Methods in Molecular Biology 1619

Springer Protocols

David W. Greening
Richard J. Simpson *Editors*

Serum/Plasma Proteomics

Methods and Protocols

Second Edition

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Serum/Plasma Proteomics

Methods and Protocols

Second Edition

Edited by

David W. Greening

Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC, Australia

Richard J. Simpson

Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC, Australia

Humana Press

Editors
David W. Greening
Department of Biochemistry and Genetics
La Trobe Institute for Molecular Science
La Trobe University
Melbourne, VIC, Australia

Richard J. Simpson
Department of Biochemistry and Genetics
La Trobe Institute for Molecular Science
La Trobe University
Melbourne, VIC, Australia

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-7056-8 ISBN 978-1-4939-7057-5 (eBook)
DOI 10.1007/978-1-4939-7057-5

Library of Congress Control Number: 2017944282

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

Human blood is one of the most informative and important proteomes from a clinical perspective. Blood, plasma, and serum are the predominant samples used for diagnostic analyses in clinical practice and are available in biobanks from thousands of clinical studies. The fact that blood constituents, primarily proteins, reflect diverse physiological, pathological, and pharmacological states makes them of great clinical significance—i.e., blood can be considered the "window of physiology and disease". The ease with which blood (especially its plasma/serum components) can be sampled in a noninvasive manner makes it a logical choice for diagnostic screening applications. Characterization of plasma and serum proteins (both in qualitative and quantitative terms) should provide a foundation for the discovery of candidate markers for disease diagnosis and development of new therapeutics. Mass spectrometry (MS)-based proteomics is a technology capable of discovering biomarkers in blood. However, MS-based blood proteomics is extremely challenging for a number of reasons, most prominently the significant large dynamic range of protein abundances. Over the past several years, we have witnessed the advent of more powerful proteomics technologies and fractionation strategies that allow identification and accurate, multiplexed, quantitative monitoring of a diverse range of components from blood proteome. Improvements in discovery-based proteomics are focused toward increased multiplex quantitation and overcoming sample complexity through fractionation, increased mass ranges through improvements in labeling efficiency, and reduced user costs. Advances in targeted proteomics are being directed toward biomedical research and clinical applications, such as large-scale quantification, improvements in method development, throughput, data processing and analysis, and the utilization of fast-scanning high-resolution accurate-mass instruments to analyze low-abundance proteins in complex biological matrices. Further developments in informatics analyses, software developments, and computational tools are providing insights into large data sets and open-source data along with large-scale application of bioinformatics. Collectively, these improvements have, in part, fueled the quest for the discovery and monitoring of novel blood-based biomarkers and their modifications during disease.

This updated volume describes recent developments in blood proteomics—providing key insights and recommendations into processing and handling strategies, fractionation, posttranslational modification analyses, antibody-based approaches, and key developments in discovery and targeted proteomics toward clinical assay development. Part I of this volume comprises three chapters devoted to blood collection, handling and processing, and storage. Part II relates to fractionation strategies for in-depth blood proteome analysis and posttranslational modifications and includes seven chapters. Part III relates to proteome analyses of blood cell components, including platelets and red blood cells, circulating extracellular vesicles/exosomes, and related biofluids, while Part IV relates to detailed protocols for performing discovery and targeted antibody-based quantitative assays. Part V provides key insights into mass spectrometry-based discovery (global) and targeted (multiple/selected reaction monitoring) approaches. Part VI relates to studies focusing on key proteomics-based developments in biomarker discovery utilizing blood products.

To aid blood proteome researchers, we also include current standard operating procedures (SOPs) for plasma and serum collection for the purpose of clinical research, measured concentrations of many plasma proteins from quantitative assays, and reference ranges for blood tests. We also include a detailed overview of reference ranges for current blood tests. Such reference ranges for blood tests are studied within the field of clinical chemistry. These aids are appended at the end of the volume.

An Updated Serum/Plasma Proteomics is a comprehensive resource of 35 chapters and protocols for areas—pre-analytical through to analytical—of plasma and serum proteomics to assess human health and disease. This updated volume, contributed by leading experts in the field, complements the initial volume *Serum/Plasma Proteomics* and provides a valuable foundation for the development and application of blood-based proteomics.

David W. Greening
Richard J. Simpson

Contents

Contributors

Sunil S. Adav • *School of Biological Sciences, Nanyang Technological University, Singapore, Singapore*

Nuzhat Ahmed • *Fiona Elsey Cancer Research Institute, Ballarat, VIC, Australia; Federation University, Ballarat, VIC, Australia*

Gloria Alvarez-Llamas • *Department of Immunology, IIS-Fundacion Jimenez Diaz, Madrid, Spain*

Sandra I. Anjo • *CNC.IBILI, University of Coimbra, Coimbra, Portugal; CNC—Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal; Faculty of Sciences and Technology, University of Coimbra, Coimbra, Portugal*

Burcu Ayoglu • *Affinity Proteomics, SciLifeLab, School of Biotechnology, KTH—Royal Institute of Technology, Solna, Sweden*

Montserrat Baldan-Martin • *Department of Vascular Physiopathology, Hospital Nacional de Paraplejicos (HNP), SESCAM, Toledo, Spain*

Maria G. Barderas • *Department of Vascular Physiopathology, Hospital Nacional de Paraplejicos (HNP), SESCAM, Toledo, Spain*

Kurt T. Barnhart • *Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA, USA*

Lynn A. Beer • *The Wistar Institute, Philadelphia, PA, USA*

Larissa Belov • *School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia*

Tue Bjerg Bennike • *Department of Pathology, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA*

O. Giles Best • *School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia; Kolling Institute of Medical Research, Royal North Shore Hospital, St. Leonards, NSW, Australia*

Santosh D. Bhosale • *Turku Centre for Biotechnology, University of Turku, Turku, Finland*

Elin Birgersson • *Affinity Proteomics, SciLifeLab, School of Biotechnology, KTH—Royal Institute of Technology, Solna, Sweden*

Edmond J. Breen • *Australian Proteome Analysis Facility (APAF), Macquarie University, Sydney, NSW, Australia*

Qichen Cao • *National Center for Protein Sciences Beijing, State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing, China; Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, China*

Ana Sofia Carvalho • *Computational and Experimental Biology Group, CEDOC—Chronic Diseases Research Center, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisboa, Portugal*

Uta Ceglarek • *Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University Hospital Leipzig, Leipzig, Germany; LIFE-Leipzig Research Center for Civilization Diseases, Leipzig University, Leipzig, Germany*

Emily I. Chen • *Herbert Irving Comprehensive Cancer Center, Proteomics Shared Resource, Columbia University Medical Center, New York, NY, USA; Department of Pharmacology, Columbia University Medical Center, New York, NY, USA*
Dong Huey Cheon • *Center for Theragnosis, Biomedical Research Institute, Korea Institute of Science and Technology, Seoul, Republic of Korea; Interdisciplinary program of Integrated Biotechnology, Sogang University, Seoul, Republic of Korea*
Katelyn M. Chessler • *Department of Biochemistry and Molecular Genetics, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA*
Nicolas Christinat • *Lipid Biology, Nestlé Institute of Health Sciences, Lausanne, Switzerland*
Richard I. Christopherson • *School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia*
Ornella Cominetti • *Nestlé Institute of Health Sciences SA, Lausanne, Switzerland*
John Corthésy • *Nestlé Institute of Health Sciences SA, Lausanne, Switzerland*
Fernando de la Cuesta • *Department of Vascular Physiopathology, Hospital Nacional de Paraplejicos (HNP), SESCAM, Toledo, Spain*
Angelo D'Alessandro • *Department of Biochemistry and Molecular Genetics, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA*
Loïc Dayon • *Nestlé Institute of Health Sciences SA, Lausanne, Switzerland*
Julia Dittrich • *Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University Hospital Leipzig, Leipzig, Germany; LIFE-Leipzig Research Center for Civilization Diseases, Leipzig University, Leipzig, Germany*
Tea Dodig-Crnković • *Affinity Proteomics, School of Biotechnology, SciLifeLab, KTH—Royal Institute of Technology, Stockholm, Sweden*
Monika Dzieciatkowska • *Department of Biochemistry and Molecular Genetics, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA*
Antonio Núñez Galindo • *Nestlé Institute of Health Sciences SA, Lausanne, Switzerland*
Tamar Geiger • *Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel*
Henrique Girao • *Institute of Biomedical Imaging and Life Sciences (IBILI), Faculty of Medicine, University of Coimbra, Coimbra, Portugal; CNC.IBILI, University of Coimbra, Coimbra, Portugal*
David R. Goodlett • *Turku Centre for Biotechnology, University of Turku, Turku, Finland; Department of Pharmaceutical Science, University of Maryland, Baltimore, MD, USA*
Viviana Greco • *Proteomics and Metabonomics Unit, Fondazione Santa Lucia, IRCCS, Rome, Italy*
David W. Greening • *Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Bundoora, Melbourne, VIC, Australia*
Hongbo Gu • *Proteomic Service Group, Cell Signaling Technology, Danvers, MA, USA*
Shabarni Gupta • *Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, India*
Susannah Hallal • *School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia*
Kirk C. Hansen • *Department of Biochemistry and Molecular Genetics, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA*
Michal Harel • *Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel*

Cecilia Hellström • *Affinity Proteomics, School of Biotechnology, SciLifeLab, KTH—Royal Institute of Technology, Stockholm, Sweden*
Mun-Gwan Hong • *Affinity Proteomics, School of Biotechnology, SciLifeLab, KTH—Royal Institute of Technology, Stockholm, Sweden*
Mashanipalya G. Jagadeeshaprasad • *Proteomics Facility, Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune, India; Academy of Scientific and Innovative Research (AcSIR), New Delhi, India*
Xiaoying Jia • *Proteomic Service Group, Cell Signaling Technology, Danvers, MA, USA*
David Juncker • *Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, QC, Canada; Department of Biomedical Engineering, McGill University, Montreal, QC, Canada; McGill University and Genome Quebec Innovation Center, McGill University, Montreal, QC, Canada; Micro and Nanobioengineering Laboratory, Department of Biomedical Engineering, McGill University, Montreal, QC, Canada*
Jenny Kim Kim • *Herbert Irving Comprehensive Cancer Center, Proteomics Shared Resource, Columbia University Medical Center, New York, NY, USA*
Antonius Koller • *Herbert Irving Comprehensive Cancer Center, Proteomics Shared Resource, Columbia University Medical Center, New York, NY, USA*
Arvind M. Korwar • *Proteomics Facility, Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune, India*
Petri Kouvonen • *Turku Centre for Biotechnology, University of Turku, Turku, Finland*
Oskar Knittelfelder • *Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany*
Christoph Krisp • *Department of Chemistry and Biomolecular Sciences, Australian Proteome Analysis Facility (APAF), Macquarie University, Sydney, Australia*
Mahesh J. Kulkarni • *Proteomics Facility, Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune, India; Academy of Scientific and Innovative Research (AcSIR), New Delhi, India*
Martin Kussmann • *Nestlé Institute of Health Sciences SA, Lausanne, Switzerland*
Bonnie Ky • *Division of Cardiovascular Medicine, University of Pennsylvania, Philadelphia, PA, USA*
Véronique Laforte • *Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, QC, Canada; Department of Biomedical Engineering, McGill University, Montreal, QC, Canada; McGill University and Genome Quebec Innovation Center, McGill University, Montreal, QC, Canada*
Riitta Lahesmaa • *Turku Centre for Biotechnology, University of Turku, Turku, Finland*
Cheolju Lee • *Center for Theragnosis, Biomedical Research Institute, Korea Institute of Science and Technology, Seoul, Republic of Korea; Department of Biological Chemistry, University of Science and Technology, Daejeon, Republic of Korea*
Ji Eun Lee • *Center for Theragnosis, Biomedical Research Institute, Korea Institute of Science and Technology, Seoul, Republic of Korea*
Huiyan Li • *Department of Biomedical Engineering, McGill University, Montreal, QC, Canada; McGill University and Genome Quebec Innovation Center, McGill University, Montreal, QC, Canada*
Qing-Run Li • *Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China*
Pengyuan Liu • *The Wistar Institute, Philadelphia, PA, USA*

Yan-Sheng Liu • *Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China*
Pik-Shan Lo • *Department of Biomedical Engineering, McGill University, Montreal, QC, Canada; McGill University and Genome Quebec Innovation Center, McGill University, Montreal, QC, Canada*
Bruno Manadas • *CNC.IBILI, University of Coimbra, Coimbra, Portugal; CNC—Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal*
K.P. Manubhai • *Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, India*
Tania Martins-Marques • *Institute of Biomedical Imaging and Life Sciences (IBILI), Faculty of Medicine, University of Coimbra, Coimbra, Portugal; CNC.IBILI, University of Coimbra, Coimbra, Portugal*
Mojgan Masoodi • *Lipid Biology, Nestlé Institute of Health Sciences, Lausanne, Switzerland; Faculty of Medicine, Department of Nutritional Sciences, University of Toronto, Toronto, ON, Canada*
Kieran Matic • *School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia*
Rune Matthiesen • *Computational and Experimental Biology Group, CEDOC—Chronic Diseases Research Center, Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Lisboa, Portugal*
Kevin W. Meyer • *Perfinity Biosciences, Inc., West Lafayette, IN, USA*
Mark P. Molloy • *Department of Chemistry and Biomolecular Sciences, Australian Proteome Analysis Facility (APAF), Macquarie University, Sydney, Australia*
Delphine Morin-Rivron • *Lipid Biology, Nestlé Institute of Health Sciences, Lausanne, Switzerland*
Robert Moulder • *Turku Centre for Biotechnology, University of Turku, Turku, Finland*
Shuvolina Mukherjee • *Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, India*
Stephen P. Mulligan • *School of Life and Environmental Sciences, University of Sydney, Sydney, Australia; Kolling Institute of Medical Research, Royal North Shore Hospital, St. Leonards, NSW, Australia*
Peter Nilsson • *Affinity Proteomics, School of Biotechnology, SciLifeLab, KTH—Royal Institute of Technology, Stockholm, Sweden*
John P. O'Grady • *Perfinity Biosciences, Inc., West Lafayette, IN, USA*
Purvi Patel • *Herbert Irving Comprehensive Cancer Center, Proteomics Shared Resource, Columbia University Medical Center, New York, NY, USA*
Luisa Pieroni • *Proteomics and Metabonomics Unit, Fondazione Santa Lucia, IRCCS, Rome, Italy*
Cristian Piras • *Department of Veterinary Medicine, University of Milan, Milan, Italy*
Derrick N. Poe • *Perfinity Biosciences, Inc., West Lafayette, IN, USA*
Xiaohong Qian • *National Center for Protein Sciences Beijing, State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing, China*
Teresa Ribeiro-Rodrigues • *Institute of Biomedical Imaging and Life Sciences (IBILI), Faculty of Medicine, University of Coimbra, Coimbra, Portugal; CNC.IBILI, University of Coimbra, Coimbra, Portugal*

Julie A. Reisz • *Department of Biochemistry and Molecular Genetics, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA*
Jianmin Ren • *Proteomic Service Group, Cell Signaling Technology, Danvers, MA, USA*
Manuel S. Rodriguez • *Advanced Technology Institute in Life Sciences (ITAV) CNRS-USR3505, Toulouse, France; University of Toulouse III-Paul Sabatier, Toulouse, France*
Luis M. Ruilope • *Unidad de Hipertension, Instituto de Investigación i+12, Madrid, Spain*
Gema Ruiz-Hurtado • *Unidad de Hipertension, Instituto de Investigación i+12, Madrid, Spain*
Susanne Sales • *Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany*
Jochen M. Schwenk • *Affinity Proteomics, SciLifeLab, School of Biotechnology, KTH—Royal Institute of Technology, Solna, Sweden*
Andrej Shevchenko • *Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany*
Richard J. Simpson • *Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC, Australia*
Ronald Sjöberg • *Affinity Proteomics, School of Biotechnology, SciLifeLab, KTH—Royal Institute of Technology, Stockholm, Sweden*
Rosemary L. Sparrow • *Transfusion Science, Melbourne, VIC, Australia; Department of Immunology and Pathology, Monash University, Melbourne, VIC, Australia*
David W. Speicher • *The Wistar Institute, Philadelphia, PA, USA*
Sanjeeva Srivastava • *Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, India*
Hanno Steen • *Department of Pathology, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA*
Matthew P. Stokes • *Proteomic Service Group, Cell Signaling Technology, Danvers, MA, USA*
Siu Kwan Sze • *Division of Structural Biology and Biochemistry, School of Biological Sciences, Nanyang Technological University, Singapore, Singapore*
Sumaiya Tanjil • *Department of Obstetrics & Gynaecology, Women's Cancer Research Centre, Royal Women's Hospital, Parkville, VIC, Australia*
Andrea Urbani • *Proteomics and Metabonomics Unit, Fondazione Santa Lucia, IRCCS, Rome, Italy; Institute of Biochemistry and Clinical Biochemistry, Catholic University of Sacred Heart, Rome, Italy*
Garikapati Vannuruswamy • *Proteomics Facility, Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune, India*
Sandra Wissmueller • *School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia*
Guowang Xu • *Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China*
Eun Gyeong Yang • *Center for Theragnosis, Biomedical Research Institute, Korea Institute of Science and Technology, Seoul, Republic of Korea*
Peiyuan Yin • *Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China*
Wantao Ying • *National Center for Protein Sciences Beijing, State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing, China*

Rong Zeng • *Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China*
Qing Zhao • *National Center for Protein Sciences Beijing, State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing, China; Department of Traditional Chinese Medicine, Hebei University, Baoding, China*
Jerry Zhou • *School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, Australia*

Part I

Updated Blood Processing and Handling Strategies

Chapter 1

Direct Assessment of Plasma/Serum Sample Quality for Proteomics Biomarker Investigation

Viviana Greco, Cristian Piras, Luisa Pieroni, and Andrea Urbani

Abstract

Blood proteome analysis for biomarker discovery represents one of the most challenging tasks to be achieved through clinical proteomics due to the sample complexity, such as the extreme heterogeneity of proteins in very dynamic concentrations, and to the observation of proper sampling and storage conditions. Quantitative and qualitative proteomics profiling of plasma and serum could be useful both for the early detection of diseases and for the evaluation of pathological status. Two main sources of variability can affect the precision and accuracy of the quantitative experiments designed for biomarker discovery and validation. These sources are divided into two categories, pre-analytical and analytical, and are often ignored; however, they can contribute to consistent errors and misunderstanding in biomarker research. In this chapter, we review critical pre-analytical and analytical variables that can influence quantitative proteomics. According to guidelines accepted by proteomics community, we propose some recommendations and strategies for a proper proteomics analysis addressed to biomarker studies.

Key words Pre-analytical and analytical variables, Sample quality control, Clinical proteomics, Mass spectrometry, Biomarkers

1 Introduction

Serum and plasma, because of their availability and stability, represent a valuable resource for proteome analysis and projects for the detection of biomarkers and for the study of the outcome of human pathologies [1, 2]. All body cells communicate through blood. Many cellular compartments and tissues collect nutrients and release at least part of their contents into the bloodstream, altering its composition. For this reason, this biological specimen is considered a key source of physiological information about the overall status of each tissue, and it can reflect the status of several pathological conditions. However, the translation of the results from research and experiments to the clinical practice still remains a critical point. The composition of serum and plasma specimens could be strongly influenced by the sample pretreatment method used

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_1,

[3–5]. More precisely, the pattern of peptides eventually observed indicates that each study has to be designed and optimized for the specific disease to be investigated, starting from the sample collection to the preprocessing steps [6, 7]. However, serum and plasma proteome analysis is influenced by a wide variety of pre-analytical factors [8, 9] that can contribute to misleading interpretation.

The lack of well-established guidelines to control pre-analytical and analytical phases to improve sample quality and datasets reliability has a negative impact for the clinical proteomics studies and biomarker discovery [10]. Even small details in specimens processing or handling could affect the outcome of the analytical results [11]. In order to avoid these problems, standard operating procedures (SOPs) are necessary to minimize differences due to sample handling [11], and, once SOPs have been defined, it is necessary to develop high-throughput, reliable, and cost-effective analytical tools for the investigation of proper SOP application in sample collection and preprocessing. One of the main difficulties in developing SOPs for proteomics protocols for biomarker discovery relies on the variability of MS profiles and datasets that are strongly connected to the type of employed instrumentation. While for most of the clinical chemistry and immunochemistry routine applications are available well-defined and specific SOPs, regarding blood proteomics, standardized procedures are still lacking. For all these reasons, an analytical and direct quality control of specimens represents a key element in determining whether the correct SOPs have been properly applied, from sample collection and handling to sample storage. This is even more important if considering the increasing number of biobanks and their important role in support of clinical research projects. In this chapter, we present a systematic evaluation of sample handling, collection storage, and processing conditions that should be taken into account to perform a proper proteomics study for biomarker discovery, verification, and validation.

2 Plasma and Serum: Composition and Their Role as Specimen with Diagnostic and Prognostic Value

The blood proteome is one of the most complex components of the human proteome. Its content reflects the physiological and pathological status of each subject. One of the most important advantages of blood samples is related to the quite simple sample withdrawal. Urine, saliva, or feces are easier to collect, but they are less informative and stable and show higher variability according to the time of collection and individual habits. A blood withdrawal takes place through a venipuncture and subsequent collection in a vial containing or not anticoagulants. This first difference in the collection technique directly brings to the formation of two different

biological specimens: plasma and serum [12, 13]. Plasma, the liquid component of blood in which blood cells are suspended, represents about 55% of total blood volume. Plasma is obtained after the centrifugation of whole blood. To avoid clotting, when the blood is withdrawn, an anticoagulant (EDTA, sodium citrate, or heparin) is added immediately after its collection; afterward, the sample is gently centrifuged to remove blood cells [14, 15]. On the other hand, serum collection is done in the absence of anticoagulants after blood clotting and centrifugation. The centrifugation step allows the removal of fibrin clot and cellular elements. Because of this simple processing, plasma sampling is less time-consuming and offers more advantages: no clotting time is requested, the volume obtained is 10–20% higher than serum sampling, and the amount of proteins is higher than serum [16, 17]. Protein profiles of plasma and serum are different because plasma contains fibrinogen protein and other clotting factors. The choice to collect plasma better than serum or vice versa depends on the purpose of sample withdrawal. In some cases, serum could be more useful because of the absence of anticoagulants; in other cases, plasma could represent a better choice for its higher stability or for the study of features of clotting cascade [18].

Blood composition includes salts, lipids, amino acids, carbohydrates, and a large sort of proteins. It collects proteins from tissues and organs from the whole body and contains more than 10,000 different protein classes [19, 20]. These numbers, however, just partially reflect the real complexity of protein species, which is determined by the huge number of different proteoforms and classes of antibodies [21]. Serum/plasma is composed of 90% of high-abundant proteins (HAPs). Albumin alone takes into account for at least 50% of the total protein content, and mainly fibrinogen and haptoglobin contribute to the composition of the remaining 40% of the whole plasma proteome [14].

Because of their abundance, HAPs can be seen as a negative player in the field of biomarker research [22]. The presence of HAPs in serum/plasma hides the remaining part of proteome content composed of middle- and low-abundance proteins (LAPs) with high clinical potential as biomarker candidates [22–25]. The region of low molecular weight proteins represents as well a valuable source of diagnostic information [21]. The serum/plasma proteome complexity is increased by the presence of degraded forms of these bigger proteins, by proteins derived by genetic polymorphisms, and by a high number of posttranslational modifications (PTMs) [26]. HAPs are very often not considered valuable as putative biomarkers [27]; however, their relative abundance and their cleaved or modified forms could precisely reflect physiological and pathological status [28]. Several studies have been performed on the role of albumin and other proteins, like apolipoprotein and haptoglobin, as markers for the early detection of different

diseases, from vascular damage [29] to sarcopenia [30] and from meningioma [31] to neurological disorders like multiple sclerosis [32]. Indeed, among them are included the most used proteins in the current clinical practice, and the interpretation of their abundance could represent a key strategy in the current diagnostic procedures. Proteomics investigations provide additional information on the posttranslational modifications of these proteins improving their role and application as biomarkers [33]. The detection of potential molecular markers from both plasma and serum sources has more recognizing potential for several diseases. Hence, novel serum or plasma protein biomarker panels revealed, by quantitative proteomics based on mass spectrometry, their prognostic value in breast cancer [34, 35].

Diseases such as cancer, cardiovascular and metabolic diseases, and autoimmune pathologies show increasing rates worldwide. Early diagnosis is considered a valuable field of intervention in order to control healthcare expenses. However, the use of serum or plasma for biomarker research projects remains a very complex task and has to deal with sample complexity and with the presence of high-abundant proteins that could provide informative value or hide the signal of low-abundant proteins. For this reason, each biomarker research project should deal with this premise and evaluate the best strategy for the research for each pathology.

3 Proteomics of Plasma and Serum for Biomarker Discovery

Biomarker discovery and the development of clinical diagnostic tests can improve the performances in early detection, clinical decision-making, and clinical outcomes. The technological improvements in the field of proteomics opened new horizons for the discovery of novel biomarkers for many existing pathologies. The first important distinction that has to be taken into account is about biomarker *discovery* and *validation*. The biomarker discovery procedure is very demanding and requires, from the proteomics point of view, expensive equipment, very well-trained personnel, and precious specimens. The biomarker validation, because of its definition (a defined clinical utility demonstrated among many patients and populations), could be even more challenging and requires a coordinated research activity among different geographical areas. It involves the recruiting of a consistent number of patients to test the biomarkers robustness [36, 37]. Because of proteomics high demand of financial and time resources, its applications are more adequate for the discovery phase in the biomarker field. However, if considering the last advances about MRM technique, it is feasible to consider proteomics for the validation steps as well [38].

There are several types of biomarkers detectable in serum; the most common and useful biomarkers are characterized by protein biomarkers and related posttranslational modifications. Besides protein biomarkers, there are as well metabolic biomarkers that are commonly used for clinical purposes. Many protein biomarkers are constantly used in the clinical routine such as alanine aminotransferase (ALT) for liver dysfunction or fecal elastase as marker of pancreatic insufficiency [37]. In the field of biomarker discovery, many studies have been performed in the past years. The discovery phases of biomarker research are mostly addressed by an untargeted proteomics approach. There are several methods that can be applied in the biomarker discovery, among them, the gel-based and the gel-free approach.

The gel-based approach is mostly related to the usage of 2D–E followed by MS analysis for protein identification. However, mono-dimensional-based approaches can be used as well and have been successfully applied for the research of biomarkers for membranous nephropathy and focal and segmental glomerulosclerosis [39], for the biomarker discovery of breast cancer [40] and acute myocardial infarction (AMI) [41]. In these last two studies, SELDI-TOF-MS together with SDS-PAGE/MALDI-TOF/TOF and immunoprecipitation/SDS-PAGE together with UHPLC-coupled quadrupole-Orbitrap mass spectrometer have been, respectively, used. SELDI-TOF-MS together with SDS-PAGE/MALDI-TOF/TOF has been used as well for the identification of apolipoprotein A-I as a potential hepatoblastoma biomarker in children [42]. Serum biomarkers to differentiate cholangiocarcinoma from benign biliary tract diseases have been as well successfully identified using a common SDS-PAGE followed by LC-MS/MS approaches [43]. These applications demonstrate the possible successful applications of mono-dimensional SDS-PAGE to the discovery phase of possible biomarkers. Two-dimensional electrophoresis (2D–E) and later 2D–DIGE play a key role in the biomarker discovery field mostly because they probably still represent the best technique for the visualization of protein isoforms or proteoforms of abundant proteins that are valuable as biomarkers. The altered expression and glycosylation pattern of serum haptoglobin and alpha-1-antitrypsin has been discovered in chronic hepatitis C, hepatitis C-induced liver cirrhosis, and hepatocellular carcinoma patients using 2D–E coupled with LC-MS [44]. Two protein isoforms of C3c complement and haptoglobin have been as well identified as putative biomarkers for the detection of Crohn's disease [45]. Among other applications in the field of biomarker discovery, this technique was used to identify novel serum biomarkers of prolonged erythropoiesis-stimulating agent (ESA) exposure (darbepoetin-α) and/or aerobic training [46] and for the detection of early-stage breast cancer. 2D–DIGE basically consists in the use of 2D–E where samples are labeled with fluorescent dyes (e.g.,

Cy3, Cy5, Cy2) prior to two-dimensional electrophoresis. The advantage of using this method in comparison with classical 2D–E is related to the lower amount of sample required and to the precision and resolution of the assay. Among biomarkers proposed and identified through this technique, there are some candidates for the detection of inflammation and innate immunity in atrophic nonunion fracture [47], in congenital disorders of glycosylation (CDG) [48], in acute kidney injury (AKI) [49], in lung squamous cell cancer (SCC) [50], and in inflammatory bowel disease [51].

Gel-free approaches include LC-MS/MS approaches both label based and label-free and represent the most used approaches for untargeted discovery phases of biomarkers. Label-free approaches have been recently used for numerous applications in untargeted proteomics for discovery phase of biomarkers. Among protein biomarkers or biomarker candidates, there are numerous and valuable studies that discovered biomarker candidates that require the validation process. Recently, Mesaros and Blair proposed the quantification of a panel of proteins through LC-MRM/MS for the diagnosis of cardiovascular diseases [39]. LC-ESI-MS/MS has been used to evaluate whether innate immune dysfunction could be somehow involved as co-cause of poststroke depression itrepresents the right strategy (PSD) [52] and, staying in the field of neurosciences, recent important studied addressed potential biomarkers for major depressive disorder (MDD) using both LC-MS(E) and SRM [53]. These represent just two examples of proposed biomarkers for neurological disorders, but applications of label-free MS in this field include the possible diagnosis of several types of cancer [54–57], inflammatory disorders [58], and many other pathologies. Label-based approaches recently used include iTRAQ and TMT (tandem mass tag). Almost all human pathologies can be studied using these approaches. Some examples related to iTRAQ refer to the discovery of putative biomarker for cancer [59, 60], depression [61], and inflammation [62]. TMT has successfully being applied to the study of gastric cancer [63] and, with a novel approach, to the workflow for the analysis of 1000 plasma samples from the multicentered human dietary intervention study "DiOGenes" [64]. In this last example, authors demonstrated the feasibility of the application of isobaric tagging to analyze large number of human plasma samples for biomarker discovery. According to this described evidences, it is becoming clearer that proteomics has in his panel several different valuable techniques extremely useful in the field of biomarker discovery. Moreover, technical implementations are providing more precise tools to improve this research field. However, the more precise is the technique to used, the more important is the role of standardization in sample processing and LC-MS platform, in order to avoid data misinterpretation due to pre-analytical and analytical variables.

4 Plasma/Serum Pre-analytical and Analytical Variables

The first step for a proteomics analysis addressed to a biomarker study is closed inside a sentence: collect a good sample. In disease marker study, where measurement of biochemical markers represents an important aid to clinicians in the early diagnosis and prognosis of diseases, there are two main sources of variability that can affect the precision and the validity of the proteomics quantitation approach. These sources are divided in two categories, pre-analytical and analytical variables. These factors are often ignored; however, they are the principle variables that influence the precision and accuracy of absolute quantitative proteomics experiments designed for biomarker discovery and validation. In this section, we provide the most critical pre-analytical and analytical variables that can influence the outcomes of quantitative proteomics experiments (Fig. 1). According to guidelines accepted by MS community, we propose some recommendations and strategies for a proper proteomics analysis addressed to biomarker studies.

4.1 Pre-analytical Factors

Pre-analytical steps include all the processes performed before the analysis of the biofluid [14, 27, 65–67]. In a quantitative plasma/serum proteomics study, three key sample-related technical factors can contribute to influence the validity of the outcome of the quantitative analysis. Beside them, there are biological variables, strictly linked to patients (age sex, life style) [19]. We have chosen to focus on technical variables that can interfere with mass spectrometry

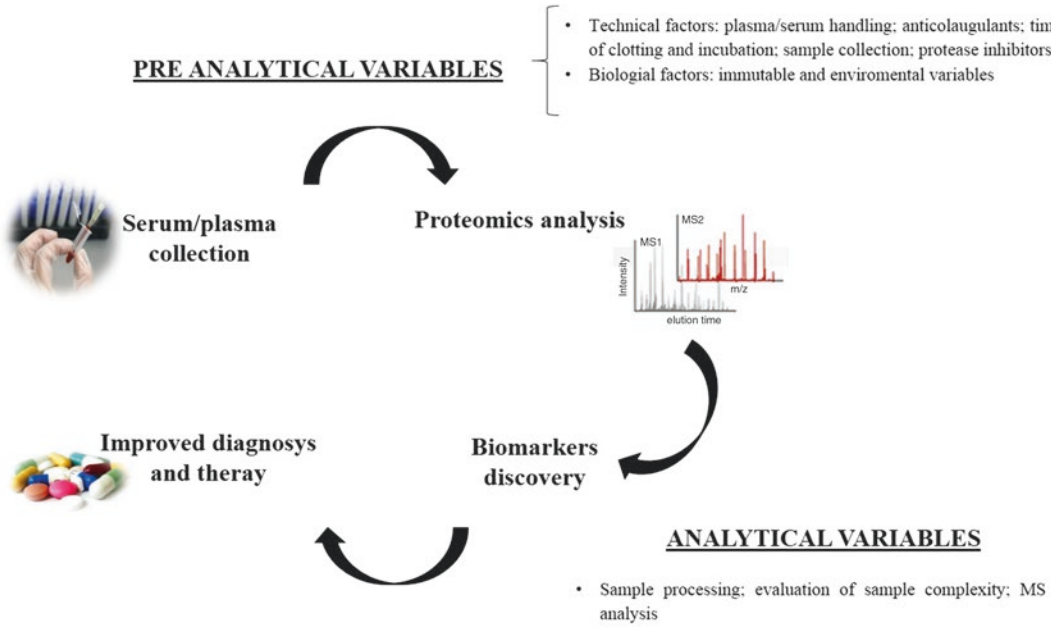


Fig. 1 Pre-analytical variables and analytical variables

analysis. The technical factors include the variables linked to sample selection (plasma or serum), sample collection (phlebotomy, anticoagulants, protease inhibitor, and tubes), and sample stability and storage (temperature, freeze-thaw cycles).

4.1.1 Sample Selection

It is a false belief that plasma and serum are the same fluid. According to a clinical chemistry point of view, serum differs from plasma only because it lacks fibrinogen, but they mostly differ for protein composition and concentration. Protein and peptide profiles of plasma and serum have been reported to be dissimilar [24, 68].

The same coagulation process for the serum production is the first source that causes differences in the serum protein content. Due to the phenomenon of clotting by the action of proteases, the coagulation events, and the enzymatic activities, a great amount of peptides has been observed in sera and not in plasma. Moreover, the use of clotting in plastic inside of clotting in glass and the choice of clot activators have different effects on the blood proteome [69, 70].

Therefore, for all these biochemical issues previously described, the current consensus in the proteomics community recommends the use of plasma for biomarker studies [72–74].

4.1.2 Sample Collection

Modes of sample withdrawal. Sample collection mode represents a critical step for two aspects, one related to patient and withdrawal and the other one linked to laboratory practice [75, 76].

The patient's posture (if he is standing, lying, or sitting), the tourniquet application, the site of withdrawal, and the use of alcohol to clean the skin can cause hemolysis and influence proteomics analysis [75, 76]. To avoid the diurnal fluctuations of blood biomarkers, blood drawl should be in the morning, before 10 a.m. fasting [77]. In the step of drawl, a 21-G needle is preferred to avoid the risk of hemolysis. The median cubital vein, usually easily found and accessed, is considered the preferred site. The skin should be washed with alcohol to evaporate in order to avoid the hemolysis consequently to blood contamination [77].

Anticoagulants. The plasma handling without clotting is obviously time-consuming; however, some other precautions are requested. In a biomarker study, the choice for serum or plasma and, if plasma, for which anticoagulants to use, represents an important step to address a proper sample preparation. Plasma can be naturally converted to serum at RT by the action of the proteases of thrombin on fibrinogen and on the other components of the coagulation cascade [14, 78, 79].

The use of anticoagulant (EDTA, heparin, citrate) is requested for plasma separation in order to protect plasma samples from clotting [14].

The Human Proteome Organization (HUPO) highlighted that there is the need of standardized guidelines for each anticoagulant [80]. For a quantitative proteomics analysis, EDTA seems to be the elective anticoagulant [81]. However in general, it is necessary to evaluate the single type of experiment, because the use of anticoagulants may affect the stability of some specific proteins [82]. In fact, the anticoagulant can also present some drawbacks [18]. For example, the heparinized samples seem to be more stable [79]; therefore, heparin can interact not only with antithrombin III factor but also with a considerable number of different proteins [78]. Moreover, being a highly charged molecule, it can influence the binding of other molecules in solution, and it is important in some particular techniques such as chromatographic separation [14, 79]. EDTA is a common anticoagulant with a negative charge. It can bind to metal ions decreasing their reactivity so it is not indicated to use EDTA in the experiments through the use of divalent cations. EDTA seems less stable in comparison to heparin [68], but it represents the right strategy to inhibit proteases that necessitate metal ions for the [81] coagulation process [68]. Another negative aspect depends on EDTA's ability to induce platelet aggregation altering the plasma proteome content [83]. Citrate is the other common anticoagulant able to bind calcium and commonly present in fluid forms in the tubes, leading to a dilutional effect after the addition of blood during sample preparation [80]. In this case, it is critical to calculate the right ratio anticoagulant/blood in order to not dilute excessively the sample [80].

Proteaseinhibitors. To avoid the protein degradation, protein inhibitors should be added to the samples. HUPO HPP highlighted the importance of the use of protease inhibitors starting to the sample collection phase. By the analysis of different peptide peaks obtained by SELDI-TOF-MS, it has been demonstrated that plasma proteomics profiling of samples treated with inhibitors is more stable with respect to the other untreated [80]. In general, serum and plasma citrate were observed to be most proteolytically active followed by plasma heparin and plasma EDTA. On the other hand, it has been demonstrated that the adding of protease inhibitors on the heparinized EDTA and citrated samples leads to a perturbation of p*I* profile evaluable. In fact, these molecules are able to modify active sites of proteases like trypsin and chymotrypsin. Moreover, they can alter serine residues on other proteins by adding an amine group, thus shifting the pI to higher values. All these modifications can be evaluated by smearing 2D gel [14].

Collection tubes. On the other hand, when the samples are collected, other factors can alter the proteome content. Silicones, plastic covering, polymeric surfactants (polyvinylpyrrolidone or polyethylene glycols), and polymeric gels adapt to regulate viscosity, can be released from tubes, and alter the serum/plasma content, interfering with peaks of sample detectable with MALDI spectra [84]. It has also been demonstrated that the same sample collected

in different tubes, like red top tubes or tiger typo tubes, can show a different proteomics profiling [85, 86].

Collection tubes preloaded with a protease inhibitor cocktail and anticoagulants have been found to produce reproducible plasma samples [70]. By contrast, no commercially available serum collection tubes have yet been developed which can reproducibly produce serum samples for proteomics studies. For this reason, HUPO has recommended using plasma instead of serum, and this is the rationale for other mass spectrometry group [70, 72–74].

4.1.3 Sample Stability and Storage

Zimmerman et al. demonstrated that plasma proteome is stable to delay in sample preparation until 1 week at 4 °C, until 25 freeze-thaw cycles and hemolysis [87]. These data are confirmed by a recent and relevant multicentric study performed by Mateos et al. [88]. They have highlighted that blood proteins in plasma are broadly insensitive to such pre-analytical variables as delayed processing or freeze-thaw cycles when analyzed at the peptide level. This is another characteristic that makes plasma better than serum sample. Although the major stability, several studies performed by Martino et al. [89], Lundbald et al. [71], and Mann [90] show a kind of diurnal variations of the concentrations of some proteins as plasminogen, transthyretin, and apolipoprotein involved in some pathways with a circadian response. For this reason, it would be better to collect the sample in the same day [71].

Serum proteome differences are strictly linked to the time of coagulation and temperature of storage [85]. During the coagulation process, peptides can be degraded, and new peptides can be formed. Apweiler et al. demonstrated that serum peptide profiling could be modified during clotting by the time (30–60 min) [91]. After 60 min, the serum proteomics profiling has been shown to be altered only if sample was stored at RT; while if it was stored on ice, it has been not shown any modification.

As recommended by HUPO, serum should be obtained 60 min after clotting at RT and subsequently stored at 4–8 °C [91] or, preferably, at −80 °C before analysis [91].

Temperature. However, as highlighted also for other fluids [10, 92, 93], the temperature is the key element for the stability of proteome and for enzymatic activity, in whole analysis, from the withdrawal to collection and from the transport to the storage. It is critical for collection, transport practice, and freeze-thaw cycles, and its effects are evident in sample quality. Several studies have been performed concerning these critical factors [94, 95].

According to Rai et al. [80], storage with liquid nitrogen would represent the ideal condition to guarantee protein stability, and when it is not possible, temperature should be set at −80 °C, immediately after the handling, preferable in small aliquots [80, 94]. For the shipment, ice cold packs and the dry ice should be used. Frequent cycles of storage should be avoided limiting to two refreezing steps [80].

4.2 Analytical Factors

During a proteomics quantitative experiment, each step of whole workflow, from sample processing to MS analysis, can be influenced from different variables. Therefore, analytical factors principally include sample processing procedures that are fundamental when biomarker studies move to translate into biomarkers candidates for multicentric and larger studies. The sample processing should include validated protocols and standardized guidelines that elucidate not only the methodological approach but also defining the initial processing of samples including protein sample pipetting, buffer composition, and sample dilution. These are two important steps for MS analysis and the bond of proteins during the analysis, detectable in MS spectra [14]. In particular, in a sample processing procedure, analytical variables principally include sample pretreatment, tryptic digestion, use of internal standard, chemical interference, and LC-MS platform [70].

Sample pretreatment. Because of the extreme heterogeneity of proteins in a very dynamic concentration, a comprehensive analysis becomes very difficult if before the sample is not fractionated or processed [96]. The wide dynamic range of protein concentration and the ratio HAPs and LAPs previously described represent critical problems for proteomics analysis [14, 27]. Hence, the reduction of proteome complexity and so the separation or removal of HAPs are essential steps to identify characteristic proteins for that contest. Immunoaffinity depletion and enrichment are common technologies useful to simplify the complexity of blood proteome for a sensitive and accurate proteomics quantitative analysis [70]. While depletion techniques are performed at proteomics level, serum and plasma enrichment is on peptide levels [97, 98]. Strategies as centrifugal ultrafiltration, solid phase extraction (dye-ligand binding, antibody based, ion exchange), and organic solvent extraction should be used [99–103]. However, these techniques are not ideal: they are not completely sensitive and specific, and they affect sample recovery removing a good portion of HAPs [100]. For this reason, the enrichment of low-abundance proteins can be performed by the use of microscopic beads or functionalized particles and peptide ligand libraries [19]. In addition, to have a more comprehensive analysis, pre-fractionation and multidimensional separation techniques are required. Among them, the most used include RP-LC, strong cation exchange chromatography, anion exchange chromatography, SDS-PAGE, and membrane-based or free-flow electrophoresis [104–106]. In general, due to these contrasting characteristics, the use of sample pretreatment technique is not recommended for biomarker studies where the putative biomarker candidate is not known among a large number of proteins. In contrast, it is more recommended in a clinical study when the biomarker is known and approved and it needs to have a quantification [70].

Tryptic digestion. In a bottom-up LC-MS analysis, trypsin is considered the elective and the most widely used protease [70]. It

shows several advantages: high specificity in cleaving amine bounds (C-term of Lys/Arg), low cost, generation of ideal peptides for size (14 mer as average), and cleavage (+1 to +3) for both efficient MS ionization and fragmentation [107, 108]. Digestion efficiency is influenced by three factors: denaturant, buffer, and grade and source of trypsin [108]. Differences in the denaturation and digestion method can influence the efficiency of digestion and consequently the outcomes of proteomics quantitative analysis [108, 109]. Urea is considered as conventional chaotrope; however, sodium deoxycholate is recommended in plasma proteomics analysis because it can be easily removed prior to LC-MS analysis centrifuging the acidified digest [110]. For the grade and source of trypsin, sequencing grade modified trypsin (from Promega, WI, USA) is considered the best among the commercially available trypsin [111]. The assessment of digestion efficiency is a key point necessary to guarantee a proper workflow of analysis. The use of stable isotope internal standard, as described below, to add to the sample at the beginning of the analytical process could be a strategy to check the efficiency of tryptic digestion.

Automation. One of the strategies to reduce the analytical variability could be the use of liquid handling robotics, which are designed to reduce human errors and the variations in pretreatment between samples. Several robotics workstations are available that can be used to perform the automatic digestion, the manipulation of small sample size, and the handling of affinity beads [70].

Internal standard. In order to achieve the strong accuracy and precision, the use of isotopically labeled internal standard (IS, standard labels with 13C/15 N isotope at C-term of Lys and Arg residues) for an absolute quantitative proteome analysis could be optimal. These lead to verify normalizing the sample preparation and instrumental-related variability [112]. IS shows the same characteristics of its homologous. Therefore, two peptides, natural and standard peptide, have the same behavior during chromatography, electrospray ionization, and collision-induced dissociation, while they are different for m/z value of precursor ion and y series. IS can be used to check the efficiency of digestion [70].

Chemical interference. In a proteomics quantitative analysis, chemical interferences can affect the outcomes of the analysis. Background interferences and high chemical noise can occur principally due to the complexity of sample characterized by a wide protein concentration range and by millions of peptides after proteomics digestion [70]. Sample pre-fractionation and longer chromatographic runs can improve the analysis [70]. Several strategies have been developed to detect interferences [113, 114].

LC-MS platform. The specific LC-MS platform used has an impact on the accuracy and reproducibility of the results. Several studies have been performed [115, 116]; the fundamental evidence is that the lower is the resolution of the instruments, the

higher is the probability to generate false positives in a MS-based analysis as well as chromatographic resolution and reproducibility [70, 115].

5 Tools for Quality Assessment of Plasma/Serum for Biomarker Discovery

In the field of biomarker discovery, in the latest years, proteomics is moving from a qualitative to a quantitative approach. However, despite all interests and advances in the knowledge of human blood proteome, the MS community continues to lack plasma and serum proteomics standardized procedures. Over the years, one of the most important aims was the development of standardized methods and system performance in targeted plasma/serum proteomics. Several guidelines have been proposed to reduce the pre-analytical variability that crop out from blood collection to handling. The guidelines related to standard operating procedures could contribute to at least reduce, if not eliminate, these source of putative errors. Proteomics becomes a powerful tool to highlight how these factors could have an impact on the outcomes of biomarker studies. By the use of SELDI-TOF-MS, Karsan et al. demonstrated that variations in a MS spectra were linked to sample collections and processing, without pathological or physiological relevance in the samples for breast cancer diagnosis [117]. Another similar study was performed by Marshall et al. that show how changes in proteomics profiles in plasma of patients with myocardial infarction were due to differences in the delaying time between the sample collection and measurements, rather than to the disease condition [94]. These kinds of studies elucidate how recognizing the effects of pre-analytical factors could be important to address proper proteomics studies for biomarker discovery.

The importance of controlling them has been highlighted in several interlaboratory study, and they have been widely reviewed. HUPO and in particular HUPO-PPP (Plasma Proteome Project) initiative have set a general standard for the comprehensive analysis of the protein constituents of human plasma and serum through the standardization of protocols from sample preparation to proteomics data analysis related to plasma proteome. The HUPO-PPP stressed that there are too many variables to consider to make a universal statement on the best plasma, but the consideration of these variables in study design, along with thorough documentation of all steps of handling and processing of the samples, can overcome or minimize or even mitigate some of these problems. HUPO proposed some pre-analytical variables that need to check through SOPs developed and followed to test these confounding factors on proteomics assay. Moreover, in addition to blood sample quality control, the reproducibility of a quantitative proteomics approach should be also checked. Irregular outcomes due to the

sample preparation as well as performance in LC-MS analysis could lead to erroneous quantification in a biomarker study.

The application of mass spectrometry for the direct analysis of blood samples has often neglecting some relevant source of errors associated with the reproducibility of gas phase reactions. In particular, the ionization phase and the fragmentation reactions are considered to be free from any experimental error [118, 119]. Though the stability of mass spectrometry technology has reached high standard, it is quite difficult to have a stable desorption ionization process between different instruments or even within the same instrument over a long period. Ionization performance criteria have been introduced [120–123], and latest MALDI-TOF-MS instruments provide embedded systems for ion source cleaning. The fragmentation pattern is a second issue, which needs to be taken into account when translating quantitative multiplexing assays between different instruments. In fact, the fragmentation pattern is dependent from the fragmentation procedure, and it is necessary to validate the fragment ion profile according to the different instruments employed. This effect has been limiting so far the wide application of spectral library matching algorithm for clustering multicenter investigations [123].

To overcome this deficiency, some groups addressed this research to develop protocols and kits for evaluating sample preparation procedures and LC platforms in MRM-based plasma quantitative proteomics analysis. In this contest, important studies have been performed. Percy et al. developed two kits to assist inter- and intra-laboratory quality control MRM experiments, the first one to test the efficacy and the robustness of LC-MRM/MS platform and the second one to test the whole analytical workflow of sample preparation [108]. More recently, Gallien et al. continued to focus on this kind of quality control study developing a simple analytical method, applicable also in different laboratories, able to assess and check the instrument performance and the reproducibility of the data, from the sample preparation, the efficiency of tryptic digestion, and chromatographic and mass spectrometry performances [124]. Quality control evaluations are necessary if quantitative plasma/serum proteomics analysis is to be used in discovery and validation of putative disease biomarkers across different research laboratories.

6 Conclusion

In recent years, quantitative proteomics analysis becomes a powerful tool to ensure and guarantee accuracy and reproducibility in a biomarker study. “Collect to a good sample” is the prerequisite for a proper proteomics analysis addressed to biomarker discovery and validation. For this to happen, a comprehensive knowledge and

understanding of those pre-analytical and analytical variables that can affect proteomics results are necessary. Some of these can be overcome by SOPs. However, the MS community continues to lack plasma and serum proteomics well-standardized proteomics procedures. The harmonization of these processes, from the assessment of the sample quality to the technology platform performance, remains one of the major challenge for the proteomics community [124]. Beside SOPs and guidelines already in force, a proper study design is requested by proteomics researchers, with particular attention to pre-analytical and analytical factors described. In this way, MS-based proteomics can achieve accuracy and reproducibility in biomarker studies.

References

1. Geyer PE et al (2016) Plasma proteome profiling to assess human health and disease. Cell Syst 2(3):185–195
2. Pieragostino D et al (2010) Pre-analytical factors in clinical proteomics investigations: impact of ex vivo protein modifications for multiple sclerosis biomarker discovery. J Proteome 73(3):579–592
3. Poulsen K et al (2012) Characterization and stability of transthyretin isoforms in cerebrospinal fluid examined by immunoprecipitation and high-resolution mass spectrometry of intact protein. Methods 56(2):284–292
4. Hubel A et al (2011) State of the art in preservation of fluid biospecimens. Biopreserv Biobank 9(3):237–244
5. Schoonenboom NS et al (2005) Effects of processing and storage conditions on amyloid β (1–42) and tau concentrations in cerebrospinal fluid: implications for use in clinical practice. Clin Chem 51(1):189–195
6. Aldred S, Grant MM, Griffiths HR (2004) The use of proteomics for the assessment of clinical samples in research. Clin Biochem 37(11):943–952
7. Yi J, Craft D, Gelfand CA (2011) Minimizing preanalytical variation of plasma samples by proper blood collection and handling. Methods Mol Biol:137–149
8. Plebani M (2006) Errors in clinical laboratories or errors in laboratory medicine? Clin Chem Lab Med 44(6):750–759
9. Del Boccio P et al (2007) Cleavage of cystatin C is not associated with multiple sclerosis. Ann Neurol 62(2):201–204
10. Greco V et al (2014) Direct analytical sample quality assessment for biomarker investigation: qualifying cerebrospinal fluid samples. Proteomics 14(17–18):1954–1962
11. Tuck MK et al (2008) Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. J Proteome Res 8(1):113–117
12. Rodak BF, Fritsma GA, Keohane E (2013) Hematology: clinical principles and applications. Elsevier Health Sciences, Amsterdam
13. Thomson JM (1984) Specimen collection for blood coagulation testing. Lab Hematol 2:833–863
14. Lista S, Faltraco F, Hampel H (2013) Biological and methodical challenges of blood-based proteomics in the field of neurological research. Prog Neurobiol 101-102:18–34
15. Ray S et al (2011) Proteomic technologies for the identification of disease biomarkers in serum: advances and challenges ahead. Proteomics 11(11):2139–2161
16. Tammen H (2008) Specimen collection and handling: standardization of blood sample collection. Methods Mol Biol (Clifton, NJ) 428:35–42
17. Tammen H, Hess R (2011) Collection and handling of blood specimens for peptidomics. Methods Mol Biol (Clifton, NJ) 728: 151–159
18. Luque-Garcia JL, Neubert TA (2007) Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. J Chromatogr A 1153(1–2):259–276
19. Thadikkaran L et al (2005) Recent advances in blood-related proteomics. Proteomics 5(12):3019–3034
20. Adkins JN et al (2002) Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. Mol Cell Proteomics 1(12):947–955

21. Liotta LA, Ferrari M, Petricoin E (2003) Clinical proteomics: written in blood. Nature 425(6961):905
22. Diamandis EP (2004) Mass spectrometry as a diagnostic and a cancer biomarker discovery tool – opportunities and potential limitations. Mol Cell Proteomics 3(4):367–378
23. Diamandis EP (2004) Analysis of serum proteomic patterns for early cancer diagnosis: drawing attention to potential problems. J Natl Cancer Inst 96(5):353–356
24. Tammen H et al (2005) Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. Proteomics 5(13):3414–3422
25. Greening DW, Simpson RJ (2011) Low-molecular weight plasma proteome analysis using centrifugal ultrafiltration. Methods Mol Biol 728:109–124
26. Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1(11):845–867
27. Hu S, Loo JA, Wong DT (2006) Human body fluid proteome analysis. Proteomics 6(23):6326–6353
28. Hortin GL, Sviridov D (2010) The dynamic range problem in the analysis of the plasma proteome. J Proteome 73(3):629–636
29. Cetın N et al (2016) Serum albumin and von Willebrand factor: possible markers for early detection of vascular damage in children undergoing peritoneal dialysis. Clin Investig Med 39(4):E111
30. Can B et al (2016) Serum markers of inflammation and oxidative stress in sarcopenia. Aging Clin Exp Res:1–8
31. ABBRITTI RV et al (2016) Meningiomas and proteomics: focus on new potential biomarkers and molecular pathways. Cancer Genomics Proteomics 13(5):369–379
32. LeVine SM (2016) Albumin and multiple sclerosis. BMC Neurol 16(1):1
33. Greifenhagen U et al (2016) Plasma proteins modified by advanced Glycation end products (AGEs) reveal site-specific susceptibilities to glycemic control in patients with type 2 diabetes. J Biol Chem 291(18):9610–9616
34. Chung L et al (2014) Novel serum protein biomarker panel revealed by mass spectrometry and its prognostic value in breast cancer. Breast Cancer Res 16(3):R63
35. Suman S et al (2016) Quantitative proteomics revealed novel proteins associated with molecular subtypes of breast cancer. J Proteome 148:183–193
36. Li D, Chan DW (2014) Proteomic cancer biomarkers from discovery to approval: it's worth the effort. Expert Rev Proteomics 11(2):135–136
37. Crutchfield CA et al (2016) Advances in mass spectrometry-based clinical biomarker discovery. Clin Proteomics 13(1):1
38. Percy AJ et al (2016) Clinical translation of MS-based, quantitative plasma proteomics: status, challenges, requirements, and potential. Expert Rev Proteomics 13(7):673–684
39. Pant P et al (2016) Serum sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis of patients with membranous nephropathy and focal and segmental glomerulosclerosis. Saudi J Kidney Dis Transpl 27(3):539
40. Sun Y et al (2016) Identification of apolipoprotein CI peptides as a potential biomarker and its biological roles in breast cancer. Med Sci Monitor 22:1152
41. Streng AS et al (2016) Development of a targeted selected ion monitoring assay for the elucidation of protease induced structural changes in cardiac troponin T. J Proteome 136:123–132
42. Zhao W et al (2015) Screening and identification of apolipoprotein AI as a potential hepatoblastoma biomarker in children, excluding inflammatory factors. Oncol Lett 10(1):233–239
43. Janvilisri T et al (2015) Novel serum biomarkers to differentiate cholangiocarcinoma from benign biliary tract diseases using a proteomic approach. Dis Markers 2015
44. Mondal G et al (2016) Altered glycosylation, expression of serum haptoglobin and alpha-1-antitrypsin in chronic hepatitis C, hepatitis C induced liver cirrhosis and hepatocellular carcinoma patients. Glycoconj J 33(2):209–218
45. Piras C et al (2014) Serum protein profiling of early and advanced stage Crohn's disease. EuPA Open Proteom 3:48–59
46. Christensen B et al (2015) Serum proteomic changes after randomized prolonged erythropoietin treatment and/or endurance training: detection of novel biomarkers. PLoS One 10(2):e0117119
47. de Seny D et al (2016) Biomarkers of inflammation and innate immunity in atrophic nonunion fracture. J Transl Med 14(1):258
48. Heywood WE et al (2016) Global serum glycoform profiling for the investigation of dystroglycanopathies & congenital disorders of glycosylation. Mol Genet Metab Rep 7:55–62
49. Wu F et al (2015) Identification of phosphorylated MYL12B as a potential plasma biomarker for septic acute kidney injury using

a quantitative proteomic approach. Int J Clin Exp Pathol 8(11):14409

50. Okano T et al (2016) Identification of haptoglobin peptide as a novel serum biomarker for lung squamous cell carcinoma by serum proteome and peptidome profiling. Int J Oncol 48(3):945–952
51. Viennois E et al (2015) Longitudinal study of circulating protein biomarkers in inflammatory bowel disease. J Proteome 112:166–179
52. Nguyen VA et al (2016) A pathway proteomic profile of ischemic stroke survivors reveals innate immune dysfunction in association with mild symptoms of depression-a pilot study. Front Neurol 7:85
53. Ruland T et al (2016) Molecular serum signature of treatment resistant depression. Psychopharmacology 233(15–16):3051–3059
54. Lin Q et al (2015) Analysis of colorectal cancer glyco-secretome identifies laminin β-1 (LAMB1) as a potential serological biomarker for colorectal cancer. Proteomics 15(22):3905–3920
55. Gu H et al (2016) Quantitative profiling of post-translational modifications by immunoaffinity enrichment and LC-MS/MS in cancer serum without immunodepletion. Mol Cell Proteomics 15(2):692–702
56. Boichenko AP et al (2014) A panel of regulated proteins in serum from patients with cervical intraepithelial neoplasia and cervical cancer. J Proteome Res 13(11):4995–5007
57. Tonry CL et al (2015) Discovery and longitudinal evaluation of candidate protein biomarkers for disease recurrence in prostate cancer. J Proteome Res 14(7):2769–2783
58. McArdle A et al (2015) Developing clinically relevant biomarkers in inflammatory arthritis: a multiplatform approach for serum candidate protein discovery. Proteomics Clin Appl 10(6):691–698
59. Lin C et al (2016) ITRAQ-based quantitative proteomics reveals apolipoprotein AI and transferrin as potential serum markers in CA19-9 negative pancreatic ductal adenocarcinoma. Medicine 95(31):e4527
60. Wang X et al (2016) iTRAQ-based proteomics screen identifies LIPOCALIN-2 (LCN-2) as a potential biomarker for colonic lateral-spreading tumors. Sci Rep 6
61. Wang Q et al (2016) iTRAQ technology-based identification of human peripheral serum proteins associated with depression. Neuroscience 330:291–325
62. Lee SE et al (2015) Plasma proteome biomarkers of inflammation in school aged children in Nepal. PLoS One 10(12):e0144279
63. Xiao H et al (2016) Differential proteomic analysis of human saliva using tandem mass tags quantification for gastric cancer detection. Sci Rep 6
64. Cominetti O et al (2015) Proteomic biomarker discovery in 1000 human plasma samples with mass spectrometry. J Proteome Res 15(2):389–399
65. Rosenling T et al (2009) The effect of preanalytical factors on stability of the proteome and selected metabolites in cerebrospinal fluid (CSF). J Proteome Res 8(12):5511–5522
66. Takehana S et al The effects of pre-analysis sample handling on human plasma amino acid concentrations. Clin Chim Acta 2016;455:68–74.
67. Mirjanic-Azaric B et al (2015) The impact of time of sample collection on the measurement of thyroid stimulating hormone values in the serum. Clin Biochem 48(18):1347–1349
68. Banks RE et al (2005) Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. Clin Chem 51(9):1637–1649
69. Caisey JD, King DJ (1980) Clinical chemical values for some common laboratory animals. Clin Chem 26(13):1877–1879
70. Percy AJ, Parker CE, Borchers CH (2013) Pre-analytical and analytical variability in absolute quantitative MRM-based plasma proteomic studies. Bioanalysis 5(22):2837–2856
71. Lundblad RL (2005) Considerations for the use of blood plasma and serum for proteomic analysis. Int J Gastroenterol 1:1–11
72. Omenn GS (2004) The human proteome organization plasma proteome project pilot phase: reference specimens, technology platform comparisons, and standardized data submissions and analyses. Proteomics 4(5):1235–1240
73. Omenn GS (2007) THE HUPO human plasma proteome project. Proteomics Clin Appl 1(8):769–779
74. Omenn GS et al (2005) Overview of the HUPO plasma proteome project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics 5(13):3226–3245
75. Lippi G et al (2006) Influence of hemolysis on routine clinical chemistry testing. Clin Chem Lab Med 44(3):311–316
76. Lippi G et al (2005) Preanalytical variability in laboratory testing: influence of the blood drawing technique. Clin Chem Lab Med 43(3):319–325

77. Rodriguez AD, Gonzalez PA (2009) Diurnal variations in biomarkers used in cardiovascular medicine: clinical significance. Rev Esp Cardiol 62(11):1340–1341
78. Capila I, Linhardt RJ (2002) Heparin-protein interactions. Angew Chem Int Ed Engl 41(3):391–412
79. Dammann CE et al (2006) Protein detection in dried blood by surface-enhanced laser desorption/ionization-time of flight mass spectrometry (SELDI-TOF MS). Biol Neonate 89(2):126–132
80. Rai AJ et al (2005) HUPO plasma proteome project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics 5(13):3262–3277
81. Ahn S-M, Simpson RJ (2007) Proteomic strategies for analyzing body fluids. In: Thongboonkerd V (ed) Proteomics of human body fluids. Springer, New York, NY, pp 3–30
82. Jambunathan K, Galande AK (2014) Sample collection in clinical proteomics—Proteolytic activity profile of serum and plasma. Proteomics Clin Appl 8(5–6):299–307
83. White JG (2000) EDTA-induced changes in platelet structure and function: clot retraction. Platelets 11(1):49–55
84. Drake SK et al (2004) Potential interferences from blood collection tubes in mass spectrometric analyses of serum polypeptides. Clin Chem 50(12):2398–2401
85. Hsieh SY et al (2006) Systematical evaluation of the effects of sample collection procedures on low-molecular-weight serum/plasma proteome profiling. Proteomics 6(10):3189–3198
86. Villanueva J et al (2005) Correcting common errors in identifying cancer-specific serum peptide signatures. J Proteome Res 4(4):1060–1072
87. Zimmerman LJ et al (2012) Global stability of plasma proteomes for mass spectrometry-based analyses. Mol Cell Proteomics 11(6):M111.014340
88. Mateos J et al (2017) Multicentric study of the effect of pre-analytical variables in the quality of plasma samples stored in biobanks using different complementary proteomic methods. J Proteome 150:109–120
89. Martino TA et al (2007) Diurnal protein expression in blood revealed by high throughput mass spectrometry proteomics and implications for translational medicine and body time of day. Am J Phys Regul Integr Comp Phys 293(3):R1430–R1437
90. Robles MS, Mann M (2013) Proteomic approaches in circadian biology, in Circadian clocks. Springer, New York, NY, pp 389–407
91. Apweiler R et al (2009) Approaching clinical proteomics: current state and future fields of application in fluid proteomics. Clin Chem Lab Med 47(6):724–744
92. Rosenling T et al (2011) The impact of delayed storage on the measured proteome and Metabolome of human cerebrospinal fluid. Clin Chem 57(12):1703–1711
93. Pieragostino D et al (2013) Oxidative modifications of cerebral transthyretin are associated with multiple sclerosis. Proteomics 13(6):1002–1009
94. Marshall J et al (2003) Processing of serum proteins underlies the mass spectral fingerprinting of myocardial infarction. J Proteome Res 2(4):361–372
95. West-Nielsen M et al (2005) Sample handling for mass spectrometric proteomic investigations of human sera. Anal Chem 77(16):5114–5123
96. Rai AJ, Vitzthum F (2006) Effects of preanalytical variables on peptide and protein measurements in human serum and plasma: implications for clinical proteomics. Expert Rev Proteomics 3(4):409–426
97. Whiteaker JR et al (2010) An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers. Mol Cell Proteomics 9(1):184–196
98. Whiteaker JR et al (2011) Evaluation of large scale quantitative proteomic assay development using peptide affinity-based mass spectrometry. Mol Cell Proteomics 10(4):110.005645
99. Ahmed N et al (2003) An approach to remove albumin for the proteomic analysis of low abundance biomarkers in human serum. Proteomics 3(10):1980–1987
100. Björhall K, Miliotis T, Davidsson P (2005) Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. Proteomics 5(1):307–317
101. Steel LF et al (2003) Efficient and specific removal of albumin from human serum samples. Mol Cell Proteomics 2(4):262–270
102. Gong Y et al (2006) Different immunoaffinity fractionation strategies to characterize the human plasma proteome. J Proteome Res 5(6):1379–1387
103. Fu Q et al (2005) A robust, streamlined, and reproducible method for proteomic analysis of serum by delipidation, albumin and IgG depletion, and two-dimensional gel electrophoresis. Proteomics 5(10):2656–2664

104. Barquinero NC (2016) Identification and quantitation of proteins in human plasma and serum by LC-MS/MS. Nat Methods PMCID:PMC3943160
105. Horn A et al (2006) Multidimensional proteomics of human serum using parallel chromatography of native constituents and microplate technology. Proteomics 6(2):559–570
106. Tang HY et al (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(13):3329–3342
107. Olsen JV, Ong S-E, Mann M (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol Cell Proteomics 3(6):608–614
108. Olsen, J.V., S.-E. Ong, and M. Mann, Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Molecular & Cellular Proteomics, 2004. 3(6): p. 608-614.
109. Proc JL et al (2010) A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. J Proteome Res 9(10):5422–5437
110. Zhou J et al (2006) Evaluation of the application of sodium deoxycholate to proteomic analysis of rat hippocampal plasma membrane. J Proteome Res 5(10):2547–2553
111. Burkhart JM et al (2012) Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. J Proteome 75(4):1454–1462
112. Hoofnagle AN, Wener MH (2009) The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. J Immunol Methods 347(1):3–11
113. Percy AJ et al (2013) Absolute quantitation of proteins in human blood by multiplexed multiple reaction monitoring mass spectrometry. Methods Mol Biol:167–189
114. Keshishian H et al (2007) Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. Mol Cell Proteomics 6(12):2212–2229
115. Lehmann S et al (2013) Quantitative clinical chemistry proteomics (qCCP) using mass spectrometry: general characteristics and application. Clin Chem Lab Med 51(5):919–935
116. Hoofnagle AN (2010) Quantitative clinical proteomics by liquid chromatography–tandem mass spectrometry: assessing the platform. Clin Chem 56(2):161–164
117. Karsan A et al (2005) Analytical and preanalytical biases in serum proteomic pattern analysis for breast cancer diagnosis. Clin Chem 51(8):1525–1528
118. Bae YJ et al (2012) Degree of ionization in MALDI of peptides: thermal explanation for the gas-phase ion formation. J Am Soc Mass Spectrom 23(8):1326–1335
119. Westmacott G et al (2002) The influence of laser fluence on ion yield in matrix-assisted laser desorption ionization mass spectrometry. Int J Mass Spectrom 221(1):67–81
120. Frey BL et al (2005) Controlling gas-phase reactions for efficient charge reduction electrospray mass spectrometry of intact proteins. J Am Soc Mass Spectrom 16(11):1876–1887
121. Bae YJ, Park KM, Kim MS (2012) Reproducibility of temperature-selected mass spectra in matrix-assisted laser desorption ionization of peptides. Anal Chem 84(16):7107–7111
122. O'Rourke MB, Djordjevic SP, Padula MP (2016) The quest for improved reproducibility in MALDI mass spectrometry. Mass Spectr Rev. doi:10.1002/mas.21515
123. Gallien S, Domon B (2015) Detection and quantification of proteins in clinical samples using high resolution mass spectrometry. Methods 81:15–23
124. Gallien, S. and B. Domon, Detection and quantification of proteins in clinical samples using high resolution mass spectrometry. Methods, 2015. 81: p. 15-23.

