregnancy diagnosed with thalassemia major. In Cyprus there is widespread approval of the program, which is also supported by the majority of the thalassemia major patients. This program gives individuals a choice, whereas if this service did not exist then that choice would not have existed either. Therefore using the suggested definition of the WHO, the Cyprus carrier screening program for thalassemia cannot be considered as eugenics.

It can also be examined why such prevention programmes are supported by the majority of thalassemia major patients. It is possible that the continual birth of affected patients will ultimately have a detrimental effect on the treatment of the existing patients. If there was no prevention program in Cyprus there would be approximately 50 affected births per year. By 2021, a total of 70,000 units of blood will be required and 17.5% of the possible 400,000 possible donors (out of a population of 600,000) could need to donate blood at least once per year (Angastiniotis, 2003), There is also the cost of iron chelation treatment including the cost of the remainder of the treatment, which will ultimately become prohibitive to the economy of Cyprus. Therefore it is not surprising that prevention programmes are supported by the majority of thalassemia major patients.

The fate of many patients in some developing countries where carrier screening programs are not developed is dismal. In many developing countries, the state does not have the resources to treat the affected patients. In Pakistan, most of these patients die because their parents cannot afford treatment. Most of these families were not given a choice, as national screening programs for thalassemia were not available. However, in Pakistan there is now religious approval for termination of pregnancy for a genetic disorder (Ahmed *et al*., 2000) and prenatal diagnosis is available in private clinics (Ahmed *et al*., 2000), although a national carrier screening program is not yet available.

Many countries now have introduced or are developing screening and counseling programs. In Iran without prevention there would be about 1,200 affected children born annually, with over 20,000 children attending treatment centers. Iran has taken on the vast task of providing national premarital screening and genetic counseling. By the end of 2001, over 2.7 million prospective couples had been screened and 10,298 at-risk couples identified and counseled. Fifty-three percent of these couples proceeded with their marriage plans, 29% of at-risk couples separated, and the remainder were still struggling with their decision (Samavat and Modell, 2004). Further recent data show that the number of couples proceeding with marriage has increased even further (A. Samavat, personal communication). Therefore the majority of couples find it unacceptable to select a partner on the basis of genetic screening information and there is a high demand for prenatal diagnosis. The earlier results of the programme were considered at the highest religious and political levels and a fatwa was issued permitting first trimester abortion when a fetus is found to have a serious genetic disorder. A national network of molecular diagnostic laboratories was set up to make prenatal diagnosis available within the health system.

Therefore, prevention services are steadily spreading globally and services are now available in several low resourced countries (Petrou and Modell, 1995). The ethos in present-day medical genetics is to help people make whatever voluntary decisions are best for them in the light of their own reproductive and other goals. There is, therefore, a decisive difference between present-day medical genetics and yesterday's eugenics.

37.6 ETHICS AND RELIGION IN GENETIC COUNSELING

As discussed earlier, genetic counseling should be nondirective, and the genetic counselor's main role is to provide people at risk with full information, give them time for consideration, and support them in making the decisions they feel to be morally right for themselves.

Some reasons why genetic counseling should be non-directive are:

- People at risk often have first-hand experience of the condition in question, unlike most of their advisors,
- They have to learn all the facts, think the issues through, and actually reach a decision that they must live with for the rest of their lives,
- The right choice for a given individual among the options actually available is likely to be determined by many factors, including their social and religious attitudes, personal experiences, economic and educational level, and family and reproductive history,
- Doctors and other professionals are no more, or less, qualified than their patients to make moral choices on issues associated with a person being made aware of their genetic risk.

Medical ethics are based on the moral, religious, and philosophical ideals, and the principles of the society in which they are practiced. It is therefore not surprising to find that what is considered ethical in one society might not be considered ethical in another. The wide range of family and social structures, religious and legal conventions, and economic resources within the Middle East may also lead to conclusions that differ between countries. Sensitive services like genetic counseling cannot be transported from one social context into another.

Truly ethical conduct consists of personal searching for relevant values that lead to an ethically inspired decision. Surveillance or audit shows how information on parental informed choice is obtained and this reflects what is ethically acceptable in a particular society (Modell *et al*., 1997, 2000b, c; Petrou *et al*., 2000). It is necessary for practitioners and critics of conduct to be sensitive to such information before they make their judgments on what is acceptable. The ethical attitude of the individual, whether a patient or a genetic counselor, is often colored by the attitude of their society. Social attitudes are influenced by theologians, demographers, family planning administrators, doctors, policymakers, sociologists, economists, and legislators. All these groups should also consider the fundamental ethical principles of genetic counseling, particularly the autonomy of the individual, and place emphasis primarily on informed parental choice by those at risk before making judgments for society.

All medical programs, including genetic prevention programs, must operate within existing legal and social frameworks. However, technology can develop rapidly, whereas legal, social, and religious attitudes evolve more slowly.

In most religions, there is a range of opinions and dilemmas. The high uptake of prenatal diagnosis in Italy, Cyprus, and Greece clearly shows that people make their own choices for what is appropriate to them. There is no right answer. The right answer is the one that is acceptable to the individual for his/her own particular reasons.

As several Muslim countries are already offering prenatal diagnosis and selective abortion to at-risk couples or are at various stages of developing these services, it is important to comment on the information that is already available on their acceptability in these countries.

For many, Muslim religion is central to daily life and influences a lot of behaviors, attitudes, practices, and policy making (Serour, 2000). Instructions regulating everyday activities of life to be followed by Muslims are called *Sharia*. The Sharia is not rigid or fixed, except for a few rules such as those concerning worship, rituals, and codes of morality. Islamic Sharia accommodates different honest opinions as long as they do not conflict with the spirit of its primary sources and are directed toward the benefit of humanity. In Muslim societies, the provisions and spirit of Islamic Sharia and the local and social conditions of the society should be taken into account when formulating rules and guidelines.

Fetal development has been viewed by Muslim theologians as occurring in three stages, each lasting 40 days: the sperm cell and ovum, the clump resembling a blood clot, and the lump of flesh (fetus). At the end of these stages, the fetus is ensouled. The belief that ensoulment starts only at 120 days does not change the fact that life starts at a much earlier stage of embryo life.

Therefore from a Muslim perspective, it is considered ethical to perform an abortion to protect the mother's life or health, or because of fetal anomaly incompatible with life (Glover, 1989; Serour *et al*., 1995). However, the stage at which termination is deemed permissible seems to vary. Some Muslim jurists do not allow abortion in any circumstances, whereas others would permit abortion in the first 120 days of fetal life if there is some reason such as danger to the mother or the fetus. The time when abortion is allowed also varies, and some jurists would allow it only at 40 days and others at 90 days.

Therefore most Muslims accept that life does not begin at conception, but believe that human life requiring protection commences some weeks, perhaps two weeks or so, after development of the primitive streak (Serour, 2001; Serour and Dickens, 2001). Therefore preimplantation genetic diagnosis is encouraged, where feasible, as an option to avoid termination of pregnancy (Serour, 2001; Serour and Dickens, 2001). The importance of preimplantation genetic diagnosis was recognized at the international workshop at The International Islamic Centre for Population Studies and Research, Al-Azhar University, Cairo, on Ethical Issues in Assisted Reproductive Technology.

It should be recognized that ethical and religious reasoning on the same issue can justify different conclusions. Therefore, it is acceptable that adherents of one preferred outcome may well acknowledge that adherents of an alternative preferred outcome are applying approaches that result in different but equally ethical conclusions.

37.7 CONSANGUINEOUS MARRIAGE

Consanguineous marriage is usually defined as marriages between people who are second cousins or closer (Bittles, 1994). The chances of inheriting two identical genes, including neutral as well as pathological genes, at a particular locus are increased if parents are close relatives (Bodmer and Cavalli-Sforza, 1978).

The question of how to provide genetic counseling in the context of societies that favor consanguineous marriage can also create dilemmas for families and health workers. Any ideal genetics program must include a sensitive and realistic approach.

Over 20% of the world population lives in communities that favor cousin marriage, and worldwide at least 1 in 12 children is born to parents who are related (Modell and Darr, 2002). Cousin marriage has been customary in many parts of the world for thousands of years, and is not always associated with Muslim religion. A custom that has been so common for so long obviously has important social functions. Currently, in Northern European populations, about 0.5% of marriages are between first cousins.

The great majority of families where the parents are related suffer no adverse effect. The reported increase in average childhood mortality and morbidity in such populations is due largely to relatively severe effects in a limited number of families, which shifts the average figures for the group as a whole. Therefore medical attempts to help families reduce genetic problems should focus on identifying families at particularly high risk, and providing them with genetic counseling and access to appropriate services.

Nevertheless, efforts have been made in some countries to discourage consanguineous marriage through public information programs that emphasize the associated genetic risks. However, this proposal is inconsistent with the ethical principles of genetic counseling, overlooks the social importance of consanguineous marriage, and is ineffective (Modell and Darr, 2002). These programs arise from the perception that recessively inherited rare diseases are unusually common in these populations, and the need for prevention seems urgent, but appropriate genetic counseling services are rudimentary or non-existent. It may then seem that altering the behavior of the population is the only possible way to reduce genetic disease incidence.

In view of the ignorance of the social causes and consequences of customary consanguineous marriage, attempts to reduce its frequency in the population as a whole on genetic grounds run the risk of doing more harm than good, by disturbing customary marriage arrangements when the majority of families would come to no harm in any case. Furthermore, the concept of non-directive genetic counseling is incompatible with a campaign against consanguineous marriage. Policies relating to consanguineous marriage should be firmly grounded in an understanding of its social role, and the possible consequences of attempts to disturb it.

Discussions in the United Kingdom with families who have had pressure against consanguineous marriage put on them led to the following conclusions:

- Pressure against cousin marriage rarely alters what people actually do, though it can make them feel uncomfortable about it.
- If people are told their children are sick because they are related, it causes great unnecessary distress, may alienate them, and makes it more difficult for them to understand the real explanation.
- Where cousin marriage is common, people are aware that most couples of cousins have perfectly healthy children. If they are told not to marry a cousin because their children may be sick, they may become confused and could lose confidence in medical advice.
- Avoiding cousin marriage does not guarantee that children will not have a congenital disorder. Unrelated couples who have affected children may lose confidence in medical advice.
- People do not attend genetic counseling if they think they will be criticized and their cultural conventions attacked (WHO EMRO, 1997; Modell and Kuliev, 1992).

37.8 WHY COUSIN MARRIAGE IS FAVORED BY SOME COMMUNITIES

Many non-European communities have a patrilineal kinship pattern (family name and property are inherited in the male line). The men and their descendants tend to stay together, especially when the family owns land and members of the extended family share responsibilities for each other. At marriage, women leave their family to enter their husband's family. Cousin marriage can soften the implications of this transition and contribute to family well-being by strengthening the woman's position within the family and promoting female networks. Darr and Modell (2002) made the following conclusions in this issue:

- Parents are able to remain closer to their children, particularly their daughters, when they marry a relative,
- A woman is likely to be comfortable with a mother-in-law who is also an aunt she has known since childhood,
- There is increased financial security, because if one partner dies the remaining partner is still a member of the family in their own right,
- Equal numbers of sons and daughters are needed within the extended family, so a daughter is not seen as a burden,
- In societies that practice segregation between the sexes, young people can get to know each other before marriage more easily if they are related,
- Expenses and exchange of property associated with marriage may be reduced.

A convention of cousin marriage could make family-oriented genetic counseling particularly effective, for two main reasons. First, unusually large numbers of carriers of the presenting disorder may be detected within the family. Thus carrier testing, for example, may permit early detection of many individuals at risk for some types of thalassemia and sickle cell anemia and identify newly affected individuals where surveillance and early treatment could be beneficial. Second, when cousin marriage is common within the family, carriers are at particularly high risk of making an at-risk marriage. Family studies following the diagnosis of a child with a disorder will identify many carriers, who should then be offered counseling about their reproductive risk. Many will already be married, and some of these may be at-risk couples, identified in time for prospective reproductive counseling. Many will be children or not yet married, and early information might permit carrier status to be taken into account to avoid further at-risk marriages within the family. This approach has been tested in Pakistan (Ahmed *et al*., 2002).

Consanguineous marriage increases the couple's chance of both carrying the same mutation. Therefore from the genetic testing point of view, a family-oriented approach to genetic screening becomes more efficient, because it simplifies the challenges of DNA testing. In a recent study in Jordan, where homozygous thalassemia births were studied, it was found that the homozygous genotype was approximately 5.5 times higher in consanguineous marriage than random mating, giving 62% homozygous births versus 11.5% by random mating (Qubbaj and Petrou, in preparation).

37.9 CONCLUSIONS

The medical profession's responsibility starts from the time prenatal diagnosis becomes available and therefore a new social responsibility is generated for every subsequent birth. This responsibility should be passed onto the parents through adequate information and genetic counseling. The availability of genetic diagnosis increases the responsibility of the medical service when providing this service to the population. It is then the parents' responsibility to make the decision. The most common single saying of couples at risk for thalassemia is: ''We have no choice: we cannot knowingly bring a child into the world to suffer.'' The other face of responsibility is guilt. From counseling at-risk couples in the United Kingdom it is known that parents want to take this responsibility of life or death. Couples with a thalassemia major child frequently want to sue the health professionals for negligence, because they have not been identified and counseled during pregnancy. Their complaint is that they have been deprived of information and choice, and this increases their difficulties because their child's suffering is now perceived as an injustice rather than as a simple misfortune. They are expressing a strong conviction that since knowledge is available, someone is responsible for the birth of their affected child and that the responsibility that should have been their own has been taken from them (Modell, 1988). This also helps them remove the anger and guilt they have.

The ethical principles governing genetic counseling need to be assessed for each country, after taking into account the social and religious structures of that country. This is particularly important for issues surrounding prenatal diagnosis and counseling in relation to customary consanguineous marriage.

REFERENCES

Ahmed, S., Saleem, M., Sultana, N., Raashid, Y., Waqar, A., Anwar, M., Modell, B., Karamat, K.A., and Petrou, M. (2000). Prenatal diagnosis of beta thalassaemia in Pakistan: experience in a Muslim country. *Prenat. Diagn.* 20, 378–383.

Ahmed, S., Saleem, M., Modell, B., and Petrou, M. (2002). Screening extended families for genetic haemoglobin disorders in Pakistan. *N. Engl. J. Med.* 347, 1162–1168.

Akinyanju, O.O. (1989). A profile of sicle cell disease in Nigeria. *Ann. NY Acad. Sci.* 565, 126–136.

Al Hamdan, A.R.N., Almazrou, Y.Y., Alswaidi, M.F., and Choudhry, A.J. (2007). Premarital screening for thalassemia and sickle cell disease in Saudi Arabia. *Genet. Med.* 9, 372–377.

Alwan, A., and Modell, B. (2003). Recommendations introducing genetic services in developing countries. *Nat. Rev. Genet.* 4, 61–68.

Angastiniotis, M.A., and Hadjiminas, M.G. (1991). Prevention of thalassaemia in Cyprus. *Lancet* 1, 369–370.

Angastiniotis, M., and Modell, B. (2003). Recommendations for introducing genetic services in developing countries. *Nat. Rev. Genet.* 4, 61–67.

Angastiniotis, M.A., Kyriakidou, S., and Hadjiminas, M. (1986). How thalassaemia was controlled in Cyprus. *World Health Forum* 7, 291–297.

Angastiniotis, A. (2003). Why prevention. In *Prevention of Thalassaemias and other Haemoglobin Disorders* (Galanello, R., Eleftheriou, A., Traeger-Synodinos, J., Old, J., Petrou, M., and Angastiniotis, M., eds), Vol. 1, pp. 1–9. Thalassaemia International Federation (TIF).

Bittles, A.H. (1994). The role and significance of consanguinity as a demographic variable. *Pop. and Develop. Rev.* 20, 561–584.

Bodmer, W.F., and Cavalli-Sforza, L.L. (1978). *Genetics Evolution and Man*. W.H. Freeman and Company, San Francisco.

Carter, C.O. (1974). Current status of genetic counselling and its assessment. In *Birth Defects* (Motulsky, A., and Lenz, W., eds). Excerpta Medica. Amsterdam.

Chattopadhyay, S. (2006). Rakter dosh-corrupting blood: the challenges of preventing thalassaemia in Bengal, India. *Soc. Sci. Med.* 63, 2661–2673.

Christianson, A., Howson, C.P., and Modell, B. (2006). *March of Dimes Birth Defects Foundation*. White Plains, New York.

Clarke, A. (1990). Genetics, ethics and audit. *Lancet* 335, 1145–1147.

Clarke, A. (1991). Is non-directive genetic counselling possible?. *Lancet* 338, 988–1001.

Constantinou, G. (2002). Quality of My Life, Conference, Optimising Thalassaemia Care: Working Towards National Standards, organised by the United Kingdom Thalassaemia Society, London, UK.

Constantinou, G. (2004). The disappearing patient. *Br. Med. J.* 328, 358–359.

Constantinou, G., Melides, S., and Modell, B. (2003). The Olivieri case. *N. Engl. J. Med.* 348, 860–863.

Darr, A., and Modell, B. (2002). Science and society: genetic counselling and customary consanguineous marriage. *Nat. Rev. Genet.* 3, 225–229.

Final Report of the (Canadian) Royal Commission on New Reproductive Technologies. Canada Communications Group (1993). Proceed with Care-Publishing. Ottawa, Canada, K1A OS9.

Fleming, A.F., Storey, J., Molineaux, L., Iroko, E.A., and Attai, E.D. (1979). Abnormal haemoglobins in the Sudan savannah of Nigeria. 1. Prevalence of haemoglobins and relationships between sickle cell trait, malaria and survival. *Ann. Trop. Med. Parasitol.* 73, 161–172.

Fletcher, J.C., Berg, K., and Tranoy, K.E. (1985). Ethical aspects of medical genetics: a proposal for guidelines in genetic counselling, prenatal diagnosis and screening. *Clin. Genet.* 27, 199–205.

Gamberini, M.R., Lucci, M., Vullo, C., Anderson, B., Canella, R., and Barrai, I. (1991). Reproductive behaviour of families segregating for Cooley's anaemia before and after the availability of prenatal diagnosis. *J. Med. Genet.* 28, 523–529.

Glover, J. (1989). *Ethics of New Reproductive Technologies: The Glover Report to the European Commission, Studies in Biomedical Policy*. Northern Illinois University, Dekalb, IL.

Harper, P. (1988). *Practical Genetic Counselling*, 3rd edition. Wright, Bristol.

Hoffbrand, V.A. (2004). Deferiprone. *Br. Med. J.* 328, 358–359.

Jauniaux, E., and Petrou, M. (2003). Fetal sampling for the haemoglobin disorders. In *Prevention of Thalassaemias and other Haemoglobin Disorders* (Galanello, R., Eleftheriou, A., Traeger-Synodinos, J., Old, J., Petrou, M., and Angastiniotis, M., eds). Chapter 6, Vol. 1, pp. 83–91. Thalassaemia International Federation (TIF).

Modell, B., Ward, R.H.T., and Fairweather, D.V.I. (1980). Effect of introducing antenatal diagnosis on the reproductive behaviour of families at risk for thalassaemia major. *Br. Med. J.* 280, 1347–1350.

Modell, B. (1988). Ethical and social aspects of fetal diagnosis for the haemoglobinopathies: a practical view. In *Prenatal Diagnosis Thalassaemia and the Haemoglobinopathies* (Loukopoulos, D., ed.). CRC Press Inc, Boca Raton, Florida.

Modell, B., and Kuliev, A.M. (1992). Social and genetic implications of customary consanguineous marriage among British Pakistanis. *Galton Institute Occasional Papers*, Second Series, No. 4. The Galton Institute, 19 Northfields Prospect, Northfields, London.

Modell, B., Petrou, M., Layton, M., Varnavides, L., Slater, C., Ward, R.H., Rodeck, C., Nicolaides, K., Gibbons, S., Fitches, A., and Old, J. (1997). Audit of prenatal diagnosis for haemoglobin disorders in the United Kingdom: the first 20 years. *BMJ* 315, 779–784.

Modell, B., Khan, M., Darlison, M., King, A., Layton, M., Old, J., Petrou, M., and Varnavides, L. (2000a). A national register for surveillance of inherited disorders: beta thalassaemia in the United Kingdom. *Bull. World Health Organ.* 79, 1006–1013.

Modell, B., Harris, R., Lane, B., Khan, M., Darlinson, M., Petrou, M., Old, M., Layton, M., and Varnavides, L. (2000b). Informed choice in genetic screening for thalassaemia during pregnancy: audit from a national confidential enquiry. *Br. Med. J.* 320, 337–341.

Modell, B., Darlison, M., Khan, M., and Harris, R. (2000c). Role of genetic diagnosis registers in ongoing consultation with the community. *Community Genet.* 3, 144–147.

Modell, B., Khan, M., and Darlinson, M. (2000d). Survival in beta thalassaemia major in the UK: data from the UK Thalassaemia Register. *Lancet* 355, 2051–2052.

Modell, B., and Darr, M. (2002). Genetic counselling and customary consanguineous marriage. *Nat. Rev. Genet.* 3, 223–229.

Modell, B., and Petrou, M. (2003). Surveillance. In *Prevention of Thalassaemias and other Haemoglobin Disorders* (Galanello, R., Eleftheriou, A., Traeger-Synodinos, J., Old, J., Petrou, M., and Angastiniotis, M., eds), Vol. 1, pp. 133–147. For the Thalassaemia International Federation (TIF).

Ng, P.C., Fok, T.F., Lee, C.H., Cheung, K.L., Li, C.K., So, K.W., Wong, W., and Yuen, P.M. (1998). Is homozygous alpha thalassaemia a lethal condition in the 1990s? *Acta Paediatr.* 87, 1197–1199.

Old, J. (1996). Haemoglobinopathies. *Prenat. Diagn.* 16, 1181–1186.

Olivieri, N.F. (2004). To the editor. *Br. Med. J.* 328, 358–359.

Patrinos, G.P., Giardine, B., Riemer, C., Miller, W., Chui, D.H., Anagnou, N.P., Wajcman, H., and Hardison, R.C. (2004). Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. *Nucleic Acids Res.* 32, D537–D541.

Petrou, M., Brugiatelli, M., Ward, R.H.T., and Modell, B. (1992a). Factors affecting the uptake of prenatal diagnosis for sickle cell disease. *J. Med. Genet.* 2, 820–823.

Petrou, M., Brugiatelli, M., Old, J., Ward, R.H.T., and Modell, B. (1992b). Alpha thalassaemia hydrops fetalis in the UK: the importance of screening pregnant women of Chinese, other South East Asian and Mediterranean extraction for alpha thalassaemia. *Br. J. Obstet. Gynaecol.* 99, 985–989.

Petrou, M., and Modell, B. (1995). Prenatal screening for haemoglobin disorders. *Prenat. Diagn.* 15, 1275–1295.

Petrou, M., Modell, B., Shetty, S., and Ward, R.H. (2000). Long term effects of prospective detection of high genetic risk on couples reproductive life: data for thalassaemia. *Prenat. Diagn.* 20, 469–474.

Samavat, A., and Modell, B. (2004). Iranean national thalassaemia screening programme. *BMJ* 329, 1134–1137.

Savulescu, J. (2004). Thalassaemia major: the murky story of deferiprone. *BMJ* 328, 358–359.

Serour, G.I., Aboulghar, M.A., and Mansour, R.T. (1995). Bioethics in medically assisted conception in the Muslim world. *J. Assist. Reprod. Genet.* 12, 559–565.

Serour, G.I. (2000). Ethical considerations of assisted reproductive technologies: a Middle Eastern perspective. *Middle East Fertil. Steril. J.* 5, 13–18.

Serour, G.I. (2001). Attitudes and cultural perspectives on infertility and its alleviation in the Middle East Area. Paper presented at the WHO Expert Group Meeting on ART, Geneva, Sept. 2001.

Serour, G.I., and Dickens, B.M. (2001). Assisted reproductive developments in the Islamic world. *Int. J. Gynecol.* 74, 187–193.

Spino, M. (2004). The true ethics of deferiprone debate. *Br. Med. J.* 328, 358–359.

Stamatoyannopoulos, G. (1974). Problems of screening and counselling in the hemoglobinopathies. In *Birth Defects* (Motulsky, A.G., and Lenz, W., eds), pp. 268–276. Excerpta Medica, Amsterdam.

Stuart, M.J., and Nagel, R.L. (2004). Sickle cell disease. *Lancet* 354, 1343–1360.

Telfer, P., Coen, P.G., Christou, S., Hadjigavriel, M., Kolnakou, A., Pangalou, E., Pavlides, N., Psiloines, M., Simamonian, K., Skordos, G., Sitarou, M., and Angastiniotis, M. (2006). Survival of medically treated thalassaemia patients in Cyprus. Trends and risk factors over the period 1980–2004. *Haematologica* 91, 1187–1192.

Telfer, P., Constantinou, G., Andreou, P., Christou, S., Modell, B., and Angastiniotis, M. (2005). Quality of life in Thalassamia. *Ann. NY Acad. Sci.* 1054, 273–282.

United Kingdom Thalassaemia Society, Standards for the Clinical Care of Children and Adults with Thalassaemia in the UK (2005).

Van Baal, S., Kaimakis, P., Phommarinh, M., Koumbi, D., Cuppens, H., Riccardino, F., Macek, M., Jr., Scriver, C.R., and Patrinos, G.P. (2007). FINDbase: a relational database recording frequencies of genetic defects leading to inherited disorders worldwide. *Nucleic Acids Res.* 35, D690–D695.

Verlinsky, Y., and Juliev, A.M. (2000). *An Atlas of Preimplantation Genetic Diagnosis*. Parthenon Publishing, New York.

Weatherall, D.J., and Clegg, J.B. (2001). *The Thalassaemia Syndromes*, 4th edition. Blackwell Scientific Publications, Oxford.

WHO (1993). Unpublished report WHO/HDP/TIF/HA/93.1 Joint WHO/TIF meeting on the prevention and control of haemoglobinopathies (7th meeting of the WHO Working Group on the Control of Hereditary Anaemias). Nicosia, Cyprus 3–4 April 1993 (unpublished report WHO/HDP/TIF/HA/93.1).

WHO (1994). *Guidelines for the control of haemoglobin disorders*. WHO, Geneva, Switzerland.

WHO/EMRO Technical Publications Series 24 (1997). Community Control of Genetic and Congenital Disorders. Available from WHO Regional Office for the Eastern Mediterranean, Alexandria, Egypt.

WHO, Report of Consultants to WHO (2001). Review of Ethical Issues in Medical Genetics, WHO/HGN/ETH/00.4.

WHO (1997). WHOQOL Measuring Quality of Life, WHO/MSA/MNH/PSF/97.4.

Zeesman, S., Clow, C.L., Cartier, L., and Scriver, C.R. (1984). A private view of heterozygosity: eight-year follow-up study on carriers of the Tay-Sachs gene detected by high-school screening in Montreal. *Am. J. Med. Genet.* 18, 769–778.

Chapter 38

Genetic Testing and Psychology

Andrea Farkas Patenaude
Division of Pediatric Oncology and Cancer Risk and Prevention Program, Dana-Farber Cancer Institute; Department of Psychiatry, Harvard Medical School, Boston MA, USA

38.1 INTRODUCTION

The completion of the comprehensive sequence of the human genome in the spring of 2003 was hailed as a ''landmark event'' and the beginning of a ''revolution in biological research'' (Collins *et al*., 2003). The huge international cooperative effort that comprised the Human Genome Project (HGP) produced a tool of enormous importance to basic scientists, clinical researchers, clinicians, and, ultimately, patients. Although a great deal already has been learned about basic biology from the HGP, including the fact that humans have only about 30,000 protein-coding genes, many fewer than originally estimated, the most interesting challenges lie ahead (International Human Genome Consortium, 2001; Feero *et al*., 2008). With the human genome sequence readily available to all, the challenge will be to ''capitalize on the immense potential of the HGP to improve human health and well-being'' (Collins *et al*., 2003).

Genetic advances are changing the practice of medicine in many significant ways. Though some may argue over the extent and the pace of changes to come as a result of the HGP (Holtzman and Marteau, 2000), it is clear that the ability to define, diagnose, detect, and prevent, treat, or cure many illness in the future will involve technologies that have their roots in recent genetic advances (Wolf *et al*., 2000; Joshi *et al.*, 2008). On the occasion of completion of the human genome sequence, the US National Human Genome Research Institute put forth a blueprint for future genomic research. The second of three major challenges laid out was labeled ''Genomics to Health'' and involves the translation of ''genome-based knowledge into health benefits''.

The goals of this part of the work of genomic scientists and clinical researchers are to:

- ''Identify genes and pathways with a role in health and disease, and determine how they interact with environmental factors,
- Develop, evaluate, and apply genome-based diagnostic methods for the prediction of susceptibility to disease, the prediction of drug response, the early detection of illness, and the accurate molecular classification of disease, and
- Develop and deploy methods that catalyze the translation of genomic information into therapeutic advances'' (Collins *et al*., 2003).

It is argued that the achievement of these lofty aims is dependent on psychological factors in a variety of ways. Whether patients are willing to participate in genetic research, whether they opt to undergo genetic counseling (see also Chapter 37) or testing to determine if they have increased disease risk, whether they are willing to adopt screening or surveillance recommendations targeted to those known or suspected of being at high risk, or whether they opt for prophylactic surgery, depend in large measure on emotional, attitudinal, and behavioral factors. The recruitment of appropriate, diverse patients for genetic research studies will necessitate improved understanding of what patients bring to genetic counseling or testing from their own family experience with illness, their perception of disease risk, and their attitudes toward genetics research and the medical community. Individuals from different ethnic and cultural communities vary widely in their attitudes toward genetics, genetic testing, and genetic medicine.

Subjective feelings often play a large role in determining how people respond to situations they consider threatening. Research suggests that individuals often act on the basis of subjective feelings about their disease risk, even after receiving counseling that informs them that their risk perceptions are overestimates of actual risk (Iglehart *et al.*, 1998). How can such responses be modified to be in tune with actual risk? What kinds of counseling techniques need to be developed to guide patients toward selecting one of the appropriate choices for mutation carriers? How can people who are found to be at increased hereditary

risk but who do not heed surveillance or screening recommendations be educated? How can people be helped to make optimal personal decisions about prophylactic surgery, which may be the risk reduction option which conveys the greatest protection against developing disease but at significant emotional cost? What is known about how individuals in high-risk families communicate about hereditary disease predisposition with their relatives?

Much of the ultimate success of the HGP in improving human health rests on whether individuals, suspected or known to be at risk for hereditary illness, feel sufficiently comfortable with the use of genetic technology and with the emotional impact of genetic information to make use of the revolutionary advances which the HGP has made possible. This chapter attempts to summarize what is currently known about psychological factors as they affect individuals and families with concern about hereditary predisposition to illness.

With the advent of genetic testing, first for Huntington's disease (HD) and over the past decade for some types of cancer, we have been able to observe reactions to the increased availability of genetic information in individuals from hereditary disease families. In the USA and in several other countries, studies of the ethical, legal, social, and psychological impact of genetic testing have been undertaken simultaneously with the molecular studies of the human genome. It is critical to learn under what circumstances genetic advances improve people's lives and health, as well as understanding under what circumstances adverse outcomes result from participation in genetic research, genetic counseling, and/or genetic testing.

38.2 GETTING TO THE TEST: AWARENESS, ACCESS, AND ADVERTISING

38.2.1 What Tests are there?

Linkage analysis for HD has been possible since 1983 (Gusella *et al*., 1983) and direct sequence analysis, a much more accurate form of genetic testing, became possible after the cloning of the *HD* gene in 1993 (Huntington's Disease Collaborative Research Group, 1993). The first cancer genetic testing was for *Rb1*, the gene responsible for hereditary forms of retinoblastoma, a relatively rare cancer of the eye, which was cloned in 1986 (Friend *et al*., 1986). Testing for *p53* mutations in members of families with Li-Fraumeni syndrome (LFS) began in the mid-1990s. LFS is a rare condition in which mutation carriers have a 90% lifetime risk of developing a number of common cancers including breast cancer, leukemia, sarcomas, and brain tumors, including a 40% chance of developing cancer in childhood (Williams and Strong, 1985). The first major cancer susceptibility genes that affected a large number of people and that achieved significant notice in the popular press were *BRCA1*, which was cloned in 1994 (Miki *et al*., 1994), and *BRCA2*, cloned in 1995 (Wooster *et al*., 1995). These two genes predispose female mutation carriers to greatly increased risks of developing both breast and ovarian cancer. Breast cancer risk for *BRCA1/2* carriers has been estimated as between 56% (Struewing *et al*., 1997) and 85% (Ford *et al*., 1994), with ovarian cancer risk estimates running from 27% (Ford *et al*., 1998) to 60% (Easton *et al*., 1995). Men who are mutation carriers also have an increased risk of breast cancer, although the risk for men is much smaller than for women (Struewing *et al*., 1995b; Wooster *et al*., 1994). Hereditary cancers typically occur at younger-than-usual ages and it is not unusual for multiple cancers to occur.

Genetic testing is also now possible for a number of colon cancer genes, such as *APC*, which predisposes to familial adenomatous polyposis (FAP), and *MLH1*, *MSH2*, *MSH6*, and *PMS2* which predispose to hereditary non-polyposis colorectal cancer (HNPCC, Giardiello *et al*., 2001; Lynch *et al*., 2008). Genetic testing is also possible for a number of quite rare cancer genetic syndromes, such as Von Hippel-Lindau syndrome and Cowden's syndrome and for some members of families with significant history of melanoma. Genetic testing may not be far off also for hereditary prostate cancer. It is estimated that between 5% and 10% of all cancers have a hereditary etiology (Offit, 1998). Because most of the psychological studies of genetic testing have been done for *HD* and more recently for *BRCA1/2* and colon cancer genes, these studies will be the major focus of the discussion.

38.2.2 Genetic Testing

Genetic testing first was done in research settings and offered to members of families known to have had unusual numbers of affected individuals (that is, individuals diagnosed with the disease). Members of these families tended to have been aware for a long time, sometimes several generations, about the unusual concentration of disease in their families. Many were from families with extremely high concentrations of hereditary cancer, since it was the most affected families who first came to the notice of clinicians and disease registries. Family members had often observed several early deaths from the disease in their family and had grown up fearing that this would be their fate as well. Prior to the availability of genetic testing, early research studies asked members of these hereditary disease families if they thought they would want to be tested when the relevant gene was found. Interest was high (typically over 70%) among members of HD and breast/ovarian syndrome families prior to cloning of the respective genes (Mastromauro *et al*., 1987; Jacopini *et al*., 1992; Lerman *et al*., 1995). Of great interest is the fact that with HD, the uptake of genetic testing since testing became available among members of affected families has been only about 5% to 20% (Harper *et al*., 2000; Meiser and

Dunn, 2001; Creighton *et al*., 2003). HD is a devastating, neurodegenerative disease. The *HD* gene is 100% penetrant, meaning that knowledge of being a mutation carrier is equivalent to knowledge that one will certainly develop the disease. There are no preventive options or cures for HD. The earlier arguments put forth by at-risk individuals that testing would enable them to plan better for the future or to achieve certainty about whether or not they were at risk apparently paled in comparison with the possible pain of discovering, before symptoms occurred, that one would, with certainty, be subject to the fate of others in the family already affected with HD.

It had been conjectured that for *BRCA1/2*, uptake would be much more similar to projected interest, since there are some measures that are recommended for mutation carriers to try to prevent or detect cancer early. However, although the uptake for *BRCA1/2* testing – for example, 27% to 43% (Lerman *et al*., 1996; Julian-Reynier *et al*., 2000; Peters *et al*., 2005; Holloway *et al*., 2008) – has been higher than for HD, it has not approached the levels reported for hypothetical interest before testing was actually possible.

Researchers are still in the process of understanding the decision making of individuals who opt not to undergo genetic testing or, often, even genetic counseling for hereditary cancer. There is some research suggesting detrimental emotional effects, such as continuing depression, among those who refuse testing (Lerman *et al*., 1998; Kash *et al*., 2000), but further research is needed among larger and more diverse samples to fully understand the psychology of ''not wanting to know'' and the barriers to acceptance of genetic counseling and testing (Foster *et al*., 2004). Some investigations have found it is patients with lower chances of being mutation carriers who are most likely to want to be tested, since testing, they hope, can serve as reassurance that they are not at increased hereditary risk of disease. It is possible to say with some certainty, however, that hypothetical estimates of interest in genetic testing are not accurate predictors of eventual uptake.

Cancer genetic testing has moved beyond the research setting into clinical testing. Clinical testing makes it likely that individuals with less extensive family histories of disease will be tested; the impact of this change has not yet been fully evaluated (Bonadona *et al*., 2005; DeMarco *et al*., 2006). Some patients newly diagnosed with breast cancer are being offered upfront genetic testing, even before treatment options are in place. Genetic testing at diagnosis increases uptake of prophylactic mastectomy among mutation carriers (Weitzel *et al*., 2003; Schwartz *et al*., 2004); the quality of life impact of ''upfront genetic testing'' is not yet established although one study suggests that patients who elect contralateral prophylactic mastectomy (CPM) after ''upfront genetic testing'' do not experience more distress than patients who do not undergo CPM in the year following breast cancer diagnosis (Tercyak *et al*., 2007). Myriad Genetics, Inc. has test kits for breast, colon, and melanoma genetic testing that can be accessed by any physician who wishes to test his/her patient, and has recently begun direct-to-consumer advertising. Some companies offer patients direct access to test kits via the internet (Williams-Jones, 2003), which raises significant issues among professional groups and ethicists about assuring that individuals undergoing testing are sufficiently educated about the implications of learning their test result and that the testing offers sufficient privacy protection (ACOG Committee Opinion No. 409, 2008; Geransar and Einsiedel, 2008). Changes in physician behavior may result from increased patient requests as a result of the advertising (Myers *et al*., 2006).

There also remain difficult issues about whether patients or even many medical professionals are able to interpret test results correctly or to give accurate information to patients (Giardiello *et al*., 1997; Wideroff *et al*., 2005; Bethea *et al*., 2008). Even though genetic counseling is recommended by most genetics professionals, there is no requirement that patients be counseled prior to receiving results. Primary care physicians appear more reluctant to refer patients for genetic assessment than specialists (Brandt, Ali, Sabel, McHugh, & Gilman, 2008) Brandt *et al*., 2008, with minority-serving physicians even less likely to make such referrals (Shields *et al*., 2008). There are programs in progress to improve the level of genetic knowledge and comfort in making referrals for primary care physicians (Burke and Emery, 2002) and assessment tools being developed to measure the success of the programs (Calefato *et al*., 2008). Some studies suggest frequent misinterpretation of results by patients, including counseled patients, particularly when test results are more ambiguous (Calzone, 2002; Hallowell *et al*., 2002). An example of ambiguous findings is indeterminate results, which can occur when an individual is tested without prior identification of a familial mutation in an affected family member. This might be the case, for example, where an Ashkenazi Jewish individual is tested for the three mutations of *BRCA1/2* genes, which are found with high frequency in that ethnic group. Some patients misinterpret the absence of these mutations to mean that they are not at increased risk; however, this cannot be assumed since there might be another mutation or gene that is causing hereditary cancers in that family. In responsible hands, patients at risk for cancer genes who receive indeterminate results are cautioned to continue to undergo screening or surveillance, which is appropriate for a mutation carrier, since they cannot be excluded as having a mutation.

Some researchers, however, have found little misinterpretation of indeterminate results, at least not in the well-educated cohorts studied to date (Dorval *et al*., 2005). Most of the patients coming forward for testing either in research or clinical settings have been well-educated Caucasian patients from higher socio-economic strata. It is clear that specialized approaches will be necessary to provide genetic

counseling to patients with less background knowledge of genetics and possibly greater skepticism about genetic privacy and the motivation of research scientists. It does seem, though, that patients with more genetics knowledge show some reservations about being tested after learning about the limitations of genetic testing, whereas less educated patients seem to overlook these limitations in embracing testing (Hughes *et al*., 1997). Development of testing protocols which take into consideration of the range of cultural backgrounds, health beliefs, and health literacy of minority patients is an ongoing challenge (Meiser *et al*., 2006, 2007; Roter *et al*., 2007).

Genetic testing does not supply all the answers that patients wish to have. There are very real limitations to what genetic testing can tell patients about the likelihood they will develop the disease in question. A positive test result for a gene that is less than 100% penetrant does not indicate for certain that the mutation carrier will ever develop the disease, or when it will occur. In some cases, a single genetic mutation predisposes the individual to a number of related diseases; testing usually will not indicate which disease the person will develop or whether they will develop more than one disease within the syndrome of diseases to which they are predisposed. Cancer genetic testing is, after all, a trade of the uncertainty about whether one is a mutation carrier, for the uncertainty, in the event that the test is positive, about which cancers might appear and when they will emerge.

38.2.3 Access to Testing

Access to genetic testing is not equally available to all. Full-sequence testing for cancer genes costs several thousand dollars and patients may opt not to bill the costs to their health insurance, either because it is not a covered service in all plans or because patients are reluctant to share test results with their insurer for fear of later discrimination. In the USA, the passage of the Genetic Information Non-Discrimination Act (GINA) prevents insurers from rejecting healthy individuals or raising health insurance premiums on the basis of personal or familial genetic predispositions. It does not, however, cover life, disability, or long-term care insurance (Hudson *et al*., 2008). In countries with a national health insurance system, discrimination is not an issue for health insurance. However, access may be limited by rationing of the number of genetic tests offered in a geographic region during a particular time period and/or very strict criteria for test eligibility. These practices may result in long waits for testing or may exclude patients with less extensive family histories.

Testing was free in most of the early research programs evaluating the uptake and reaction to genetic testing. Without free testing, it is not clear how much the cost of testing affects access or uptake. Minority populations are likely to be particularly hindered in access to genetic testing for a variety of reasons, including cost, lack of minority genetics professionals, and lack of non-English language educational materials.

38.3 INDIVIDUAL FACTORS INFLUENCING UTILIZATION OF GENETIC TESTING

Complex psychological factors govern an individual's uptake of genetic testing and the use they make of the information received. Because early research has focused on heavily affected families, much of what is known in this area pertains to members of what are often called research families or cancer families. It is not difficult to understand that personal experience with a family member who has had the disease in question and possibly died from it is likely to impact one's perception of personal risk, disease-related anxiety, and the desire to avoid a similar outcome. There are, however, other factors, including general personality characteristics, that can affect how actively a person seeks to confront knowledge of personal, hereditary disease risk and to undertake actions designed to prevent or detect illness. Increasing knowledge of these factors will enable researchers and counselors to develop more effective genetic counseling and follow-up techniques and to better advise both patients and medical providers.

38.3.1 Risk Perception

Members of high-risk cancer families typically greatly overestimate their risk for developing cancer (Kash *et al*., 2000). Those who have lost a parent to the disease are especially likely to overestimate the risk that they, themselves, will develop the cancer their parent died from. In an English study of women referred for genetic counseling to a Cancer Family History Clinic, 70% could not cite the population risk for breast cancer, but 76% perceived themselves to be at higher than population levels of risk for breast cancer (Hallowell *et al*., 1998). A study by Zakowski and coworkers (1997) found that women undergoing mammography screening (but with no abnormal findings) whose mother had died of breast cancer estimated their lifetime risk of developing breast cancer as 70% versus the 53% risk estimate of women undergoing mammography who did not have a mother die of breast cancer and the 32% risk estimate of similarly aged women in the general population not undergoing mammography studies. All these figures significantly overestimate breast cancer risk; lifetime risk for breast cancer for a woman in the US general population is about 10%. Having a first-degree relative with breast cancer raises that risk somewhat, to 1.7 to 5 times (Offit, 1998), but not to 70%.

High perceived risk of cancer is a source of motivation for seeking genetic counseling or testing, although high risk may also be accompanied by anxiety which deters

some individuals from accepting counseling or testing (McInerney-Leo *et al*., 2006). Genetic testing provides objective information about cancer risk and is in most studies an important factor affecting risk perception. Given that many women come to *BRCA1/2* genetic testing with high perceptions of their own breast cancer risk, receipt of a positive result indicating one is a carrier may not significantly raise risk perception (McInerney-Leo *et al*., 2006). Young women who are carriers are particularly likely to have their risk perception increase following genetic testing which showed them to be carriers, but this perception may decrease with time (Watson *et al*., 2005). For *BRCA1/2* mutation carriers, perception of their risk for ovarian cancer is likely to increase following receipt of a positive test result (McInerney-Leo *et al*., 2006). Those found to be negative for a known familial mutation are likely to have decreased risk perception (Schwartz *et al*., 2002; Watson *et al*., 2005; Claes *et al*., 2005; McInerney-Leo *et al*., 2006).

Similarly, perceptions of risk among carriers of genes conveying increased predisposition to colon cancer increase following disclosure of positive test results, but there are even larger, significant increases among carriers following post-testing in perception of their risk for extracolonic cancers (Hadley *et al*., 2008). Genetic counseling and testing are particularly important, then, in notifying members of high-risk cancer families about risk for cancers other than those which have primarily affected family members previously. Experience of deaths of close relatives due to an HNPCC-related cancer increases risk perception among HNPCC family members, even following cancer genetic testing (Domanska *et al*., 2007).

38.3.2 Distress

The perception that one has a high likelihood of developing a serious disease is often associated with elevated levels of emotional distress. Among both research and clinical samples of women at high risk for breast cancer, distress scores at or near clinical cut-offs have been found in a substantial proportion of subjects (Kash *et al*., 1992; Audrain *et al*., 1997). Distress is associated with higher breast cancer risk perception and lower perceptions of control over developing breast cancer. In an English study of 312 individuals reporting for counseling for hereditary breast/ovarian cancer, 22% of the women and 14% of the men scored at or above clinical cut-offs for a psychiatric disorder (Foster *et al*., 2002). In a Dutch study, individuals at risk of HD scored higher on distress measures than those at risk of hereditary cancer syndromes (Dudok deWit *et al*., 1997). Those whose scores were highest indicated that their distress was closely related to painful memories of family members who had had the familial disease, as has also been shown in other studies (Lodder *et al*., 1999; Butow *et al*., 2005). A recent study showed that irrespective of personal cancer history, French Canadian women initiating *BRCA1/2* genetic testing had higher distress scores (except depression) than women matched for age in the general population (Dorval *et al*., 2008). Cognitive emotional factors have also been found to be related to distress around hereditary disease risk. Miller and coworkers (1995) have distinguished two styles of coping with threatening information. *Monitors* are those who tend to focus on the threat, to actively seek information related to the threat, and to have difficulty perceiving that all has been done which could be done to cope with the threat. *Blunters*, in contrast, are individuals who naturally try to avoid or shield themselves from threatening information. It has been shown that anxiety during the wait for genetic testing results is higher among women who are high monitors (Tercyak *et al*., 2002), suggesting that these women may need extra support while awaiting their result. Similarly, an Israeli study of individuals undergoing colon cancer genetic testing found monitors were more distressed than non-monitors, especially when they received indeterminate or positive results (Shiloh *et al*., 2008). Cognitive style also appears to influence participation in screening, so that targeting of screening and surveillance recommendations according to the patient's cognitive style may be important to maximize participation (Miller *et al*., 1996).

38.3.3 Health Beliefs

Misconceptions can affect one's ability to realistically evaluate the increase in risk imposed by being a mutation carrier and can also affect one's thinking about the value of screening, surveillance, or prophylactic surgery options. There are many misconceptions about inherited cancer risk even among those at increased risk. In a study of 200 well-educated women with family histories of breast cancer who had had breast cancer themselves, 86% erroneously believed that 1 in 10 women has a *BRCA1/2* mutation, 62% thought that half of all breast cancer is due to *BRCA1/2* mutations (correct estimate: 5–10%), and that 56% did not think fathers could transmit *BRCA1/2* mutations to their daughters as easily as mothers could (Iglehart *et al.*, 1998).

Different ethnic or cultural groups tend to have differing values and may hold different beliefs related to medicine generally, and genetics more specifically. African-Americans have a negative history of experimentation for sickle cell disease, another genetic condition, which may reduce motivation to participate in genetic studies (see also previous chapter). In fact, African-Americans have participated in genetic studies at a lower rate than their participation in medical research generally (Royal *et al*., 2000). The fact that there are very few African-Americans genetics professionals may contribute further to the hesitancy of African-Americans to undergo genetic testing (Halbert *et al*., 2006). On the other hand, a study of 407 African-American and Caucasian women who had at least one first-degree relative with breast or ovarian cancer suggested that

the African-American women had less knowledge about cancer genetics and genetic testing for *BRCA1/2*, but were more positive about the potential benefits of testing. A very small percentage of women tested to date, however, have been from minority populations. Research is under way on how to approach African-American and other minority populations about genetic testing (Hughes *et al*., 1997; Meiser *et al*., 2001; Thompson *et al*., 2002; Brewster *et al*., 2007; Huo and Olopade, 2007). Much remains to be learned about the factors responsible for the lack of involvement of minorities in genetic medicine to date.

A recent study showed that immigrants were only 1/3 as likely to report a family history of cancer in a health survey of 5,586 households in the USA, controlling for sociodemographic and cancer knowledge variables as non-immigrants. Since family history of cancer is an important factor affecting referral for genetic counseling and testing, this finding, which may relate to less knowledge of family history and/or to customs in the native country enforcing silence and secrecy about cancer diagnoses, may play an important role in the lower uptake of genetic services by individuals from ethnic minorities (Oram *et al*., 2007). It is not acculturation *per se*, but the strength of family beliefs and openness of communication about the cancer in the family which influenced attitudes (Kenen *et al*., 2004; Barlow-Stewart *et al*., 2006; Werner-Lin *et al.,* 2007). Cultural beliefs may also affect decision making about uptake of risk-reduction options. Julian-Reynier and coworkers (2001) contrasted views on prophylactic surgery for *BRCA1/2* carriers in England, France, and French Canada. The French were the least accepting of prophylactic mastectomy of the three groups and the French and French-Canadians were less accepting of prophylactic oophorectomy (prophylactic removal of the ovaries) than the English.

38.3.4 Health Behaviors

It is often assumed that anyone known or suspected of being at high risk for disease would wish to undergo any recommended screening mechanism to detect disease early or to try to prevent the occurrence of the disease. This turns out not always to be the case. Anxiety is a powerful inhibitor and, as has been previously discussed, many individuals at high risk experience high levels of anxiety about developing cancer. As a result, compliance with recommended screening or surveillance practices is far from perfect in many groups of at-risk individuals. In a Canadian study of high-risk individuals, those who perceived themselves to be at high risk were five times less likely to adhere to screening recommendations (Ritvo *et al*., 2002).

Although most women at high risk for breast or ovarian cancer report that they get regular mammograms (Watson *et al*., 2005), many do not do breast self-examination, as recommended (Kash *et al*., 1992). An Australian study suggested that after women were informed via genetic testing that they are *BRCA1/2* mutation carriers, they were less likely to adhere to mammogram recommendations (Meiser *et al*., 2002). Contrary results have been found in several studies which indicate increased adherence to mammography recommendations following genetic testing (Claes *et al*., 2005; van Dijk *et al*., 2005).