Chapter 2

A Protocol for the Preparation of Cryoprecipitate and Cryo-depleted Plasma for Proteomic Studies

Rosemary L. Sparrow, Richard J. Simpson, and David W. Greening

Abstract

Cryoprecipitate is a concentrate of high-molecular-weight plasma proteins that precipitate when frozen plasma is slowly thawed at 1–6 °C. The concentrate contains factor VIII (antihemophilic factor), von Willebrand factor (vWF), fibrinogen, factor XIII, fibronectin, and small amounts of other plasma proteins. Clinical grade preparations of cryoprecipitate are mainly used to treat fibrinogen deficiency caused by acute bleeding or functional abnormalities of the fibrinogen protein. In the past, cryoprecipitate was used to treat von Willebrand disease and hemophilia A (factor VIII deficiency), but the availability of more highly purified coagulation factor concentrates or recombinant protein preparations has superseded the use of cryoprecipitate for these coagulopathies. Cryo-depleted plasma ("cryosupernatant") is the plasma supernatant remaining following removal of the cryoprecipitate from frozen-thawed plasma. It contains all the other plasma proteins and clotting factors present in plasma that remain soluble during cold-temperature thawing of the plasma. This protocol describes the clinical-scale preparation of cryoprecipitate and cryo-depleted plasma for proteomic studies.

Key words Plasma, Cryoprecipitate, Antihemophilic factor, Factor VIII, Cryo-depleted, Fibrinogen, Proteomics, Cryosupernatant

1 Introduction

Cryoprecipitate (also known as cryoprecipitated antihemophilic factor) was first described in the mid-1960s as a method to concentrate factor VIII (antihemophilic factor) from plasma for use in patients with hemophilia, von Willebrand disease, or hypofibrinogenemia [1, 2]. Cryoprecipitate is the insoluble concentrate of high-molecular-weight plasma proteins that precipitate when frozen plasma is slowly thawed at 1–6 °C [2]. Cryoprecipitate is enriched for plasma coagulation proteins, in particular factor VIII, fibrinogen, von Willebrand factor, factor XIII, and fibronectin. Small amounts of other plasma proteins, such as immunoglobulins, may also be present.

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_2,

Cryoprecipitate is a standard blood transfusion component manufactured by most blood transfusion services [3]. Clinically, the main indications for use of cryoprecipitate are for the treatment of fibrinogen deficiency (hypofibrinogenemia), caused by significant bleeding due to trauma, massive transfusion or disseminated intravascular coagulation, or dysfibrinogenemia arising from functional abnormal fibrinogen [3, 4]. Previously, cryoprecipitate was used to treat hemophilia and von Willebrand disease, but with the advent of specific coagulation factor concentrates, it has been relegated to second-line therapy for these diseases. Over the years, attempts have been made to improve the yield in cryoprecipitate by the use of techniques such as thaw-siphon [5] and the effect of temperature freezing and thawing [6] or by the use of various additives, such as heparin [7] and sodium citrate [8, 9].

Cryo-depleted plasma (cryosupernatant) is the plasma supernatant remaining following removal of the precipitated cold-insoluble proteins (i.e., cryoprecipitate) from frozen-thawed plasma. It is therefore significantly depleted of fibrinogen, factor VIII, von Willebrand factor, factor XIII, and fibronectin but contains all the other plasma proteins and clotting factors in similar concentrations as the original plasma. Clinically cryo-depleted plasma is used for plasma exchange in thrombotic thrombocytopenic purpura and in situations requiring rapid temporary reversal of warfarin anticoagulant therapy [10].

In this protocol, the preparation of small research-scale cryoprecipitate and cryo-depleted plasma is outlined and is based on the procedures used by blood transfusion services for the preparation of clinical-scale cryoprecipitate [10].

2 Materials

2.1 Blood Collection, Plasma Preparation, and Storage

1. Whole blood collection tubes containing citrate anticoagulant (e.g., BD Vacutainer 4.5 mL tube with 0.5 mL 3.2% sodium citrate anticoagulant, BD Biosciences #366415; BD Vacutainer 8.5 mL tube with 1.5 mL acid citrate dextrose (ACD) Sol A anticoagulant (22.0 g/L trisodium citrate, 8.0 g/L citric acid, and 24.5 g/L dextrose), BD Biosciences #364606) (*see* **Note 1**).
2. Blood collection needles compatible with the blood collection tubes (e.g., BD Vacutainer® Safety-Lok™ Blood Collection Set #367283; 23G butterfly needle with attached sterile tubing).
3. Alcohol and swabs for disinfection of the venipuncture site.
4. Personal protective equipment, gloves, gown, and eye safety glasses.
5. Disposal container for biological hazards.
6. Polypropylene tubes (1.5 mL, 15 mL).

7. Labels for blood sample tubes.
8. Tube storage rack.
9. Centrifugation unit (with swing-bucket rotor, compatible with 1.5/15 mL tubes, programmable temperature setting; range 4–25 °C).
10. Pipettes.
11. Freezer (−20 °C or lower).

2.2 Cryoprecipitate/ Cryo-depleted Plasma Preparation

1. Refrigerator or water bath set at 4 ± 2 °C.
2. Centrifuge, refrigerated (swing-bucket rotor, compatible with 1.5/15 mL tubes, programmable temperature setting).
3. 0.9% saline solution.

3 Methods

3.1 Blood Collection/ Phlebotomy

1. Blood collection must only be performed by personnel trained in phlebotomy/venipuncture. Safety precautions for the collection and handling of blood must be employed at all times (*see* **Note 2**). Particular care must be taken with insertion of the needle into the vein to limit the possibility of activation of the coagulation factors in the blood, which could compromise the quality of the blood sample.
2. It is important to collect the volume of blood specified for the particular type of blood collection tube to ensure the correct blood/anticoagulant ratio is achieved (*see* **Note 3**).
3. After blood collection, gently mix the blood by inverting the tube several times to ensure thorough mixing with the anticoagulant. For thorough mixing of blood collected into citrate tubes, it is recommended to invert the tube 3–4 times, while ACD tubes should be inverted eight times.
4. Blood samples should be maintained at temperate conditions (i.e., 20–24 °C) and centrifuged within 4 h of blood collection. Superior factor VIII yields are obtained from blood that is maintained at 20–24 °C before processing.
5. To separate the plasma, centrifuge the blood samples at 1200 × *g* for 10 min at 22 °C. If needed, RCF for a centrifuge can be calculated. For an online calculator tool, refer to *http://www.currentprotocols.com/tools/g-forcerpm-conversion-tool*.
6. After centrifugation, the plasma layer will be the upper layer of the separated blood, and the cellular fractions are the lower layers. The plasma should be a clear, straw-yellow-colored fluid (*see* **Note 4**). Mononuclear cells and platelets form a thin whitish layer (buffy coat) that settles directly on top of the red blood cell layer.

7. Carefully collect the plasma layer with an appropriate transfer pipette without disturbing the buffy coat layer. Do not attempt to collect all the plasma. Do not allow the tip of the transfer pipette within 5 mm of the buffy coat layer, and avoid touching the wall of the tube with the pipette. This helps to avoid inadvertent contamination of the plasma with cells that may only be softly sedimented in the buffy coat-plasma interface (*see* **Note 4**). If more than one tube of blood is collected from the same donor, pool the plasma samples from both tubes into a 15 mL polypropylene tube. If necessary, aliquot plasma into smaller volumes. A practical minimum volume is 1–1.5 mL.
8. Close the tube tightly and place on ice or immediately freeze by placing in the freezer. This process should be completed within 30 min of centrifugation. Plasma should be frozen as quickly as possible to minimize loss of labile coagulation factors, such as factor VIII. Frozen plasma should be stored at below −20 °C.

3.2 Cryoprecipitation

1. For the preparation of cryoprecipitate, remove tube(s) of frozen plasma from the freezer, and immediately place in a thermostatically controlled water bath or refrigerator set at 1–6 °C. Slowly thaw the plasma until it becomes "slushy" (required time will depend on volume of plasma being thawed) (*see* **Note 5**). Optimum temperature for cryoprecipitate formation is 3 °C.
2. Immediately sediment precipitated proteins in a refrigerated centrifuge (1–6 °C) at 5000 × *g* for 15 min. A white precipitate should be evident in the bottom of the tube.
3. Carefully remove the supernatant (*Note*: this is the cryo-depleted plasma). If this cryo-depleted plasma is required, aliquot into separate polypropylene tube(s). Leave a small amount of plasma above the deposited cryoprecipitate (5–10% v/v, 50–100 μL for 9–10 mL blood collection volume). If the cryoprecipitate or cryo-depleted plasma is not required immediately, freeze at −20 °C (*see* **Note 6**).

3.3 Thawing of Cryoprecipitate/Cryo-depleted Plasma

1. Thaw cryoprecipitate or cryo-depleted plasma in a water bath at 30–37 °C. The cryoprecipitate should be evenly dissolved at warming temperature.
2. Cryoprecipitate can be suspended in diluent, such as 0.9% saline, at 20–24 °C. Cryoprecipitate should be maintained at 20–24 °C and used within 6 h of thawing. Progressive functional decline of labile proteins, such as factor VIII, occurs following thawing (*see* **Notes 7–10**).
3. Thawed cryo-depleted plasma can be maintained at 2–6 °C.

4 Notes

1. *Anticoagulant.* For all physiological coagulation studies, sodium citrate and ACD are the anticoagulants of choice. Cryoprecipitate prepared by blood transfusion services for clinical use is prepared from whole blood collected into citrate-phosphate-dextrose (CPD) anticoagulant or from plasma collected into ACD anticoagulant by apheresis [11]. Cryoprecipitate prepared from CPD-anticoagulated plasma has been shown to give a higher yield of factor VIII compared to ACD-anticoagulated plasma [11]. The preparation of cryoprecipitate from plasma anticoagulated with non-citrate anticoagulants (e.g., heparin) may yield a different profile of precipitated proteins [12].
2. *Safety.* All blood and biological specimens and materials coming in contact are considered biohazards. Use gloves, gowns, eye protection, other personal protective equipment, and controls to protect from blood splatter, blood leakage, and potential exposure to blood-borne pathogens. Use aseptic technique and sterile disposables (tubes, pipettes) throughout to prevent contamination of the blood. Risk factors for possible transfusion transmissible infections should be rigorously screened prior to blood collection. Handle as if capable of transmitting infection, and dispose of with proper precautions in accordance with federal, state, and local regulations. Refer to your institutional regulations regarding the screening of blood for specific infectious disease markers (i.e., HIV, hepatitis B, hepatitis C, etc.). Discard all blood collection materials in biohazard containers approved for their disposal.
3. *Evacuated blood collectiontubes* (e.g., BD Vacutainers) are manufactured to draw the blood volume specified for the particular tube. Filling is complete when vacuum no longer continues to draw blood into the tube. Partially filled tubes will not have the correct blood/anticoagulant ratio and should not be used for the purpose of plasma sample preparation. If a citrated blood sample is the first sample to be collected from the donor, it is important to first collect a small volume of blood into a discard tube. This ensures that the "dead" volume in the needle/tubing set is filled with blood prior to the citrated tube being connected, thus ensuring the correct blood volume is drawn into the tube.
4. *Quality control of plasma sample.* It is recommended that following the preparation of the plasma sample, the quality of the plasma is monitored for hemolysis, clarity, and contamination with cells and/or platelets. Manual or automated cell counting methods should be validated by your institution or testing laboratory.

Normal plasma should be straw yellow in color, and the clarity should be relatively clear. Plasma from female donors who are taking oral contraceptives can have a green coloration and is considered normal [13]. Pink or red coloration of the plasma is indicative of hemolysis of red blood cells and is not normal [14]. Hemolysis could be due to poor collection or handling of the blood sample or due to donor-related medical factors. Milky opaque plasma is due to raised lipid content [15]. Plasma lipids can be transiently raised in a healthy individual due to the recent consumption of a high-fat meal (postprandial-induced lipemia). Except in the instance of postprandial-induced lipemia, very opaque plasma is not normal. For protein chemistry studies of normal plasma, samples with high lipid content or hemolysis should be avoided. Plasma that has been previously thawed and refrozen will give inferior yields if used for the preparation of cryoprecipitate.

5. *How to prevent cryoprecipitation when thawing plasma.* If plasma is to be used for coagulation studies, it is important to prevent cryoprecipitation from occurring. To avoid cryoprecipitation of the cold-insoluble proteins when thawing replete plasma, plasma must be thawed quickly at 37 °C. This can be achieved by placing the frozen plasma samples in a 37 °C water bath or dry heating system set at 37 °C. Such equipment must be maintained and kept clean to avoid inadvertent bacterial contamination of plasma samples.
6. *Cryoprecipitate and cryo-depleted plasmastorage.* The prepared cryoprecipitate and cryo-depleted plasma can be used up to 12 months when stored at or below −18 °C [3]. When cryoprecipitate is thawed and held at 20–24 °C, the functional levels of non-labile proteins, such as fibrinogen and factor XIII, remain stable for up to 72 h, while labile proteins such as factor VIII decline within 24 h [16].
7. *Cryoprecipitate content and specifications.* According to blood transfusion guidelines, a unit of cryoprecipitate prepared from the plasma of a standard 450–500 mL CPD-anticoagulated whole blood donation should contain at least 150 mg of fibrinogen and a minimum of 80 international units (IU) of factor VIII [3, 10, 11, 17]. This equates to 30–70% of the factor VIII/vWF and fibrinogen content of the original plasma. Proportionally similar yields should be achieved when cryoprecipitate is prepared from smaller starting volumes of plasma providing care is taken with the processing, freezing, and thawing of the plasma samples as described in this protocol.
8. *Effect of ABO blood group.* Plasma/cryoprecipitate from blood group O individuals has lower levels of factor VIII and vWF than A, B, or AB blood groups.

9. *Effect of additional treatment steps on coagulation factor proteins.* Some blood transfusion services supply plasma units that have undergone additional treatment steps to further minimize the already very low risk of infectious disease transmission or adverse reactions by blood transfusion. Various pathogen reduction technologies have been developed, including treatments with solvent/detergent or photoactivating agents such as methylene blue, riboflavin, or psoralen [18]. Removal of leukocytes by filtration before processing of whole blood is commonly performed by blood transfusion services. A trade-off for the increased safety rendered by these treatments is a reduced yield and activity of plasma coagulation factors, including fibrinogen and factor VIII [18–20]. The nature of any biochemical changes that occur to the coagulation proteins following pathogen reduction treatment is yet to be fully determined.
10. *Extracellular vesicles.* Cryoprecipitate has been reported to contain a significant enrichment of extracellular vesicles (also called microparticles or microvesicles) present in normal plasma [21] (Chan and Sparrow, unpublished findings). Moreover, high levels of extracellular vesicles in cryoprecipitate may contribute to its therapeutic effects in bleeding patients [21]. It is yet to be determined whether extracellular vesicles in cryoprecipitate are biologically functional, such as in hemostasis, inflammation, and/or allo-immunoreactivity [22].

References

1. Pool JG, Gershgold EJ, Pappenhagen AR (1964) High-potency Antihaemophilic factor concentrate prepared from Cryoglobulin precipitate. Nature 203:312
2. Pool JG (1965) Preparation and testing of antihemophilic globulin (factor 8) sources for transfusion therapy in hemophilia. Description of a new sterile concentrate process for blood banks. Scand J Clin Lab Invest 17(84):70–77
3. Callum JL, Karkouti K, Lin Y (2009) Cryoprecipitate: the current state of knowledge. Transfus Med Rev 23:177–188
4. Erber WN, Perry DJ (2006) Plasma and plasma products in the treatment of massive haemorrhage. Best Pract Res Clin Haematol 19: 97–112
5. Mason EC (1978) Thaw-siphon technique for production of cryoprecipitate concentrate of factor VIII. Lancet 2:15–17
6. Rock GA, Tittley P (1979) The effects of temperature variations on cryoprecipitate. Transfusion 19:86–89
7. Rock GA, Cruickshank WH, Tackaberry ES, Palmer DS (1979) Improved yields of factor VIII from heparinized plasma. Vox Sang 36: 294–300
8. Shanbrom E, Owens WJ (2001) A novel method to enhance the formation of cryoprecipitate from fresh plasma. Blood 98:60a
9. Yousef H, Neurath D, Freedman M, Rock G (2006) Cryoprecipitate production: the use of additives to enhance the yield. Clin Lab Haematol 28:237–240
10. Fung MK, Grossman BJ, Hillyer CD, Westhoff CM, American Association Of Blood Banks (2014) Technical manual, 18th edn. American Association of Blood Banks, Bethesda, MD
11. Wensley RT, Snape TJ (1980) Preparation of improved cryoprecipitated factor VIII concentrate. A controlled study of three variables affecting the yield. Vox Sang 38:222–228
12. Meilin E, Sela S, Kristal B (2008) Heparin cryoprecipitation reduces plasma levels of non-traditional risk factors for atherosclerosis in vitro. Blood Purif 26:238–248
13. Elkassabany NM, Meny GM, Doria RR, Marcucci C (2008) Green plasma-revisited. Anesthesiology 108:764–765

14. Brunel V, Larson T, Peschanski N, Cauliez B (2012) Evaluation of haemolysis in emergency department samples requesting high sensitivity troponin T measurement. Ann Clin Biochem 49:509–510
15. Peffer K, de Kort WL, Slot E, Doggen CJ (2011) Turbid plasma donations in whole blood donors: fat chance? Transfusion 51:1179–1187
16. Green L, Backholer L, Wiltshire M, Platton S et al (2016) The hemostatic properties of thawed pooled cryoprecipitate up to 72 hours. Transfusion 56:1356–1361
17. Levitt J (2014) Standards of blood banks and transfusion services, 23rd edn. American Association of Blood Banks, Bethesda, MD, p 28
18. Prowse C (2009) Properties of pathogen-inactivated plasma components. Transfus Med Rev 23:124–133
19. Coene J, Devreese K, Sabot B, Feys HB et al (2014) Paired analysis of plasma proteins and coagulant capacity after treatment with three methods of pathogen reduction. Transfusion 54:1321–1331
20. Chan KS, Sparrow RL (2014) Microparticle profile and procoagulant activity of fresh-frozen plasma is affected by whole blood leukoreduction rather than 24-hour room temperature hold. Transfusion 54:1935–1944
21. George JN, Pickett EB, Heinz R (1986) Platelet membrane microparticles in blood bank fresh frozen plasma and cryoprecipitate. Blood 68:307–309
22. Morel O, Morel N, Jesel L, Freyssinet JM et al (2011) Microparticles: a critical component in the nexus between inflammation, immunity, and thrombosis. Semin Immunopathol 33:469–486

Chapter 3

Preparation of Platelet Concentrates for Research and Transfusion Purposes

David W. Greening, Richard J. Simpson, and Rosemary L. Sparrow

Abstract

Platelets are specialized cellular elements of the blood that play central roles in physiologic and pathologic processes of hemostasis, wound healing, host defense, thrombosis, inflammation, and tumor metastasis. Activation of platelets is crucial for platelet function that includes a complex interplay of adhesion, signaling molecules, and release of bioactive factors. Transfusion of platelet concentrates is an important treatment component for thrombocytopenia and bleeding. Recent progress in high-throughput mRNA and protein profiling techniques has advanced the understanding of platelet biological functions toward identifying novel platelet-expressed and secreted proteins, analyzing functional changes between normal and pathologic states, and determining the effects of processing and storage on platelet concentrates for transfusion. It is important to understand the different standard methods of platelet preparation and how they differ from the perspective for use as research samples in clinical chemistry. Two simple methods are described here for the preparation of research-scale platelet samples from whole blood, and detailed notes are provided about the methods used for the preparation of platelet concentrates for transfusion.

Key words Platelet concentrate, Platelet rich plasma, Buffy coat, Apheresis, Transfusion, Proteomics, Research, Protocol, SOP

1 Introduction

Platelets, the smallest of the human blood cellular elements (~3.6 × 0.7 μm), are central players in hemostasis and thrombosis. In addition, platelets are involved in clot retraction, vessel constriction and repair, inflammation including promotion of atherosclerosis, host defense, and even tumor growth/metastasis [1–3]. They are produced by differentiation of the bone marrow-derived megakaryocytes and are released as anucleated fragments into the circulation [4]. Being anuclear, platelets are not cells in the strict meaning of the word, although they are often referred to as blood cells. Circulating platelets have a discoid shape; are the second most numerous cellular element, normally circulating between 150 and 450 × 10^9/L; and have the lowest specific gravity of

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_3, © Springer Science+Business Media LLC 2017

formed blood cells [1, 2]. Their shape and small size, combined with blood flow rheology, allow platelets to access the edge of blood vessels, thereby enabling them constantly to survey vascular integrity. Complex, regulated reactions occur between platelets, von Willebrand factor, collagen, and soluble coagulation factors in regions of disturbed vasculature. These changes induce platelet adherence to vessel walls and platelet activation, which leads to platelet aggregation, procoagulant activity, spreading, microparticle release, and formation of a primary hemostatic plug [1, 2].

Since platelets lack nuclear DNA and their genome consists of a subset of megakaryocyte-derived mRNA transcripts, they represent a simplified biological model when investigating cell function [4]. While valuable information may be gathered from studies of mRNA [5, 6], the rapid signaling and regulatory events in platelets are not governed by, or dependent on, alterations in gene expression [7]. In contrast, proteomics can provide the necessary tools for characterizing, at the protein level, the vicissitudes of platelet function. Most proteomic efforts can be grouped into several distinct categories: (1) cataloging the spectrum of proteins that comprise the normal, resting (quiescent) platelet proteome, (2) characterizing proteins released from activated (functional) platelets, and (3) identifying specific platelet sub-proteomes (reviewed in [8]) (i.e., membrane [9, 10], granules [11], phosphorylation [12, 13], or functional endpoints in response to external stimuli (i.e., microparticles [14, 15] and releasate [16, 17]), as well as in applications for transfusion [18–20] and disease [21–23]).

The separation of platelets from whole blood is based on the differential densities of the various cellular elements when blood is subjected to defined centrifugation forces. Platelets, being the smallest and lightest cellular elements of blood, remain suspended in the liquid plasma when whole blood is centrifuged at a low centrifugal force. Protocols for the preparation of platelets rely on this characteristic, including those used by blood transfusion services for the preparation of platelet concentrates (PCs) for clinical use. Transfusion of PCs is indicated for the treatment of thrombocytopenia and bleeding, caused by hematological disease, the effects of chemotherapy, and postoperative bleeding.

PCs for transfusion can be prepared by three different methods including (1) platelet-rich plasma-platelet concentrates (PRP-PC), (2) buffy coat-platelet concentrates (BC-PC), and (3) apheresis-platelet concentrates (apheresis-PC). For the preparation of PRP-PC, an initial soft centrifugation produces PRP, which is separated from white cells and red cells, and the PRP is centrifuged at a higher *g* force to pellet the platelets. This methodology is typically used for preparing research platelet samples from smaller volumes of whole blood. Disadvantages of this method are that it is difficult to avoid aspirating some white cells and red cells with the PRP [9], and the platelets are hard spun against the surface of the

container, which can cause increased platelet activation and/or damage. In contrast, platelets prepared by the BC-PC method are not subjected to being pelleted, but rather are cushioned by the red cells and then centrifuged at much lower centrifugation force to suspend the platelets in the supernatant, consequently limiting the platelet pelleting and reducing the level of cellular contaminants [24]. For apheresis-PC, a specialized automated blood cell separator is used whereby the blood donor is directly connected to the machine and whole blood is drawn, immediately mixed with anticoagulant and separated into components; the target component is collected into a separate bag, while the other components are returned to the donor [25].

In developed countries, blood for transfusion is collected, processed, and distributed by licensed blood transfusion services that operate in a highly regulated environment governed by strict codes of practice similar to those that apply to the manufacturers of medicinal products (i.e., the code of good manufacturing practice, cGMP). These regulations are designed to ensure standardization of all procedures to maximize quality and safety of blood transfusion components, such as PCs [26].

For this protocol, two research-scale methods for the preparation of platelets from whole blood are described that yield a minimally manipulated platelet specimen suitable for use in proteomic analysis and other research applications. An overview of the methods used by blood transfusion services to prepare PCs from whole blood donations and by single-donor apheresis collection is detailed in the Subheading 4.

2 Materials

2.1 Blood Collection and Platelet Sample Preparation

1. Acid citrate dextrose (ACD) blood collection tubes (e.g., Whole Blood Glass Vacutainer® Tube with Anticoagulant, ACD Sol A, 8.5 mL, BD Biosciences #364606) (*see* **Note 1**).
2. Alcohol and swabs for disinfection of the venipuncture site.
3. Gloves, gown, and eye protection.
4. Disposal container for biological hazards.
5. Polypropylene tubes for processing and storage (15 mL).
6. Tube rack.
7. Labels for blood sample tubes.
8. Transfer pipettes and wide aperture.
9. Centrifuge (swing bucket rotor, compatible with 15 mL tubes, programmable temperature setting, set at 22 °C).
10. Platelet wash buffer (*see* **Note 2**).

3 Methods

3.1 Blood Collection

1. For the preparation of normal, resting platelets, the blood donor must be healthy, with no signs of infection or inflammation, and must not have taken medications that have antithrombotic (anticlotting) or anti-inflammatory effects, such as aspirin or ibuprofen.
2. Blood collection must only be performed by personnel trained in phlebotomy/venipuncture. Safety precautions for the collection and handling of blood must be employed at all times (*see* **Note 3**). Particular care must be taken with insertion of the needle into the vein to limit the possibility of activation of the hemostasis/coagulation system, which could compromise the quality of the blood sample. For routine venipuncture procedures, a 21 G needle is recommended to minimize shear stress upon the collected blood specimen.
3. It is important to collect the volume of blood specified for the particular blood collection tube to ensure the correct blood/anticoagulant ratio is achieved (*see* **Note 4**).
4. After blood collection, gently invert the tube 3–4 times to ensure thorough mixing with the ACD anticoagulant.
5. Blood samples must be maintained at temperate conditions (i.e., 20–24 °C) and platelet preparations prepared within 4 h of blood collection.

3.2 Platelet-Rich Plasma (PRP) Preparation

1. Platelets are extremely labile and are very easily activated during sample preparation. It is important to limit the extent of manipulation of the blood sample to avoid unintentional activation of the platelets. All procedures must be performed at 20–24 °C to maintain platelet quality and viability. Platelets undergo cold storage-induced activation if subjected to temperatures below 20 °C [27, 28]. The temperature of all equipment (e.g., centrifuge) and wash buffers (if used) must be between 20 and 24 °C prior to use.
2. The PRP is separated from whole blood (typical volume 8.5 mL) by light spin centrifugation at 110 × *g* for 15 min at 22 °C, without brake. Under these centrifugation conditions, the platelets will remain suspended in the plasma (upper fraction, yellow-colored fluid), while the white cells and red cells will be softly sedimented in the lower fraction. Using a wide-aperture transfer pipette, carefully collect only the upper 40% of the PRP to limit contamination by white cells and red cells. Do not let the tip of the pipette touch the sides of the tube. Place the PRP in a fresh tube. Take a small aliquot of the collected PRP sample to use for quality assessment, which should be performed as quickly as possible (*see* **Note 5**).

3. To sediment the platelets in the PRP sample, centrifuge at a higher centrifugation rate (1000 × *g* for 15 min, 22 °C. without brake). Carefully remove the plasma supernatant and discard, if not required (alternatively the plasma can be stored frozen at < −20 °C). If necessary, the platelets can be washed with an appropriate physiological buffer to remove residual plasma proteins (*see* **Note 2**), using the same centrifugation conditions, but care must be taken with the washing and resuspension steps to limit activation of the platelets (*see* **Note 6**).
4. The platelets should be used immediately upon isolation. For example, prepare a platelet lysate using a relevant protocol suitable for subsequent proteomic analysis. One 8.5 mL tube of whole blood from a donor with a normal platelet count (i.e., 150–450 × 10^9 platelets/L) should be sufficient to yield enough platelets to prepare a sample for proteomic analysis.

3.3 Buffy Coat (BC) Platelet Preparation

1. Fractionate whole blood by centrifuging at 2000 × *g* for 10 min at 22 °C, without brake. This will separate the blood into an upper plasma layer, a lower red cell layer, and a thin white-colored interface called the BC, containing the majority of the platelets and white cells.
2. Carefully aspirate the plasma with a transfer pipette, being careful not to disturb the BC layer. Discard the plasma if not required, or place in a separate clean tube, and store frozen at < −20 °C for other purposes.
3. Using a wide-aperture transfer pipette, carefully aspirate the BC layer. Aspirate slowly, using a circular motion, to pull all the BC layer into the transfer pipette. Some contamination with the underlying red cells is unavoidable, although care should be taken to minimize the amount of red cells collected.
4. Suspend the BC in at least 5 mL of platelet wash buffer (*see* **Note 2**).
5. Centrifuge the BC suspension using a light centrifugation at 520 × *g* for 6 min at 22 °C, without brake. This sediments the white cells and red cells and leaves the platelets in suspension.
6. Collect the platelet-rich supernatant with a clean transfer pipette. Be careful not to disturb the cell layer, which contains the white cells and red cells. Only collect the yellow-colored supernatant. Do not allow the tip of the pipette to touch the sides of the tube. Place the supernatant into a fresh tube.
7. Take a small aliquot of the collected platelet sample to use for quality assessment, which should be performed as quickly as possible (*see* **Note 5**).
8. To sediment the platelets, if necessary, centrifuge at a higher centrifugation rate (1000 × *g* for 15 min, 22 °C, without

brake). Carefully remove the supernatant and discard. If necessary, the platelets can be washed with an appropriate physiological buffer to remove residual plasma proteins (*see* **Note 2**), using the same centrifugation conditions, but care must be taken with the washing and resuspension steps to limit activation of the platelets (*see* **Note 6**).

9. The platelets should be used immediately upon isolation. For example, prepare a platelet lysate using a relevant protocol suitable for subsequent proteomic analysis. One 8.5 mL tube of whole blood from a donor with a normal platelet count (i.e., 150–450 × 10^9 platelets/L) should be sufficient to yield enough platelets to prepare a suitable sample for proteomic analyses.

3.4 How Blood Transfusion Services Prepare PCs

For an overview of the procedures used by blood transfusion services to prepare PCs, *see* **Note 7** (PRP-PC and BC-PC), **Note 8** (apheresis-PC), and **Notes 9–14** for details of additional processes.

4 Notes

1. *Anticoagulant.* ACD is the preferred anticoagulant for preparation of platelets as it preserves cell quality and function. The ACD used in the blood collection tubes suggested here (i.e., #364606, 8.5 mL, BD Biosciences) contains 22.0 g/L trisodium citrate, 8.0 g/L citric acid, and 24.5 g/L dextrose (known as ACD Solution A), which is similar to the ACD-A formulation used by blood transfusion services for the preparation of apheresis-platelet concentrates [29]. CTAD (citrate, theophylline, adenosine, dipyridamole) is an alternative citrate-based anticoagulant that contains inhibitors of platelet activation (theophylline and dipyridamole) and may be useful if platelet activation must be minimized [30]. CTAD vacutainer tubes are available from BD Biosciences (#367947, 4.5 mL).
2. *Platelet wash buffer.* Tyrode's salt solution is a suitable physiological buffer for washing platelets: 8.0 g/L NaCl, 0.214 g/L $MgCl_2$•$6H_2O$, 0.2 g/L KCl, 1.0 g/L $NaHCO_3$, 1.0 g/L glucose, and pH 7.2. Use freshly prepared buffer, warmed to 20–24 °C. If required, an inhibitor of platelet activation can be added to the wash buffer immediately prior to use. A suitable inhibitor is 0.02 U/mL of apyrase (adenosine 5′-triphosphate diphosphohydrolase, EC 3.6.1.5) (final concentration) (e.g., apyrase, ATPase ≥200 units/mg protein, lyophilized powder, premium quality, Sigma-Aldrich #A6535).
3. *Safety.* All blood and biological specimens and materials coming in contact are considered biohazards. Use gloves, gowns, eye protection, other personal protective equipment, and con-

trols to protect from blood splatter, blood leakage, and potential exposure to blood-borne pathogens. Use aseptic technique and sterile disposables (tubes, pipettes) throughout to prevent contamination of the blood. Risk factors for possible transfusion transmissible infections should be rigorously screened for prior to blood collection. Handle as if capable of transmitting infection and dispose of with proper precautions in accordance with federal, state, and local regulations. Refer to your institutional regulations regarding the screening of human blood for specific infectious disease markers (i.e., HIV, hepatitis B, hepatitis C, etc.). Discard all blood collection materials in biohazard containers approved for this purpose.

4. *Evacuated blood collectiontubes* (e.g., BD Vacutainers®) are manufactured to draw the blood volume specified for the particular tube. Filling is complete when vacuum no longer continues to draw blood into the tube. Partially filled tubes will not have the correct blood/anticoagulant ratio and should not be avoided for the purpose of platelet sample preparation.
5. *Platelet sample quality assessment*. The PRP sample should be assessed to determine the platelet count as well as the level of contaminating white cells and red cells. This is best performed on an automated hematology analyzer. If access to a hematology analyzer is not possible, a qualitative assessment can be performed by examining the sample by light microscopy for the presence of contaminating white cells and red cells, which are considerably larger in size than platelets (i.e., white cells, 8–15 μm diameter; red cells, 6–8 μm diameter; platelets, 2–4 μm diameter). Other platelet quality parameters that could be considered include assessment of the expression of CD62P as a marker of platelet activation [31]. An important note is to process the platelet sample as quickly as possible and minimize transition and storage time prior to quality assessment. Delayed assessment can affect the accuracy of the results, such as platelet yield, viability, and level of platelet activation. Temperatures below room temperature can cause cold-induced platelet activation [27, 28].
6. *Washing and resuspension of pelleted platelets.* Platelets are very delicate and are easily damaged or activated by handling. Avoid strong forces such as rapid pipetting or vigorous mixing. As with any form of cell washing, a proportion of platelets will be lost and there is an increased risk of platelet activation. To remove residual plasma from the pelleted platelets, the gentlest wash procedure is simply to rinse the platelet pellet without resuspension of the platelets. Slowly and gently trickle the wash buffer down the side of the tube to avoid disturbing the pellet. If more thorough washing of the platelets is required, very gently and slowly resuspend the platelets in the wash buffer

using a wide-aperture transfer pipette. Freshly resuspended platelets have a tendency to remain in clumps, but the clumps should disperse when the suspension is allowed to rest for a short time at room temperature, providing the platelets have not become activated. Centrifuge at 1000 × *g* for 15 min, 22 °C, without brake. Repeat wash step once, if required.

7. *Whole blood-derived, pooled-donor platelet transfusion components.* The following is a brief description of the procedures used by blood transfusion services to prepare whole blood-derived PCs for clinical use. Whole blood is collected into a sterile soft-plasticized PVC blood collection bag system that consists of a series of interconnected bags (including a gas-permeable bag for the storage of platelets) joined by flexible tubing, which enables closed system (aseptic) processing. Whole blood (450 mL ± 10%) is collected into the primary collection bag containing 63 mL (± 10%) citrate-phosphate-dextrose (CPD) anticoagulant or alternative formulations such as CPDA-1 and CP2D [26]. Blood collection must be completed within 12 min. Longer collection times are indicative of poor venous access, collapsed vein, or insufficient hydration of the donor, all of which can adversely affect the quality of the collected blood. Whole blood for the preparation of PCs is maintained at 20–24 °C and must be processed within 24 h of collection (an 8 h max limit applies in the USA). Two different methods can be used to prepare PCs from whole blood: the PRP method or the BC-PC method. The PRP method is used in the USA, while the BC-PC method is used in many countries in Europe, the UK, Canada, and Australia. The BC-PC method can be used in conjunction with synthetic protein-free platelet storage solutions as an alternative to plasma as the suspending fluid [32]. Further detail about methods used for the preparation of PCs is reviewed by Vassallo and Murphy [25].

   For the PRP method, the whole blood donation is centrifuged by a soft spin (2000 × *g*, 3 min at 22 °C) to sediment the white cells and red cells, and the PRP is expressed into an attached separate platelet storage bag using a blood bag pressing device. Typically, this process will also include passing the PRP through a specialized leukocyte-depletion filter to remove the white cells, in a process known as pre-storage leukoreduction. The PRP is recentrifuged using a heavier spin (5000 × *g*, 5 min at 22 °C) to sediment the platelets, and the excess plasma is expressed into an attached separate storage bag, leaving behind approximately 40–70 mL of plasma in which the platelets are resuspended. Typically, a unit of platelets from a single whole blood donation should contain approximately 5.5×10^{10} platelets. Individual donor platelet units are stored on a reciprocating platform agitator (50–70 oscillations/min) at 20–24 °C. In order to prepare a therapeutic dose of PRP-PC

($>3 \times 10^{11}$ platelets), 4–6 individual donor PRP units must be pooled. Pooling is done either 24 h after preparation of the PRP and subsequently stored for use within 5–7 days of collection, or pooling is performed just prior to issue to the patient.

For the BC-PC method, whole blood is collected into a quadruple "top-bottom" bag system [24, 25]. The whole blood unit is centrifuged by a hard spin ($5000 \times g$ for 10 min at 22 °C) that sediments the platelets into the BC layer. The plasma and red cells are expressed from the top and bottom of the bag, respectively, into separate storage bags using a semi-automated blood component separator, which leaves behind the platelet-rich BC in the original collection bag. BCs are rested for several hours or overnight, and then concentrates from four to six donors are pooled and suspended in either plasma or platelet additive solution (~300 mL) and centrifuged by a soft spin ($520 \times g$, 6 min at 22 °C), which sediments the white cells and leaves the platelets in suspension. The pooled BC-PC is passed through a leukocyte-reduction filter into a 1.3 L gas-permeable platelet storage bag, which is then placed on a reciprocating platform agitator at 20–24 °C and stored for use within 5–7 days of collection.

8. *Apheresis, single-donor platelet transfusion components.* In contrast to the whole blood methods (described in **Note 7**), which require pooling of platelets from four to six donors to achieve a therapeutic dose of platelets ($>3 \times 10^{11}$ platelets), platelets collected by apheresis yield a therapeutic dose from a single donor [25]. Platelet collection by apheresis uses a specialized programmable instrument that processes over 3.5 L of the donor's blood in a sterile closed system, extracts the platelets into a gas-permeable platelet collection bag, and returns the remainder of the blood to the donor (sometimes with saline fluid replacement). Blood processed by apheresis is anticoagulated with ACD. Leukocyte reduction is incorporated into the automated processing protocol. Typically, apheresis platelets are suspended in plasma although protocols that use platelet additive solutions as the storage medium are becoming available. The procedure normally takes 1.5–2 h, with the final volume of apheresis-PC between 100 and 400 mL. Storage and shelf life of apheresis-PCs are the same as PRP and BC-PCs. The majority of the PCs produced in the USA are by apheresis collection, and it is increasingly being used in other Western countries.

9. *Storage of platelet transfusion components and storage effects.* PCs for transfusion are stored at 20–24 °C with continuous gentle horizontal agitation (reciprocating platform agitator, set at 50–70 oscillations/min) in specifically designed platelet

storage bags that permit O_2 and CO_2 exchange to maintain platelet viability and quality [33]. This combination of storage container, agitation, suspension medium (100% plasma or platelet additive solution/plasma mix), and temperature permits satisfactory preservation of platelets for up to 7 days [34]. Because PCs are stored at temperate conditions, there is an increased risk of bacterial growth that can occur for various reasons, most often being donor derived. Consequently this risk limits the storage time for PCs, which is typically 5–7 days from collection [33, 34].

During storage, platelets undergo numerous physicochemical changes, known as the platelet storage lesion, which reduces the quality, function, and viability of the stored platelets. These changes include platelet activation, shape change, release of bioactive factors and granules, shedding of microparticles, and apoptosis/necrosis [34].

10. *Leukofiltration of PCs.* Removal of white cells from PCs is achieved by filtration using specialized filters that reduce the residual white cells by the order of >4 $\log_{10}$ [35]. In the USA, residual white cells must be $<5 \times 10^6$/PC, while in Europe and other countries, the limit is $<1 \times 10^6$/PC. Pre-storage leukoreduction of PCs has been widely implemented to reduce the risk of adverse transfusion reactions, alloimmunization, and transmission of white cell-borne viruses [35].
11. *Pathogen reduction of PCs.* Strategies to minimize the risk of transmission of infectious diseases are a fundamental aspect of the blood banking industry [36]. Pathogen inactivation (PI) technologies that inactivate transfusion transmissible infectious agents, such as viruses and bacteria, and that are suitable for the treatment of PCs are being implemented by blood transfusion services in various jurisdictions. For PCs, treatments have been developed with Psoralen/ultraviolet-A light or Riboflavin/ultraviolet-B light [37]. Both products are CE marked in Europe, and the Psoralen/ultraviolet-A process has recently been approved for use for PCs and plasma in the USA. The treatments cause some reduction in platelet function and viability, but PI-treated PCs meet standard acceptance criteria for clinical use [38].
12. *Irradiation of PCs.* Gamma irradiation is used to inactivate residual white cells present in transfusion components. Irradiated blood components are prescribed for patients at risk of transfusion-associated graft-versus-host disease (TA-GvHD), including immunosuppressed patients or patients that are partially tissue matched (haploidentical) with the transfusion donor. Platelets are relatively unaffected by gamma irradiation at the prescribed dosage of 25–50 Gy routinely used to irradiate blood components. Consequently the shelf life of PCs is unchanged [39].

13. *Washing and cryopreservation of PCs.* In rare circumstances, PCs need to be washed prior to transfusion to avoid adverse transfusion reactions in hypersensitive recipients [40]. Platelets can be cryopreserved, typically using 5–6% dimethyl sulfoxide (DMSO) and frozen at either −80 °C or liquid nitrogen. Upon thawing and subsequent washing, there is significant loss of platelets and changes in platelet quality [41, 42]. Cryopreserved PCs are a specialized product and are not routinely available.
14. *PC quality assessment.* The standard indicator of PC quality is pH (at expiry, 6.4–7.4). A range of other tests can be used, including functional assays (e.g., extent of shape change, adhesion, aggregation, and thrombus formation/retraction), measures of platelet activation and viability (e.g., CD62P expression, swirling index, release of bioactive factors and granules, procoagulant activity, release of annexin V and microparticles, mitochondrial activity), and metabolic indices (e.g., pH, lactate, ATP content, hypotonic shock response) [43–45].

References

1. Goto S, Hasebe T, Takagi S (2015) Platelets: small in size but essential in the regulation of vascular homeostasis - translation from basic science to clinical medicine. Circ J 79(9): 1871–1881
2. McFadyen JD, Kaplan ZS (2015) Platelets are not just for clots. Transfus Med Rev 29(2): 110–119
3. Franco AT, Corken A, Ware J (2015) Platelets at the interface of thrombosis, inflammation, and cancer. Blood 126(5):582–588
4. Sim X et al (2016) Understanding platelet generation from megakaryocytes: implications for in vitro-derived platelets. Blood 127(10):1227–1233
5. Schubert S, Weyrich AS, Rowley JW (2014) A tour through the transcriptional landscape of platelets. Blood 124(4):493–502
6. Clancy L, Freedman JE (2015) The role of circulating platelet transcripts. J Thromb Haemost 13(Suppl 1):S33–S39
7. Howes JM (2013) Proteomic profiling of platelet signalling. Expert Rev Proteomics 10(4):355–364
8. Di Michele M, Van Geet C, Freson K (2012) Recent advances in platelet proteomics. Expert Rev Proteomics 9(4):451–466
9. Greening DW et al (2008) Comparison of human platelet-membrane cytoskeletal proteins with the plasma proteome: towards understanding the platelet-plasma nexus. Proteomics Clin Appl 2:63–77
10. Premsler T et al (2011) Phosphoproteome analysis of the platelet plasma membrane. Methods Mol Biol 728:279–290
11. Zufferey A et al (2014) Characterization of the platelet granule proteome: evidence of the presence of MHC1 in alpha-granules. J Proteome 101:130–140
12. Zahedi RP et al (2008) Phosphoproteome of resting human platelets. J Proteome Res 7(2):526–534
13. Zimman A et al (2014) Phosphoproteomic analysis of platelets activated by pro-thrombotic oxidized phospholipids and thrombin. PLoS One 9(1):e84488
14. Milioli M et al (2015) Quantitative proteomics analysis of platelet-derived microparticles reveals distinct protein signatures when stimulated by different physiological agonists. J Proteome 121:56–66
15. Shai E et al (2012) Comparative analysis of platelet-derived microparticles reveals differences in their amount and proteome depending on the platelet stimulus. J Proteome 76:287–296
16. Cini C et al (2015) Differences in the resting platelet proteome and platelet releasate between healthy children and adults. J Proteome 123:78–88

17. Velez P et al (2015) A 2D-DIGE-based proteomic analysis reveals differences in the platelet releasate composition when comparing thrombin and collagen stimulations. Sci Rep 5:8198
18. Dzieciatkowska M et al (2015) Proteomics of apheresis platelet supernatants during routine storage: gender-related differences. J Proteome 112:190–209
19. Kamhieh-Milz J et al (2016) Secretome profiling of apheresis platelet supernatants during routine storage via antibody-based microarray. J Proteome 150:74–85
20. Prudent M et al (2014) Proteome changes in platelets after pathogen inactivation—an interlaboratory consensus. Transfus Med Rev 28(2): 72–83
21. Donovan LE et al (2013) Exploring the potential of the platelet membrane proteome as a source of peripheral biomarkers for Alzheimer's disease. Alzheimers Res Ther 5(3):32
22. Garcia A (2016) Platelet clinical proteomics: facts, challenges, and future perspectives. Proteomics Clin Appl 10(8):767–773
23. Shah, P., et al., Platelet glycoproteins associated with aspirin-treatment upon platelet activation. Proteomics, 2016 9:27452734.
24. Murphy S (2005) Platelets from pooled buffy coats: an update. Transfusion 45(4):634–639
25. Vassallo RR, Murphy S (2006) A critical comparison of platelet preparation methods. Curr Opin Hematol 13(5):323–330
26. Greening DW et al (2010) International blood collection and storage: clinical use of blood products. J Proteome 73(3):386–395
27. Murphy S, Gardner FH (1969) Effect of storage temperature on maintenance of platelet viability--deleterious effect of refrigerated storage. N Engl J Med 280(20):1094–1098
28. Wood, B., et al., Refrigerated storage of platelets initiates changes in platelet surface marker expression and localization of intracellular proteins. Transfusion, 2016;56(10):2548–2559.
29. Greening DW et al (2009) Enrichment of human platelet membrane-cytoskeletal proteins for proteomic analysis. In: Pierce M (ed) Proteomic analysis of membrane proteins: methods and protocols. Methods in molecular medicine series. Humana, Louisville, KY, pp 245–258
30. Neufeld M, Nowak-Gottl U, Junker R (1999) Citrate-theophylline-adenine-dipyridamol buffer is preferable to citrate buffer as an anticoagulant for flow cytometric measurement of platelet activation. Clin Chem 45(11): 2030–2033
31. Curvers J et al (2008) Flow cytometric measurement of CD62P (P-selectin) expression on platelets: a multicenter optimization and standardization effort. Transfusion 48(7): 1439–1446
32. van der Meer PF (2016) PAS or plasma for storage of platelets? A concise review. Transfus Med 26(5):339–342
33. Thomas, S., Platelets: handle with care.. Transfus Med, 2016.
34. Smethurst PA (2016) Aging of platelets stored for transfusion. Platelets 27(6):526–534
35. Bassuni WY, Blajchman MA, Al-Moshary MA (2008) Why implement universal leukoreduction? Hematol Oncol Stem Cell Ther 1(2):106–123
36. Dodd RY (2012) Emerging pathogens and their implications for the blood supply and transfusion transmitted infections. Br J Haematol 159(2):135–142
37. Devine DV, Schubert P (2016) Pathogen inactivation technologies: the advent of pathogen-reduced blood components to reduce blood safety risk. Hematol Oncol Clin North Am 30(3):609–617
38. Kaiser-Guignard J et al (2014) The clinical and biological impact of new pathogen inactivation technologies on platelet concentrates. Blood Rev 28(6):235–241
39. Moroff G, Luban NL (1997) The irradiation of blood and blood components to prevent graft-versus-host disease: technical issues and guidelines. Transfus Med Rev 11(1):15–26
40. Kelley WE et al (2009) Washing platelets in neutral, calcium-free, Ringer's acetate. Transfusion 49(9):1917–1923
41. Johnson L et al (2016) Refrigeration and cryopreservation of platelets differentially affect platelet metabolism and function: a comparison with conventional platelet storage conditions. Transfusion 56(7):1807–1818
42. Slichter SJ et al (2014) Review of in vivo studies of dimethyl sulfoxide cryopreserved platelets. Transfus Med Rev 28(4):212–225
43. Albanyan AM, Harrison P, Murphy MF (2009) Markers of platelet activation and apoptosis during storage of apheresis- and buffy coat-derived platelet concentrates for 7 days. Transfusion 49(1):108–117
44. Cardigan R, Turner C, Harrison P (2005) Current methods of assessing platelet function: relevance to transfusion medicine. Vox Sang 88(3):153–163
45. Tynngard N (2009) Preparation, storage and quality control of platelet concentrates. Transfus Apher Sci 41(2):97–104

Part II

Updated Fractionation Strategies for In-depth Blood Proteome Analysis

Chapter 4

Bead-Based and Multiplexed Immunoassays for Protein Profiling via Sequential Affinity Capture

Elin Birgersson, Jochen M. Schwenk, and Burcu Ayoglu

Abstract

Antibody microarrays offer high-throughput immunoassays for multiplexed analyses of clinical samples. For such approaches, samples are either labeled in solution to enable a direct readout on the single binder assay format or detected by matched pairs of capture and detection antibodies in dual binder assay format, also known as sandwich assays. Aiming to benefit from the flexibility and capacity offered by single binder assay readout and the specificity and sensitivity of dual binder assays, we developed a multiplexed dual binder procedure that is based on a sequential, rather than combined, antigen binding. The method, entitled dual capture assay (DCA), is composed of an initial antigen capture by antibodies on beads, followed by labeling of captured protein targets on beads, combinatorial elution steps at high and low pH, and a readout using a secondary bead array. Compared to classical single binder assays, the described method demonstrated several advantages such as reduced contribution of off-target binding, lower noise levels, and improved correlation when comparing with clinical reference values. This procedure describes a novel and versatile immunoassay strategy for proteome profiling in body fluids.

Key words Affinity proteomics, Antibody arrays, Plasma profiling, Suspension bead array, Antibody selectivity, Assay sensitivity

1 Introduction

By investigating the human proteome in clinical specimen, we aim to acquire a deeper understanding of mechanisms of diseases and to establish clinical applications in the biomarker field [1, 2]. Alongside mass spectrometry, affinity reagents (e.g., antibodies) have proven to be versatile tools for the discovery and validation of protein biomarkers and the characterization of human tissues, cells, and body fluids [3]. Combined with highly parallelized and miniaturized technologies, antibodies are applicable to advanced protein profiling of large study sets, complex sample compositions, and biological specimens of limited availability. The discovery-oriented biomarker analysis of blood-derived serum and plasma is commonly a single binder assay based on a direct sample labeling approach

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_4,

combined with antibodies immobilized on glass or plastic slides [4] or magnetic color-coded microspheres [5–7]. These assays offer high-throughput and multiplexing capacities but do not match the sensitivity and selectivity of a dual binder assay [8]. To create an alternative immunoassay, we developed the so-called dual capture assay (DCA) [9]. The procedure combines multiplexing capabilities of a single binder assay and demonstrates decreased levels of off-target binding and improved background levels. The assay is centered around the concept of sequential protein capture [10], enabling to process samples with one selected set of antibodies twice or by combining two complementary sets of antibody chronologically. The enriched and labeled targets are eluted by the sequential use of high and low pH followed by neutralization. Magnetic microspheres provide the suitable solid support for both multiplexed enrichment and readout. Automated handling of the procedure is possible with magnetic bead handling devices and preferred in terms of assay performance and throughput. Figure 1 presents an overview of the assay workflow, and Fig. 2 presents an exemplary protein profile of a plasma sample analyzed using the dual capture assay format.

2 Materials

2.1 Coupling of Antibodies on Beads

1. Beads: MagPlex® magnetic microspheres (Luminex Corp).
2. Plates: 96-well half-area flat bottom polystyrene plates.
3. Plate shaker (Grantbio, PHMP-4).
4. Automated plate washer (BioTek, EL406) (*see* **Note 1**).
5. Activation buffer (1×): 100 mM monobasic sodium phosphate, pH 6.2, store at 4 °C for up to 1 month and at −20 °C for long term.
6. Prepare aliquots of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) in screw-capped tubes and store at −20 °C (*see* **Note 2**).
7. Prepare sulfo-N-hydroxysuccinimide (NHS) aliquots in screw-capped tubes and store at 4 °C (*see* **Note 2**).
8. Coupling buffer: 100 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 5.0, store at 4 °C for up to 1 month and at −20 °C for long term.
9. Wash buffer: 0.05% (v/v) Tween20 in 1× PBS pH 7.4 (PBS-T).
10. Antibodies (per 5×10^5 beads per ID): Dilute the antibodies to 20 µg/ml in coupling buffer. Diluted antibodies can be stored at 4 °C for up to 24 h (*see* **Note 3**).

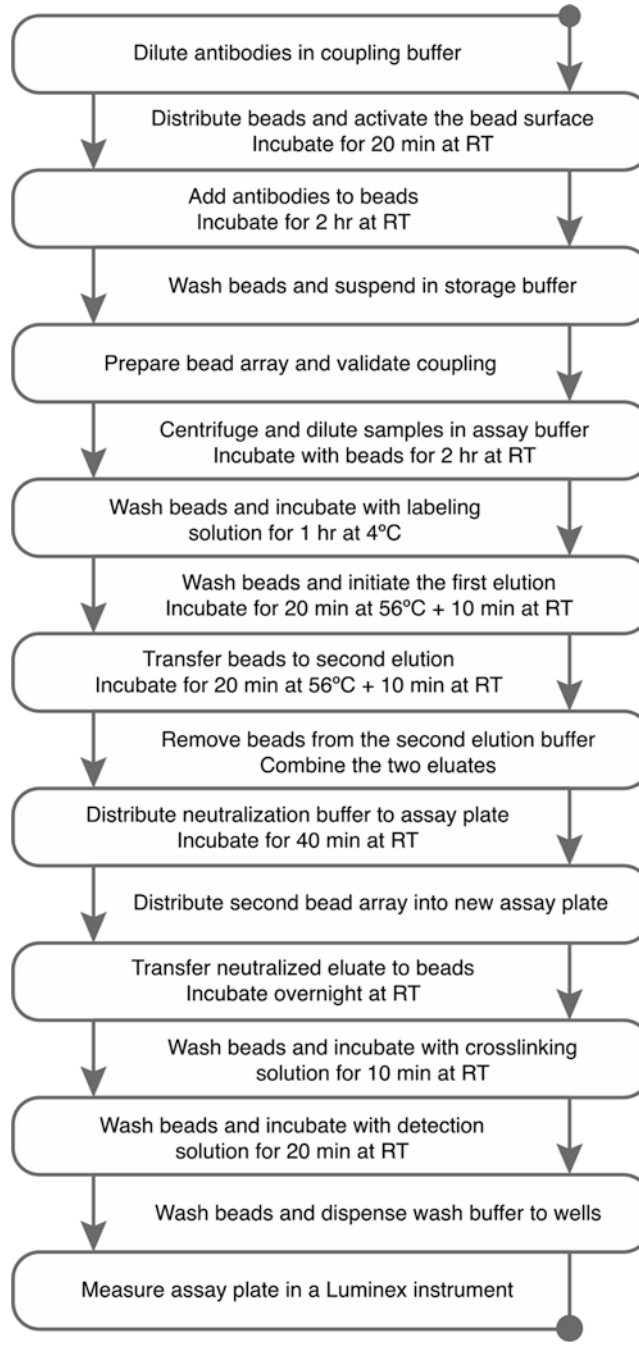


Fig. 1 Workflow for plasma and serum profiling with dual capture assay

11. Storage buffer (10×): Blocking reagent for ELISA (BRE). Dissolve to 2.7 mg/ml in 1× PBS. Store at −20 °C for long term and 4 °C for up to 1 month when diluted 1:10 in Milli-Q water and supplemented with 0.1% (v/v) ProClin 300™ (Supelco Analytical).

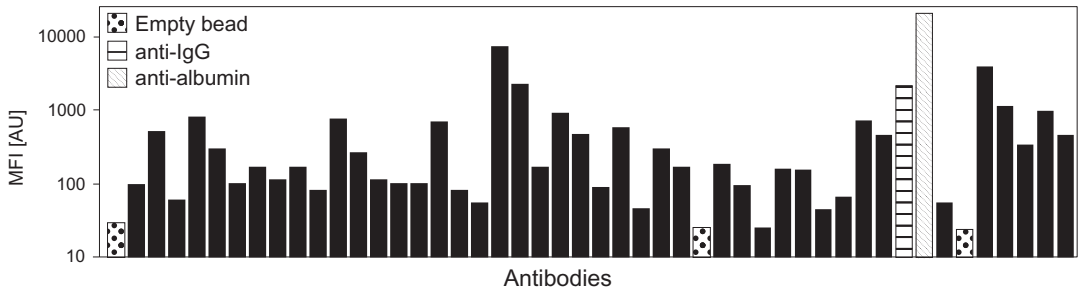


Fig. 2 Protein profile of a plasma sample analyzed with the dual capture assay. A 1:3 diluted plasma sample was analyzed with a 48-plex antibody array including antibodies against various human proteins including albumin and IgG. The *y*-axis displays the MFI values for the antibodies included in the array

2.2 Preparation of the Bead Array and the Coupling Efficiency Test

1. Plates: 96-well half-area flat bottom polystyrene plates.
2. Low-binding microcentrifuge tubes.
3. Magnetic tube holder (Dynal).
4. Sonication bath.
5. Storage buffer (10×): Blocking reagent for ELISA (BRE).
6. Antibody detection solution: R-phycoerythrin (R-PE) conjugated anti-species antibodies (e.g., Jackson ImmunoResearch). Dilute to 0.25 μg/ml in PBS-T directly before use and protect fluorophore from light (*see* **Note 4**).
7. Wash buffer: PBS-T.
8. Luminex instrument (FlexMap3D allowing for a multiplexing of up to 500 antibodies and assay readout in a 384-well plate or LX200 allowing for a multiplexing of up to 100 antibodies and assay readout in a 96-well plate) .

2.3 First Protein Capture and Labeling on Beads

1. Assay plates: Skirted 96-well PCR plate.
2. Assay buffer (1×): Prepare 0.1% (w/v) casein, 0.5% (w/v) polyvinyl alcohol, and 0.8% (w/v) polyvinylpyrrolidone (all Sigma) in 1× PBS. Store at −20 °C for long term and 4 °C for up to 1 month when supplemented with 0.05% (v/v) ProClin 300™ (Supelco Analytical). Add purified IgG derived from the host antibody species to a final concentration of 0.5 mg/ml and 0.05% (v/v) Tween20 before usage.
3. Magnetic particle processor (Thermo Scientific) (*see* **Note 5**).
4. Wash buffer: PBS-T.
5. Labeling solution: Dissolve sulfo-N-hydroxysuccinimide-polyethylene oxide biotin (NHS-PEG_4-Biotin, Thermo Scientific) directly before use to 40 mg/ml in dimethyl sulfoxide (DMSO, Sigma). Dilute solution in 1× PBS 0.005% (v/v) Triton X-100 to a final biotin concentration of 0.5 mg/ml (*see* **Note 6**).

2.4 Combinatorial Elution of Captured Proteins

1. Assay plates: Skirted 96-well PCR plate.
2. Magnetic particle processor (Thermo Scientific) (*see* **Note 5**).
3. Wash buffer: PBS-T.
4. Water bath.
5. Ring magnet (E&K Scientific).
6. Elution buffer 1: 0.1 M glycine-NaOH (Sigma), pH 10.0, supplemented with 0.05% (v/v) Tween20.
7. Elution buffer 2: 2.5% acetic acid (VWR), pH 3.0, supplemented with 0.05% (v/v) Tween20.
8. Neutralization buffer (1×): Prepare 2 M TRIS with 0.2% (w/v) casein, 1.0% (w/v) polyvinyl alcohol, and 1.6% (w/v) polyvinylpyrrolidone (all Sigma) in 1× PBS, pH 7.95. Store at −20 °C for long term and 4 °C for up to 1 month when supplemented with 0.05% (v/v) ProClin 300™ (Supelco Analytical). Add purified IgG derived from the host antibody species to a final concentration of 0.5 mg/ml and 0.05% (v/v) Tween20 before usage.

2.5 Second Capture and Assay Readout

1. Assay plates: 96-well half-area flat bottom polystyrene plates (Greiner Bio-One).
2. Plate shaker (Grantbio, PHMP-4).
3. Automated plate washer (BioTek, EL406) (*see* **Note 1**).
4. Wash buffer: PBS-T.
5. Cross-linking solution (10×): Prepare 4% paraformaldehyde (PFA, Alfa Aesar) solution and store at 4 °C for up to 1 month. Dilute to 0.4% in PBS-T prior to usage.
6. Detection solution: Dilute R-PE conjugated streptavidin (SAPE, Invitrogen) to 0.5 μg/ml in PBS-T directly before use and protect from light.
7. Luminex instrument (FlexMap3D allowing for a multiplexing of up to 500 antibodies and assay readout in a 384-well plate or LX200 allowing for a multiplexing of up to 100 antibodies and assay readout in a 96-well plate).

3 Methods

3.1 Coupling of Antibodies on Beads

The following protocol describes antibody immobilization on magnetic beads. The instructions can be applied to non-magnetic beads, with the exception of procedures involving liquid removal and washes based on the magnetic force to retain the particles. For those processes, we suggest to centrifuge beads into pellet or utilization of filter-bottomed plates (with pore size smaller than bead diameter) and a vacuum device. For a low number of bead IDs to

be immobilized with antibodies, we suggest to perform the coupling protocol in microcentrifuge tubes. If the number of IDs exceeds 24, we advise to follow a plate-based coupling protocol as described below:

1. Dilute the antibodies individually in coupling buffer to the required concentration. We recommend preparing 100 µl of diluted antibodies at 20 µg/ml per 5×10^5 beads per ID.
2. Distribute the range of bead IDs (e.g., 40 µl = 5×10^5 beads per ID) into the wells of a 96-well half-area bottom plate and wash the beads with 1 × 80 µl of activation buffer (*see* **Note 7**).
3. Dispense 50 µl of activation buffer into each well.
4. Prepare fresh solutions of NHS and EDC separately at 50 mg/ml in activation buffer. Calculate 0.5 mg of each of the two chemicals per well and prepare a mixture of volume enough for all wells and which consists of one part of NHS solution, one part of EDC solution, and three parts of activation buffer (*see* **Note 8**).
5. Distribute 50 µl of the prepared EDC-NHS activation solution to each well.
6. Incubate for 20 min at ambient room temperature in dark, under permanent and gentle mixing on a plate shaker (650 rpm), and wash thereafter with 2 × 100 µl coupling buffer.
7. Continue without interruption by applying the diluted antibody to the activated beads, and incubate for 2 h at ambient room temperature in dark, under permanent and gentle mixing on a plate shaker (650 rpm).
8. Wash the coupled beads 2× with 100 µl wash buffer and add 50 µl of storage buffer (1×) to each well. Store beads at 4 °C in the dark overnight.