Colon cancer screening compliance among the general population is lower than screening for breast cancer, and high-risk populations are also below ideal levels for colon cancer screening. For individuals from HNPCC families, colonoscopy is recommended every three years beginning around age 25 (Vasen *et al*., 1998). Among a group of adult patients studied after genetic testing for colon cancer genes, 45% to 48% had never had a colonoscopy and 28% had never had any colon cancer screening test (Johnson *et al*., 2002). Genetic testing appeared to be a catalyst for those found to be positive to undergo regular colon screening, but appeared to deter those who were negative from getting regular colonoscopic screening, as recommended in population guidelines for those at normal risk. In another study of 42 individuals with a personal or family history fulfilling clinical criteria for hereditary colorectal cancer (HNPCC or FAP) who presented for genetic counseling, only 64% had undergone appropriate surveillance for colorectal cancer (Stoeffel *et al*., 2003). Among the women in that study at risk of endometrial manifestations of hereditary colorectal cancer syndromes, only 25% had had appropriate screening. The authors felt for a variety of reasons that these were likely overestimates of compliance with screening and surveillance recommendations for those at risk of hereditary colorectal cancers.

It will be critical to address the cancer worry and general anxiety of those who are found to be mutation carriers. As evidence grows about the efficacy of various screening, chemopreventive, and surgical options for those at increased risk, it will be important to help high-risk individuals cope with the anxiety generated by their awareness of their genetic status in order to understand and, if desired, to take advantage of the available options. It is relatively recently that data has emerged showing that one of the more controversial preventive surgery options, prophylactic mastectomy, conveys 90% risk reduction for breast cancer (Hartmann *et al*., 1999). Prophylactic oophorectomy has also been found not only to reduce the risk of ovarian or endometrial cancer, but has also been shown to have a risk-reducing effect on breast cancer (Rebbeck, 2001).

Increased data will help patients decide about which options they are willing to undertake, but it is clear that psychological factors will continue to be important in all such decisions. As a result, it will be critical that providers recognize the need for patients to discuss the risks and benefits of preventive options, but that they also have access to psychological counseling as needed to help evaluate the personal acceptability of various courses of surgical risk reduction.

38.4 GETTING THE GENETIC TEST RESULT: PERSONAL IMPACT AND PROFESSIONAL COMMUNICATION

The disclosure of a genetic test result is a moment of high drama. It is a moment of importance not only for the individual patient and their (often) accompanying spouse, but it has implications for children, born and unborn, and grandchildren. There may also be implications, though less directly, for siblings, cousins, and other relatives. It may be the end of many years of wondering if one has a hereditary predisposition or it may come as the response to a relatively recent education about hereditary illness, in some cases occasioned by a recent cancer diagnosis. Some results are more definitive than others. Learning one is truly freed from the family curse by virtue of a true negative result can be a moment of blissful relief. Learning, alternatively, that one is a mutation carrier, can answer for some people who have had cancer the question of why they became ill, but it opens many other questions about the subsequent risks for the patient him/herself and questions about risk to offspring. For unaffected patients, learning that they are mutation carriers can be a fearsome moment, conjuring images of ill or deceased family members. These individuals often express disappointment that they could not be released from their cancer worry as they had imagined they would be if they had tested negative. The intensity of their immediate, emotional response often makes it difficult for them to pay attention to the recommendations for mutation carriers offered by the genetic counselor. For others, the result they receive is indeterminate. There can be both disappointment that the answer they received is unclear and relief that at least, for the time being, a mutation was not found. The latter response is somewhat irrational, since an indeterminate finding does not assume that a mutation is not present, only that one could not be found with present testing methods. In each case, emotions are aroused and often patients recount that little is heard beyond the result itself.

38.4.1 Testing Positive

The model for much of how genetic testing has been conducted to date has been the HD research programs, which included a large number of interviews prior to taking blood for the genetic test and great care around the delivery of the result. There had been considerable worry when genetic testing for HD was initiated about whether patients learning that they were HD mutation carriers would experience extreme, adverse emotional reactions, including suicide. Concern was so high because of the unusually high suicide rate (4–8 times population rates) among HD patients (Schoenfeld *et al*., 1984; DiMaio *et al*., 1993). Outcome studies from the HD programs have shown that among more than 4,000 patients tested worldwide for HD, less than 1% either have attempted or committed suicide following disclosure of their test result (Almqvist *et al*., 1999); those who committed suicide already had symptoms of HD. These figures are relatively reassuring, but support efforts to define which patients are most likely to respond adversely and to develop psychological counseling models of use to such patients (Robins-Wahlin *et al*., 2000). Factors predisposing to adverse outcomes following HD genetic test disclosure were having a psychiatric history of five years duration or less and being unemployed (Almqvist *et al*., 1999).

The follow-up of patients undergoing cancer genetic testing has shown, too, that extreme adverse outcomes are rare (Butow *et al*., 2003). Most patients who are found to be carriers of deleterious mutations respond with sadness, but do not develop clinical symptoms of depression or anxiety. Mean scores of most samples of tested individuals have been within the normal range (Lerman *et al*., 1996; Croyle *et al*., 1997). Little difference has been noted in psychological outcomes between research samples and self-referred clinic samples (Schwartz *et al*., 2002).

On the other hand, subclinical effects of increased intrusive thinking about cancer and cancer worry have been found in many individuals who have tested positive for a cancer predisposition gene (Meiser *et al*., 2002). Also, about a quarter to a third of many of the samples of individuals undergoing cancer genetic testing have shown elevated, clinical levels of depression or anxiety following disclosure (Grosfeld *et al*., 1996; Lodder *et al*., 2001; Bonadona *et al*., 2002). Some predictors of higher distress have been the test status of other family members. Lodder and coworkers (2001) found that non-carriers who had sisters who had been found to be mutation carriers were more depressed than other non-carriers. Similarly, Smith and coworkers (1999) found that men who were the first to be genetically tested in their families and had other siblings awaiting results were more distressed than carriers whose siblings were all test-negative. Female carriers who were the only positive members of their sibship (i.e. their siblings were all negative) scored as high as recently diagnosed cancer patients on the Impact of Events Scale (Horowitz *et al*., 1979), a measure assessing cancer-related worry. Women carriers whose sisters had not yet received results also were quite distressed.

Disease status also affects psychological outcomes. The ability to correctly anticipate the level of one's emotional reaction to a forthcoming event is thought of as a predictor of good adjustment to that event. In a study of women tested for *BRCA1/2*, some of whom had had cancer and some who were unaffected (i.e., had not had cancer), the cancer patients found it significantly more difficult to anticipate their emotional reaction following disclosure. Cancer patients tended to feel that learning that they were mutation carriers would not be as eventful for them as it turned out to be. Underestimation of the emotional impact of disclosure was associated with significantly increased distress levels six months later (Dorval *et al*., 2000).

38.4.2 Testing Negative

Of interest is that in both the HD and cancer genetic testing outcome studies, a subset of individuals testing negative have been found to experience moderate levels of distress (Myers *et al*., 1997). These paradoxical reactions were largely unanticipated by researchers who believed that one of the major benefits of genetic testing was that, for dominant genes, it could free 50% of the patients from fear of hereditary disease predisposition. Relief is prominent in the reactions of most individuals testing negative, although many (but not all) say they also feel guilty toward family members who have had the disease or who are mutation carriers. More extreme negative reactions have occasionally been seen in individuals who are significantly distressed after learning they are not mutation carriers. In some cases, particularly within HD samples, this distress has been associated with regret for having lived so many years with the expectation of getting HD and of having foregone opportunities for further education or other experiences because of that, now proved faulty, expectation (Huggins *et al*., 1992).

There are also very rare cases reported in the literature of Munchausen syndrome whereby individuals falsely claim to have family histories suggestive of hereditary syndromes for whatever secondary emotional gains they believe might accrue to at-risk individuals (Kerr *et al*., 1998).

Another unexpected reaction of individuals testing negative for mutations in cancer genes or being told that their cancer risk is not sufficient to recommend genetic testing has been an unwillingness to accept the recommendations. Some of the people insist on getting tested or find it difficult to give up the increased screening they had received before testing when it was thought they could be mutation carriers. Individuals testing negative for FAP have, in some cases, stated their reluctance to discontinue the early and more frequent colonoscopic screening recommended for those at high risk, stating insufficient trust in the genetic test result (Rhodes *et al*., 1991; Bleiker *et al*., 2003). In another study, many of the women who were told on the basis of their family history that they were not eligible for *BRCA1/2* genetic testing because their risk levels were too low reacted with anger rather than relief (Bottorff *et al*., 2000).

38.4.3 Consequences of Testing

The majority of genetically tested individuals express no regret at being tested, even if they are found to be carriers. Carriers express sadness and disappointment, but also express the belief that knowing they are at increased hereditary risk will help motivate them to take good care of themselves physically and to be regularly screened by medical professionals. Concern for their offspring is a major factor in the guilt and sadness that they feel. Those who are negative feel particular relief on behalf of their children and grandchildren who are spared continuing worry about hereditary predisposition.

An important question, of course, is whether knowledge of one's genetic test status correlates with appropriate levels of screening behavior or with uptake of preventive options. Because this is a relatively new area of medicine, decisions about screening, surveillance, and preventive options are complicated by imperfect data about their efficacy in preventing cancer in mutation carriers. There are no preventive options for individuals at risk for HD. Nonetheless, it is of interest to learn how testing changes health behaviors. A Canadian 18-month follow-up study of 79 women who had tested positive for a *BRCA1/2* mutation found that 57% said that they had altered their screening practices following disclosure (Ritvo *et al*., 2002). Especially younger (age <50) women were likely to report change in health behaviors. Prophylactic mastectomy had been elected by 21% of the women and prophylactic oophorectomy by 40%. Lerman and coworkers (2000) showed how mutation carriers' initial intentions regarding prophylactic surgery did not translate, at least over the year following disclosure, into actions. At disclosure, 31% intended to consider prophylactic mastectomy and 40% intended to undergo prophylactic oophorectomy. Six months later, only 1% had undergone prophylactic mastectomy and 2% had had prophylactic oophorectomy. At one year, 3% of the women had had prophylactic mastectomy and 13% had undergone prophylactic oophorectomy. In that same group, only 21% had had a CA-125 test, which is recommended at least annually to *BRCA1/2* mutation carriers to try to detect ovarian cancer in its earliest stages. Two-thirds of the women had had a mammogram in the past year, a percentage unchanged from before testing. More recent reports show somewhat higher rates for uptake of prophylactic surgeries, especially oophorectomy in women near or over age 40. Rates reported vary significantly from country to country and vary, too, with the specifics of the population studied. Rates of oophorectomy following genetic testing were 14% in a New Zealand study (Phillips *et al*., 2006), 44% in a Dutch study (Madalinska *et al*., 2005) to 75% in a Belgian study (Claes *et al*., 2005). Prophylactic mastectomy remains a low uptake option with rates ranging from 9% (Claes *et al*., 2005) to 11% in a US study (Ray *et al*., 2005) and 27% in a Canadian cohort (Metcalfe *et al*., 2005), to a high of 54% in a 2002 Dutch study (Lodder *et al*., 2002). Genetic testing of women newly diagnosed with unilateral breast cancer is increasing uptake of prophylactic contralateral mastectomy (Weitzel *et al*., 2003; Schwartz *et al*., 2004; Graves *et al*., 2007).

Although many good intentions to optimize and personalize health screening behaviors in the face of disclosure of a positive genetic test were expressed just after disclosure, many of these intentions were often not acted upon in the year following disclosure. This raises questions about the timeframe in which patients are expected to make

such health behavior changes, but also about the need for follow-up counseling to encourage active screening and surveillance and to help individuals with consideration of more extreme preventive options. There is much that needs to be learned about how patients perceive recommendations for screening and surveillance, especially once it has been established that they are mutation carriers. In addition, researchers and genetic professionals also need to understand more about pretesting levels of distress (depression, anxiety, and disease-related worry) and their relation to utilization of health behaviors following disclosure.

38.5 FAMILY COMMUNICATION

Genetic medicine is family medicine. Genetic professionals advise patients to consider carefully the ramifications of genetic information on other family members. Both before and after genetic testing, there are reasons why open communication with family members about family history and hereditary risk may be advisable. In order to provide the medical information genetics professionals require to calculate the risk that an individual carries a mutation, it is often necessary to talk to relatives about which family members were affected by disease at what ages and with what outcomes. Women are typically the keepers of such family information. Because of the many premature deaths in families with hereditary illness, there are often gaps in knowledge about family history. An individual seeking genetic counseling may have to talk to relatives with whom they ordinarily have little contact in order to complete the family history to a point where the counselor will have enough information to calculate risk.

Gathering family history information in order to create a family illness pedigree may require talking to relatives on both sides of the family, which sometimes includes uncomfortable contacts with long-lost relatives. Family traditions may include not talking very directly about illness and doing so may open old emotional wounds. Some family members are completely estranged, in some cases, making completion of a full pedigree impossible. Because the hallmark of hereditary illness is the involvement of multiple generations of family members, family issues tend to surface more readily with regard to genetic testing than in other kinds of medical endeavors.

The relevance of one person's genetic information to other family members can set up difficult alliances or the tendency to avoid relatives with whom one is not comfortable discussing highly personal information, such as a genetic test result.

The accuracy of family information can affect genetic professionals' ability to correctly analyze the likelihood that there is a hereditary predisposition to disease in a particular family. Documentation of age of onset of disease is particularly important in such determinations, since, in many cases, early age of onset is a hallmark of hereditary illness. Also, knowledge of exactly which cancers a family member had may be critical as cancer syndromes link certain cancers to hereditary mutations. There is more confusion about some cancers than others. Knowledge about breast cancer within a family, for example, tends to be more accurate than colon cancer, which may be confused with other gastrointestinal cancers. In a study by Glanz and coworkers (1999), which identified colon cancer patients from a medical registry, 25% of the first-degree relatives of these patients denied that they had a relative who had had colon cancer. Schneider and workers (2004) found that over half of the members of families presented their family history with insufficient accuracy to have led providers to suspect that they were an LFS family. These reports suggest the need to verify diagnoses (other than breast cancer) from medical records of the family members, if possible, as part of the genetic counseling process.

Although most inaccuracies are due to misconceptions or unintentional error, there are occasions when family embarrassment about illness or other factors leads to deliberate falsification of the family history. One of the risks of gathering family history data is the risk of uncovering misassigned paternity; that is, that the putative father of a child is not, in fact, the actual father. Patients should be helped to understand that the experience of searching for family history data may be more emotional and/or upsetting than anticipated.

Once genetic test results are available and disclosed to the patient, considerations about the implications of the result for other family members come into play. Genetics professionals recommend quite strongly that results be shared with family members, especially when screening recommendations likely would be influenced by the genetic information. For example, if a man in his 50s is found to be a mutation carrier for HNPCC, it would be important for his children to know that they should initiate colonoscopic screenings at age 25, not at age 50, as recommended for the general population. Similarly, if the man is negative, his children could use that information to decide that they did not require colonoscopy until they are 50 years of age.

In most cases, information about genetic testing is readily shared, but this is not universal. There are rifts in many families that prevent parents and children from communicating about difficult subjects such as health risks. Some do not talk at all. Divorce, disgruntlement, and embarrassment can all inhibit or prevent open discussion of such mutually relevant medical information. Other family crises including disease-related deaths may impede immediate disclosure and discussion of genetic test results. It can be difficult to decide how and when to inform relatives. Since the testing of children for adult-onset disorders generally is discouraged, except when results would influence uptake of preventive options, which would begin in childhood, parents often postpone telling children about the parent's

test result. Early research in this area, however, has shown that it is the family's style of communicating generally that typically governs whether and when children are told (Tercyak *et al*., 2002).

Some family members do not wish to know about their own or others' genetic test status. This can raise problems when that family member is a twin or an obligate carrier of a mutation identified in another relative. Some family members are fearful that telling a relative could be quite upsetting and bring back unpleasant memories or arouse feelings of guilt for having passed on a deleterious mutation.

Providers need help in advising patients about the communication of results to family members. As the genetic component of many diseases is recognized, the complexity of such communication will increase. Both patients and providers will find it valuable to utilize CD-ROMs and other aids to communicate complicated genetic information to relatives, but this will not replace the need for sensitive patient-provider communication and awareness of the complex psychological responses, which the dissemination of genetic information within a family can evoke.

38.6 FUTURE CHALLENGES: MORE GENES, LESS CLARITY

The translation of twenty-first century genetic advances into clinical practice will necessitate ever-closer collaboration between molecular geneticists, clinicians, and social scientists (Khoury *et al*., 2007). Most of the genes for which genetic testing is currently available are inherited in Mendelian fashion, that is, in relatively simple ways. It is very likely that genes discovered in the future that convey susceptibility to a host of common conditions, such as heart disease, diabetes, and psychiatric disorders, will have their effects by virtue of complex combinations with other genes or in interaction with environment factors (Feero *et al*., 2008).

For patients to understand such complex patterns of inheritance will require carefully constructed teaching tools. Targeting of the information offered to the cognitive style of the individual may help as might the availability of adjunct materials, like CD-ROMs. But there will always be a need for a sympathetic and respectful clinician who can help the patient make links between their family history, their genetic risk, and the possibilities of reducing adverse outcomes of hereditary predisposition.

Genetic medicine holds great promise for identifying individuals at increased disease risk and for treating patients with greater efficacy and fewer side effects through pharmacogenetics (see also Chapter 22). Patient education about genetics will necessitate having well-informed, well-trained, culturally competent professionals to address the different genetic counseling, medical, and psychological questions that arise (Meiser *et al*., 2001). It is a challenge for professionals to educate themselves about this rapidly changing field (Emery and Hayflick, 2001). Professional organizations, such as the National Coalition for Health Professional Education in Genetics (NCHPEG) and Genetic Resources on the Web (GROW), are developing and disseminating resources to foster increased genetic literacy among health professionals in preparation for an expected exponential growth in genetic medicine in the decades to come.

In addition to the facts of hereditary risk and delineation of the risk and benefits of genetic testing, it will be important also to discuss the psychological factors governing patient and family response to genetic information. It is only when patients feel empowered and not threatened by such information that they can make optimal use of the advances in genetic medicine.

REFERENCES

Almqvist, E.W., Bloch, M., Brinkman, R., Craufurd, D., and Hayden, M.R. (1999). A worldwide assessment of the frequency of suicide, suicide attempts, or psychiatric hospitalization after predictive testing for Huntington disease. *Am. J. Hum. Genet.* 64, 1293–1304.

Audrain, J., Schwartz, M.D., Lerman, C., Hughes, C., and Peshkin, B.N. (1997). Psychological distress in women seeking genetic counseling for breast-ovarian cancer risk: the contributions of personality and appraisal. *Ann. Behav. Med.* 19, 370–377.

Barlow-Stewart, K., Yeo, S.S., Meiser, B., Goldstein, D., Tucker, K., and Eisenbruch, M. (2006). Toward cultural competence in cancer genetic counseling and genetics education: lessons learned from Chinese-Australians. *Genet. Med.* 8, 24–32.

Bernhardt, C., Schwan, A.M., Kraus, P., Epplen, J.T. and Kuntsmann, E. (2009). Decreasing uptake of predictive testing for Huntington's disease in a German center: 12 years' experience (1993–2004). *Eur. J. Hum. Genet*. 17, 295–300.

Bethea, J., Qureshi, N., Drury, N., and Guilbert, P. (2008). The impact of genetic outreach education and support to primary care on practitioners' confidence and competence in dealing with familial cancers. *Community Genet.* 11, 289–294.

Bleiker, E.M.A., Menko, F.H., Taal, B.G., Kluijt, I., Wever, L.D.V., Gerritsma, M.A., Vasen, H.F.A., and Aaronson, N.K. (2003). Experience of discharge from colonoscopy of mutation negative HNPCC family members. *J. Med. Genet.* 40, e55.

Bonadona, V., Saltel, P., Desseigne, F., Mignotte, H., Saurin, J.-C., Wang, Q., Sinilnikova, O., Giraud, S., Freyer, G., Plauchu, H., Puisieux, A., and Lasset, C. (2002). Cancer patients who experienced diagnostic genetic testing for cancer susceptibility: reactions and behavior after the disclosure of a positive test result. *Cancer Epidemiol. Biomar. Prev.* 11, 97–104.

Bonadona, V., Sinilnikova, O.M., Chopin, S., Antoniou, A.C., Mignotte, H., Mathevet, P., Bremond, A., Martin, A., Bobin, J.-Y., Romestaing, P., Raudrant, D., Rudigoz, R.-C., Leone, M., Chauvin, F., Easton, D.F., Lenoir, G.M., and Lasset, C. (2005). Contribution of BRCA1 and BRCA2 germ-line mutations to the incidence of breast cancer in young women: results from a prospective population-based study in France. *Genes Chromosome. Cancer* 43, 404–413.

Bottorff, J.L., Balneaves, L.G., Buxton, J., Ratner, P.A., McCullum, M., Chalmers, K., and Hack, T. (2000). Falling through the cracks: women's experience of ineligibility for risk of breast cancer. *Can. Fam. Physician.* 46, 1449–1456.

Brandt, R., Ali, Z., Sabel, A., McHugh, T., and Gilman, P. (2008). Cancer genetics evaluation: barriers to and improvements for referral. *Genet. Test.* 12, 9–12.

Brewster, K., Wileyto, E.P., Kessler, L., Collier, A., Weathers, B., Stopfer, J.E., Domchek, S., and Halbert, C.H. (2007). Sociocultural predictors of breast cancer risk perceptions in African American breast cancer survivors. *Cancer Epidemiol. Biomar. Prev.* 16, 244–248.

Burke, W., and Emery, J. (2002). Genetics education for primary-care providers. *Nat. Rev. Genet.* 3, 561–566.

Butow, P.N., Lobb, E.A., Meiser, B., Barratt, A., and Tucker, K.M. (2003). Psychological outcomes and risk perception after genetic testing and counselling in breast cancer: a systematic review. *Med. J. Aust.* 178, 77–81.

Butow, P., Meiser, B., Price, M., Bennett, B., Tucker, K., Davenport, T., and Hickie, I.The Kathleen Cuningham Consortium for Research into Familial Breast Cancer Psychosocial Group (2005). Psychological morbidity in women at increased risk of developing breast cancer: a controlled study. *Psychooncology* 14, 196–203.

Calefato, J.M., Nippert, I., Harris, H.J., Kristoffersson, U., Schmidtke, J., Ten Kate, L.P., Anionwu, E., Benjamin, C., Challen, K., Plass, A. M., Harris, R., and Julian-Reynier, C. (2008). Assessing educational priorities in genetics for general practitioners and specialists in five countries: factor structure of the Genetic-Educational Priorities (Gen-EP) scale. *Genet. Med.* 10, 99–106.

Calzone, K.A., and Biesecker, B.B. (2002). Genetic testing for cancer predisposition. *Cancer Nurs.* 25, 15–25.

Claes, E., Evers-Kiebooms, G., Decruyenaere, M., Denayer, L., Boogaerts, A., Philippe, K., and Legius, E. (2005). Surveillance behaviour and prophylactic surgery after predictive testing for hereditary breast/ovarian cancer. *Behav. Med.* 31, 93–105.

Collins, F.S., Green, E.D., Guttmacher, A.E., and Guyer, M.S. (2003). A vision for the future of genomics research. *Nature* 422, 835–847.

Committee on Genetics, American College of Obstetrics and Gynecology: Committee on Ethics (2008). American College of Obstetrics and Gynecology Opinion No. 409: Direct-to-consumer marketing of genetic testing. *Obstet. Gynecol.* 111, 1493–1494.

Creighton, S., Almqvist, E.W., MacGregor, D., Fernandez, B., Hogg, H., Beis, J., Welch, J.P., Ridell, C., Lokkesmoe, R., Khalifa, M., MacKenzie, J., Sajoo, A., Farrell, S., Robert, F., Shugar, A., Summers, A., Meschino, W., Allingham-Hawkins, D., Chiu, T., Hunter, A., Allanson, J., Hare, H., Schweeen, J., Collins, L., Sanders, S., Greenberg, C., Cardwell, S., Lemire, E., MacLeod, P., and Hayden, M.R. (2003). Predictive, prenatal and diagnostic genetic testing for Huntington's disease: the experience in Canada from 1987 to 2000. *Clin. Genet.* 63, 462–475.

Croyle, R.T., Smith, K.R., Botkin, J.R., Baty, B., and Nash, J. (1997). Psychological responses to BRCA1 mutation testing: preliminary findings. *Health Psychol.* 16, 63–72.

DeMarco, T.A., Loffredo, C.A., Sampilo, M.L., and Tercyak, K.P. (2006). On using a cancer center registry to identify newly affected women eligible for hereditary breast cancer syndrome testing: practical considerations. *J. Genet. Couns.* 15, 129–136.

DiMaio, L., Squitieri, F., Napolitano, G., Campanella, G., Trofatter, J.A., and Conneally, P.M. (1993). Suicide risk in Huntington's disease. *J. Med. Genet.* 30, 293–295.

Domanska, K., Nilbert, M., Soller, M., Silfverberg, B., and Carlsson, C. (2007). Discrepancies between estimated and perceived risk of cancer among individuals with hereditary non-polyposis colorectal cancer. *Genet. Test.* 11, 183–186.

Dorval, M., Patenaude, A.F., Schneider, K.A., Kieffer, S.A., DiGianni, L., Kalkbrenner, K., Bromberg, J.I., Basili, L.A., Calzone, K., Stopfer, J., Weber, B.L., and Garber, J.E. (2000). Anticipated versus actual emotional reactions to disclosure of results of genetic tests for cancer susceptibility: findings from *p53* and *BRCA1* testing programs. *J. Clin. Oncol.* 18, 2135–2142.

Dorval, M., Gauthier, G., Maunsell, E., Dugas, M.J., Rouleau, I., Chiquette, J., Planate, M., Laframboise, R., Gaudet, M., Bridge, P.J., and Simard, J. (2005). No evidence of false reassurance among women with an inconclusive BRCA1/2 genetic test result. *Cancer Epidemiol. Biomar. Prev.* 14, 2862–2867.

Dorval, M., Bouchard, K., Maunsell, E., Chicquette, J., Camden, S., Dugas, M.J., and Simard, J. the INHERIT BRCA Group (2008). Health behaviors and psychological distress in women initiating BRCA1/2 genetic testing: comparison with control population. *J. Genet. Couns.* 17, 314–326.

Dudok deWit, A.C., Tibben, A., Duivenvoorden, H.J., Frets, P.G., Zoeteweij, M.W., Losekoot, M., van Haeringen, A., Niermeijer, M.F., and Passchier, J. (1997). Psychological distress in applicants for predictive DNA testing of autosomal dominant, heritable, late onset disorders. *J. Med. Genet.* 34, 382–390.

Easton, D.F., Ford, D., and Bishop, D.T. the Breast Cancer Linkage Consortium (1995). Breast and ovarian cancer incidence in *BRCA1*-mutation carriers. *Am. J. Hum. Genet.* 56, 265–271.

Emery, J., and Hayflick, S. (2001). The challenge of integrating genetic medicine into primary care. *BMJ* 322, 1027–1030.

Feero, W.G., Guttmacher, A.E., and Collins, F.S. (2008). The genome gets personal - almost. *JAMA* 299, 1351–1352.

Ford, D., Easton, D.F., Bishop, D.T., Narod, S., and Goldgar, D.E. (1994). Risks of cancer in BRCA1-mutation carriers. Breast cancer linkage consortium. *Lancet* 343, 692–695.

Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M.D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T.R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B.A., Gayther, S.A., and Zelada-Hedman, M. for the Breast Cancer Linkage Consortium (1998). Genetic heterogeneity and penetrance analysis of the *BRCA1* and *BRCA2* genes in breast cancer families. *Am. J. Hum. Genet.* 62, 676–689.

Foster, C., Evans, D.G.R., Eeles, R., Eccles, D., Ashley, S., Brooks, L., Davidson, R., Mackay, J., Morrison, P.J., and Watson, M. (2002). Predictive testing for BRCA1/2: attributes, risk perception, and management in a multi-centre clinical cohort. *Br. J. Cancer* 86, 1209–1216.

Foster, C., Evans, D.G., Eeles, R., Eccles, D., Ashley, S., Brooks, L., Cole, T., Cook, J., Davidson, R., Gregory, H., Mackay, J., Morrison, P.J., and Watson, M. (2004). Non-uptake of predictive genetic testing for BRCA1/2 among relatives of known carriers: attributes, cancer worry, and barriers to testing in a multicenter clinical cohort. *Genet. Test.* 8, 23–29.

Friend, S.H., Bernards, R., Rogelj, S., Weinberg, R.A., Rapaport, J.M., Albert, D.M., and Dryja, T.P. (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* 323, 643–646.

Geransar, R., and Einsiedel, E. (2008). Evaluating online direct-to-consumer marketing of genetic tests: informed choices or buyers beware. *Genet. Test.* 12, 13–23.

Giardiello, F.M., Brensinger, J.D., Petersen, G.M., Luce, M.C., Hylind, L.M., Bacon, J.A., Booker, S.V., Parker, R.D., and Hamilton, S.R. (1997). The use and interpretation of commercial APC gene testing for familial adenomatous polyposis. *N. Engl. J. Med.* 336, 823–827.

Giardiello, F.M., Brensinger, J.D., and Petersen, G. (2001). American Gastroenterology Association technical review on hereditary colorectal cancer and genetic testing. *Gastroenterology* 121, 199–213.

Glanz, K., Grove, J., Le Marchand, L., and Gotay, C. (1999). Underreporting of family history of colon cancer: correlates and implications. *Cancer Epidemiol. Biomar. Prev.* 8, 635–639.

Graves, K.D., Peshkin, B.N., Halbert, C.H., DeMarco, T.A., Isaacs, C., and Schwartz, M.D. (2007). Predictors and outcomes of contralateral prophylactic mastectomy among breast cancer survivors. *Breast Cancer Res. Treat.* 104, 321–329.

Grosfeld, F.J.M, Lips, C.J.M., Ten Kroode, H.F.J., Beemer, F., Van Spijker, H.G., and Brouwers-Smalbraak, G.J. (1996). Psychosocial consequences of DNA analysis for MEN type 2. *Oncology* 10, 141–157.

Gusella, J.F., Wexler, N.S., Conneally, P.M., Naylor, S.L., Anderson, M.A., Tanzi, R.E., Watkins, P.C., Ottina, K., Wallace, M.R., and Sakaguchi, A.Y. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306, 234–238.

Hadley, D.W., Jenkins, J.F., Steinberg, S.M., Liewehr, D., Moller, S., Martin, J.C., Calzone, K.A., Soballe, P.W., and Kirsch, I.R. (2008). Perceptions of cancer risks and predictors of colon and endometrial cancer screening in women undergoing genetic testing for Lynch syndrome. *J. Clin. Oncol.* 26, 948–954.

Halbert, C.H., Kessler, L., Stopfer, J.E., and Wileyto, E.P. (2006). Low rates of acceptance of BRCA1 and BRCA2 test results among African-American women at increased risk for hereditary breast cancer. *Genet. Med.* 8, 576–582.

Hallowell, N., Statham, H., and Murton, F. (1998). Women's understanding of their risk of developing breast/ovarian cancer before and after genetic counseling. *J. Genet. Counsel.* 7, 345–364.

Hallowell, N., Foster, C., Ardern-Jones, A., Eeles, R., Murday, V., and Watson, M. (2002). Genetic testing for women previously diagnosed with breast/ovarian cancer: examining the impact of BRCA1 and BRCA2 mutation searching. *Genet. Test.* 6, 79–87.

Harper, P.S., Lim, C., and Craufurd, D. on behalf of the UK Huntington's Disease Prediction Consortium (2000). Ten years of presymptomatic testing for Huntington's disease: the experience of the UK Huntington's disease prediction consortium. *J. Med. Genet.* 37, 567–571.

Hartmann, L.C., Schaid, D.J., Woods, J.E., Crotty, T.P., Myers, J.L., Arnold, P.G., Petty, P.M., Sellers, T.A., Johnson, J.L., McDonnell, S.K., Frost, M.H., and Jenkins, R.B. (1999). Efficacy of bilateral prophylactic mastectomy in women with a family history of breast cancer. *N. Engl. J. Med.* 340, 77–84.

Holloway, S.M., Bernhard, B., Campbell, H., and Lam, W.W. (2008). Uptake of testing for BRCA1/2 mutations in South East Scotland. *Eur. J. Hum. Genet.* 16, 906–912.

Holtzman, N.A., and Marteau, T.M. (2000). Will genetics revolutionize medicine? *N. Engl. J. Med.* 343, 141–144.

Horowitz, M., Wilner, N., and Alvarez, W. (1979). Impact of events scale: a measure of subjective stress. *Psychosom. Med.* 41, 209–218.

Hudson, K.L., Holohan, M.K., and Collins, F.S. (2008). Keeping pace with the times – the genetic information non-discrimination act of 2008. *New Engl. J. Med.* 358, 2661–2663.

Huggins, M., Bloch, M., Wiggins, S., Adam, S., Suchowersky, O., Trew, M., Klimek, M., Greenberg, C.R., Eleff, M., Thompson, L.P., Knight, J., MacLeod, P., Girard, K., Theilmann, T., Hedrick, A., and Hayden, M.R. (1992). Predictive testing for Huntington disease in Canada: adverse effects and unexpected results in those receiving a decreased risk. *Am. J. Med. Genet.* 42, 508–515.

Hughes, C., Gomez-Caminero, A., Benkendorf, J., Kerner, J., Isaacs, C., Barter, J., and Lerman, C. (1997). Ethnic differences in knowledge and attitudes about BRCA1 testing in women at increased risk. *Patient Educ. Counsel* 32, 51–62.

Huntington's Disease Collaborative Research Group (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971–983.

Huo, D., and Olopade, O.I. (2007). Genetic testing in diverse populations: are researchers doing enough to get out the correct message? *JAMA* 298, 2910–2911.

Iglhehart, J.D., Miron, A., Rimer, B.K., Winer, E.P., Berry, D., and Shildkraut, J.M. (1998). Overestimation of hereditary breast cancer risk. *Ann. Surg.* 228, 375–384.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Jacopini, G.A., D'Amico, R., Frontali, M., and Vivona, G. (1992). Attitudes of persons at risk and their partners toward predictive testing. In *Psychosocial Aspects of Genetic Counseling* (Evers-Kiebooms, G., Fryns, J.P., and Cassiman, J.J., eds), pp. 113–117. Wiley-Liss, New York.

Johnson, K.A., Trimbath, J.D., Petersen, G.M., Griffin, C.A., and Giardiello, F.M. (2002). Impact of genetic counseling and testing on colorectal screening behavior. *Genet. Test.* 6, 303–306.

Joshi, V.A., and Kucherlapati, R. (2008). Genetics and genomics in the practice of medicine. *Gastroenterol.* 134, 1284–1288.

Julian-Reynier, C., Sobol, H., Sevilla, C., Nogues, C., and Bourret, P. the French Cancer Genetic Network (2000). Uptake of hereditary breast/ovarian cancer genetic testing in a French national sample of *BRCA1* families. *Psychooncology* 9, 504–510.

Kash, K.M., Holland, J.C., Halper, M.S., and Miller, D.G. (1992). Psychological distress and surveillance behaviors of women with a family history of breast cancer. *J. Natl. Cancer Inst.* 84, 24–30.

Kash, K.M., Ortega-Verdejo, K., Dabney, M.K., Holland, J.C., Miller, D.G., and Osborne, M.P. (2000). Psychosocial aspects of cancer genetics: women at high risk for breast and ovarian cancer. *Sem. Surg. Oncol.* 18, 333–338.

Kenen, R., Ardern-Jones, A., and Eeles, R. (2004). We are talking, but are they listening? Communication patterns in families with a history of breast/ovarian cancer (HBOC). *Psychooncology* 13, 335–345.

Kerr, B., Foulkes, W.D., Cade, D., Hadfield, L., Hopwood, P., Serruya, C., Hoare, E., Narod, S.A., and Evans, D.G. (1998). False family history of breast cancer in the family cancer clinic. *Eur. J. Surg. Oncol.* 24, 275–279.

Khoury, M.J., Gwinn, M., Yoon, P.W., Dowling, N., Moore, C.A., and Bradley, L. (2007). The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention?. *Genet. Med.* 9, 665–674.

Lerman, C., Seay, J., Balshem, A., and Audrain, J. (1995). Interest in genetic testing among first-degree relatives of breast cancer patients. *Am. J. Med. Genet.* 57, 385–392.

Lerman, C., Narod, S., Schulman, K., Hughes, C., Gomez- Caminero, A., Bonney, G., Gold, K., Trock, B., Main, D., Lynch, J., Fulmore, C., Snyder, C., Lemon, S.J., Conway, T., Tonin, P., Lenoir, G., and Lynch, H. (1996). *BRCA1* testing in families with hereditary breast-ovarian cancer: a prospective study of patient decision making and outcomes. *JAMA* 275, 1885–1892.

Lerman, C., Hughes, C., Lemon, S.J., Main, D., Snyder, C., Durham, C., Narod, S., and Lynch, H.T. (1998). What you don't know can hurt you: adverse psychologic effects in members of BRCA1-linked and BRCA2-linked families who decline genetic testing. *J. Clin. Oncol.* 16, 1650–1654.

Lerman, C., Hughes, C., Croyle, R.T., Main, D., Durham, C., Snyder, C., Bonney, A., Lynch, J.F., Narod, S.A., and Lynch, H.T. (2000). Prophylactic surgery decisions and surveillance practices one year following *BRCA1/2* testing. *Prev. Med.* 31, 75–80.

Lodder, L.N., Frets, P.G., Trijsburg, R.W. et al. (2002). One year follow-up of women opting for presymptomatic testing for BRCA1 and BRCA2: emotional impact of the test outcome and decisions on risk management (surveillance or prophylactic surgery). *Breast Cancer Res. Treat.* 73, 97–112.

Lodder, L., Frets, P.G., Trijsburg, R.W., Meijers-Heijboer, E.J., Klijn, J.G.M., Duivenvoorden, H.J., Tibben, A., Wagner, A., van der Meer, C.A., Devilee, P., Cornelisse, C.J., and Niermeijer, M.F. (1999). Presymptomatic testing for BRCA1 and BRCA2: how distressing are the pretest weeks? *J. Med. Genet.* 36, 906–913.

Lodder, L., Frets, P.G., Trijsburg, R.W., Meijers-Heijboer, E.J., Klijn, J.G.M., Duivenvoorden, H.J., Tibben, A., Wagner, A., van der Meer, C.A., van den Ouweland, A.M., and Niermeijer, M.F. (2001). Psychological impact of receiving a *BRCA1/BRCA2* test result. *Am. J. Med. Genet.* 98, 15–24.

Lodder, L.N., Frets, P.G., Trijsburg, R.W., Meijers-Heijboer, E.J., Klijn, J.G., Seynaeve, C., van Geel, A.N., Tilanus, M.M., Bartels, C.C., Verhoog, L.C., Brekelmans, C.T., Burger, C.W., and Niermeijer, M.F. (2002). One year follow-up of women opting for presymptomatic testing for BRCA1 and BRCA2: emotional impact of the test outcome and decisions on risk management (surveillance or prophylactic surgery). *Breast Cancer Res. Treat.* 73, 97–112.

Lynch, H.T., Lynch, J.F., Lynch, P.M., and Attard, T. (2008). Hereditary colorectal cancer syndromes: molecular genetics, genetic counseling, diagnosis and management. *Fam. Cancer* 7, 27–39.

Madalinska, J.B., Hollenstein, J., Bleiker, E., van Beurden, M., Valdimarsdottir, H.B., Massuger, L.F., Gaarenstroom, K.N., Mourits, M.J., Verheijen, R.H., van Dorst, E.B., van der Putten, H., van der Velden, K., Boonstra, H., and Aaronson, N.K. (2005). Quality-of-life effects of prophylactic salpingo-oophorectomy versus gynecologic screening among women at increased risk of hereditary ovarian cancer. *J. Clin. Oncol.* 23, 6890–6893.

Mastromauro, C., Myers, R.H., and Berkman, B. (1987). Attitudes toward presymptomatic testing in Huntington's disease. *Am. J. Med. Genet.* 26, 271–282.

McInerney-Leo, A., Hadley, D., Kase, K.G., Giambaressi, T.R., Struewing, J.P., and Biesecker, B.B. (2006). BRCA1/2 testing in hereditary breast and ovarian cancer families III: risk perception and screening. *Am. J. Med. Genet. A.* 140, 2198–2206.

Meiser, B., and Dunn, S. (2001). Psychological effect of genetic testing for Huntington's disease: an update of the literature. *West. J. Med.* 174, 336–340.

Meiser, B., Eisbruch, M., Barlow-Stewart, K., Tucker, K., Steel, Z., and Goldstein, D. (2001). Cultural aspects of cancer genetics: setting a research agenda. *J. Med. Genet.* 38, 425–429.

Meiser, B., Butow, P., Friedlander, M., Barratt, A., Schnieden, V., Watson, M., Brown, J., and Tucker, K. (2002). Psychological impact of genetic testing in women from high-risk breast cancer families. *Eur. J. Cancer* 38, 2025–2031.

Meiser, B., Gaff, C., Julian-Reynier, C., Biesecker, B.B., Esplen, M.J., Vodermaier, A., and Tibben, A. (2006-2007). International perspectives on genetic counseling and testing for breast cancer risk. *Breast Dis.* 27, 109–125.

Metcalfe, K.A., Esplen, M.J., Goel, V., and Narrod, S.A. (2005). Predictors of quality of life in women with a bilateral prophylactic mastectomy. *Breast J.* 11, 65–69.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, M.L., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P.K., Norris, F.H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J.C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Goldgar, D., Wiseman, R., Kamb, A., and Skolnick, M.H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* 266, 66–71.

Miller, S.M., Roussi, P., Caputo, G.C., and Kruus, L. (1995). Patterns of children's coping with an aversive dental treatment. *Health Psychol.* 14, 236–246.

Miller, S.M., Shoda, Y., and Hurley, K. (1996). Applying cognitive-social theory to health-protective behavior: breast selfexamination in cancer screening. *Psychol. Bull.* 119, 70–94.

Myers, R.H., Taylor, C.A., and Sinsheimer, J.A. (1997). Genetic testing for neuropsychiatric disease: experiences from 8 years of genetic testing for Huntington's disease. In *Progress in Alzheimer's Disease and Similar Conditions* (Heston, L.L., ed.), pp. 137–160. American Psychiatric Press, Washington DC.

Myers, M.F., Chang, M.H., Jorgensen, C., Whitworth, W., Kassim, S., Litch, J.A., Armstrong, L., Bernhardt, B., Faucett, W.A., Irwin, D., Mouchawar, J., and Bradley, L.A. (2006). Genetic testing for cancer susceptibility to breast and ovarian cancer: evaluating the impact of a direct-to-consumer marketing campaign on physicians' knowledge and practices. *Genet. Med.* 8, 361–370.

Nebert, D.W., Zhang, G., and Vesell, E.S. (2008). From human genetics and genomics to pharmacogenetics and pharmacogenomics: past lessons, future directions. *Drug Metab. Rev.* 40, 187–224.

Offit, K. (1998). *Clinical Cancer Genetics: Risk Counseling and Management*. Wiley-Liss, New York.

Oram, H., Cote, M.L., Gonzalez, H.M, Aunderwood, W., III, and Schwartz, A.G. (2007). Family history of cancer: is it an accurate indicator of cancer risk in the immigrant population? *Cancer* 112, 399–406.

Peters, N., Domchek, S.M., Rose, A., Polis, R., Stopfer, J., and Armstrong, K. (2005). Knowledge, attitudes, and utilization of BRCA1/2 testing among women with early-onset breast cancer. *Genet. Test.* 9, 48–53.

Phillips, K.A., Jenkins, M.A., Lindeman, G.J., McLachlan, S.A., McKinley, J.M., Weideman, P.C., Hopper, J.L., and Friedlander, M. L.kConFab Investigators (2006). Risk-reducing surgery, screening and chemoprevention practices of BRCA1 and BRCA2 mutation carriers: a prospective cohort study. *Clin. Genet.* 70, 198–206.

Quaid, K. A., Sims, S.L., Swenson, M.M., Harrison, J.M., Moskowitz, C., Stepanov, N., Suter, G.W., and Westphal, B.J. (2008). Living at risk: concealing risk and preserving hope in Huntington disease. *J. Genet. Couns.* 17, 117–128.

Ray, J.A., Loescher, L.J., and Brewster, M. (2005). Risk-reduction surgery decisions in high-risk women seen for genetic counseling. *J. Genet. Couns.* 14, 473–484.

Rebbeck, T.R. (2001). Prophylactic oophorectomy in BRCA1 and BRCA2 mutation carriers. *J. Clin. Oncol.* 18, 100S–103S.

Rhodes, M., Chapman, P.D., Burn, J., and Gunn, A. (1991). Role of a regional register for familial adenomatous polyposis: experience in the northern region. *Br. J. Surg.* 78, 451–452.

Ritvo, P., Irvine, J., Robinson, G., Brown, G., Murphy, K.J., Matthew, A., and Rosen, B. (2002). Psychological adjustment to familial-genetic risk assessment for ovarian cancer: predictors of non-adherence to surveillance recommendations. *Gynecol. Oncol.* 84, 72–80.

Robins-Wahlin, T.B., Backman, L., Lundin, A., Haegermark, A., Winblad, B., and Anvret, M. (2000). High suicidal ideation in persons testing for Huntington's disease. *Acta Neurol. Scand.* 102, 150–161.

Roter, D.L., Erby, L.H., Larson, S., and Ellington, L. (2007). Assessing oral literacy demand in genetic counselling dialogue: preliminary test of a conceptual framework. *Soc. Sci. Med.* 65, 1442–1457.

Royal, C., Baffoe-Bonnie, A., Kittles, R., Powell, I., Bennett, J., Hoke, G., Pettaway, C., Weinrich, S., Vijayakumar, S., Ahaghotu, C., Mason, T., Johnson, E., Obeikwe, M., Simpson, C., Mejia, R., Boykin, W., Roberson, P., Frost, J., Faison-Smith, L., Meegan, C., Foster, N., Furbert-Harris, P., Carpten, J., Bailey-Wilson, J., Trent, J., Berg, K., Dunston, G., and Collins, F. (2000). Recruitment experience in the first phase of the African-American hereditary prostate cancer (AAHPC) study. *Ann. Epidemiol.* 10, S68–S77.

Schoenfeld, M., Myers, R.H., Cupples, L.A., Berkman, B., Sax, D.S., and Clark, E. (1984). Increased rate of suicide among patients with Huntington's disease. *J. Neurol. Neurosurg. Psych.* 47, 1283–1287.

Schneider, K.A., DiGianni, L.M., Patenaude, A.F., Klar, N., Stopfer, J.E., Calzone, K.A., Li, F.P., Weber, B.L., and Garber, J.E. (2004). Accuracy of cancer family histories: comparison of two breast cancer syndromes. *Genet. Test.* 8, 222–228.

Schwartz, M.D., Peshkin, B.N., Hughes, C., Main, D., Isaacs, C., and Lerman, C. (2002). Impact of *BRCA1/BRCA2* mutation testing on psychologic distress in a clinic-based sample. *J. Clin. Oncol.* 20, 514–520.

Schwartz, M.D., Lerman, C., Brogan, B., Peshkin, B.N., Halbert, C.H., DeMarco, T., Lawrence, W., Main, D., Finch, C., Magnant, C., Pennanen, M., Tsangaris, T., Willey, S., and Issacs, C. (2004). Impact of BRCA1/BRCA2 counseling and testing on newly-diagnosed breast cancer patients. *J. Clin. Oncol.* 22, 1823–1829.

Shields, A.E., Burke, W., and Levy, D.E. (2008). Differential use of available genetic tests among primary care physicians in the United States: results of a national survey. *Genet. Med.* 10, 404–414.

Shiloh, S., Koehly, J., Jenkins, J., Martin, J., and Hadley, D. (2008). Monitoring coping style moderates emotional reactions to genetic testing for hereditary nonpolyposis colorectal cancer: a longitudinal study. *Psychooncology* 17, 746–755.

Smith, K.R., West, J.A., Croyle, R.T., and Botkin, J.R. (1999). Familial context of genetic testing for cancer susceptibility: moderating effect of siblings' test results on psychological distress one to two weeks after *BRCA1* mutation testing. *Cancer Epidemiol. Biomar. Prev.* 8, 385–392.

Stoeffel, E.M., Garber, J.E., Grover, S., Russo, L., Johnson, J., and Syngal, S. (2003). Cancer surveillance is often inadequate in people at high risk for colorectal cancer. *J. Med. Genet.* 40, e54.

Struewing, J.P., Hartge, P., Wacholder, S., Baker, S.M., Berlin, M., McAdams, M., Timmerman, M.M., Brody, L.C., and Tucker, M.A. (1997). The risk of cancer associated with specific mutations of *BRCA1* and *BRCA2* among Ashkenazi Jews. *N. Engl. J. Med.* 336, 1401–1408.

Tercyak, K.P., Peshkin, B.N., DeMarco, T.A., Brogan, B.M., and Lerman, C. (2002). Parent-child factors and their effect on communicating BRCA1/2 results to children. *Patient Educ. Couns.* 47, 145–153.

Tercyak, K.P., Peshkin, B.N., Brogan, B.M., DeMarco, T., Pennanen, M.F., Willley, S.C, Magnant, C.M., Rogers, S., Isaacs, C., and Schwartz, M.D. (2007). Quality of life after contralateral prophylactic mastectomy in newly diagnosed high-risk breast cancer patients who underwent BRCA1/2 gene testing. *J. Clin. Oncol.* 25, 285–291.

Thompson, H.S., Valdimarsdottir, H.B., Duteau-Buck, C., Guevarra, J., Bovbjerg, D.H., Richmond-Avellanda, C., Amarel, D., Godfrey, D., Brown, K., and Offit, K. (2002). Psychosocial predictors of BRCA counseling and testing decisions among urban African-American women. *Cancer Epidemiol. Biomar. Prev.* 11, 1579–1585.

van Dijk, S., Otten, W., Timmermans, D.R., van Asperen, C.J., Meijers-Heijboer, H., Tibben, A., Breuning, M.H., and Kievit, J. (2005). What's the message? Interpretation of an uninformative BRCA1/2 test result for women at risk of familial breast cancer. *Genet. Med.* 7, 239–245.

Vasen, H.F., van Ballegooijen, M., Buskens, E., Kleibeuker, J.K., Taal, B.G., Griffioen, G., Nagengast, F.M., Menko, F.H., and Meera Khan, P. (1998). A cost-effectiveness analysis of colorectal screening of hereditary nonpolyposis colorectal carcinoma gene carriers. *Cancer* 82, 1632–1637.

Watson, M., Kash, K.M, Homewood, J., Ebbs, S., Murday, V., and Eeles, R. (2005). Does genetic counseling have any impact on management of breast cancer risk? *Genet. Test.* 9, 167–174.

Weitzel, J.N., McCaffrey, S.M., Nedelcu, R., MacDonald, D.J., Blazer, K.R., and Cullinane, C.A. (2003). Effect of genetic cancer risk assessment on surgical decisions at breast cancer diagnosis. *Arch. Surg.* 138, 1323–1328.

Werner-Lin, A.V. (2007). Danger zones: risk perceptions of young women from families with hereditary breast and ovarian cancer. *Fam. Process* 46, 335–349.