3.2 Preparation of the Bead Array and the Coupling Efficiency Test

Combine equal volumes of selected bead identities to create a suspension bead array stock. Theoretically, suspending the starting amount of 5×10^5 beads per ID in 50 µl of storage buffer after the coupling procedure yields a bead concentration of 50,000 beads per ID in 5 µl. For this dual capture assay, we recommend preparing a mixture of beads with a final bead concentration of 1250 beads per ID/5 µl. Thus, in this example, each bead ID should be diluted 40 times in the final bead stock. The volume of the mixture of beads should be prepared so that the volume is in excess: 5 µl × number of assay wells +20% extra volume (*see* **Note 9**). To confirm a successful coupling, the immobilization efficiency should be evaluated through an initial test with fluorescently labeled anti-species-specific antibodies.

1. Calculate the final volume of bead array mixture needed for the assay. From each bead ID, transfer a volume corresponding to 2.5% of the final volume into a low-binding microcentrifuge tube.

2. Adjust the volume of the bead array mixture with 1× storage buffer. Use a magnetic tube holder if any extra volume of buffer needs to be removed.
3. Vortex and sonicate the bead array mixture before usage.
4. Dispense 5 µl of mixture of beads in an assay plate and add 50 µl of prepared antibody detection solution.
5. Incubate for 20 min at ambient room temperature in dark, under permanent and gentle mixing on a plate shaker (650 rpm), and wash thereafter with 3 × 100 µl wash buffer.
6. Dispense a final of 100 µl wash buffer and measure the fluorescence intensities with Luminex instrumentation.
7. Store bead array at 4 °C and in dark until sample analysis.

3.3 First Protein Capture and Labeling on Beads

1. Thaw serum or plasma samples at 4 °C (*see* **Note 10**).
2. Vortex and centrifuge samples at 1811 RCF for 2 min before preparing a 1:3 dilution in assay buffer.
3. Distribute 5 µl of previously prepared bead array mixture to each well of an assay plate (*see* **Note 11**).
4. Add 30 µl of diluted samples to the assay plate with beads, and incubate for 2 hr. at 4 °C with vortex intervals every 15–25 min to avoid bead aggregation (*see* **Note 12**).
5. Prepare an assay plate with 30 µl labeling solution and three plates with 100 µl wash buffer the KingFisher Flex process. Set the instrument protocol to 3× wash for 1 min and thereafter bead release into the plate with labeling solution (*see* **Note 13**).
6. Incubate beads in the labeling solution for 1 h at 4 °C with vortex intervals every 15–25 min to avoid bead aggregation.

3.4 Combinatorial Elution of Captured Proteins

1. To terminate the biotinylation and initiate the protein elution, distribute 15 µl of elution buffer 1 in an assay plate and prepare three assay plates with 100 µl wash buffer. Apply protocol using the KingFisher Flex including 3× wash for 1 min and final bead release into elution buffer 1.
2. Seal elution plate properly before heat treatment in water bath at 56 °C for 20 min followed by 10 min at ambient room temperature (*see* **Note 14**).
3. Prepare a new assay plate with 15 µl of elution buffer 2 and program the KingFisher Flex to transfer beads from elution buffer 1 to plate with elution buffer 2.
4. Store the protein eluate from first elution at 4 °C while the beads undergo their second elution at 56 °C for 20 min followed by 10 min at ambient room temperature (*see* **Note 14**).

5. Finally, the beads in elution buffer 2 are transferred to a plate with wash buffer.
6. The eluates are combined by dispensing 13 µl from one elution plate into the other. Thereafter 30 µl of neutralization buffer is distributed into each well, and the assay plate is incubated for 40 min at ambient room temperature (*see* **Note 15**).

3.5 Second Capture and Assay Readout

1. Dilute previously used bead mixture 2.5× to 500 beads per ID in5 µl, and dispense 5 µl to each well of a 96-well half-area flat bottom assay plate (*see* **Note 16**).
2. Transfer 50 µl of the neutralized eluate to wells with new bead array, and incubate overnight at ambient room temperature in dark, under permanent and gentle mixing on a plate shaker (650 rpm) (*see* **Note 15**).
3. Wash 3× 100 µl wash buffer and add 50 µl of cross-linking solution. Thereafter incubate for 10 min at ambient room temperature in dark, under permanent and gentle mixing on a plate shaker (650 rpm).
4. Wash 3× 100 µl with wash buffer, followed by a dispense of 50 µl of detection solution per well and incubation for 20 min at ambient room temperature in dark, under permanent and gentle mixing on a plate shaker (650 rpm).
5. A final 3× 100 µl wash with wash buffer takes place before 100 µl wash buffer is dispensed to the wells and plate is ready for readout.
6. Select the Luminex instrument settings according to bead identities included in second bead array as well as at least 50 beads per ID and well. We suggest using the "median florescence intensity" to further process and evaluate your data.

4 Notes

1. The washing steps described in this procedure are carried out on an automated plate washer (BioTek, EL406) suitable for handling magnetic beads. If no such instrument is available, a plate magnet (LifeSep, 96F) and a vacuum device (Gilson Safe Aspiration Station) can be used instead.
2. EDC and NHS are both highly hygroscopic substances and need to be equilibrated to ambient room temperature for at least 30 min before opening the vials.
3. Employ solutions of purified antibodies and avoid other stabilizing proteins or other amine-based buffers as they might reduce the coupling efficiency.

4. Other fluorescent dyes than R-PE, such as Alexa555, Alexa532, or Cy3, can be applied as well, but they have been shown to yield lower signal intensities. Different suppliers for R-PE conjugates can also be compared to achieve the desired assay performance.
5. A magnetic particle processor has been shown suitable for washing steps and transferring of beads throughout the procedures. If no such instrument is available, a plate magnet (LifeSep, 96F) can be used in its place.
6. Do not interrupt the process after dissolving biotin in PBS-Triton X-100, as biotin is susceptible to hydrolysis and will begin to lose its activity when in contact with water-based buffers. The plate can be held on ice to counteract the hydrolysis reaction.
7. At all times, try to minimize the light exposure, especially to direct sunlight, as the internal fluorescence of the beads as well as reporter fluorophore could be bleached. During incubations, protect the plates with an opaque cover and/or place the plates into a light-tight box.
8. Do not interrupt the process after dissolving EDC and NHS, as these substances are susceptible to hydrolysis. We recommend distributing the solution to beads without any delay to avoid the risk of reduced coupling efficiency.
9. The required number of beads should be adjusted for each assay procedure. To ensure a sufficient number of beads, a count with Luminex instrument can be of help before further analysis. If not all bead IDs are present in a bead array mixture, a nonexisting bead ID can be included in the software protocol for the coupling test. This would allow the instrument to count for the specified time-out seconds and provide a better overview of the amount of each bead ID in the final bead array.
10. We prefer a thawing procedure at 4 °C (e.g., refrigerators). The thawing time depends on the size of the sample cohort and volumes of plasma or serum. For volumes >200 µl, thawing overnight at 4 °C is recommended.
11. If aggregation of beads has occurred, vortex the beads followed by a sonication for 3 min. Safety measures regarding the handling of sonication baths are to be taken in consideration.
12. We recommend distribution of the bead array (i.e., 5 µl) into the well first and then adding the larger volume of sample material (i.e., 30 µl). This will promote better suspension of the beads in the sample well.
13. As mentioned previously, biotin is susceptible to hydrolysis and will begin to lose its activity when dissolved in PBS-Triton X-100. To optimize the procedure, biotin should be prepared

directly before contact with the beads. If the KingFisher Flex instrument is used, we recommend including the function "Pause" in the software protocol for enabling a late insertion of the plate with labeling solution.

14. Heat treatment has proven to optimize the elution of proteins in both low and high pH. According to our experience, the preferred combination of elution steps is to first introduce high pH and thereafter low pH. Nevertheless, the optimal elution procedure might vary between each protein target and should be evaluated before initiating the sample assays.
15. We recommend mounting the plate on a ring magnet to ensure no beads are transferred along with the elution solutions. This is to avoid any interference of the first bead array once the secondary bead array is introduced.
16. This assay enables two alternatives for the second capture step: Either the same bead array can be introduced again (as described here) or an additional set of coupled beads can be selected to investigate the performance of antibody pairs toward the same protein target.

Acknowledgments

We thank all members of the Biobank Profiling group at SciLifeLab, as well as the entire staff of the Human Protein Atlas. This study was funded by grants from Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, and the KTH Center for Applied Proteomics (KCAP) funded by the Erling-Persson Family Foundation. The authors declare no conflict of interest.

References

1. Solier C, Langen H (2014) Antibody-based proteomics and biomarker research-current status and limitations. Proteomics 14:774–783
2. Fu Q, Schoenhoff FS, Savage WJ, Zhang P, Van Eyk JE (2010) Multiplex assays for biomarker research and clinical application: translational science coming of age. Proteomics Clin Appl 4:271–284
3. Uhlen M, Fagerberg L, Hallstrom BM et al (2015) Tissue-based map of the human proteome. Science 347:1260419–1260419
4. Ayoglu B, Haggmark A, Neiman M et al (2011) Systematic antibody and antigen-based proteomic profiling with microarrays. Expert Rev Mol Diagn 11:219–234
5. Bystrom S, Ayoglu B, Haggmark A et al (2014) Affinity proteomic profiling of plasma, cerebrospinal fluid, and brain tissue within multiple sclerosis. J Proteome Res 13:4607–4619
6. Ayoglu B, Chaouch A, Lochmüller H et al (2014) Affinity proteomics within rare diseases: a BIO-NMD study for blood biomarkers of muscular dystrophies. EMBO Mol Med 6:918–936
7. Bachmann J, Burté F, Pramana S et al (2014) Affinity proteomics reveals elevated muscle proteins in plasma of children with cerebral malaria. PLoS Pathog 10:e1004038
8. Landegren U, Vänelid J, Hammond M et al (2012) Opportunities for sensitive plasma proteome analysis. Anal Chem 84:1824–1830
9. Ayoglu B, Birgersson E, Mezger A et al (2016) Multiplexed protein profiling by sequential affinity capture. Proteomics 16:1251–1256
10. Zhou S, Lu X, Chen C, Sun D (2010) An immunoassay method for quantitative detection of proteins using single antibodies. Anal Biochem 400:213–218

Chapter 5

Affinity Proteomics for Fast, Sensitive, Quantitative Analysis of Proteins in Plasma

John P. O'Grady, Kevin W. Meyer, and Derrick N. Poe

Abstract

The improving efficacy of many biological therapeutics and identification of low-level biomarkers are driving the analytical proteomics community to deal with extremely high levels of sample complexity relative to their analytes. Many protein quantitation and biomarker validation procedures utilize an immunoaffinity enrichment step to purify the sample and maximize the sensitivity of the corresponding liquid chromatography tandem mass spectrometry measurements. In order to generate surrogate peptides with better mass spectrometric properties, protein enrichment is followed by a proteolytic cleavage step. This is often a time-consuming multistep process. Presented here is a workflow which enables rapid protein enrichment and proteolytic cleavage to be performed in a single, easy-to-use reactor. Using this strategy Klotho, a low-abundance biomarker found in plasma, can be accurately quantitated using a protocol that takes under 5 h from start to finish.

Key words Immunoaffinity, Mass spectrometry, Hybrid LBA/LC-MS, Thermal denaturation, Immunocapture, Streptavidin, Immunoprecipitation, Trypsin

1 Introduction

Sensitivity and specificity are two major driving needs in the analysis of proteins in plasma. As it becomes increasingly evident how complex the analysis of biological samples can be, the specificity of mass spectrometry over traditional ligand binding assays (LBAs) is gaining traction both in research and clinical settings [1, 2]. Coupling this specificity to sensitivity gained using antibody-based capture and enrichment has enabled researchers to analyze plasma for very low-abundant biomarkers [3–5]. One method of performing this capture is to biotinylate an antibody targeting the protein of interest and then bind that antibody to a bead or other surface that has been coated with streptavidin. The target may subsequently be captured onto this surface by the antibody. The targeted is then eluted off of the surface or subjected to digestion—proteolysis—while still bound.

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_5,

Affinity capture for mass spectrometry analysis can be performed before digestion [3], after digestion [4, 5], or both before and after [6], depending on the sensitivity needs and the nature of the system being analyzed [3]. One major hindrance to more widespread adoption of these protocols is their time and labor-intensive nature, regularly taking 3 days and numerous reaction steps to complete [5, 7]. Here we demonstrate a simple method employing an initial capture followed by digestion for the rapid detection of the low-abundance plasma biomarker Klotho.

2 Materials

Prepare all solutions using ultrapure water (deionized water with a sensitivity of 18 MΩ-cm at 25 °C or better) and analytical grade reagents. Prepare and store all reagents at room temperature unless otherwise indicated.

Due to the low protein concentrations seen toward the end of the procedure, it is recommended that all steps after the affinity be performed in low protein binding materials to prevent interference from nonspecific adsorption to the plastics.

2.1 Biotinylation Solution

1. Weigh out 0.5 mg of N-hydroxysuccinimide biotin in a 1.5 mL tube. Add 1 mL dimethyl sulfoxide (DMSO).

2.2 Antibody Preparation

1. Purified monoclonal antihuman Klotho antibody (a-hKlotho) purchased from R&D Systems (*see* **Note 1**).
2. PBS solution, 10 mM: In a 1 L bottle, combine 0.26 g potassium phosphate monobasic, 2.17 g potassium phosphate dibasic heptahydrate, 8.71 g sodium chloride, and 800 mL water. Adjust the pH to 7.4 using sodium hydroxide or hydrochloric acid and bring the volume to 1 L with water. Do *not* add azide (*see* **Note 2**).

2.3 Control Samples: Pure and in Matrix

1. Purified recombinant human Klotho (rhKlotho) at a concentration of 1 mg/mL purchased from R&D Systems.
2. Calibration samples in murine plasma: Add 0.95 μL of rhKlotho to 1.9 mL plasma to make 500 ng/mL rhKlotho in plasma. Perform serial dilutions as noted in Table 1.

2.4 Capture and Digestion Kit with Thermally Stable Trypsin

1. SMART Digest Immuno Affinity kit – Streptavidin (SDIA) from Thermo Fisher Scientific.
   (a) Wash buffer: Empty the kit wash concentrate solution packet into a 1 L bottle. Make up to 1 L with water.

Table 1
Serial dilutions for the creation of a calibration curve for 500 µL samples in triplicate

Starting solution	Diluent (murine plasma) (mL)	Final concentration (ng/mL)
0.95 µL stock rhKlotho (1 mg/mL)	1.9	500
281 µL 500 ng/mL solution	2.527	50
739 µL 50 ng/mL solution	1.6	15.8
468 µL 50 ng/mL solution	1.872	10
739 µL 15.8 ng/mL solution	1.6	5
506 µL 10 ng/mL solution	1.094	3.16
506 µL 5 ng/mL solution	1.094	1.58

2. Protein LoBind 1.5 mL microcentrifuge tubes from Eppendorf (Hauppauge, NY).
3. An Eppendorf ThermoMixer C with 1.5 mL head and ThermoTop heated lid or an Eppendorf ThermoMixer F1.5 and ThermoTop heated lid (*see* **Note 3**).

2.5 LC-MS/MS Analysis

1. LC-MS/MS equipped with a C18 column, divert valve, and quantitation software.
2. Mobile Phase A: In a 1 L bottle add 980 mL of water, 20 mL of acetonitrile, and 1 mL of formic acid.
3. Mobile Phase B: In a 1 L bottle add 450 mL of acetonitrile, 50 mL of water, and 0.5 mL of formic acid.

3 Methods

Carry out all steps at room temperature unless otherwise noted.

3.1 Antibody Biotinylation

1. Dilute the a-hKlotho stock solution into a working concentration of 100 µg/mL by adding 50 µL of the 1 mg/mL solution to 450 µL of PBS.
2. To the working solution of a-hKlotho, add 2.5 µL of 0.5 mg/mL biotinylation solution and mix thoroughly (*see* **Note 4**).
3. Allow the solution to react with continuous mixing, 1400 rpm on an orbital mixer, for at least 2 h.

3.2 Immunoaffinity Capture

1. To 1.5 mL tubes add (each) 500 µL of sample, 10 µL of 100 µg/mL biotinylated a-hKlotho, and 30 µL of SDIA and mix thoroughly.
2. Allow the solution to react with continuous mixing, 1400 rpm on an orbital mixer, for at least 2 h (*see* **Note 5**).

3.3 Matrix Removal

1. Centrifuge the samples at ≥3000 × *g* for at least 1 min and decant 480 µL of the supernatant.
2. To the resin with captured target, add 450 µL of wash buffer. Centrifuge the samples again and decant 450 µL of the supernatant. Do this five times in total.

3.4 Target Digestion

1. To the resin add 150 µL of SDIA Digest Buffer and mix thoroughly (*see* **Note 6**).
2. Incubate at 70 °C and 1400 rpm for 90 min on the ThermoMixer C with heated lid (*see* **Notes 7** and **8**).
3. Remove samples from the ThermoMixer C and allow them to cool for 5 min before centrifuging them as before and decanting the supernatant into a protein LoBind holder for analysis on the LC-MS/MS.

3.5 LC-MS/MS Analysis

1. Create a method on the LC-MS/MS tracking the peptides FSISWAR (parent mass 433.7, fragment mass 519.3 and 632.4) and LQDAYGGWANR (parent mass 625.8, fragment mass 660.3 and 1009.4) using a gradient of 2–50% Mobile Phase B (*see* **Note 8**). Start the method with holding the gradient at 2% Mobile Phase B and diverting flow to waste for 1 min to avoid loading salts onto the MS.
2. Inject 50 µL of the digested supernatant for analysis.
3. Analyze the resultant peak areas (*see* **Note 9**). Figure 1 demonstrates the separation achieved by the gradient.

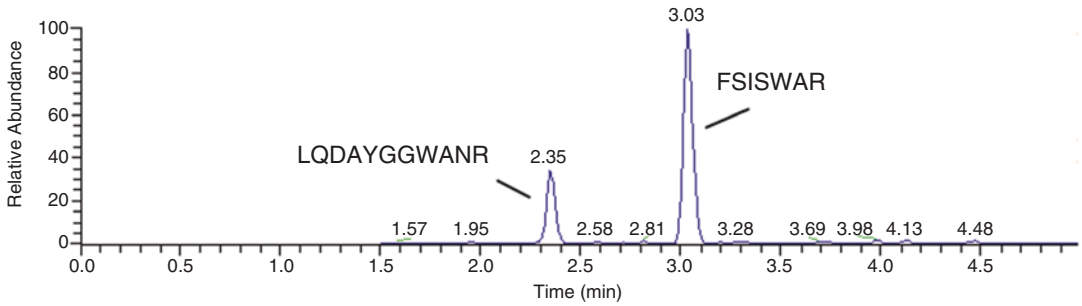


Fig. 1 Chromatogram of Klotho peptides FSISWAR and LQDAYGGWANR

4 Notes

1. Antibody selection is a critical component of developing a successful immunoaffinity assay. In order to achieve maximum sensitivity, an antibody must have both good activity and selectivity toward the target protein of interest. In the course of developing this assay, three different antibodies were tested. Of these, the two not presented here both exhibited poor sensitivity in a plasma matrix. Antibody binding can be tested by adding pure protein and a sufficient amount of biotinylated antibody to bind the protein to the SDIA, incubating for 4 h with mixing, then transferring the supernatant to a different set of beads for digestion and analyzing the resulting solution for the presence of the peptides of interest.
2. Other buffers may be used for biotinylation provided that they are at a non-denaturing pH—usually between 6 and 8—and amine-free. For this reason, azide must be avoided. The amines will react with the NHS biotin and compete with the amine residues of the antibody for biotinylation.
3. While other heater/shaker apparatuses may be substituted, multiple different units were tested for their uniformity, and only the newer Eppendorf heater shakers with a heated lid did not show any significant disparity across the plate. Other formats may also be used (such as 96 deep-well plates), but the heated lid is essential for uniformity of the digestion step.
4. The optimum amount of biotin per protein is noted in Table 2. Too much biotin will leave remaining free biotin to swamp out streptavidin binding sites on the SDIA. Too little will fail to sufficiently biotinylate the antibody for capture. This titering eliminates the need to dialyze or otherwise remove any unreacted biotin from the antibody.
5. The capture time is experimentally determined using a high concentration of the target protein. Using a tuned digestion

Table 2
NHS-biotin guide for protein biotinylation

NHS-biotin addition (μg)							
		Protein molarity (nM)					
Sample volume (mL)	1	0.1	1	5	10	25	100
		0.4	4.1	20.5	40.95	102.4	409.7
	0.5	0.2	2.05	10.25	20.5	51.2	204.85
	0.25	0.1	1	5.1	10.25	25.6	102.4
	0.1	0.05	0.4	2.05	4.1	10.25	40.95

time (*see* **Note 7**), perform the capture for varying lengths of time, e.g., 30-min intervals over a span of 4 h, to determine the point at which the target is fully captured. This is determined by observing when the detection level becomes asymptotic with respect to signal over time. Fifteen minutes may be added to account for any variances in sample matrices.

6. If stable isotope-labeled peptides are being used to control for potential variances in ionization downstream, they should be added at this point in the protocol.
7. The digestion method can be tuned by making eight samples of 50 μL of a high concentration of the pure protein (e.g., ≥1 μg/mL) and digesting them for varying lengths of time, e.g., 15-min intervals over a span of 2 h, to determine the point at which the target is fully digested. This is determined by observing when the detection level becomes asymptotic with respect to signal over time.
8. Initial method development may be performed by digesting the purified target protein for 2 h with SDIA in a manner similar to **Note 5** in order to have peptides to test. Although this may not be the optimal time, it should be sufficient to present the peptides of interest.
9. If no peaks are seen, troubleshoot the method starting at the end and working back to the beginning. First confirm that the LC-MS system is working using manufacturer recommended controls. Next, ensure that the LC-MS detection method is able to detect the peptides of interest using purified peptides or digested pure protein (*see* **Note 8**). Then ensure that sufficient time is being given to digest the sample (*see* **Note 7**). Finally, ensure that the antibody is able to capture the protein (*see* **Note 1**) and that sufficient sufficient time is being given for the protein to be captured from the matrix (*see* **Note 5**).

Conflict of Interest

Perfinity Biosciences is the manufacturer of the SMART Digest Immunoaffinity kit for Thermo Fisher Scientific.

References

1. Stevenson L, Garofolo F, DeSilva B, Dumont I, Martinez S, Rocci M, Amaravadi L, Brudny-Kloeppel M, Musuku A, Booth B, Dicaire C, Wright L, Mayrand-Provencher L, Losauro M, Gouty D, Arnold M, Bansal S, Dudal S, Dufield D, Duggan J, Evans C, Fluhler E, Fraser S, Gorovits B, Haidar S, Hayes R, Ho S, Houghton R, Islam R, Jenkins R, Katori N, Kaur S, Kelley M, Knutsson M, Lee J, Liu H, Lowes S, Ma M, Mikulskis A, Myler H, Nicholson B, Olah T, Ormsby E, Patel S, Pucci V, Ray C, Schultz G, Shih J, Shoup R, Simon C, Song A, Neto JT, Theobald V, Thway T, Wakelin-Smith J, Wang J, Wang L, Welink J,

Whale E, Woolf E, Xu R (2013) 2013 white paper on recent issues in bioanalysis:'hybrid'-the best of LBA and LCMS. Bioanalysis 5(23):2903–2918

2. Dufield D, Neubert H, Garofolo F, Kirkovsky L, Stevenson L, Dumont I, Kaur S, Xu K, Alley SC, Szapacs M, Arnold M, Bansal S, Haidar S, Welink J, Le Blaye O, Wakelin-Smith J, Whale E, Ishii-Watabe A, Bustard M, Katori N, Amaravadi L, Aubry AF, Beaver C, Bergeron A, Cai XY, Cojocaru L, DeSilva B, Duggan J, Fluhler E, Gorovits B, Gupta S, Hayes R, Ho S, Ingelse B, King L, Lévesque A, Lowes S, Ma M, Musuku A, Myler H, Olah T, Patel S, Rose M, Schultz G, Smeraglia J, Swanson S, Torri A, Vazvaei F, Wilson A, Woolf E, Xue L, Yang TY (2014) 2014 white paper on recent issues in bioanalysis: a full immersion in bioanalysis (part 2-hybrid LBA/LCMS, ELN & regulatory agencies' input). Bioanalysis 6(23):3237–3249
3. Ackermann BL, Berna MJ (2014) Coupling immunoaffinity techniques with MS for quantitative analysis of low-abundance protein biomarkers. Expert Rev Proteomics 4(2):175–186
4. Anderson NL, Anderson NG, Haines LR, Hardie DB, Olafson RW, Pearson TW (2004) Mass spectrometric quantitation of peptides and proteins using stable isotope standards and capture by anti-peptide antibodies (SISCAPA). J Proteome Res 3(2):235–244
5. Hoofnagle AN, Becker JO, Wener MH, Heinecke JW (2008) Quantification of thyroglobulin, a low-abundance serum protein, by immunoaffinity peptide enrichment and tandem mass spectrometry. Clin Chem 54(11):1796–1804
6. Palandra J, Finelli A, Zhu M, Masferrer J, Neubert H (2013) Highly specific and sensitive measurements of human and monkey interleukin 21 using sequential protein and tryptic peptide immunoaffinity LC-MS/MS. Anal Chem 85(11):5522–5529
7. Kuhn E, Addona T, Keshishian H, Burgess M, Mani DR, Lee RT, Sabatine MS, Gerszten RE, Carr SA (2009) Developing multiplexed assays for troponin I and interleukin-33 in plasma by peptide immunoaffinity enrichment and targeted mass spectrometry. Clin Chem 55(6):1108–1117

Chapter 6

Characterization of the Low-Molecular-Weight Human Plasma Peptidome

David W. Greening and Richard J. Simpson

Abstract

The human plasma proteome represents an important secreted sub-proteome. Proteomic analysis of blood plasma with mass spectrometry is a challenging task. The high complexity and wide dynamic range of proteins as well as the presence of several proteins at very high concentrations complicate the profiling of the human plasma proteome. The peptidome (or low-molecular-weight fraction, LMF) of the human plasma proteome is an invaluable source of biological information, especially in the context of identifying plasma-based markers of disease. Peptides are generated by active synthesis and proteolytic processing, often yielding proteolytic fragments that mediate a variety of physiological and pathological functions. As such, degradomic studies, investigating cleavage products via peptidomics and top-down proteomics in particular, have warranted significant research interest. However, due to their molecular weight, abundance, and solubility, issues with identifying specific cleavage sites and coverage of peptide fragments remain challenging. Peptidomics is currently focused toward comprehensively studying peptides cleaved from precursor proteins by endogenous proteases. This protocol outlines a standardized rapid and reproducible procedure for peptidomic profiling of human plasma using centrifugal ultrafiltration and mass spectrometry. Ultrafiltration is a convective process that uses anisotropic semipermeable membranes to separate macromolecular species on the basis of size. We have optimized centrifugal ultrafiltration (cellulose triacetate membrane) for plasma fractionation with respect to buffer and solvent composition, centrifugal force, duration, and temperature to facilitate recovery >95% and enrichment of the human plasma peptidome. This method serves as a comprehensive and facile process to enrich and identify a key, underrepresented sub-proteome of human blood plasma.

Key words Blood, Plasma, Proteome, Low-molecular weight, LMF, LMW, Ultrafiltration, Peptidome, Proteomics, Degradome

1 Introduction

Human plasma is one of the most informative and important proteomes from a clinical perspective. For example, characteristic changes in protein levels in plasma are indicative of many clinical conditions, including severe liver disease, hemolytic anemia and Down's syndrome, schizophrenia, Alzheimer's disease, amyotrophic lateral sclerosis, and Creutzfeldt-Jakob disease [1].

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_6,

Hence, characterization of plasma proteins (both in qualitative and quantitative terms) should provide a foundation for the discovery of candidate markers for disease diagnosis and development of new therapeutics. However, human plasma is limited by its dynamic range of protein abundances (ten orders of magnitude between the least abundant (1–5 pg/mL, e.g., interleukins, cytokines) and most abundant (35–70 × 10^9 pg/mL, e.g., albumin, IgG; [2]). For example, albumin and immunoglobulin G constitute approximately 51–71% and 8–26% of the total protein content in human plasma, respectively [3]. This complexity creates extensive difficulties in the use of many proteomic separation tools (e.g., free-flow electrophoresis, 10^5 [4]) for the identification of low-abundance species directly in plasma (overview: [2]). The strategies that have been most frequently used to overcome this issue of dynamic range are to fractionate the plasma proteome into smaller subsets (sub-proteomes) and/or to deplete one or more of the major proteins. Immunoaffinity is an established method that addresses the dynamic range of plasma by specific depletion of high-abundance proteins [5]. However, although the efficiency of immunodepletion ranges from 96 to 99%, the remaining concentration of albumin, for example, would still be ~50–1000 μg/mL—a value ~10^4-fold higher than blood CEA levels (~5 ng/mL) and 5 × 10^6-fold higher than blood IL-6 levels (~10 pg/mL). Hence, mass spectrometry-based detection of most already known biomarkers in blood requires deployment of additional separation/enrichment technologies.

The peptidome (also known as degradome or low-molecular-weight fraction, LMF) is defined as a subset of the proteomes that consist of peptides formed through the proteolysis of proteins to release latent bioactive peptides. This targeted proteolytic digestion of proteins can release peptides that have distinct activity compared to the precursor protein. The proteolysis directed toward proteins occurs through the catalytic activity of system protease enzymes and protease inhibitors. It is well known that protease activity is extremely selective and the enzyme is directed to a site-specific sequence of amino acids to be able to cleave the protein into multiple peptides, and in addition each individual protease has optimal conditions under which it can exert its activity. Known plasma polypeptides, such as the defensins, and bioactive peptides like glucagon, insulin, growth hormone, and neuropeptides are involved in a variety of biological functions. The LMF also contains proteolytic peptide fragments of several abundant proteins such as albumin, transthyretin, and apolipoproteins [6, 7]. To this aim, the plasma or serum proteome has been the focus of recent attempts to identify low-abundance and low-molecular-weight endogenous peptides which hold diagnostic and prognostic potential [8–12].

Extracellular proteolysis represents a dynamic role in cell regulation, signaling, and tissue homeostasis. In cancer, proteolytic activity is an important component regulating intercellular communication throughout the surrounding microenvironment, with altered proteolysis promoting deregulated tumor growth, tissue remodeling, inflammation, tissue invasion, and metastasis [13]. Ectodomain shedding and intramembrane proteolysis are becoming critical elements of many diverse intra- and intercellular signaling events mediated by small peptide fragments [12]. The presence of bioactive components in human plasma and serum as a consequence of cancer and other pathological processes has been well documented [6, 8–10, 14–18]. Recently, efforts have been directed toward developing high-throughput proteomic screens to identify protease cleavage events and protease substrates in complex biological samples. However, due to their molecular weight (<3 kDa) and solubility, issues with identifying specific cleavage sites and coverage of peptide fragments remain challenging. Peptidomics is currently focused toward comprehensively studying peptides cleaved from precursor proteins by endogenous proteases [19–29].

Of the studies investigating the plasma peptidome, issues with membrane selectivity, centrifugal conditions, buffers and solvents, and filtrate heterogeneity and contamination with abundant, high-M_r plasma proteins have limited enrichment and characterization of the peptidome [11, 19, 30–33]. Centrifugal ultrafiltration has been the most widely used method to extract peptides and remove proteins with higher molecular weights from plasma/serum based on a size-exclusion filtration mechanism [30, 31, 34–36]. Typically, membranes have a mean pore size between 10 and 500 Å (or 1.0 and 50 nm). Here, we report a strategy to selectively and rapidly isolate and identify the human plasma peptidome using centrifugal ultrafiltration and mass spectrometry (Fig. 1).

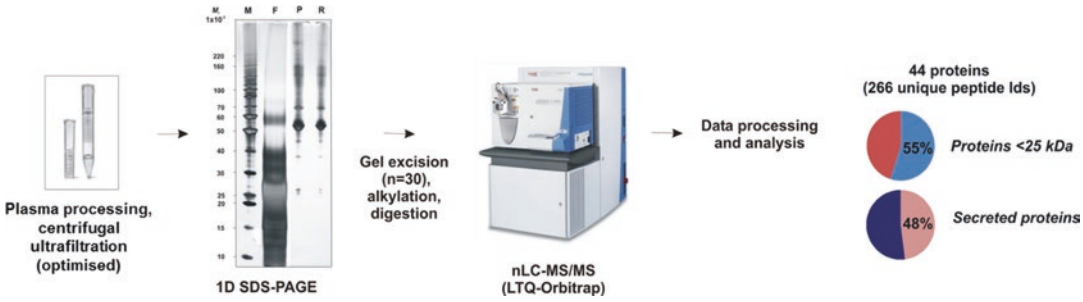


Fig. 1 Isolation and characterization strategy for the comprehensive analysis of the low-molecular-weight human plasma peptidome. The efficiency of the optimized method is demonstrated by the identification of the low-abundance classically secreted proteins cystatin-C (plasma concentration 0.62–1.02 μg/mL), CXCL7, serine protease inhibitor A3, and cystatin-M. In all, 48% of the proteins identified in the rapid processing and analysis of the plasma peptidome are known secreted proteins

2 Materials

Throughout the protocol, Milli-Q deionized water (HPLC grade, ≥18 MΩ) should be used for making up all aqueous solutions. All washing, lysis, and HPLC buffers should be prepared using clean glassware on the day analysis is to be performed.

2.1 Blood Collection, Plasma Preparation, and Storage (See Notes 1–3)

1. EDTA Blood Collection Tubes (e.g., BD Vacutainer).
2. Polypropylene tubes (1.5 mL, 15 mL).
3. Freezer (−80 °C or lower).
4. Gloves, gown, eye protection.
5. Pipettes.
6. Disposal container for contaminated tubes.
7. Centrifugation unit (either/or benchtop/swing-bucket rotor—compatible with 1.5/15 mL tubes, programmable temperature setting; range 4–25 °C).
8. Labels for blood sample tubes.
9. Alcohol (70% (v/v) aqueous ethanol) and swabs for cleaning venipuncture site.
10. Micro BCA Protein Assay Kit, sufficient reagent to perform 480 standard tube assays or 3200 microplate assays (Pierce).
11. Water bath or incubator set at 37 °C.

2.2 Centrifugal Ultrafiltration

1. Centrifugation unit (benchtop series – compatible with 1.5 mL tubes, programmable temperature setting; range 4–25 °C).
2. Centrifugation unit (swing-bucket rotor—compatible with 15 mL tubes, programmable temperature setting; range 4–25 °C).
3. Centrifugal ultrafiltration membranes—Vivaspin® 2 MWCO of 20,000 (cellulose triacetate (CTA), Sartorius Stedim Biotech, Aubagne, France) (*see* **Notes 4** and **5**). For selecting the correct NMWL of the filtration membrane device, *see* **Notes 6** and **7**.
4. Acetonitrile, HPLC grade (Fisher, A998–1 or equivalent) (*see* **Note 8**).
5. Water, HPLC grade (Fisher, W5–1 or equivalent) .

2.3 SDS-PAGE

1. Laemmli non-reducing sample buffer (0.2 M Tris–HCl, 40% (v/v) aqueous glycerol, 4% (w/v) SDS, trace bromophenol blue).
2. Heat block—up to 95 °C (compatible with 1.5 mL centrifuge tubes).
3. NuPAGE® LDS sample buffer (Invitrogen), stored at 4 °C.
4. 1D gel apparatus (Invitrogen Novex Mini-Cell).

5. Precast SDS polyacrylamide 12-well, 1.5 mm gel (4–12% Bis-Tris precast gel, Invitrogen).
6. 20× NuPAGE® MES SDS running buffer (Invitrogen): 50 mM MES, pH 7.2, 50 mM Tris-NaOH, 0.1% SDS, 1 mM EDTA, pH 7.3, stored at room temperature (RT). Add 25 mL 20× running buffer to 475 mL water for 1× SDS running buffer.
7. BenchMark Unstained Protein Ladder or Mark12 Unstained Standard (Thermo Fisher Scientific), stored at 4 °C.

2.4 Protein Visualization

1. SilverSNAP® Stain Kit II gel stain, sufficient reagents to stain up to 20 SDS-PAGE mini-gels (Pierce).
2. Fixing solution, 30% (v/v) aqueous ethanol containing 10% (v/v) aqueous acetic acid (>99.7%, Sigma-Aldrich, Saint Louis, MO).
3. Personal Densitometer SI (Molecular Dynamics).
4. Coomassie R-250, 1 L sufficient reagent for up to 50 mini-gels (Imperial Protein Stain, Pierce Biotechnology) (*see* **Note 9**).
5. ImageQuant™ software (Molecular Dynamics) .

2.5 In-Gel Digestion and Peptide Extraction

1. Imperial™ Protein Stain (Pierce, Thermo Fisher Scientific).
2. GridCutter (The Gel Company, San Francisco, CA).
3. Protein LoBind Tubes—1.5 mL microcentrifuge tube (low protein binding)—or Protein LoBind Plates.
4. 100 mM ammonium bicarbonate: 0.4 g NH4HCO3 in 50 mL water. Prepare fresh for every digest.
5. 50 mM ammonium bicarbonate/acetonitrile (1:1 v/v).
6. 50 mM ammonium bicarbonate in water.
7. 10 mM DTT (dithiothreitol) in 100 mM ammonium bicarbonate (7.5 mg DTT). Pre-weighed DDT can be stored at −20 °C.
8. 50 mM IAA (iodoacetamide) in 100 mM ammonium bicarbonate (10 mg IAA).
9. Trypsin solution: dissolve content of a 20 μg vial (V5111, 5 × 20 μg, Promega) in 1.5 mL of trypsin buffer (10 mM ammonium bicarbonate, 10% (v/v) acetonitrile) and keep on ice. The concentration of trypsin is 13 ng/μL. 2 mL trypsin buffer: 10 mM ammonium bicarbonate, 10% (v/v) acetonitrile.
10. 5% (v/v) formic acid in water.
11. Extraction buffer: 0.25 mL 5% (v/v) formic acid, 0.25 mL water, 0.5 mL acetonitrile.
12. Thermomixer temperature range up to 56 °C.
13. Thermostat oven at 37 °C.

14. Sonicator.
15. Vacuum centrifuge (lyophilizer).
16. STAGE Tips/desalting column – remove small disks (2–3) of C18 Empore filter using a 22 G flat-tipped syringe and ejecting disks into P200 pipette tips. Ensure that the disk is securely wedged in the bottom of the tip. Condition the columns (wet membrane) for each sample by using extraction buffer.
17. MS sample vials with snap lid (#THC11141190, snap ring vial with glass insert, Thermo Fisher Scientific) .

3 Methods

3.1 Blood Collection (See Notes 1–3)

1. It is important to obtain the required volume of blood using each specific blood collection tube type. This is essential to ensure the blood to anticoagulant ratio is not exceeded. The blood collection should be completed within 5 ± 2 min from the starting time.
2. After blood collection, gently mix the unit by inverting the tube 8–10 times.
3. Label the donor collection tube(s). If storage is required, do so immediately at −20 °C.
4. Thawing of the plasma sample on the day of use should be performed at 37 °C (not at room temperature or on ice) (*see* **Notes 10** and **11**). This is to prevent the formation of cryoprecipitate.
5. The protein concentration of the plasma sample when thawed should be determined. For consistency, the bicinchoninic acid (BCA) protein assay, using bovine serum albumin (BSA) as a standard, should be used [37].

3.2 Centrifugal Ultrafiltration

1. Prepare centrifugal filter membranes according to manufacturer's instructions by rinsing in 15 mL of HPLC grade water at 2000 × *g* for 10 min (*see* **Note 12**). Set centrifugal temperature to 20 °C. Twist off the lock cap and remove the inner tube (filtrate collector). Make sure not to touch or bend the membrane. If the device is not to be used immediately, store at 4 °C with Milli-Q water covering the membrane surface.
2. Dilute 100 μL of thawed plasma with 900 μL 10% (v/v) aqueous acetonitrile and allow to stand at RT for 2 min (*see* **Note 8**). Centrifuge each plasma sample (with a counterbalance) at 14,000 × *g* for 2 min at RT to precipitate any insoluble material that may clog the filters.
3. Apply the supernatant to the prepared centrifugal filter(s) and samples placed in an M4 swing-bucket rotor and centrifuged

(with a counterbalance) at 4000 × *g* for 35 min at 20 °C (*see* **Note 5**). A small aliquot (50 µL) of the sample was set aside in order to assess LMF recovery. This sample is stored at −80 °C.

4. The retentate (retained fraction, ~5% initial volume) should be removed and stored separately. The filtrate (flow-through fraction, ~90–95% initial volume) volume can either be removed by pipette or the filtrate recovered by inverting the tube and centrifuging at 2000 × *g* for 1 min.
5. The LMF recoveries of the filter membrane can be analyzed by BCA protein assay [37], comparing the initial plasma concentration to the concentration and volume of both the retained (retentate) and filtered (filtrate) samples. Typical recoveries for this experiment should be in the range 94–97% (three experimental replicates) (Table 1). Retentate samples are stored at −80 °C.
6. The plasma LMF filtrates are lyophilized to dryness by centrifugal lyophilization and resuspended in Laemmli nonreducing sample buffer.

3.3 SDS-PAGE Analyses

1. A plasma LMF protein sample (50 µg) is mixed with prewarmed NuPAGE® LDS sample buffer (in the ratio sample/buffer, 2:1).

Table 1
Plasma peptidomefractionation efficiency using various centrifugal ultrafiltration devices

Centrifugal ultrafiltration membrane	Protein recoveries[a]		
	Filtrate (mg, %)	Retentate (mg, %)	Overall recovery (mg, %)
Membrane A Microcon®, Millipore (30K)	0.20 (3.0%)	5.6 (83.6%)	5.8 (86.6%)
Membrane B Centrisart®, Sartorius (20K)	2.9 (43.3%)	3.4 (50.7%)	6.3 (94.0%)
Membrane C Amicon Ultra®, Millipore (30K)	1.4 (20.9%)	4.8 (71.6%)	6.2 (92.5%)
Membrane D Vivaspin®, Sartorius (20K)	0.3 (4.0%)	6.1 (91.0%)	6.3 (94.0%)

[a]100 µL plasma (67 mg/mL, 6.7 mg) was diluted to 1000 µL (various buffers utilized, refer to Subheadings 2 and 3) and loaded onto each prepared filtration device. The amount of protein recovered in the filtrate and retentate is expressed as a percentage of the initial plasma protein concentration loaded onto each filtration device. The volume of the filtrates and retentates have been adjusted to a total volume of 1000 µL; values shown in parentheses represent the percentage of the initial plasma protein volume loaded onto each filtration device. Overall recovery represents the summation of protein recovery in both the filtrate and retentate for each membrane; values shown in parentheses represent the summation of filtrate and retentate recoveries expressed as a percentile. Each value is representative of experiments performed in triplicate

2. The sample mixture is heated for 5 min on a heat block at 95 °C and cooled (2 min) prior to sample loading.
3. Separation is performed using a precast 12-well SDS polyacrylamide gel (4–12% Bis-Tris precast gel).
4. 500 mL of 1 × MES SDS running buffer is prepared – approximately 200 mL in the upper (inner) buffer compartment and 300 mL in the lower (outer) buffer compartment.
5. Samples are loaded into defined gel lanes. BenchMark protein standards (5 μL) are used for molecular weight comparison.
6. Protein separation is performed at 150 V (constant voltage) until tracking dye reaches the bottom of the gel (approx. 75 min).
7. Immediately following electrophoresis the gel should be washed with water and stained with colloidal Coomassie R-250, as described elsewhere [38]. Destain background with water.

3.4 In-Gel Digestion and Peptide Extraction

1. After staining, gel sections are excised (using either scalpel or gel excision tool with slices ~1.0–1.5 mm) from a single lane.
2. Individual gel lanes are placed on a clean glass plate and cut into equal slices (20 × 2 mm) using a GridCutter or clean scalpel and individual gel slices subjected to in-gel reduction, alkylation, and trypsinization.
3. For sample reduction, microcentrifuge tubes are centrifuged briefly, heated using a Thermomixer at 56 °C (700 rpm for 15 min), and the solution discarded. 200 μL acetonitrile is added and gel pieces should shrink and take an opaque white color. Remove acetonitrile and let air-dry for 5 min in thermomixer at 56 °C. Add 50 μL fresh DTT solution and incubate at 56 °C for 30 min.
4. For sample alkylation, set thermomixer to 22 °C and remove DTT solution completely. Immediately add 70 μL IAA solution, and incubate for 20 min in thermomixer 22 °C (700 rpm) covered by aluminum foil.
5. The IAA solution is then removed, 300 μL acetonitrile added for 2 min, then removed. Add 100 μL of 50 mM ammonium bicarbonate/acetonitrile (1/1) and incubate for 30 min in thermomixer with light mixing at RT. The sample should be lightly centrifuged and supernatant removed.
6. 300 μL acetonitrile is added to the samples, removed completely, and air-dried for 5 min.

3.5 Tryptic Digestion and STAGE Tips Desalting

1. For tryptic digestion, add enough trypsin to cover the dry gel pieces (typically, 50–60 μL depending on gel volume). Store all samples immediately on ice. After 30 min, check if all solution is absorbed and add more trypsin, if necessary. Gel pieces should be completely covered with trypsin.

2. Leave gel pieces for another 30 min to saturate with trypsin and add 20 μL of 50 mM ammonium bicarbonate to cover the gel pieces.
3. Place tubes with gel pieces into the thermostat oven and incubate samples overnight at 37 °C.
4. Withdraw supernatants to low protein-binding tubes, and use a pipette with fine gel loader tip (reuse these tips for all following peptide collection steps).
5. Add extraction buffer (100 μL) to each tube and sonicate for 15 min. Collect supernatants into the corresponding tube. Repeat.
6. Dry down samples (covered with Vacufilm with four pin holes) in vacuum centrifuge for 15 min.
7. For STAGE Tips desalting, prepare as many desalting columns as necessary by punching out small disks (2–3) of C18 Empore filter using a 22 G flat-tipped syringe and ejecting the disks into P200 pipette tips. Ensure that the disks are securely wedged in the bottom of the tip. Careful preparation of these STAGE Tips devices will ensure effective filtration.
8. Condition columns by forcing methanol through (50 μL) and check whether the STAGE Tips are leaky.
9. Remove any remaining organic solvent in the column by forcing buffer A through the disk (×2) (40 μL).
10. Adjust pH of peptide sample to pH <2.5 using 2% (v/v) TFA.
11. Force the acidified peptide sample through the C18-StageTip column.
12. Wash the column with buffer A. Elute the peptides from the C18 material using 20–30 μL buffer B. Elute directly into a microfuge tube or autosampler plate. Repeat elution.
13. Carefully dry samples in the SpeedVac without heating, until all acetonitrile has evaporated (~2–3 μL final volume). Note to not completely overdry/dehydrate peptide sample due to issues with sample loss and resolubilization.
14. Mix the sample (1:1) with sample buffer up to 8 μL.
15. Withdraw to MS specific vial for analysis (typically 3 μL loaded representing ~3 μg sample). To determine peptide concentration, a spectrophotometer analysis can be obtained based on 215 nm absorbance and comparison with known standard.
16. Short-term storage at 4 °C (within 2 weeks) or long-term −80 °C (up to 18 months).

3.6 MS/MS Analysis

1. RP-HPLC is performed on a nanoAcquity® (C18) 150 × 0.15 mm i.d. reversed-phase UPLC column (Waters), using an Agilent 1200 HPLC, coupled online to an LTQ-Orbitrap mass spectrometer equipped with a nanoelectrospray ion source [39].
2. RP-HPLC column is developed with a linear 60 min gradient with a flow rate of 0.8 μL/min at 45 °C from 0 to 100% solvent B where solvent A was 0.1% (v/v) aqueous formic acid and solvent B was 0.1% (v/v) aqueous formic acid/60% (v/v) acetonitrile.
3. The mass spectrometer was operated in data-dependent mode where the top 20 most abundant precursor ions in the survey scan (300–2500 Th) were selected for MS/MS fragmentation. Survey scans were acquired at a resolution of 120,000 at m/z 400. Unassigned precursor ion charge states and singly charged species were rejected, and peptide match disabled. The isolation window was set to 3 Th and selected precursors fragmented by CID with normalized collision energies of 25. Maximum ion injection times for the survey scan and MS/MS scans were 20 and 60 ms, respectively, and ion target values were set to 3E6 and 1E6, respectively. Dynamic exclusion was activated for 90 s.

3.7 Data Processing and Analysis

1. Raw data was processed using MaxQuant [40] (v1.1.1.25) and searched with Andromeda using human-only (UniProt) sequence database.
2. Data was searched with a parent tolerance of 10 ppm, fragment tolerance of 0.5 Da, and minimum peptide length 7. Database search parameters as follows: fixed modification, carboxymethylation of cysteine (+58 Da), variable modifications, NH_2-terminal acetylation (+42 Da), methionine oxidation (+16 Da).
3. FDR was 1% at the peptide and protein levels, and data examined with label-free quantitation (LFQ) [41]. LFQ intensities for all unique and razor peptides were included, with zero intensity values replaced with a constant value of 1 to calculate fold change ratios. LFQ intensity values were normalized for protein length and fold change ratios calculated.
4. Contaminants and reverse database identifications were excluded from further data analysis. Proteins commonly identified in both replicate experiments were used to compare against other cell samples.
5. Proteins were correlated with prediction of nonclassical protein secretion (SecretomeP 2.0) (http://www.cbs.dtu.dk/services/SecretomeP/) and also the Secreted Protein Database (http://spd.cbi.pku.edu.cn/spd_search.php).
6. Other resources to classify identified proteins based on several predictive algorithms included SignalP 4.1 (http://www.cbs.

dtu.dk/services/SignalP/), TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/), Gene Ontology (GO) (http://www.geneontology.org/index.shtml?all/), and the UniProt database (http://www.uniprot.org/).

7. For an in-depth resource of the Human Plasma Proteome Project, including extensive collection of *raw* data, refer to PeptideAtlas (http://www.peptideatlas.org/hupo/hppp/).

4 Notes

1. **Sample hemolysis**. The release of cellular material due to hemolysis into serum/plasma may introduce additional confounding factors. We recommend that if hemolysis (pink to red tinge in serum/plasma sample) is observed following centrifugation, this information should be recorded. It is recommended that hemolyzed samples are not used for proteomic/peptidomic analyses.
2. **Monitoring pre/post-analytical variation**. In 2005, the HUPO PPP report detailed an extensive analysis of the variables affect the stability of plasma [42]. These included (a) the anticoagulant used in collection tube types (e.g., EDTA and ascorbate), (b) sample processing times, (c) temperatures at which blood specimens were processed and stored, (d) sample storage parameters, and (e) thaw-refreeze cycles, associated with obtaining human plasma and serum samples for proteomic analyses directed toward clinical research. It is of upmost importance that for diagnostic use, these variables are controlled and monitored at all times, from blood collection as an anticoagulated or coagulated source to processing, handling, and storage [43–45]. Recently, it has been shown that biomarker validation studies should use standardized collection conditions and use multiple control groups to detect and correct for potential biases associated with sample collection [46].
3. **Data points**. For blood handling, it is important to note also (a) the date and time of blood collection, (b) the number and volume of samples/aliquots prepared, (c) the date and time placed at −80 °C, (d) the date and time of shipping, (e) any freeze-thaw cycles that occur, and (f) variations or deviations from the standard operating protocol, problems, or issues which arise.
4. **Centrifugal ultrafiltrationmembrane devices**. A wide range of centrifugal filters are commercially available for concentrating and filtrating protein solutions, removing small solutes, and/or buffer exchanging. These devices consist (mostly) of two chambers separated by a semipermeable membrane. These

membranes can be composed of different chemistries and different orientations depending on their application (*see* **Note 7**). Under centrifugal force, solvent and solute molecules smaller than the NMWL readily pass through the membrane (filtrate) (*see* **Note 6**). Vertical or angular membrane configuration reduces concentration polarization (membrane fouling) and allows high flow rates for optimal solvent passage even with high proteinaceous solutions. The direction of the centrifugal force and flow rate of solute differs between membrane devices used. We evaluated plasma peptidome fractionation efficiency using various centrifugal ultrafiltration devices, noting significant differences in the protein recoveries of different devices and membrane types (Table 1). The separation efficiency of these four filtration devices were assessed by 1D–SDS-PAGE using a starting volume of 100 μL of human plasma. As noted in Fig. 2, significant amounts of high-molecular and abundant

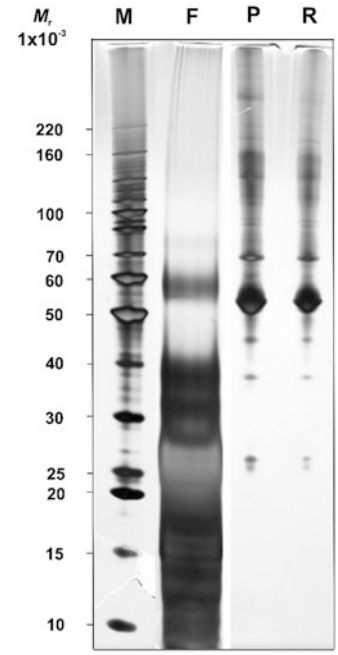


Fig. 2 Low-molecular-weight/peptidome analysis of human plasma. One hundred microliters (100 μL) of platelet-poor plasma was diluted 1:9 with 900 μL of 10% ACN, pH 8.5 v/v. This sample was fractionated using a prepared low protein-binding Vivaspin 2 20K MWCO membrane filter at 4000 × *g* until 95% of the input plasma had passed through the 20K (cellulose triacetate) filter. Aliquots of whole (*Lane P*) or ultrafiltered plasma (filtrate, *Lane F*, and retentate, *Lane R*) were subjected to 1-DE and stained using silver staining. *Lane M*, BenchMark molecular weight marker

plasma proteins are retained by the optimized ultrafiltration membrane device (device D), enabling selective enrichment and recovery of the low-molecular-weight plasma proteome. Additional information regarding centrifugal ultrafiltration membrane devices and their membrane chemistry can be obtained from www.millipore.com/andwww.sartorius.com/.

5. **Optimized centrifugal ultrafiltration**. Conditions for each plasma sample should be optimized. Conditions provided in this protocol are the combined effect of analyzing multiple filter membrane units, with conditions optimized with respect to plasma buffer and solvent compositions, centrifugal force, duration, and temperature. Typically, plasma LMF should represent 95% of the initial supernatant applied to the filtration devices. The amount of protein recovered in the filtrate and retentate can be calculated as a percentage of the initial plasma protein concentration loaded (Table 1).
6. **Appropriate membranes—selecting the NMWL**. Ultrafiltration membranes are not absolute in their pore size (NMWL) ratings. Separation occurs as a result of differences in the filtration rate of different components across the membrane in response to a given pressure. Unlike UF membranes, microporous membranes have a precisely controlled pore size that ensures quantitative retention of particles and biomolecules greater than the pore size of the membrane. In selecting the most effective membrane for filtration applications, a rule has been developed to rapidly calculate the appropriate membrane pore size (NMWL). It is a simple calculation based on the molecular weight of the desired protein to be concentrated or removed in the retentate unit (upper level of the membrane apparatus). The "rule of 1.5–2" requires a membrane cutoff approximately two times smaller than the desired proteins' molecular weight. For example, to remove proteins of ~65,000 MW and greater, use a 30,000 NMWL regenerated cellulose membrane. Typically, this results in >90–95% recovery of the filtrate, containing proteins/peptides <65,000 MW. Other factors to consider when determining an optimal membrane include flow rate, also known as flux, solute concentration, solute composition, and temperature.
7. **Centrifugal ultrafiltrationmembrane chemistries**. For a detailed overview, please refer to [47]:

   Polyethersulfone—General purpose membrane, providing excellent performance with most solutions when retentate recovery is of primary importance. Polyethersulfone membranes exhibit no hydrophobic or hydrophilic interactions and are usually preferred for their low fouling characteristics, exceptional flux, and broad pH range.

Cellulose triacetate—High hydrophilicity and very low non-specific binding characterize this membrane. These membranes are preferred for sample cleaning and protein removal and when high recovery of the filtrate solution is of primary importance.

Regenerated cellulose/Hydrosart—These membranes demonstrates the same properties as regenerated cellulose, but with the added benefit of enhanced performance characteristics and extremely low protein binding, making it the membrane of choice for applications such as concentration and desalting of immunoglobulin fractions.

8. **Disrupting protein-protein interactions**. A low concentration of organic solvent (typically, 5–10% acetonitrile) is added to buffers to disrupt high-M_r protein-protein interactions. For chemical compatibility of membranes, be careful to read each company's manual prior to operation (based on 2 h membrane contact time). Normally, small uncomplexed proteins and peptides (i.e., less than 30 K) are rapidly cleared from the circulation through enzymatic degradation and uptake by the reticuloendothelial system or by glomerular filtration, which discriminates on the basis of molecular size and charge [48]. It is believed that the circulation half-life of the LMF fraction is directly related to its binding affinity to large high-abundance carrier proteins [7, 10].
9. **Coomassie dye staining**. The Coomassie dyes (R-250 and G-250) bind to proteins through ionic interactions between dye sulfonic acid groups and positive protein amine groups. Coomassie R-250, the more commonly used on the two dyes, can detect protein levels down to 0.1 μg. Additionally, Coomassie R-250 does not require methanol/acetic acid fixation and destaining.
10. **Cryoprecipitateformation**. A cryoprecipitate is often formed if the fresh-frozen plasma unit is slowly thawed at temperatures just above freezing (1–6 °C), typically in a water bath or a refrigerator. The product is then centrifuged at low speed (typically 5000 × *g*) to remove the majority of the precipitate. The formation of the cryoprecipitate can be avoided by thawing at 37 °C.
11. **Plasma thaw process**. Thawing of plasma can be achieved in various ways, the most common of which uses a recirculating water bath. This carries a risk of bacterial contamination and must be maintained according to a controlled sterile environment. Denaturation of plasma proteins can be avoided by using a dry-heating apparatus.
12. **Prerinsing membranes**. Most ultrafiltration membrane devices contain trace amounts of glycerine/sodium azide. If

this interferes with subsequent sample analysis, prerinse the device extensively with buffer or Milli-Q water through the concentrator. If interference still persists, rinse the membrane with 0.1 M NaOH followed by repeated centrifugation with buffer or Milli-Q water.

Acknowledgments

This work was supported, in part, by the National Health and Medical Research Council (NHMRC) of Australia project grant #1057741 (R.J.S.), La Trobe University Leadership RFA Grant (D.W.G.), and La Trobe Institute for Molecular Science Biomedical Fellowship (D.W.G.). We acknowledge the La Trobe University-Comprehensive Proteomics Platform for providing infrastructure and expertise.

References

1. Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1:845–867
2. Omenn GS, States DJ, Adamski M et al (2005) Overview of the HUPO plasma proteome project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics 5:3226–3245
3. Putnam RW (1975) The plasma proteins. Academic Press, New York, NY
4. Moritz RL, Ji H, Schutz F et al (2004) A proteome strategy for fractionating proteins and peptides using continuous free-flow electrophoresis coupled off-line to reversed-phase high-performance liquid chromatography. Anal Chem 76:4811–4824
5. Bjorhall K, Miliotis T, Davidsson P (2004) Comparison of different depletion strategies for improved resolution in proteomic analysis of human serum samples. Proteomics 5:307–317
6. Lowenthal MS, Mehta AI, Frogale K et al (2005) Analysis of albumin-associated peptides and proteins from ovarian cancer patients. Clin Chem 51:1933–1945
7. Zhou M, Lucas DA, Chan KC et al (2004) An investigation into the human serum "interactome". Electrophoresis 25:1289–1298
8. Lopez MF, Mikulskis A, Kuzdzal S et al (2005) High-resolution serum proteomic profiling of Alzheimer disease samples reveals disease-specific, carrier-protein-bound mass signatures. Clin Chem 51:1946–1954
9. Albrethsen J, Bogebo R, Gammeltoft S et al (2005) Upregulated expression of human neutrophil peptides 1, 2 and 3 (HNP 1-3) in colon cancer serum and tumours: a biomarker study. BMC Cancer 5:8
10. Liotta LA, Petricoin EF (2006) Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. J Clin Invest 116:26–30
11. Lai ZW, Petrera A, Schilling O (2015) The emerging role of the peptidome in biomarker discovery and degradome profiling. Biol Chem 396:185–192
12. Greening DW, Kapp EA, Ji H et al (2013) Colon tumour secretopeptidome: insights into endogenous proteolytic cleavage events in the colon tumour microenvironment. Biochim Biophys Acta 1834:2396–2407
13. Mason SD, Joyce JA (2011) Proteolytic networks in cancer. Trends Cell Biol 21: 228–237
14. Antwi K, Hostetter G, Demeure MJ et al (2009) Analysis of the plasma peptidome from pancreas cancer patients connects a peptide in plasma to overexpression of the parent protein in tumors. J Proteome Res 8:4722–4731
15. Bedin C, Crotti S, Ragazzi E et al (2015) Alterations of the plasma Peptidome profiling in colorectal cancer progression. J Cell Physiol 231(4):915-925
16. Shen Y, Tolic N, Liu T et al (2010) Blood peptidome-degradome profile of breast cancer. PLoS One 5:e13133
17. Bassani-Sternberg M, Barnea E, Beer I et al (2010) Soluble plasma HLA peptidome as a

potential source for cancer biomarkers. Proc Natl Acad Sci U S A 107:18769–18776

18. Karpova MA, Moshkovskii SA, Toropygin IY et al (2010) Cancer-specific MALDI-TOF profiles of blood serum and plasma: biological meaning and perspectives. J Proteomics 73:537–551
19. Mahboob S, Mohamedali A, Ahn SB et al (2015) Is isolation of comprehensive human plasma peptidomes an achievable quest? J Proteomics 127:300–309
20. Bery A, Leung F, Smith CR et al (2014) Deciphering the ovarian cancer ascites fluid peptidome. Clin Proteomics 11:13
21. Di Meo A, Pasic MD, Yousef GM (2016) Proteomics and peptidomics: moving toward precision medicine in urological malignancies. Oncotarget 7(32):52460–52474
22. Gelman JS, Sironi J, Berezniuk I et al (2013) Alterations of the intracellular peptidome in response to the proteasome inhibitor bortezomib. PLoS One 8:e53263
23. Jia C, Lietz CB, Ye H et al (2013) A multiscale strategy for discovery of novel endogenous neuropeptides in the crustacean nervous system. J Proteomics 91:1–12
24. Kalaora S, Barnea E, Merhavi-Shoham E et al (2016) Use of HLA peptidomics and whole exome sequencing to identify human immunogenic neo-antigens. Oncotarget 7:5110–5117
25. Labots M, Schutte LM, van der Mijn JC et al (2014) Mass spectrometry-based serum and plasma peptidome profiling for prediction of treatment outcome in patients with solid malignancies. Oncologist 19:1028–1039
26. Sasaki K, Sato K, Akiyama Y et al (2002) Peptidomics-based approach reveals the secretion of the 29-residue COOH-terminal fragment of the putative tumor suppressor protein DMBT1 from pancreatic adenocarcinoma cell lines. Cancer Res 62:4894–4898
27. E-Kobon T, Thongararm P, Roytrakul S et al (2016) Prediction of anticancer peptides against MCF-7 breast cancer cells from the peptidomes of *Achatina fulica* mucus fractions. Comput Struct Biotechnol J 14: 49–57
28. Wang F, Zhu J, Hu L et al (2012) Comprehensive analysis of the N and C terminus of endogenous serum peptides reveals a highly conserved cleavage site pattern derived from proteolytic enzymes. Protein Cell 3:669–674
29. Xu Z, Wu C, Xie F et al (2015) Comprehensive quantitative analysis of ovarian and breast cancer tumor peptidomes. J Proteome Res 14:422–433
30. Harper RG, Workman SR, Schuetzner S et al (2004) Low-molecular-weight human serum proteome using ultrafiltration, isoelectric focusing, and mass spectrometry. Electrophoresis 25:1299–1306
31. Tirumalai RS, Chan KC, Prieto DA et al (2003) Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics 2:1096–1103
32. Di Girolamo F, Alessandroni J, Somma P et al (2009) Pre-analytical operating procedures for serum low molecular weight protein profiling. J Proteomics 73(3):667–677
33. Tammen H, Schulte I, Hess R et al (2005) Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. Proteomics 5:3414–3422
34. Zheng X, Baker H, Hancock WS (2006) Analysis of the low molecular weight serum peptidome using ultrafiltration and a hybrid ion trap-Fourier transform mass spectrometer. J Chromatogr A 1120:173–184
35. Hu L, Li X, Jiang X et al (2007) Comprehensive peptidome analysis of mouse livers by size exclusion chromatography prefractionation and nanoLC-MS/MS identification. J Proteome Res 6:801–808
36. Jung WW, Phark S, Oh S et al (2009) Analysis of low molecular weight plasma proteins using ultrafiltration and large gel two-dimensional electrophoresis. Proteomics 9:1827–1840
37. Smith PK, Krohn RI, Hermanson GT et al (1985) Measurement of protein using bicinchoninic acid. Anal Biochem 150:76–85
38. Neuhoff V, Arold N, Taube D et al (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie brilliant blue G-250 and R-250. Electrophoresis 9:255–262
39. Greening DW, Simpson RJ (2010) A centrifugal ultrafiltration strategy for isolating the low-molecular weight (< or =25K) component of human plasma proteome. J Proteomics 73:637–648
40. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372
41. Luber CA, Cox J, Lauterbach H et al (2010) Quantitative proteomics reveals subset-specific

viral recognition in dendritic cells. Immunity 32:279–289

42. Rai AJ, Gelfand CA, Haywood BC et al (2005) HUPO plasma proteome project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics 5:3262–3277
43. Banks RE, Stanley AJ, Cairns DA et al (2005) Influences of blood sample processing on low-molecular-weight proteome identified by surface-enhanced laser desorption/ionization mass spectrometry. Clin Chem 51:1637–1649
44. Ferguson RE, Hochstrasser D, Banks R (2007) Impact of preanalytical variables on the analysis of biological fluids in proteomic studies. Proteomics Clin Appl 1:739–746
45. West-Norager M, Kelstrup CD, Schou C et al (2007) Unravelling in vitro variables of major importance for the outcome of mass spectrometry-based serum proteomics. J Chromatogr B Analyt Technol Biomed Life Sci 847:30–37
46. Thorpe JD, Duan X, Forrest R et al (2007) Effects of blood collection conditions on ovarian cancer serum markers. PLoS One 2:e1281
47. Cheryan M (1998) Ultrafiltration and microfiltration handbook. CRC Press, Boca Raton, FL
48. Torchilin VP, Lukyanov AN (2003) Peptide and protein drug delivery to and into tumors: challenges and solutions. Drug Discov Today 8:259–266

Chapter 7

In-Depth, Reproducible Analysis of Human Plasma Using IgY 14 and SuperMix Immunodepletion

Lynn A. Beer, Bonnie Ky, Kurt T. Barnhart, and David W. Speicher

Abstract

Identification of cancer and other disease biomarkers in human plasma has been exceptionally challenging due to the complex nature of plasma and the presence of a moderate number of high- and medium-abundance proteins which mask low-abundance proteins of interest. As a result, immunoaffinity depletion formats combining multiple antibodies to target the most abundant plasma proteins have become the first stage in most plasma proteome discovery schemes. This protocol describes the use of tandem IgY 14 and SuperMix immunoaffinity depletion to reproducibly remove >99% of total plasma protein. This greatly increases the depth of analysis of human plasma proteomes. Depleted plasma samples can then be analyzed in a single high-resolution LC-MS/MS run on a Q Exactive Plus mass spectrometer, followed by label-free quantitation. If greater depth of analysis is desired, the depleted plasma can be further fractionated by separating the sample for a short distance on a 1D SDS gel and cutting the gel into uniform slices prior to trypsin digestion. Alternatively, the depleted plasma can be reduced, alkylated, and digested with trypsin followed by high-pH reversed-phase HPLC separation.

Key words Proteomics, Plasma biomarkers, Major protein depletion, SuperMix

1 Introduction

Human plasma is considered to be a promising resource for proteomic discovery of disease biomarkers because most cells in the body shed proteins into the blood. Accordingly, the plasma proteome is expected to contain valuable information regarding the physiological condition of most tissues and organs in an individual [1–3]. However, proteomic discovery of low-abundance biomarkers for cancer and other diseases has been extremely challenging due to the complex nature of the plasma proteome. Specifically, the presence of a few very high-abundance (mg/mL) proteins, i.e., albumin and immunoglobulins, transferrin, etc., constitutes approximately 65–80% of total plasma protein, while most of the remaining 20% of plasma protein is comprised of a moderate number of medium-abundance (μg/mL) proteins. Additionally, the

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_7,

molecular heterogeneity of many proteins can further impede the identification of low-abundance biomarkers [3, 4].

To overcome the hurdles caused by high- and medium-abundance plasma proteins, immunoaffinity depletion has become a major focus in proteomic studies and is generally the first step in most proteomic discovery workflows. The Multiple Affinity Removal System (MARS), developed by Agilent in 2003, was one of the first commercial products to combine polyclonal antibodies targeting more than one or two abundant proteins from plasma [5, 6]. More recently, manufacturers have produced antibody columns against 7, 12, 14, or 20 abundant plasma proteins in both spin-column and LC-column formats. While many multicomponent depletion formats are available, antibody columns that remove 14, 20, or more medium- and high-abundance proteins are most effective at removing ~95% of total protein and hence substantially improving protein profiling capacities [7–9]. In our experience, most commercially available high-level depletion columns (Table 1) perform comparably in that they remove >98–99% of most targeted proteins and ~95% of total protein. However, if serum albumin is present at 40 mg/mL and 99% is removed, the remaining 0.4 mg/mL is still a medium-abundance protein. Additionally, even after the highest abundance proteins are removed, the remaining medium-abundance proteins can still mask the identification of lower abundance proteins. Thus, in 2008 a new technology comprised of avian polyclonal IgY antibodies was developed by GenWay Biotech [8]. This novel SuperMix column included a mixture of IgY antibodies that were produced to the flow-through of their "Top 12" IgY column. More recently, a "Top 14" column has been developed, and when the IgY 14 and SuperMix columns are connected in tandem, they can effectively capture 14 high-abundance proteins and up to 60 additional medium-abundance proteins [10], effectively removing >99% of the total protein. This reduction in total protein allows larger amounts of depleted plasma volumes to be analyzed in downstream analytical separations and greatly improves depth of analysis in proteome studies. For instance, when we analyzed depleted plasma samples that were digested with trypsin and analyzed in a single high-resolution LC-MS/MS analysis, ~500 additional proteins were identified in the IgY 14/SuperMix depleted sample compared with the other depletion methods (Fig. 1a). Importantly, when these datasets were compared to a list of low-abundance plasma proteins from the Plasma Proteome Database (PPD) that included 879 proteins with reported concentrations of 100 ng/mL or less [11], the IgY 14/SuperMix flow-through identified nearly 3–6 times more low-abundance proteins compared with the other depletion methods (Fig. 1b).

However, a major concern involved with extensive depletion of plasma is the loss of nontargeted proteins of interest, especially

Table 1
High-abundance proteins targeted in top 14, top 20, and SuperMix columns[a]

Gene name	Targeted protein	ProteoPrep 20 (Sigma-Aldrich)	MARS 14 (Agilent)	IgY 14 (Sigma-Aldrich)	SuperMix[b] (Sigma Aldrich)
ALB	Albumin	+	+	+	
IGHG	IgG	+	+	+	
SERPINA1	Alpha 1-antitrypsin	+	+	+	
IGHA	IgA	+	+	+	
IGHM	IgM	+	+	+	
TF	Transferrin	+	+	+	
HP	Haptoglobulin	+	+	+	
FGG, FGA	Fibrinogen	+	+	+	+
A2M	Alpha 2-Macroglobulin	+	+	+	
C3	Complement C3	+	+	+	+
ORM1,ORM2	Orosomucoid	+	+	+	
APOA1	Apolipoprotein AI	+	+	+	
APOA2	Apolipoprotein A2	+	+	+	
APOB	Apolipoprotein B	+		+	
TTR	Transthyretin	+	+		
CP	Ceruloplasmin	+			+
C4	Complement C4	+			+
C1Q	Complement C1q	+			+
IGHD	IgD	+			
PLG	Plasminogen	+			

[a]ProteoPrep 20, MARS 14, and IgY 14 targets are as listed by the manufacturers. SuperMix targets are as cited in ref. 10
[b]The additional SuperMix targeted proteins are summarized in Fig. 2a

low-abundance proteins that may bind either nonspecifically to antibodies on the column or to targeted carrier proteins such as albumin [12]. Also of concern is the reproducibility of the depletion method. Assuming that some nontargeted proteins will inevitably be removed by the column, the degree to which they are removed from run to run becomes important. That is, if a protein is present at equal levels in two plasma samples and 80% of the protein is removed in one sample and 40% is removed in another, the residual protein may be falsely interpreted as an apparent biomarker as the difference of that protein across the samples would appear to be about threefold.

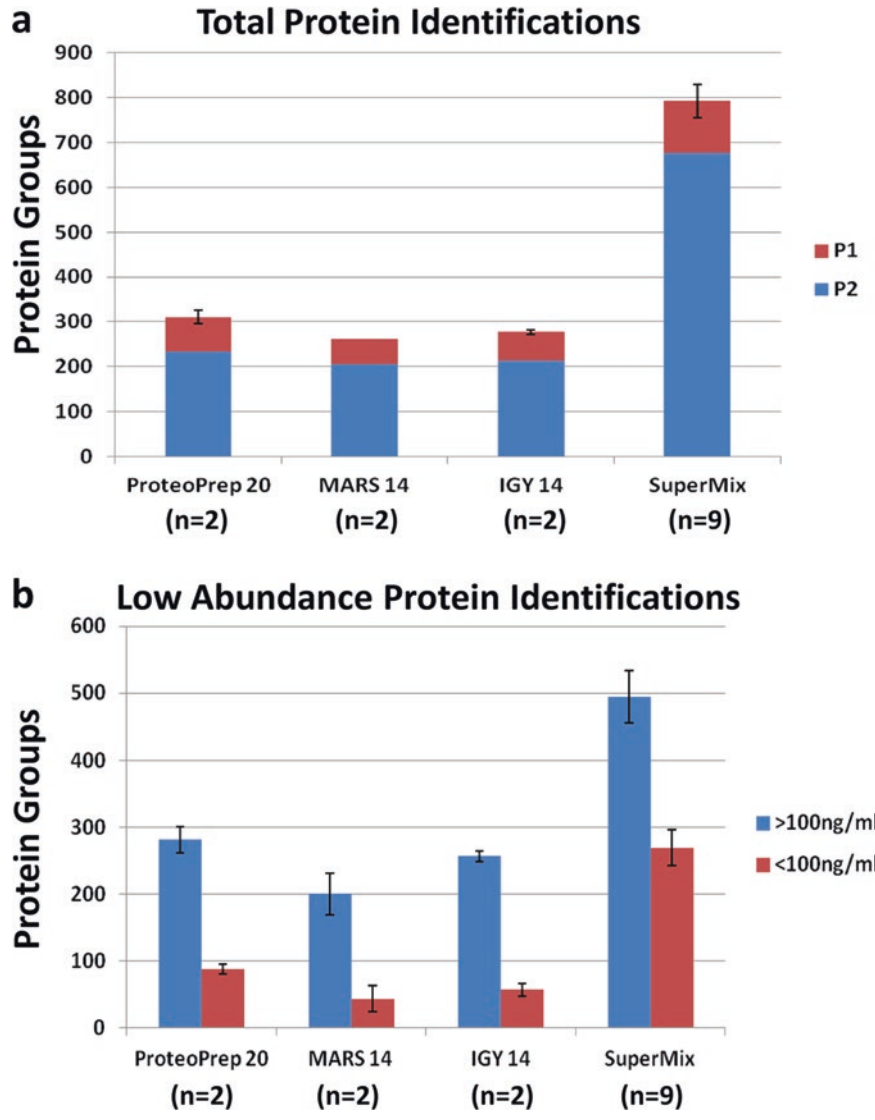


Fig. 1 Identification of total proteins (**a**) and reported low-abundant (<100 ng/mL) proteins (**b**) using alternative depletion methods in a single fraction plasma proteome analysis. (**a**) Human plasma samples were depleted using ProteoPrep 20 LC (n = 2), Agilent 14 LC (n = 2), IgY 14 LC (n = 2), or tandem IgY 14/SuperMix LC columns (n = 9). Following depletions, samples were run on SDS-PAGE and digested with trypsin, and equal amounts (1 μg) of digested plasma was analyzed in a single 4 h LC-MS/MS run on a Q Exactive instrument followed by label-free analysis with MaxQuant software. P1 = proteins identified by a single peptide, P2 = proteins identified by two or more peptides. (**b**) Protein groups from panel (**a**) having at least two-peptide identification were cross-referenced to a list of low-abundance plasma proteins from the Plasma Proteome Database (PPD) that included 879 proteins with reported concentrations of 100 ng/mL or less. Averages and standard error bars are shown in both panels

The following protocol describes immunoaffinity depletions using the IgY 14/SuperMix tandem columns. In an independent analysis, we have assessed the reproducibility of the IgY 14/SuperMix flow-through and the SuperMix bound fractions from triplicate depletions of three plasma pools. Depleted plasma samples were run on short (0.5 cm) 1D SDS gels, digested with trypsin, and analyzed in a single high-resolution LC-MS/MS run on a Q Exactive Plus mass spectrometer, followed by subsequent label-free quantitation.

2 Materials

2.1 Affinity Depletion of Plasma

1. Human plasma (*see* **Note 1**).
2. HPLC or FPLC system (e.g., an ÄKTApurifier UPC 10, GE Healthcare).
3. Seppro® IgY 14 LC5 column (Sigma-Aldrich, 12.7 × 39.5 mm inner diameter, 5 mL bed volume, storage 2–8 °C).
4. Seppro® SuperMix LC2 column (6.4 × 63 mm inner diameter, 2 mL bed volume, storage 2–8 °C).
5. 10× dilution buffer: 100 mM Tris–HCl with 1.5 M NaCl, pH 7.4 (storage, 2–8 °C).
6. 10× stripping buffer: 1 M glycine, pH 2.5 (storage, 2–8 °C).
7. 10× neutralization buffer: 1 M Tris–HCl, pH 8.0 (storage, 2–8 °C).
8. 2 mg/mL Leupeptin.
9. 1 mg/mL Pepstatin-A.
10. 150 mM phenylmethylsulfonyl fluoride (PMSF).
11. Stericup® GP Express Plus membrane, 0.22 μm filter units, 500 mL capacity.
12. Amicon Ultrafree®-MC, 0.45 μm microcentrifuge filters, 500 μL capacity.
13. Milli-Q® (Millipore) water or equivalent.
14. Microcentrifuge.

2.2 Protein Concentration

1. Amicon Ultracel®-10 K NMWL centrifugal filter units, 4 mL (Millipore).
2. Refrigerated centrifuge (4 °C) with capacity for up to 15 mL centrifuge tubes.
3. Colorimetric protein assay (e.g., BCA Protein Assay Kit, Thermo Fisher Scientific).

2.3 Reduction/Alkylation Prior to SDS-PAGE

1. 10% (w/v) SDS.
2. 1 M Tris–HCl, pH 8.5.

3. 1 M dithiothreitol.
4. 0.5 M iodoacetamide in 100 mM Tris–HCl, pH 8.6.
5. 37 °C thermostatically controlled incubator/shaker.

2.4 1D SDS-PAGE

1. 5× protein-solubilizing buffer: 1 M sucrose, 15% (w/v) SDS, 313.5 mM Tris–HCl, 11 mM Na_2EDTA, 5% (v/v) 2-mercaptoethanol, and 2% (v/v) saturated bromophenol blue solution.
2. NuPAGE® MES running buffer, 20× (Thermo Fisher Scientific), storage 2–8 °C.
3. NuPAGE® 10% Bis-Tris gel, 1 mm, ten-well (Thermo Fisher Scientific), storage 2–8 °C.
4. 1 cc insulin syringe.
5. India ink.
6. Mini Gel Tank.
7. Heat block set to 90 °C.

2.5 Colloidal Coomassie® Blue Staining

1. Novex® Colloidal Blue Staining Kit (Thermo Fisher Scientific). Individual components include:
   (a) Stainer A.
   (b) Stainer B.
2. Fixing solution: 50% (v/v) methanol/10% (v/v) acetic acid in water.
3. Methanol (Optima® 0.2 μm filtered)
4. Glacial acetic acid.
5. Milli-Q® (Millipore) water or equivalent.
6. Staining trays.

2.6 In-Gel Trypsin Digestion

1. PCR hood with built-in laminar flow and equipped with a HEPA filter and lightbox.
2. 0.1% (v/v) trifluoroacetic acid/50% methanol in water.
3. Methanol (Optima® 0.2 μm filtered).
4. Stainless steel razor blades.
5. Jewelers microforceps.
6. Destain solution: 50% (v/v) acetonitrile in 0.2 M ammonium bicarbonate.
7. Sequencing grade-modified trypsin (Promega).
8. Trypsin working solution: 0.02 μg/μL trypsin in 40 mM ammonium bicarbonate.
9. Trypsin wash buffer: 0.03% (v/v) formic acid in 40 mM ammonium bicarbonate.

10. Resuspension solution: 0.1% (v/v) formic acid (LC-MS grade, Sigma-Aldrich) in 5% (v/v) acetonitrile (HPLC grade, Thomas Scientific).
11. 0.5 mL Eppendorf tubes.
12. Autosampler sample tubes.
13. Vacuum filtration system
14. SpeedVac® centrifuge.
15. 37 °C thermostatically controlled incubator/shaker.

2.7 LC-MS/MS

1. Solvent A: Milli-Q water containing 0.1% (v/v) formic acid (LC-MS grade).
2. Solvent B: Acetonitrile (HPLC grade, Thomas Scientific) containing 0.1% (v/v) formic acid (LC-MS grade).
3. 180 μm i.d. × 2 cm UPLC Symmetry trap column packed with 5 μm C18 resin (Waters).
4. Reversed-phase 75 μm i.d. × 25 cm length analytical column packed with 1.7 μm C18 resin [e.g., nanoACQUITY UPLC BEH C18 Column (Waters)].
5. UPLC equipped with a chilled microvolume autosampler, 10 μL injection loop, and column heater maintained at 40 °C.
6. Q Exactive Plus or Q Exactive HF mass spectrometer (Thermo Fisher Scientific).

3 Methods

3.1 FPLC Affinity Depletion of Plasma

The following protocol describes FLPC immunoaffinity depletion using the IgY 14 and SuperMix columns connected in tandem producing a flow-through fraction which has passed through both columns. The columns are then disconnected and each column is eluted separately. The IgY 14 bound fraction is neutralized and frozen for storage at −20 °C, and in selected experiments, the SuperMix bound fraction is neutralized and analyzed in parallel with the flow-through fraction for direct comparison (*see* **Note 2**).

1. Prepare the three 1× mobile-phase buffers. Allow the three 10× buffers (dilution, stripping, and neutralization) to come to room temperature. If any precipitation is observed, mix gently to dissolve. Separately dilute each buffer tenfold with Milli-Q® water and filter the diluted buffers through 0.22 μm filter units. The final volumes needed per plasma depletion are dilution, ~100 mL; stripping, ~40 mL; and neutralization, ~25 mL. 1× buffers are stable for 1 week at r.t.
2. Before connecting the columns, purge the lines with the mobile-phase buffers and flush the system with 1× dilution buffer (*see* **Note 3**).

3. Connect and equilibrate the columns. For LC systems without a column control valve, connect the columns with a linear connection between the IgY 14 LC 5 and SuperMix LC2 columns (*see* **Note 4**). Connect the column inlet tubing to the top of the IgY 14 column first and then attach the detector inlet solvent line to the bottom of the SuperMix column. Equilibrate the columns with 1× dilution buffer for 20 min at 2 mL/min until a flat baseline is observed.
4. Prior to the first use of the columns, or if the columns have not been used for several days, two blank runs (injecting 500 μL of 1× dilution buffer) should be performed. If the columns are being used daily, a single full-length blank gradient (*see* Table 1) prior to the first plasma injection each day should be sufficient.
5. Dilute plasma (typically 100 μL; *see* **Note 5**) fivefold with 1× dilution buffer and filter with a prerinsed 0.45 μm microcentrifuge filter for 1 min. at 9000 × *g*. Keep the filtered sample on ice until ready for use (*see* **Note 6**).
6. Inject 500 μL of diluted and filtered plasma at a flow rate of 0.5 mL/min for 15 min followed by a wash with 1× dilution buffer at 1.5 mL/min for 10 min.
7. Collect and pool the flow-through fraction (*see* **Note 7**) as it comes off the columns, and keep the pooled sample on ice.
8. Add protease inhibitors to the pooled flow-through fraction (*see* **Note 8**). Final concentrations should be 1 μg/mL Leupeptin, 1 μg/mL Pepstatin-A, and 0.15 mM PMSF.
9. Disconnect and cap both ends of the SuperMix column. Then reconnect the detector inlet solvent line to the bottom of the IgY 14 column.
10. Elute bound proteins from the IgY 14 column with 1× stripping buffer at 2.0 mL/min for 13 min. Collect and pool the bound fractions.
11. Add a small volume (0.1× fraction volume) of 10× neutralization buffer to the pooled bound fraction to bring the pH to approximately 8.0 and store on ice. For long-term storage and future analysis, the IgY 14 bound fraction should be concentrated, frozen, and stored at −20 °C.
12. Neutralize the IgY 14 column with 1× neutralization buffer at 2.0 mL/min for 6 min.
13. Re-equilibrate the IgY 14 column with 1× dilution buffer at 2.0 mL/min for 7 min.
14. Disconnect and cap both ends of the IgY 14 column. Then reconnect the column inlet tubing to the top of the SuperMix column first and attach the detector inlet solvent line to the bottom of the SuperMix column.

15. Elute bound proteins from the SuperMix column with 1× stripping buffer at 1.5 mL/min for 8 min. Collect and pool the bound fractions.
16. Add a small volume (0.1× fraction volume) of 10× neutralization buffer to the pooled bound fraction and store on ice. For long-term storage, the SuperMix bound fraction should be concentrated, frozen, and stored at −80 °C for further analysis. *See* **Note 2** regarding analysis of the SuperMix bound fraction.
17. Neutralize the SuperMix column with 1× neutralization buffer at 1.5 mL/min for 7 min.
18. Re-equilibrate the column with 1× dilution buffer at 1.5 mL/min for 8 min.
19. Reconnect the linear connection between the IgY 14 and SuperMix columns and reattach the detector inlet solvent line to the bottom of the SuperMix column.
20. Inject the next sample and repeat **steps 5–19** until all samples are depleted.
21. Disconnect the columns and store at 2–8 °C in 1× dilution buffer containing 0.2% (v/v) sodium azide (*see* **Note 9**).
22. Replace the mobile-phase buffers with ultrapure water and flush the system, including detector, for 30 min at 1 mL/min followed by 20% (v/v) ethanol in water for 30 min at 1 mL/min (*see* **Note 10**).

3.2 Protein Concentration

Concentration of the IgY 14/SuperMix flow-through by ultrafiltration is preferred over ethanol or acetone precipitation because the very low protein concentration can result in significant losses during protein precipitation. If desired, the bound fractions may also be concentrated in parallel with the flow-through fraction. To prevent proteolysis, ultrafiltration should be performed in a refrigerated centrifuge (4 °C), and protease inhibitors should be added to samples prior to concentration (*see* **Note 8**).

1. Add 4 mL of the IgY 14/SuperMix flow-through fraction to a prerinsed Amicon Ultracel®-10K NMWL centrifugal filter units and centrifuge at 4000 × *g* for 15 min.
2. Remove the filtrate, add additional sample, and repeat step 1 until the entire sample is concentrated to <50 µL (*see* **Note 11**).
3. Measure the protein concentration with a BCA assay or equivalent (*see* **Note 12**).
4. Freeze the concentrated sample and store at −80 °C if sample is not immediately further analyzed.

3.3 Reduction/Alkylation Prior to SDS-PAGE

Reduction and alkylation of the depleted plasma samples prior to 1D SDS-PAGE will save time in the downstream in-gel digestion protocol. This protocol assumes that 100 µL of human plasma has

been depleted and the flow-through fraction has been concentrated to ~50 µL. The final sample volume, after adding reduction and alkylation reagents, will be ~80 µL.

1. Thaw depleted and concentrated flow-through fraction.
2. Add 8 µL of 10% SDS and 8 µL of 1 M Tris, pH 8.5. Adjust sample to a final concentration of 1% SDS and 100 mM Tris–HCl, pH 8.5.
3. Reduce samples with 1.6 µL of 1 M DTT (final concentration, 20 mM). Incubate for 1 h at 37 ° C with shaking.
4. Alkylate proteins by adding 9.6 µL of 0.5 M iodoacetamide/100 mM Tris–HCl, pH 8.5 (final concentration, 60 mM iodoacetamide). Incubate for 1 h at 37 °C in the dark, with shaking.
5. Quench the reaction by adding 4 µL of 1 M DTT (final concentration, 50 mM). Incubate for 15 min at 37 °C with shaking.

3.4 1D SDS-PAGE

The following protocol is for SDS-PAGE using pre-case NuPAGE® gels and buffers, followed by staining with Colloidal Coomassie® Blue. The samples are run for a short distance (0.5 cm; *see* **Note 13**) on the gel so the entire proteome can be digested and analyzed in a single fraction. To reduce keratin contamination, gloves should be worn at all times, and all gel tanks and staining trays should be washed with a 1% solution of mild detergent and rinsed thoroughly with Milli-Q water prior to running gels.

1. Add 5× protein-solubilizing buffer to reduced and alkylated plasma samples (final concentration, 1× protein solubilization buffer).
2. Mark a precast 1 mm, 10% Bis-Tris 10-well NuPAGE gel cassette at a point that is 0.5 cm from the bottom of the sample wells.
3. Prepare 1× MES running buffer: dilute 50 mL of 20× NuPAGE MES buffer with 950 mL of Milli-Q water.
4. Heat the plasma samples at 90 °C for 2 min.
5. Assemble the gel in the Mini Gel Tank (*see* **Note 14**) and add running buffer to the upper and lower chambers. Load a Benchmark MW standard in the first lane and plasma samples in the following lanes (*see* **Note 15**).
6. Run the gel at 200 V with constant voltage. Stop the electrophoresis when the dye front has reached the 0.5 cm marking.
7. Remove the gel from the electrophoresis unit and cassette and carefully mark the exact point of the dye front on the outer edges of the gel using a syringe filled with a small volume of India ink

3.5 Colloidal Coomassie® Blue Staining

1. Prepare fixing solution (100 mL per gel): 50% (v/v) methanol/10% (v/v) acetic acid in water. Place the gel in fixing solution for 10 min, with gentle agitation on a shaker.
2. Prepare staining solution (95 mL per gel): 20% (v/v) methanol/20% (v/v) Stainer A in water. Remove fixing solution and add staining solution to the staining tray. Shake the gel in the staining solution (without Stainer B) for 10 min.
3. Add the Colloidal Coomassie® Stainer B (5 mL per gel container) to the existing staining solution. Shake the gel in the staining solution between 3 and 12 h.
4. Remove staining solution and replace with 200 mL of ultrapure water to destain. The gel will have a clear background after destaining for several hours, but the water can be changed several times to accelerate the destaining process.

3.6 In-Gel Trypsin Digestion

To reduce airborne keratin contamination, we generally perform in-gel digestions in a PCR hood equipped with laminar flow and a HEPA filter that contains a small lightbox for visualization of gels. This protocol describes in-gel digestions performed in individual Eppendorf tubes.

1. Wash all Eppendorf and autosampler tubes twice with 0.1% (v/v) trifluoroacetic acid/50% (v/v) methanol in water, followed by two rinses with 100% methanol, and allow to air dry under an aluminum foil cover.
2. Slice each 0.5 cm gel lane lengthwise into six uniform ~1 × 5 mm slices using a stainless steel razor blade. Several adjacent lanes (2–3) of the same sample may be combined after digestion to maximize protein load.
3. Using microforceps, transfer the excised gel slices from one sample into two separate pre-cleaned 0.5 mL Eppendorf tubes, i.e., three gel slices per tube (*see* **Note 16**).
4. Destain gel slices for 30 min with 100 µL of 200 mM ammonium bicarbonate/50% acetonitrile at 37 °C, with gentle shaking. Remove buffer with vacuum aspiration and discard.
5. Dry gel slices for approximately 20–30 min using a SpeedVac® evaporator.
6. Rehydrate gel bands by adding 45 µL of trypsin working solution. Incubate 16–18 h at 37 °C in a thermostatically controlled incubator.
7. Transfer the digested protein extract into a clean 0.5 mL Eppendorf tube.
8. Add 20 µL of wash buffer to the digested gel slices and incubate for 30 min at 37 °C.
9. Transfer the second extract into the 0.5 mL Eppendorf tube containing extract 1.

10. Pool all extracts corresponding to the same sample tubes (e.g., combine two tubes for a single sample lane, combine four tubes for duplicate sample lanes, etc.) and snap-freeze samples in liquid nitrogen or a dry ice/ethanol bath.
11. Lyophilize pooled digest samples using a SpeedVac® evaporator.
12. Resuspend samples in a desired volume of resuspension solution, 0.1% (v/v) formic acid in 5% (v/v) acetonitrile, and vortex to dissolve (*see* **Note 17**).
13. Transfer samples into pre-cleaned autosampler sample vials.

3.7 LC-MS/MS

This protocol describes analysis of an entire proteome in a single fraction performed using an extended 4-h gradient on a Q Exactive Plus instrument.

1. Connect a 75 μm i.d. × 25 cm column packed with BEH C18 resin, 1.7 μm particle size, to a UPLC system that is interfaced with the mass spectrometer. To avoid temperature fluctuations, the column should be contained in a thermostatted compartment maintained at 40 °C or higher.
2. Set the flow rate to 200 nL/min and equilibrate the column with 3% Solvent B.
3. Inject 4–8 μL (*see* **Note 18**) of each trypsin digest at 5 μL/min over a 4 min period with 100% Solvent A. The HPLC gradient is as follows (*see* **Note 19**):

   5% to 70% Solvent B over 225 min.

   70% to 80% Solvent B over 5 min.

   Hold at 80% B for 5 min.

   Return to 5% B over 1 min.

   Hold at 5% B for 15 min.
4. The following parameters are used for MS/MS data acquisition and downstream label-free analysis with MaxQuant (*see* **Note 20**):
   (a) Scan ranges: 400–2000 *m/z*.
   (b) Full MS scan: 70,000 resolution in profile mode.
   (c) Full MS AGC target: $3e^6$.
   (d) Full MS maximum IT: 30 ms.
   (e) MS^2 scan: 17,500 resolution in centroid mode.
   (f) MS^2 AGC target: $1e^5$.
   (g) MS^2 maximum IT: 60 ms.
   (h) Isolation window: 1.5 *m/z*.
   (i) Number of MS/MS: 20 most abundant with data-dependent scans.

(j) Charge exclusion of $z = 1$ and unassigned.

(k) Peptide match: preferred.

(l) Dynamic exclusion: 30 s.

4 Notes

1. Studies have shown that, for proteome analyses, plasma is preferred over serum because the clotting processes involved in serum collection can be highly variable and activates proteases that can fragment other proteins. In addition, using EDTA as an anticoagulant is preferred over others such as heparin or citrate [4, 13]. Heparin acts as an anticoagulant through activation of antithrombin III, while citrate and EDTA inhibit coagulation and other enzymatic processes by chelate formation with metal-dependent enzymes. EDTA was shown to be more consistent and a better chelator of calcium in an in-depth analysis of specimen collection parameters for plasma proteome studies [4, 13, 14].
2. Previous studies report that the SuperMix column removes ~45 medium-abundance proteins with >99% depletion efficiency [10]. However, the SuperMix columns do not target a defined group of proteins. These columns are made by immunizing chickens with a complex mixture of antigens from the flow-through fractions of IgY 12 or 14 depleted plasma samples [8]. Therefore, the polyclonal IgY antibodies generated and used for the SuperMix columns are likely to vary substantially from batch to batch, both because they are polyclonal and because the immunogen is extremely complex. Therefore, the medium-abundance proteins that will be removed by such columns are likely to vary substantially in different studies analyzing SuperMix column depletion. Nonetheless, the potential batch-to-batch variation should not be a concern if a single column is used to deplete all samples in a given study. This is because the most important consideration when using abundant protein depletions in biomarker studies is reproducibility rather than on the specific medium abundant proteins that are removed. Consistent with the expected variability of different SuperMix columns, our systematic analysis of the IgY 14/SuperMix flow-through and the SuperMix bound fraction showed that only 39 out of 45 previously reported "targeted" moderate-abundance proteins captured by the SuperMix column [10] were removed in our study as summarized below. The majority of these proteins showed approximate depletion efficiencies of >90%, but 14 proteins had approximate depletion efficiencies ranging from ~45 to 88% (Fig. 2a). In this analysis, % depletion efficiency was calculated for each protein based on normalized MS intensities of

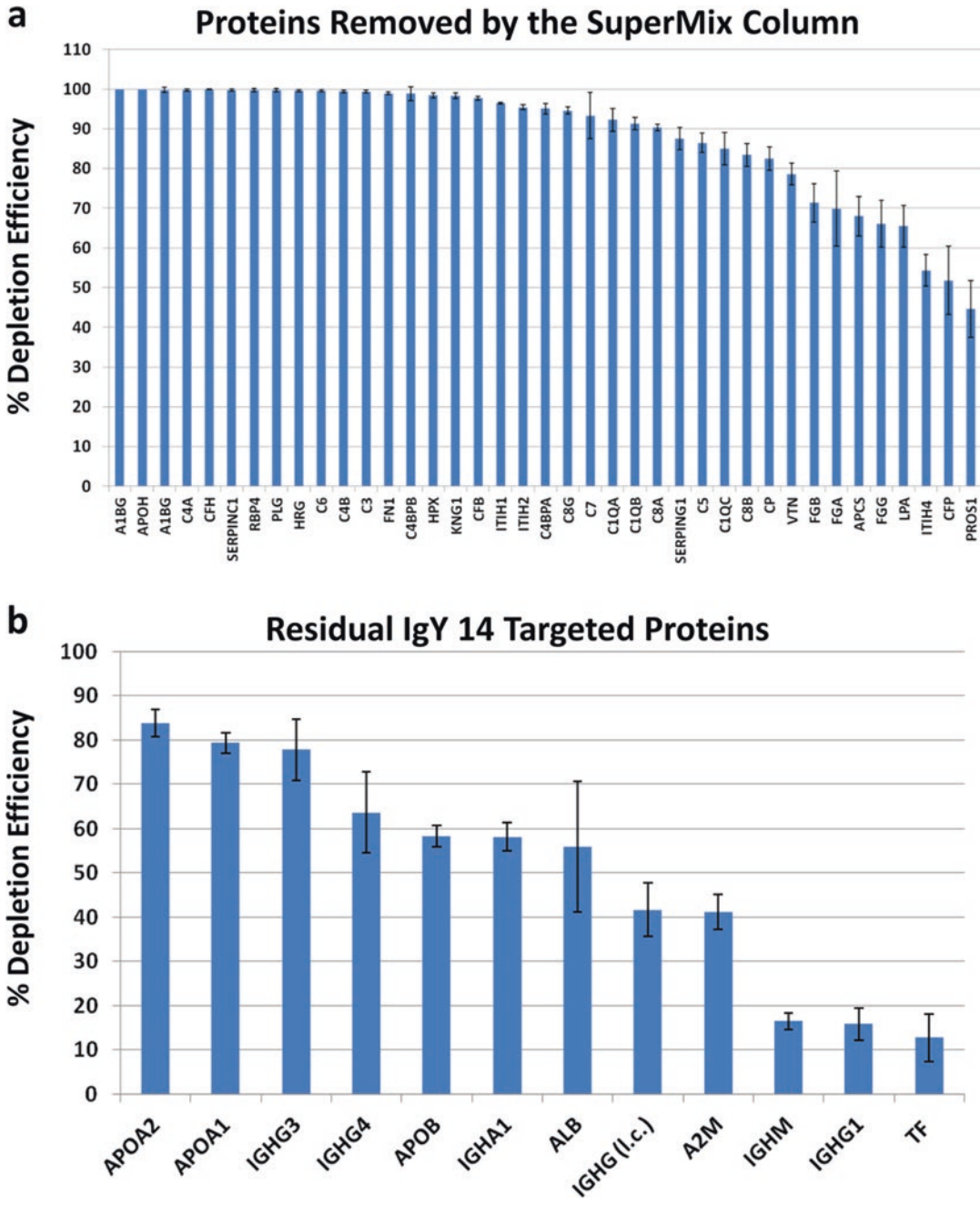


Fig. 2 (**a**) Approximate depletion efficiencies of medium-abundance proteins captured by the SuperMix column (*see* ref. 10). The IgY 14/SuperMix flow-through (FT) and SuperMix bound (B) fractions from triplicate depletions of three plasma pools ($n = 9$) were quantitated using MaxQuant. % depletion efficiencies (%DE) were calculated for each protein based on normalized MS intensities of the bound fraction and flow-through fractions, i.e., %DE = [B/(B + FT)] × 100. (**b**) Depletion efficiencies of the residual high-abundance proteins targeted by the IgY 14 column, remaining after IgY 14/SuperMix depletion. Averages and standard error bars are shown in both panels. *l.c.* light chain

proteins in the bound fraction and flow-through fractions, i.e., % depletion efficiency = [bound/(bound + flow-through)] × 100. However, these percentages will be approximate and are likely to underestimate depletion efficiency because the composition of the depleted fraction is much simpler, which will increase the depth of analysis for small amounts of targeted proteins that are not completely removed. We also found residual amounts of several of the IgY 14 targeted proteins remaining in the SuperMix bound and flow-through fractions (Fig. 2b). These proteins had approximate capture efficiencies ranging from ~12 to 85%, with the highest variability attributed to albumin and the targeted immunoglobulin isotypes. It should be noted that this analysis was performed using only the bound fraction from the SuperMix column, and it is expected that the majority of these 14 targets will be found in the IgY 14 bound fraction; therefore, overall removal of these proteins is expected to be far higher than the further removal produced by the SuperMix column.

We also identified an additional 65 proteins in the SuperMix bound fraction with approximate depletion efficiencies ranging from >10 to 100%. The 12 proteins with >80–100% depletion efficiencies were nonspecifically, but reproducibly, removed by the SuperMix column (Fig. 3). Nine of these proteins are medium-abundance proteins with reported concentrations ranging from 2 to 850 μg/mL [11]. The remaining

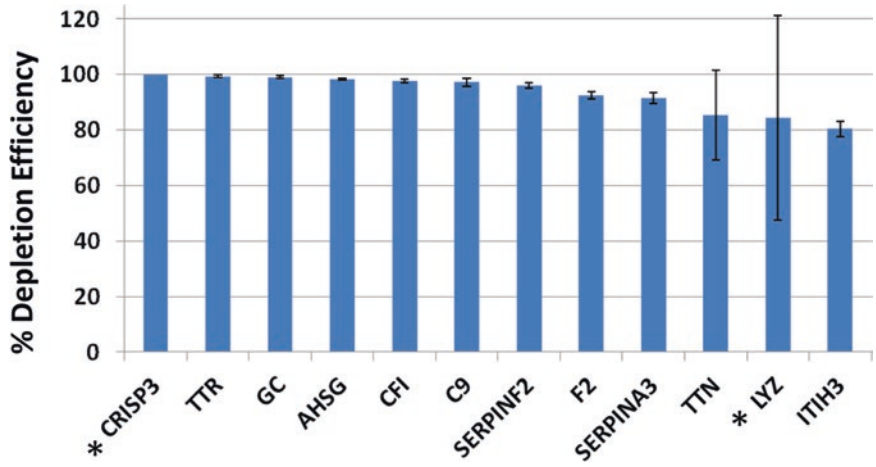


Fig. 3 Proteins removed by the SuperMix column by >80%. The IgY 14/SuperMix flow-through (FT) and SuperMix bound (B) fractions from triplicate depletions of three plasma pools ($n = 9$) were quantitated using MaxQuant. Approximate % depletion efficiencies (%DE) were calculated for each protein based on normalized MS intensities of the bound fraction and flow-through fractions, i.e., %DE = [B/(B + FT)] × 100. *Asterisk* indicates proteins with average normalized MS intensity $<3 \times 10^7$, which may be near the detection limit of the mass spectrometer

53 proteins had capture efficiencies ranging from 10 to 79%, including 12 low-abundance proteins having reported protein concentrations <100 ng/mL. These results are not uncommon because, as described above, the SuperMix antibodies do not target a defined group of proteins, but rather rely on a host's response to the complex antigen pool from the IgY 14 flow-through [8, 14]. Also, some proteins may be in the SuperMix bound fraction due to interactions with the high- and medium-abundance proteins that specifically bind to the immobilized antibodies. Because of these potential losses, it is recommended that the SuperMix column be considered as a fractionation scheme and both the flow-through and bound fractions should be analyzed if comprehensive proteome analysis is desired [10, 14, 15].

3. The IgY 14 and SuperMix columns have a maximum operation pressure of 350 psi, including the column pressure and the system back pressure. However, the antibody-modified resin can only withstand 100 psi. Because most HPLC systems require pressures greater that 100 psi to operate properly, it is preferable to use an FPLC-type system which can operate at low pressures. It is important to check the back pressure of the instrument first, before attaching the column, by running 1× dilution buffer through the system at 2 mL/min. If the system back pressure is more than 300 psi, the pressure can be reduced by using PEEK tubing with a larger inner diameter (i.d.).
4. There are several different configurations in which the IgY 14 and SuperMix columns can be used. Each column can be used as a single-stage depletion format, or a high-throughput tandem separation can be set up by using a six-port control valve to couple the two columns. This configuration will allow automated collection of the flow-through, IgY bound, and SuperMix bound in three separate fractions. Alternatively, as this protocol describes, for LC systems without a control valve, the columns can be connected in tandem to collect the flow-through fraction, and the columns can then manually be disconnected from each other and reconnected to the system individually to collect separate IgY 14 and SuperMix bound fractions. Finally, for blank runs (no sample injection), or in cases where it is not necessary to have separate IgY 14 and SuperMix bound fractions, the columns can be connected in tandem during the entire LC gradient resulting in single flow-through and elution fractions, respectively (Fig. 4a, Table 2).
5. The loading capacity of the IgY 14/SuperMix column is stated as being 90–110 µL of human plasma or serum. However, because total plasma concentrations and concentrations of specific targeted proteins can vary substantially from sample to sample, it is recommended to load slightly less than the maximum specified capacity; hence, we typically inject 80–100 µL of plasma to prevent overloading of the column.

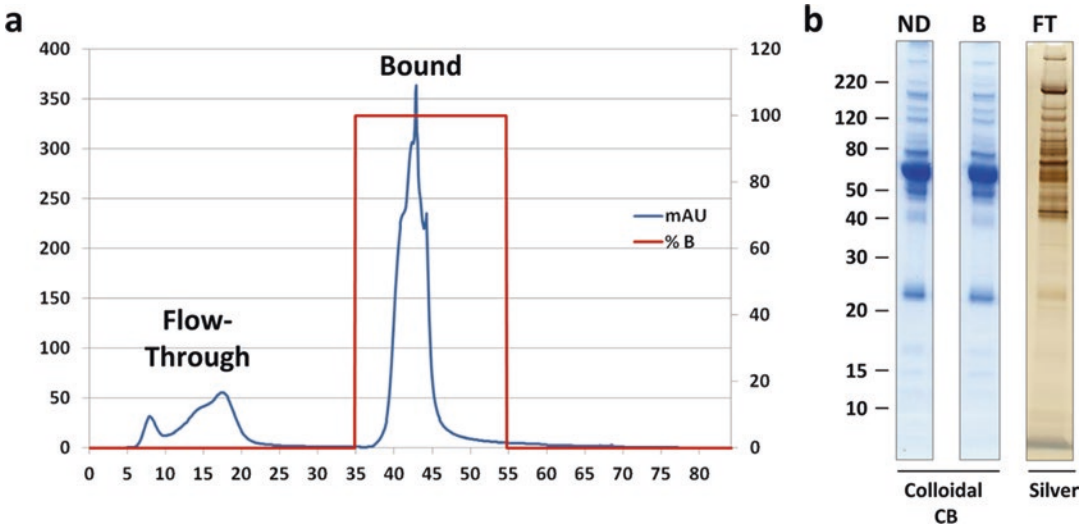


Fig. 4 (**a**) Typical chromatogram from IgY 14/SuperMix tandem column depletion with collection of a single-bound fraction. (**b**) Depletion of human plasma (100 μL) on tandem IgY 14/SuperMix columns. Proportional plasma loads of non-depleted human plasma (ND), pooled IgY 14/SuperMix bound fraction (B), and pooled IgY 14/SuperMix flow-through fraction (FT) were analyzed on Colloidal Coomassie® Blue- or silver-stained gels

Table 2
FPLC timetable for IGY 14/SuperMix tandem column depletion (single elution)

Cycle	Time (min)	1× dilution buffer	1× stripping buffer	1× neutralization buffer	Flow Rate (mL/min)	Max Pressure (MPa)
Sample Loading	0	100	0	0	0.5	2.5
Wash	6	100	0	0	0.5	2.5
Wash	24	100	0	0	1.5	2.5
Elution	34	0	100	0	1.5	2.5
Neutralization	54	0	0	100	1.5	2.5
Re-equilibration	64	100	0	0	1.5	2.5
End	84	0	0	0	0	2.5

6. A number of considerations that should be noted during plasma sample preparation are (a) the number of depletions and/or (b) different plasma samples that will be run per day. If the multiple aliquots of the same plasma sample are being depleted in a series of runs, it is best to thaw and filter the entire sample at once and store it on ice until all depletions are complete for optimal reproducibility. If multiple samples from different donors are being depleted, it is recommended that a

single sample is thawed and filtered shortly before its injection. To minimize proteolysis, thaw and prepare the next sample while the previous sample is running.

7. To monitor depletion efficiency, it is useful to reserve aliquots of both the individual fractions from the flow-through and bound fractions, as well as the pooled samples to be run on analytical gels. In general, proportional loads of non-depleted plasma and the IgY 14/SuperMix bound fraction can be analyzed using Colloidal Coomassie® Blue stain, while silver stain is needed to visualize the un-concentrated flow-through fraction (Fig. 4b).
8. To minimize proteolysis in the flow-through fraction, it is advised to remove the unbound fractions quickly from the fraction collector and store on ice immediately, rather than allowing them to sit at room temperature for extended periods of time. It is also beneficial to add protease inhibitors after pooling and prior to subsequent concentration steps.
9. Proper storage and monitoring of the antibody LC columns is critical to maintain optimal column performance. The manufacturer recommends flushing and storage of the column with 1× dilution buffer containing 0.01% sodium azide to prevent microbial growth in the column. Likewise, it is important to track the total number of blank and depletion cycles run on each column, either in single-stage format or connected in tandem. To evaluate column performance, Shi et al. assess the intensity ratios between the three peaks, as well as peak shapes and peak elution times [10]. It is preferable to have a large stock of aliquots of the same plasma sample stored at −80 °C for repeated quality control analyses because we have observed that peak shapes can vary slightly among different patient plasma samples; however, peak shapes from the same sample should be reproducible over the lifetime of the column. The columns are recommended for 100 uses each, but if column capacity has moderately declined, it is usually adequate to reduce plasma loads (i.e., 80 μL vs. 100 μL) to achieve a similar degree of depletion as was observed with new columns.
10. To prevent bacterial growth or corrosion due to halogen salt-containing buffers, the FPLC system should be thoroughly flushed with water, followed by 20% ethanol after the final run of each day.
11. Due to the high salt concentration in the 1× dilution buffer, the manufacturer recommends buffer exchange of the IgY 14/SuperMix flow-through to a volatile buffer such as ammonium bicarbonate if the samples will be lyophilized immediately following depletion. However, because we are running SDS-PAGE in the following step and we do not observe band distortion from high salts, we omit the buffer exchange step.

12. In our experience, the protein concentration of the unbound fraction from 100 µL IgY 14/SuperMix depleted plasma after concentration to a final volume of ~50 µL is typically around 1.5–2.0 µg/µL; however, plasma concentration ranges in patients can often be variable. Therefore, it is recommended to measure the protein concentration of each sample after ultrafiltration, but before adding reduction and alkylation buffers which can interfere with the protein assay.
13. Extensive fractionation is often necessary to achieve good depth of analysis in proteome studies. However, we have found that running samples on 1D SDS-PAGE gels for 0.5 cm can yield a single fraction proteome for higher-throughput analyses. When coupled with an extended high-resolution 4 h LC-MS/MS gradient on a high-speed mass spectrometer such as a Q Exactive Plus, we can identify ~4000 proteins in a cell lysate. However, the plasma proteome has a much wider dynamic range of concentrations even after IgY 14/SuperMix depletion; therefore, only about ~800–1000 plasma proteins can be identified in a single 4 h run.
14. When running more than one short distance (e.g., 1–4 cm) on the gel, we advise running the gels side by side in separate electrophoresis chambers because two gels in the same electrophoresis unit can migrate slightly differently. Hence, it may be difficult to achieve equal gel migration in a standard gel chamber (e.g., XCell *SureLock*™ Mini-Cell, Thermo Fisher Scientific). Alternatively, the Mini Gel Tank now available from Thermo Fisher Scientific can accommodate up to two gels in a convenient side-by-side format where migration between gels is more consistent.
15. Protein loads on gels should be maximized to increase the total sample digested and analyzed by LC-MS/MS. Ten-well NuPAGE® Bis-Tris gels allow up to 25 µL to be loaded per lane. Therefore, the entire concentrated IgY 14/SuperMix flow-through from one 100 µL depletion can be loaded in two or three lanes, and digests can be pooled and lyophilized after digestion.
16. It is not recommended to use more than three ~1 × 5 mm gel slices per tube. Increasing the number of gel slices will increase gel volume, and consequently more trypsin solution will be absorbed as the gels rehydrate during the overnight incubation. This effect will also increase the risk of trapping of peptides in the large gel volume.
17. Resuspension volume will depend on the concentration of the digested sample. This concentration is calculated based on the amount of protein loaded onto the gel (based on protein assay) and after assuming 50% sample loss following 1D SDS-PAGE and

trypsin digestion. We typically inject 1.0–2.0 μg of trypsin digests onto the mass spectrometer. Therefore, a volume that will yield ~1 μg in 4 μL of buffer would be practical because up to 2.0 μg of the same sample can be re-injected if necessary, without exceeding the maximum volume of the autosampler loop.

18. As described above (*see* **Note 17**), we typically inject 1.0–2.0 μg of trypsin digest onto the mass spectrometer. This protein amount is calculated based on amount of protein loaded onto the gel (based on protein assay) and after assuming 50% sample loss after 1D SDS-PAGE and trypsin digestion.
19. The sample run time is 245 min. To help minimize carryover between samples, a "rapid" blank should be run between samples. A reasonable "rapid" blank cycle time is ~30 min, giving a total analysis time per sample of 275 min.
20. For a detailed protocol describing label-free quantitation using MaxQuant, please reference Chapter 23 in this book [16].

Acknowledgment

This work was supported by NIH Grants RO1HD076279, RO1CA131582, and WW Smith Charitable Trust Grants H1205 and H1305 (D.W. Speicher), PA Department of Health Commonwealth Universal Research Enhancement (CURE) Program Grant (B. Ky), as well as CA10815 (NCI core grant to the Wistar Institute).

References

1. Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD (2003) Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics 2(10):1096–1103. doi:10.1074/mcp.M300031-MCP200. M300031-MCP200 [pii]
2. Jacobs JM, Adkins JN, Qian WJ, Liu T, Shen Y, Camp DG 2nd, Smith RD (2005) Utilizing human blood plasma for proteomic biomarker discovery. J Proteome Res 4(4):1073–1085. doi:10.1021/pr0500657
3. Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1(11):845–867
4. Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, Apweiler R, Haab BB, Simpson RJ, Eddes JS, Kapp EA, Moritz RL, Chan DW, Rai AJ, Admon A, Aebersold R, Eng J, Hancock WS, Hefta SA, Meyer H, Paik YK, Yoo JS, Ping P, Pounds J, Adkins J, Qian X, Wang R, Wasinger V, CY W, Zhao X, Zeng R, Archakov A, Tsugita A, Beer I, Pandey A, Pisano M, Andrews P, Tammen H, Speicher DW, Hanash SM (2005) Overview of the HUPO plasma proteome project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics 5(13):3226–3245. doi:10.1002/pmic.200500358
5. Pieper R, Su Q, Gatlin CL, Huang ST, Anderson NL, Steiner S (2003) Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. Proteomics 3(4):422–432. doi:10.1002/pmic.200390057
6. Echan LA, Tang HY, Ali-Khan N, Lee K, Speicher DW (2005) Depletion of multiple high-abundance proteins improves protein profiling capacities of human serum and

plasma. Proteomics 5(13):3292–3303. doi:10.1002/pmic.200401228

7. Hoffman SA, Joo WA, Echan LA, Speicher DW (2007) Higher dimensional (Hi-D) separation strategies dramatically improve the potential for cancer biomarker detection in serum and plasma. J Chromatogr B Analyt Technol Biomed Life Sci 849(1–2):43–52. doi:10.1016/j.jchromb.2006.10.069. S1570-0232(06)00885-3 [pii]
8. Qian WJ, Kaleta DT, Petritis BO, Jiang H, Liu T, Zhang X, Mottaz HM, Varnum SM, Camp DG 2nd, Huang L, Fang X, Zhang WW, Smith RD (2008) Enhanced detection of low abundance human plasma proteins using a tandem IgY12-SuperMix immunoaffinity separation strategy. Mol Cell Proteomics 7(10):1963–1973. doi:10.1074/mcp.M800008-MCP200. M800008-MCP200 [pii]
9. Lin B, White JT, Wu J, Lele S, Old LJ, Hood L, Odunsi K (2009) Deep depletion of abundant serum proteins reveals low-abundant proteins as potential biomarkers for human ovarian cancer. Proteomics Clin Appl 3(7):853–861. doi:10.1002/prca.200800141
10. Shi T, Zhou JY, Gritsenko MA, Hossain M, Camp DG, 2nd, Smith RD, Qian WJ (2012) IgY14 and SuperMix immunoaffinity separations coupled with liquid chromatography-mass spectrometry for human plasma proteomics biomarker discovery. Methods 56 (2):246–253. doi:10.1016/j.ymeth.2011.09.001. S1046-2023(11)00165-4 [pii]
11. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, Sharma R, Ahmad Khan A, Balakrishnan L, Sahasrabuddhe NA, Kumar S, Jhaveri BN, Sheth KV, Kumar Khatana R, Shaw PG, Srikanth SM, Mathur PP, Shankar S, Nagaraja D, Christopher R, Mathivanan S, Raju R, Sirdeshmukh R, Chatterjee A, Simpson RJ, Harsha HC, Pandey A, Prasad TS (2014) Plasma proteome database as a resource for proteomics research: 2014 update. Nucleic Acids Res 42(Database issue):D959–D965. doi:10.1093/nar/gkt1251. gkt1251 [pii]
12. Gundry RL, Fu Q, Jelinek CA, Van Eyk JE, Cotter RJ (2007) Investigation of an albumin-enriched fraction of human serum and its albuminome. Proteomics Clin Appl 1(1):73–88. doi:10.1002/prca.200600276
13. Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, Schuchard MD, Mehigh RJ, Cockrill SL, Scott GB, Tammen H, Schulz-Knappe P, Speicher DW, Vitzthum F, Haab BB, Siest G, Chan DW (2005) HUPO plasma proteome project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics 5(13):3262–3277. doi:10.1002/pmic.200401245
14. Juhasz P, Lynch M, Sethuraman M, Campbell J, Hines W, Paniagua M, Song L, Kulkarni M, Adourian A, Guo Y, Li X, Martin S, Gordon N (2011) Semi-targeted plasma proteomics discovery workflow utilizing two-stage protein depletion and off-line LC-MALDI MS/MS. J Proteome Res 10(1):34–45. doi:10.1021/pr100659e
15. Patel BB, Barrero CA, Braverman A, Kim PD, Jones KA, Chen DE, Bowler RP, Merali S, Kelsen SG, Yeung AT (2012) Assessment of two immunodepletion methods: off-target effects and variations in immunodepletion efficiency may confound plasma proteomics. J Proteome Res 11(12):5947–5958. doi:10.1021/pr300686k.
16. Beer LA, Liu P, Ky B, Barnhart KT, Speicher DW (2017) Efficient Quantitative Comparisons of Plasma Proteomes Using Label-Free Analysis with MaxQuant. Methods Mol Biol 1619.

Chapter 8

Low-Molecular-Weight Plasma Proteome Analysis Using Top-Down Mass Spectrometry

Dong Huey Cheon, Eun Gyeong Yang, Cheolju Lee, and Ji Eun Lee

Abstract

While human plasma has a wealth of diagnostic information regarding the state of the human body in heath and disease, low molecular weight (LMW) proteome (<30 kDa) has been shown to contain a rich source of diagnostic biomarkers. Here we describe a protocol for top-down proteomic analysis to identify and characterize the LMW proteoforms present in four types of human plasma samples without immunoaffinity depletion and with depletion of the top two, six, and seven high-abundance proteins. Each type of plasma sample was first fractionated based on molecular weight using gel-eluted liquid fraction entrapment electrophoresis (GELFrEE). Then, the GELFrEE fractions containing up to 30 kDa were subjected to nanocapillary–LC–MS/MS, and the high-resolution MS and MS/MS data were processed using ProSightPC software. As a result, a total of 442 LMW proteins and cleaved products, including those with posttranslational modifications (PTMs) and single amino acid variations (SAAVs), were identified with a threshold E-value of 1×10^{-4} from the four types of plasma samples.

Key words Human plasma proteome, Low molecular weight, Top-down mass spectrometry, Gel-eluted liquid fraction entrapment electrophoresis (GELFrEE), Posttranslational modification (PTM), Single amino acid variation (SAAV), Proteoforms

1 Introduction

While human plasma serves as an invaluable source for disease diagnosis, low molecular weight (LMW) plasma proteome (<30 kDa), which is composed of either small proteins such as hormones, cytokines, and growth factors or peptides derived from the proteolytic degradation of larger proteins, has attracted attention in the field of biomarker discovery [1–3]. The possibility of the LMW components as diagnostic biomarkers was initially observed from peak profiling experiments using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) or surface-enhanced laser desorption/ionization (SELDI)-TOF MS [4, 5]. Although the profiling platform based on MALDI– or SELDI–TOF MS has been combined with tandem

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_8,

mass spectrometry for identification of the LMW species [6, 7], it usually focuses on identifying the LMW components showing the differential changes of peak profiles between the control and disease states, which does not result in identifying a great number of LMW components present in the plasma or serum samples.

The enrichment strategies such as centrifugal ultrafiltration followed by bottom-up mass spectrometric analyses have shown to effectively identify a large number of LMW proteins from the plasma or serum samples [8–10]. Hundreds of proteins belonging to LMW plasma or serum proteome were successfully identified from the bottom-up proteomic analyses of the LMW fraction. However, there is a limit in discovering LMW proteins that undergo posttranslational modifications (PTMs) and endogenous proteolytic cleavages associated with disease states because the enzymatic digestion used for bottom-up analysis eliminates the information of intact proteins that naturally occur in plasma or serum.

Top-down mass spectrometric analysis in which intact proteins are directly ionized and fragmented in a mass spectrometer enables a full characterization of the primary structure of a protein and therefore can differentiate diverse protein isoforms called proteoforms, arising from genetic variations, alternative splicing, endogenous proteolysis, and PTMs [11]. While there have been great advances in top-down proteomics for analyzing complex protein mixtures [12, 13], the technologies have not yet been widely applied to clinical samples such as plasma or serum. Here, we report a protocol of top-down mass spectrometric analysis of LMW proteome (<30 kDa) present in four types of human plasma samples without immunoaffinity depletion and with depletion of the top two, six, and seven high-abundance proteins based on our recent work [14]. Prior to top-down MS, the four types of plasma samples were fractionated using continuous tube gel electrophoresis, known as gel-eluted liquid fraction entrapment electrophoresis (GELFrEE), in which proteins are constantly eluted from a sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) tube gel column based on molecular weight (MW) and collected in liquid form [15, 16]. Then, the GELFrEE fractions containing up to 30 kDa were subjected to nanocapillary–LC–MS/MS, and the high-resolution MS and MS/MS data were processed using ProSightPC software, resulting in identification of 442 LMW proteoforms with molecular weight ranges of 1.2–28 kDa.

2 Materials

All solutions were prepared using HPLC grade water.

2.1 Plasma Preparation

1. Protease and phosphatase inhibitor cocktail tablets.
2. Bicinchoninic acid (BCA) assay kit.

2.2 Depletion of High-Abundance Proteins from Plasma Samples

1. ProteoExtract™ Albumin/IgG removal kit (CALBIOCHEM): albumin removal column, immunoglobulin G (IgG) removal column, 10× binding buffer (250 mM sodium phosphate, pH 7.4), elution buffer for albumin (25 mM sodium phosphate, pH 8.0, 2 M NaCl), and elution buffer for IgG (250 mM citric acid).
2. MARS-6 and MARS-7 columns (4.6 × 50 mm, Agilent Technologies).
3. Buffer A (salt-containing neutral buffer, pH 7.4, Agilent Technologies) and Buffer B (urea buffer, pH 2.2, Agilent Technologies) used for the removal of the top six and seven high-abundance plasma proteins.
4. HPLC system: Agilent 1100 series (Agilent Technologies).
5. Amicon Ultracel-3 centrifugal filter: 3 kDa cutoff.
6. Reduction buffer (1 M dithiothreitol (DTT) stock solution): 0.1542 g of DTT was dissolved in 1 mL of water.
7. Alkylation buffer (0.5 M iodoacetamide (IAA) stock solution): 0.0925 g of IAA was dissolved in 1 mL of water.
8. 0.22 μm polyvinylidene fluoride (PVDF) centrifugal filter.

2.3 GELFrEE Fractionation

1. Glass tube: 6 mm o.d. × 6.0 cm.
2. Resolving gel buffer (1.5 M Tris–HCl, pH 8.8): 181.71 g of Tris base was dissolved in 900 mL of water and 6 M HCl was added to adjust pH. Then, water was added to make a final volume of 1 L.
3. Stacking gel buffer (0.5 M Tris–HCl, pH 6.8): 60.57 g of Tris base was dissolved in 900 mL of water and 6 M HCl was added to adjust pH. Then, water was added to make a final volume of 1 L.
4. 5× sample buffer: 1 g of sodium dodecyl sulfate (SDS) and 0.05 g of bromophenol blue were added to 5 mL of 0.5 M Tris–HCl (pH 6.8), and then 5 mL of glycerol was finally added to the solution.
5. GELFrEE running buffer: 0.025 M Tris, 0.192 M glycine, and 0.1% SDS.
6. 30% acrylamide/Bis solution.
7. 10% ammonium persulfate (APS): 1 g of APS was dissolved in 10 mL of water.
8. Tetramethylethylenediamine (TEMED).
9. Eight-channel multiplexed device consisting of a cathode chamber, eight gel columns, a collection chamber for each gel column, and an anode chamber (*see* **Note 1**).

2.4 Liquid Chromatography–Mass Spectrometry

1. NanoLC 2D system (Eksigent Technologies).
2. Mobile phase A: 0.2% formic acid and 99.8% water.
3. Mobile phase B: 0.2% formic acid and 99.8% acetonitrile.
4. Orbitrap XL mass spectrometer (Thermo Fisher Scientific).
5. PLRP-S 1000 Å 5 μm resin (Agilent Technologies).
6. PicoTip emitters (New Objective).

2.5 MS Data Analysis

1. ProSightPC 3.0: search engine for protein identification and characterization (Thermo Fisher Scientific).

3 Methods

The workflow for top-down mass spectrometric analysis of the LMW proteome present in human plasma samples is seen in Fig. 1. The current protocol includes plasma sample preparation without immunoaffinity depletion and with depletion of the top two, six, and seven high-abundance proteins, molecular weight-based separation, nanocapillary–LC–MS/MS, and high-resolution MS and MS/MS data processing for identification of proteoforms.

3.1 Human Plasma Sample Preparation

1. Protease and phosphatase inhibitor cocktails were added to individual plasma sample (*see* **Note 2**).
2. Protein concentration of the plasma sample was determined using the bicinchoninic acid (BCA) method according to the manufacturer's instructions (*see* **Note 3**).

3.2 Depletion of High-Abundance Proteins

For top-down proteomic analysis of plasma samples with the removal of the top two high-abundance proteins (albumin and IgG), the top six high-abundance proteins (albumin, IgG, immunoglobulin A (IgA), serotransferrin (TRFE), haptoglobin (HPT), and alpha-1 antitrypsin (A1AT), and the top seven high-abundance proteins (albumin, IgG, IgA, TRFE, HPT, A1AT, and fibrinogen) high-abundance proteins, the human plasma sample was depleted of its high-abundance proteins using three different immunoaffinity columns.

3.2.1 Depletion of Top Two High-Abundance Proteins

1. 180 μL of plasma was diluted with 180 μL of 10× binding buffer, which was provided by the manufacturer, and 1440 μL of water.
2. After a syringe was filled with 6 mL of 1× binding buffer, which was diluted from 10× binding buffer using water, without introducing air bubble, it was connected to an albumin removal column. Then, gentle pressure was applied so that 1× binding buffer was passed through each column for column equilibration at a flow rate of 0.25 mL/min. The process was repeated for the other albumin removal column and an IgG removal column. Two albumin and one IgG removal columns were connected together prior to sample loading. The flow-through fraction was discarded.