Wideroff, L., Vadaparampil, S.T, Greene, M.H., Taplin, S., Olson, L., and Freedman, A.N. (2005). Hereditary breast/ovarian and colorectal cancer genetics knowledge in a national sample of US physicians. *J. Med. Genet.* 42, 749–755.

Williams-Jones, B. (2003). Where there's a web, there's a way: commercial genetic testing and the Internet. *Community Genet.* 6, 46–57.

Williams, W.R., and Strong, L.C. (1985). Genetic epidemiology of soft tissue sarcomas in children. In *Familial Cancer* (Mueller, H., and Webers, W., eds). Basel Switzerland, S. Karger.

Wolf, C.R., Smith, G., and Smith, R.L. (2000). Science, medicine, and the future: pharmacogenetics. *BMJ* 320, 987–990.

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., Micklem, G., Barfoot, R., Hamoudi, R., Patel, S., Rice, C., Biggs, P., Hashim, Y., Smith, A., Connor, F., Arason, A., Gudmundsson, J., Ficenec, D., Kelsell, D., Ford, D., Tonin, P., Bishop, D.T., Spurr, N., Ponder, B.A.J., Eeles, R., Peto, J., Devilee, P., Cornelisse, C., Lynch, H., Narod, S., Lenoir, G., Egilsson, V., Barkadottir, R.B., Easton, D.F., Bentley, D.R., Futreal, P.A., Ashworth, A., and Stratton, M.R. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378, 789–792.

Zakowski, S.G., Valdimarsdottir, H.B., Bovbjerg, D., Borgen, P., Holland, J., and Kash, K. (1997). Predictors of intrusive thoughts and avoidance in women with family histories of breast cancer. *Ann. Behav. Med.* 19, 362–369.

Chapter 39

General Considerations Concerning Safety in Biomedical Research Laboratories

William E. Grizzle, Walter C. Bell and Jerry Fredenburgh
Department of Pathology, University of Alabama at Birmingham, USA

39.1 INTRODUCTION

The operation of biomedical research laboratories safely requires attention to many complex issues that depend on the goals and activities of each specific laboratory. For example, if a laboratory works with samples of human or animal tissues and/or fluids, then precautions to minimize the exposure of all laboratory personnel to blood-borne human pathogens are required. Most countries and provinces have regulations related to safety precautions that may need to be followed by laboratories. In the USA, regulations specify approaches necessary to protect employees from blood-borne pathogens, general laboratory chemicals and specific chemicals such as formaldehyde. In contrast, if no radioactive material is stored or used in a laboratory, a component of a safety plan focused on radiological safety may be unnecessary. Safety issues related to fire, electrical, and physical safety must be considered by all laboratories; however, regulatory requirements related to these areas of safety are usually promulgated by regional and local governments. Each laboratory should determine which areas of safety are relevant and develop a safety program to protect its employees from exposure to all laboratory hazards. This includes following local, regional, national, and international regulations related to safety as well as education in safety issues that may apply to a specific laboratory/organization.

The approaches to safety discussed in this chapter are based primarily upon regulations that govern laboratory safety in the USA. Nevertheless, the approaches to safety discussed can be generalized to any biological/biomedical research laboratory.

The safe operation of a laboratory that focuses on biomedical research is difficult. No single source of information can provide all the information and guidance necessary to develop an adequate safety program/plan to protect laboratory personnel from various hazards. The approaches and information presented in this chapter are not adequate to ensure the safety of all laboratory personnel or to ensure that a laboratory is able to meet regulatory or accreditation standards in safety. The authors are providing only a starting point, and such a starting point is not static in that there are new issues related to laboratory safety that develop on a daily basis.

39.2 HELP IN UNDERSTANDING REGULATORY ISSUES AND GUIDELINES IN LABORATORY SAFETY: INTERNATIONAL, NATIONAL, REGIONAL, AND LOCAL

There are numerous aids available for developing safety plans for biomedical laboratories. This is due to the extensive governmental regulations focused on protecting the health and safety of laboratory employees. Along with regulations, many governmental and non-governmental organizations provide guidance concerning how to meet specific regulations related to safety and maintaining a safe working environment. For example, web-based resources which may help in understanding the regulations in the USA and some other countries concerning safety are listed in Table 39.1, while general informational resources, such as review articles (Beekmann and Doebbeling, 1997; Cardo and Bell, 1997; Grizzle and Fredenburgh, 2001; Grizzle and Polt, 1988; Padhye *et al*., 1998; Richmond *et al*., 1996; Sewell, 1995) and books (Table 39.2) are available that may aid in developing a safety program. Biomedical laboratories outside the USA also will find many of these educational sites to be useful in developing their safety programs.

TABLE 39.1 Internet resources on safety.

Website	Organization	Topics
General safety		
http://www.osha.gov	Occupational Safety and Health Administration, Department of Labor, USA	Current developmental and operational regulations; technical information; prevention information; training information; links relevant to other sites
http://www.rmlibrary.com/db/lawosha.htm		Occupational safety laws of all 50 states
http://www.healthsystem.virginia.edu/internet/epinet/	University of Virginia, International Health Care Worker Safety Center	Surveillance data
http://www.healthsystem.virginia.edu/internet/epinet/about_epinet.cfm	Exposure Prevention Information Network (EpiNet)	Surveillance information
http://www.cap.org	College of American Pathologists	General and technical information; laboratory management; laboratory safety
http://www.lbl.gov/ehs/pub3000	Lawrence Berkeley National Laboratory Health and Safety	Health and safety manual
http://www.clsi.org/	National Committee for Clinical Laboratory Standards name changed to Clinical and Laboratory Standards Institute (CLSI)	General and technical information; forums; safety; links
Biological safety		
http://www.cdc.gov	Centers for Disease Control and Prevention, Atlanta, GA	Surveillance data; prevention information; technical information; biohazards; links
http://www.cdc.gov/ncpdcid/	National Center for Preparedness, Detection, and Control of Infectious Diseases (NCPDCID)	These four websites replace the National Center for Infectious Diseases
http://www.cdc.gov/vaccines/about/default.htm	National Center for Immunization and Respiratory Diseases (NCIRD)	
http://www.cdc.gov/nczved/	National Center for Zoonotic, Vector-Borne, and Enteric Diseases (NCZVED)	
http://www.cdc.gov/nchhstp/	National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP)	
http://www.fda.gov/cber	Food and Drug Administration, Center for Biologics Evaluation and Research	Information on recalls, withdrawals; and safety issues concerning biologics
http://www.absa.org	American Biological Safety Association (ABSA)	Technical information
http://npic.orst.edu/	The services previously provided by NAIN (National Antimicrobial Information Network) are now offered by the National Pesticide Information Center (NPIC)	Technical information on disinfectants; links
http://www.epa.gov/	Environmental Protection Agency	
http://www.defra.gov.uk/	UK Department for Environment, Food and Rural Affairs (DEFRA)	Surveillance data on BSE in Europe; technical information on prions
http://www.cjd.ed.ac.uk	The National Creutzfeldt-Jakob Disease Surveillance Unit (NCJDSU)	Surveillance data on CJD; technical information on prions; links to other related sites
Chemical safety		
http://www.cdc.gov/niosh/database.html	National Institute for Occupational Safety and Health (NIOSH)	Databases and information resource links and publications

(*Continued*)

TABLE 39.1 (Continued)

Website	Organization	Topics
http://www.ilo.org/public/english/protection/safework/cis/products/icsc/dtasht/index.htm	International Occupational Safety and Health Information Center (CIS)	Chemical database; International Chemical Safety Cards (ICSC)
http://response.restoration.noaa.gov/chemaids/react.html	Office of Response and Restoration (OR&R)	Chemical reactivity worksheet; chemical database of reactivity of substances or mixtures of substances
http://www.cdc.gov/niosh/chem-inx.html	Master Index of Occupational Health Guidelines for Chemical Hazards (NIOSH)	Guidelines for chemical hazards of specific chemicals
Electrical safety		
http://www.lbl.gov/ehs/pub3000/CH08.html	Lawrence Berkeley National Laboratory Health and Safety	Electrical safety program
http://web.princeton.edu/sites/ehs/labguide/sec_5.htm	Princeton University	Laboratory electrical safety program
http://www.ehs.uconn.edu/Word%20Docs/Electrical%20Safety%20in%20the%20Lab.pdf	University of Connecticut Environmental Health and Safety	Electrical safety in the laboratory
Fire safety		
http://www.lbl.gov/ehs/pub3000/CH12.html	Lawrence Berkeley National Laboratory Health and Safety	Fire prevention and protection program
http://www.sunysb.edu/ehs/	Stony Brook University Environmental Health and Safety	Laboratory fire safety hazard assessment and work practices
Radiological safety		
http://www.lbl.gov/ehs/pub3000/CH21.html	Lawrence Berkeley National Laboratory Health and Safety	Radiation safety program
http://www.jmu.edu/safetyplan/radiology/advisorycommittee.shtml	James Madison University	Radiation protection program

39.3 GENERAL CONSIDERATIONS IN LABORATORY SAFETY

A safety program should be developed and utilized to minimize the chance of injuries to employees in the work environment. As a first step, the likelihood and potential sources of specific injuries for each employee must be determined separately. These risks will depend upon the procedures/activities that an employee performs (i.e. his job) as well as all locations in which the employee is likely to spend time. For example, while office personnel usually would not be expected to be exposed to biohazards, such exposure might occur if the secretary/office personnel pick up contaminated material to be typed, filed, or transferred in a room in which biohazards are present or come into direct contact with surfaces (e.g. benches) or equipment (e.g. computers) which might be contaminated. When these conditions exist for office personnel a safety plan must be written to address specific work hazards that might impose a danger to these employees.

Both the employee and his direct supervisor should identify and agree to the employee's potential sources (risks) of injury. The safety plan should be developed so that the likelihood of injury from any of the potential sources of injury can be minimized via modifications of procedures or changes in engineering. Examples of changes in engineering would include an improvement of ventilation within a specific area in which volatile chemicals are in use.

39.4 TRAINING IN SAFETY

Training in each area of safety (e.g. biohazards and chemical hazards) follows the same general approaches and has the same general requirements. Training in each area of safety that is important to a specific employee should be provided before the employee begins work in the laboratory. Training must be provided by trainers who are knowledgeable in the area of safety and should be presented in a language and at an educational level that is appropriate for the employees being trained. The training also should be appropriate for the risks to which each employee may be exposed (e.g. for sites not using or storing radioactive material, no training in radiological safety may be required). Thus there will be a need for different levels of training in safety based upon the needs and requirements of each specific employee. This is especially important for employees or contractual personnel who may not be performing risky tasks, but are exposed to risky environments.

An example is janitorial personnel who may clean areas of the laboratory. They must be educated as to safety issues associated with the spaces they clean. The training should be updated periodically for each employee according to governmental regulations; in the USA, training of employees in specific areas of safety (e.g. blood-borne pathogens) must be updated annually and a record of the training must be documented in the records of each employee. Records of employee training should be maintained according to regulations; in the USA records must be kept for at least three years.

39.5 SAFETY INFRASTRUCTURE

Regulations may vary as to who is ultimately responsible for employee safety; however, in the USA, the chief executive officer (CEO) of each organization has total responsibility for the safe operation of all components of the organization; depending on the country, the CEO may be subject to civil and/or criminal penalties depending on safety violations, the extent of any injuries resulting from safety violations/problems, and prior efforts to correct problems in safety.

The CEO usually appoints a safety committee (SC) which is responsible for developing and implementing and periodically reviewing the overall safety plan (SP) of the institution. The periodic review of the safety plan includes evaluation of the success of the safety plan and required modifications of the safety plan to correct problems with safety or training in safety. The SC appoints a safety officer to administer the safety program. Very large organizations that have many separate large divisions may have separate divisional safety committees and safety officers, especially when specific hazards are limited to only one division (e.g. formaldehyde may be used in only one small area of one of multiple buildings. Specifically, a large clinical laboratory that is part of a university may frequently have a separate safety committee, safety officer, and safety plan. Usually, the administrator responsible for areas with increased safety concerns appoint these separate safety officers; however, the safety committee, officer, and safety plan are always subordinate to the principal safety committee, officer, and safety plan.

The safety officer (SO) is responsible for day-to-day issues related to safety; the safety officer establishes and monitors the training program in safety and ensures the compliance of all employees with training in safety. The SO monitors the safety plan and evaluates all incidents related to safety, especially all injuries of employees. The SO must evaluate all incidents/injuries and must recommend to the safety committee changes to the safety program to prevent recurrence of such incidents and injuries. In the day-to-day operations of the safety program, the safety officer works closely with area supervisors to monitor and to ensure local safety. While a working safety infrastructure is important, the responsibility for safe operation of an organization falls primarily on each and every employee.

Biomedical laboratories are potentially very dangerous working environments compared to other environments because so many potential dangers exist in the typical biomedical laboratory. Not only are there biohazards associated with specimens of human and/or animal tissues or fluids, but also with cell-lines or cellular components developed or produced from humans, animals, or microorganisms. The type of laboratory may limit the type of biohazards to which employees are exposed; however, it would be a rare biomedical laboratory that does not contain a wide variety of chemicals, some of which may be potential carcinogens and allergens such as formaldehyde; others may be teratogens, toxins and irritants, or may be flammable/explosive. Similarly, any facility/space will have physical hazards, electrical hazards, and fire hazards. Also, many laboratories will use some type of radioactivity (e.g. α-, β- or γ-emitters), and all forms of radioactivity are dangerous.

If a laboratory operates within safety infrastructure as described, the safety officer together with the laboratory supervisor(s) will monitor safety in the laboratory. If such an infrastructure does not exist at an organization, it may be necessary to establish a mini-infrastructure in safety that follows the safety approach outlined previously. Specifically, a senior technologist may be appointed as a safety officer. This individual would then be responsible for developing a safety plan, training laboratory personnel in safety, and disposal of dangerous materials.

The safety plan plus the general details of the safety requirements and national regulations related to safety should be available to all employees. Providing specific protocols and details to ensure safety in laboratories, especially those handling human and/or animal tissues, chemicals and radioactivity, are beyond the scope of this chapter. Multiple books and articles are devoted to specific safety information, which are appropriate for safety training and for establishing or improving a safety program (see also section 39.2 and Table 39.2).

39.5.1 Biological safety

Animal and especially all human tissues are inherently dangerous and must be handled with universal precautions (Grizzle and Fredenburgh, 2001). All employees of organizations handling or processing tissues must be educated in the dangers of tissues as well as governmental regulations which apply to handling or being exposed to human and animal tissues. Similarly, those transferring tissues to other individuals or laboratories should require that anyone who receives the tissues as well as all those at the receiving site who may handle or contact the tissues are educated in the potential dangers of human tissue (e.g. what to do to avoid

TABLE 39.2 Books on safety.

Block, S.S. (2001). *Disinfection, Sterilization, and Preservation*, 5th edition. Lea and Febiger, Philadelphia, PA.
Bloom, B.R. (1994). *Tuberculosis: Pathogenesis, Protection and Control*. American Society for Microbiology Press, Washington, DC, pp. 85–110.
Centers for Disease Control and Prevention/National Institutes of Health (1999). *Biosafety in Microbiological and Biomedical Laboratories*, 4th edition. US Government Printing Office: US Department of Health and Human Services, Public Health Service, CDC and NIH, Washington, DC. http://www.cdc.gov/od/ohs/pdffiles/4th%20BMBL.pdf
Fleming, D.O., Richardson, J.H., Tulis, J.J., and Wesley, D. (1995). *Laboratory Safety: Principles and Practices*, 2nd edition. American Society for Microbiology Press, Washington, DC.
Fredenburgh, J.L., and Grizzle, W.E. (1993). *Safety and Compliance in the Histology Laboratory: Biohazards to Toxic Chemicals* (available only at workshops).
Furr, A.K. (2000). *CRC Handbook of Laboratory Safety*, 5th edition. CRC Press, Boca Raton, FL.
Heinsohn, P.A., Jacobs, R.R. and Concoby, B.A. (1996). *AIHA Biosafety Reference Manual*, 2nd edition. American Industrial Hygiene Association, Fairfax, VA.
Kent, P.T., and Kubica, G.P. (1985). *Public Health Mycobacteriology. A Guide for the Level III Laboratory*. US Department of Health and Human Services, Public Health Service, CDC, Atlanta, GA.
Kubica, G.P., and Dye, W.E. (1967). *Laboratory Methods for Clinical and Public Health Mycobacteriology*. Public Health Service Publication, Number 1547. US Department of Health, Education, and Welfare. United States Government Printing Office, Washington, DC.
Lieberman, D.F. (1995). *Biohazards Management Handbook*. Marcel Dekker, New York. Biosafety in the Laboratory: Prudent practices for the handling and disposal of infectious materials.
Miller, B.M. (1986). *Laboratory Safety Principles and Practices*. American Society for Microbiology, Washington, DC.
Richmond, J.Y. (1997). *Designing a Modern Microbiological/Biomedical Laboratory: Lab Design and Process and Technology*. American Biological Safety Association.
Richmond, J.Y. (2000a). *Anthology of Biosafety I: Perspectives on Laboratory Design*. American Biological Safety Association, Mundelein, IL.
Richmond, J.Y. (2000b). *Anthology of Biosafety II: Facility Design Considerations*. American Biological Safety Association, Mundelein, IL.
Richmond, J.Y. (2000c). *Anthology of Biosafety III: Application of Principles*. American Biological Safety Association, Mundelein, IL.
Wald, P.H., and Stave, G. (2001). *Physical and Biological Hazards of the Workplace*, 2nd edition. Van Nostrand Reinhold, New York. Preventing Occupational Disease and Injury.

exposures via cuts, sticks, splashes, oral, and respiratory transmissions and actions to take if exposures occur); they also must be familiar with applicable governmental regulations related to exposure to human tissues.

The general issues in training in safety apply to training in biohazards (e.g. training before beginning work, periodically thereafter and proper record-keeping). The requirements/standards of the educational program in safety as to biohazards may vary; for example, some may require training in how various pathogenic organisms are transmitted and affect man, including the course of the disease.

All human tissues and to a lesser extent animal tissues, whether fixed, fresh, frozen, freeze dried, or paraffin embedded, should be considered as biohazardous. As the extent of alteration of tissue increases (e.g. fresh → frozen → fixed → paraffin embedded) the risks from various infective agents are usually reduced with the greatest reduction occurring following fixation and tissue processing to paraffin; however, certain agents such as prions (e.g. the infective agents for Creutzfeldt-Jakob disease, mad cow disease, deer/elk wasting disease, scappie) may still be infective even when tissues are fixed in a wide range of fixatives (e.g. 10% neutral buffered formalin) and processed to paraffin blocks. Also, spores of certain bacteria such as anthrax which may be present in animal tissues (e.g. pelts) as well as specific dirt/soils may be infective for decades. Thus, all human and animal tissues independent of their physiochemical state should be treated with universal precautions, i.e. should be handled as if infected with agents that may be pathogenic to humans (Grizzle and Fredenburgh, 2001).

Other biological hazards in laboratories may include systems for transfecting cells with specific genetic products. Biological hazards may result when the transfection system includes self-replicating viral vectors. The safety of such systems always should be considered. Also, reagents

and research products derived from humans or animals may be biohazardous. This is especially true when produces are derived from pooled samples from many humans or animals. For example, growth hormone extracted from human pituitaries may be contaminated in Creutzfeldt-Jakob disease. Similarly, reagents from humans (serum) may also be contaminated with biohazardous agents. Also, products from cows (sheep) such as albumin and bovine serum have in the past been infected with prions (e.g. mad cow disease) and/or other agents and have subsequently been recalled.

With increased international travel, newly identified infectious agents may spread rapidly through populations – severe acute respiratory syndrome (SARS), which developed in China and spread rapidly to multiple countries; similarly, monkey pox and West Nile virus which have been acquired in or transferred to North America from exotic pets. Similarly, with the possibility of bioterrorism (see also Chapter 28), agents that may be encountered very rarely such as anthrax and small pox could represent agents to which laboratory personnel are exposed. When there is notice of a ''new'' disease which may be studied and diagnosed or to which employees may be exposed, a plan to deal with such conditions should be developed. These laboratory safety issues are in addition to those prescribed for the usual pathogens potentially encountered in a laboratory dealing with human and animal tissues.

In the USA, regulation 29 CFR Part 1910.1030 ''Occupational Exposure to Bloodborne Pathogens'' specifies the general requirements for infrastructure as well as other organizational requirements needed to protect employees from blood-borne pathogens such as hepatitis B. Any laboratory in the USA that deals with human tissues may need to meet the requirements of these regulations. The development of a safety program in biohazards is outlined in Table 39.3.

General approaches to bio-safety include identifying potential hazards without regard to standard operational procedures (SOPs) or to the use of safety equipment. After potential hazards/dangers are identified, SOPs should be modified to reduce the likelihood of injuries from biohazards. SOPs should include requirements for frequent washing of hands, for using safety equipment, for providing employees with hepatitis B vaccinations and related preventive medical support. All employees should be trained with respect to minimizing biohazards. Medical support should be provided when employees are injured and work practices should be changed based on the analysis of safety incidents to prevent future injuries. The safety plan should be evaluated yearly and improved to prevent the recurrence of injuries. Clear, detailed records of the above approaches to bio-safety should be maintained.

Most laboratories that handle human tissues for research will at some time be queried as to whether or not an individual from whom tissues were obtained is infected with HIV, hepatitis B, C, or D, West Nile virus, Creutzfeldt-Jakob prions, etc. This query may be prior to use of the tissues and/or after an employee is exposed to infection by the tissue. Organizations that provide tissues for research usually will not have permission to test patients and/or their tissue for such agents. A negative test does not ensure safety of the tissues for that pathogen and also there are no tests for some pathogens. Recipients of tissue specimens must agree to educate employees as to the dangers of tissues with which they come into contact, and not to test tissues for human pathogens. When an employee is exposed (cut or needle stick) to human biohazards, immediate medical care is necessary. If the hepatitis B titer is low, a re-vaccination should be considered. Also, treatment with hyper-immune gamma globulin might be warranted. After acute medical care is provided, and appropriate medical advice provided,

TABLE 39.3 The key steps in developing a biological safety program.

1. Identify requirements related to biohazards promulgated by governmental and laboratory accrediting organizations; collect up-to-date information as to bio-safety from likely sources. Use this information in developing or updating an overall safety program and in training programs related to biohazards.
2. Develop the organizational infrastructure necessary to develop and maintain a safety program including a training program in biohazards.
3. Identify risks and general issues of bio-safety for each employee in your laboratory. This includes identification of work activities and the safety issues of each work activity as well as risks associated with various work spaces.
4. Develop written guidelines to ensure bio-safety based on published information, federal, state, and local regulations as well as local and consultant experience. These guidelines should be reviewed and updated periodically and modified as soon as possible to correct any identified problems or incidents. Maintain records of personnel safety incidents, their evaluation, and actions to prevent future similar incidents.
5. Develop and implement a training program of which a major focus is bio-safety and maintain records of employee training.
6. Develop a resource that provides medical support for prevention of disease (e.g. vaccination) as well as medical aid to employees who have been exposed to specific pathogens.

the employee can then be monitored for evidence of those human pathogens for which testing is available.

In addition to considering biological dangers associated with human tissues and materials that may be used within the laboratory, some biohazardous conditions may develop within the laboratory independent of the use of human or animal tissues. For example, the ventilation system, cooling system, condensation drains, drains of sinks, animal debris, bedding, incubators, cell culture, or other growth media, refrigerators/freezers, and/or unconsumed food may develop colonization with fungal species or other biohazardous agents that especially may affect laboratory personnel who may have compromised immune systems (e.g. employees with HIV infections or those taking immunosuppressive agents such as steroids for asthma or arthritis or chemotherapeutic drugs for cancer). Also, some common fungi, such as *Cryptococcus* sp., may affect apparently healthy individuals. These same agents via circulating spores may also compromise biological experiments. All biological laboratories should maintain strict standards of cleanliness with, for example, periodic decontamination of the drains of sinks and prohibition of use of food and drinking material in the laboratory. Material and debris should be cleared from the laboratory.

39.5.2 Chemical Safety

There are several federal regulations related to chemical safety in the USA that may affect laboratories. These include Occupational Exposure to Hazardous Chemicals in Laboratories (29CFR 1910.1450), the Hazard Communication Standard (29 CFR 1910.1200), and the Formaldehyde Standard (29 CFR 1910.1048). Most laboratories must abide by the Occupational Exposure to Hazardous Chemicals in Laboratories law (29CFR 1910.1450). This law mandates that employers develop a written chemical hygiene plan; it is the core of the standard. The chemical hygiene plan must be capable of protecting employees from hazardous chemicals in the laboratory and capable of keeping chemical exposures below the action level or in its absence the permissible exposure limit (PEL). Table 39.4 consists of the designated elements that a chemical hygiene plan must include. A laboratory involved with fixation of tissue specimens should also understand and incorporate into its plan the Formaldehyde Standard.

Most laboratories deal with relatively small amounts of most chemicals; nevertheless, even small amounts of specific chemicals may be very dangerous. Extremely small amounts of some chemicals may be toxic or may be carcinogens or teratogens. Similarly, ether, especially old ether which has oxidized after opening or picric acid which has dried out, may constitute explosive hazards. Accidental or unaware combinations of chemicals may cause spontaneous combustion as well as accelerated heating which may cause boiling and splashing, or even explosions. Personnel should avoid direct contact with even small quantities of carcinogens, teratogens, and/or highly toxic agents such as cyanides.

All chemicals that are commercially purchased should have material safety data sheets (MSDS) available for reference for employees who potentially will come into contact with these chemicals. MSDS sheets are prepared by the manufacturer and are available from the manufacturer of the chemical. The MSDS lists the various hazards of the specific chemical to which the MSDS apply. This includes toxicity, explosive potential and categories of danger (e.g. strong oxidizers). The MSDS also specifies procedures to minimize toxic exposures as well as contact information so that additional information on the chemical can be obtained rapidly.