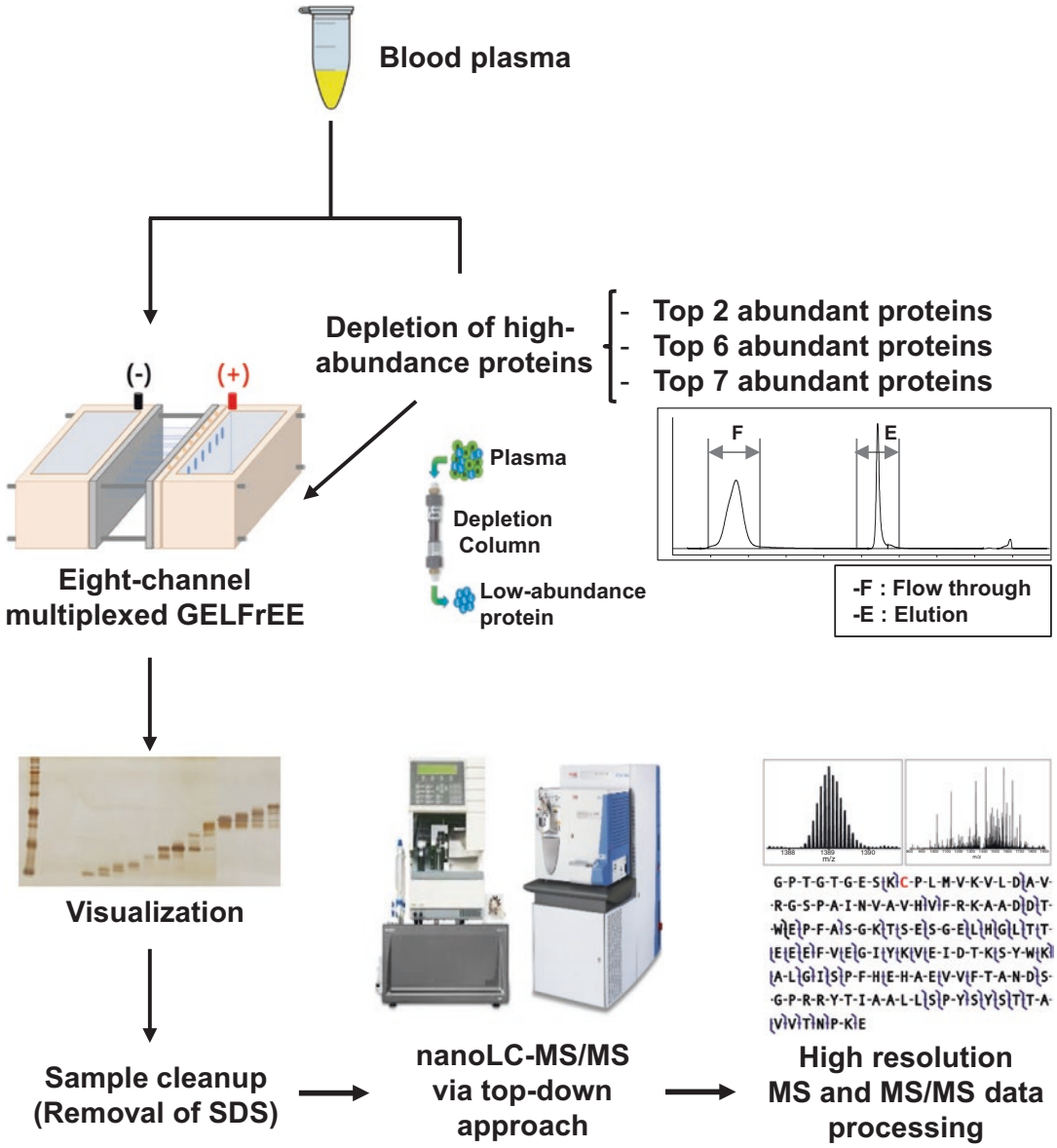


Fig. 1 Workflow for top-down mass spectrometric analysis of four types of human plasma samples. Four types of human plasma without depletion of high-abundance proteins and with depletion of the top two, six, and seven abundant proteins were fractionated using eight-channel multiplexed GELFrEE. The GELFrEE fractions containing up to 30 kDa were subjected to nanoLC-MS/MS, and the resulting high resolution MS and MS/MS data are processed using ProSightPC software tailored for top-down analysis. Reproduced from [14] with permission from the publisher

3. A new syringe was filled with the diluted sample and connected to the removal columns. Then, gentle pressure was applied so that the diluted sample was slowly loaded to the removal columns at a flow rate of 0.1 mL/min. The flow-through fraction was collected.
4. After a syringe filled with 6 mL of 1× binding buffer was connected to the removal columns, gentle pressure was applied so

that the binding buffer was passed through the removal columns at a flow rate of 0.25 mL/min. The flow-through fraction was also collected and combined with the flow-through fraction that was previously collected.

5. The albumin and IgG removal columns were disconnected prior to elution of bound proteins. After a new syringe filled with 6 mL of albumin elution buffer was connected to the albumin removal columns, gentle pressure was applied so that the albumin elution buffer was passed through them at a flow rate of 0.25 mL/min. When the bound proteins need to be analyzed, the eluted fraction can be collected. This process was also repeated for the IgG removal column using 3 mL of IgG elution buffer. Then, the individual removal columns were equilibrated with 2 mL of 1× binding buffer and stored at 4 °C.

3.2.2 Depletion of Top Six and Seven High-Abundance Proteins

1. Buffers A and B, which were provided by the manufacturer, were filtered using 0.45 μm regenerated cellulose membrane and then degassed for 10 min.
2. 20 μL of plasma was diluted with 80 μL of buffer A and then filtered using 0.22 μm polyvinylidene fluoride (PVDF) centrifugal filter at 16,000 × *g* for 1 min.
3. MARS-6 or MARS-7 column was connected to Agilent 1100 HPLC system and equilibrated with buffer A for 20 min at a flow rate of 1 mL/min.
4. 100 μL of the diluted plasma sample was injected onto a MARS-6 or MARS-7 column at a flow rate of 0.25 mL/min and then separated using the following gradient conditions: 0 min 100% buffer A (0.25 mL/min), 0–9 min 100% buffer A (0.25 mL/min), 9–9.01 min 100% buffer B (1 mL/min), 9.01–12.5 min 100% buffer B (1 mL/min), 12.5–12.6 min 100% buffer A (1 mL/min), and 12.6–20 min 100% buffer A (1 mL/min).
5. The flow-through fraction, which was not bound to the MARS-6 or MARS-7 column, was collected (*see* **Note 4**).
6. Another 100 μL of the diluted plasma sample was injected again onto a MARS-6 or MARS-7 column and repeated for the removal of the top six or seven high-abundance proteins.
7. When the depletion process was done, MARS-6 or MARS-7 column was equilibrated with buffer A for 7.4 min at a flow rate of 1 mL/min and kept at 4 °C.

3.2.3 Sample Enrichment

1. The flow-through fractions obtained from the depletion of top two, six, and seven high-abundance proteins were diluted with threefold with 10 mM Tris–HCl (pH 7.5) and concentrated using Amicon Ultracel-3 centrifugal filters (3 kDa cutoff). Prior to concentration of each depleted plasma sample, a centrifugal filter device was first rinsed two times with 450 μL of

water 14,000 × *g* for 10 min and conditioned two times with 450 µL of 10 mM Tris–HCl (pH 7.5) at 14,000 × *g* for 10 min. The concentrated samples having approximately 100 µL were recovered by inverting the filter device and centrifuged at 2000 × *g* for 2 min.

2. The protein concentration of the concentrated samples was determined using the BCA assay kit.

3.3 GELFrEE Fractionation of Four Types of Plasma Samples

3.3.1 GELFrEE Fractionation

1. One end of a glass tube was tightly covered with Parafilm (2 cm × 2 cm).
2. In order to cast 17.5% T for the resolving gel, 1.7 mL of water, 2.5 mL of 1.5 M Tris–HCl (pH 8.8), and 5.8 mL of 30% acrylamide/Bis (37.5:1) solution were first mixed in a 15 mL falcon tube, and then 50 µL of 10% APS solution and 5 µL of TEMED were finally added. Then, 849 µL of the resolving gel buffer was slowly added to the glass tube in order to make a 3 cm length of resolving gel, and 100 µL of 2-methyl-2-butanol was immediately added to the top layer of the resolving gel.
3. After the resolving gel was polymerized, 1.5 cm-long stacking gel was cast to 4% T. 6.1 mL of water, 2.5 mL of 0.5 M Tris–HCl (pH 6.8), and 1.3 mL of 30% acrylamide/Bis solution (37.5:1) were first mixed in a 15 mL falcon tube, and then 50 µL of 10% APS solution and 10 µL of TEMED were finally added. After the residual 2-methyl-2-butanol over the resolving gel was removed by inverting the glass tube on clean absorbent paper, 300 µL of the stacking gel buffer was slowly added to the top of the resolving gel, and 100 µL of 2-methyl-2-butanol was immediately added to the top layer of the stacking gel. After the stacking gel was polymerized, the residual 2-methyl-2-butanol was also removed as described above (*see* **Note 5**).
4. 350 µg of the depleted plasma sample was diluted to 77 µL using Tris–HCl (pH 7.5). Then, 0.5 µL of 1 M DTT solution was added to the depleted plasma sample followed by incubation for 35 min at 56 °C to reduce cysteine residues. Then, 2.5 µL of 500 mM IAA solution was added to the depleted plasma sample followed by incubation for 30 min at room temperature in the dark for alkylation (*see* **Note 6**).
5. 80 µL of the plasma sample mixed with 5× sample buffer and heated for 10 min at 95 °C. Then, 100 µL of each sample was loaded onto a SDS-polyacrylamide tube gel column (*see* **Note 7**).
6. The cathode and anode chambers of eight-channel multiplexed GELFrEE device were filled with fresh running buffer. 150 µL of running buffer was also added to each collection chamber of the GELFrEE device. Then, the eight-channel multiplexed GELFrEE device was operated with a constant application of 240 V in a stop and go cycle, collecting fractions from each gel columns at defined time points by transferring the solution in

each collection chamber to a siliconized microcentrifuge tube (*see* **Note 8**). After collecting the fractions at each time point, 150 μL of fresh running buffer was introduced into each collection chamber, and the power supply was resumed to continue separation. 16 to 18 GELFrEE fractions containing up to 30 kDa were collected for each type of plasma sample.

7. Each fractionation was visualized by silver staining of an SDS-PAGE slab gel with 8 μL of each 150 μL GELFrEE fraction (Fig. 2).

3.3.2 Sample Processing for Top-Down Mass Spectrometric Analysis

1. The GELFrEE fractions with similar molecular weight ranges collected from eight to 24 channel replicates of GELFrEE were typically combined and concentrated using an Amicon Ultracel-3 centrifugal filter (*see* **Note 9**). Prior to concentra-

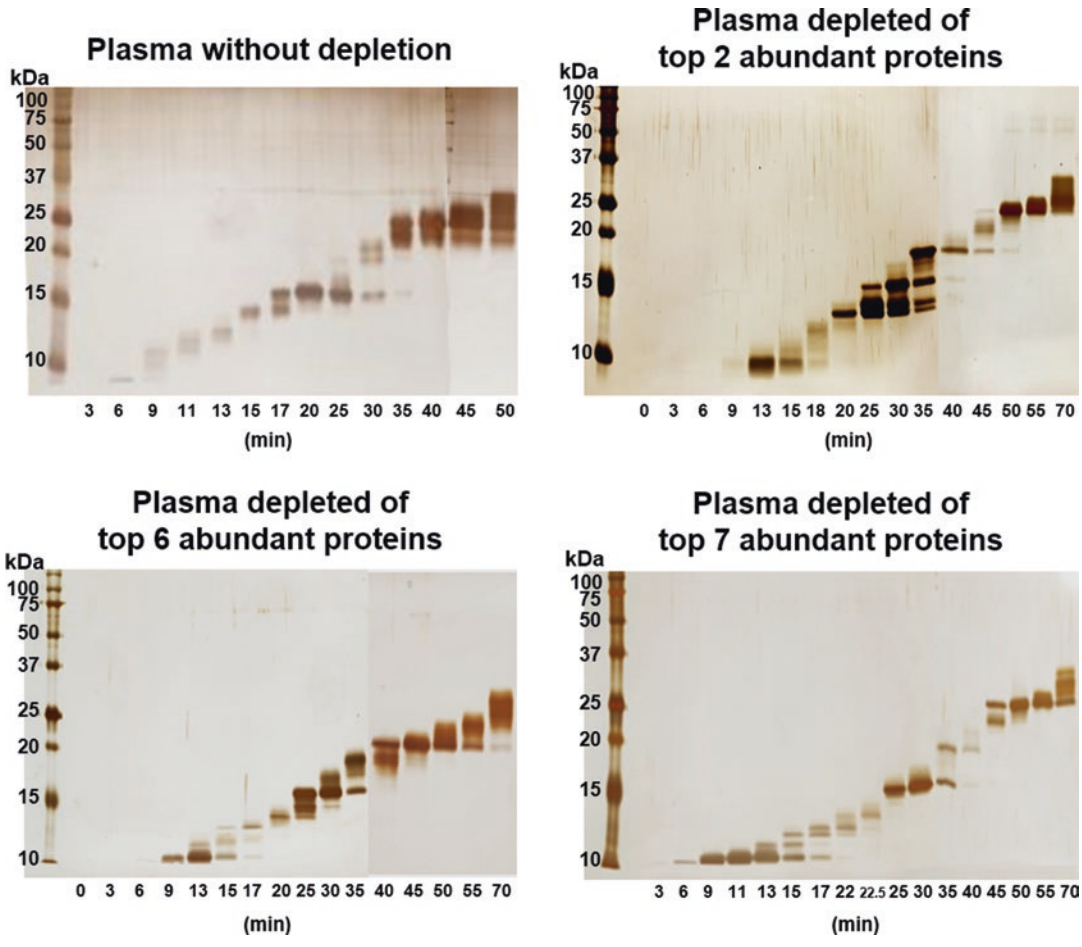


Fig. 2 Slab gel visualizations of GELFrEE fractionation for LMW proteome (<30 kDa) present in four types of human plasma samples. Plasma samples without depletion of high-abundance proteins and with depletion of the top two, six, and seven abundant proteins were fractionated using 17.5% tris-glycine GELFrEE. Reproduced from [14] with permission from the publisher

tion of the fractionated samples, the centrifugal filter devices were rinsed two times with 450 μL of water at 14,000 × *g* for 10 min and conditioned two times with 450 μL of 10 mM Tris–HCl (pH 7.5) at 14,000 × *g* for 10 min. GELFrEE fractions were diluted threefold with 10 mM Tris–HCl (pH 7.5) and centrifuged at 14,000 × *g* for 10 min. The concentrated solute having approximately 100 μL was recovered by inverting the filter device and centrifuged at 2000 × *g* for 2 min.

2. For the removal of SDS from the fractionated samples, the concentrated samples were precipitated using chloroform/methanol/water precipitation [17]. 400 μL of methanol was added to each sample and vortexed vigorously for 1 min. 100 μL of chloroform was then added and vortexed vigorously again. 300 μL of water was added and vortexed again. The samples were centrifuged at 16,000 × *g* for 20 min. After centrifugation, the top aqueous/methanol layer was carefully pipetted off and discarded, while the protein pellet over chloroform layer was not touched. 400 μL of methanol was added to the protein pellet over chloroform layer and gently mixed with the protein pellet. Then, the samples were centrifuged again at 16,000 × *g* for 20 min. The supernatant was carefully removed, while the protein pellet was not disturbed. Then, 400 μL of methanol was added again to the protein pellet and gently mixed with the protein pellet. Then, the samples were centrifuged again. After the supernatant was removed, residual solvent was allowed to dry in a fume hood. Then, the protein pellets were resuspended with 20 μL of solution consisting of solution 0.2% formic acid, 94.8% water, and 5% acetonitrile.

3.4 Liquid Chromatography–Mass Spectrometry

1. Trap (150 μm i.d. × 3 cm) and analytical (75 μm i.d. × 10 cm) columns were packed with PLRP-S media (1000 Å, 5 μm).
2. Typically, 5–10 μL of sample was injected onto a trap column using an autosampler and separated on an analytical column (1000 Å, 5 μm) with 350 nL/min. The typical gradient conditions: 0 min 95% buffer A (100% water with 0.2% formic acid) and 5% buffer B (100% acetonitrile with 0.2% formic acid), 0–5 min 5–20% B, 5–10 min 20–21% B, 10–55 min 21–30% B, 55–70 min 30–40% B, 70–78 min 40–52% B, 78–83 min 52–85% B, 83–88 min 85–5% B, and 88–100 min at 5% B.
3. Data were collected on an LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA) using the Orbitrap mass analyzer with AGC targets of 1×10^6 for MS (4–16 microscans, 60,000 or 100,000 resolving power at m/z 400, typical scan range of 800–1600 m/z) and MS/MS (4–16 microscans, 60,000 or 100,000 resolving power at m/z 400).

4. The spray voltage was set to 2.5 kV, and the temperature of the heated capillary was held at 250 °C.
5. Fragmentation was achieved using data-dependent collision-induced dissociation (CID) or source-induced dissociation (SID). In most cases, GELFrEE fractions with molecular ranges up to 13 kDa were subjected to CID fragmentation, and the ones with molecular ranges from 13 to 30 kDa were subjected to SID fragmentation. CID was pursued with a 15 or 25 m/z isolation window for either the most and fourth intense ions or the third and sixth intense ions from the previous full MS scans to decrease the chances of fragmenting the different charge states originated from the same proteins. MS/MS settings for the CID were as follows, minimum signal threshold = 1000 counts, normalized collision energy = 41%, activation Q = 0.4, and activation time = 100 ms. Dynamic exclusion was enabled with a repeat count of 1, an exclusion duration of 480 s, and a repeat duration of 120 s. SID utilized a potential of 75 V, and data were collected with a scan range of 400–1800 m/z.

3.5 Top-Down MS Data Analysis

1. Each LC-MS/MS file was analyzed using ProSightPC 3.0 program.
2. Intact precursor and fragment masses from LC-MS/MS files were determined using Xtract algorithm within ProSightHT of ProSightPC software to determine monoisotopic neutral masses from high-resolution precursor and fragment ion spectra and compiled into a ProSight upload file (.puf). From precursor selection criterion within ProSightHT, multiplexing mode was also selected (*see* **Note 10**).
3. Each .puf file was searched in absolute mass mode via an iterative search tree method. The first absolute mass search was initiated against a shotgun-annotated human proteome database containing PTMs, known alternative splice forms, coding single nucleotide polymorphisms (cSNPs), and peptide cleavage events (UniProt release 2012_06, 10,535,964 proteoforms) with 10,000 Da precursor window and 10 ppm fragment mass tolerance (*see* **Note 11**).
4. For initial searches that did not identify a protein below an E-value cutoff of 1×10^{-4}, a second absolute mass search took place against a simplified database including N-terminal acetylation and initial methionine cleavage (UniProt release 2012_06, 472,735 proteoforms) with 100,000 Da precursor window and 10 ppm fragment mass tolerance.
5. As for LC–MS/MS files generated from GELFrEE fractions containing up to 15 kDa, the .puf files were additionally searched in biomarker search mode against a simplified database (UniProt release 2012_06, 237,388 proteoforms) with

2.2 Da precursor window and 10 ppm fragment tolerance (*see* **Note 12**). At least four matched fragment ions and an E-value lower than 1×10^{-4} were required for protein identification [12, 13] (Fig. 3).

6. A Sequence Gazer tool in ProSightPC software was used to manually determine PTMs or single amino acid variations (SAAVs) (Fig. 4).

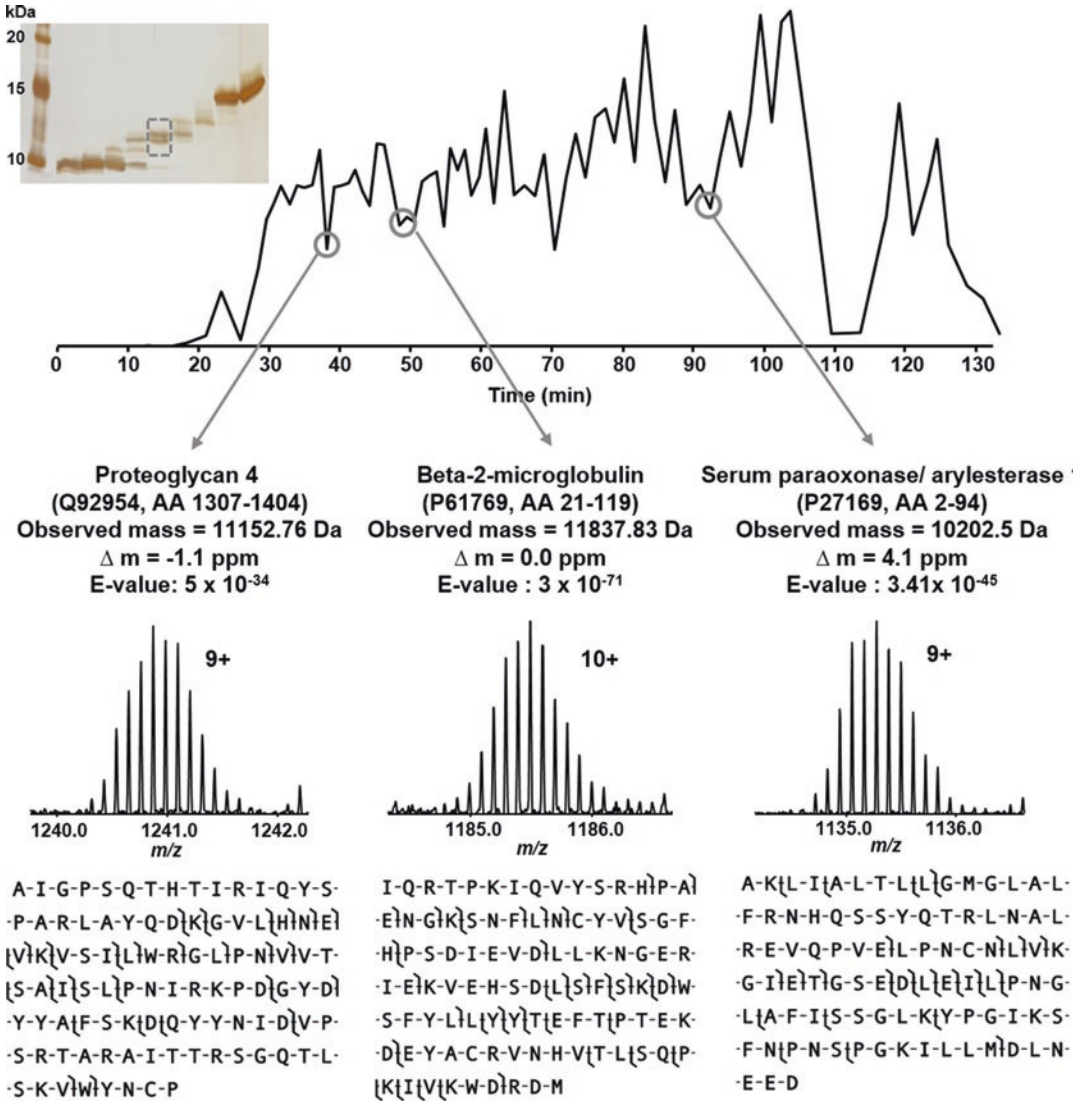


Fig. 3 An example of identification of LMW proteoforms from a GELFrEE fraction of plasma sample depleted of its seven high-abundance proteins via top-down approach. A total ion chromatogram is shown with intact mass spectra and graphical fragmentation maps for beta-2-microglobulin and cleaved products of proteoglycan 4 and serum paraoxonase/arylesterase 1. Reproduced from [14] with permission from the publisher

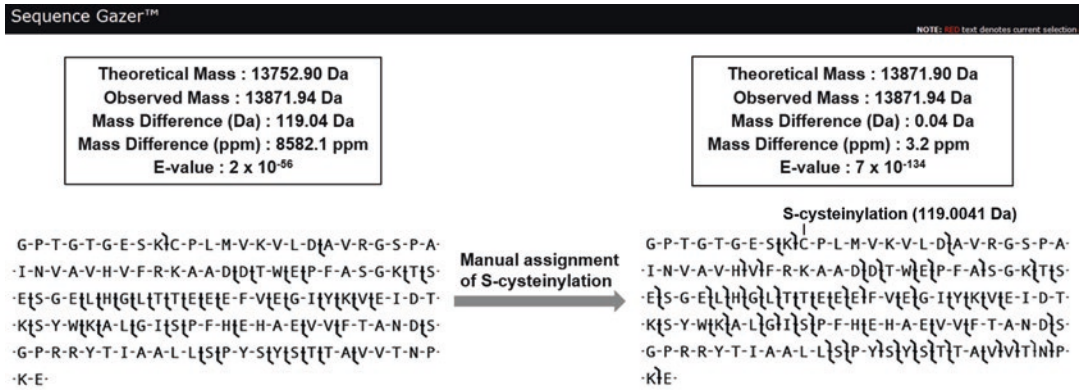


Fig. 4 Graphical fragmentation map of transthyretin (TTR) with S-cysteinylation that was manually assigned from Sequence Gazer tool in ProSightPC software. TTR protein was originally identified with one b-ion and 40 y-ions and with 119.04 Da of intact mass difference. According to UNIMOD (www.unimod.org), the mass difference can result from S-cysteinylation (119.0041 Da). When the mass value was added on the cysteine residue of the protein sequence from Sequence Gazer tool, 36 b-ions and one y-ion were additionally matched with intact mass difference of 3.2 ppm

4 Notes

1. Eight-channel multiplexed device for separating proteins based on molecular weight was fabricated as previously described [16]. GELFREE8100 device (Expedeon) can be used for the same purpose prior to top-down mass spectrometric analysis.
2. The individual human plasma sample was collected as suggested by the HUPO Plasma Proteome Project [18]. It is important to avoid freeze-thaw cycles for plasma samples without addition of protease inhibitors because the degraded products produced from sample processing steps can be identified via top-down approach. Protease and phosphatase inhibitor cocktails can be added immediately after the plasma samples are prepared from blood or added when they are first thawed.
3. When the plasma samples are needed to be pooled for top-down mass spectrometric analysis, they may be pooled with equal amounts prior to analysis.
4. The peak of flow-through fraction was typically seen from 1.5 min to 6 min from the chromatogram of the MARS-6 or MARS-7 column.
5. The resolving and stacking gels were usually polymerized within 1 h.

6. The plasma sample without depletion of high-abundance proteins was not subjected to reduction and alkylation because immunoglobulin light chains were supposed to be produced under reducing conditions and to be present in the GELFrEE fraction around 25 kDa. Thus, the reduction of disulfide bonds using DTT for the plasma sample without removal of high-abundance proteins was omitted in order to maximize the identification of plasma proteoforms that were endogenously present in the molecular weight region.
7. As for the protein loading amount to an SDS-PAGE tube gel, 850 μg of plasma sample without depletion of high-abundance proteins and 350 μg of the depleted plasma samples were loaded onto each SDS-polyacrylamide gel column because 850 μg of the plasma sample without depletion revealed similar separation efficiency of LMW fractions (<30 kDa) compared to that obtained from 350 μg of the depleted plasma samples. Since the high-abundance proteins including albumin and IgG constitute over 75% of the total proteins of plasma sample without depletion of highly abundant proteins, it is assumed that the amount of LMW proteoforms present in the plasma sample without removal of high-abundance proteins is relatively smaller than that present in the same amount of the depleted plasma samples.
8. After the entire portion of the blue dye had entered the collection chamber, the first fraction was collected.
9. While a GELFrEE fraction can be analyzed by top-down mass spectrometry, we found that a combination of multiple GELFrEE fractions with similar molecular weight ranges resulted in identification of more number of LMW proteoforms due to more intense MS and MS/MS signals.
10. From multiplexing mode, multiple precursor masses were selected within an isolation range as multiple precursors, based on an intensity cutoff (set at 10% here) relative to the base peak of the analysis window. The multiplexing mode allows identification of multiple precursors that are fragmented together in the same isolation window.
11. Absolute mass search involves matching the observed precursor mass to theoretical intact masses from a database within a user-specified tolerance and then comparing the observed fragment masses to those calculated from possible forms within a user-specified tolerance.
12. Biomarker search involves matching an observed mass to the theoretical masses of possible subsequences from the database within a precursor mass tolerance and then comparing the observed fragment masses to those calculated from the candidate subsequences within a user-specified tolerance.

Acknowledgments

The authors gratefully acknowledge the technical assistance from the Kelleher Group at Northwestern University. This work was supported by grants from the Multi-omics Program (2012M3A9B9036679) and the Brain Research Program (2015M3C7A1064795) funded through NRF supported by the Korean Ministry of Science, ICT and Future Planning and a KIST institutional program.

References

1. Liotta LA, Petricoin EF (2006) Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. J Clin Invest 116(1):26–30
2. Petricoin EF, Belluco C, Araujo RP, Liotta LA (2006) The blood peptidome: a higher dimension of information content for cancer biomarker discovery. Nat Rev Cancer 6(12):961–967
3. Drake RR, Cazares L, Semmes OJ (2007) Mining the low molecular weight proteome of blood. Proteomics Clin Appl 1(8):758–768
4. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA (2002) Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359(9306):572–577
5. de Noo ME, Deelder A, van der Werff M, Ozalp A, Mertens B, Tollenaar R (2006) MALDI-TOF serum protein profiling for the detection of breast cancer. Onkologie 29(11): 501–506
6. Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, Fleisher M, Lilja H, Brogi E, Boyd J, Sanchez-Carbayo M, Holland EC, Cordon-Cardo C, Scher HI, Tempst P (2006) Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest 116(1):271–284
7. Ebert MPA, Niemeyer D, Deininger SO, Wex T, Knippig C, Hoffmann J, Sauer J, Albrecht W, Malfertheiner P, Rocken C (2006) Identification and confirmation of increased fibrinopeptide A serum protein levels in gastric cancer sera by magnet bead assisted MALDI-TOF mass spectrometry. J Proteome Res 5(9):2152–2158
8. Greening DW, Simpson RJ (2010) A centrifugal ultrafiltration strategy for isolating the low-molecular weight (<or=25K) component of human plasma proteome. J Proteome 73(3):637–648
9. Harper RG, Workman SR, Schuetzner S, Timperman AT, Sutton JN (2004) Low-molecular-weight human serum proteome using ultrafiltration, isoelectric focusing, and mass spectrometry. Electrophoresis 25(9):1299–1306
10. Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD (2003) Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics 2(10):1096–1103
11. Smith LM, Kelleher NL, Consortium for Top Down P (2013) Proteoform: a single term describing protein complexity. Nat Methods 10(3):186–187
12. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL (2011) Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature 480(7376):254–258
13. Catherman AD, Durbin KR, Ahlf DR, Early BP, Fellers RT, Tran JC, Thomas PM, Kelleher NL (2013) Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. Mol Cell Proteomics 12(12):3465–3473
14. Cheon DH, Nam EJ, Park KH, Woo SJ, Lee HJ, Kim HC, Yang EG, Lee C, Lee JE (2016) Comprehensive analysis of low-molecular-weight human plasma proteome using top-down mass spectrometry. J Proteome Res 15(1):229–244
15. Tran JC, Doucette AA (2008) Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. Anal Chem 80(5):1568–1573
16. Tran JC, Doucette AA (2009) Multiplexed size separation of intact proteins in solution phase for mass spectrometry. Anal Chem 81(15):6201–6209
17. Wessel D, Flugge UI (1984) A method for the quantitative recovery of protein in dilute solution

in the presence of detergents and lipids. Anal Biochem 138(1):141–143

18. Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, Schuchard MD, Mehigh RJ, Cockrill SL, Scott GB, Tammen H, Schulz-Knappe P, Speicher DW, Vitzthum F, Haab BB, Siest G, Chan DW (2005) HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics 5(13):3262–3277

Chapter 9

Identification of Post-Translational Modifications from Serum/Plasma by Immunoaffinity Enrichment and LC-MS/MS Analysis Without Depletion of Abundant Proteins

Hongbo Gu, Jianmin Ren, Xiaoying Jia, and Matthew P. Stokes

Abstract

Immunoaffinity enrichment combined with LC-MS/MS enables identification of Post-translational modifications (PTMs) from serum/plasma samples without abundant protein depletion. Here we described the workflow in details in identifying various types of PTMs such as lysine acetylation and arginine methylation from cancer serum. The method described is compatible with all common proteomic analysis platforms and quantitative methods.

Key words Post-translational modification (PTM), Immunoaffinity purification (IAP), Serum/plasma, LC-MS/MS

1 Introduction

As noninvasive bio-fluid, serum/plasma has been preferred samples for identifying biomarkers due to ease of collection and richness in proteins and metabolites. Combining high-abundance protein immune depletion and LC-MS/MS analysis, researchers have identified and quantified thousands of proteins from serum/plasma samples [1]. However, majority of the effort in serum/plasma proteome research has been focused on total proteome analysis. Currently, there exists very limited data about Post-translational modifications (PTMs) in serum/plasma beyond glycosylation [2, 3]. As one of the most important mechanisms for regulating protein functions, PTMs including phosphorylation, acetylation, ubiquitination, and methylation have been identified and validated as critical for signaling transduction, protein degradation, and transcriptional regulation [4, 5]. Thanks to the fast development of enrichment methods including metal ion-based enrichment and antibody-based enrichment, peptides bearing

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_9,

PTMs can now be efficiently purified from complex digest of whole cell/tissue proteome; the enriched modified peptides can be analyzed in a high-throughput manner by LC-MS/MS for both qualitative and quantitative analysis [6–8]. However, one major challenge associated with identifying PTMs in serum/plasma is the large quantity of starting materials for immunoaffinity enrichment of PTM containing peptides. Because the top 20 abundant proteins take up most of the proteome in serum/plasma [9], it would be very costly to perform immuno-depletion of abundant proteins prior to enrichment for PTM containing peptides. Given the high specificity and enrichment efficiency of PTMScan workflow for PTM containing peptide enrichment, we have developed a robust procedure to achieve the identification of various types of PTMs from serum/plasma samples without immuno-depletion [10]. The method incorporates highly specific immunoaffinity purification of modified peptides from complex digest of serum/plasma digest followed by LC-MS/MS analysis of enriched peptides. This method allows PTM profiling from a reasonable volume of serum (~250 μL for multiple PTM enrichments). Among the PTMs surveyed, lysine acetylation (AcK) and arginine monomethylation (Rme) were identified as the more prevalent PTMs in cancer patients' sera. These PTMs were profiled in sera from patients with acute myelogenous leukemia (AML), breast cancer (BC), and non-small cell lung cancer (NSCLC). At 1% FDR, we have identified 796 unique AcK sites and 808 unique Rme sites in the sera of 12 cancer patients.

2 Materials

1. Urea lysis buffer: to prepare 20 mL urea lysis buffer, weight sequanal grade urea 10.8 g. Add 2 mL 200 mM HEPES stock solution, pH 8.0, and then add water to 20 mL. Rotate the mixture at room temperature until all urea powder dissolved (*see* **Note 1**). The final concentration of urea is 9 M in 20 mM HEPES buffer, pH 8.0.
2. 1.25 M dithiothreitol (DTT) solution: weight 193 mg DTT powder; add water to 1 mL to make DTT stock solution.
3. 100 mM iodoacetamide (IAA) solution: weigh 95 mg IAA powder; add water to 5 mL to make IAA stock solution (*see* **Note 2**).
4. Trypsin stock solution for first round digestion: dissolve 100 mg trypsin lyophilized powder (Worthington Biochemical) in 100 mL 1 mM HCl, and then aliquot to 1 mL/tube (*see* **Note 3**).

5. Immunoaffinity bead for PTM enrichment: for each serum/plasma sample, take one tube of bead from PTMScan kit (Cell Signaling Technology). Rinse the bead by 1 mL PBS buffer for four times; after the final rinse, leave small amount of PBS to cover the bead.
6. Immunoaffinity purification (IAP) buffer: IAP stock (10×) solution is included in PTMScan kit. IAP buffer (1×) contains 50 mM MOPS, 10 mM sodium phosphate, and 50 mM NaCl, pH 7.2.
7. IAP wash buffer: mix 1 mL NP-40 and 1 mL 10× IAP buffer, and then add water to the final volume of 10 mL, so the concentration of NP-40 is 1% in 1× IAP buffer.
8. C_{18}-stagetip: squeeze one tiny piece of Empore™ C_{18} material by a 18 × 4 needle with blunt end (Cadence Science); repeat this step and push two layers of C_{18} material in the needle into a 10 μL pipette tip by a capillary tubing (*see* **Note 4**).
9. C_{18}-stagetip conditioning buffer: 50% acetonitrile in 0.1% TFA.

   C_{18}-stagetip washing buffer: 0.1% TFA.

   C_{18}-stagetip eluting buffer: 40% acetonitrile in 0.1% TFA.
10. Ammonium bicarbonate (AMBIC) stock solution: weigh 80 mg AMBIC, and add HPLC grade water to 1 mL. The final concentration of AMBIC is 1 M.
11. Trypsin solution for secondary digestion: mix 190 μL HPLC grade water, 10 μL HPLC grade acetonitrile, and 6 μL 1 M AMBIC stock solution. The pH of digestion buffer should be around 8.0. Then add 5 μL sequence grade trypsin solution (Promega) to 75 μL digestion buffer to make final concentration of trypsin of 25 ng/μL (*see* **Note 5**).

3 Methods

3.1 Digestion of Serum/Plasma

1. Centrifuge serum/plasma at 16,000 × *g* for 15 min at 4 °C. Take about 250 μL supernatant mixed with 500 μL urea lysis buffer. Vortex the mixture and centrifuge the mixture again at 16,000 × *g* for 15 min 4 °C.
2. Add 2.7 μL DTT stock solution into the serum/plasma sample and incubate the mixture at 56 °C for 30 min to reduce disulfide bonds.
3. Cool the mixture to room temperature, and then add 75 μL freshly prepared IAA stock solution. Incubate the mixture at room temperature for 15 min at dark.
4. Dilute the sample fourfold with 20 mM HEPES buffer pH 8.0. Add 30 μL trypsin solution for first round of digestion with rotation at room temperature overnight.

5. Stop the digestion by adding 150 μL 20% TFA solution. Vortex the acidified digest solution at room temperature for 15 min, and then centrifuge the digest solution at 16,000 × *g* for 10 min.
6. Prepare Sep-Pak classic C_{18} columns (Waters) for desalting the digest solution: connect one Sep-Pak cartridge with a 10 mL syringe; add 5 mL acetonitrile to the syringe; add 10 mL 0.1% TFA in the syringe. After all solution drains through the syringe, the cartridge is ready for sample loading (*see* **Note 6**).
7. Load cleared supernatant of digest solution to the syringe to drain through the cartridge. Then add 1 mL, 3 mL, and 8 mL of 0.1% TFA stepwise to wash the syringe. Then add 2 mL of 5% acetonitrile in 0.1% TFA to the syringe to wash the cartridge.
8. Elute clean peptides from the cartridge by adding 1 mL, 3 mL, and 6 mL of 40% acetonitrile in 0.1% TFA to the syringe. Freeze eluent in −80 °C freezer for overnight before lyophilization to obtain peptide powder.

3.2 Immunoaffinity Purification

1. Dissolve lyophilized peptide powder by 1.5 mL 1× IAP buffer. Vortex the peptide solution vigorously and centrifuge the solution at 16,000 × *g* at 4 °C for 10 min. Check if pH of the peptide solution is around 7.2 (*see* **Note 7**).
2. Transfer cleared supernatant of the peptide solution to the vial of washed PTMScan bead. Rotate the mixture at 4 °C for 2 h.
3. Centrifuge the mixture at 2000 × *g* for 30 s to spin down the bead, and transfer the supernatant to a new tube.
4. Add 1.5 mL IAP wash buffer to the bead and rotate the mixture for 30 min at 4 °C. Repeat the wash step for a total of three times.
5. Remove residue of wash buffer after the final wash step by gel loading tip. Then add 1 mL ice-cold HPLC grade water to the bead. Shake the tube up and down several times to rinse off residual IAP wash buffer. Repeat the HPLC grade water wash for a total of three times. Remove residue of HPLC grade water after the final wash step by gel loading tip.
6. Elute enriched peptides by adding 40 μL 0.15% TFA solution. Tap the tube gently to keep the bead suspending at room temperature for 10 min. Transfer the eluent by a gel loading tip to a new tube. Repeat the elution step by adding 35 μL 0.15% TFA solution and combine eluents.
7. Prepare the C_{18}-stagetip: condition the stagetip by passing 50 μL C_{18}-stagetip conditioning buffer at 1500 × *g* for 1 min; wash the stagetip by passing 50 μL C_{18}-stagetip washing buffer at 1500 × *g* for 1 min; repeat this step.

8. Load eluent from **step 6** to the stagetip and pass the eluent at 1500 × *g* for 2 min; repeat this step. Wash the stagetip by passing 50 μL C_{18}-stagetip washing buffer at 1500 × *g* for 1 min; repeat this step.
9. Elute clean peptides from the stagetip by passing 10 μL C_{18}-stagetip eluting buffer at 750 × *g* for 1 min; repeat this step. Dry the eluents under vacuum.
10. Resuspend dried peptides by 10 μL trypsin solution for secondary digestion. Incubate the peptide solution at 37 °C for 2 h. Stop the reaction by adding 1 μL 5% TFA and 40 μL 0.1% TFA, and perform another round of C_{18}-stagetip cleaning of peptides. Now the peptide sample is ready for LC-MS/MS.

3.3 LC-MS/MS Analysis

1. Resuspend dry peptides in 0.125% formic acid; the volume of resuspension buffer depends on the number of injections and injection volume. For example, enriched peptides from 250 μL serum/plasma are enough for three LC-MS/MS injections. Sample can be resuspended by 12.5 μL 5% ACN in 0.1% TFA for three injections at 4 μL each.
2. Enriched peptides are separated using a 120-min linear gradient of acetonitrile in 0.125% formic acid delivered at 280 nL/min from 3% to 30%.
3. Tandem mass spectra are collected in a data-dependent manner with an LTQ-Orbitrap Elite mass spectrometer running Xcalibur 2.0.7 SP1 using a top 20 MS/MS method, a dynamic repeat count of one, and a repeat duration of 30 s. The isolation window is set at 1.0 Da with a normalized collision energy of 35%. Real-time recalibration of mass error was performed using lock mass with a singly charged polysiloxane ion m/z = 371.101237.
4. MS/MS spectra were evaluated using SEQUEST (*see* **Note 8**). Files were searched against the Swiss-Prot *Homo sapiens* FASTA database concatenated with corresponding reverse database. A mass accuracy of ±5 ppm was used for precursor ions and 1 Da for product ions. Enzyme specificity was limited to trypsin, with at least one tryptic terminus required per peptide and up to four mis-cleavages allowed. Cysteine carboxamidomethylation was specified as a static modification and oxidation of methionine residue, and the appropriate PTMs were allowed as variable modifications for each enrichment sample set. Reverse decoy databases were included for all searches to estimate false discovery rates and filtered using a 1% FDR at peptide level. Example of MS/MS spectrum of acetylation on K298 of albumin was shown in Fig. 1.
5. The workflow described above is compatible with all major quantitative proteomic analysis methods including iTRAQ, TMT, and label-free quantification [8, 10].

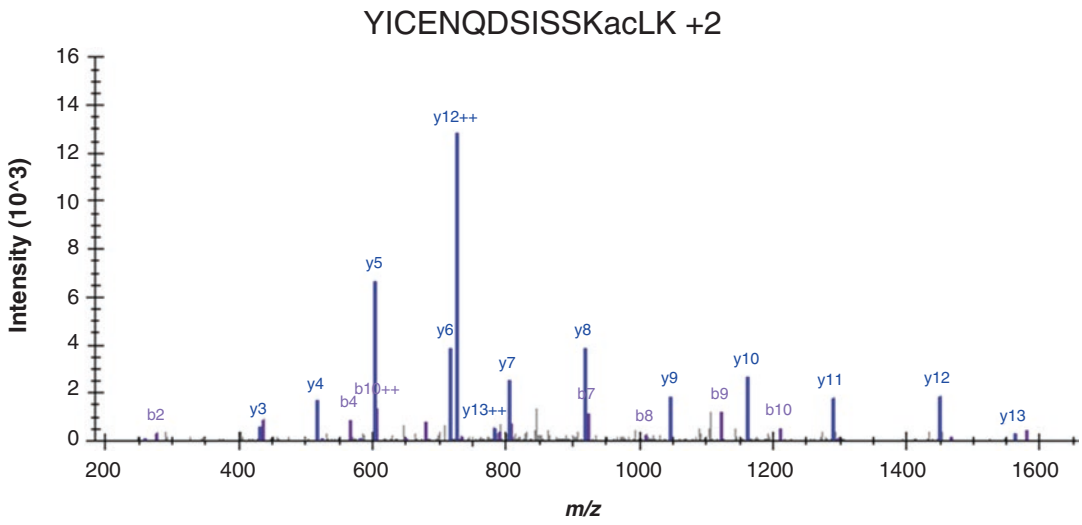


Fig. 1 A typical MS/MS spectrum of lysine acetylation on K298 on the protein albumin. Site-determining ions (b12, y2, and y3 ions) indicate the location of acetylation (ac) in the sequence: YICENQDSISSKacLK

4 Notes

1. It may take long to dissolve urea powder. To accelerate the process, warm water bath (37 °C) can be used. Hot water is not recommended as urea generated cyanate under high temperature, which causes protein/peptide carbamylation.
2. IAA solution needs to prepare fresh and short before use. Avoid light when adding water to IAA powder.
3. Avoid freeze/thawing of trypsin stock solution. The trypsin from Worthington Biochemical is for the first round of digestion. For the purpose of cost-saving, we recommend using lower-grade trypsin for the first round digestion. Although certain level of mis-cleavages is expected, the issue can be efficiently minimized by using sequencing grade trypsin for the second round digestion (Subheading 3.2, **step 10**) of enriched peptides.
4. Detailed procedure for making stop-and-go extraction tips (stagetips) can be found in Rappsilber et al. [11]. The loading capacity of the stagetip described in the current method is about 2 μg peptides, which is enough for each single immunoaffinity purification of PTM containing peptides from ~250 μL serum/plasma.
5. The purpose of secondary digestion after IAP is to (1) minimize mis-cleavage during first round of digestion and (2) cut antibody or antibody fragments left in the enriched peptide mixture. Our data showed that online analytical column can be better reserved when samples were treated with secondary digestion.

6. Air bubbles could be stacked at the connection of cartridge and syringe, which can be removed by using gel loading tip. It is recommended to remove air bubbles by gel loading tip every time when new buffer or samples are loaded.
7. Insufficient lyophilization could lead to trace amount of TFA left in the peptide powder, which causes slightly lower pH of the peptide solution. If necessary, adjust the pH of the peptide solution to 7.2 by 1 M Tris base buffer.
8. All search engines capable of analyzing data of high mass accuracy can be used to search acquired data including Mascot, MaxQuant, X! Tandem, Comet, and OMSSA. If MS/MS spectra are acquired with high mass accuracy such as HCD mode of Q Exactive mass spectrometer, fragment ion tolerance should be set to 0.02 Da.

References

1. Keshishian H, Burgess MW, Gillette MA, Mertins P, Clauser KR, Mani DR, Kuhn EW, Farrell LA, Gerszten RE, Carr SA (2015) Multiplexed, quantitative workflow for sensitive biomarker discovery in plasma yields novel candidates for early myocardial injury. Mol Cell Proteomics. doi:10.1074/mcp.M114.046813
2. Berven FS, Ahmad R, Clauser KR, Carr SA (2010) Optimizing performance of glycopeptide capture for plasma proteomics. J Proteome Res 9(4):1706–1715. doi:10.1021/pr900845m
3. Chen S, Lu C, Gu H, Mehta A, Li J, Romano PB, Horn D, Hooper DC, Bazemore-Walker CR, Block T (2012) Aleuria Aurantia Lectin (AAL)-reactive immunoglobulin G rapidly appears in sera of animals following antigen exposure. PLoS One 7(9):e44422. doi:10.1371/journal.pone.0044422
4. Huang H, Lin S, Garcia BA, Zhao Y (2015) Quantitative proteomic analysis of histone modifications. Chem Rev 115(6):2376–2418. doi:10.1021/cr500491u
5. Prabakaran S, Lippens G, Steen H, Gunawardena J (2012) Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. Wiley Interdiscip Rev Syst Biol Med 4(6):565–583. doi:10.1002/wsbm.1185
6. Hornbeck PV, Chabra I, Kornhauser JM, Skrzypek E, Zhang B (2004) PhosphoSite: a bioinformatics resource dedicated to physiological protein phosphorylation. Proteomics 4(6):1551–1561. doi:10.1002/pmic.200300772
7. Sharma K, D'Souza RC, Tyanova S, Schaab C, Wisniewski JR, Cox J, Mann M (2014) Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. Cell Rep 8(5):1583–1594. doi:10.1016/j.celrep.2014.07.036
8. Svinkina T, Gu H, Silva JC, Mertins P, Qiao J, Fereshetian S, Jaffe JD, Kuhn E, Udeshi ND, Carr SA (2015) Deep, quantitative coverage of the lysine acetylome using novel anti-acetyllysine antibodies and an optimized proteomic workflow. Mol Cell Proteomics. doi:10.1074/mcp.O114.047555
9. Mitchell P (2010) Proteomics retrenches. Nat Biotechnol 28(7):665–670. doi:10.1038/nbt0710-665
10. Gu H, Ren JM, Jia X, Levy T, Rikova K, Yang V, Lee KA, Stokes MP, Silva JC (2016) Quantitative profiling of post-translational modifications by immunoaffinity enrichment and LC-MS/MS in cancer serum without immunodepletion. Mol Cell Proteomics 15(2):692–702. doi:10.1074/mcp.O115.052266
11. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2(8):1896–1906. doi:10.1038/nprot.2007.261

Chapter 10

Identification of Core-Fucosylated Glycoproteome in Human Plasma

Qichen Cao, Qing Zhao, Xiaohong Qian, and Wantao Ying

Abstract

The core-fucosylated (CF) glycoproteins are widely distributed in mammalian tissues and regulated under pathological conditions, especially in cancer progression. The Food and Drug Administration (FDA) has approved the core-fucosylated α-fetoprotein as a biomarker for the early diagnosis of hepatocellular carcinoma (HCC). An approach for identifying CF glycoproteins has significantly practical value. Here we introduce a novel method for identification of CF glycoproteome in human plasma. The method integrates tandem glycopeptide enrichment, stepped fragmentation, and "glycan diagnostic ion"-based spectrum refinement. With this method, the productivity of identifying CF glycopeptides will be significantly improved. We anticipate that this method could be widely utilized to explore the CF glycoproteins and their regulation under physiological or pathological condition.

Key words Core fucosylation, Glycoproteomics, Human plasma, HILIC, Lectin affinity, Stepped fragmentation, Mass spectrometry

1 Introduction

Core fucosylation (CF), characterized as an α-1, 6 fucose substitution on the innermost N-acetylglucosamine (GlcNAc) of the pentasaccharide core of N-linked glycans, is a special glycosylation pattern of proteins with many biological or pathological functions. Recent investigations linked the CF glycoproteins to hepatocellular carcinoma (HCC) [1, 2], pancreatic cancer [3, 4], lung cancer [5, 6], ovarian cancer [7], and prostate cancer [8]. The core-fucosylated α-fetoprotein (AFP-L3) has been approved by the Food and Drug Administration (FDA) for the early diagnosis of HCC. Based on our previous research, three CF glycopeptides are increased in HCC patients' plasma [9].

Many efforts have been made to identify CF glycoproteins [1, 2, 10–13]. Briefly, lectin affinity followed with hydrophilic interaction chromatography (HILIC) or low molecular weight (MW) cutoff is conducted for CF glycopeptide enrichment.

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_10,

Neutral loss-dependent MS3 or ETD scans are used to generate the characteristic CF glycopeptide spectra. However, several problems still exist that may limit the efficiency and productivity of these methods. Firstly, the MW cutoff is less specific toward glycopeptides because large nonglycopeptides or small glycopeptides may retain on or pass through the filter. Secondly, the neutral loss-triggered MS3 or ETD scan speed is slow, and the MS3 scan is run on an ion trap-type mass spectrometer with a low mass accuracy. Moreover, the diagnostic ions of glycans were always lost due to the low mass cutoff of the ion trap.

Here we introduced the novel method for precise and large-scale identification of CF glycopeptides from human plasma samples. Three key steps are combined to this method (Fig. 1): (1) "Stepped MS2 fragmentation" function is used to obtain high accuracy fragment ions from both the glycan and peptide of a simplified CF glycopeptide. The high-speed merit of the MS2 scan for a Q Exactive mass spectrometer is also retained. (2) The "glycan diagnostic ion"-based spectrum optimization method is employed to improve the efficiency and accuracy of the discovery of CF glycopeptides (Fig. 2). (3) Tandem application of hydrophilic interaction chromatography (HILIC) and lectin affinity

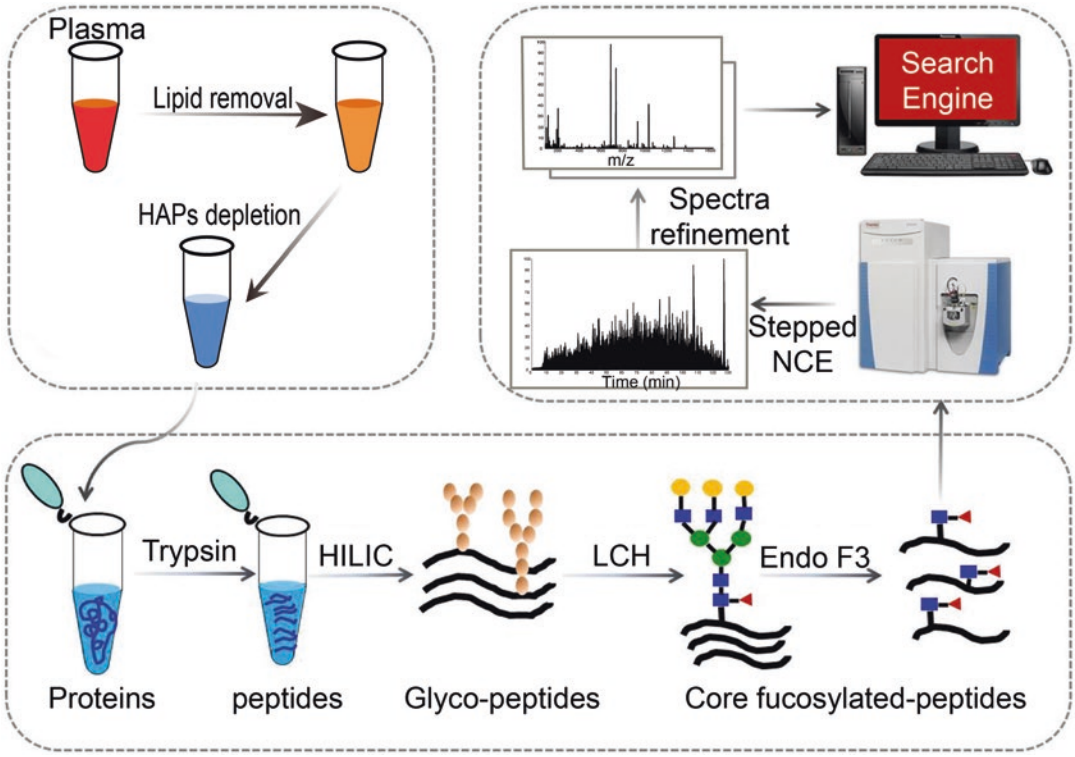


Fig. 1 Schematic workflow for identification of core-fucosylated glycoproteome in human plasma. The method integrates tandem glycopeptides enrichment, stepped MS fragmentation, and "glycan diagnostic ion"-based spectra refinement. *HAPs:* high-abundance proteins

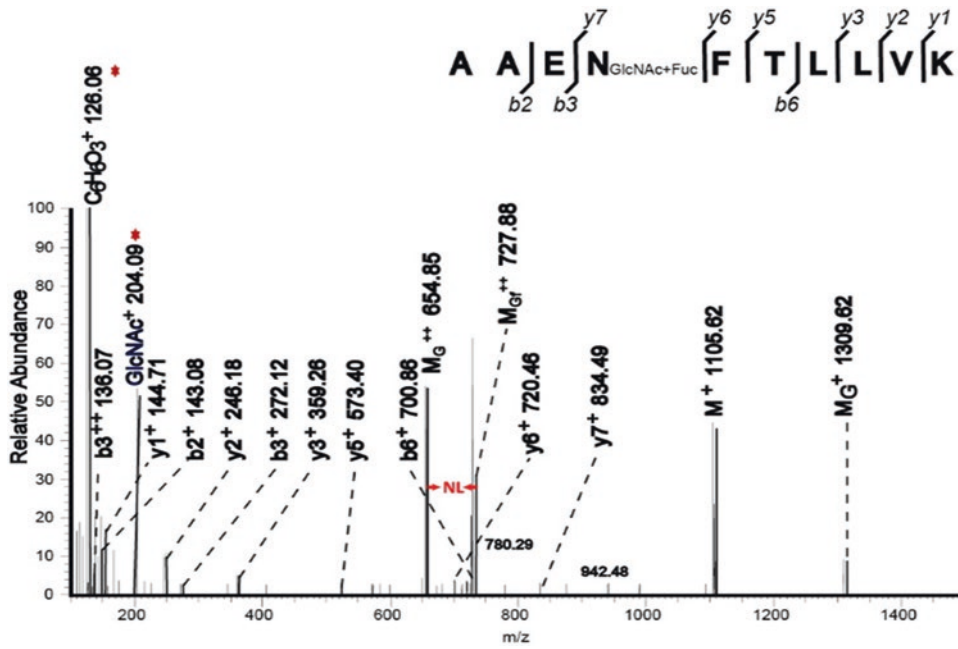


Fig. 2 The characteristic MS2 spectra of simplified CF glycopeptide generated by stepped fragmentation. *MGF*$^{++}$ the parent ion attached with GlcNAc and fucose residue; *MG*$^{++}$ *or MG*$^{+}$ the fragment parent ion attached with a GlcNAc residue; *GlcNAc*$^{+}$ the diagnostic ion of GlcNAc; *y6*$^{+}$ the y-type ion without glycan; *red NL* neutral loss

chromatography is adopted to improve the efficiency and reproducibility of CF glycopeptide enrichment.

With this method, the capability of identifying CF glycopeptides will be significantly increased. The feasibility and reproducibility of this method have been demonstrated through the analysis of CF glycoproteomes of plasma from HCC patients. We anticipate that this strategy could be utilized for a variety of applications, in which CF glycoproteins need to be identified under different physiological or pathological conditions.

2 Materials

All of the solutions involved in this procedure were prepared with analytical grade reagents and ultrapure water. The binding buffer for lectin affinity enrichment was stored and working under 4 °C. The other solutions were all freshly prepared before usage (if not mentioned specially).

2.1 Depletion of Plasma High-Abundance Proteins

1. ProteoExtract™ Albumin/IgG Removal Kit (Calbiochem, USA) is used in this method to remove high-abundance proteins from plasma sample.
2. Centrifugal filter device: Amicon® Ultra-15 (10 KD) (Millipore) is used in this method.

2.2 Protein Digestion

1. Protein denature solution: 8 M urea dissolved in 50 mM ammonium bicarbonate solution (pH 8.0).
2. 1 M Dithiothreitol (DTT) solution in water.
3. 1 M Iodoacetamide (IAA) solution in water, stored and used under darkness.
4. 50 mM Ammonium bicarbonate solution in water (pH 8.0).
5. Calcium chloride ($CaCl_2$, Sigma): 1 M solution in water.
6. Formic acid (FA, Sigma).
7. Trypsin solution: 0.5 μg/μL trypsin (e.g., sequencing grade trypsin lyophilized powder) in 50 mM acetic acid solution (prepared just before usage).

2.3 Peptide Desalting

1. C18 solid-phase extraction (SPE) cartridge: sorbent bed weight 50 mg (e.g., Sep-Pak C18 cartridges, Waters).
2. Desalting wash buffer: 0.1% FA (*v/v*) dissolved in water.
3. Desalting elution buffer: acetonitrile (ACN)-water-FA (49.9: 50: 0.1, *v/v/v*).

2.4 Glycopeptide Enrichment by HILIC

1. 0.1% TFA: 0.1% trifluoroacetic acid (TFA) (*v/v*) dissolved in water.
2. Binding buffer for HILIC (BB_H): ACN-water (TFA) (80:19.8:0.2, *v/v/v*).
3. Elution buffer for HILIC ((EB_H): 0.1 % TFA (*v/v*) dissolved in water.
4. HILIC sorbent: Venusil HILIC silica (5 μm ID) from Agela Technologies (Tianjin, China) is used as an example in this method.

2.5 CF Glycopeptide Enrichment by Lectin

1. Lectin-conjugated agarose (*see* **Note 1**): Lentil Lectin Sepharose 4B (GE Healthcare) is used in this method.
2. Binding buffer for lectin affinity enrichment (BB_L): 0.5 M NaCl, 1 mM $MnCl_2$, and 1 mM $CaCl_2$ dissolved in 20 mM Tris–HCl (pH 7.45), stored and worked under 4 °C.
3. Elution buffer for lectin affinity enrichment (EB_L): 1 M acetic acid dissolved in water.

2.6 Endoglycosidase Digestion

1. Endo F3: Endoglycosidase F3 (Endo F3, Sigma-Aldrich) is used in this method (*see* **Note 2**).
2. Digestion buffer for Endo F3 (DB_E): 0.1 M ammonium acetate dissolved in water adjusted to pH 4.5 with acetic acid (*see* **Note 3**).

2.7 NanoLC-MS/MS Analysis

1. Mobile phase A: 0.1% FA (*v/v*) dissolved in water.
2. Mobile phase B: ACN-water-FA (98:1.9:0.1, *v/v/v*).

3. Commercially available fused silica capillary C18 HPLC column (75 μm ID, 10 cm length).
4. NanoLC-MS/MS system: Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo, USA) equipped with a nanoelectrospray ionization source and an Easy-nLC 1000 high-performance liquid chromatography system (Thermo, USA) is used as an example in this method.

2.8 MS Data Analysis

1. MS raw file format conversion: MSConvert from ProteoWizard (3.0.5009) is used in this method.
2. Search engine: Mascot (version 2.3, Matrix Science) is used in this method.

3 Methods

All procedures are carried out at room temperature unless otherwise noted. When handling blood products, always take appropriate personal protective equipment to minimize biohazards from aerosols.

3.1 Sample Preparation

The blood sampling and collection procedure is approved by the clinical ethics committee.

3.1.1 Depletion of High-Abundance Proteins (HAPs) from Human Plasma

1. Prechill centrifuge and rotors to 4 °C.
2. If necessary, thaw plasma at room temperature.
3. Add 50 μL plasma to a microcentrifuge tube and centrifugate at ~14,000 × *g* for 10 min under 4 °C. Place plasma on wet ice and aspirate the lipid layer on the top of the plasma.
4. ProteoExtract™ Albumin/IgG Removal Kit (Calbiochem, USA) is used for plasma HAP depletion according to the operation manual (*see* **Note 4**).
5. Transfer the collected HAPs depletion sample to a 10 KD (15 mL) centrifugal filter device. Add 10 mL water to the plasma and centrifugate at 4,000 × *g* for 30 min under 4 °C.
6. Insert a pipettor into the filter device and withdraw the sample.
7. Measure the protein concentration using BCA assay.
8. Inspect the plasma sample by SDS-PAGE analysis (Fig. 3).
9. Lyophilize the plasma sample and store under −80 °C for further analysis.

3.1.2 Digestion of HAP Depleted Plasma Sample Using Trypsin

1. In a 1.5 mL microtube, dissolve ~1 mg HAPs depleted plasma sample in 100 μL protein denature solution.
2. Add 1 μL DTT solution to the plasma sample and incubate the sample under 37 °C for 4 h.
3. Add 4 μL IAA solution to the sample and incubate the sample at room temperature in darkness for 30 min.

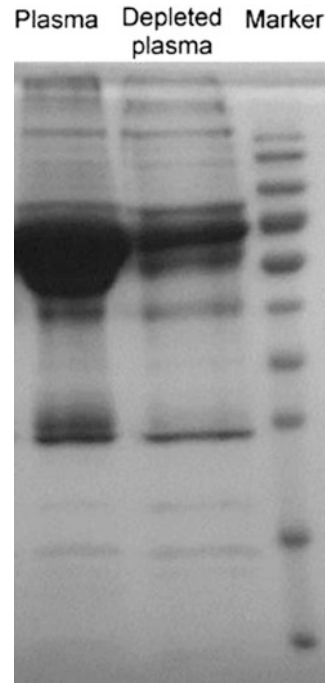


Fig. 3 Depletion of high-abundance proteins from human plasma. The plasma sample is treated with ProteoExtract™ Albumin/IgG Removal Kit (Calbiochem, USA)

4. Add additional 1 μL DTT solution to the plasma sample and incubate at room temperature for 30 min.
5. Add 900 μL ammonium bicarbonate solution (50 mM, pH 8.0) and 1 μL Ca_2Cl solution to the sample.
6. Add 20 μL trypsin solution, gently vortex, and incubate under 37 °C for 4 h. Be sure to place trypsin back on ice.
7. Add additional 20 μL trypsin solution, gently vortex, and incubate under 37 °C for 12 h.
8. Add 5–10 μL FA to stop the digestion.
9. Centrifuge the sample at 14,000 × *g* for 10 min. Transfer the peptides in the supernatant to another microtube.
10. Place one C18 SPE cartridge (e.g., Waters Sep-Pak C18 Vac cartridge) on the SPE vacuum manifold. The vacuum should be pre-adjusted to make the flow rate up to ~1 drop/s. Sequencely add 1 mL ACN and 1 mL desalting wash buffer to flow through the sorbent.
11. Turn off the vacuum, load the peptides onto the SPE cartridge, and wait for the peptide to flow through the sorbent by gravity.
12. Turn on the vacuum, and add 2 mL desalting wash buffer to flow through the sorbent.
13. Place a new microtube under the SPE cartridge. Add 400 μL desalting elution buffer to the cartridge for twice. Collect and combine the eluted fractions.

14. Dry the desalted peptides under vacuum and store under −80 °C if necessary.

3.2 Enrichment of Glycopeptides by HILIC

1. Take 50 mg HILIC sorbent into a 600 μL microtube, add 500 μL 0.1% TFA (*v*/*v*), gently shake for 15 mins, and discard the supernatant.
2. Add 400 μL BB_H solution to the sorbent, gently shake for 15 min, spin the microtube, and discard the supernatant. Repeat for three times.
3. Dissolve 1.0 mg lyophilized peptides in 100 μL BB_H solution, and centrifuge at 14,000 × *g* for 5 min. Transfer the supernatant to the HILIC sorbent, and gently vortex. Seal the microtube with Parafilm properly and shake for 2 h at room temperature.
4. Spin the microtube at 100 rpm for 1 min and discard the supernatant.
5. Add 400 μL BB_H solution to the sorbent, and gently shake for 5 min. Spin the microtube at 100 rpm for 1 min and discard the supernatant. Repeat for three times.
6. Add 400 μL EB_H solution to the sorbent, gently vortex, and shake the microtube for 15 min. Spin the microtube and transfer the supernatant to a new 1.5 mL microtube. Repeat for three times and combine the supernatant solution.
7. Centrifuge the eluted glycopeptides at 14,000 × *g* for 5 min, and discard the pellets. Dry the supernatant containing glycopeptides under vacuum.

3.3 Enrichment of CF Glycopeptides by LCH Lectin

1. Take 100 μL LCH 50% slurry into a 600 μL microtube, add 400 μL BB_L solution, and gently shake for 5 min under 4 °C. Spin the microtube at 100 rpm for 1 min and discard the supernatant. Repeat for three times.
2. Dissolve the HILIC isolated glycopeptides in 100 μL BB_L solution, and centrifuge at 14,000 × *g* for 5 min under 4 °C. Transfer the supernatant onto the LCH lectin resin. Gently vortex and shake overnight under 4 °C.
3. Spin the microtube at 100 rpm for 1 min and discard the unbounded peptides. Add 400 μL BB_L solution and gently shake for 5 min under 4 °C. Repeat for three times.
4. Spin the microtube and discard the supernatant. Add 150 μL EB_L solution, and gently shake for 15 min. Spin the microtube and collect the supernatant to a new 1.5 mL microtube. Repeat for three times and combine the supernatant solution.
5. Centrifuge the microtube at 14,000 × *g* for 5 min and discard the pellets. Dry the CF glycopeptides under vacuum.

3.4 Endoglycosidase Digestion of CF Glycopeptides

1. Dissolve the isolated CF glycopeptides in 50 μL DB_E solution (*see* **Notes 3** and **5**), and centrifuge at 14,000 × *g* for 5 min. Transfer the supernatant into a new microtube, add 0.02 UN Endo F3 (*see* **Note 2**), and incubate under 37 °C overnight.
2. Desalt the simplified glycopeptides with C18 SPE cartridge followed in Subheading 3.1.2, **steps 10–14**.
3. Transfer the eluent to 1.5 mL low protein binding microcentrifuge tubes and dry the desalting peptides under vacuum.
4. Store dried samples under −80 °C before mass spectrometric analysis.

3.5 NanoLC-MS/MS Analysis

An example configuration of the nanoLC-MS/MS system is Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo, USA) equipped with a nanoelectrospray ionization source and an Easy-nLC 1000 high-performance liquid chromatography system (Thermo, USA). The data-dependent acquisition (DDA) and "stepped NCE" acquisition mode are employed (*see* **Note 6**). The method can be adapted to other high-resolution tandem mass spectrometers with a beam-type collision cell (e.g., Q-TOF/HCD-Orbitrap/Q-FT ICR).

3.5.1 Sample Preparation for NanoLC-MS/MS Analysis

1. Dissolve the dried sample in 10 μL mobile phase A.
2. Centrifuge the sample at 14,000 × *g* for 5 min.
3. Transfer the peptide samples to the compatible vials for the autosampler coupled to the nanoLC-MS/MS system. Ensure that no bubbles are present in the solution.

3.5.2 Chromatographic Conditions

1. Load 5 μL sample on the trapping column for 10 min with mobile phase A at the flow rate 2 μL/min.
2. Peptides are eluted from the trap column onto the analytical column and separated by the gradient detailed in Table 1.

3.5.3 Mass Spectrometer Parameters

Spray voltage: +2000 V.

Capillary temperature: 320 °C.

MS1 scan range: 300–1400 *m/z*.

MS1 automatic gain control (AGC): $3e^6$.

MS1 resolution: 70,000.

MS1 intensity threshold: $6.3e^3$.

MS1 maximum ion injection time: 60 ms.

MS1 isolation window: 3.0 *m/z*.

Top N: 20.

Normalized collision energy (NCE): 22.

Stepped NCE: 54.5%.

Table 1
Chromatography gradient and flow rate for nanoLC-MS/MS analysis of CF glycopeptides

Time (min)	Mobile phase A (%)	Mobile phase B (%)	Flow rate (nL/min)
0	95	5	300
2	92	8	300
60	70	30	300
65	50	50	300
68	20	80	300
73	20	80	300
74	95	5	300
84 (end)	95	5	300

MS2 resolution: 17,500.

MS2 automatic gain control (AGC): $5e^4$.

MS2 maximum ion injection time: 80 ms.

Dynamic exclusion time: 18 s.

3.6 MS Data Analysis

3.6.1 Peak Picking and Spectra Refinement for CF Glycopeptides

1. Convert the MS raw data file into proper editable format (e.g., MGF, mzML, mzXML).
2. Pick up the MS2 spectra for which precursor ion and the neutral loss ion with the mass shift of 146.0579 Da (10 mmu window) coexisted and also containing the diagnostic ions of the GlcNAc including m/z 126.0555, m/z 138.0555, m/z 186.0761, and m/z 204.0872 (10 mmu window).
3. Subtract the residue mass of fucose (146.0579 Da) from the precursor mass and remove the precursor ion peak (4 Da window) from the fragment ion list.
4. Remove the diagnostic ions of the GlcNAc including m/z 126.0555, m/z 138.0555, m/z 186.0761, and m/z 204.0872 (10 mmu window) from the MS2 peak list.
5. Merge the selected CF glycopeptide candidate spectra for database searching.

3.6.2 Database Searching

Preprocessed spectra file from Subheading 3.6.1 is searched against a human protein sequence database (e.g., UniProt) using Mascot. Fixed modification contains carbamidomethylation (+57.0215 Da) on Cys residues. Variable modifications contain oxidation (+15.9949 Da) of Met residues, acetylation (+42.0106 Da) of protein N-terminals, and a HexNAc (+203.0794 Da) variable addition to Asn

residues. Trypsin is specified with at most two missed cleavage sites. The tolerance of the parent ions is 20 ppm and that of the fragment ions is 20 mmu. The other settings include target-decoy database searching strategy for estimation of false discovery rate (FDR). The HexNAc-modified Asn (agreeing with glycosylation sequence motif of NX(S/T/C)) is considered as the core-fucosylated glycosite.

4 Notes

1. Many lectins could bind to core (α1-6)-fucosylated glycans including *Lens culinaris* lectin (LCA/LCH), *Aleuria aurantia* lectin (AAL), *Aspergillus oryzae* lectin (AOL), *Pisum sativum* lectin (PSA), etc. LCA and PSA can be used as specific probes for detecting core-fucosylated, mono- and biantennary *N*-glycans, such as those attached to AFP-L3. However, AOL and AAL also bound to α1-2-, α1-3-, and α1-4-fucosylated glycans with a range of affinities. One should select the proper lectin(s) based on the glycoform of the analytes.
2. Endoglycosidase F3 cleaves asparagine-linked biantennary and triantennary complex N-linked oligosaccharides. There is no activity on oligomannose and hybrid molecules. Therefore, Endo H may be used as a complement for more thorough coverage of CF glycopeptides [14–17].
3. The pH of digestion buffer should be optimized according to the endoglycosidase used for glycan simplification.
4. The dynamic range of protein concentrations in plasma exceeds ten orders of magnitude and brings huge challenge for detection of medium- and low-abundance proteins in proteomic analyses. The top ten most abundant plasma proteins account for approximately 90% of the total protein content. Thus, in almost all studies on plasma proteome, the first step is the depletion of high-abundance proteins (HAPs). However, it should also be noted that many HAPs (e.g., IgG, alpha-1 antitrypsin) also contain the core-fucosylated modification.
5. The pH of digestion buffer should be carefully inspected.
6. In the "stepped NCE" acquisition mode, the Q Exactive MS (Thermo) performs three stepped fragmentations on the same precursor ion with low, median, and high NCE value (10, 22, and 34 for this method). All of the fragment ions from the three stepped fragmentations are collected and injected into the Orbitrap for detection in a single scan. Similar function is termed as CES (collision energy spread) for TripleTOF 5600 MS (AB Sciex).

Acknowledgments

This work was supported by the grants from the National Key Program for Basic Research of China (2014CBA02001, 2016YFA0501300) and the National Natural Science Foundation of China (81530021, 21505151).

References

1. Block TM, Comunale MA, Lowman M et al (2005) Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. Proc Natl Acad Sci U S A 102:779–784
2. Comunale MA, Lowman M, Long RE et al (2006) Proteomic analysis of serum associated fucosylated glycoproteins in the development of primary hepatocellular carcinoma. J Proteome Res 5:308–315
3. Okuyama N, Ide Y, Nakano M et al (2006) Fucosylated haptoglobin is a novel marker for pancreatic cancer: a detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation. Int J Cancer 118:2803–2808
4. Barrabes S, Pages-Pons L, Radcliffe CM et al (2007) Glycosylation of serum ribonuclease 1 indicates a major endothelial origin and reveals an increase in core fucosylation in pancreatic cancer. Glycobiology 17:388–400
5. Geng F, Shi BZ, Yuan YF et al (2004) The expression of core fucosylated E-cadherin in cancer cells and lung cancer patients: prognostic implications. Cell Res 14:423–433
6. Wang X, Inoue S, Gu J et al (2005) Dysregulation of TGF-beta1 receptor activation leads to abnormal lung development and emphysema-like phenotype in core fucose-deficient mice. Proc Natl Acad Sci U S A 102:15791–15796
7. Saldova R, Royle L, Radcliffe CM et al (2007) Ovarian cancer is associated with changes in glycosylation in both acute-phase proteins and IgG. Glycobiology 17:1344–1356
8. Tabares G, Radcliffe CM, Barrabes S et al (2006) Different glycan structures in prostate-specific antigen from prostate cancer sera in relation to seminal plasma PSA. Glycobiology 16:132–145
9. Zhao Y, Jia W, Wang J et al (2011) Fragmentation and site-specific quantification of core fucosylated glycoprotein by multiple reaction monitoring-mass spectrometry. Anal Chem 83:8802–8809
10. Comunale MA, Wang M, Hafner J et al (2009) Identification and development of fucosylated glycoproteins as biomarkers of primary hepatocellular carcinoma. J Proteome Res 8:595–602
11. Cao Q, Zhao X, Zhao Q et al (2014) Strategy integrating stepped fragmentation and glycan diagnostic ion-based spectrum refinement for the identification of core fucosylated glycoproteome using mass spectrometry. Anal Chem 86:6804–6811
12. Chen R, Wang F, Tan Y et al (2012) Development of a combined chemical and enzymatic approach for the mass spectrometric identification and quantification of aberrant N-glycosylation. J Proteomics 75:1666–1674
13. Jia W, Lu Z, Fu Y et al (2009) A strategy for precise and large scale identification of core fucosylated glycoproteins. Mol Cell Proteomics 8:913–923
14. Tarentino AL, Plummer TH Jr (1994) Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from Flavobacterium meningosepticum. Methods Enzymol 230:44–57
15. Plummer TH Jr, Phelan AW, Tarentino AL (1996) Porcine fibrinogen glycopeptides: substrates for detecting endo-beta-N-acetylglucosaminidases F2 and F3(1). Anal Biochem 235:98–101
16. Tarentino AL, Quinones G, Plummer TH Jr (1995) Overexpression and purification of non-glycosylated recombinant endo-beta-N-acetylglucosaminidase F3. Glycobiology 5: 599–601
17. Trimble RB, Tarentino AL, Plummer TH Jr et al (1978) Asparaginyl glycopeptides with a low mannose content are hydrolyzed by endo-beta-N-acetylglucosaminidase H. J Biol Chem 253:4508–4511

Part III

Updated Proteome Analysis of Blood Cell Components, Vesicles, and Blood-related Fluids

Chapter 11

Proteomic Analysis of Blood Extracellular Vesicles in Cardiovascular Disease by LC-MS/MS Analysis

Montserrat Baldan-Martin, Fernando de la Cuesta, Gloria Alvarez-Llamas, Gema Ruiz-Hurtado, Luis M. Ruilope, and Maria G. Barderas

Abstract

Extracellular vesicles are membrane vesicles related to cell communication. These vesicles consist of proteins, RNA, and microRNA and are an interesting and important tool to understand the processes taking place in the secreting cell, especially in diseases in which its release is often enhanced. The used of blood extracellular vesicles in cardiovascular disease as a low invasive, easily accessible source of circulating markers could give us important information related to pathological process even more with the use of proteomic analysis. In this chapter, we describe a protocol to isolate and proteomic analyze extracellular vesicles from blood associated with cardiovascular disease.

Key words Extracellular vesicles, Proteomics, Cardiovascular disease, LC-MS/MS, Biomarkers

1 Introduction

Importance of extracellular vesicles (EVs) in cell/cell communication has been highlighted in the recent years and has placed this field into the spotlight of scientific research. EVs include two different types of vesicles according to size and secretion pathway: (1) microvesicles (MVs) or microparticles (MPs), 50 nm to 1 μm vesicles which bud directly from the plasma membrane, and (2) exosomes, 50–120 nm vesicles released by fusion of multivesicular endosomes (MVEs) to the plasma membrane [1]. Circulating cells release EVs to the bloodstream, as well as endothelial cells, due to their constant interaction with the blood flow. Furthermore, correlation of blood levels of EVs from platelet and endothelial origin with increased cardiovascular risk has indeed showed the utility of blood EVs as biomarkers of cardiovascular disease [2].

EVs carry proteins, RNA (mRNA, miRNA, ncRNA, etc.), and even DNA fragments [3]. Cargo from circulating EVs can provide

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_11,

very useful information and offer a very interesting source of biomarkers of the pathological state of the donor cell [4]. Therefore, a biomarker carried by EVs could be easily analyzed by a low invasive liquid biopsy. In this sense, differential proteomic analysis of circulating EVs in cardiovascular patients has proved to offer a wide collection of proteins specifically enriched in EVs from patients with worse diagnosis and thus constitutes a state-of-the-art strategy for analyzing cardiovascular risk [5]. For that purpose, a combination of isobaric labeling, such as isobaric tags for relative and absolute quantitation (iTRAQ), and liquid chromatography coupled to mass spectrometry (LC-MS) seems ideal and offers the possibility of analyzing thousands of proteins from different samples in a single run.

In this chapter, we describe the workflow for isolating EVs from human blood samples, check purity of this fraction, extract its proteome, and perform differential analysis providing with a subset of low invasive biomarkers associated with the cardiovascular disease under study (Fig. 1).

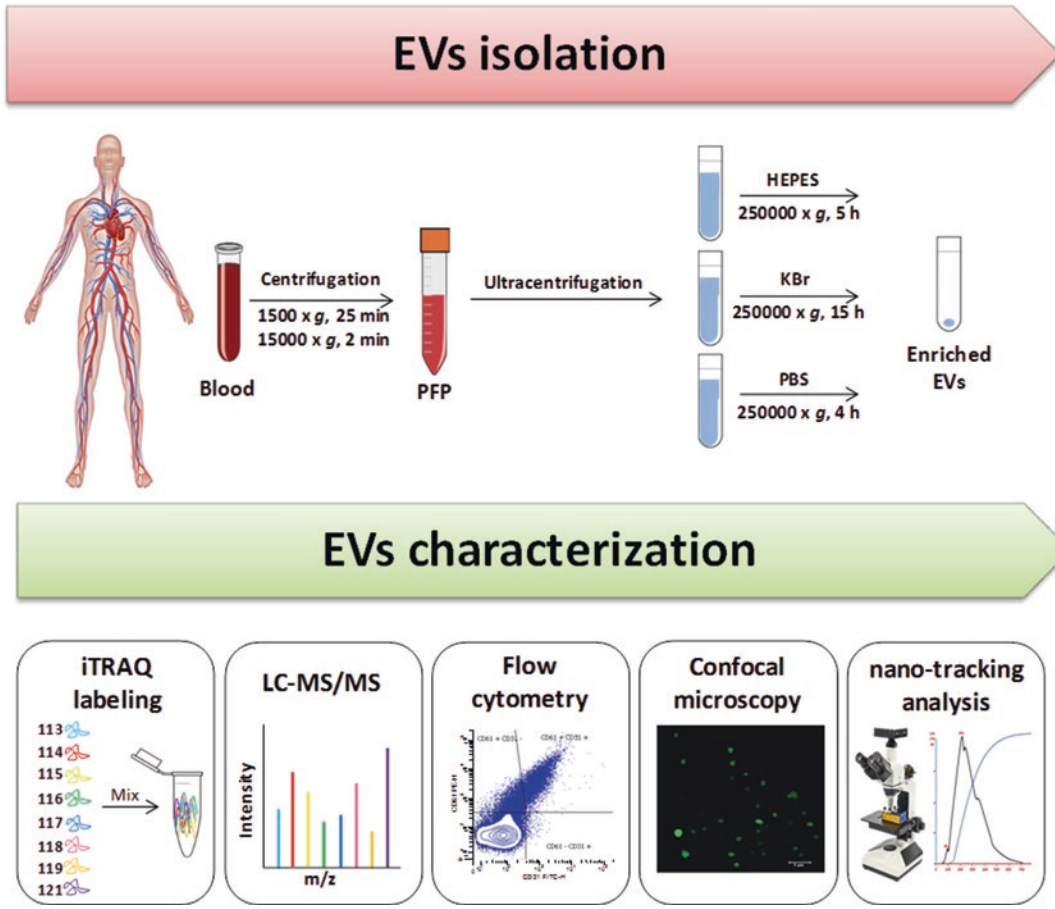


Fig. 1 Analysis of blood extracellular vesicles in cardiovascular disease. The schematic workflow illustrates the isolation of extracellular vesicles from plasma by two centrifugations steps to obtain PFP and then three additional ultracentrifugation steps in different buffers. The characterization of extracellular vesicles was carried out by iTRAQ labeling, LC-MS/MS, flow cytometry, confocal microscopy and nano-tracking analysis (*PFP* platelet free plasma)

2 Materials

2.1 Blood Collection and Extracellular Vesicle Isolation

1. Sodium citrate blood collection tubes.
2. Optima L-100 XP ultracentrifuge (Beckman Coulter).
3. SW41 rotor (Beckman Coulter).
4. 1 M HEPES.
5. 250 mM KBr.
6. Phosphate-buffered saline (PBS) buffer pH 7.2: NaCl (99.5% purity), KCl (99% purity), Na_2HPO_4 (99.5% purity), KH_2PO_4 (99% purity).
7. 14 ml of blood collected from each patient using sodium citrate tubes.

2.2 Flow Cytometry

1. PBS.
2. Citrate.
3. Blocking buffer 2×: 2% BSA, 2% goat normal serum (GNS) in 20 mM PBS-citrate.
4. FcR blocking reagent (Miltenyi Biotec).
5. PE Mouse Anti-Human CD61 (BD Biosciences).
6. Mouse IgG1 K Iso Control PE (eBioscience).
7. Megamix-Plus SSC beads (Biocytex).
8. FACSCANTO II (BD Biosciences).

2.3 Confocal Microscopy

1. PE Mouse Anti-Human CD61 (BD Biosciences).
2. 5% bovine serum albumin (BSA) in double-distilled water.
3. 4% formaldehyde in PBS.
4. Coverslips.
5. Mowiol 4-88 (Sigma-Aldrich).
6. Leica TCS SP5 (Leica Microsystems).

2.4 Nano-tracking Analysis (NTA)

1. 1 ml syringe.
2. PBS.
3. 70% ethanol.
4. NanoSight LM14C (Malvern).
5. Compressed air duster.
6. NTA software (Malvern).

2.5 iTRAQ Labeling

1. Lysis buffer: 7 M urea, thiourea 2 M, CHAPS 4%, SDS 1%, DTT 1%.
2. Bradford protein assay reagent (Bio-Rad).
3. BSA.

4. iTRAQ reagent kit, 8-plex (AB SCIEX).
5. Isopropanol.
6. 0.5 M triethylammonium bicarbonate (TEAB, pH 8.5).
7. Dissolution buffer: 1 M triethylammonium bicarbonate.
8. Milli-Q water.

2.6 Protein Digestion

1. −20 °C acetone.
2. Solubilization buffer: 50 Mm Tris 50, 4% SDS, 50 mM DTT 50, pH 8.
3. SDS-PAGE gel (*see* **Note 1**).
4. Dithiothreitol (DTT) (Bio-Rad).
5. Iodoacetamide (IAA) (Bio-Rad).
6. 60 ng/ml modified trypsin (Promega) at 12:1 protein/trypsin (w/w) ratio.
7. 50 mM ammonium bicarbonate, containing 10% acetonitrile, pH 8.8.

2.7 Liquid Chromatography Coupled to Mass Spectrometry (LC-MS/MS)

1. C-18 reversed-phase nano-column (75 μm ID × 50 cm, 2 μm particle size, Acclaim PepMap RSLC, 100 C18 (Thermo Fisher Scientific).
2. Oasis HLB-MCX columns (*see* **Note 2**).
3. Q Exactive mass spectrometer (Thermo Fisher Scientific).

3 Methods

3.1 Blood Collection and Extracellular Vesicle Isolation

1. Collect 14 ml of blood from each patient using sodium citrate tubes.
2. Obtain platelet-free plasma (PFP) by two serial centrifugation steps: 1500 × *g*, 25 min, and 15,000 × *g*, 2 min at 20 °C.
3. Isolate EVs by ultracentrifugation: ultracentrifuge three times at 250,000 × *g* in a SW41 rotor of an Optima L-100 XP ultracentrifuge (Beckman Coulter) using the following buffers: (1) the first one with HEPES during 5 h, (2) the second ultracentrifugation with KBr during 15 h, and (3) the last one with PBS during 4 h (*see* **Note 3**).
4. Suspend in 500 μl of PBS for flow cytometry, confocal microscopy, and NTA. For proteomic analysis, suspend in lysis buffer.

3.2 Flow Cytometry

1. Aliquot 50 μl of the EVs of each sample and prepare an isotype aliquot of 50 μl with a mixture of all samples as negative control.
2. Add 50 μl of blocking buffer 2× to each tube. Incubate on ice for 20 min.

3. Block each sample with 5 μl of FcR blocking buffer (*see* **Note 4**). Incubate on ice for 10 min.
4. Add 5 μl of PE-conjugated antibody against CD61 and 2 μl of FITC-conjugated antibody against CD31 to each sample. For the isotype control, add 1.25 μl of IgG-PE and 0.5 μl of IgG-FITC (*see* **Note 5**).
5. Incubate for at least 30 min on ice in the dark.
6. Dilute the EVs 1:1 in 20 mM PBS.
7. Vortex and spin.
8. EV gate was defined following MegaMix SSC Plus beads manufacture indications.
9. Analyze by flow cytometry in a FACSCANTO II following manufacture conditions.

3.3 Confocal Microscopy

1. Suspend aliquots of 50 μl from the isolated EVs in 4% formaldehyde.
2. Place in coverslips and air dry.
3. Block samples using 5% BSA for 1 h.
4. Incubate for 16 h with 10 μl of PE-conjugated antibody.
5. Mount coverslips onto slides using Mowiol 4-88.
6. Analyze images in the confocal microscope.

3.4 Nano-tracking Analysis (NTA)

1. Suspend pellet from last ultracentrifugation step in 1 ml of PBS.
2. Clean NanoSight chamber injecting 70% ethanol with a 1 ml syringe and letting the liquid come out through the outlet tube. Open then the chamber and dry it thoroughly with compressed air.
3. Load all sample in a 1 ml syringe without needle, inject trough the inlet tube, and check if the chamber is filling with liquid looking through the glass window. Fill until first droplet comes out of the outlet tube. Let syringe plugged to the system during analysis.
4. Switch on laser beam and focus microscope at the sample (*see* **Note 6**).
5. Take a video of a duration enough as to get at least 700 tracks (*see* **Note 7**).
6. Analyze video with the NTA software. A report will be generated showing the number of EVs/ml, size range, mean and mode size, etc.
7. Repeat 3–5 times after injecting more sample with the syringe, as to have technical replicates.

3.5 EV Lysis, Solubilization, and Digestion of Cargo Proteins

1. Lyse the EVs isolated from ultracentrifugation using 500 μl of lysis buffer and perform two 2 min sonication steps.
2. Centrifuge at 15,000 × *g*, 5 min, and collect the supernatant.
3. Quantify the proteins using Bradford protein assay.
4. Precipitate 150 μg total protein in cold acetone (*see* **Note 8**).
5. Discard supernatant and suspend pellet in solubilization buffer.
6. Perform 12% SDS-PAGE at 25mA/gel and stop the electrophoresis when the front dye barely passes from the stacking gel into the resolving gel to concentrate proteins in one band (*see* **Note 9**).
7. Excise the protein band.
8. Reduce with 10 mM DTT during 15 min and alkylate cysteine residues with 55 mM IAA in 50 mM ammonium bicarbonate during 15 min at room temperature.
9. Digest overnight at 37 °C with 60 ng/ml modified trypsin at 12:1 protein/trypsin (w/w) ratio in 50 mM ammonium bicarbonate, acetonitrile (10%), pH 8.8.

3.6 iTRAQ Labeling

1. Follow iTRAQ labeling manufacturer's protocol (AB SCIEX). In order to maximize labeling efficiency, the volume of the sample digest must be less than 50 μl and the sample concentration must be at 1–5 μg/μl (*see* **Note 10**).
2. Reconstitute dried peptide samples with 30 μl of TEAB.
3. Vortex and sonicate for 2 min to facilitate peptide sample dissolution and spin 10 s.
4. Label each sample using 8-plex iTRAQ Reagents Multiplex Kits (Applied Biosystems) according to manufacturer's protocol (113,114, 115, 116, 117, 118, 119, and 121) to room temperature (*see* **Note 11**).
5. Check the pH of each sample. Make sure it is between 7.8 and 8.5 (*see* **Note 12**). The final concentration of ethanol must be 70% v/v.
6. Incubate the tubes in the dark at room temperature for 1 h.
7. Add 100 μl of Milli-Q water to each tube to quench the iTRAQ reaction. Incubate at room temperature for 30 min.
8. Combine the contents of all iTRAQ reagent-labeled sample tubes into one tube. Vortex and spin.
9. Dry completely the tube containing all the combined iTRAQ mixes in a centrifugal vacuum concentrator; the time will depend on the equipment used. Now, the iTRAQ-labeling peptides from samples, control, and patients are ready for identification and relative quantitation.

3.7 LC-MS/MS Differential Analysis

1. Analyze the iTRAQ-labeled peptides onto a C-18 reversed-phase nano-column in a continuous acetonitrile gradient consisting of 0–30% B in 360 min and 50–90% A in 3 min at a flow rate of 200 nL/min (*see* **Note 13**).
2. Elute from the reversed-phase column to an emitter nanospray needle for real-time ionization and peptide fragmentation in a Q Exactive mass spectrometer.
3. Perform spectra analysis for peptide identification with Proteome Discoverer version 1.4.0.29 using SEQUEST-HT (Thermo Fisher Scientific) using manufacturer conditions. The parameters were selected as follows: trypsin digestion with two maximum missed cleavage sites; precursor and fragment mass tolerances of 2 Da and 0.02 Da, respectively; carbamidomethyl cysteine as fixed modification; and methionine oxidation as variable modifications. For iTRAQ-labeled peptides, N-terminal and Lys iTRAQ modifications were selected as a fixed modification. False discovery rate (FDR) was estimated using inverted databases. Only peptides with at least 95% were used to quantify the relative abundance.
4. Perform protein quantification from reporter ion intensities and statistical analysis with QuiXoT software, based on a statistical model previously described (*see* **Note 14**).
5. Search significant proteins of interest against Vesiclepedia and EVpedia databases to check for enrichment (*see* **Note 15**).

4 Notes

1. Gels for SDS-PAGE can be obtained from commercial vendors or they can be cast in the laboratory following standard procedures [6].
2. This step is optional, just in case an additional fractionation step prior to LC-MS/MS analysis is desirable to increase the number of proteins identified (proteome coverage).
3. The subsequent ultracentrifugation steps are performed to avoid contamination with proteins from plasma (majority proteins). Addition of citrate is important to prevent aggregation of EVs.
4. The use of FcR blocking increases the specificity of immunofluorescent staining with antibodies, since it blocks unwanted binding of antibodies to human Fc receptor-expressing cells.
5. Amount of antibodies and isotype controls per test was reduced four times after optimization in order to avoid background.
6. Choose an area with no background and as much particles as possible. Region closer to laser inlet into the chamber offers normally best performance.

7. Measures with <700 tracks, which means particles tracked, are considered non-reliable. This is calculated after taking the video; therefore, after analyzing a video, one could realize the length was not enough for achieving >700 tracks and may need to increase shooting time.
8. Acetone should be pre-cooled at −20 °C, and precipitation procedure should be preferentially carried out on ice to maximize recovery.
9. The electrophoresis is stopped before separation of protein mixture into discrete bands along the gel takes place. In this way, all proteins concentrate in a unique band, eliminating sample contaminants and facilitating reproducibility for comparisons.
10. In order to maximize labeling efficiency, the volume of the sample digest must be less than 50 μl and the protein concentration must be at 1–5 μg/μl; this peptide concentration range is the optimal for labeling the iTRAQ reagent. If the volume of the sample digest is greater than 50 μl, dry the sample in a centrifugal vacuum concentrator and then reconstitute with 30 μl dissolution buffer.
11. Samples labeled with iTRAQ reagents at room temperature for 1 h previously reconstituted with 70 μl of isopropanol.
12. The pH must be between 7.8 and 8.5, because the iTRAQ reaction occurs at basic pH. If your sample is at pH below the indicated, then add up to 5 μl of 0.5 M TEAB to increase the pH at or above 7.8.
13. For increasing proteome coverage, samples can be fractionated by cation exchange chromatography (Oasis HLB-MCX columns) into, e.g., six fractions and analyzed by LC-MS/MS using the same system and conditions described before.
14. In this model protein log2 ratios are expressed in the form of the standardized variables, i.e., in units of standard deviation according to their estimated variances (Zq values). Cutoff for signification was set at Zq = ±1.5 [7, 8].
15. EVpedia [9] and Vesiclepedia [10] are an integrated and comprehensive compendium of molecular data identified in different classes of EVs.

Acknowledgments and Sources of Funding

This work was supported by grants from the Instituto de Salud Carlos III (FIS PI070537, IF08/3667-1, PI11-02239, PI14/01917, PI11/01401, PI11/02432, PI13/01873, PI13/01746, PI13/01581, CP15/00129, PI14/01650, PI14/01841) by PT13/0001/0013,

IDCSalud (3371/002), Fundación Conchita Rábago de Jiménez Díaz, Fundacion Senefro, and Redes Temáticas de Investigación Cooperativa (FONDOS FEDER, RD06/0014/1015, RD12/0042/0071).

References

1. Xu R, Greening DW, Zhu HJ, Takahashi N, Simpson RJ (2016) Extracellular vesicle isolation and characterization: toward clinical application. J Clin Invest 126(4):1152–1162
2. Lacroix R, Dubois C, Leroyer AS, Sabatier F, Dignat-George F (2013) Revisited role of microparticles in arterial and venous thrombosis. J Thromb Haemost 11(Suppl 1):24–35
3. Desrochers LM, Antonyak MA, Cerione RA (2016) Extracellular vesicles: satellites of information transfer in cancer and stem cell biology. Dev Cell 37(4):301–309
4. Raposo G, Stoorvogel W (2013) Extracellular vesicles: exosomes, microvesicles, and friends. J Cell Biol 200(4):373–383
5. Vélez P, Parguiña AF, Ocaranza-Sánchez R, Grigorian-Shamagian L, Rosa I, Alonso-Orgaz S, de la Cuesta F, Guitián E, Moreu J, Barderas MG, González-Juanatey JR, García Á (2014) Identification of a circulating microvesicle protein network involved in ST-elevation myocardial infarction. Thromb Haemost 112(4):716–726
6. Martín-Rojas T, Gil-Dones F, Lopez-Almodovar LF, Padial LR, Vivanco F, Barderas MG (2012) Proteomic profile of human aortic stenosis: insights into the degenerative process. J Proteome Res 11(3):1537–1550
7. García-Marqués F, Trevisan-Herraz M, Martínez-Martínez S, Camafeita E, Jorge I, Lopez JA et al (2016) A novel systems-biology algorithm for the analysis of coordinated protein responses using quantitative proteomics. Mol Cell Proteomics 15(5):1740–1760
8. Navarro P, Trevisan-Herraz M, Bonzon-Kulichenko E et al (2014) General statistical framework for quantitative proteomics by stable isotope labeling. J Proteome Res 13:1234–1247
9. Kim DK, Lee J, Kim SR, Choi DS, Yoon YJ, Kim JH et al (2015) EVpedia: a community web portal for extracellular vesicles research. Bioinformatics 31:933–939
10. Kalra H, Simpson RJ, Ji H, Aikawa E, Altevogt P, Askenase P et al (2012) Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation. PLoS Biol 12:e1001450

Chapter 12

Targeted Approach for Proteomic Analysis of a Hidden Membrane Protein

Tania Martins-Marques, Sandra I. Anjo, Teresa Ribeiro-Rodrigues, Bruno Manadas, and Henrique Girao

Abstract

Given the properties of plasma membrane proteins, namely, their hydrophobicity, low solubility, and high resistance to digestion and extraction, their identification by traditional mass spectrometry (MS) has been a challenging task. Hence, proteomic studies involving the transmembrane protein connexin43 (Cx43) are scarce. Additionally, studies demonstrating the presence of proteins embedded in the lipid bilayer of extracellular vesicles (EVs) are difficult to perform and require specific changes and fine adjustments in the experimental and technical procedure to allow their detection by MS. In this review, we provide a detailed description of the protocol we have used to detect Cx43 in EVs of human peripheral blood. This includes some of the modifications that we have introduced in order to improve the detection of Cx43 in EVs, including an optimization of vesicle isolation, Cx43 purification, MS acquisition data, and further analysis.

Key words SWATH-MS, Proteomics, Extracellular vesicles, Cx43, Serum

1 Introduction

Although several proteomic analyses have been performed to investigate the protein content of extracellular vesicles (EVs), the gap junction protein connexin43 (Cx43) had never been identified before by mass spectrometry (MS) (*see* refs. 1, 2). In order to assess the presence of Cx43 in EVs, we first had to optimize the detection parameters, using an enriched sample of immunopurified Cx43 from cellular extracts. Once the conditions were optimized, we performed a targeted analysis in EVs secreted by mammalian cells in culture, before we proceed to the evaluation of Cx43 in circulating EVs.

Untargeted proteomics has a central role in qualitative characterization of the protein content of a sample, due to its capacity to identify a large number of proteins. The success of shotgun proteomics is mainly linked to the combination of the separation power of liquid chromatography (LC) coupled to tandem MS

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_12,

(LC-MS/MS) acquisition (*see* ref. 3). Protein identification is achieved through the identification of the peptides formed by enzymatic digestion of the samples. For this purpose, MS instruments operate in an information-dependent acquisition (IDA) mode, where the fragment ion spectra (specific signatures of the peptides) are acquired for selected precursor ions (intact peptides). Since only a limited number of precursor ions can be selected according to their relative intensity, this method largely depends on sample complexity, being difficult to identify low abundant proteins or proteins difficult to digest, such as membrane proteins (*see* refs. 4–6). This can be overcome by separation and fractionation techniques, which result in an improvement of efficiency and sensitivity of MS-based proteomics.

Conversely, MS acquisition methods for targeted analysis can be an alternative to the usual IDA method used for protein detection. Although multiple reaction monitoring (MRM) acquisition mode constitutes the gold standard for targeted analysis, it requires a laborious process of development and validation (*see* refs. 7–10). Therefore, data-independent acquisition (DIA) methods, as the sequential windowed data-independent acquisition of the total high-resolution mass spectra (SWATH-MS), where fragmentation data are acquired by repeatedly cycling through sequential isolation windows over the whole chromatographic elution range, can be used as alternatives to MRM. This analysis generates a complete recording of the fragment ion spectra of all peptides detectable in a biological sample, in which the precursor ion signals are within a user-defined m/z vs. retention time (t_R) window (*see* refs. 3, 11, 12). Hence, the SWATH-MS windows can be adapted to improve the selectivity of the method, by adjusting the precursor windows to particular targets. Moreover, SWATH-MS can be easily combined with the information obtained from protein identification by shotgun proteomics to interpret the data and reduce the time required for analysis (*see* ref. 3).

2 Materials

2.1 Serum Isolation

1. Plastic serum tubes, with clot activators (e.g., BD Vacutainer® SST II Plus, BD Biosciences).
2. Refrigerated centrifuge.

2.2 EV Isolation from Serum Samples by Differential Ultracentrifugation

1. Human serum (or plasma, *see* **Note 1**).
2. Phosphate-buffered saline (PBS): 137 mM NaCl, 2.7 mM KCl, 1.5 mM KH_2PO_4, 8.1 mM Na_2HPO_4, pH 7.4.
3. Refrigerated centrifuge.
4. Ultracentrifuge and fixed-angle or swinging-bucket rotor.
5. Appropriate polypropylene centrifuge tubes and microcentrifuge tubes.
6. 0.22-μm filters and syringes.

2.3 EV Isolation from Serum Samples Using Commercially Available Kits

1. Human serum (or plasma, *see* **Note 1**).
2. Total Exosome Isolation Reagent from serum (Life Technologies).
3. Refrigerated centrifuge.
4. Microcentrifuge tubes.

2.4 Nanoparticle Tracking Analysis (NTA)

1. Isolated EVs.
2. PBS (*see* **Note 2**).
3. 1 mL syringes (slip tip).
4. Malvern NanoSight NS300 (sCMOS camera; 405 nm laser) (NanoSight Ltd., Amesbury, United Kingdom) and analysis software (NTA software 2.3.5).

2.5 Transmission Electron Microscopy (TEM)

1. Isolated EVs.
2. Formvar-carbon-coated EM grids (e.g., TAAB Laboratories).
3. Forceps.
4. PBS.
5. 1% Glutaraldehyde.
6. Uranyl oxalate, pH 7.
7. Methylcellulose-uranyl acetate, pH 4: 9:1 solution of 2% methylcellulose to 4% uranyl acetate, mixed just before use.
8. Transmission electron microscope (e.g., Tecnai G2 Spirit BioTWIN electron microscope (FEI, Hillsboro, OR, USA)).

2.6 Immunoblot

1. Isolated EVs.
2. Protein lysis buffer (e.g., RIPA buffer: 150 mM NaCl, 50 mM Tris–HCl, pH 7.4, 1% NP-40, 0.1% SDS pH 7.4), supplemented with protease inhibitors, just before use (protease inhibitor cocktail (Roche), 2 mM phenylmethanesulfonyl fluoride (PMSF), 10 mM iodoacetamide, and 2 mM sodium orthovanadate).
3. Probe sonicator.
4. Refrigerated centrifuge.
5. Protein concentration assay (e.g., BCA protein assay).
6. Sample buffer, reducing or non reducing (i.e., with or without dithiothreitol (DTT) or β-mercaptoethanol).
7. Reagents and equipment for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblot.
8. Digital imaging system (e.g., VersaDoc, Bio-Rad, Hercules, CA, USA) and analysis software (e.g., Image J, National Institutes of Health, NIH).

2.7 Immunopurification of Cx43 for Method Development

1. HEK-293 cells overexpressing V5-tagged Cx43 (or similar) (*see* **Note 3**).
2. PBS.
3. RIPA buffer.
4. Cell scrapers and microcentrifuge tubes.
5. Refrigerated centrifuge.
6. Primary antibodies against Cx43 (e.g., AB0016, Sicgen) and non-specific antibodies (e.g., anti-GFP; AB0020, Sicgen).
7. Sepharose beads (e.g., protein G-Sepharose 4 Fast Flow).
8. Laemmli buffer (4×): 62.5 mM Tris-HCl, pH 6.8, 4% SDS, 10% (w/v) glycerol, 0.005% bromophenol blue.

2.8 Protein Separation by Short-GeLC or Complete SDS-PAGE and Colloidal Coomassie Protein Staining

1. Isolated EVs and/or immunopurified Cx43.
2. Acrylamide.
3. Precast gels with polyacrylamide gradient from 4% to 20%.
4. Reagents and equipment for SDS-PAGE.
5. Laemmli buffer.
6. Fixation solution: 10% (v/v) of 85% orthophosphoric acid, 10% (w/v) ammonium sulfate, and 20% (v/v) methanol.
7. Coomassie Brilliant Blue G-250.
8. Orbital shaker.

2.9 In-Gel Digestion and Peptide Extraction

1. Stained gel.
2. 20% (w/v) SDS.
3. Acetate sheet.
4. Laminar flow hood.
5. Scalpel blade.
6. Microcentrifuge tubes and low-binding microcentrifuge tubes.
7. Ammonium bicarbonate.
8. Acetonitrile (ACN).
9. Destaining solution: 50 mM ammonium bicarbonate and 30% ACN.
10. Concentrator (vacuum centrifuge).
11. 0.01 μg/μL trypsin (in 10 mM ammonium bicarbonate).
12. Formic acid (FA).
13. LC grade water.
14. Solutions for peptide extraction (in LC grade water)—solution A, 30% ACN/1% FA; solution B, 50% ACN/1% FA; solution C, 98% ACN/1% FA.

2.10 Peptide Cleanup by C18 Solid-Phase Extraction

1. Evaporated peptide mixture.
2. Benchtop centrifuge and concentrator.
3. Sonicator with cup horn (e.g., 750 W sonicator).
4. SPE tips with C18 matrix (e.g., OMIX tip C18 100 µL, Agilent, Santa Clara, CA, USA) and Combitips (with precut end).
5. Low-binding microcentrifuge tubes.
6. ACN solutions (50% ACN; 2% ACN/1% FA; 70% ACN/0.1% FA).

2.11 LC-MS/MS Data Acquisition in IDA and Semi-Targeted SWATH-MS

1. Purified peptides.
2. iRT peptides (iRT Kit, Biognosys, Schlieren, Switzerland).
3. Mobile phases for sample preparation: 2% ACN in 0.1% FA.
4. Mobile phases for high-performance liquid chromatography (HPLC): A (0.1% FA in water) and B (0.1% FA in ACN).
5. Benchtop centrifuge.
6. Sonicator with cup horn (e.g., 750 W sonicator).
7. Vials and precut snap rings.
8. ChromXP™ C18AR reversed phase column (300 µm ID × 15 cm length, 3 µm particles, 120 Å pore size, Eksigent).
9. Linear ACN gradient: 2–35% ACN in 0.1% FA.
10. HPLC system: nanoLC Ultra 2D (Sciex).
11. Mass spectrometer: TripleTOF™ 5600 system operated by Analyst® TF 1.7 with an electrospray ionization source (Sciex).

2.12 Data Processing

1. ProteinPilot™ software or similar for protein database search (Sciex).
2. MASCOT search engine (Matrix Science; www.matrixscience.com/cgi/search_form.pl?FORMVER=2&SEARCH=MIS).
3. PEAKS Studio v4.5, SP2 (Bioinformatics Solutions Inc.).
4. PeakView™ v2.0.01 with SWATH™ processing plug-in (Sciex).

3 Methods

3.1 Serum Isolation

In this section, we describe in detail the procedure for isolation of EVs from human serum samples. Several methods have been described for the purification of EVs from different biological fluids; however we will only be focused on differential ultracentrifugation and in the use of commercially available EV isolation kits (*see* **Note 4**). The methods described below could also be applied in EV isolation from human plasma or cultured cells, with appropriate adaptations (*see* **Notes 1** and **5**) (*see* refs. 1, 13).

1. Collect the blood sample (*see* **Note 6**).
2. Leave the tube in a standing position for about 20–30 min, enabling the blood to clot.
3. Centrifuge at 1,000 × *g*, for 10 min.
4. Remove the serum (supernatant), and store at 4 °C (max. 1 week, *see* **Note 7**), prior to the isolation of EVs, following one of the methods described in Subheadings 3.2 or 3.3 (*see* **Note 8**).

3.2 EV Isolation from Serum Samples by Differential Ultracentrifugation

1. Dilute serum samples with an equal volume of PBS and centrifuge at 2,000 × *g*, for 30 min.
2. Ultracentrifuge the supernatant at 12,000 × *g*, for 45 min.
3. Carefully pour the supernatant and ultracentrifuge at 110,000 × *g*, for 2 h.
4. Resuspend the pellet in 1 mL PBS, and transfer it into a clean tube. Dilute the specimen in PBS and filter through a 0.22-μm filter, to remove large vesicles.
5. Ultracentrifuge at 110,000 × *g*, for 70 min.
6. Repeat **step 5**.
7. Resuspend the EV pellet in an appropriate buffer (*see* **Note 9**).

3.3 EV Isolation from Serum Samples Using Commercially Available Kits

1. Centrifuge the serum sample at 2,000 × *g*, for 30 min, to remove cells and cellular debris.
2. Transfer the supernatant to a new tube, without disturbing the pellet.
3. Mix the serum:reagent (5:1), by vortexing until there is a homogeneous solution. Note that the solution should have a cloudy appearance.
4. Incubate for 30 min, at 2–8 °C.
5. Centrifuge at 10,000 × *g*, for 10 min, at room temperature.
6. Resuspend the EV pellet in an appropriate buffer (*see* **Note 9**).

3.4 Characterization of Serum EVs

In this section, we describe some of the methods usually performed to characterize EV-enriched samples, which can be used as a complement of large-scale proteomic analyses. NTA gives an accurate idea of the vesicle size distribution of the sample; however it can also detect protein aggregates (*see* refs. 14, 15). TEM allows to assess not only the size of the vesicles but also the cup-shaped characteristic morphology of EVs. Western blotting permits the detection of proteins present in EVs, including the protein of interest and others commonly used as markers of EVs (*see* ref. 13).

3.4.1 NTA

1. Set up the equipment according to the manufacturer's instructions (*see* **Note 10**).
2. Dilute the EV pellet in 1 mL of particle-free PBS, at a concentration between 2 and 8 × 10^8 vesicles/mL (*see* **Notes 6** and **11**).

3. Transfer the EV suspension to a clean 1 mL syringe, load it onto the instrument, and pump the sample until it reaches the reading chamber.
4. Set and adjust the camera levels, and focus and center the image position, in order to achieve a clear view of the particles, with under 30% particles being colored (indicates saturated pixels) (*see* **Note 12**).
5. Using the SOP function, collect a minimum of five repeated measures. If a syringe pump is available, measurements can be performed under low flow (typically 10–20). If not, advance the sample between each repeat (*see* **Note 13**).
6. For analysis, select an appropriate "Detect Threshold" setting to process the captured videos. Each analysis gives the mean, mode, and median size vesicle, as well as particle concentration (*see* **Note 14**) (Fig. 1a).

3.4.2 TEM

1. Place a 5 μL drop of fixed EV suspension on Formvar-carbon-coated EM grids (set 2–3 grids for EV suspension). Let the membranes adsorb for 20 min in a dry environment. Alternatively, place a drop on clean *Parafilm* and, with forceps, gently position a Formvar-carbon-coated nickel grid on top of each drop for 30–60 min. Assure that the grid is positioned with the coating side facing the drop containing EVs.
2. Wash the grid by sequentially transferring it to the top of 100 μL drops of PBS placed on clean *Parafilm*. Between washes, dry the excess PBS with absorbing paper, holding it

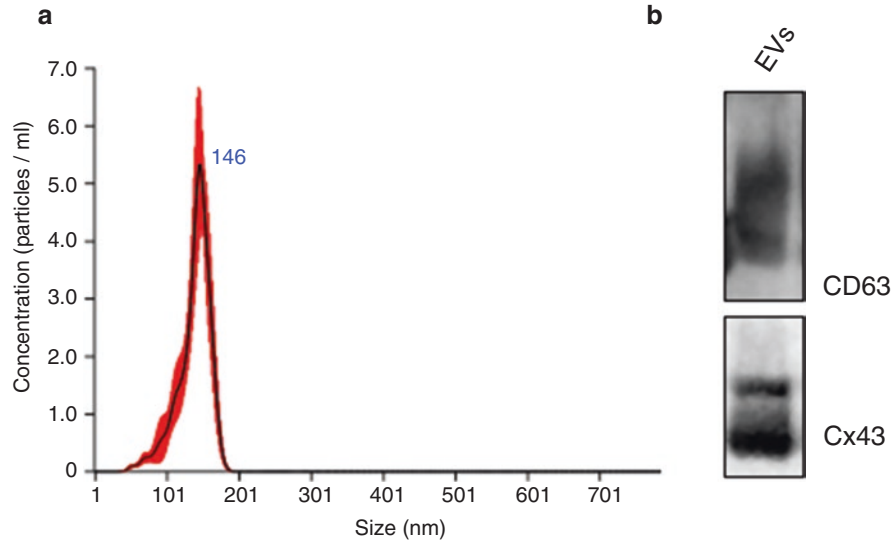


Fig. 1 Characterization of serum EVs. (**a**) NTA analysis of human serum EVs concentration and size distribution. Size mode: 146 nm. Graph depicts average concentration/size (*black line*) ± standard deviation (SD) (*red bars*) from five technical replicates. (**b**) Characterization of serum EVs was performed by immunoblot under non reducing conditions. The presence of Cx43 and the EV marker CD63 were analyzed

closely to the side of the grid, avoiding contact with the coated area (*see* **Note 15**).

3. Place the grid in a 50 μL drop of 1% glutaraldehyde for 5 min.
4. Transfer the grid to a 100 μL drop of ddH_2O and let it stand for 2 min. Repeat 7×, for a total of eight washes.
5. Contrast the sample by transferring the grid to a 50 μL drop of uranyl oxalate solution for 5 min.
6. Embed the sample by placing the grid in a 50 μL drop of methylcellulose-uranyl acetate for 10 min on ice (drops can be placed in a glass dish covered with *Parafilm* on ice).
7. Remove excess liquid by gently using a filter paper and positioning the grid on a paper with the coated side up. Let it air-dry for 5–10 min.
8. Observe the preparations with an electron microscope or store the grids in an appropriate grid box.

3.4.3 Immunoblot

1. Sonicate lysed EVs (*see* **Note 9c**) with two 10 s pulses (30 s in between pulses) using a probe sonicator (keep the samples in an ice bath and keep the probe away from the sample-air interface to minimize foaming).
2. Centrifuge the sample at 13,000 × *g*, for 5 min at 4 °C, and transfer the supernatant to a new microcentrifuge tube.
3. Measure the total amount of protein, using a commercially available assay. Add denaturing sample buffer, either reducing or non reducing, as appropriate, and load 20–50 μg of protein well (*see* **Notes 6** and **16**).
4. Separate and transfer the proteins by gel electrophoresis and electroblotting.
5. Block and wash the membrane before probing with antibodies against EV proteins (*see* **Notes 17** and **18**).
6. Detect the specific protein by chemiluminescence, using a digital imaging system and analysis software (Fig. 1b).

3.5 Immunopurification of Cx43 for Method Development

All the indicated steps should be performed at 4 °C, using ice-cold buffer solutions.

1. Place the cell culture plates on ice, remove cell culture medium gently, and wash the cells with ice-cold PBS (3×).
2. Remove the PBS and add RIPA buffer (1 mL/100 mm^2 plate).
3. Scrape the cells using a plastic cell scraper, and transfer the cell suspension into a microcentrifuge tube. Leave the cells on ice for 30 min.
4. Centrifuge at 1,000 × *g*, for 10 min, to remove dead cells and cellular debris (*see* **Note 19**).

5. Discard the cell pellet, and incubate the supernatants with 30 μg goat polyclonal antibodies directed against Cx43, overnight, with gentle agitation. Non-specific antibodies should be used as technical control (control IP).
6. Add 120 μg protein G-Sepharose to the samples, and incubate for 1.5 h.
7. Centrifuge the protein G-Sepharose sediments, and wash 3× with 1 mL RIPA buffer.
8. Elute the immunopurified Cx43 with 1× Laemmli buffer and denature the samples for 5 min, at 95 °C (*see* **Notes 20** and **21**).

3.6 Protein Separation by Short-GeLC or Complete SDS-PAGE and Colloidal Coomassie Protein Staining

Gel digestion is advantageous for the analysis of complex protein samples, including membrane-enriched samples, such as EVs. Preparatory electrophoresis allows the use of more stringent buffers (such as higher content of detergents), important to solubilize membrane proteins and to remove contaminants that may interfere with the digestion.

1. Denature the samples (prepared in Subheadings 3.2, 3.3, or 3.5; *see* **Note 9d**) as indicated in **step 8** from Subheading 3.5. Allow samples to reach room temperature, and add acrylamide to a final concentration of 1% (v/v), to alkylate the reduced cysteines.
2. Load the samples in a precast polyacrylamide gel. Leave one empty lane between samples, to avoid cross contamination, and fill empty lanes with an equal volume of 1× Laemmli buffer (*see* **Notes 22** and **23**).
3. Run short-GeLC at 110 V (constant voltage), for 15 min, or SDS-PAGE at 150–200 V (constant voltage) until the tracking dye reaches the bottom of the gel.
4. Place the gel into a container with ddH_2O.
5. Remove the gel from the water, add the fixation solution, and keep under low-speed agitation.
6. Add 100 mg Coomassie to the solution using a strainer, to prevent the formation of clusters (*see* ref. 14), and incubate with agitation for 1–2 h until visualize the protein staining.
7. Discard the solution in an appropriate disposal recipient, and transfer the gel to a new box with ddH_2O. Incubate under agitation using an orbital shaker at low speed; change the ddH_2O until the gel background is clear.
8. Store in ddH_2O at 4 °C (for long-term storage, add NaN_3 at a final concentration of 0.1%) (*see* **Note 24**).

3.7 In-Gel Digestion and Peptide Extraction

1. Wash gloves and an acetate sheet with a 20% SDS solution.
2. Transfer the gel to the acetate sheet, in a laminar flow hood (*see* **Note 25**).

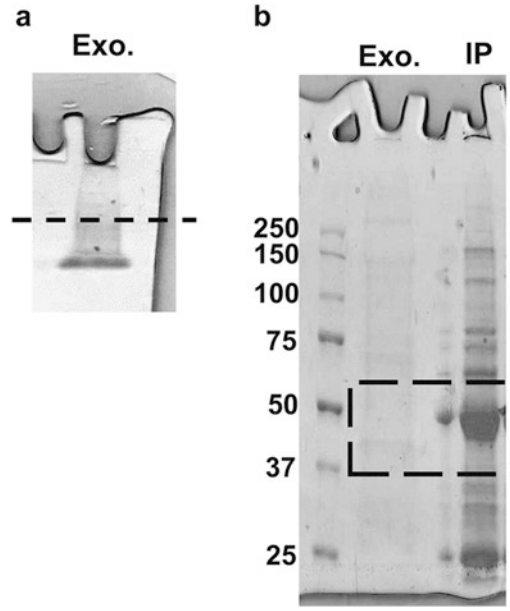


Fig. 2 Sample preparation for untargeted and targeted analysis of EV proteome. (**a**) Short-GeLC approach used in untargeted analysis of EVs (EVs). A partial electrophoretic separation of the isolated EVs was visualized after colloidal Coomassie Blue staining. The lane was sliced (as indicated by the *black dashed line*) for independent processing and protein identification by IDA experiments. (**b**) Complete SDS-PAGE for protein separation and isolation of Cx43. For targeted analysis, only the region containing the protein of interest was sliced and analyzed (between 37 and 60 kDa, *dashed square*). A positive control (immunopurified Cx43, Cx43-IP) and the sample of interest were processed in parallel and analyzed by a semi-targeted SWATH-MS method

3. For targeted experiments, cut the band or region of interest (for Cx43, cut from 36 to 60 kDa) with a scalpel blade (*see* Fig. 2b).
4. Cut the band/region into small pieces and transfer to a microcentrifuge tube with 1 mL of ddH_2O.
5. To destain the gel pieces, remove the water from the tube, add 1 mL of destaining solution, and agitate for 15 min, at 25 °C (~850 rpm); discard the supernatant, and repeat this process until all the stain has been removed.
6. Add 1 mL of water and shake for 10 min, at 25 °C (~850 rpm).
7. Dry the gel pieces using a concentrator.
8. Add trypsin (*see* Material list, Subheading 2.9) until all the pieces are covered with the solution (use 30 μL/band to 100 μL for a larger region of the gel).

9. Incubate for 10–15 min at 4 °C, until the gel pieces are rehydrated. Add ammonium bicarbonate solution to cover the gel pieces, and incubate overnight at room temperature.
10. Remove excess solution from gel pieces (containing trypsin and some peptides), and transfer to a low-binding microcentrifuge tube.
11. Add 40 μL of solution A (*see* Materials list, Subheading 2.7), and keep under agitation (~1050 rpm), for 15 min at 25 °C.
12. Transfer the solution with peptides to the low-binding microcentrifuge tube from **step 10**.
13. Repeat the **steps 11** and **12** using the solutions B and C (*see* Materials list, Subheading 2.9).
14. Evaporate the extracted peptides, almost completely, using a concentrator.
15. Proceed to C18 cleanup protocol (*see* ref. 4).

3.8 Peptide Cleanup by C18 Solid-Phase Extraction

1. Add 100 μL of 2% ACN/1% FA to the evaporated peptide mixture (*see* **Note 26**), and sonicate for 2 min (*on/off* pulses with low amplitude).
2. Add 2× 100 μL of 50% ACN, on the top of the C18 tip.
3. Push through the sample with the help of a precut Combitip, and discard the flow through.
4. Equilibrate the tip matrix with 3× 100 μL of 2% ACN in 1% FA; discard the flow through (*see* **step 3**).
5. Apply the sample on top of the tip matrix, and transfer to the microcentrifuge tube from **step 4**. Repeat this step, 4×.
6. Wash the tip by adding 100 μL of 2% ACN/1% FA (*see* **step 5**).
7. Elute peptides with 4× 100 μL of 70% ACN/0.1% FA.
8. Collect the eluted peptides in a new low-binding microcentrifuge tube, and evaporate the samples.

3.9 LC-MS/MS Data Acquisition in IDA and Semi-Targeted SWATH-MS

3.9.1 Sample Preparation

1. Spike cleaned samples with iRT peptides, used as internal standards to account for sample losses and/or RT alignment (*see* **Note 27**).
2. Resuspend samples in 30 μL of mobile phase (*see* Materials list, Subheading 2.11). Vortex, spin and sonicate for 5 min (*on/off* pulses at 20% intensity).
3. Centrifuge for 5 min, at 14,000 × *g*, to remove insoluble material.
4. Transfer the collected sample to a proper vial for LC-MS/MS.

3.9.2 LC Method

1. Inject 5–10 μL of the sample (of 30 μL samples), depending on its concentration.
2. Resolve the peptide mixture on a C18 reversed phase column at 5 μL/min.
3. Elute peptides into the mass spectrometer with a 25 min linear ACN gradient (*see* Materials list, Subheading 2.11; *see* **Notes 28** and **29**).

3.9.3 IDA Method for Identification and Development of the Semi-Targeted Method

1. Set the mass spectrometer with the proper parameters.
2. Scan full spectra from 350 to 1250 *m/z*, for 250 ms. Scan up to 20 MS/MS spectra from 100 to 1500 *m/z*, for 75 ms accumulation time each.
3. For fragmentation, isolate the candidate ions that have (1) a charge state between +2 and +5 and (2) counts above a minimum threshold of 70 counts/s. Exclude the candidate ions for 15 s after one MS/MS spectra is collected. Use rolling collision energy spread of 5 eV (*see* **Note 30**).

3.9.4 Semi-Targeted SWATH-MS Acquisition Method

For SWATH-MS-based experiments, the mass spectrometer is operated in a looped product ion mode and specifically tuned to allow quadrupole resolution of a specific mass selection. By using an isolation width plus 1 Da and by containing 1 *m/z* of overlap, a complete transmission is achieved (*see* ref. 3).

1. Scan full spectra from 350 to 1250 *m/z*, for 250 ms accumulation time.
2. Use a defined number of overlapping windows, covering a representative number of peptides of the protein of interest, previously identified. For detection of EV-Cx43, a set of 12 overlapping windows of 10 Da width was used to cover Cx43 and iRT peptides (Table 1).

Table 1
SWATH windows used for the semi-targeted analysis (adapted from [1])

	Peptide sequence	Peptide *m/z*	t_R	SWATH window (*m/z*)
Cx43	YGIEEHGK	466.7256	6.2	464–474
	VQAYSTAGGK	491.2515	5.83	483–494
	SDPYHATTGPLSPSK	519.9237	10.98	513–524
	TYIISILFK	549.3325	24.7	543–554
iRT peptides	TGFIIDPGGVIR	622.8525	21.0609	613–624
	TPVISGGPYYER	669.836	14.4352	663–674
	GDLDAASYYAPVR	699.3371	16.1109	693–704

3. Scan MS/MS spectra from 100 to 1500 *m/z*, for 250 ms accumulation time each. Use rolling collision energy spread of 15 eV, with a collision energy for each window determined for a charge +2 ion centered upon the window.

3.10 Data Processing

3.10.1 Protein Identification for Untargeted Analysis of EV Proteome

1. Perform peptide identification by searching the IDA files of the EV sample (Fig. 2a, from short-GeLC) in ProteinPilot™ using the following parameters:
   (1) Protein database, canonical UniProtKB/Swiss-Prot complete proteome database with the sequences of the iRT peptides (database must be in fast format); (2) alkylating agent, acrylamide; (3) enzyme, trypsin; (4) special factors, gel-based ID; and (5) indication of the equipment used to acquire the data.
2. Perform a false discovery rate (FDR) analysis using the target-decoy approach (on ProteinPilot™), which will assess the quality of the identifications. Positive identifications should be considered when both proteins and peptides identified reach a 5% local FDR confidence (*see* refs. 15, 16).

3.10.2 Identification of Cx43

To a more comprehensive identification, different search engines, such as ProteinPilot™, PEAKS Studio, or MASCOT, should be used.

1. To perform the searches in ProteinPilot™ (*see* **step 1** of Subheading 3.10.1).
2. To perform the searches in PEAKS Studio:
   (1) Protein database, canonical UniProtKB/Swiss-Prot complete proteome database; (2) acrylamide adduct in cysteines as fixed modification and methionine oxidation as variable modification; and (3) 20 ppm mass tolerance of precursor ions and 0.1 Da mass tolerance for fragment ions.
3. To perform the searches in MASCOT search engine:
   (1) Protein database, canonical UniProtKB/Swiss-Prot complete proteome database; (2) enzyme, trypsin with one missed cleavage; (3) propionamide as fixed modification and methionine oxidation as variable modification; and (4) 20 ppm mass tolerance of precursor ions and 0.6 Da mass tolerance for fragment ions.
4. To consider as positive identification proteins that have more than one peptide hit with individual score above 95% of confidence or based on a single peptide hit with a minimum individual score of 95% and a minimum sequence tag of three amino acids (four consecutive peaks in the MS/MS spectrum) (Fig. 3a).

3.10.3 Setup of the Semi-Targeted SWATH-MS Method

1. Build a list with the precursors *m/z* of the peptides identified for Cx43 and iRT peptides used to RT alignment.

2. Generate the SWATH windows (10 Da width) for each peptide by summing and subtracting 5 Da to the *m/z* of a particular peptide (*see* **Notes 31** and **32**).

3.10.4 Generation of a Specific Library of Precursor and Fragment Ions for Cx43

Create the library of precursors and fragment ions for Cx43 and iRT peptides, combining the peptides identified in all search engines (*see* Subheading 3.10.2).

1. Select the peptides identified in ProteinPilot™, uploading the group file obtained in the database search into the SWATH™ processing plug-in for PeakView™. Save library as tab-delimited text file (Fig. 3b). For selection of the proper peptides,

a

MGDWSALGKLLDKVQAYSTAGGKVWLSVLFIFRILLLGTAVESAWGDEQSAFRCNTQQPG
CENVCYDKSFPISHVRFWVLQIIFVSVPTLLYLAHVFYVMRKEEKLNKKEEELKVAQTDG
VNVEMHLKQIEIKKFKYGIEEHGKVKMRGGLLR**TYIISILFK**SVFEVAFLLIQWYIYGFS
LSAVYTCKRDPCPHQVDCFLSRPTEKTIFIIFMLVVSLVSLALNIIELFYVFFKGVKDRV
KGR**SDPYHATTGPLSPSK**DCGSPKYAYFNGCSSPTAPLSPMSPPGYKLVTGDRNNSSCRN
YNK**QASEQNWANYSAEQNRMGQAGSTISNSHAQPFDFPDDNQNAK**KVAAGHELQPLAIVD
QRPSSRASSRASSRPRPDDLEI

b

Q1	Q3	t_R detected	protein name	Stripped sequence	prec_z	frg_type	frg_z	frg_nr	decoy	confidence	shared	N
519.923	678.3513	11.05	Cx43	SDPYHATTGPLSPSK	3	y	2	13	FALSE	0.98	FALSE	1
519.923	418.2296	11.05	Cx43	SDPYHATTGPLSPSK	3	y	1	4	FALSE	0.98	FALSE	1
519.923	203.0663	11.05	Cx43	SDPYHATTGPLSPSK	3	b	1	2	FALSE	0.98	FALSE	1
519.923	548.2933	11.05	Cx43	SDPYHATTGPLSPSK	3	y	2	11	FALSE	0.98	FALSE	1
519.923	887.4833	11.05	Cx43	SDPYHATTGPLSPSK	3	y	1	9	FALSE	0.98	FALSE	1
623.225	598.368	21.06	iRT	TGFIIDPGGVIR		y	1	6	FALSE	0.95	FALSE	2
623.225	713.395	21.06	iRT	TGFIIDPGGVIR		y	1	7	FALSE	0.95	FALSE	2
623.225	826.479	21.06	iRT	TGFIIDPGGVIR		y	1	8	FALSE	0.95	FALSE	2

Fig. 3 Sequence coverage of the targeted proteins. (**a**) Indication of the peptides identified (*highlighted regions*) in the sequence of Cx43 and used for the semi-targeted SWATH-MS method. The peptides identified cover 17.3% of the sequence. (**b**) Library of precursor and fragment ions of the peptides from the targeted protein and the internal standards. Mandatory parameters required for the generation of a library of precursors and fragment ions are depicted. Example of the library for detection of Cx43. Q1, precursor (intact peptide) *m/z*; Q2, fragment ion *m/z*; t_R detected, retention time of the peptide; stripped_sequence, peptide sequence without modification; prec_z and fragm_z, precursor and fragment ion charge (*z*), respectively; frag_typ, type of the fragment ion, usually b- or y-ios; decoy, inverted sequence, usually is *false*; confidence, peptide confidence; shared, peptide is shared by two or more proteins, usually indicates *false*; and N, indicated the protein group—different proteins must have different N

identify a number of proteins to analyze that contained the group number of the targeted protein and the iRT peptides. Exclude peptides with biological modifications and/or peptides shared between different protein entries/isoforms.

2. Filter the library generated in **step 1** in order to contain only the information correspondent to the targeted protein and the iRTs, eliminating all the information about other proteins also identified in the IDA experiment.
3. For peptides identified in MASCOT or PEAKS, select the fragment ions manually (Fig. 4). The selected precursors and respective fragment ions must be added to the tab-delimited text file created in **step 1**.