While the dangers of most chemicals are identified in material safety data sheets, combinations of chemicals that may be serious hazards may not be specified clearly. It may be obvious that concentrated strong acids (HCI) should

TABLE 39.4 Mandated elements of a chemical hygiene SOP plan relevant to safety and health must be developed and the following procedures need to be addressed.

1. A written emergency plan should be established to address chemical spills. The plan should include consideration of prevention, containment, clean-up, waste disposal, and disposal of chemically contaminated materials used during the clean-up.
2. A policy to monitor the effectiveness of ventilation and to minimize exposure to potentially dangerous vapors.
3. A policy in reference to ventilation failure, evacuation, medical care, reporting of chemical exposure incidents, and chemical safety drills.
4. Policies prohibiting eating, drinking, smoking, gum chewing, and application of cosmetics in the laboratory should be developed.
5. Policies should be developed to prohibit storing food and/or beverages in storage areas or laboratory refrigerators.
6. Mouth pipetting and mouth suctioning for starting a siphon must be prohibited.
7. Personal protection should be mandated; all persons including visitors must wear appropriate eye protection. Suitable gloves should be worn where there is a potential for contact with toxic chemicals. Gloves should be inspected before using and washed before removing. The use of contact lenses in the laboratory should be avoided.
8. After handling hazardous materials, hands and other possible areas of exposed skin should be washed.

never be combined directly with concentrated bases, e.g. especially strong bases (e.g. NaOH) without dilution and extreme care; however, mixing strong oxidizers (e.g. potassium permanganate) with materials with high carbon content, e.g. ethylene glycol, may cause spontaneous fires and such mixes should also be approached with great care.

These are just a few of the examples which should be included in the educational program in chemical safety.

For biomedical research laboratories, most MSDS do not address specific uses or conditions to which a specific chemical may be exposed; due to this, one should always practice caution. It is common in laboratories to make specific solutions and/or compounds for use in a particular protocol. The following provisions should be applied to chemical substances prepared in the laboratory for specific uses:

- All solutions and/or compounds made in the laboratory must be labeled and include the ingredients and hazards as well as the date prepared.
- If the composition of a chemical substance which is produced exclusively for the laboratory's use is known, the employer must determine if it is a hazardous chemical. If the chemical substance is determined to be hazardous, the employer must provide appropriate training as required for all hazardous chemicals.
- If the chemical produced is a byproduct whose exact composition is not known, the employer shall assume that the substance is hazardous. The employer shall provide appropriate training as required for hazardous chemicals.
- If the chemical substance is produced for another laboratory or site outside of the source laboratory, the employer is required to prepare a material safety data sheet for the other laboratory(s) and follow labeling requirements.

All chemical exposures should be minimized. Few laboratory chemicals are without hazards, universal precautions for handling all laboratory chemicals should be adopted. Some particularly hazardous chemicals will require specific guidelines. The following is a list of prudent rules that should be included in a chemical safety plan:

- Implement an overall chemical hygiene program.
- Skin contact with chemicals should always be avoided.
- Avoid underestimated risks of exposure to chemicals with no known significant hazard.
- Take special precautions with chemicals that present moderate or severe hazards.
- Assume that any mixture will be more toxic than its most toxic component and that all substances of unknown toxicity are toxic.
- Work within adequate ventilation.
- Observe chemical exposure limits when available.
- Know the current legal requirements concerning regulated substances.
- Ensure that all individuals working in the laboratory area have appropriate training concerning hazardous chemicals.
- Take care during maintenance of equipment potentially contaminated with hazardous chemicals.
- Monitor exposure to hazardous chemicals of all employees.

When individuals work with chemicals, the chemical safety plan should ensure that potential injuries are avoided by proper work procedures, by proper clothing and safety equipment, and by extensive education. Such approaches to chemical safety are regulated in the USA by the Department of Labor and are addressed by laws such as the Occupational Exposure to Hazardous Chemicals in Laboratories (29CFR1910.1450) and the Hazard Communication Standard (29CFR1910.1200).

There also are special laws that address particularly dangerous chemicals such as formaldehyde (the Formaldehyde Standard 29CFR1910.1048). This law has specific safety requirements for the preparation and use of formaldehyde.

While laboratories outside the USA may be governed by separate regulations, the above US laws may provide aids in developing a chemical safety program. Also, material safety data sheets are international and understanding such sheets is a great step in laboratory safety.

39.5.3 Electrical Safety

Electrical injuries can be avoided by ensuring that all equipment is grounded and testing them when first purchased and yearly thereafter. Electrical base plugs must be in good condition and grounded. Electrical work should be done with great care ensuring that all areas are protected by removal of fuses and with written warnings at the fuse box. Frequently, personal electrical appliances such as radios, hairdryers, etc., may be ignored when testing for grounding and represent significant dangers. Also, great care should be taken with electrical appliances/equipment around water sources, especially sinks and bathrooms/showers.

39.5.4 Fire Safety

Fire safety can be evaluated by inviting an inspection by the local fire department. Prior to such inspections and at least yearly, fire drills should be practiced and emergency exit pathways should be posted at all room exits. Obviously, emergency exits should never be blocked, obstructed or locked and hallways must not be obstructed or cluttered. Similarly, access to fire blankets, showers, and fire equipment must not be impeded.

Flammable agents should be stored appropriately including storage of large amounts of flammable agents only in a fire cabinet. The amount per container and volume of flammable agents stored on site in laboratories may

be regulated at the local level as well as by accreditation agencies. Fire cabinets or areas where small amounts of flammable chemicals are stored should not be located near exits of rooms/areas.

Smoking should be regulated carefully; similarly, furniture, rugs, and equipment should be constructed of nonflammable material. Regulations for types of doors to serve as fire barriers should be followed as should fire requirements for construction of buildings that house specific activities (e.g. laboratories).

39.5.5 Physical Safety

The physical safety of employees is a consideration that must be considered in all organizations and for all employees. Physical safety ranges from preventing falls to ensuring employees are not physically injured or intimidated by other individuals, either employees or non-employees. Much of a plan for ensuring physical safety involves careful maintenance of the physical plant and facilities. Tears in rugs, broken steps and water, soap, paraffin and other slippery substances on floors, and inappropriate use of ladders or chairs as ladders, all may lead to unnecessary falls. Similarly, unrestrained gas cylinders, unbalanced file cabinets, and inadequately secured shelves all can lead to injuries via falling or moving agents or structures. Also, included in causes of physical injuries are repetitive action injuries and back injuries secondary to inappropriate lifting as well as temperature burns both cold (e.g. liquid nitrogen) and hot.

Great care should be taken with the overall security of the workplace; this includes limiting access to the workplace by unauthorized personnel. There should be no tolerance of threats to employees, especially by other employees. The protection of employee safety from others may extend outside the direct work environment to areas surrounding the workplace, including parking areas.

Physical injuries that are more difficult to avoid include minor cuts (e.g. paper), bumps, and strains due to inattentive actions. However, such minor injuries should not be compounded by exposure, for example of broken skin to biohazards. The other hazards that can be prevented or ameliorated (use of gloves to avoid burns) should be addressed in the overall safety program.

39.5.6 Radiological Safety

Laboratories that purchase, store, and/or use radioactive material should have a radiological safety plan. For laboratories requiring a radiological safety plan, the personnel who utilize or come into contact with radioactive material require extensive training as well as the availability of specific equipment to monitor for the extent of radiation and radioactive contamination. As with other areas of safety, the training must be periodically updated and should be appropriate for language and education of personnel. Everyone who has access to the areas where radiation is used or stored should have appropriate training. Depending on the danger of the radioisotopes used, some employees, usually those with professional degrees, may hold a local ''license'' to use specific isotopes and these licensed personnel are responsible for the direction and supervision of all employees who work under their supervision and all spaces to which they are assigned.

Although radiological safety is a very specialized area, it can be generalized (Table 39.5). Surveys of all laboratory areas which can detect types of radioactive material being used in the laboratory should be performed periodically. Gamma survey meters are an aid indicating contamination by strong β- and γ-emitters but periodic surveys of work,

TABLE 39.5 Essential elements of a radiological safety program.

1. Train in radiological hazards all personnel who have access to areas where radioactivity is used and/or stored and update training periodically.
2. Monitor usage of radioactivity and maintain accurate records as to the usage, disposal, and inventory of radioactive nucleides.
3. Develop a plan to prevent and to minimize radioactive contamination.
4. Develop a plan to survey periodically for radioactive contamination all areas of the laboratory.
5. Monitor radiation exposure of all employees exposed to radiation.
6. Develop a plan to minimize the exposure of all personnel to radiation using appropriate shielding, safety equipment, and optimal procedures.
7. Develop a plan to ensure safe and secure storage of radioactive material.
8. Develop a plan to contain and clean up major radioactive spills.
9. Record and evaluate all incidents involving radioactivity.
10. Maintain careful records of all aspects of radiological safety plan.

floor, and storage areas using swipes are critical to laboratory safety. Contamination is easily spread throughout a building and even to the homes of laboratory personnel. When occupying unsurveyed laboratory space or using unsurveyed equipment (e.g. refrigerators and freezers), these should be certified as radiation free before occupancy or use.

Personnel should be shielded from exposure to radiation, particularly from β- and γ-emitters (e.g. P^{32}, Co^{60}). Not only does this require shielding but also appropriate laboratory clothing and safety equipment. Contamination is avoided by quickly and effectively containing all radioactive spills/contamination and cleaning up all spills or release of radioactivity.

39.6 CONCLUSIONS

The safe operation of a laboratory depends upon all employees working to maintain a safe working environment. This requires establishing a safety infrastructure including developing and monitoring of an effective safety plan and training as to the safety hazards associated with special types of laboratory work – biohazards, chemical hazards, and radiation hazards. All laboratories should consider issues in physical, electrical, and fire safety. Laboratory safety can be maintained with the careful evaluation of all safety incidents and modification of the safety plan to prevent similar incidents.

ACKNOWLEDGEMENTS

This work is supported in part by the Early Detection Research Network (5U24CA86359-08) to William E. Grizzle, the SPORE in Breast Cancer at UAB (2P50CA089019-06A2), the SPORE in Pancreatic Cancer at UAB (5P20CA101955-05), the UAB Skin Disease Research Center (5P30AR509048-04), the Cooperative Human Tissue Network, Southern Division (5U01CA44968-18S2) to William E. Grizzle, and the Cardiovascular Medical Research and Education Fund Pulmonary Hypertension Breakthrough Initiative to William E. Grizzle.

REFERENCES

Beekmann, S.E., and Doebbeling, B.N. (1997). Frontiers of occupational health. New vaccines, new prophylactic regimens, and management of the HIV-infected worker. *Infect. Dis. Clin. North Am.* 11, 313–329.

Cardo, D.M., and Bell, D.M. (1997). Bloodborne pathogen transmission in health care workers. Risks and prevention strategies. *Infect. Dis. Clin. North Am.* 11, 331–346.

Grizzle, W.E., and Polt, S.S. (1988). Guidelines to avoid personnel contamination by infective agents in research laboratories that use human tissues. *J. Tissue Cult. Methods,* 8, 191–200.

Grizzle, W.E., and Fredenburgh, J. (2001). Avoiding biohazards in medical, veterinary and research laboratories. *Biotech. Histochem.* 76(4), 183–206.

Padhye, A.A., Bennett, J.E., McGinnis, M.R., Sigler, L., Fliss, A., and Salkin, I.F. (1998). Biosafety considerations in handling medically important fungi. *Med. Mycol.* 36, 258–265.

Richmond, J.Y., Knudsen, R.C., and Good, R.C. (1996). Biosafety in the clinical mycobacteriology laboratory. *Clin. Lab. Med.* 16, 527–550.

Sewell, D.L. (1995). Laboratory-associated infections and biosafety. *Clin. Microbiol. Rev.* 8, 389–405.

Chapter 40

Quality Management in the Laboratory

David Burnett
Lindens Lodge, Bradford Place, Penarth, United Kingdom

40.1 INTRODUCTION

The introduction of international standards for quality and competence in laboratories and a growing interest in the accreditation of medical laboratories throughout the world has led to an increasing need in medical laboratories for an understanding of quality management and quality management systems.

The International Organization for Standards (ISO) has defined *quality* as ''the degree to which a set of inherent characteristics fulfils requirements'' (ISO 9000:2000). This sort of definition is almost impenetrable to those not familiar with the world of standards, but relating it to a clinical situation can help bring it to life. For example, a patient in a hypoglycemic coma has a requirement for a blood glucose measurement and the inherent characteristics of that measurement are that it be done on the correct specimen in an accurate and timely manner and be properly interpreted in order to provide a quality result or service.

Quality management in the laboratory involves specifying the requirements posed by differing clinical situations and ensuring that the inherent characteristics of the laboratory measurement or examination fulfill the specified requirements, or in other words, it is fit for its intended purpose. A quality management system is the mechanism for quality management and aims to ensure quality. The preamble to Standard A4 in the Clinical Pathology Accreditation (UK) Ltd ''Standards for the Medical Laboratory'' (http://www.cpa-uk.co.uk; Burnett *et al*., 2002) describes a quality management system as providing ''… the integration of organisational structure, processes, procedures and resources needed to fulfil a quality policy and thus meet the needs and requirements of users''. It is this all-embracing concept of a quality management system that this chapter seeks to emphasize.

40.2 INTERNATIONAL STANDARDS AND THEIR ROLE IN ACCREDITATION

40.2.1 Elements of Accreditation

The author, in his book *A Practical Guide to Accreditation in Laboratory Medicine*, recognizes four elements in accreditation systems (Burnett, 2002): the *accreditation body*, which oversees the assessments and grants accreditation; the *standards*, with which a laboratory has to comply in order to gain accreditation; the *assessors*, who establish compliance with the standards by conducting an assessment; and the *laboratory*, which is required, or voluntarily seeks, to comply with the standards. Central to any effective system of accreditation are clearly written and objectively verifiable standards.

40.2.2 International Standards for Quality and Competence

The purpose of international standards is to ''reflect the quest for quality and promote harmonisation of practice from laboratory to laboratory, from country to country'' (http://www.iso.org). There are three standards published by different groups within ISO that deal with quality and competence and are useful in different situations. These standards and their application are shown in Table 40.1.

ISO 9001:2000 is conceptually important in establishing a working quality management system but does not include competence requirements, whereas ISO/IEC 17025:2005 does include competence requirements but does not address issues specific to medical laboratories. The standard of choice for medical laboratories is ISO 15189:2007. It takes into account the special constraints imposed by the medical environment, the pre- and post-examinations aspects of the

TABLE 40.1 International standards for quality and competence.

Title of the standard	Application
ISO 9001:2000 Quality management systems – Requirements	For certification of quality management systems
ISO/IEC 17025:2005 General requirements for the competence of testing and calibration laboratories	For accreditation of the quality and competence of testing and calibration laboratories
ISO 15189:2007 Medical laboratories – Particular requirements for quality and competence	For accreditation of the quality and competence of medical laboratories

work, the role of diagnostic manufacturers, and emphasizes the essential contribution that medical laboratory services make to patient care through advisory, interpretative, and educational services.

A third edition of ISO 15189 will be published in late 2008/early 2009 and will set out the requirements for quality and competence in more accessible fashion than the current standard. As with the first two editions of ISO 15189 it owes its origins to ISO 9001 and ISO 17025; however, this new edition aligns its quality requirements more closely to ISO 9001 and its technical competences are specifically related to the medical laboratory environment.

40.2.3 Accreditation Bodies and Standards

Until the official publication of ISO 15189:2003, accreditation bodies throughout the world were adopting ISO/IEC 17025:1999 as the basis for assessing performance of testing and calibration laboratories. It was seen as applicable to all laboratories whether they tested for pesticide residues in fresh fruit or examined a renal biopsy for the evidence of disease. Some accreditation bodies have been reluctant to adopt a standard specific for medical laboratories, but increasely throughout the world medical laboratories are using ISO 15189 as the standard of choice. The General Assembly of the International Laboratory Accreditation Cooperation (http://www.ilac.org) has stated that when ISO 15189 is published, ''medical laboratories may be accredited to that standard as an alternative to ISO/IEC 17025''. The resolution is a compromise between those accrediting bodies that wanted a single generic standard for all laboratories and those who understand the desire of the medical laboratory community to be accredited to a standard specifically designed for medical laboratories.

40.2.4 The Principle of Subsidiarity and Guidelines to Standards

Although harmonization of practice in medical laboratories is important there is more than one way to achieve a particular goal. In the European community, the principle of subsidiarity requires the community to act ''only if and so far as the objectives of the proposed action cannot be sufficiently achieved by Member States, and can therefore, by reason of the scale or the effects of the proposed action, be better achieved by the Community''. Subsidiarity in its original philosophical meaning[1] as being ''concerned with fostering social responsibility'' can be translated as saying that architects of standards and of accreditation schemes must be wary of aggregating to themselves rights that may not be acceptable to the ''fourth element of accreditation'' (Burnett and Blair, 2001; Burnett, 2002), the laboratories to be accredited.

Different situations require different solutions. After national accreditation bodies have decided to adopt a particular standard as a basis for accreditation they often develop secondary documents that facilitate the assessment process and take into account local practice. A good example of these are the standards and guidelines developed by the National Pathology Advisory Council (NPACC) of Australia which include ''Requirements for Quality Management in Medical Laboratories'' (2007) and ''Laboratory Accreditation Standards and Guidelines for Nucleic Acid Detection and Analysis'' (2006), available free of charge from http://www.health.gov.au/npaac. In countries with limited resources or where accreditation is voluntary, benefit can be gained from development of standards or guidelines for national use and this principle of subsidiarity can then be seen as promoting the long-term aim of harmonization.

40.3 A PROCESS-BASED APPROACH TO QUALITY MANAGEMENT SYSTEMS

The starting point for developing a framework for process-based quality management of a medical laboratory lies in the introduction to ISO 9001. It promotes the adoption of ''a process approach when developing, implementing and improving the effectiveness of a quality management system'' in order ''…to enhance customer satisfaction by meeting customer requirements''. Process is described as ''an activity using resources, managed in order to enable

[1] The principle of subsidiarity in its original philosophical meaning was expressed by Pope Pius XI in his Encyclical letter in 1931: "It is an injustice, a grave evil and disturbance of right order for a larger and higher association to arrogate to itself functions which can be performed efficiently by smaller and lower sections".

the transformation of inputs into outputs''. In the context of a medical laboratory this translates into consultation with users, receiving a request for an examination, carrying out the work, and reporting the results, with interpretation where appropriate.

Within any organization (e.g. a medical laboratory) there are numerous interrelated or interacting processes, and it is ''the identification and interactions of these processes and their management'' that is referred to as a process approach. It is the adoption of this approach that creates a process-based quality management system. Figure 40.1, adapted from ISO 9001, represents a model of such a system.

It is helpful to translate some of the terms used in this model into language more familiar to medical laboratory professionals. It can be viewed in two different ways. First, the user (*customer*) has requirements that are formulated in consultation with laboratory management (*5. Management responsibility*) and the laboratory responds by carrying out preexamination, examination, and post-examination processes (*7. Product realization*) to produce a report (*product*) for the user. Depending on whether their requirements have been met or not, users may be defined as satisfied or dissatisfied.

The second view is that of a process model in which laboratory management (*5. Management responsibility*) creates a quality system (*4. Quality management system*) and uses resources, staff, equipment, and so on (*6. Resource management*) to carry out preexamination, examination, and post-examination processes (*7. Product realization*) to fulfill the requirements of the user. The preexamination, examination, and post-examination processes are evaluated continually and improvements made as appropriate (*8. Measurement, analysis and improvement*). Evaluation and continual improvement activities include, for example, assessment of users' needs and requirements, internal audit of the examination processes, and review of participation in external quality assessment schemes.

One of the unsatisfactory aspects of ISO 15189 is that the management (quality management system) and technical (competence) requirements of the standard are presented in two separate sections. This has the unfortunate effect that it is very difficult for laboratories to discover the dynamic relationships between quality and competence requirements. Figure 40.2 resolves this dilemma by placing the appropriate clauses of the standard into a simplified version of the process-based model presented in Fig. 40.1.

40.4 BUILDING A QUALITY MANAGEMENT SYSTEM

In defining a *management system* as a ''system to establish policies and objectives and to achieve those objectives'', ISO 9000:2000 ''Quality management system – Fundamentals and vocabulary'' draws attention to the fact that the overall management system of an organization can include different management systems, for example systems for quality, environmental, or financial management. Thus it defines a *quality management system* as a ''management system to direct and control an organization with regard to quality''. This section discusses the steps to be

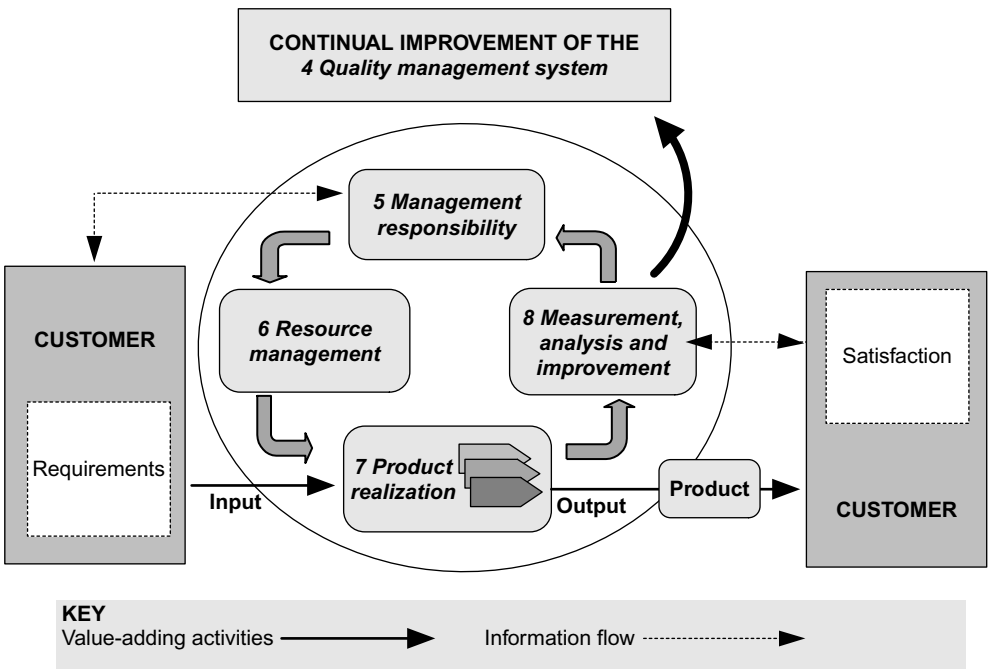


FIGURE 40.1 The ISO 9001:2000 model for a process-based quality management system. The numbers 4–8 in the figure correspond to the main clauses of ISO 9001:2000.

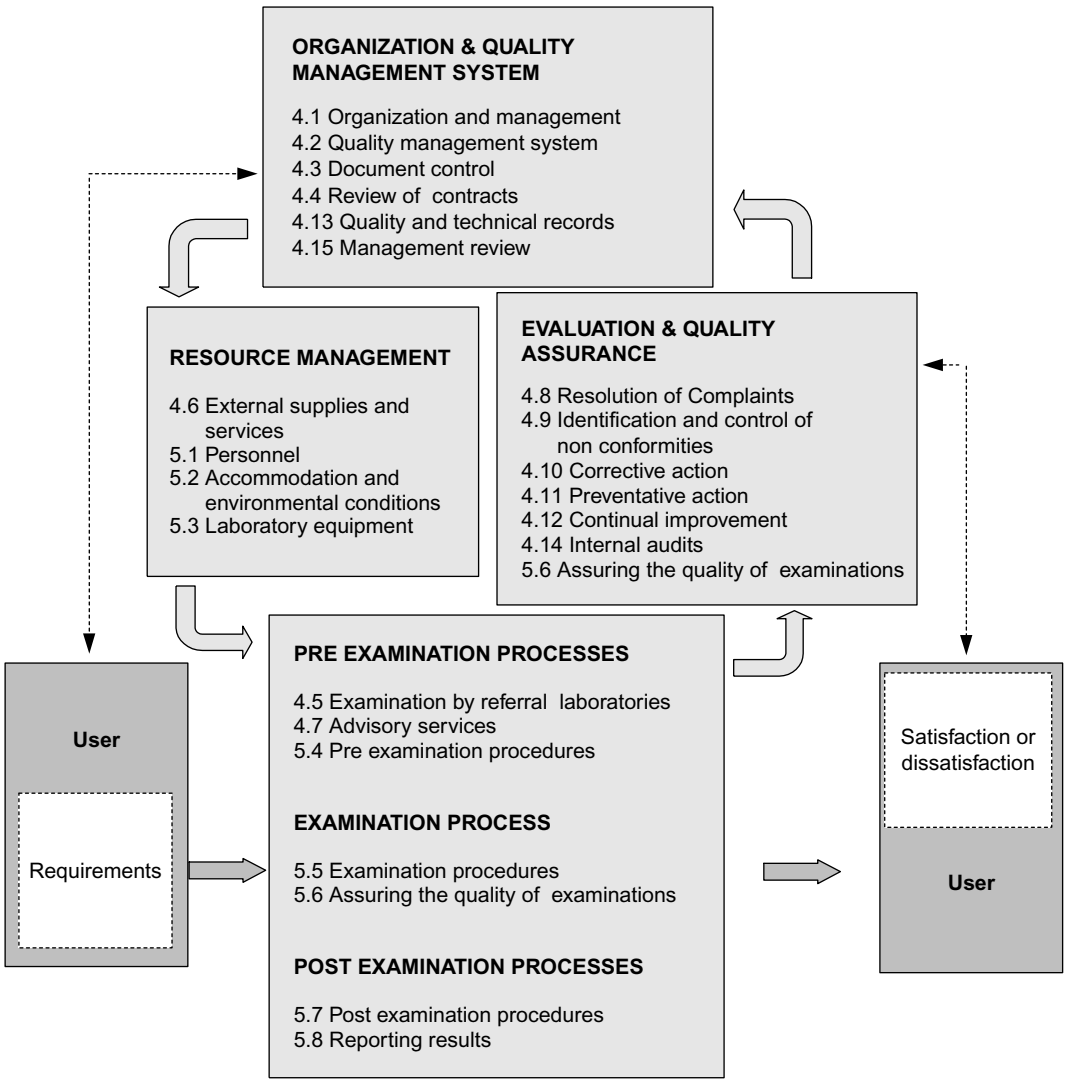


FIGURE 40.2 The requirements of ISO 15189:2003 placed in a process-based quality management system. The numbers 4.1–4.14 and 5.1–5.8 in the figure correspond to the main clauses of ISO 15189:2003.

taken to establish, control, review, and improve a quality management system (QMS).