3.10.5 Detection of Cx43 in EVs Using the Semi-Targeted SWATH-MS Data (with SWATH™ Processing Plug-In for PeakView™)

1. Upload the file correspondent to the specific library created in the previous section (Fig. 3b).
2. Import the SWATH files of (1) a positive control (immunopurified Cx43, Subheading 3.5), (2) the samples of interest (iso-

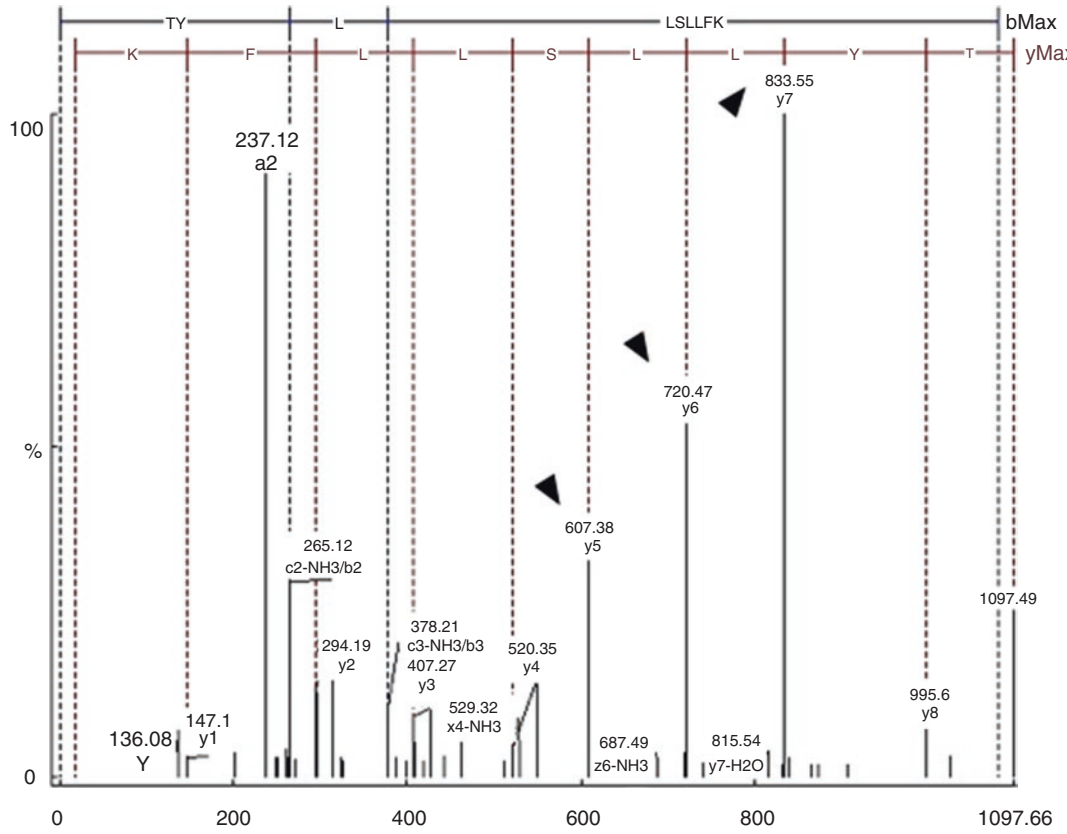


Fig. 4 Fragmentation spectrum (MS/MS spectrum) obtained in PEAKS for manual selection of the fragment ions to use in the specific library to detect a targeted protein. Example of a MS/MS spectrum of an identified Cx43 peptide. The fragment ions selected (*arrowhead*) should correspond to the most intense ones. Peptide sequence: TYIISILFK. Peptide *m/z*: 549.335. Selected fragment ions: y7 = 833.55, y6 = 720.47, and y5 = 607.38

lated EVs), and (3), if possible, a negative control (e.g., a Cx43 knockdown/knockout).

3. If a RT alignment is necessary, select the iRT peptides using the "RT+Cal" icon. This will add a new protein to the list (RT calibration protein), which should be selected to apply the RT calibration to the remaining data set.
4. Define the processing setting:
   (1) Peptide filter, use up to the maximum number of peptides in the library with up to five transitions per peptide; (2) XIC extraction window (min) should be adjusted to accommodate entire chromatographic peaks, usually around 3–5 min; and (3) XIC width (ppm or Da), dependent on instrument mass error, usually around 0.02 Da.
5. Peptides are confirmed by finding and scoring peak groups (Fig. 5), which are a set of fragment ions for a given peptide, following the criteria previously described (*see* ref. 12). Peak group confidence threshold is determined based on a FDR analysis using the target-decoy approach. 1% extraction FDR threshold should be used for positive identifications.
6. Protein positive detection/identification is considered if at least two peptides were confirmed in the sample of interest.

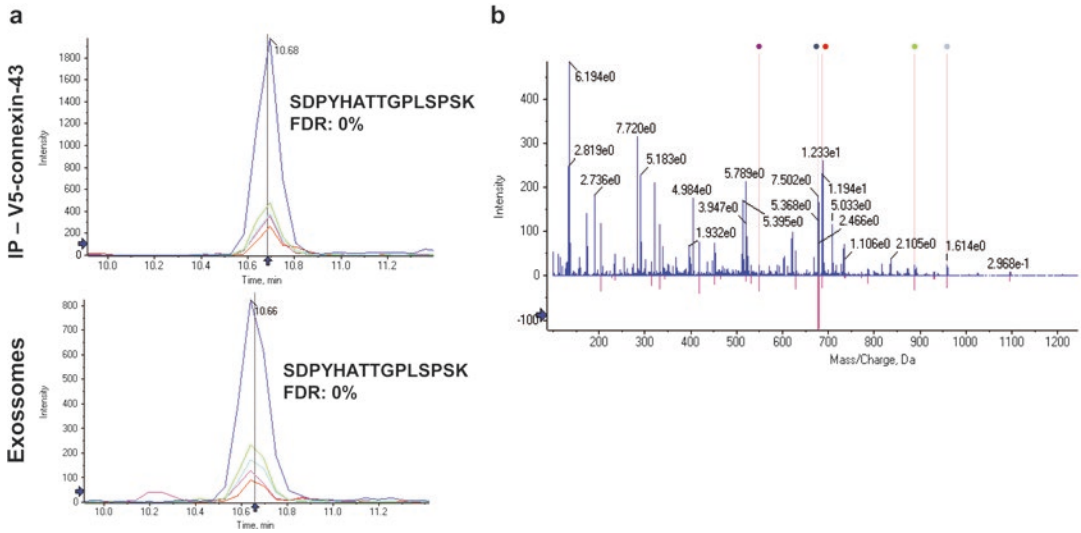


Fig. 5 Detection of Cx43 in EVs by semi-targeted SWATH-MS approach. (**a**) Peak group (extracted ion traces of chosen fragments) of one Cx43 peptide previously identified in IDA experiments of the Cx43-enriched sample (Cx43-IP; *top*) and EVs (*bottom*). For each peak group (peptide), the peptide sequence and FDR value, obtained using SWATH™ processing plug-in for PeakView™ and calculated according with the criteria described in Lambert *et al.* (*see* ref. 12), are identified. (**b**) Combined fragmentation spectrum obtained in the analysis of the EVs using the targeted SWATH method. Each spectrum represents the MS/MS spectrum obtained in the SWATH-MS acquisition (*blue*) and the theoretical spectrum (*pink*). The transitions used are indicated with *colored balls* according to its respective extracted ion trace color (in **a**). Adapted from [1]

4 Notes

1. It is possible to purify EVs from either serum or plasma samples. However, it is recommended to use serum samples, given that a large portion of contaminant circulating proteins can be eliminated during the clotting step. In either case, keep in mind that specific blood collection tubes (with or without anti-clotting agents) have to be used.
2. PBS should be checked by NTA to be particle-free before use. If particles are detected, PBS can be passed through a 0.22-μm filter, checked again and used as vehicle (if negative for particles). Alternatively, particle-free water can be used as vehicle.
3. In order to have a sufficient amount of immunopurified Cx43 to perform a targeted proteomic analysis, it is recommended to use approximately 20 mg of total protein as input for Cx43 immunoprecipitation, which corresponds to ~8 confluent 100 mm^2 dishes of HEK-293 cells, overexpressing V5-tagged Cx43. For control IP, the same amount of protein should be used, and the same protocol should be followed, but replacing the specific anti-Cx43 antibody by a nonspecific antibody, such as anti-GFP (*see* ref. 2).
4. Differential centrifugation methods are still gold standards for EV purification (*see* ref. 13). However, ultracentrifugations at 110,000 × *g* may provoke co-sedimentation of other contaminants, such as protein aggregates, lipoproteins, and extracellular proteins nonspecifically bound to EVs. Further purification, using sucrose or iodixanol gradients can be performed in order to eliminate these contaminants. The use of commercially available kits, such as the one applied here, have emerged as good alternatives. However, these approaches, whose principle is to restrain water molecules, forcing less-soluble components (including EVs) to precipitate, have a similar risk of coprecipitation of non-vesicular contaminants (*see* refs. 17, 18).
5. To isolate EVs from cultured cells, differential ultracentrifugation is the preferred method (*see* refs. 1, 13):
   (a) Grow cells (e.g., HEK-293) in cell culture medium (e.g., Dulbecco's Modified Eagle Medium, DMEM), until 70–90% confluency is reached. Then, wash the cells with PBS, and replace the medium for an EV-depleted medium, for 24–48 h (*see* **step 5b**). Harvest the conditioned culture medium and centrifuge at 300 × *g*, for 10 min, to remove dead cells and cellular debris. Pour the supernatant into clean centrifuge tubes and centrifuge at 16,500 × *g*, for 20 min. Collect the supernatant and pass it through a 0.22-μm filter unit. Ultracentrifuge the filtrates at 120,000 × *g*, for 70 min. Resuspend the EV pellets in an appropriate buffer solution (*see* **Note 9**).

(b) EV-depleted medium is prepared by ultracentrifugation of 50% FBS (diluted in DMEM), at 120,000 × *g*, for 16 h. Carefully pour the supernatant into clean tubes and dilute it to a final concentration of 10% FBS in DMEM. Given that serum contains abundant extracellular proteins that can coprecipitate with purified EVs, it is recommended that the EV-producing cells were grown in 2–5% serum, to reduce the amount of contaminants found by proteomic analysis. Alternatively, serum-free medium can be used; however, it can compromise cell viability or alter the proteome profile of EVs produced by the cells in such stressful environment.

6. Proteomic analyses require a large amount of protein, therefore when using cell cultures, it is necessary to use around 1×10^9 cells. On the other hand, serum/plasma have greater yields. Typically, the ultracentrifugation method yields about ten times less total protein than commercially available kits. In the latter, it is possible to obtain ~500 μg protein/mL serum, which corresponds to ~10^{11}–10^{12} particles/mL serum (*see* refs. 17, 19).
7. After serum isolation, purification of EVs should be performed as soon as possible, in order to minimize degradation of EVs.
8. Up to this point, and unless stated otherwise, the entire procedure should be carried out at 4 °C, to prevent degradation of EVs and preserve protein integrity.
9. The choice of the most appropriate buffer to resuspend EV pellets is ultimately dictated by the downstream applications of EVs. Make sure that the buffer components do not inhibit chemical reactions and/or enzymes required in further functional experiments or biochemical analyses. Some examples are the following:
   (a) To characterize EVs present in serum by NTA, EV-containing pellets should be resuspended in particle-free PBS.
   (b) For TEM characterization, EVs should be resuspended in PBS, followed by fixation in 2% (w/v) paraformaldehyde (PFA; dilute in PBS from premade 16% (w/v) (Electron Microscopy Sciences, Hatfield, PA, USA)). EVs in 2% PFA can be stored up to 1 week at 4 °C before further processing.
   (c) For WB analysis, EV pellets should be resuspended in PBS, after which a 2× RIPA buffer (or similar) should be added.
   (d) For proteomic analysis, EV pellets should be resuspended in RIPA buffer, followed by denaturation in 1× Laemmli buffer.
   (e) Since phosphate-based buffers inhibit certain enzymes, including alkaline phosphatase, resuspend EV pellets in Tris-based buffers (e.g., Tris-buffered saline, TBS), whenever alkaline phosphatase has to be used.

10. Before starting, clean sample chamber surfaces using particle-free water and a lens wipe or tissue. Rinse syringes to be used, by filling and discarding diluent solution three times. After setting the chamber, clean the system by pumping through it three times its total volume (~1 mL). If a syringe pump is available, all the washes and readings can be performed under controlled and automatic flow. If pumping, manually ensure that flow is slow and that excessive pressure is not being exerted. Record and/or control in the cell (can be done automatically, depending on instrument model).
11. Pilot studies to test optimal dilution sample may be required to ensure that particle concentration is in the region of 1×10^8 particles/mL (manufacturer's recommendations; *see* ref. 20).
12. When available, autosetup function available in the software can be used to obtain the appropriate camera level and focus. To accurately track the vesicles, they must be visualized as single points of light. The NTA software is unable to effectively track very large vesicles or those with confounding Newton rings. In this case, gain can be reduced accordingly.
13. The duration of each capture depends on the number of particles visible in the field of view; 60 s is usually sufficient if 20–60 particles can be tracked in the field of view in any single frame. A minimum of 500 total particle tracks per measurement should be collected.
14. The settings are optimal when each white dot that should be considered as a particle is labelled with a red cross. When the detection threshold is too high, some particles are not included in the data profile (white dot will not contain red dot). However, if the detection threshold is set too low, noise will be included in the final data.
15. The membrane side of the grid should be kept wet during all steps, while the reverse should be maintained dried.
16. The presence of proteins, such as CD9, CD63, CD81, and ICAM-1, should be performed in non reducing conditions (i.e., sample buffer without DTT or β-mercaptoethanol).
17. Due to the lack of specific markers, proteins that are particularly enriched in EVs are commonly used to certify the presence of EVs (some being specific to the cell of origin). Examples of these proteins are tetraspanins (e.g., CD9, CD63, and CD81), proteins involved in multivesicular biogenesis (e.g., Tsg101 and Alix), flotillin, clathrin, Hsc70, A33, CD3, MHC classes I and II, and ICAM-1. It is also recommended to assess the presence of markers of other cell compartments that produce vesicles such as the endoplasmic reticulum (e.g., calnexin and Grp78) and the Golgi apparatus (e.g., GM130). The absence of these proteins indicates no or little contamination with vesicles of other compartments.

18. If lysates of cells that produce EVs are not available (like in the case of the use of human blood samples), EVs and/or whole cell lysates from the same species can be used as technical control.
19. The centrifugation time and force should vary according to the cell type and/or the proteins of interest for the immunoprecipitation protocol. In the case of Cx43, low-speed centrifugations should be used to prevent loss of gap junctional Cx43.
20. Sample buffer composition can be adjusted to increase the desired solubility of the proteins. In general, an increase in detergent and/or reducing agents can help to promote more efficient solubilization. This may be particularly important in the case of proteins that are difficult to solubilize, including membrane proteins.
21. Protein denaturation promotes a more efficient protein solubilization.
22. Gradient gels should be used to promote a better separation of proteins from complex mixtures. Furthermore, the use of precast gels and commercially available solutions will reduce sample contamination with keratin.
23. The volume of the samples should be similar; if necessary adjust the volume with 1× Laemmli buffer.
24. The staining is very useful to verify the amount of protein loaded and to define which bands should be analyzed by MS.
25. Direct contact of the band with any potential dirty surface should be avoided. Always wash the gloves after touching any potential dirty surface.
26. Micropipette tips adapted for SPE applications have a known amount of sorbent, do not exceed the correct amount of sample to be applied; otherwise some loss of peptides of interest can occur due to matrix saturation.
27. Concentration of iRT peptides in the sample can be 10× lower than the recommended by the manufacturer.
28. The length of the gradient should be adjusted according to sample complexity and application. For samples with a reduced complexity, use a 25 min gradient, while for samples with higher complexity, use a 45 min gradient.
29. A typical LC method should comprise the following phases: (1) equilibration of the column (this phase is linked to the injection step), (2) gradient of organic solvent (could be linear or stepwise), (3) column wash with high organic content, and (4) re-equilibration of the column with high hydrophilic content of the mobile phase.

30. Some parameters of the acquisition method, such as the number of candidate ions and the accumulation time, should be adjusted according to the sample and the chromatographic peak width. This would enable the acquisition of the maximum information within a cycle time, compatible with the chromatographic separation. A minimum of 8 points should be acquired across the chromatographic peak to obtain a good peak profile. In order to determine the compatible cycle time, the peak width should be divided by 8 (or by the desired number of points per peak). With the indicated chromatographic conditions, the cycle time is usually around 3 s.
31. The width of the SWATH windows can be adjusted to limit the interferences, reducing the number of precursor included in an acquisition window. The minimum width is 3 Da.
32. In the most recent versions of the Analyst TF (version 1.7), the SWATH windows defined can be directly uploaded from a tab-delimited text file. For older versions, build a method covering the *m/z* range of all the peptides analyzed and with the SWATH window width desired. Eliminate the unnecessary windows.

Acknowledgments

This work was supported by the Portuguese Foundation for Science and Technology (FCT) grants, FCT-UID/NEU/04539/2013, PTDC/NEU-NMC/0205/2012, POCI-01-0145-FEDER-007440, and PTDC/NEU-SCC/7051/2014, by REDE/1506/REM/200 and by HealthyAging2020 CENTRO-01-0145-FEDER-000012-N2323. TMM was supported by PD/BD/106043/2015, SIA by SFRH/BD/81495/2011, and TRR by PD/BD/52294/2013.

References

1. Soares AR, Martins-Marques T, Ribeiro-Rodrigues T et al (2015) Gap junctional protein Cx43 is involved in the communication between extracellular vesicles and mammalian cells. Sci Rep 5:13243
2. Martins-Marques T, Anjo SI, Pereira P et al (2015) Interacting network of the gap junction (GJ) protein Connexin43 (Cx43) is modulated by ischemia and reperfusion in the heart. Mol Cell Proteomics 14:3040–3055
3. Gillet LC, Navarro P, Tate S et al (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics 11:O111.016717
4. Anjo SI, Lourenço AS, Melo MN et al (2016) Unraveling mesenchymal stem cells' dynamic secretome through nontargeted proteomics profiling. Methods Mol Biol 1416: 521–549
5. Ong S-E, Mann M (2005) Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol 1:252–262

6. Granvogl B, Plöscher M, Eichacker LA (2007) Sample preparation by in-gel digestion for mass spectrometry-based proteomics. Anal Bioanal Chem 389:991–1002
7. Finlay EM, Games DE, Startin JR et al (1986) Screening, confirmation, and quantification of sulphonamide residues in pig kidney by tandem mass spectrometry of crude extracts. Biomed Environ Mass Spectrom 13:633–639
8. Liebler DC, Zimmerman LJ (2013) Targeted quantitation of proteins by mass spectrometry. Biochemistry 52(22):3797–3806
9. Boja ES, Fehniger TE, Baker MS et al (2014) Analytical validation considerations of multiplex mass-spectrometry-based proteomic platforms for measuring protein biomarkers. J Proteome Res 13:5325–5332
10. Boja ES, Rodriguez H (2012) Mass spectrometry-based targeted quantitative proteomics: achieving sensitive and reproducible detection of proteins. Proteomics 12:1093–1110
11. Liu Y, Hüttenhain R, Surinova S et al (2013) Quantitative measurements of *N*-linked glycoproteins in human plasma by SWATH-MS. Proteomics 13:1247–1256
12. Lambert J-P, Ivosev G, Couzens AL et al (2013) Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. Nat Methods 10:1239–1245
13. Théry C, Amigorena S, Raposo G et al (2006) Isolation and characterization of exosomes from cell culture supernatants and biological fluids. In: Current protocols in cell biology. John Wiley & Sons, Inc., Hoboken, NJ, pp 3.22.1–3.22.29
14. Dragovic RA, Gardiner C, Brooks AS et al (2011) Sizing and phenotyping of cellular vesicles using nanoparticle tracking analysis. Nanomedicine 7:780–788
15. Momen-Heravi F, Balaj L, Alian S et al (2012) Alternative methods for characterization of extracellular vesicles. Front Physiol 3:354
16. Candiano G, Bruschi M, Musante L et al (2004) Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. Electrophoresis 25:1327–1333
17. Tang WH, Shilov IV, Seymour SL (2008) Nonlinear fitting method for determining local false discovery rates from decoy database searches. J Proteome Res 7:3661–3667
18. Sennels L, Bukowski-Wills J-C, Rappsilber J (2009) Improved results in proteomics by use of local and peptide-class specific false discovery rates. BMC Bioinformatics 10:179
19. Li M, Zeringer E, Barta T et al (2014) Analysis of the RNA content of the exosomes derived from blood serum and urine and its potential as biomarkers. Philos Trans R Soc Lond B Biol Sci 369:654–659
20. Turay D, Khan S, Diaz Osterman CJ et al (2016) Proteomic profiling of serum-derived exosomes from ethnically diverse prostate cancer patients. Cancer Invest 34:1–11

Chapter 13

Red Blood Cells in Clinical Proteomics

Ana Sofia Carvalho, Manuel S. Rodriguez, and Rune Matthiesen

Abstract

Red blood cells (RBCs) are known for their role in oxygen and carbon dioxide transport. The main function of RBCs is directly linked to many diseases that cause low oxygen levels in tissues such as congenital heart disease in adults, chronic obstructive pulmonary disease, sleep apnea, sickle cell disease, etc. Red blood cells are a direct target for a number of parasitic diseases such as malaria (*Plasmodium*) and similar parasites of the phylum Apicomplexa (*Toxoplasma*, *Theileria*, *Eimeria*, *Babesia*, and *Cryptosporidium*). RBC membrane components, in particular, are suitable targets for the discovery of drugs against parasite interaction. There is also evidence that RBCs release growth and survival factors, thereby linking RBCs with cancer. RBCs are abundant and travel throughout the body; consequently changes in RBC proteome potentially reflect other diseases as well. This chapter describes erythrocyte isolation from blood and its fractionation into RBC membrane and soluble cytosolic fractions. Alternative procedures for mass spectrometry analysis of RBC membrane proteome will be presented.

Key words Red blood cell, Proteome, Membranar proteins, Mass spectrometry, Proteases, Infection

1 Introduction

The average human has 5 L of blood. Red blood cells also known as erythrocytes are the most common cell in blood. RBCs synthesized in the bone marrow enter circulation where they gradually degrade and are consequently removed by macrophages in the spleen and liver after 120 days [1]. A complete blood count (CBC) is a routine medical assessment that measures a number of blood parameters that can hint to a number of pathologies such as anemia, infection, inflammatory diseases, and malignancy. For example, hematocrit measures the fraction of total blood cells that constitute RBCs. A low hematocrit count, due to a drop in RBCs production in the bone marrow, can be a consequence of a decrease in erythropoietin or malfunctioning bone marrow caused by toxins or cancer [1]. Therefore, more detailed molecular investigation of RBCs potentially holds the promise to contain markers for a number of diseases.

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_13,

Mature RBCs in mammals lack nucleus, thereby making genomics less relevant compared to proteomics. RBC's important physiologic function, lack of internal organelles, and easy of obtaining make them attractive for proteomics analysis [2]. Currently proteomics studies can be divided into two subcellular fractions, membranar fraction [3–6], cytosolic fraction [7–10], or both [11, 12]. In addition to the quantitative protein changes upon different diseases, changes in posttranslational modifications have been observed. For example, proteolytic and oxidative damage of membrane skeletal proteins of spectrin has been reported during blood storage, [4] and the oxidative state of peroxiredoxin 2 has been demonstrated to change upon sleep apnea and upon positive airway pressure (PAP) treatment [13]. Furthermore, mature RBCs contain intact proteasomes [2] and ubiquitin is abundant in erythrocytes, as well as many of the proteins identified in RBCs can be ubiquitinated [11]. The role of ubiquitination in RBCs is not fully elucidated, and the ubiquitome have not been globally explored [14]. The human and *Plasmodium falciparum* tandem ubiquitin-binding entity (TUBE) proteome were recently profiled over the asexual intraerythrocytic developmental cycle (IDC) of *Plasmodium falciparum*, and ubiquitin proteasome factors were found to be highly abundant [15]. We, therefore, speculate that the combination of the protocols in this chapter can be merged with previously published protocols for TUBE enrichment of ubiquitinated proteins [16] and anti-diglycine-based enrichment of peptides potentially originating from ubiquitinated proteins [17].

2 Materials

2.1 Isolation of Erythrocytes from Peripheral Whole Blood and Fractionation of RBCs

1. Lysis buffer: 5 mM Na_2HPO_4, 8 mM EDTA, pH 8.
2. Protease inhibitor cocktail tablets (Complete mini, Roche).
3. EDTA tubes, plastic pipettes, and centrifugation tubes.
4. Microcentrifuge with fixed-angle rotor.

2.2 RBC Membrane Extraction

Lysis buffer: 5 mM phosphate buffer, 8 mM EDTA, pH 8.

100 mM Na_2CO_3, pH 11.

25-gauge needle.

Microcentrifuge with fixed-angle rotor.

2.3 SDS-PAGE

XCell SureLock™ Mini-Cell

Tris-Glycine Gels (SDS-PAGE) 4–12%

Tris-Glycine SDS Running Buffer (10×)

Tris-Glycine SDS Sample Buffer (2×)

Reducing Agent (10×)

2.4 MS Sample Preparation of Membrane RBCs

Wash solution: 50% (v/v) methanol, 5% (v/v) acetic acid in water.

100 mM ammonium bicarbonate.

50 mM ammonium bicarbonate.

10 mM DTT in 100 mM ammonium bicarbonate.

100 mM iodoacetamide in 100 mM ammonium bicarbonate.

Trypsin solution: dissolve 20 μg of trypsin sequence grade (Promega) in 1000 μL of ice cold 50 mM ammonium bicarbonate.

Extraction buffer: 50% (v/v) acetonitrile and 5% (v/v) formic acid in water.

Reconstitution solution: 5% (v/v) formic acid in water.

Microcentrifuge with fixed-angle rotor.

SpeedVac Concentrator.

3 Methods

3.1 Isolation of Erythrocytes from Peripheral Whole Blood

1. Draw blood into blood collection tubes containing anticoagulant (*EDTA*, sodium citrate, sodium heparin, or lithium heparin) (*see* **Note 1**).
2. Invert tubes carefully ten times to mix blood and anticoagulant and incubate for 72–96 h, 4 °C, without shaking to allow maturation of reticulocytes to RBCs.
3. Samples should undergo centrifugation immediately. This should be carried out for a minimum of 10 min at 1000 × *g*, 4 °C.
4. After centrifugation, remove the plasma (top layer), the buffy coat, in the interface between the plasma and the RBCs, which contains white blood cells and platelets. The remaining red fraction corresponds to the RBCs.
5. Resuspend RBCs in 10 mL of 0.9% (w/v) NaCl in 5 mM phosphate buffer, pH 8.
6. Mix thoroughly by pipetting up and down.
7. Centrifuge 10 min at 1000 × *g*, 4 °C, and discard upper liquid phase.
8. Repeat **steps 5**–7 three times (*see* **Note 2**).
9. Resuspend RBCs in 3 mL of 0.9% (w/v) NaCl in 5 mM phosphate buffer, pH 8 (*see* **Note 3**).

3.2 Fractionation of RBCs

1. Add 100 μL of erythrocytes (RBCs) to a microcentrifuge tube.
2. Add 900 μL of ice cold 5 mM phosphate buffer, 8 mM EDTA, pH 8 containing protease inhibitors.

3. Incubate for 1 h at 4 °C with gentle shaking.
4. Centrifuge for 30 min at 25,000 × *g*, 4 °C.
5. The supernatant constitutes the cytosolic or soluble fraction, and the pellet constitutes the crude fraction of ghost RBCs (*see* **Note 4**).
6. The cytosolic fraction can be stored at −80 °C until further analysis.
7. For MS analysis of the cytosolic fraction, depletion of hemoglobin is highly recommended using, for example, HemoVoid™—Hemoglobin Depletion Reagent Kit (Biotech Support Group) as described by the manufacture protocol. Trypsin digestion of the depleted fraction can be performed as in Subheading 3.5. Alternatively in solution digestion or filter-aided sample preparation (FASP) can be used [18].

3.3 RBC Membrane Extraction

1. Add 1 mL of ice cold 5 mM phosphate buffer, 8 mM EDTA, pH 8 containing protease inhibitors to crude fraction of ghost RBCs.
2. Incubate for 30 min at 4 °C with gentle shaking.
3. Centrifuge for 30 min at 25,000 × *g*, 4 °C, and discard supernatant.
4. Repeat **steps 1–3** until obtaining a "whitish" pellet (*see* **Note 5**).
5. Resuspend the pellet in 1 mL 100 mM Na_2CO_3 (pH 11) and pass five times through a 25-gauge needle.
6. Incubate for 30 min, 4 °C with agitation.
7. Centrifuge for 90 min at 245,000 × *g* and discard supernatant.
8. Repeat **steps 5–7**.

3.4 SDS-PAGE

1. Resuspend pellet (Subheading 3.3, **step 8**) in 10 μL of 10% (wt/vol) sodium dodecyl sulfate and vortex.
2. Pipette 8 μL of pellet solution to a microcentifuge tube.
3. Add 10 μL of Tris-Glycine SDS Sample Buffer (2×).
4. Add 2 μL Reducing Agent (10×).
5. Heat samples at 85 °C for 2 min. Load the samples onto the gel immediately.
6. Load 10 μL of the sample on a precast Tris-Glycine 4% to 12% polyacrylamide gel (NuPAGE, Invitrogen) in Tris-Glycine SDS Running Buffer 1×, and run the gel for 10 min at 125 V.
7. Remove the gel from the cassette and rinse the gel with 100 mL deionized water three times for 5 min.
8. Stain the gel with enough SimplyBlue™ SafeStain (Invitrogen) to cover the gel, for 1 h at room temperature with gentle

shaking. Bands will begin to develop within minutes. After incubation, discard the stain.

9. Wash the mini-gel with 100 mL of water for 1–3 h. Discard the washing solution.
10. Place the gel in a clean surface and keep the gel hydrated with deionized water.

3.5 MS Sample Preparation of Membrane RBCs

1. Cut the gel band from the gel with a sharp scalpel and further split it into smaller pieces (1–2 mm^3).
2. Transfer the gel pieces into a 1.5 mL microcentrifuge tube.
3. Wash the gel pieces with 100–200 μL of 50% (v/v) methanol and 5% (v/v) acetic acid overnight at room temperature (*see* **Note 6**).
4. Remove the wash solution from the gel pieces with a plastic pipette and discard.
5. Wash the gel pieces with 100–200 μL of 50% (v/v) methanol and 5% (v/v) acetic acid for further 2–3 h at room temperature.
6. Remove the wash solution from the gel pieces with a plastic pipette and discard.
7. Dehydrate the gel pieces with 100–200 μL of acetonitrile for 5 min at room temperature. The gel pieces will become white and with smaller size.
8. Remove the acetonitrile from the gel pieces with a plastic pipette and discard (*see* **Note 7**).
9. Dry the gel pieces in a vacuum centrifuge at room temperature for 2–3 min.
10. Reduce the proteins with 30 μL of 10 mM DTT in ammonium bicarbonate for 30 min at room temperature.
11. Remove the DTT solution from the sample with a plastic pipette and discard.
12. Alkylate the proteins with 30 μL of 100 mM iodoacetamide in ammonium bicarbonate for 30 min at room temperature.
13. Remove the iodoacetamide solution from the sample with a plastic pipette and discard.
14. Dehydrate the gel pieces with 100–200 μL of acetonitrile for 5 min at room temperature. The gel pieces will become white and with smaller size.
15. Remove the acetonitrile from the gel pieces with a plastic pipette and discard.
16. Rehydrate the gel pieces in 200 μL of 100 mM ammonium bicarbonate, incubating the samples for 10 min at room temperature.

17. Remove the ammonium bicarbonate from the sample with a plastic pipette and discard.
18. Dehydrate the gel pieces with 200 μL acetonitrile with a plastic pipette and discard, for 5 min at room temperature.
19. Remove the acetonitrile from the sample with a plastic pipette and discard.
20. Dry the gel pieces in a vacuum centrifuge at room temperature for 2–3 min.
21. Prepare a trypsin solution, 20 ng/μL, on ice by adding 1000 μL of ice cold 50 mM ammonium bicarbonate to 20 μg of trypsin.
22. Rehydrate the gel pieces on ice for 10 min with 30 μL of trypsin solution. Vortex occasionally. The gel pieces must rehydrate by increasing size and drop the white color.
23. Collect the gel pieces in the bottom of the tube by centrifuging the tube for 30 s.
24. Add 5–10 μL of 50 mM ammonium bicarbonate to the gel pieces to ensure they are rehydrated. Mix the sample and collect the gel pieces in the bottom of the tube by centrifuging the tube for 30 s. Digest overnight at 37 °C.
25. To extract the peptides from the protein digestion:

    Add 30 μL of 50 mM ammonium bicarbonate to the digestion sample and incubate the sample for 10 min at room temperature mixing occasionally. Collect the gel pieces in the bottom of the tube by centrifuging the tube for 30 s. Pipette the sample carefully and transfer it to a microcentrifuge tube.

    Add 30 μL of extraction buffer to the digestion sample and incubate the sample for 10 min at room temperature mixing occasionally. Collect the gel pieces in the bottom of the tube by centrifuging the tube for 30 s. Pipette the sample carefully and transfer to the previous microcentrifuge tube. Repeat this step by adding a second 30 μL aliquot of extraction buffer.
26. Reduce the volume of the peptide sample to ~25 μL by evaporation in a vacuum centrifuge at room temperature.
27. Desalt the peptide sample using a Stage Tip [18] and reconstitute the peptides in 10–20 μL of 5% formic acid. The sample can be readily analyzed by LC-MS/MS.

3.6 Data Analysis

A number of programs are available for identification. Quantitation of proteins in the form of spectral counting, ion currents, or intensity-based absolute quantification (iBAQ) [19] values can be obtained from programs such as MaxQuant [20], VEMS [21], and X!Tandem and related tools [22]. Statistical analysis of the quantitative data is conveniently done in the R statistical programing language using packages such as the limma package and related packages [23]. The protein identifications can be validated against

previous published large-scale LC-MS data sets of RBCs [3–12] and online databases [24]. The identified membrane proteins can be grouped based on predicted membrane-spanning domains [25]. Another way to validate the identifications from whole RBCs, cytosolic RBCs, or membrane RBCs is to perform complete functional analysis [26] to define the level of enrichment statistically, for example, proteins from RBCs versus other blood cells or membrane proteins versus proteins from other subcellular fractions.

4 Notes

1. EDTA and heparin sulfate are widely used anticoagulants. The effect of anticoagulants in the analysis of blood components such as plasma has been reported. Effects of blood sampling must be considered prior to mass spectrometry analysis such as the choice of anticoagulant, the preparation, and the storage time prior to sample analysis.
2. Purity and quality of RBC samples are crucial. During the isolation procedure, lysis of RBC can occur; therefore, repeated washes are recommended. This will also help to eliminate plasma proteins dissolved in the supernatant. Additionally, plasma proteins can be eliminated by removing the top layer after each centrifugation washing step.
3. To further access purity and quality of RBC samples, check for contamination of other blood cell types. Whole blood isolated RBC samples can contain white blood cells and reticulocytes. A detailed method for the assessment has been described [11].
4. Preparation of ghost RBCs from freshly isolated cells produces a high-quality sample compared with long-term storage RBC.
5. High pure RBC ghost samples appear whitish, and washing steps must be repeated until a colorless supernatant is observed.
6. After rehydration, observe the gel pieces and if necessary add 10 µL aliquots to complete gel rehydration. During in-gel digestion procedure, avoid that the gel pieces dry.
7. Carefully remove acetonitrile before each rehydration step, particularly in the digestion with trypsin as acetonitrile can inhibit the protease activity.

Acknowledgments

RM is supported by FCT investigator program 2012 (IF/01002/2012). ASC is supported by grant SFRH/BPD/85569/2012 funded by Fundação para a Ciência e Tecnologia.

References

1. Dean L (2005) Chapter 1. National Center for Biotechnology Information: Bethesda, MD. https://www.ncbi.nlm.nih.gov/books/NBK2263/
2. Goodman SR, Kurdia A, Ammann L, Kakhniashvili D, Daescu O (2007) The human red blood cell proteome and interactome. Exp Biol Med (Maywood) 232:1391–1408. doi:10.3181/0706-MR-156
3. van Gestel RA et al (2010) Quantitative erythrocyte membrane proteome analysis with Blue-native/SDS PAGE. J Proteomics 73:456–465. doi:10.1016/j.jprot.2009.08.010
4. D'Amici GM, Rinalducci S, Zolla L (2007) Proteomic analysis of RBC membrane protein degradation during blood storage. J Proteome Res 6:3242–3255. doi:10.1021/pr070179d
5. De Palma A et al (2010) Extraction methods of red blood cell membrane proteins for Multidimensional Protein Identification Technology (MudPIT) analysis. J Chromatogr A 1217:5328–5336. doi:10.1016/j.chroma.2010.06.045
6. Pesciotta EN et al (2012) A label-free proteome analysis strategy for identifying quantitative changes in erythrocyte membranes induced by red cell disorders. J Proteomics 76:194–202. doi:10.1016/j.jprot.2012.08.010
7. D'Amici GM, Rinalducci S, Zolla L (2011) An easy preparative gel electrophoretic method for targeted depletion of hemoglobin in erythrocyte cytosolic samples. Electrophoresis 32:1319–1322. doi:10.1002/elps.201000659
8. Walpurgis K et al (2012) Validated hemoglobin-depletion approach for red blood cell lysate proteome analysis by means of 2D PAGE and Orbitrap MS. Electrophoresis 33:2537–2545. doi:10.1002/elps.201200151
9. Ringrose JH et al (2008) Highly efficient depletion strategy for the two most abundant erythrocyte soluble proteins improves proteome coverage dramatically. J Proteome Res 7:3060–3063. doi:10.1021/pr8001029
10. Pesciotta EN et al (2015) In-depth, label-free analysis of the erythrocyte cytoplasmic proteome in diamond blackfan anemia identifies a unique inflammatory signature. PLoS One 10:e0140036. doi:10.1371/journal.pone.0140036
11. Pasini EM et al (2006) In-depth analysis of the membrane and cytosolic proteome of red blood cells. Blood 108:791–801. doi:10.1182/blood-2005-11-007799
12. Pasini EM et al (2008) Deep coverage mouse red blood cell proteome: a first comparison with the human red blood cell. Mol Cell Proteomics 7:1317–1330. doi:10.1074/mcp.M700458-MCP200
13. Feliciano A et al (2016) Evening-morning peroxiredoxin-2 redox/oligomeric state changes in obstructive sleep apnea red blood cells: correlation with polysomnographic and metabolic parameters. Biochim Biophys Acta 1863(2):621–629. doi:10.1016/j.bbadis.2016.11.019
14. Chakrabarti A, Halder S, Karmakar S (2016) Erythrocyte and platelet proteomics in hematological disorders. Proteomics Clin Appl 10:403–414. doi:10.1002/prca.201500080
15. Mata-Cantero L et al (2016) New insights into host-parasite ubiquitin proteome dynamics in P. falciparum infected red blood cells using a TUBEs-MS approach. J Proteomics 139:45–59. doi:10.1016/j.jprot.2016.03.004
16. Azkargorta M, Escobes I, Elortza F, Matthiesen R, Rodriguez MS (2016) TUBEs-mass spectrometry for identification and analysis of the ubiquitin-proteome. Methods Mol Biol 1449:177–192. doi:10.1007/978-1-4939-3756-1_9
17. Swaney DL et al (2013) Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. Nat Methods 10:676–682. doi:10.1038/nmeth.2519
18. Wisniewski JR, Zougman A, Mann M (2009) Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. J Proteome Res 8:5674–5678. doi:10.1021/pr900748n
19. Mann K, Edsinger E (2014) The Lottia gigantea shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. Protein Sci 12:28. doi:10.1186/1477-5956-12-28
20. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372. doi:10.1038/nbt.1511
21. Carvalho AS et al (2014) Global mass spectrometry and transcriptomics array based drug profiling provides novel insight into glucosamine induced endoplasmic reticulum stress. Mol Cell Proteomics 13:3294–3307. doi:10.1074/mcp.M113.034363
22. Duncan DT, Craig R, Link AJ (2005) Parallel tandem: a program for parallel processing of tandem mass spectra using PVM or MPI and X!Tandem. J Proteome Res 4:1842–1847. doi:10.1021/pr050058i

23. Law CW, Alhamdoosh M, Su S, Smyth GK, Ritchie ME (2016) RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Res 5:1408. doi:10.12688/f1000research.9005.1
24. Hegedus T et al (2015) Inconsistencies in the red blood cell membrane proteome analysis: generation of a database for research and diagnostic applications. Database (Oxford) 2015:bav056. doi:10.1093/database/bav056
25. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. doi:10.1006/jmbi.2000.4315
26. Carvalho AS, Molina H, Matthiesen R (2016) New insights into functional regulation in MS-based drug profiling. Sci Rep 6:18826. doi:10.1038/srep18826

Chapter 14

High-Throughput Quantitative Lipidomics Analysis of Nonesterified Fatty Acids in Plasma by LC-MS

Nicolas Christinat, Delphine Morin-Rivron, and Mojgan Masoodi

Abstract

Nonesterified fatty acids are important biological molecules which have multiple functions such as energy storage, gene regulation, or cell signaling. Comprehensive profiling of nonesterified fatty acids in biofluids can facilitate studying and understanding their roles in biological systems. For these reasons, we have developed and validated a high-throughput, nontargeted lipidomics method coupling liquid chromatography to high-resolution mass spectrometry for quantitative analysis of nonesterified fatty acids. Sufficient chromatographic separation is achieved to separate positional isomers such as polyunsaturated and branched-chain species and quantify a wide range of nonesterified fatty acids in human plasma samples. However, this method is not limited only to these fatty acid species and offers the possibility to perform untargeted screening of additional nonesterified fatty acid species.

Key words Lipidomics, Liquid chromatography, Mass spectrometry, Nonesterified fatty acids, Branched-chain fatty acids, Human plasma

1 Introduction

Fatty acids are a large family of carboxylic acids having an aliphatic chain of various length and degree of saturation. In nature, they are known for being primary components of complex lipids such as phospholipids, acyl glycerols, or sphingolipids, but they can also be found in the circulation as free, nonesterified fatty acids (NEFA). NEFA are involved in multiple biological processes and have been linked to a various human diseases such as brain disease [1], obesity [2], diabetes [3], and insulin resistance [4]. Measuring and understanding the profile of circulating nonesterified fatty acids are extremely important to elucidate their biological functions and properties, and for that reason this topic has been considerably studied over the past years [5].

Traditionally fatty acid profiles have been measured using gas chromatography (GC) coupled with flame ionization detection (FID) or mass spectrometry (MS) [6]. Recently liquid chromatography

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_14,

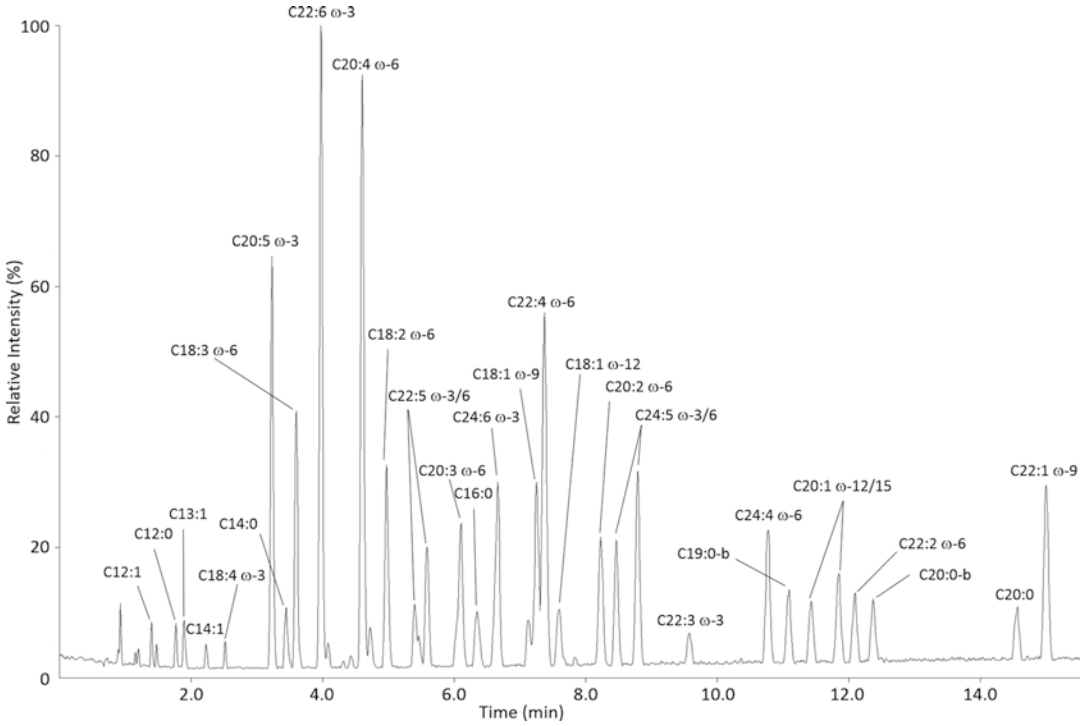


Fig. 1 Base peak chromatogram from calibration solution presenting the separation of detected NEFA

(LC) coupled with mass spectrometry has been successfully implemented as an alternative to GC. The main advantage of LC-MS(/MS) over GC-MS is the possibility to measure fatty acids without having to derivatize them prior to analysis while maintaining good method sensitivity. This does not only greatly simplify and accelerate sample preparation but also broadens the field of applications to new matrices [7–9].

LC-MS(/MS) methods for quantification of fatty acids in biological matrices have been abundantly described in the scientific literature [10]. These methods often target a subclass of NEFA, but recently methods profiling a broader range of NEFA have been published [8, 11]. Koletzko and co-workers have presented a method for the quantification of 30–40 NEFA species in various biofluids [12–14]. This method however lacks the resolving power to distinguish between positional isomers and, being a targeted method, does not allow for screening of unknown NEFA species.

In this chapter, we present a high-throughput nontargeted LC-MS method for the quantification of medium to very long chain NEFA in human plasma. It allows for the separation of most NEFA positional isomers and has been validated for a wide panel of NEFA species covering the classes of saturated and unsaturated but also branched-chain fatty acid (Fig. 1) [15]. Furthermore, the method is not limited to quantification of these species as high-resolution mass spectrometry detection offers the possibility to simultaneously screen for additional unknown NEFA species.

2 Materials

2.1 Samples

50 μL human plasma collected on EDTA tubes.

2.2 Standard Solution Preparation

Prepare all solution using HPLC or LC-MS grade solvents (methanol, isopropanol, chloroform, and acetonitrile), 18 MΩ water (Milli-Q water), and 99.99% purity ammonium acetate powder. Store all stock solution at −20 °C.

1. A mixture of fatty acid standards (GLC-566) (Nu-Chek Prep, Inc., Elysian, MN, USA). The whole content of the ampule (~100 mg) is accurately weighted and dissolved in 10 mL methanol. The solution is further diluted with methanol until a concentration of 30–250 μM of each fatty acid is reached (NEFA mix 1).
2. 12-Methyltetradecanoic acid, 14-methylhexadecanoic acid, 15-methylhexadecanoic acid, 10-heptadecenoic acid, petroselinic acid, and pristanic acid are purchased from Sigma-Aldrich (St. Louis, MO, USA). 5-Eicosenoic acid and 8-eicosenoic acid are obtained from Nu-Chek Prep, Inc. (Elysian, MN, USA) and 13-methyltetradecanoic acid, stearidonic acid, phytanic acid, mead acid, 4,7,10,13,16-docosapentaenoic acid, 9,12,15,18-tetracosatetraenoic acid, 6,9,12,15,18-tetracosapentaenoic acid, 9,12,15,18,21-tetracosapentaenoic acid, and nisinic acid are purchased from Larodan Fine Chemicals AB (Malmoe, Sweden). 10 mM stock solutions are prepared separately in methanol or chloroform and subsequently combined to get a 100 μM stock solution in methanol (NEFA mix 2). A summary of all fatty acid standards with their exact mass and retention time is presented in Table 1. (*See* **Note 1**.)
3. Octanoic-d_{15} acid, decanoic-d_{19} acid, dodecanoic-d_{23} acid, hexadecanoic-d_{31} acid, octadecanoic-d_{35} acid, 10-pentadecenoic acid, and 10,13-nonadecadienoic acid are used as internal standards and purchased either from Sigma-Aldrich (St. Louis, MO, USA) or from Nu-Chek Prep, Inc. (Elysian, MN, USA). 100 μM individual stock solutions of each compound in methanol are prepared. A 10 μM internal standard solution is prepared by mixing appropriate amount of 100 μM individual stock solutions in methanol.

2.3 LC-MS Instrumentation and Software

1. FTN I-Class UPLC system (Waters Corporation, Milford, MA, USA).
2. Waters ACQUITY UPLC CSH C18, 130 Å, 1.7 μm, 150 × 2.1 mm column equipped with a column in-line filter.
3. Solvent A: 10 mM ammonium acetate in water/acetonitrile (2:3, v/v).

Table 1
Fatty acid standards used in the sample preparation (*see* Note 2)

Fatty acid	Abbreviation	Exact mass [−H⁺]	RT [min]	Dynamic range [μM]
Octanoic acid	C8:0	143.1078	0.91	0.158–63.269
Nonanoic acid	C9:0	157.1234	1.02	0.072–14.415
Decanoic acid	C10:0	171.1391	1.21	0.132–13.242
Hendecanoic acid	C11:0	185.1547	1.47	0.092–3.673
10-Undecenoic acid	C11:1 $\omega - 1$	183.1391	1.16	0.031–12.379
Lauric acid	C12:0	199.1704	1.89	0.142–28.467
11-Dodecenoic acid	C12:1 $\omega - 1$	197.1547	1.41	0.086–17.253
Tridecanoic acid	C13:0	213.186	2.52	0.027–5.321
12-Tridecenoic acid	C13:1 $\omega - 1$	211.1704	1.77	0.054–10.743
Myristic acid	C14:0	227.2017	3.45	0.125–49.941
Myristoleic acid	C14:1 $\omega - 5$	225.186	2.23	0.025–5.038
Pentadecanoic acid	C15:0	241.2173	4.72	0.071–14.115
12-Methyltetradecanoic acid	C15:0-anteiso	241.2173	4.31	0.040–4.021
13-Methyltetradecanoic acid	C15:0-iso	241.2173	4.43	0.040–4.037
Palmitic acid	C16:0	255.233	6.34	0.222–22.239
Palmitoleic acid	C16:1 $\omega - 7$	253.2173	4.09	0.022–4.483
Margaric acid	C17:0	269.2486	8.26	0.042–8.434
14-Methylhexadecanoic acid	C17:0-anteiso	269.2486	7.67	0.040–8.016
15-Methylhexadecanoic acid	C17:0-iso	269.2486	7.84	0.040–8.052
10-Heptadecenoic acid	C17:1 $\omega - 7$	267.233	5.46	0.020–4.050
Stearic acid	C18:0	283.2643	10.35	0.080–8.018
Vaccenic acid	C18:1 $\omega - 7$	281.2486	7.13	0.040–8.075
Oleic acid	C18:1 $\omega - 9$	281.2486	7.26	0.040–20.189
Petroselenic acid	C18:1 $\omega - 12$	281.2486	7.6	0.020–8.139
Linoleic acid	C18:2 $\omega - 6$	279.233	4.98	0.016–16.267
α-Linolenic acid	C18:3 $\omega - 3$	277.2173	3.46	0.020–4.096
γ-Linolenic acid	C18:3 $\omega - 6$	277.2173	3.62	0.012–12.289
Stearidonic acid	C18:4 $\omega - 3$	275.2017	2.52	0.020–4.052
Pristanic acid	C19:0-b	297.2799	11.14	0.020–0.812
Arachidic acid	C20:0	311.29555	14.55	0.029–2.919

(continued)

Table 1
(continued)

Fatty acid	Abbreviation	Exact mass [−H⁺]	RT [min]	Dynamic range [μM]
Phytanic acid	C20:0-b	311.29555	12.39	0.020–0.808
Gondoic acid	C20:1 ω – 9	309.2799	11.11	0.018–7.346
8-Eicosenoic acid	C20:1 ω – 12	309.2799	11.45	0.020–8.116
5-Eicosenoic acid	C20:1 ω – 15	309.2799	11.87	0.020–8.116
11,14-Eicosadienoic acid	C20:2 ω – 6	307.2643	8.23	0.022–11.091
11,14,17-Eicosatrienoic acid	C20:3 ω – 3	305.2486	6.03	0.011–11.164
Dihomogamma linolenic acid	C20:3 ω – 6	305.2486	6.12	
Mead acid	C20:3 ω – 9	305.2486	6.66	0.008–4.079
Arachidonic acid	C20:4 ω – 6	303.233	4.62	0.019–3.746
5,8,11,14,17-Eicosapentaenoic acid	C20:5 ω – 3	301.2173	3.24	0.011–2.263
Erucic acid	C22:1 ω – 9	337.3112	15	0.027–2.695
13,16-Docosadienoic acid	C22:2 ω – 6	335.2956	12.09	0.014–6.778
13,16,19-Docosatrienoic acid	C22:3 ω – 3	333.2799	9.54	0.007–6.818
Adrenic acid	C22:4 ω – 6	331.2643	7.38	0.027–2.744
7,10,13,16,19-Docosapentaenoic acid	C22:5 ω – 3	329.2486	5.41	0.007–6.902
4,7,10,13,16-Docosapentaenoic acid	C22:5 ω – 6	329.2486	5.6	0.008–4.067
4,7,10,13,16,19-Docosahexaenoic acid	C22:6 ω – 3	327.233	3.99	0.017–3.472
9,12,15,18-Tetracosatetraenoic acid	C24:4 ω – 6	359.2956	10.77	0.020–8.067
9,12,15,18,21-Tetracosapentaenoic acid	C24:5 ω – 3	357.2799	8.48	0.008–8.046
6,9,12,15,18-Tetracosapentaenoic acid	C24:5 ω – 6	357.2799	8.81	0.008–4.015
Nisinic acid	C24:6 ω – 3	355.2643	6.68	0.008–0.801

4. Solvent B: 10 mM ammonium acetate in acetonitrile/isopropanol (1:1, v/v).
5. LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany).

6. Xcalibur Software 2.2 SP1 QuanBrowser module (Thermo Fisher Scientific, Bremen, Germany) for targeted quantitative analysis.
7. LipidSearch software (Mitsui Knowledge Industry, Tokyo, Japan) for unknown screening.

2.4 Other Equipment

1. Star robotic unit (Hamilton Bonaduz AG, Bonaduz, Switzerland) equipped with a cooling deck.
2. Thermomixer Comfort C.
3. Eppendorf 5810R centrifuge.
4. Concentrator plus SpeedVac.
5. HTS PAL liquid handler.

3 Methods

3.1 Standard Preparation

In brown glass vials, prepare a series of ten calibrators covering the range of NEFA endogenous concentrations in human plasma (approximately 0.005–70 μM) by mixing different volumes of the two external standard stock solutions NEFA mix 1 and NEFA mix 2 (*see* **Note 3**).

Add 20 μL of 10 μM internal standard solution and adjust the volume of each standard to 1 mL with water/acetonitrile (1/1, v/v).

3.2 Protein Precipitation Reagent Preparation

In a 50 mL volumetric flask, place 55 μL of each of the internal standards 100 μM stock solutions and fill the flask with isopropanol. These 50 mL of precipitation reagent are sufficient for extraction of 96 samples (complete DWP). The final concentration of each internal standard in the extract will be approximately 200 nM.

3.3 Plasma Sample Preparation

Place 50 μL of plasma in a 96 DWP (Eppendorf AG, Hamburg, Germany) and add 450 μL of precipitation reagent to precipitate proteins (*see* **Note 4**).

Shake the plate for 30 min at 700 rpm and centrifuge for 10 min at 453 × *g*.

Collect 150 μL of supernatant and transfer it to a 96-well PCR plate (Eppendorf AG, Hamburg, Germany).

Evaporate solvent under vacuum and reconstitute in 75 μL acetonitrile/water (1:1, v/v).

Shake the plate for 5 min at 453 × *g* in the thermomixer maintained at 4 °C.

Seal the plate with an aluminum foil and place it in the autosampler for analysis.

3.4 Liquid Chromatography Analysis

Set autosampler temperature to 4 °C and column oven temperature to 55 °C. Inject 1 μL of calibrator solutions or plasma extract. Set flow rate to 450 μL/min and perform separation using the following binary gradient: starting conditions 90% solvent A, 10% B for 2 min. Linearly increase solvent B to 46% over 12 min and subsequently to 100% over 3.5 min. Rinse column at 100% B for 3 min, return to initial conditions, and allow equilibration for 2.5 min.

3.5 Mass Spectrometry Settings

Perform analysis using heated electrospray ionization in negative FTMS mode over the mass range of 110–380 Da with a resolution of 60,000. Set spray voltage to −3 kV; heater and capillary temperatures to 300 °C and 350 °C, respectively; and sheath and auxiliary gas flow rates to 35 AU and 10 AU, respectively (*see* **Note 5**).

3.6 Mass Spectrometric Data Processing

Perform signal detection, integration, and quantification using the following parameters: mass tolerance 5 ppm, retention time window 30 s, peak detection algorithm ICIS, smoothing points 3, and tailing factor 1.5. If necessary, manually modify peak integration.

Generate calibration curves by plotting the peak area ratio versus the expected concentration. Curves should be linearly fitted with a weighting factor of $1/x$ and ignoring the origin. For each NEFA calibration curve, select a minimum of six—out of the ten—calibrators according to the range of physiological concentrations. The acceptance criteria are a minimum correlation coefficient (R^2) of 0.99 and an accuracy of 85–115% compared to the nominal value.

For unknown NEFA screening, provide the software program with experiment parameters such as formed adducts ($[M-H^+]$), type of instrument (Orbitrap), mass tolerance (5 ppm), and selected class of lipid (fatty acids) and upload raw data files. Once processing is over, examine extracted chromatograms to confirm proper peak picking. Eventually each potential hit has to be confirmed by comparison with a known standard (*see* **Note 6**).

4 Notes

1. In addition to the species shown in the Table 1, GLC-566 also contains behenic acid (C22:0), tricosanoic acid (C23:0), lignoceric acid (C24:0), and nervonic acid (C24:1 $\omega - 9$). However, they have been excluded from the presented method since it was not possible to quantify them accurately.
2. Fatty acids are named according to the number of carbon atoms and number of double bonds in their aliphatic chain, using the convention "C number of carbon/number of double bond." To differentiate isomers, the position of the first double bond is indicated by "ω–x" where x is the first double bond carbon atom from the methyl end of the chain. For instance,

C16:0 and C18:3 $\omega - 3$ stand for palmitic acid and α-linolenic acid, respectively. For branched fatty acids, the position of the methyl group is indicated by the suffix iso or anteiso.

3. ETA and dihomogamma linolenic acid (C20:3 $\omega - 3/6$) are not chromatographically separated and thus can't be quantified individually. Their concentration has to be reported as the sum of their individual concentrations.
4. For better accuracy in calibrator preparation, it is recommended to use a HTS PAL liquid handler for standard preparation.
5. All sample pipetting steps of plasma sample preparation are performed using a Hamilton Star robotic. With this platform, extraction of 96 samples is typically performed within 3 h. To avoid sample degradation, the temperature of the liquid handler, shaker, and centrifuge is maintained at 4 °C.
6. The mass spectrometer is calibrated every 4 days following manufacturer specifications. If a better mass accuracy is required for low masses, trifluoroacetate ([M−H⁺] 112.985590 m/z) can be used as an additional signal for mass calibration of the MS instrument in negative mode. Trifluoroacetate is a background ion, and its signal is commonly detected in commercial calibration solution.
7. Both screening and quantitative measurements can be performed simultaneously, without affecting each other. Potentially, NEFA having 8–24 carbon atoms in their chain can be screened.

References

1. Bazinet RP, Laye S (2014) Polyunsaturated fatty acids and their metabolites in brain function and disease. Nat Rev Neurosci 15(12):771–785. doi:10.1038/nrn3820
2. Boden G (2008) Obesity and free fatty acids. Endocrinol Metab Clin North Am 37(3):635–646 . doi:10.1016/j.ecl.2008.06.007viii-ix
3. Hirasawa A, Tsumaya K, Awaji T, Katsuma S, Adachi T, Yamada M, Sugimoto Y, Miyazaki S, Tsujimoto G (2005) Free fatty acids regulate gut incretin glucagon-like peptide-1 secretion through GPR120. Nat Med 11(1):90–94. doi:10.1038/nm1168
4. Kahn SE, Hull RL, Utzschneider KM (2006) Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature 444(7121):840–846. doi:10.1038/nature05482
5. Li M, Yang L, Bai Y, Liu H (2014) Analytical methods in lipidomics and their applications. Anal Chem 86(1):161–175. doi:10.1021/ac403554h
6. Christie WW (1998) Gas chromatography–mass spectrometry methods for structural analysis of fatty acids. Lipids 33(4):343–353
7. Pettinella C, Lee SH, Cipollone F, Blair IA (2007) Targeted quantitative analysis of fatty acids in atherosclerotic plaques by high sensitivity liquid chromatography/tandem mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci 850(1-2):168–176. doi:10.1016/j.jchromb.2006.11.023
8. Kamphorst JJ, Fan J, Lu W, White E, Rabinowitz JD (2011) Liquid chromatography-high resolution mass spectrometry analysis of fatty acid metabolism. Anal Chem 83(23):9114–9122. doi:10.1021/ac202220b
9. Lacaze JP, Stobo LA, Turrell EA, Quilliam MA (2007) Solid-phase extraction and liquid chromatography--mass spectrometry for the determination of free fatty acids in shellfish. J Chromatogr A 1145(1-2):51–57. doi:10.1016/j.chroma.2007.01.053

10. Hu C, van der Heijden R, Wang M, van der Greef J, Hankemeier T, Xu G (2009) Analytical strategies in lipidomics and applications in disease biomarker discovery. J Chromatogr B 877(26):2836–2846. doi:10.1016/j.jchromb.2009.01.038
11. Li X, Franke AA (2011) Improved LC-MS method for the determination of fatty acids in red blood cells by LC-orbitrap MS. Anal Chem 83(8):3192–3198. doi:10.1021/ac103093w
12. Hellmuth C, Weber M, Koletzko B, Peissner W (2012) Nonesterified fatty acid determination for functional lipidomics: comprehensive ultrahigh performance liquid chromatography-tandem mass spectrometry quantitation, qualification, and parameter prediction. Anal Chem 84(3):1483–1490. doi:10.1021/ac202602u
13. Hellmuth C, Demmelmair H, Schmitt I, Peissner W, Bluher M, Koletzko B (2013) Association between plasma nonesterified fatty acids species and adipose tissue fatty acid composition. PLoS One 8(10):e74927. doi:10.1371/journal.pone.0074927
14. Fugmann M, Uhl O, Hellmuth C, Hetterich H, Kammer NN, Ferrari U, Parhofer KG, Koletzko B, Seissler J, Lechner A (2015) Differences in the serum nonesterified Fatty Acid profile of young women associated with a recent history of gestational diabetes and overweight/obesity. PLoS One 10(5):e0128001. doi:10.1371/journal.pone.0128001
15. Christinat N, Morin-Rivron D, Masoodi M (2016) High-Throughput Quantitative Lipidomics Analysis of Nonesterified Fatty Acids in Human Plasma. J Proteome Res 15(7):2228–2235. doi:10.1021/acs.jproteome.6b00198

Chapter 15

Simultaneous Enrichment of Plasma Extracellular Vesicles and Glycoproteome for Studying Disease Biomarkers

Sunil S. Adav and Siu Kwan Sze

Abstract

To detect disease at an early stage and to develop effective disease treatment therapies, reliable biomarkers of diagnosis, disease progression, and its status remain a research priority. A majority of disease pathologies are primarily associated with different subsets of cells of different tissues, discrete compartments, and areas. These subsets of cells release glycoproteins and specific extracellular vesicles (EVs) including microvesicles and exosomes that carry bioactive cargoes of proteins, nucleic acids, and metabolites. Body fluids like blood plasma are considered as a golden source of disease biomarkers since it contains glycoprotein and EVs released by almost all cell types. The contents of glycoproteome and EV cargo change with cell status, and they act as mirror of cell's intracellular events and status; hence, EVs and glycoproteins are promising disease biomarkers. However, their abundance in blood plasma remains low posing a serious technical problem in their identification and quantification. Until recently, technical advances and exhaustive research devised a technique for either enrichment of plasma glycoprotein or EVs, but no methodologies exist that can enrich and identify both plasma glycoprotein and EVs. To overcome this technical challenge, a method that can eliminate high-abundance entities without depleting disease-modifying molecules is required. Therefore, here we describe the detailed protocol of simultaneous enrichment of glycoproteins and EVs from blood plasma by prolonged ultracentrifugation coupled to electrostatic repulsion-hydrophilic interaction chromatography (PUC-ERLIC) and their identification and quantification by mass spectrometry-based proteomic technique.

Key words Extracellular vesicle, Glycoprotein, Disease biomarker, Prolonged ultracentrifugation, Electrostatic repulsion-hydrophilic interaction chromatography, Proteomics, LC-MS/MS

1 Introduction

Blood plasma is considered as a golden source of disease biomarkers since it contains proteins, glycoproteins, secretory molecules, and membrane-derived vesicles released from various organs and tissues during both healthy and disease conditions. Membrane-derived vesicles also called as EVs play vital roles in a plethora of processes including cellular communication, the maintenance of homeostasis, and the development and progression of pathologic conditions like cancer [1] and are considered as a promising source

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_15,

of biomarkers of diagnostic and prognostic value [2]. EVs affect the physiology of neighboring recipient cells through intracellular signaling and play a major role in immune regulation and cell-cell communication through their cargos [3]. The role of EVs in numerous physiological processes and pathological disorders has been implicated rendering them as promising targets for clinical biomarker discovery [4, 3]. Being as less invasive body fluid, blood plasma remains an incredible source of disease biomarkers, but high-abundance proteins like albumin, hemoglobin, fibrinogen, etc. present a barrier for detection of medium- and low-abundance EV protein biomarkers in plasma. Hence, technologies that get rid of such high-abundance proteins without reducing disease-relevant proteins or other secretory molecules could be a perfect approach for detection of a diagnostic biomarker.