The National Pathology Accreditation Advisory Council (NPAAC) has published a valuable document, ''Guidelines for quality systems in medical laboratories'' (available at http://www.health.gov.au/npaac). Figure 40.3 is adapted from the quality system flowchart illustrated in the document and shows a structured approach to the establishment, control, review, and improvement of a QMS.

Although the NPAAC document was designed as a practical guide to ISO 17025:1999 it is equally applicable to ISO 15189. The figure shows four elements of building a quality management system – establishment, control, review, and improvement – as a cycle. In each element, a distinction is drawn between management responsibility, which focuses on the establishment and review elements; and organizational responsibility, which focuses on the control and improvement elements. In a medical laboratory, management responsibility would equate with laboratory management, the equivalent term in ISO 9001 being top management. This responsibility is executive in character, in contrast to organizational responsibility, which is corporate and is the responsibility of an organization such as the medical laboratory. In the next two sections the different elements of building a quality management system are establishment, control, review, and improvement.

40.5 ESTABLISHMENT AND CONTROL

40.5.1 Action in Quality Management

As has been seen within the establishment and control stages of creating a QMS, there is a sequence of action in quality management. This sequence is illustrated in a pyramidal

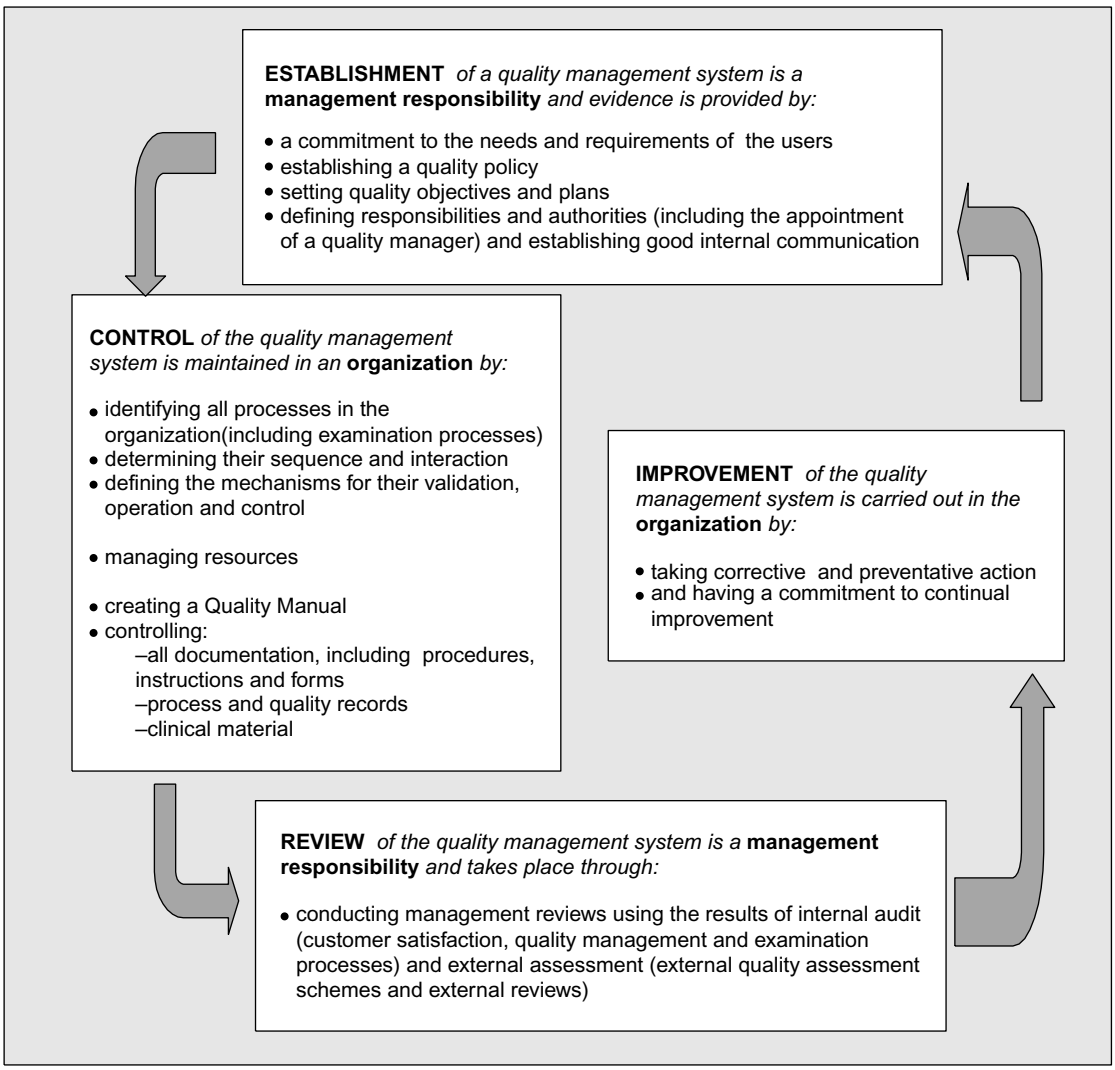


FIGURE 40.3 Building a quality management system.

form on the left-hand side of Fig. 40.4. The first step in the sequence is the creation of policies that can be defined as the overall intentions and direction of an organization. The second step, objectives and plans, involves making plans and setting objectives to enable the fulfillment of the intentions expressed in the policies. The third step, processes, involves the definition of the activities needed to carry out the intentions, and the fourth step, procedures, is the practical way in which intentions are translated into action. The fifth and final step, records (made on forms), provide evidence, on a day-to-day basis, that procedures have been carried out correctly and that intentions have been fulfilled.

In terms of a medical laboratory, this sequence would translate as follows. The quality policy of the laboratory includes a commitment to the reporting of results of examinations in a timely manner. The supplier of the laboratory computer system announces the release of a module for reporting of results. Laboratory management establishes the installation of this module as a quality objective for the next financial year. Planning for this development requires the inclusion of the resource implications in the business plan. Its impact on postexamination processes is defined, and procedures and forms reviewed and where necessary revised.

40.5.2 Documentation and the Need for Evidence

Evidence of action in quality management is adduced from the documentation that is produced and illustrated on the right-hand side of Fig. 40.4. The primary requirement for evidence is to enable the laboratory to reconstruct its examination and other processes as a result of questions asked by users of the laboratory concerning its performance. The other side of the evidence coin is the need of assessors to obtain evidence to enable them to assess a laboratory's compliance with standards.

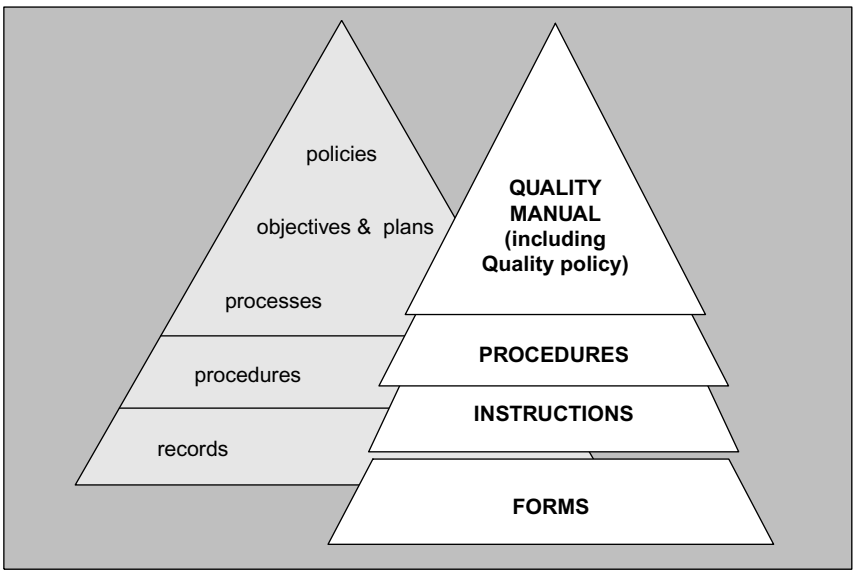


FIGURE 40.4 Hierarchy of documentation.

40.5.3 The Hierarchy of Documentation

Figure 40.4 shows that at the top of this hierarchy is the *quality manual*, the road map to the whole documentation of the laboratory. In practical terms, it should be no more than 25 pages in length. It should contain the *quality policy(s)*, and describe the *processes* that take place in the laboratory in order to fulfill the requirements of particular standards. Examples of such processes are the procurement of equipment, the examination of specimens, and the reporting of results. A policy can be defined as ''setting out the commitment of an organization to follow a particular course of action''. A pathology laboratory can have a single policy statement, which is inclusive of all aspects of its work, or there can be a number of separate policies relating to different aspects of the way in which a laboratory works.

The quality policy itself will be subject to periodic review but is unlikely to change significantly unless the primary purpose of the laboratory were to change. However, in order to pursue and maintain a particular course of action, *objectives* have to be set and *plans* executed. In contradistinction to the quality policy, objectives and plans are constantly changing in response to the changing needs and requirements of the users. The processes involved should be set out in the quality manual, particularly in relation to the annual management review. Throughout the quality manual there should be references to *procedures*, which form the second level in the hierarchy of documentation. Procedures are the practical way in which policies are translated into action and are often called SOPs, or standard operating procedures. The quality policy should refer to management, quality evaluation, health and safety, laboratory methods, and so on, and procedures are needed that relate to the same areas.

In the same way that the quality manual refers to procedures, so procedures can contain references to *(working) instructions*. This third level of documentation involves the practical day-to-day work instructions that are needed near the work situation for easy reference. For example, they might describe starting up or closing down a hematology analyzer, a microbiology plate-pouring machine, or a recipe for staining a slide. Instructions can be part of a procedure or referenced in a procedure and published separately or both in the procedure and separately. The advantage of having them separate is that any changes to instructions do not require a change to the procedure.

The final level in the hierarchy of documentation are the *forms*. These forms (and the records created using them) are a crucial part of quality management; they are the evidence that a procedure and/or related instructions have been carried out.

If the procedure or instructions require something to be recorded on a form, the form should be referred to in the procedure. The forms or records do not necessarily have to be created as hard copy (a paper record). An electronic record can be created by completing a form on a computer screen in the laboratory or a consultant's office, by anybody who has the correct authorization identity. Records, whether hard copy or electronic, have to be readily accessible for inspection. In a medical laboratory, request forms and test reports are an example of such documentation. Records of any information or data, such as patients, results, minutes of meetings, quality control data, or the result of an audit, should be made on forms of an approved format and not on the backs of envelopes or the cuffs of laboratory coats!

A practical example of this hierarchy of documentation would be a statement in a quality policy requiring ''the use of examination procedures that will ensure the highest achievable quality of all tests performed''. A procedure produced as a result of such a policy statement might be

a procedure for measuring hemoglobin (Hb) A1c. In the quality manual, reference would be made to where a list of examination procedures can be found. The procedure might refer to working instructions for starting the HbA1c analyzer and for closing it down and these could be published separately and displayed near the analyzer for easy reference. If the analyzer is interfaced to a laboratory computer, then an example of a form would be a computer-generated worksheet to assist with checking in samples. Additionally, a computer file that holds the patient details and results should be regarded as a record. Such computer-held data need to be as easily accessible on demand as any paper record.

All the documents referred to in this hierarchy must be subject to control, as described next. The concept of documents can be extended to include specifications, calibration tables, charts, text books, posters, notices, memoranda, software, drawings, and so on. They can be on various media, whether hard copy or electronic, and they may be digital, analog, photographic, or written.

The preparation of required documentation might appear to be a daunting task for a medical laboratory, but if approached in a practical manner it provides the basis of effective quality management of the laboratory.

40.5.4 Document Control

Control of documents requires that they are approved for adequacy prior to issue, are reviewed and updated as required, are available at point of use, remain legible and uniquely identifiable, and that unintended use of obsolete documents is prevented. The purpose of regularly reviewing documents is to ensure that they remain fit for their intended purpose.

An inherent part of document control is a document register or master index of documentation. It is important to decide at an early stage whether the document register should be a manual paper record, a homemade spread sheet or database, or an off-the-shelf (albeit customizable) commercial product. This is perhaps the most important decision that any laboratory can make in building a QMS. The author has experience of all three approaches but would unequivocally opt for the off-the-shelf commercial product providing it was well tested and a reasonable price. One such product is Q-Pulse produced by Gael Quality, which in the author's experience is robust, reasonably priced, user friendly, and is being progressively developed. It can be downloaded and used on a 30-day trial basis from the Gael Quality (http://www.gaelquality.com).

40.5.5 Control of Records and Clinical Material

A major feature of all quality management systems is the need to control process and quality records and, in the case of medical laboratories, clinical material. ISO 9001 says ''records shall be established and maintained to provide evidence of conformity to requirements and of the effective running of the QMS''. It further stipulates that they shall remain legible, readily identifiable, and retrievable, and that a procedure be established to define the controls necessary for identification, storage, protection, retrieval, retention time, and disposition of records. ISO 9001, being a quality standard, simply refers to retention and storage of quality records, whereas the laboratory focused standards ISO 17025 and ISO 15189 refer to the retention and control of quality and process records, and in the case of ISO 15189, clinical material as well. Whether the requirement is for control of clinical material or records, there are three distinct issues to be considered:

1. Are the records being retained going to serve a useful purpose, for example to reconstruct an examination, or to audit corrective action?
2. What are the relevant retention times?
3. How should the material be kept?

40.6 REVIEW AND IMPROVEMENT

40.6.1 Evaluation and Continual Improvement

Inspection of the review and improvement sections of building a QMS (Fig. 40.3) indicate that any laboratory should be evaluating its activities constantly and seeking to continually maintain and improve quality. Evaluation and continual improvement could be regarded as synonymous with quality assurance, but it seems increasingly uncertain what is meant by the term ''quality assurance''. The difficulty seems to arise from the meanings of the words assure and ensure. To try to ensure the quality of something is ''to make sure or certain'' of its quality, whereas to assure, ''to give confidence to oneself or others'', seems a relatively impotent activity if it is seen from the point of view of the user clinician.

40.6.2 Internal Audit and External Assessment

ISO defines audit as a ''systematic independent and documented process for obtaining evidence and evaluating it objectively to determine the extent to which audit criteria are fulfilled''.

Three different types of audit can be distinguished. The first is an *internal audit* conducted by the laboratory itself (or occasionally on behalf of a laboratory by an outside auditor) on some aspect of laboratory activity such as the accuracy of transcription of data from a request form into the laboratory information system, or whether all members of staff have up-to-date job descriptions. The second is *external audit* (sometimes termed assessments), conducted by some person or bodies interested in the organization

such as a purchasing authority, or by external independent organizations such as CPA(UK) Ltd or a regulatory authority. A third type of audit, not shown in orthodox classifications, is *cooperative audit* conducted between the laboratory and another party for mutual benefit. Examples of cooperative audit are clinical audit or customer satisfaction surveys and benchmarking activities. Schemes for external quality assessment that are run on a primarily educational basis can in some senses be regarded as cooperative audit or equally well classified as external audit. Audits provide an important mechanism for the detection and investigation of nonconformity.

40.6.3 Nonconformities/Potential Nonconformities and Remedial/Corrective Action or Preventive Action

Nonconformities can be identified in two distinct ways. First, from a (reactive*) audit* that takes place in response to a problem occurring in the conduct of an established process and leading to the need for *remedial* and/or *corrective action* and thus contributing to the maintenance of quality or to *continual improvement*; and second, a proactive *audit* that identifies a *nonconformity* that again requires *remedial* or *corrective action.* Proactive audit or evaluation can also identify *potential nonconformities* that require *preventive* action, that is action that removes the root cause of the potential nonconformity.

An example of the first situation might be as follows. An inspection of the first results from a new batch of quality control material being introduced on an analyzer showed that results at all three levels for each analyte were approximately 20% lower than expected (*a nonconformity*). Investigation (*an audit*) revealed that although the freeze-dried material had been reconstituted with 5 ml of reconstituting fluid as per the documented procedure, the manufacturer had changed the reconstitution volume from 5 ml to 4 ml without sending out a notice to this effect. All vials wrongly reconstituted were immediately removed (*remedial action*). Following this incident all personnel involved had the matter drawn to their attention and the procedure was altered and an adverse incident report dispatched to the Medical Devices Agency UK, with a copy to the manufacturer (*corrective action*). These actions contribute to ensuring the quality of examinations (*continual improvement*).

An example of a proactive audit producing a *potential nonconformity* would be a review of the rotation of staff through work sections. If this system were found not to be working this could result in a *potential nonconformity* in that staff absences would have the potential to make sections of the laboratory vulnerable. *Preventive action* would be required to ensure that this did not happen.

40.6.4 Continual Improvement

Both corrective and preventive actions are at the core of continual improvement. Examples of approaches to continual improvement are shown in Fig. 40.5 as what the author has termed cycles of continual improvement. The intention

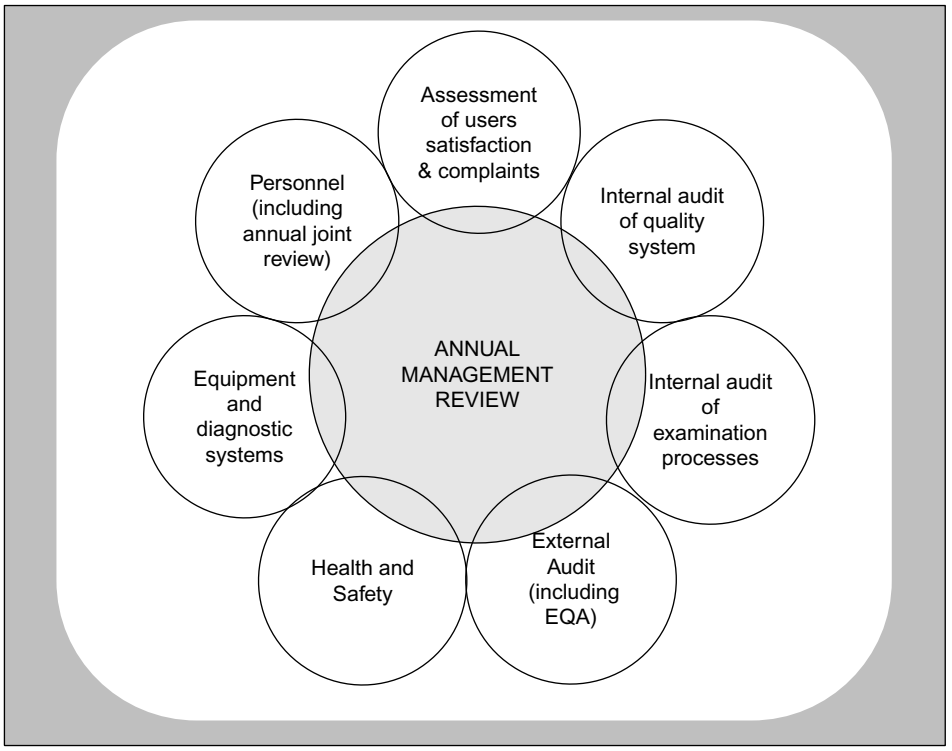


FIGURE 40.5 Cycles of continual improvement.

TABLE 40.2 Inputs to a management review.

• Assessment of user satisfaction and complaints
• Internal audit of quality management system
• Internal audit of examination processes
• External quality assessment reports
• Reports of assessments by outside bodies
• Status of preventive, corrective, and improvement actions
• Major changes in organization and management, resource (including staff), or process

of the diagram is to represent at the center the management review as the core focus of all continual improvement activity. The circles around the central circle represent individual circles of continual improvement focused on specific topics; for example, with *personnel*, the activity is the annual joint review of staff, with *internal audit of examination processes*, the vertical audit of examinations and with *equipment and diagnostic systems*, the procurement of *in vitro* diagnostic devices (IVDs).

An important question to answer at this point is when and how often should these activities take place. These circles of continual improvement should carry on throughout the year and most of the nonconformities discovered have to be resolved in a reasonably short time span for the process to be effective.

The nonconformities that are thrown up during the day-to-day activities of quality management are the ''grist to the mill'' (defined in common English usage as ''anything that can be turned to profit or advantage'') of continual improvement, or the cogs in the cycles of continual improvement.

However, during the course of a year, issues that require the formal setting of new objectives and detailed planning will be identified and these properly go forward as items for consideration at the (annual) *management review*. If the results from an *external audit* – Quality Assessment Scheme (or Proficiency Testing Scheme) – indicate a problem with an examination, it is no good waiting until the management review for its resolution, whereas the requirement for new service provision may have to wait for the capital purchase of the appropriate IVD or the recruitment of new staff.

40.6.5 Management Review

All the standards referred to in Table 40.1 have clauses entitled *management review(s)*. In ISO 9001, the responsibility for its conduct is defined as being with top management, in ISO 17025 with the laboratory's executive management, and in ISO 15189 with laboratory management. The time interval for this activity is not defined in ISO 9001, but notes in the other standards suggest that a typical period for conducting a management review is once every 12 months. It is important that the review should be seen as having inputs and outputs and the CPA(UK)Ltd, Standard A11 ''Management review'', has specific information regarding required inputs; Table 40.2 is based on these requirements.

The management review is a crucial part of a quality management system of a laboratory. It sets overall objectives for the following year and within the laboratory they are translated into objectives for the staff and thus into the staff joint reviews that identify the training needs of those staff. Continual improvement underpins the continuing provision of a quality service that aims to meet the needs and requirements of the user.

REFERENCES

Burnett, D., and Blair, C. (2001). Standards for the medical laboratory – harmonization and subsidiarity. *Clin. Chim. Acta,* 309, 137–145.

Burnett, D. (2002). *A Practical Guide to Accreditation in Laboratory Medicine*. (http://www.acb.org.uk). ACB Venture Publications, London.

Burnett, D., Blair, C., Haeney, M.R., Jeffcoate, S.L., Scott, K.W., and Williams, D.L. (2002). Clinical pathology accreditation: standards for the medical laboratory. *J. Clin. Pathol.* 55, 729–733.

ISO 9000:2000. Quality management systems – Fundamental and vocabulary.

National Pathology Advisory Council (2001). *Guidelines for quality systems in medical laboratories*. (http://www.health.gov.au/npaac). NPACC, Canberra, Australia.

National Pathology Advisory Council (2006). *Laboratory accreditation standards and guidelines for nucleic acid detection and analysis*. (http://www.health.gov.au/npaac). NPACC, Canberra, Australia.

National Pathology Advisory Council (2007). *Requirements for quality management in medical laboratories*. (http://www.health.gov.au/npaac). NPACC, Canberra, Australia.

This page intentionally left blank

Glossary

5′ nucleotidase: A structure-dependent single-stranded nuclease, used for genotyping and sequence copy-number determination reactions, generally related to the flap endonucleases, such as FEN-1.

Acrylamide: A material that is polymerized to make electrophoretic gels for separation of mixtures of macromolecules.

Allele drop-out: Failure of amplification of one allele when performing PCR on a single cell, which can be detected only in heterozygous cells.

Allele: A version of a gene or any DNA sequence that differs from the corresponding normal gene or DNA sequence by one or more base change(s), brought about by a mutation.

Allele-specific oligoprimer: A pair of oligonucleotide primers that contain basically the same sequence, but one part of the pair differs from the other in that one bears the normal allele sequence and the other bears the DNA change that specifies a mutated allele.

Amplicon: Fragment of DNA flanked by PCR primers, thus amplified by the PCR.

Amplification refractory mutation system (ARMS): A PCR-based mutation detection method, which allows amplification of a single allele, depending on the primer used.

Amplimer: One of a pair of primers needed to amplify DNA by the PCR.

Aneuploid cell: A cell, whose number of chromosomes differs from the normal chromosome number for the species by one or more chromosomes.

Aneuploidy: An irregular number of chromosomes or chromosomal regions that are not an exact multiple of the haploid chromosome number for a cell or organism, which can be higher or lower compared to the wild-type situation.

Annealing: Spontaneous alignment of two single DNA strands, resulting in the formation of a double-stranded DNA molecule.

Autoradiography: A process in which a pattern is formed onto a film, which corresponds to the location of radioactive materials incorporated into cell structures or DNA molecules, when these cell structures or DNA molecules are placed next to a film.

Autosome: Any chromosome other than a sex chromosome.

Base pair (bp): A pair of complementary nucleotide bases in a duplex DNA or RNA molecule.

Bioactive: A substance in food that confers specific health benefits. Bioactives may be recognized macro-nutrients (e.g. fiber) or micro-nutrients (e.g. vitamin C), but, increasingly, bioactive substances are being identified with no previously known role in nutrition.

Biomarker: A substance found in the body (ideally in an easily sampled tissue) or characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

Bisulfite treatment: A chemical reaction involving hydrolytic deamination of non-methylated cytosines to uracils – while methylated cytosines are resistant to conversion – and resulting in uracils being replaced by their DNA analog thymine following PCR amplification, hence allowing assessment of the methylation status at a given position, manifested in the ratio C (former methylated cytosine) to T (former non-methylated cytosine) that can be analyzed as a virtual C/T polymorphism in the bisulfite-treated DNA.

Blunt ends: Ends of duplex DNA molecules with no overhangs due to restriction or removal by single-strand specific nucleases.

Broad spectrum PCR: Pivotal PCR tests that aim at amplification of DNA sequences that are universally conserved among a variety of bacterial species for the characterization of microbial communities or the identification of novel microbial species.

Cap: A 7-methylguanosine molecule, which is added 5′ to the pre-mRNA.

CELI nuclease: A mismatch-specific nuclease isolated from celery, representative of a family of plant bifunctional DNase/RNase, induced during senescence.

Cell separation and concentration methods: Methods allowing the arrest of selective cell subpopulations from the complex sample matrix through the use of electric, magnetic, or acoustic fields.

Centromere: The constricted region of the chromosome where the sister chromatids are joined and where the spindle fibers are associated for chromosome movement during mitosis and meiosis.

Charge-coupled device (CCD) camera: A camera whose silicon chip is divided into light-sensitive pixels that can be counted upon photon exposure.

Chlamydia trachomatis: Intracellular bacterial species best known as a sexually transmitted agent, which can cause infertility in women due to pelvic inflammatory disease, one of the most prevalent sexually transmitted diseases.

Chorionic villus sampling (CVS): A placental sampling procedure for obtaining fetal tissue for chromosome and DNA analysis to assist in prenatal diagnosis of genetic disorders.

Cis-acting element: A DNA (or RNA) sequence, on which sequence-specific DNA (or RNA) binding proteins are bound.

Cleavage stage embryo: Stage in between the fertilized egg and the morula stage (day 1–day 3 after fertilization). The divisions of the cells are characterized by the lack of the G-phase, so that the cells become smaller and the total volume of the embryo does not increase.

Cloning vector: The DNA vehicle (plasmid, cosmid, or phage chromosome) used to carry the cloned DNA fragment.

COBAS Amplicor: Completely automated system suited for the detection of bacterial pathogens, such as *C. trachomatis*, by PCR.

Codon: A specific DNA or corresponding RNA sequence of three base pairs, which encode for a particular amino acid, start, or termination signal.

Comparative genomic hybridization (CGH): Molecular cytogenetic whole genome scanning technique based on the cohybridization of labeled test- and control-DNA to normal metaphase spreads, followed by the computer-assisted quantification of fluorescence intensities over the entire length of each chromosome, resulting in a "copy-number karyotype" reflecting chromosomal imbalances.

Competitive oligoprimers: A pair of allele-specific oligonucleotides that are both used in a polymerase chain reaction so that, by the resulting competition between them for annealing to the substrate, amplification of only (or mainly) the "correct" (that is, the fully matching) sequence is promoted by each competitive oligoprimer.

Composite exoninc regulatory elements of splicing (CERES): Exonic sequences with overlapping splicing enhancer and silencer function.

Copy-number variation (CNV): A copy-number change involving a DNA fragment that is approximately 1 kb or larger, excluding changes involving the insertion or deletion of transposable elements.

Core database: A collection of published mutations from all genomic loci.

Cystic fibrosis: A potentially lethal autosomal recessive genetic disorder, affecting the secretory glands.

Degenerate genetic code: A genetic code in which some amino acids are encoded by more than one codon each.

Deletion: Removal of a DNA fragment, from a single nucleotide to a chromosomal segment or even an entire chromosome.

Denaturation: The separation of two DNA strands in a DNA double helix, using heat, chemicals, or extremes of pH.

Deoxyribonucleotide (dNTP): A, C, T, or G triphosphates, used as building blocks by DNA polymerase to make a new DNA strand copied from a template.

Diploid: A cell or organism with two complete sets of homologous chromosomes.

DNA cloning: Insertion of a DNA fragment into a vector molecule, such as a plasmid, cosmid, and so on, which is then replicated to generate many copies.

DNA ligases: A group of enzymes that catalyze the formation of a phosphodiester linkage between the 3′- and 5′-ends of two DNA fragments.

DNA polymerases: Several enzymes that can synthesize new DNA strands from a DNA template, by adding nucleotides to the growing DNA chain in a 5′ to 3′ direction.

DNA sequencing: A technique for determining the array of nucleotides in a fragment of DNA.

Electrochemical detection: Method to quantitate an analyte by measuring the electrical current or voltage generated as a result of a chemical reaction involving the analyte.

Electrophoresis: Migration of charged molecules in an electric field and the separation of the components of a mixture of proteins, DNA, or RNA within an agarose or polyacrylamide gel.

Endonuclease: An enzyme that catalyses the hydrolysis (cleavage) of phosphodiester bonds between nucleotides in a sequence of DNA or RNA.

Engraftment: Time-point when transplanted hematopoietic stem cells grow and develop a sufficient number of hematopietic cells (leucocytes, erythrocytes, thrombocytes).

Enzyme-linked immunosorbent assay (ELISA): A variety of assays in which an enzyme is attached to the antigen (substance of interest) or to the antibody, and the enzyme converts a substrate into a colored or fluorescent product to demonstrate that the antigen-antibody binding has occurred in the reaction.

Epigenetics: The study of mitotically (and in some cases meiotically) heritable changes of a phenotype, such as the gene expression of specific cell types that are not due to changes in the genetic code.

Ethidium bromide: A molecule that intercalates into the DNA double helix, which mostly serves to visualize DNA fragments under ultraviolet (UV) light transillumination.

Exon: The coding sequence of a gene, which is transcribed into mRNA and translated into protein.

Expressed sequence tag (EST): A sequence-tagged site, derived from a cDNA clone, used to identify genes in genomic analysis.

Expression profiling: Identification of transcriptional differences between two different RNA samples, usually providing a unique pattern, also known as expression signature, for certain disease types, such as tumors.

Familial adenomatous polyposis (FAP): An autosomal dominant human disease, caused by germ-line mutations in the adenomatous polyposis coli (APC) gene.

Fluorescence in situ hybridization (FISH): Molecular cytogenetic approach based on the duplex formation of complementary, single-stranded, and fluorescence-labeled nucleic acid probes with target material (interphase cell; metaphase spread) allowing the exploration of the presence, number, and distinct location of genetic material *in situ.*

Fluorescence resonance energy transfer (FRET): A detection method, based on a distance-dependent interaction between

the electronic excited states of two dye molecules, in which the energy of an excited fluorophore (the donor) is passed to a nearby fluorophore (the acceptor) through resonance, then released as a photon that is detected by a fluorimeter.

Forensic science: The application of science in the investigation of legal matters.

Frameshift mutation: The insertion or deletion of nucleotide(s), resulting in a disruption of the translational reading frame.

Gene therapy: The insertion of genetically corrected or wild-type genes into cells especially to replace defective genes in the treatment of genetic disorders or to provide a specialized disease-fighting function.

Gene: The fundamental physical and functional unit of heredity, which carries information from one generation to the next, consisting of the coding region and a regulatory sequence that make its transcription possible.

Genetic locus: A gene or gene family, or a non-coding DNA sequence that resides at a specific position of a specific chromosome of an organism.

Genetic test: A test that is capable of revealing the zygosity of the specific allelic constitution at a genetic locus in the diploid cells of an organism.

Genetically modified organism (GMO): An organism containing an additional trait encoded by an introduced gene or genes, which generally produces a protein that confers the trait of interest.

Genotype: The genetic composition of the entire cell or, more commonly, of a set of genes.

Gradient: A gradual change in a quantitative property over a specific distance.

Haploid: A state in which only one chromosome or chromosome set resides in a cell, such as the Y-chromosome in human males.

Haplotype: Group of polymorphic changes closely linked on one chromosome that are inherited as a unit.

Hemizygous gene: A gene that is present in only one copy in a diploid organism, such as the X-linked genes in a male organism.

Hemoglobin (Hb): The oxygen-transporting protein in the blood of most animals.

Hereditary hemochromatosis: A human autosomal recessive disease, resulting in iron overload and irreversible tissue damage, due to increased iron absorption.

Hereditary non-polyposis colorectal cancer (HNPCC): One of the most common predispositions to cancer.

Heteroduplex: A double-stranded DNA molecule that contains one or more nucleotides that are not in the correct base pairing conformation, formed by annealing of single strands from different sources, which can show abnormalities such as loops or buckles.

Heterozygosity: A measure of the genetic variation in a population, indicated as the frequency of heterozygotes for a specific gene.

Heterozygote: An individual with a heterozygous gene pair.

High-throughput PCR: A PCR system that can be applied simultaneously in a large number of samples (large-scale) for the amplification of a normal or mutant genetic locus.

Homologous chromosomes: Chromosomes that pair with each other at meiosis.

Homozygote: An individual with a homozygous gene pair.

Huntington disease: A lethal autosomal dominant human disease, characterized by nerve degeneration, with late-age onset.

Hybrid minigene: A simplified version of a gene that contains the sequences to be studied (i.e. exon and flanking intron sequences) placed in a heterologous context.

Hybridization: The annealing of complementary, single-stranded nucleic acids (DNA, RNA).

Insertional translocation: The insertion of a segment from one chromosome into another non-homologous one.

Interphase: The cell cycle stage between nuclear divisions, consisting of the G1, S, and G2 phases, when chromosomes are extended and functionally active.

Intracytoplasmic sperm injection: A procedure that involves sperm injection into the oocyte's cytoplasm to obtain fertilization, developed to treat severe male infertility and used to prevent contamination during preimplantation genetic diagnosis and failure of fertilization.

Intron or intervening sequence: A DNA fragment within a gene, which is initially transcribed, but not included in the mRNA, due to its removal by splicing.

***In situ* proximity ligation assay (*in situ* PLA)**: The proximity ligation assay adapted for localized detection of proteins, protein complexes, and post-translational modifications of proteins in fixed cells and tissues.

Karyotype: The entire chromosome complement of an individual or cell, as can be seen during mitotic metaphase.

Kilobase (kb): 1,000 nucleotides.

Linkage: The association of genes on the same chromosome, tending due to their close physical proximity to be inherited together.

Linkage disequilibrium (LD): The non-random associations of alleles at two or more loci, meaning alleles that are located close to each other on the same chromosome have the tendency to occur and be inherited together.

Locus-specific database: A collection of all published and unpublished mutations from a specific gene.

Loss of heterozygosity (LOH): Loss of one allele at a chromosomal locus, which may imply the presence of a tumor suppressor gene at that site.

Melting curve: Originally used to denote the hypochromic shift between double-stranded and single-stranded nucleic acids, caused by greater electron delocalization in the double-strand state accounted for by stacking forces, a surrogate

measurement now commonly obtained being a thermal profile of SybrGreen (A-T intercalation) fluorescence differences between the double- and single-stranded states.

Melting temperature (T_m): The temperature in which a double-stranded DNA fragment is denatured.

Messenger RNA (mRNA): A mature RNA molecule, resulting from gene transcription by RNA polymerase, which specifies the order of amino acids during its translation to protein.

Metabolomics: The study of all the small molecules produced in a cell or organism as a result of metabolism.

Metaphase: The intermediate stage of nuclear division in the cell cycle in which the highly condensed chromosomes align along the equatorial plane between the two poles of the dividing cell, a feature that is exploited for the identification of chromosome aberrations.

Methylation-specific PCR (MSP): A technique allowing the amplification of virtually any CpG sites after bisulfite treatment.

Microarrays: An orderly arrangement of usually thousands of defined DNA molecules, immobilized onto a solid surface, such as glass or membrane.

Microfluidics: Networks of interconnected channels and chambers equipped with isolation valves and fluid pumps and fabricated in polymer, glass, silicon, or ceramic substrates, developed for miniaturized chemical analysis and field-deployable sensor applications.

Micro-PCR: Performing DNA amplification in miniature microfluidic chambers (chips) using polymerase chain reaction leading to the reduction of the sample volume and acceleration of thermal cycling, and integration of the amplification stage with subsequent detection platforms.

Microsatellite instability (MSI): A frequent, if not obligatory, surrogate marker of underlying functional inactivation of one of the human DNA mismatch repair genes, characterized be length alterations of oligonucleotide repeat sequences that occur somatically in human tumors.

Minimal residual disease (MRD): Remnant of a tumor or cancer after primary, potentially curative therapy, which can be detected by PCR with a sensitivity of one in 104 to 105 cells.

Minisatellite DNA: Repetitive DNA sequence, based on a repeated sequence core, used for DNA fingerprinting.

Mismatch: A position within a double-stranded DNA molecule, in which the two bases found opposite each other do not comply with the DNA pairing rule of AG and C-T.

Missense mutation: A mutation that alters a codon, resulting in a different amino acid.

Mitochondrial DNA: A closed circular, extranuclear chromosome, approximately 16,569 base pairs in size in humans, residing in the mitochondrion, and containing genes for two ribosomal RNAs, 22 transfer RNAs, and 13 proteins, predominately for oxidative phosphorylation.

Molecular cytogenetics: Application of molecular biology techniques to cytogenetic preparations as metaphase spreads and interphase nuclei for the identification of chromosome abnormalities and RNA expression.

Molecular diagnostics: The identification of a genetic disorder at the molecular (DNA) level, using molecular biology techniques.

Molecular genetics: The study of the molecular processes, which govern gene structure and function.

Mosaicism: The presence of two or more genetically different cell populations in one organism.

Multilocus sequence typing (MLST): A method of bacterial genetic polymorphism identification by comparative sequence analysis of several housekeeping genes, in which point mutations can be used to calculate precise genetic distances between isolates of any given bacterial species.

Multiplex amplification: Coamplification of multiple target sequences in one single amplification reaction.

Multiplex PCR: A PCR containing sets of forward and reverse amplification primers for more than one allele of a genetic locus or for different genetic loci.

Mutagenesis: The process that produces a gene or a chromosome set differing from the wild type.

Mutant allele: An allele differing from the wild-type one.

Mutation: A transmittable change in the DNA sequence of a gene or anywhere in the genome, with relation to a reference sequence, with or without consequences for the phenotype of the organism.

MutHLS nuclease: An evolutionally conserved enzyme complex of *Escherichia coli* used in DNA replication linked mismatch repair.

Nitrocellulose filter: A type of filter used to attach DNA fragments for hybridization.

Nonsense mutation: A mutation that produces a stop codon, resulting in the premature termination of the protein chain.

Northern blot: Transfer of electrophoretically separated RNA molecules from a gel onto a filter, which is then immersed in a solution containing a labeled probe that will bind to the RNA of interest.

Nucleotide: The basic building block of nucleic acids, composed of a nitrogen base, a sugar, and a phosphate group, joined in pairs by hydrogen bonds.

Nutrigenomics: The study of the interaction between an individual's genetic makeup and their diet, often achieved using genomics technologies.

Oligonucleotide ligation assay (OLA): An assay in which genetic variants are analyzed by utilizing the covalent joining of oligonucleotide probes by DNA ligase and the detection of ligation by specific labels attached to the probes.

Oligonucleotide: Short linear sequence of nucleotides, often used as probes to hybridize to the target DNA in genetic tests or as primers to promote DNA synthesis in a polymerase chain reaction.

On-chip sample preparation: An integrated sample processing within microfluidic channels and chambers, consisting of cell concentration, lysing, nucleic acid purification, and amplification, leading to the process automation and assay miniaturization.

Open reading frame (ORF): A DNA segment, beginning with a start codon and ending with a stop codon, which is presumed to be the coding sequence of a gene.

Perpendicular electrophoresis: The first step of the optimization process of a TGGE or DGGE analysis, in order to verify the reversible melting behavior of the DNA fragment and to determine its T_m under the experimental conditions.

Pharmacogenetics: The assessment of clinical efficacy, safety, and tolerability profile of a drug in groups of individuals that differ with regard to certain DNA-encoded characteristics, which if indeed associated with a differential response or phenotype may allow prediction of individual drug response.

Pharmacogenomics: The evaluation of the differential effects of a number of drugs or chemical compounds in the process of drug discovery, with regard to inducing or suppressing the expression of transcription of genes in an experimental setting.

Phenotype: The observed outward manifestations of a specific genotype.

Phenylketonuria (PKU): An autosomal recessive human metabolic disease, caused by a mutation in a gene encoding a phenylalanine-processing enzyme, which leads to mental retardation if not treated.

Philadelphia chromosome: A translocation between the long arms of chromosomes 9 and 22, often found in the white blood cells of patients with chronic myeloid leukemia.

Phosphodiester bond: The bond between a sugar group and a phosphate group, resulting in the sugar phosphate backbone of DNA.

Photolithography: A method for etching silicon wafers using a LASER (light amplification through stimulated emission of radiation) in conjunction with a mask to perform feature addressing.

Physical mapping: The identification of the positions of cloned genomic fragments.

Plasmid: Autonomously replicating extrachromosomal DNA molecule, which serves as DNA cloning vehicles.

Polar bodies: By-products of the female meiosis. The first is extruded at maturation of the oocyte, leaving the oocyte haploid, and the second is extruded after fertilization and contains chromatids.

Polymerase chain reaction (PCR): A method for the exponential amplification of a specific DNA fragment from a template by multiple rounds of DNA synthesis.

Polymerase Chain Reaction-Restriction Fragments Length Polymorphism (PCR-RFLP): A two-primer polymerase chain reaction followed by digestion with a class II restriction enzyme (one lacking restriction modification activity), allowing for variant visualization through sizing on an electrophoretic gel.

Polymorphism: A gene or any DNA sequence that appears with more than one allele in a population of an organism, namely the normal allele, or reference sequence, and one or more mutated alleles, provided that the least frequent allele is found with an incidence of at least 1% in this population.

Primer extension: Method of using a short deoxynucleotide primer precisely complementary to genomic or mRNA sequence, in the presence of an enzyme with $3' > 5'$ polymerase activity, cofactors and deoxynucleoside triphosphates, to extend the oligonucleotide to a $5'$ template terminus, in which successful polymerization can be used to discern sequence variants from each other, since extension generally fails if the primer does not precisely match the $3'$ base in the initial template.

Primer: A short single-stranded RNA or DNA that is used as the starting point for chain elongation by the DNA polymerase, when bound to a single-stranded template.

Probe: Defined nucleic acid fragments that can be used to identify specific DNA molecules bearing the complementary sequence, usually through autoradiography.

Promoter: A regulator region in short distance from the $5'$ end of a gene that acts as the binding site for RNA polymerase and other transcription factors.

Proteome: The complete set of proteins in an organism.

Proteomics: The analysis of the protein complement present in a cell, organ, or organism at any given time.

Proto-oncogene: A normal gene that can become an oncogene (a potentially cancer-inducing gene) due to mutations or increased expression.

Proximity ligation assay (PLA): A technique for quantifying proteins and protein complexes by converting a multiple target recognition event, for instance by two antibodies, into a DNA-based signal that can be read out using, e.g., quantitative PCR.

Pulsed-field gel electrophoresis (PFGE): An electrophoretic technique in which the gel is subjected to electrical fields alternating between different angles, allowing efficient separation of very large DNA fragments through the gel.

Pyrosequencing: A real-time sequencing method that uses light release as the detection signal for nucleotide incorporation into a target DNA strand, allowing for the direct assessment of DNA sequence.

Quality control management: A process involving the identification of the requirements posed by differing clinical situations and ensuring that the inherent characteristics required for a laboratory measurement or examination are fulfilled.

Real-time PCR: A PCR assay in which the PCR product is measured continuously throughout the amplification process, for example the TaqMan assay.

Recessive allele: An allele whose phenotypic effect is expressed only in the homozygous state.

Reciprocal translocation: The situation when parts of chromosomes are exchanged.

Reference material: A standardized material to compare the quality and performance of testing regimen.

Rehybridization: The anti-parallel realignment of two complementary DNA single strands to form the double helix.

Resolvase: An enzyme that hydrolyzes phosphodiester bonds in a DNA strand crossover complex undergoing DNA recombination to separate the DNA strands.

Restriction endonucleases: A variety of enzymes, extensively used in genetic engineering, that recognize specific target DNA sequences and hydrolyzes the phosphodiester bond at these points.

Restriction fragment length polymorphism (RFLP) analysis: Detection of different sizes or numbers of restriction fragments, often as a result of presence or absence of restriction sites, which can be used as markers in chromosome mapping.

Retroviral integration site analysis: Analysis of the junction between the retroviral vector sequence and the genomic sequence to identify the retrovirus integration site in retroviral-mediated gene transfer.

Retroviral vector: An artificial DNA construct derived from a retrovirus, used to insert sequences into an organism's chromosomes.

Retrovirus: An RNA virus that replicates by first being converted into double-stranded DNA.

Reverse dot-blots: Allele-specific oligonucleotide hybridization matricies involving covalent or hydrogen bonding attachment of interrogator sequences to a solid support, such as a nylon- or plastic-backed membrane. In general, muliplexed PCR products are incubated with the prepared matrix after denaturation, permitted to hybridize, and visualized using radioactive, chemiluminescent, or chromogenic techniques.

Reverse transcription-PCR (RT-PCR): RNA amplification by PCR, following copying of the RNA to cDNA by reverse transcription.

RNA interference (RNAi): A general mechanism for inhibiting gene expression, by small antisense RNA molecules, also known as small interfering RNAs (siRNA).

Robertsonian translocation: The situation when two acrocentric chromosomes translocate to become one large chromosome.

Rolling circle amplification (RCA): A method of amplifying a circular DNA template into tandemly linked copies of the DNA circle, using a polymerase such as the phi29-polymerase.

Sampling: A statistical method to collect samples containing the desired organisms.

Satellite DNA: Any type of highly repetitive DNA.

Short tandem repeat (STR) loci: Also known as microsatellites, STR loci are highly repetitive DNA elements of individual repeat motifs, each motif ranging from two to seven base pairs in length, whose difference among individuals is the basis for the population variation of these loci.

Sickle cell anemia: A potentially lethal autosomal recessive human inherited disorder, caused by a mutation in the gene encoding for the oxygen transporting hemoglobin molecule, resulting in sickle-shaped red blood cells.

Single nucleotide polymorphism (SNP): Any polymorphic variation at a single base pair position between individuals of the same species, including restriction fragment length polymorphisms and transitions or transversions that reach a frequency of greater than 1% in any given population.

Solid phase: Attachment of DNA to solid support, such as silica beads, which remains intact during reactions and purification steps.

Southern blot: Transfer of electrophoretically separated DNA fragments from a gel to a filter, which is then immersed in a solution containing a labeled probe that will bind to a fragment of interest.

Splicing enhancer and silencer: Short RNA nucleotide sequences, generally six to eight base pairs in length, located in the exon or in the intron that enhance or inhibit splicing, respectively.

Splicing: The reaction that removes introns and joins exons together, resulting in the mature mRNA.

SR proteins: Regulatory splicing factors that contain a serine-arginine-rich region.

Staphylococcus aureus: A bacterial species, known for its capacity to colonize the human nostrils, that in case of immune compromisation is the best-known opportunistic pathogen, also well known for its capacity to collect a variety of antimicrobial resistance traits. Methicillin resistant *S. aureus* or MRSA is the best-known hospital pathogen worldwide.

Stop codons: The terminating signals for translation of an mRNA into protein.

Telomere: Specialized DNA structure [(TTAGGG)n], capping the end of a chromosome involved in replication and stability of linear DNA molecules.

Tool Command Language (Tcl) and Toolkit (Tk): An interpreted software script syntax commonly used to instruct automata movements; Tk for graphical user interfaces provides a complementary graphical user interface to widgets being instructed.

Transcription: A process by which a DNA template is copied to RNA by RNA polymerase.

Transcriptome: A set of all messenger RNA (mRNA) molecules, or transcripts, produced in one or a population of cells, reflecting the genes that are being actively expressed at any given time.

Transduction: The process whereby foreign DNA is introduced into another cell via a viral vector.

Transgene: A gene that has been transferred by any of a number of genetic engineering techniques from one organism to another.

Transition: A type of nucleotide substitution involving the replacement of a purine with another purine or of a pyrimidine with another pyrimidine, for example GÆA or CÆT.

Translocation: Relocation and interchange of a chromosomal segment after breakage and attachment to a different nonhomologous chromosome in the genome.

Transversion: A type of nucleotide substitution involving the replacement of a purine with a pyrimidine or vice versa, for example, GÆT or CÆA.

Tumor suppressor gene: A gene encoding for a protein that suppresses tumor formation, speculated to function as negative regulators of cell proliferation.

Variable number tandem repeats (VNTR): Variations in the number of tandem repeats of DNA sequences found at specific loci in different populations, often used as markers.

Vector FISH: Chromosomal mapping of proviral sequences by FISH, using a retroviral vector plasmid DNA as a probe.

Western blot: Transfer of protein molecules, separated by electrophoresis, from a gel to a filter, which can be probed with a labeled antibody to detect a specific protein.

Wild type: The genotype or phenotype that is found in nature or in the standard laboratory stock for a given organism.

Y-chromosome: The smallest of the human chromosomes, approximately 60 million base pairs including the gene conveying the male gender, parentally transmitted from father to son.

Zona pellucida: Transparent protein structure surrounding the cleavage stage embryo.

Zygosity: The state of genetic identity, that is, normal versus mutant, of the two alleles of a gene or DNA sequence, in the somatic cells of a diploid organism, specified as either normal, heterozygous, or homozygous mutant.

Zygote: The unique diploid cell, formed by the fusion of an oocyte and a sperm cell, that will divide mitotically to create a differentiated diploid organism.

This page intentionally left blank

Index

D

I

L

M

N

O

P

Q

R

S

T