Protein glycosylation remains one of the most important post-translational modifications of proteins secreted in plasma comprising up to half of all circulating proteins, and it has been linked to protein folding, quality control, sorting, degradation, and secretion [5, 6]. A majority of secretory and membranous proteins that have been detected in plasma are glycoproteins; therefore, the plasma glycoproteome is one of the major sub-proteomes that is highly enriched with disease biomarkers. Again, plasma glycoproteome has significant clinical value, as most secreted biomarkers are glycosylated [7, 8], e.g., biomarkers including HER2 in breast cancer, PSA in prostate cancer, CEA in colorectal cancer, CA-125 in ovarian cancer, and alpha-fetoprotein in hepatocellular carcinoma have been found to be glycosylated [4, 9]. Further, the distribution and degree of glycosylation of glycoproteins is significantly altered by several diseases, and hence quantitative analysis of plasma glycoproteins could be valuable in biomarker discovery study. But the glycoproteomic study has been hampered by technical challenges like difficulties in identification of glycoproteins in complex samples. Hence, an ideal technical approach for identification of disease biomarkers needs to be directed to glycosylated proteins and to wide diverse types of EVs.

Both glycoproteins and EVs have tremendous potential in diagnostic biomarkers; however, they are presented in low abundance in blood plasma. Therefore, sensitive proteomic technology coupled with effective enrichment strategies that selectively isolate the targeted molecules by getting rid of background plasma proteins is necessary to facilitate the identification of the low-abundance biomarkers. Considering the great importance and persistent interest of glycoproteins and EVs in disease biomarkers and drug targeting, simultaneous enrichment of both components from blood plasma by prolonged ultracentrifugation-electrostatic repulsion-hydrophilic interaction chromatography (PUC-ERLIC) method coupled to LC-MS/MS analysis has been developed [10, 11]. Using PUC-ERLIC methodology, soluble proteins and aggregated proteins

from human brain tissue have also been explored [12]. Thus, PUCERLIC enrichment method coupled to mass spectrometry-based proteomic analysis facilitates identification of low-abundance aggregated proteins, glycoproteins, and EVs in clinical samples.

2 Materials

1. Prepare all solutions using ultrapure water (Milli-Q, prepared by purifying deionized water, to attain a sensitivity of 18 MΩ cm at 25 °C).

2.1 Glycoprotein and EV Purification and Protein Digestion

1. Plasma samples from healthy controls and patients.
2. 1× phosphate buffer saline (PBS): Prepare 1× PBS by diluting commercially available 10× PBS.
3. 25 × 89 mm polycarbonate ultracentrifuge tubes and Beckman L100-XP Ultracentrifuge (Beckman Coulter, Brea, CA) with type 50.2 Ti rotor.
4. Lysis buffer: 8 M urea and 50 mM ammonium acetate (pH 6.0).
5. Cryo-EM and western blotting system.
6. Bicinchoninic acid (BCA) protein assay kit and spectrophotometer.
7. Ammonium acetate buffer (50 mM): Add 25 ml water to 100 ml bottle by using a 50 ml graduated measuring cylinder. Weigh 0.192 g ammonium acetate, transfer to glass bottle, dissolve it, adjust pH 6.0, and make up the volume to 50 ml by water.
8. 100 mM dithiothreitol (DTT): Prepare DTT by dissolving 0.015 g DTT into 1 ml 50 mM ammonium acetate buffer (pH 6.0) (*see* **Note 1**).
9. 0.5 M iodoacetamide (IAA): Prepare IAA by dissolving 0.09 g IAA in ammonium acetate buffer (*see* **Note 1**).
10. Enzyme: Sequencing-grade trypsin (Promega, Madison, USA; catalogue number: V5111)
11. 10% formic acid (FA): Take 100 µl FA and dilute it to 1 ml by Milli-Q water
12. Vacuum centrifuge, i.e., SpeedVac.

2.2 Desalting of Tryptic Peptides

1. Sep-Pak C18 cartridge (Waters, Milford, MA).
2. Methanol (analytical grade).
3. 0.1% trifluoroacetic acid: Dilute 10 µl TFA to 10 ml by Milli-Q water.
4. Vacuum centrifuge, i.e., SpeedVac.

2.3 ERLIC Fractionation and Enrichment of Glycopeptides

1. Mobile phase A: 80% ACN containing 0.1% FA.
2. Mobile phase B: 30% ACN containing 2% FA.
3. PolyWAX LP weak anion-exchange HPLC column (4.6 × 200 mm, 5 μm, 300 Å; PolyLC).
4. Vacuum centrifuge, i.e., SpeedVac.
5. Enzyme: PNGase F (New England Bio Labs, Beverly, MA). Prepare it in 50 mM ammonium acetate (pH 5.0).

2.4 LC-MS/MS Analysis

1. Mobile phase A (0.1% FA in HPLC water) and mobile phase B (0.1% FA in ACN).
2. LTQ-FT Ultra linear ion trap mass spectrometer (Thermo Fisher Scientific Inc., Bremen, Germany) or any suitable LC-MS/MS system with Dionex or other suitable HPLCs with autosampler.
3. Zorbax peptide trap column (Agilent Technologies, Santa Clara, CA) or any suitable peptide trap for online peptide concentration and desalting.
4. Capillary column (75 μm × 10 cm) packed with C18 AQ (5 μm, 300 Å; Bruker-Michrom, Billerica, MA) or any suitable C18 capillary column for peptide separation.
5. ADVANCE™ CaptiveSpray™ source (Bruker-Michrom) or any nano-electrospray.

2.5 Mass Spectrometric Data Analysis

1. Proteome Discoverer™ (PD, version 1.4 software) and Mascot Server (version 2.4.1, Matrix Science, Boston, MA) or any protein sequence database search software.
2. UniProt human database released on or after 29 November 2013 (can be downloaded from http://www.uniprot.org/proteomes/UP000005640).

3 Methods

3.1 Glycoprotein and EV Purification by PUC and Protein Digestion

1. Centrifuge blood (approximately 2–3 ml) collected into vacutainer tube containing anticoagulant [ethylenediaminetetraacetic acid, (EDTA)] tubes for 30 min at 2000 × *g* to remove cells. Collect plasma without disturbing pellet and preserve it at −20 or −80 °C until subsequent proteomic processing.
2. Take plasma samples from freezer and thaw them on ice. Take plasma (1 ml) sample and dilute it to 5 ml with 1× PBS buffer and centrifuge at 3000 × *g* for 10 min to remove intact cells and cellular debris (*see* **Note 2**).
3. Transfer the supernatant into a polycarbonate ultracentrifuge tube and ultracentrifuge at 200,000 × *g* for 18 h at 4 °C using ultracentrifuge.

4. Collect supernatant and pellet. Pellet contains enriched glycoproteins and EVs while supernatant contains soluble proteins. Resuspend the pellet in 1× PBS and ultracentrifuge again at 200,000 × *g* (18 h, 4 °C) to remove residual contaminants.
5. Resuspend the enriched secretory and extracellular vesicle-enriched glycoproteins in 0.5 ml lysis buffer. The size and enrichment of EVs were evaluated by cryo-EM (*see* Fig. 1a, b) and western blotting (Fig. 1c). Both soluble and resuspended pellet samples can be processed for proteomic analysis.
6. Quantify protein content of samples in 96-well plate or in Eppendorf tube by adopting the BCA spectrometric assay technique.
7. Reduce disulfide bonds by incubating 300 μg protein in 20 mM DTT (if sample volume is 50 μl, then add 10 μl 100 mM DDT to bring its final concentration to 20 mM) for 3 h at 30 °C (*see* **Note 3**).
8. Then, alkylate the protein samples in the dark using 55 mM IAA (if sample volume is 50 μl, then add 5 μl 0.5 M IAA to bring its final concentration to 55 mM) for 1 h at room temperature (*see* **Note 4**).
9. Dilute the reduced and alkylated protein sample using 50 mM ammonium acetate buffer (pH 6.0) to bring down urea concentration to or below 1 M (*see* **Note 5**).

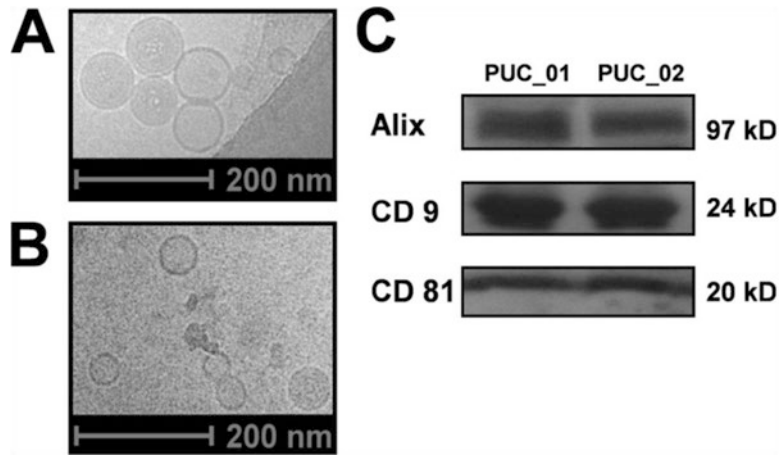


Fig. 1 Plasma extracellular vesicles isolated by ultracentrifugation [(**a**) Electron micrographs of harvested extracellular vesicles (sizes 50–100 nm, recorded at 23,500× magnification with a defocus of −6 μM). (**b**) Electron micrographs of extracellular vesicles recorded on carbon (23,500× magnification with a defocus of −6 μM). (**c**) Western blot analyses of harvested proteins using extracellular vesicle markers Alix, CD9, and CD81 (Adapted from Cheow et al. [11])]

10. Add sequencing-grade trypsin at a 1:100 ratio (w/w, trypsin: protein) and incubate at 37 °C for overnight. Stop the reaction by adding 10% FA to bring pH down to 2.0.
11. Dry the peptides using vacuum centrifuge, i.e., SpeedVac (*see* **Note 6**).

3.2 Desalting of Tryptic Peptides

1. Activate the Sep-Pak C18 cartridge column with two column volumes of methanol (*see* **Note 7**).
2. Equilibrate the column with two column volumes of 0.1% TFA.
3. Dissolve the dried tryptic peptides into 1 ml of 0.1% TFA and load peptide solution onto the cartridge using a pipette.
4. Wash the cartridge with two to four column volumes of 0.1% TFA.
5. Elute the peptides with 1 ml 70% ACN containing 0.1% TFA.
6. Concentrate the eluent-containing tryptic peptides using vacuum concentrator (SpeedVac) till dryness.

3.3 ERLIC Fractionation and Enrichment of Glycopeptides

1. Prepare mobile phase A and B.
2. Dissolve vacuum-dried peptides in 200 µl mobile phase A (*see* **Note 8**).
3. Connect PolyWAX LP weak anion-exchange column on a HPLC system and condition it by mobile phase B for 30 min and then mobile phase A for 30 min at 1 ml/min flow rate.
4. Inject the peptides reconstituted in 200 µl mobile phase A and establish the 60 min gradient, starting with 3 min of 100% A, 17 min of 0–8% B, 25 min of 8–45% B, and 10 min of 45–100% B, followed by 5 min at 100% B at constant flow rate of 1 ml/min. Collect 60 fractions using a fraction collector.
5. Record the UV spectra of the peptides at 280 nm.
6. Combine fractions into 15 pooled fractions and dry them using vacuum centrifuge. Reconstitute the peptides in 60 µl 50 mM ammonium acetate (pH 5.0) buffer to minimize experimentally induced deamidation [13, 14]. Dilute 1 unit of PNGase F (1 unit of PNGase F can be used for 1 mg protein) with 50 mM ammonium acetate buffer and add 2 µl to each fraction to deglycosylate the peptides. Incubate this reaction mixture at 37 °C for 6 h and then dry them in a vacuum concentrator prior to reconstitution in a solvent (3% ACN, 0.1% FA) for LC-MS/MS analysis.

3.4 LC-MS/MS Analysis

1. Prepare mobile phases A and B.
2. Inject approximately 2 µg of peptides from each fraction via the Dionex autosampler, concentrate it into a Zorbax peptide trap column, and subsequently separate in a capillary column packed with C18. LC-MS/MS utilized LTQ-FT Ultra linear

ion trap mass spectrometer (Thermo Fisher Scientific Inc., Bremen, Germany) coupled with a Dionex Ultimate 3000 RSLCnano system in our study (*see* **Note 9**).

3. Maintain the flow rate at 300 nl/min. Establish a 60 min gradient using mobile phases A and B, starting with 1 min of 5–8% B, 44 min of 8–32% B, 7 min of 32–55% B, 1 min of 55–90% B, and 2 min of 90% B, followed by re-equilibration in 5% B for 5 min.
4. Ionize the peptide sample using an electrospray potential of 1.5 kV in an ADVANCE™ CaptiveSpray™ source (Bruker-Michrom). Set LTQ-FT Ultra to perform data acquisition in the positive ion mode. Record full MS scan (350–1600 m/z range) in the FT-ICR cell at a resolution of 100,000 and a maximum ion accumulation time of 1000 ms. Set an automatic gain control (AGC) target for FT at 1×10^6, and activate precursor ion charge state screening. Use a linear ion trap to collect peptides and measure the fragments generated by collision-induced dissociation (CID). Use default AGC setting in the linear ion trap (full MS target: 3.0×10^4, MSn: 1×10^4). Select ten most intense ions above a 500-count threshold for MS2 fragmentation by CAD. Use a maximum ion accumulation time of 200 ms. Activate dynamic exclusion for this process. For CID, set the activation Q at 0.25, and activation time 30 ms, isolation width (m/z) 2.0, and normalized collision energy 35%.

3.5 Mass Spectrometric Data Analysis

1. Mass spectrometric data analysis uses Proteome Discoverer™ (PD, version 1.4 software), connected to Mascot Server (version 2.4.1, Matrix Science, Boston, MA) in our study.
2. Use UniProt human database released on or after 29 November 2013 that contains 88,421 sequences and 35,070,517 residues or higher improved version of dataset. Human protein database in FASTA format can be downloaded from http://www.uniprot.org/proteomes/UP000005640.
3. Use target-decoy search strategy for estimation of false discovery rate (FDR).
4. Consider peptides identified with false discovery rate (FDR) of <1% for further analysis.
5. Restrict the search to a maximum of two missed trypsin cleavages, peptide precursor mass tolerances of 5.1 ppm, and 0.8 Da mass tolerances for fragment ions.
6. Set carbamidomethylation (+57.021 Da) of cysteine residues as fixed modification.
7. Set oxidation (+15.995 Da) of methionine residues and deamidation (+0.984 Da) of asparagine and glutamine residues as dynamic peptide modifications.

8. To maximize accuracy, calculate protein and peptide-relative quantities based on the average area of the three most abundant unique peptides per protein.
9. Export mascot searched data to csv file format and then further process it in Microsoft Excel.

4 Notes

1. The prepared solutions of DTT and IAA are stable for 1 month if stored at −20 °C. However, IAA is light sensitive, and hence fresh preparation of both reducing and alkylation reagents is recommended for better results.
2. Use a clean working table bench and lab coat and wear a mask when handling samples for proteomics to avoid keratin contamination.
3. Most protocol reduces the protein samples by DTT at 56 °C for 60 min. However, at higher temperature, the urea molecule decomposes to cyanic acid which further may react with side chains of lysine and arginine and N-terminal amino groups to form carbamylated residues. So, longer reduction time (2 h) at lower temperature (37 °C) ensures the complete reduction and avoids artifactual modifications of amino acids.
4. IAA is light sensitive and needs to be protected from light. Use aluminum foil to cover the tube that contains prepared IAA. After adding IAA into a protein sample, cover the samples by aluminum foil or incubate the samples at a dark place.
5. Urea is used as the denaturant because of its dispersive action that help to solubilize insoluble protein. Trypsin digests denatured proteins much more readily than proteins in their native form. However, high concentration of urea reduces digestion efficiency of trypsin; hence, it's necessary to dilute the test solution to bring final urea concentration to 1 M or below 1 M.
6. During drying the peptide sample, to avoid temperature-induced artifactual peptide modifications, do not set the temperature of vacuum centrifuge above 30 °C.
7. Sep-Pak C18 Vac cartridges contain a hydrophobic, reversed-phase, silica-based bonded phase that has potential to adsorb analytes of even weak hydrophobicity from aqueous solutions. These syringe barrel-type cartridges can be used with vacuum manifolds; however, for better results do not apply vacuum during sample loading.

8. Add 200 μl mobile phase A to dried peptides and then mix it properly on vortex for 10 min. Then centrifuge at 15,000 × *g* and take supernatant for HPLC fractionation.
9. The peptides can be analyzed using any suitable LC-MS/MS system. It is not restricted by LTQ-FT that was used in the original study.

Acknowledgment

This work is in part supported by the Singapore Ministry of Education (MOE-Tier 2 ARC9/15 and MOE-Tier 1 RGT15/13) and the NTU-NHG Ageing Research Grant (ARG/14017).

References

1. Ahn YH, Kim KH, Shin PM, Ji ES, Kim H, Yoo JS (2012) Identification of low-abundance cancer biomarker candidate TIMP1 from serum with lectin fractionation and peptide affinity enrichment by ultrahigh-resolution mass spectrometry. Anal Chem 84(3):1425–1431
2. Revenfeld AL, Baek R, Nielsen MH, Stensballe A, Varming K, Jorgensen M (2014) Diagnostic and prognostic potential of extracellular vesicles in peripheral blood. Clin Ther 36(6): 830–846
3. Thery C, Ostrowski M, Segura E (2009) Membrane vesicles as conveyors of immune responses. Nat Rev Immunol 9(8):581–593
4. Chaput N, Thery C (2011) Exosomes: immune properties and potential clinical implementations. Semin Immunopathol 33(5):419–440
5. Rudd PM, Dwek RA (1997) Glycosylation: heterogeneity and the 3D structure of proteins. Crit Rev Biochem Mol Biol 32(1):1–100
6. Kuster B, Krogh TN, Mortz E, Harvey DJ (2001) Glycosylation analysis of gel-separated proteins. Proteomics 1(2):350–361
7. Berven FS, Ahmad R, Clauser KR, Carr SA (2010) Optimizing performance of glycopeptide capture for plasma proteomics. J Proteome Res 9(4):1706–1715
8. Seelenmeyer C, Wegehingel S, Lechner J, Nickel W (2003) The cancer antigen CA125 represents a novel counter receptor for galectin-1. J Cell Sci 116(Pt 7):1305–1318
9. Cheow ES, Cheng WC, Lee CN, de Kleijn D, Sorokin V, Sze SK (2016) Plasma-derived extracellular vesicles contain predictive biomarkers and potential therapeutic targets for myocardial ischemic injury. Mol Cell Proteomics 15(8):2628–2640
10. Adav SS, Hwa HH, De Kleijn D, Sze SK (2015) Improving blood plasma glycoproteome coverage by coupling ultracentrifugation fractionation to electrostatic repulsion-hydrophilic interaction chromatography enrichment. J Proteome Res 14(7):2828–2838
11. Cheow ESH, Sim KH, De Kleijn D, Lee CN, Sorokin V, Sze SK (2015) Simultaneous enrichment of plasma soluble and extracellular vesicular glycoproteins using prolonged ultracentrifugation-electrostatic repulsion-hydrophilic interaction chromatography (PUC-ERLIC) approach. Mol Cell Proteomics 14(6):1657–1671
12. Adav SS, Gallart-Palau X, Tan KH, Lim SK, Tam JP, Sze SK (2016) Dementia-linked amyloidosis is associated with brain protein deamidation as revealed by proteomic profiling of human brain tissues. Mol Brain 9(1):20
13. Hao P, Ren Y, Alpert AJ, Sze SK (2011) Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. Mol Cell Proteomics 10(10): O111 009381
14. Hao P, Ren Y, Datta A, Tam JP, Sze SK (2015) Evaluation of the effect of trypsin digestion buffers on artificial deamidation. J Proteome Res 14(2):1308–1314

Chapter 16

Lipidomics of Human Blood Plasma by High-Resolution Shotgun Mass Spectrometry

Susanne Sales, Oskar Knittelfelder, and Andrej Shevchenko

Abstract

Clinical lipidomics is an emerging biomarker discovery approach that compares lipid profiles under pathologically and physiologically normal conditions. Here we describe a method for the absolute (molar) quantification of more than 200 molecules from 14 major lipid classes from 5 μL of human blood plasma using high-resolution top-down shotgun mass spectrometry. Because of its technical simplicity and robustness, the protocol lends itself for high-throughput clinical lipidomics screens.

Key words Blood plasma, Lipids, Shotgun lipidomics, LipidXplorer, Mass spectrometry

1 Introduction

Blood plasmaanalysis is a basic method of clinical chemistry, laboratory diagnostics, and, since recently, biomarker discovery. A typical blood test may report more than 30 clinically relevant indices; however, only four of them (total triacylglycerols (TAG), total cholesterol (Chol), and the cholesterol content in HDL and LDL fractions) are directly reflecting the status of lipid homeostasis. Human blood plasma is now being extensively studied by lipidomics (reviewed in [1, 2]) and currently the most exhaustive analysis performed by LIPID MAPS consortium determined the molar concentration of 588 individual lipids from 21 major lipid classes [2]. Because of its unique molecular specificity, sensitivity and throughput lipidomics was employed in clinical screens to identify individual lipids and lipid classes whose plasma concentration was specifically affected by obesity [3], type 1 [4] and type 2 [5] diabetes, insulin resistance [6], hypertension [7], cardiovascular disease [8, 9], Alzheimer's disease [10], and schizophrenia [11, 12]. Correlating full lipidome profiles with the clinical status of patients

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_16,

and dynamics of disease progression shed light on molecular mechanisms of complex metabolic disorders and the role of covariate pathophysiological, genetic, and dietary factors. It also leads to the identification of promising molecular biomarkers for early diagnostics of metabolic disorders, personalized evaluation of disease severity and clinical prognosis, and monitoring the individual response toward dietary, therapeutic, or surgery treatment [13, 14, 15, 16].

Clinical lipidomics is an emerging field (reviewed in [17]), and standard operation procedures for quantifying lipids in biofluids and biopsies are yet to be established. Lipids can be identified and quantified by various means of mass spectrometry [18]. However, advances in the analytical instrumentation do not circumvent the need in complete, quantitative, and unbiased extraction of lipids from clinically relevant biomaterials. Quantification of recovered molecules is another critical step in a lipidomics pipeline. Clinical screens often report fold changes in abundances of lipids in samples from patients in comparison to an arbitrary control cohort. However, there is a considerable advantage in reporting absolute (molar) concentrations of individual lipids. Contrary to fold changes, molar concentrations determined in different projects and laboratories could be directly compared, which improves the consistency of clinically important findings and their concordance with common clinical chemistry indices. Once made available in a public domain along with the relevant anthropometric and clinical chemistry indices of study cohort members, molar concentration of individual lipids is a valuable resource that could be independently interpreted in various biological and clinical contexts. However, absolute (molar) quantification is technically challenging and critically depends on the availability and quality of internal standards [19]—typically, synthetic lipids of the same lipid classes comprising rare or unnatural fatty acid/fatty alcohol moieties that do not occur in the analyzed samples [20]. Therefore, to achieve robust and consistent absolute quantification, a lipidomics pipeline should rely upon an independent validation of the chemical purity and molar concentration of stocks of internal standards made from different batches of commercial synthetic lipids.

Here we present a shotgun lipidomics protocol to quantify the molar concentration of 207 molecules from 14 major lipid classes including glycerophospholipids, glycerolipids, sphingolipids, cholesterol esters, and free cholesterol in human plasma [21]. It relies upon lipid extraction by methyl *tert*-butyl ether followed by direct infusion of total lipid extracts into a high-resolution tandem mass spectrometer Q Exactive. Lipids are identified and quantified in FT MS spectra acquired from batches of plasma extracts by LipidXplorer software [22].

2 Materials and Lipid Standards

Use ACS or LC-MS grade solvents for lipid extraction and analysis; they could be stored at RT. Synthetic lipid standards were purchased from Avanti Polar Lipids, Inc. (Alabaster AL) or Sigma–Aldrich Chemie (Munich, Germany). Here lipid species are annotated by their classes and the number of carbon atoms and double bonds in their fatty acid/fatty alcohol moieties. Acronyms for lipid classes are as follows: cholesterol (Chol), cholesteryl ester (CholE), triacylglycerol (TAG), diacylglycerol (DAG), phosphatidylcholine (PC), *lyso*-phosphatidylcholine (LPC), phosphatidylcholine ether (PC O-), *lyso*-phosphatidylcholine ether (LPC O-), phosphatidylethanolamine (PE), *lyso*-phosphatidylethanolamine (LPE), phosphatidylethanolamine ether (PE O-), phosphatidylinositol (PI), sphingomyelin (SM), and ceramide (Cer).

2.1 Solvent Mixtures

Methyl *tert*-butyl ether (MTBE)/methanol (MeOH) (5:1.5 (v/v)) was used for lipid extraction.

7.5 mM ammonium formate dissolved in isopropanol/methanol/chloroform (4:2:1 (v/v/v)) (further termed MS mix) was used for nanoflow electrospray analyses of lipid extracts using a TriVersa NanoMate robotic ion source (Advion BioSciences, Ithaca NY).

2.2 Internal Standards for Lipid Quantification

2.2.1 Quantitative LIPID MAPS Standards (QLMS)

QLMS of TAG 54:3, DAG D5 mix II, PC 31:1, PE 31:1, PI 31:1, SM 30:1:2, and Cer 35:1:2 are supplied by the manufacturer in sealed glass ampoules as methanol solutions with the exactly known concentration and stored at −20 °C until used. After opening the ampoules unused standards were transferred into a 2 mL glass vial (Supelco, Bellefonte, PA) and stored at −20 °C.

2.2.2 Internal Standards (IS)

IS for quantifying 14 lipid classes were prepared as described in Subheading 3.2 and Table 1 (*see* **Note 1**).

2.3 Blood Plasma

EDTA plasma was prepared by 10 min centrifugation at 4 °C and 3000 × *g* of blood samples collected after overnight fasting. Upon collection, plasma samples were immediately shock-frozen in liquid nitrogen and stored at −80 °C until analyzed (*see* **Note 2**).

3 Methods

The shotgun quantification workflow is shown in Fig. 1; all operations should be carried out at 4 °C in a cold room and samples and solvents stored on ice, unless specified otherwise.

Table 1
Internal standards for plasma lipid quantification

Lipid class	Internal standard	Solvent for making the standard stock solution	Amount in 700 μL MTBE/MeOH [pmol]	Ionization mode	Molecular ion	*m/z*
CholE	CholE 12:0	1:1 Heptane/isopropanol (v/v)	6199	Positive	$[M+NH4]^+$	586.5563
Chol	Chol D7	1:1 Heptane/isopropanol (v/v)	4743	Positive	$[M+NH4]^+$	411.4319
TAG	TAG 36:0	1:1 Heptane/isopropanol (v/v)	1720	Positive	$[M+NH4]^+$	656.5829
DAG	DAG 24:0	1:1 Heptane/isopropanol (v/v)	366	Negative	$[M+HCOO]^-$	483.3685
SM	SM 30:1:2	Isopropanol + 1% H_2O	712	Negative	$[M+HCOO]^-$	691.5019
PC	PC 25:0	Isopropanol + 1% H_2O	1987	Negative	$[M+HCOO]^-$	680.4502
PE	PE 25:0	54:16:6 MTBE/MeOH/H_2O (v/v)	272	Negative	$[M-H]^-$	592.3978
PI	PI 32:0	Isopropanol + 1% H_2O	195	Negative	$[M-H]^-$	809.5180
LPC	LPC 13:0	Isopropanol + 1% H_2O	487	Negative	$[M+HCOO]^-$	498.2831
LPE	LPE 13:0	54:16:6 MTBE/MeOH/H_2O (v/v)	425	Negative	$[M-H]^-$	410.2307

3.1 Preparation of Internal Standards

3.1.1 Validation of the Concentration of Internal Standard Stocks

"Quantitative LIPID MAPS standards" (QLMS) are produced, aliquoted, quantified, and shipped in sealed vials by Avanti Polar Lipids; however, they are expensive for using in large-scale clinical lipidomics screens. At the same time, concentrations calculated from the weighted amount of dry powder of commercial lipids are often inconsistent and lead to systematic quantification errors, especially if the measurements rely on several stocks of standards prepared at different times. A practical solution described here is to employ QLMS to determine the exact lipid concentrations in larger volumes of self-prepared stocks of lipid standards and use them for quantifying plasma lipids.

1. Pipette a volume equivalent to 50 pmol of the QLMS into 5 wells of a 96-well plate (Eppendorf, Hamburg, Germany).
2. To each well, add different amounts (10, 25, 50, 75, and 100 pmol) of the quantified standard; mix and determine its concentration by shotgun analysis relative to the known concentration of QLMS. Concentrations determined in five independent measurements are averaged (*see* **Note 3**).

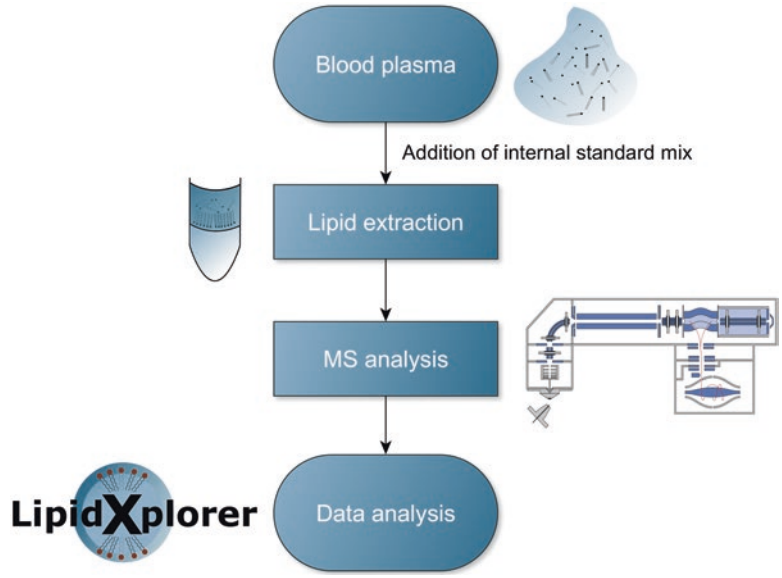


Fig. 1 Workflow of quantitative top-down shotgun lipidomics analysis of blood plasma. Thaw blood plasma at 4 °C and extract lipids with MTBE/MeOH mixture containing internal standards. Dilute 10 μL aliquot of the total extract ten times with isopropanol/methanol/chloroform (4:2:1 (v/v/v)) containing 7.5 mM ammonium formate, and infuse into a Q Exactive mass spectrometer using robotic nanoflow ion source TriVersa NanoMate. FT MS spectra are acquired at high mass resolution and interpreted by LipidXplorer software that identifies and quantifies plasma lipids

3.2 Preparation of the IS Mixture for Quantifying Plasma Lipids

1. Rinse all glass vials and bottles with MTBE, MeOH, and H_2O (*see* **Note 4**).
2. Pipette the volume containing the required amounts of each standard into a rinsed glass vial or bottle (*see* **Note 5**).
3. Vortex briefly and incubate the internal standard mix for ca. 1 h at room temperature to ensure that all standards are completely dissolved (*see* **Note 6**).

3.3 Lipid Extraction

1. Pipette 5 μl of plasma into a 2 mL "safe lock" tube (Eppendorf) placed on ice (*see* **Note 7**).
2. Add 700 μL MTBE/MeOH (5:1.5 (v/v)) supplemented with the internal standards mixture as described in Subheading 3.2. Vortex briefly.
3. Shake on an Eppendorf shaker (Thermomixer *comfort*) for 1 h at 4 °C.
4. Add 140 μL of H_2O. Vortex briefly.
5. Shake on an Eppendorf shaker for 15 min at 4 °C.
6. Spin down for 15 min at 12,000 × *g* on a desktop centrifuge (*see* **Note 8**).
7. Transfer 500–550 μL of the upper organic phase into a rinsed 2 mL glass vial (Supelco). Store at −20 °C until analyzed.

3.4 Shotgun Lipidomics and Data Processing

3.4.1 Shotgun Lipidomics Analysis

Mass spectrometric analyses were performed on a Q Exactive instrument (Thermo Fisher Scientific, Bremen, Germany) equipped with a robotic nanoflow ion source TriVersa NanoMate (Advion BioSciences, Ithaca NY, USA) using nanoelectrospray chips with the diameter of spraying nozzles of 4.1 μm. The ion source is controlled by the Chipsoft 8.3.1 software (Advion BioSciences).

1. Spin glass vials containing plasma lipid extracts for ca. 30 min at 3400 × *g* at 4 °C (*see* **Note 9**).
2. Pipette 10 μL of the lipid extract into a 96-well plate (Eppendorf), add 90 μL 7.5 mM ammonium formate in isopropanol/methanol/chloroform (4:2:1 (v/v/v)), and mix.
3. Set up ionization voltage of +0.96 kV for positive and −0.96 kV for negative ion mode, respectively; backpressure 1.25 psi for both modes should be maintained during polarity switching [23]; temperature of the ion transfer capillary 200 °C, and *S*-lens RF level set at 50% (*see* **Note 10**).
4. Acquire FTMS spectra for each sample for the total time of 5.7 min within the range of m/z 400–1000 starting from 0.02 to 1.5 min in positive and from 4.2 to 5.7 min in negative mode at the target mass resolution of Rm/z_{200} = 140,000 and automated gain control (AGC) of 10^6.
5. Acquire FTMS/MS spectra within the range of m/z 120–440 from 1.5 to 4.0 min in positive mode at Rm/z_{200} = 140,000 by fragmenting the precursor ions m/z 404.3892 and m/z 411.4325 of Chol and Chol-d7 internal standard, respectively, and detecting their specific fragments at m/z 369.3521 and m/z 376.3954. Set the number of acquired microscans to 1, width of precursor isolation window to 0.8 Da, normalized collision energy to 12.5%, and AGC to 5×10^4.

3.4.2 Data Analysis by LipidXplorer

In this workflow lipids are identified by LipidXplorer software [22, 24]. The software and the installation guide for LipidXplorer is at https://wiki.mpi-cbg.de/wiki/lipidx/index.php/Main_Page and the operational manual in ref. [24]

1. Collect all acquired *.raw files in one folder.
2. Convert *.raw files to *.mzML using MSConvert software (http://proteowizard.sourceforge.net/). Select peak picking and binary encoding precision 32-bit.
3. Start LipidXplorer and select the folder containing the *.mzML mass spectra.
4. Set import settings according to Table 2 and import the spectra (*see* **Note 11**).
5. Proceed to the Run panel and pick the molecular fragmentation query language (MFQL) files specifying the method of identification for each lipid class. Set the mass tolerance for MS

Table 2
Major settings of LipidXplorer software used for plasma lipid quantification

Instrument settings	Standard validation		Plasma		Cholesterol
Import source					
Ionization mode	Positive	Negative	Positive	Negative	Positive
Mode	FT MS	FT MS	FT MS	FT MS	FT MS/MS
Selection window [Da]	-	-	-	-	0.8
Time range [min]	0.1–1.0	0.1–1.0	0.1–1.5	4.2–5.7	1.5–4.0
Calibration masses	586.5563	680.4502	586.5563	680.4502	586.5563
	656.5829	592.3978	656.5829	592.3978	656.5829
m/z range	400–1000	400–1000	400–1000	400–1000	120–450
Resolution	140,000	140,000	140,000	140,000	140,000
Tolerance [ppm]	10	10	10	10	10
Threshold[a]	2500	400	2500	400	2500
Resolution gradient[a]	−63	−63	−63	−63	−270
Min occupation	0	0	0	0	0

[a]Threshold and resolution gradient settings are instrument and experiment dependent (see LipidXplorer manual [6] for details)

and MS/MS modes at 5 ppm. Run LipidXplorer. The results are exported as a *.csv file containing the molecular ions of identifed lipid species and their abundances. (*see* **Note 12**).

6. Lipids are quantified by comparing the isotopically corrected intensities of their molecular ions with the intensity of corresponding internal standards. Only lipids whose monoisotopic peaks are detected with the signal-to-noise ratio above the value of 10 should be quantified (*see* **Note 13**).

4 Notes

1. Concentrations of lipid standards are adjusted such that, after mixing with plasma, they will be close to concentrations of endogenous lipids of the same class. Solvent composition is optimized to avoid using chloroform for both plasma lipid extraction and making stocks of lipid standards.
2. The same protocol is also applicable to blood serum.
3. Use SM 30:1:2 for quantifying SM 35:1:2 and SM 35:1:2 for quantifying SM 30:1:2. Currently QLMS are not available for PC O- (PC O- lipids were quantified using the PC standard),

LPC O- (using the LPC standard), and PE O- (using the PE standard). Concentrations of Chol, CholE, and LPE in self-made stocks only relied on weighted amounts of lipids. We typically weight no less than 5 mg of the lipid powder to minimize the weighting error.

4. Only use glass since even at −80 °C extended storage of stock solutions in plasticware increases background. Always rinse glassware with all solvents used for extraction and dissolving the recovered lipids. Pipette standard stock solutions with glass capillaries. Before pipetting, stocks should be conditioned to room temperature. Check that all standards are completely dissolved, sonicate, and gently heat them (up to 30 °C), if necessary.
5. Prepare a volume of IS mix that is sufficient to extract lipids from all samples in the batch, including blanks. The amounts of IS (in pmol) required for the analysis of one plasma sample shown in Table 1 are adjusted to the recommended volume of extraction solvent of 700 μL.
6. IS mix can be stored at −20 °C.
7. Thaw plasma samples at 4 °C or on ice. Vortex briefly before pipetting. Do not thaw plasma samples more than twice to avoid loss of lipids. Always make a blank sample by performing all extraction operations without adding plasma—this helps to control chemical background. Mind that background might differ between different batches of plasticware and organic solvents.
8. Spin down to completely separate the two liquid phases.
9. Spin down the upper organic phase once again. Some samples may look turbid because of residual water. Water (mainly, salts that it inevitably contains) compromises the spray stability, especially in negative mode. Acquire spectra in technical duplicates. Chill down the plate holder to 6 °C (this option is available at the TriVersa NanoMate ion source).
10. These settings are instrument dependent and are given here for orientation only.
11. At this step a master scan file (*.sc) is created and saved in the same folder with *.raw files. Use m/z of at least one IS as calibration mass. Thresholds should be adjusted in each series of experiments (see operation manual [24] for details).
12. Molecular fragmentation query language (MFQL) queries are compiled for PC, PC O-, LPC, LPC O-, PE, PE O-, LPE, PI, SM, TAG, DAG, Cer, Chol, and CholE lipid classes and are available at the LipidXplorer wiki site: https://wiki.mpi-cbg.de/wiki/lipidx/index.php/Main_Page.
13. Reference concentrations for healthy young Caucasian males and females are in ref. [21]. Other plasma datasets with absolute

concentrations are in refs [2, 10, 23]. However, the health status of study subjects in [2, 10, 23] is unknown and therefore lipid concentrations, particularly of Chol, CholE and TAG, could deviate from the "healthy" reference values ref. [21] substantially.

References

1. Quehenberger O, Dennis EA (2011) The human plasma lipidome. N Engl J Med 365:1812–1823
2. Quehenberger O, Armando AM, Brown AH et al (2010) Lipidomics reveals a remarkable diversity of lipids in human plasma. J Lipid Res 51:3299–3305
3. Pietilainen KH, Sysi-Aho M, Rissanen A et al (2007) Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects - a monozygotic twin study. PLoS One 2:e218
4. Oresic M, Simell S, Sysi-Aho M et al (2008) Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. J Exp Med 205:2975–2984
5. Rhee EP, Cheng S, Larson MG et al (2011) Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest 121: 1402–1411
6. Tagami S, Inokuchi Ji J, Kabayama K et al (2002) Ganglioside GM3 participates in the pathological conditions of insulin resistance. J Biol Chem 277:3085–3092
7. Graessler J, Schwudke D, Schwarz PE et al (2009) Top-down lipidomics reveals ether lipid deficiency in blood plasma of hypertensive patients. PLoS One 4:e6261. doi:10.1371/journal.pone.0006261
8. Fernandez C, Sandin M, Sampaio JL et al (2013) Plasma lipid composition and risk of developing cardiovascular disease. PLoS One 8:e71846. doi:10.1371/journal.pone.0071846
9. Tarasov K, Ekroos K, Suoniemi M et al (2014) Molecular lipids identify cardiovascular risk and are efficiently lowered by simvastatin and PCSK9 deficiency. J Clin Endocrinol Metab 99:E45–E52
10. Han X, Rozen S, Boyle SH et al (2011) Metabolomics in early Alzheimer's disease: identification of altered plasma sphingolipidome using shotgun lipidomics. PLoS One 6:e21643. doi:10.1371/journal.pone.0021643
11. Wood PL, Filiou MD, Otte DM et al (2014) Lipidomics reveals dysfunctional glycosynapses in schizophrenia and the G72/G30 transgenic mouse. Schizophr Res 159:365–369
12. Oresic M, Seppanen-Laakso T, Sun D et al (2012) Phospholipids and insulin resistance in psychosis: a lipidomics study of twin pairs discordant for schizophrenia. Genome Med 4:1. doi:10.1186/gm300
13. De Leon H, Boue S, Szostak J et al (2015) Systems biology research into cardiovascular disease: contributions of lipidomics-based approaches to biomarker discovery. Curr Drug Discov Technol 12:129–154
14. Smilowitz JT, Zivkovic AM, Wan YJ et al (2013) Nutritional lipidomics: molecular metabolism, analytics, and diagnostics. Mol Nutr Food Res 57:1319–1335
15. Gräßler J, Kopprasch S, Passauer J et al (2013) Differential effects of lipoprotein apheresis by lipidfiltration or dextran sulfate adsorption on lipidomic profile. Atheroscler Suppl 14:151–155
16. Graessler J, Bornstein TD, Goel D et al (2014) Lipidomic profiling before and after Roux-en-Y gastric bypass in obese patients with diabetes. Pharmacogenomics J 14:201–207
17. Hyotylainen T, Oresic M (2015) Optimizing the lipidomics workflow for clinical studies-practical considerations. Anal Bioanal Chem 407:4973–4993
18. Wang C, Wang M, Han X (2015) Applications of mass spectrometry for cellular lipid analysis. Mol Biosyst 11:698–713
19. Moore JD, Caufield WV, Shaw WA (2007) Quantitation and standardization of lipid internal standards for mass spectroscopy. Methods Enzymol 432:351–367
20. Wang M, Wang C, Han X (2016) Selection of internal standards for accurate quantification of complex lipid species in biological extracts by electrospray ionization mass spectrometry-what, how and why? Mass Spectrom Rev. doi:10.1002/mas.21492
21. Sales S, Graessler J, Ciucci S et al (2016) Gender, contraceptives and individual metabolic predisposition shape a healthy plasma lipidome. Sci Rep 6:27710. doi:10.1038/srep27710

22. Herzog R, Schwudke D, Schuhmann K et al (2011) A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language. Genome Biol 12:R8. doi:10.1186/gb-2011-12-1-r8
23. Schuhmann K, Almeida R, Baumert M et al (2012) Shotgun lipidomics on a LTQ Orbitrap mass spectrometer by successive switching between acquisition polarity modes. J Mass Spectrom 47:96–104
24. Herzog R, Schwudke D, Shevchenko A (2013) LipidXplorer: software for quantitative shotgun lipidomics compatible with multiple mass spectrometry platforms. Curr Protoc Bioinformatics 43:14.12.1–14.12.30. doi:10.1002/0471250953.bi1412s43

Chapter 17

Proteomics Analysis of Circulating Serum Exosomes

Antonius Koller, Purvi Patel, Jenny Kim Kim, and Emily I. Chen

Abstract

Proteomics characterization of biofluids, such as urine and plasma, has been explored for the discovery of predictive, prognostic, and mechanistic biomarkers of diseases and tissue injury. Here we describe comprehensive characterization of protein cargos from cell-derived secreted vesicles (extracellular vesicles or exosome) for biomarker discovery using the mass spectrometry-based technology.

Key words Serum biomarkers, Proteomics analysis of biofluids, Circulating exosome, Mass spectrometry

1 Introduction

Recent advances in proteomic technologies incorporating mass spectrometry (MS) for biomarker discovery show great promise in providing comprehensive knowledge of molecular profiles from complex biological samples such as biofluids from patients [1–6]. Serum or plasma is an ideal source to identify diagnostic, prognostic, and mechanistic biomarkers because it can be obtained routinely from patients. However, analyzing the blood proteome is known to be analytically challenging due to the inherent limitations in the dynamic range of proteins in circulation [7, 8]. Therefore, we propose to perform proteomics analysis on secreted extracellular vesicles such as exosome since they have been reported to remain intact in biofluids during long-term storage and therefore can serve as an excellent reservoir for biomarker discovery. In the past few years, it has become increasingly clear that exosomes have specialized functions [9]. Proteomic cataloguing of exosomes from diverse cell types has revealed a common set of membrane and cytosolic proteins, suggesting the evolutionary importance of these membrane particles [9, 10]. These studies have also demonstrated that an exosome is actively secreted by live cells, which represents a new type of intercellular messenger. Findings from our studies further support the feasibility of identifying protein cargo in a serum exosome as biomarkers of treatment-induced toxicity and systemic

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_17,

injury [11, 12]. Using the MS-based proteomics technology, we identified between 400 and 800 proteins in serum exosome known to be in different subcellular locations (Fig. 1a) and have a wide range of biological functions (Fig. 1b). In this chapter, we described a method to prepare a serum exosome for comprehensive quantitative proteomics analysis, which is compatible with a variety of different mass spectrometry-based analytical platforms (Fig. 2).

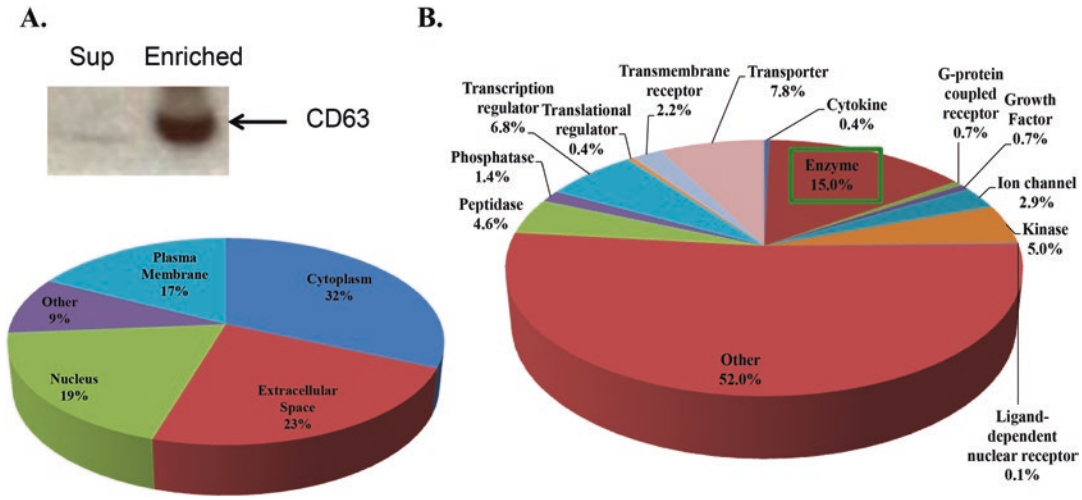


Fig. 1 Proteomics characterization of isolated serum exosome. (**a**) CD63 was used to detect the enrichment of exosome in the enriched fraction versus the supernatant by Western blot analysis. (**b**) Classification of proteins identified in a small cohort of human patients based on subcellular location (*left*) and known biological functions (*right*)

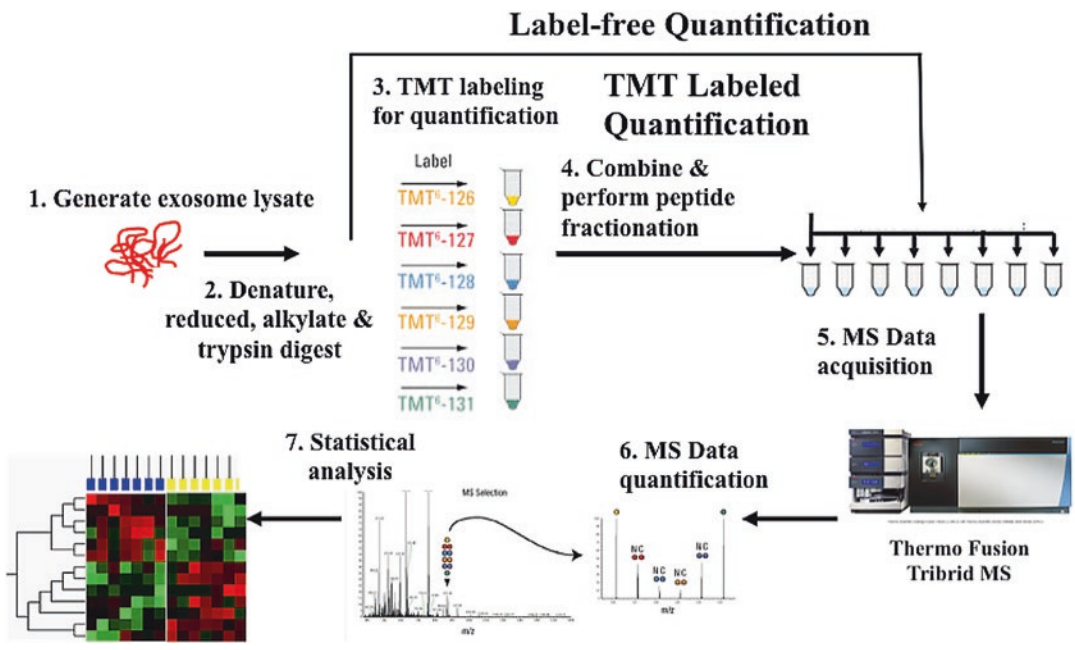


Fig. 2 An illustration of proteomics workflow to analyze protein cargo in isolated serum exosome

2 Materials

Prepare all the solutions using ultrapure water and analytical/HPLC grade reagents. Prepare and store all the reagents at room temperature (unless indicated otherwise). The amount of the reagents can be scaled up depending on the number of experiments.

2.1 Serum Exosome Isolation

1. Total exosome isolation (from serum) kit: Invitrogen by Life Technologies, catalg#4478360.
2. 100 mM triethylammonium bicarbonate (TEAB): Mix 1 mL of 1 M TEAB buffer with 9 mL of HPLC grade water.

2.2 Western Blot Analysis

1. Milli-Q water.
2. 1× running buffer: Mix 50 mL of 10× Tris-Glycine-SDS running buffer with 450 mL water.
3. 4–20% Novex Tris-Glycine mini gel.
4. Electrophoresis system (e.g., XCell SureLock Mini-Cell Electrophoresis System).
5. Loading/sample buffer: (e.g., NuPage LDS sample buffer; 4× or higher).
6. Transfer apparatus: (e.g., standard wet transfer apparatus OR semidry transfer system (iBlot)).
7. Clean plastic case.
8. 1× phosphate-buffered saline with Tween 20 (1× PBST): Add 500 µL of Tween 20–450 mL of 1×PBS (i.e., 0.1% v/v Tween 20 in PBS).
9. 0.1% Ponceau S. staining solution (0.1% (x/v) Ponceau S in 1% (v/v) acetic acid):
   Mix 10 mL water with 0.3 mL glacial acetic acid. Add 33 mg Ponceau S. Add water to make the volume up to 30 mL.
10. 5% w/v nonfat dry milk solution: Mix 2.5 g of nonfat dry milk with 1× PBST to make up the volume to 50 mL. Make sure the solution does not have any lumps. The excess can be stored in 4 °C fridge for up to 3 days.
11. Antibody dilution buffer (5% BSA/1× PBST): Mix 2.5 g of BSA with 1× PBST to make up the volume up to 50 mL.
12. Primary antibody: CD63 antibody (Santa Cruz Biotech; catalog#Sc15363).
13. Secondary antibody: HRP-conjugated anti-rabbit IgG (Jackson ImmunoResearch; catalog#111-035-003).
14. Detection reagent: ECL kit (e.g., HyGLO HRP detection kit).
15. X-ray films: Classic blue autoradiography film.

2.3 In Solution Digestion, TMT Labeling, and High pH Reversed-Phase Peptide Fractionation

1. HPLC grade water.
2. 1.5 mL nonstick tube.
3. 100 mM ammonium bicarbonate: Mix 50 μL of 1 M ammonium bicarbonate with 450 μL of water.
4. 200 mM dithiothreitol (DTT): Prepare 1 M DTT first with 100 mM ammonium bicarbonate. Dilute 1 M DTT further to 200 mM DTT with 100 mM ammonium bicarbonate (*label-free*).
5. 200 mM Tris(2-carboxyethyl)phosphine (TCEP): Add 70 μL 0.5 M TCEP to 70 μL water. Then add 35 μL of 1 M TEAB (*TMT labeling*).
6. 100 mM iodoacetamide (IAA): Prepare 1 M IAA first with 100 mM ammonium bicarbonate. Dilute 1 M IAA further to 100 mM IAA with 100 mM ammonium bicarbonate.
7. ACS methanol (TMT labeling).
8. ACS chloroform (TMT labeling).
9. Sequencing grade modified trypsin (1 μg/μL) made in 100 mM ammonium bicarbonate.
10. Pierce Quantitative Fluorometric Peptide Assay Kit: (Pierce catalog# 23290) (TMT labeling).
11. TMT10plex Isobaric Label Reagent Set: (Thermo Fisher Scientific; 0.2 mg vials; catalog#90309) (TMT labeling).
12. 5% hydroxylamine: Add 50 μL of 50% hydroxylamine to 450 μL of 100 mM TEAB (TMT labeling).
13. 0.1% trifluoroacetic acid (TFA): Mix 10 μL of TFA to 10 mL of water.
14. Pierce High pH Reversed-Phase Peptide Fractionation Kit: (Pierce catalog# 84868).
15. Acetonitrile.

3 Method

Mix samples using a vortex mixer. Do not overdry the samples in the SpeedVac as this makes the pellet harder to resolubilize.

3.1 Serum Exosome Isolation

1. Quickly thaw 100 μL of serum on ice and centrifuge the serum at 2000 rcf for 30 min at 4 °C. Transfer the supernatant to a new 1.5 mL nonstick tube without disturbing the pellet, and place it on ice.
2. Add 40 μL of the exosome reagent from the kit to the supernatant. Mix the serum/reagent mixture well by vortexing and pipetting up and down until the solution is homogenous. The solution should have cloudy appearance. Incubate the solution

on ice or at 4 °C for 30 min. After incubation, centrifuge the solution at 10,000 rcf for 10 min at room temperature.

3. Collect the supernatant (flow-through fraction) in a new 1.5 mL nonstick tube. Save the flow-through fraction for exosome enrichment analysis in Subheading 3.2. The majority of exosomes should be in the pellet.
4. Gently rinse the pellet gently with 500 μL of cold PBS. Be careful not to dissolve the pellet in 1× PBS completely. Then, centrifuge it 10,000 rcf for 10 min at room temperature. Save the wash for additional analysis if desired and save the total exosome pellet.
5. Homogenize the total exosome pellet in 150 μL of urea lysis buffer (4 M urea and protease inhibitor cocktail in 100 mM ammonium bicarbonate). The pellet is sticky, so make sure that the pellet is completely dissolved in the buffer. Centrifuge the sample at 10,000 rcf for 5 min at room temperature. Quantify the sample protein by using the appropriate protein quantification method (*see* **Note 1**). Store the remaining sample at −80 °C.

3.2 Confirmation of Serum Exosome Enrichment by Western Blot Analysis

1. Bring the precast, 4–20% Tris-Glycine mini gel, from 4 °C fridge (*see* **Note 2**). Remove the gel cassette from the pouch and rinse with water (*see* **Note 3**). Peel of the tape covering the slot on the back of the gel cassette. Insert the gel in to the buffer chamber so that the shorter "well" side of the cassette faces toward the buffer core. Fill the buffer chamber with 1× running buffer enough to completely cover the sample wells. Pull the comb out of the cassette in one fluid motion to expose the gel-loading wells.
2. Prepare the samples (exosome lysate and the flow-through fraction) by adding loading/sample buffers (final 1×) and heating the samples at 95 °C for 5 min to denature proteins (*see* **Note 4**). Centrifuge the heated samples at maximum speed in a tabletop centrifuge for 1 min to bring all the samples to the bottom of the tubes.
3. Load the exosome lysate and the flow-through in the gel wells using a gel-loading tip. Load a pre-stained protein molecular weight marker in the first well.
4. Electrophorese the gel at 125 V until the dye front has reached the bottom of the gel.
5. Following electrophoresis, open up the gel plates using a spatula. Rinse the gel with water and transfer carefully in a clean container with freshwater. Immediately proceed to transferring of the gel using either a standard wet transfer technique or semidry/rapid transfer technique (i.e., iBlot) (*see* **Note 5**). Follow the manufacturer instruction for proper protein transferring.

6. Transfer the blotted membrane in a clean plastic case. Wash the membrane in water for 5 min (*see* **Note 6**). To check for success of transfer, wash the membrane in 1× PBST, and immerse the membrane in sufficient amount of 0.1% Ponceau S staining solution. Stain the membrane with Ponceau S solution for 5 min on a benchtop orbital shaker at room temperature. After staining, wash extensively in water until the water is clear and protein bands are well defined. Take an image if possible.
7. Proceed to the blocking step. Block the membrane with 5% nonfat milk solution. Incubate it in the cold room for 1 h on a benchtop orbital shaker. Wash the membrane every 15 min with 1× PBST and vigorous rotation. Perform at least three washes.
8. Dilute the primary antibody (CD63 antibody; 1:1000 dilution) in 5% BSA. Incubate the membrane in the primary antibody solution for 1 h at room temperature with gentle shaking (*see* **Note 7**). Remove the primary antibody solution and wash the membrane every 15 min with 1× PBST and vigorous rotation. Perform at least three washes.
9. Prepare the secondary antibody solution (HRP-conjugated anti-rabbit; 1:10,000) in 5% BSA. Incubate the membrane for 1 h at room temperature on a tabletop orbital shaker. Remove the secondary antibody solution, and wash the membrane every 15 min with 1× PBST and vigorous rotation. Perform at least three washes.
10. For HRP-conjugated antibodies, enhanced chemiluminescence kits are traditionally used as substrates with varying detection limits. For signal development, follow the manufacturer instructions. Remove excess reagent and cover the membrane in transparent plastic wrap (*see* **Note 8**). Take the blot to the developing room and place the membrane between the covers of a propylene sheet protector. Gently smooth out any air pockets. Switch off the lights and place the blue X-ray film on top of the membrane. Expose the film for 30 s and then develop. Repeat the exposure, varying the time as needed for optimal detection.

3.3 Preparing Exosome Lysate for Proteomics Analysis Using the Label-Free Approach for Quantification

1. Take 120 μg of the exosome protein sample (*see* **Notes 9** and **10**), and add 6 μL of 1 M ammonium bicarbonate to the sample (final concentration 50 mM). Add 3 μL of 200 mM dithiothreitol to the sample (final concentration 5 mM). Vortex for 10 s and incubate for 30 min at room temperature. Add 12 μL of 100 mM iodoacetamide to the sample (final concentration 10 mM). Vortex for 10 s and incubate for 30 min at room temperature in the dark.
2. Add 3 μg of trypsin for digestion (1:40 enzyme-to-protein ratio). Centrifuge the sample at maximum speed for 5 min at room temperature.

3. Add water to adjust the final volume to 120 µL. Incubate the sample at 37 °C overnight in an air-circulating incubator (*see* **Note 11**). The protein concentration after the final step of digestion should be ~1 µg/µL.
4. The next day, proceed to quantify the peptides generated from the digest using the Pierce Quantitative Fluorometric Peptide Assay Kit.
5. Equilibrate kit components to room temperature for at least 45 min before opening and using the kit.
6. Centrifuge the samples at maximum speed for 5 min to collect the liquid at the bottom of the tube.
7. Using the procedure in Table 1 to prepare a serial dilution of the Peptide Digest Assay Standard for generating a standard curve (fluorescence unit vs. µg/mL). Dilute the Peptide Digest Assay Standard in clean nonstick microfuge tubes preferably using the same diluent as the sample(s). The method in Table 1 will provide sufficient volume to run a 9-point standard curve (from 0 to 1000 µg/mL) in triplicate. Peptide standard concentrations are provided in µg/mL.
8. Prepare one dilution (1:8) using 5 µL peptide sample enough to run them in triplicate. Include blank wells that contain only the Fluorometric Peptide Assay Buffer and the Fluorometric Peptide Assay Reagent from the kit.
9. Pipette 10 µL of each standard or sample a replicate into each well of the clear-bottom black 96-well microplate in triplicates. Add 70 µL of Fluorometric Peptide Assay Buffer to each well.

Table 1
Preparing a serial dilution of the Peptide Digest Assay Standard

Centrifuge tubes	Volume of diluent (µL)	Volume of digest (µL)	Final standard concentration for peptide mixtures (µg/mL)
A	0	150 of stock	1000
B	75	75 of tube A dilution	500
C	75	75 of tube B dilution	250
D	75	75 of tube C dilution	125
E	75	75 of tube D dilution	62.5
F	75	75 of tube E dilution	31.3
G	75	75 of tube F dilution	15.6
H	75	75 of tube G dilution	7.8
Blank	75	0	0

Add 20 µL of Fluorometric Peptide Assay Reagent to each well (*see* **Note 12**). Incubate the reaction at room temperature for 5 min (*see* **Note 13**).

10. Measure the fluorescence using EX/EM at 390 nm/475 nm using a fluorescent microplate reader. Use the standard curve to determine the peptide concentration of each sample.
11. Once the peptide concentration is determined, take 100 µg peptide sample(s) in separate tube and dry them using the SpeedVac. Then proceed to perform High pH Reversed-Phase Peptide Fractionation (Pierce Thermo Fisher Scientific).

3.4 Preparing Exosome Lysate for Proteomics Analysis Using TMT Isobaric Mass Tag Labeling Approach for Quantification

The following procedures are for labeling nine serum exosome samples:

1. Pipette 30 µg from each exosome lysate into nine different tubes with the corresponding sample name on the top of each tube.
2. A reference sample is created by combining 3.4 µg of each of the nine samples to a total of 30 µg. At this point there should be ten tubes each with 30 µg proteins.
3. Add 100 mM TEAB to adjust the final volume of each sample to 50 µL.
4. Add 2.5 µL of 200 mM TCEP (final concentration 10 mM) to each sample and incubate the samples at 55 °C for 1 h.
5. Add 2.5 µL of 375 mM IAA (final concentration 18.8 mM) to each sample, and incubate the samples for 30 min in the dark at room temperature.
6. Perform methanol-chloroform-water precipitation (*see* **Note 14**). Resuspend the protein pellet from each sample in 26.25 µL of 100 mM TEAB.
7. Add 3.75 µL of 200 ng/µL trypsin (750 ng) to each sample (1:40 enzyme-to-protein ratio). Digest these samples overnight at 37 °C in an air-circulating incubator.
8. The next day, proceed to quantify the peptides generated from the digest using Pierce Quantitative Fluorometric Peptide Assay Kit (follow the same procedure as in method Subheading 3.3, **steps 6–9**).
9. Once the peptide concentration is determined, take 20 µg peptide from each sample and proceed to TMT10plex isobaric mass tag labeling.
10. Centrifuge the peptide samples at maximum speed for 5 min at room temperature.
11. Bring one 0.2 mg vial from each TMT10plex isobaric mass tags (126, 127C, 127N, 128C, 128N, 129C, 129N, 130C, 130N, and 131) from the TMT10plex isobaric mass labeling kit and equilibrate them at room temperature for 5 min before

using them (*see* **Note 15**). Add 20 μL of 100% ACN to each tag and let the tag dissolve at room temperature for 5 min (*see* **Note 16**).

12. Add 20 μg of each peptide sample to the reconstituted TMT tags (*see* **Note 17**). Add 1 M TEAB to each sample to adjust the final volume to 45 μL (final concentration 100 mM TEAB, *see* **Note 18**). Mix the samples and centrifuge at maximum speed for 1 min in a tabletop centrifuge at room temperature.
13. Incubate the labeling reaction for 1 h at room temperature. Add 1 μL of 5% hydroxylamine (final 0.1%) to the samples. Incubate for 15 min at room temperature to quench the reaction. Mix the samples and centrifuge at maximum speed for 1 min in a tabletop centrifuge at room temperature.
14. Take 1 μL of each labeled sample to test the labeling efficiency by mass spectrometry (MS) analysis. Store the remaining samples in −80 °C without drying until the labeling efficiency and equal amount of labeled peptides from each sample are verified.
15. A normalization step can be done after the preliminary MS analysis to ensure an equal amount of labeled peptide is combined from each sample. Then proceed to perform High pH Reversed-Phase Peptide Fractionation (Pierce Thermo Fisher Scientific). Since the capacity of each C18 column from the kit is about 100 μg, mix 10 μg of labeled peptides from each sample, and save the other half of the labeled peptides for future use by drying them individually using the SpeedVac and store them at −80 °C.

3.5 High pH Reversed-Phase Peptide Fractionation

1. Prepare the elution solutions according to Table 2 (**label-free quantification**) or Table 3 (**TMT isobaric labeling quantification**). Resuspend 100 μg of either labeled or unlabeled peptides in 300 μL of 0.1% TFA solution.
2. Bring one C18 spin column from the kit and remove protective white tip (bottom). Place the column in a 2.0 mL tube. Centrifuge at 5000 rcf for 2 min. Discard the liquid.
3. Remove top screw cap and load 300 μL of ACN into the column. Replace the cap and put into a 2.0 mL tube and centrifuge at 5000 rcf for 2 min. Discard the acetonitrile solution.
4. Condition the spin column twice with 0.1% TFA. Discard the 0.1% TFA solution.
5. Place the spin column into a new 2.0 mL tube. Load the peptide sample onto the column and place the top cap. Centrifuge the column at 3000 rcf for 2 min. Repeat the loading of the flow-through material onto the same column, and collect the flow-through from the second loading material in the same tube. This is the flow-through fraction. Transfer it in a new 0.5 mL nonstick tube and save for MS analysis.

Table 2
Preparation of elution solutions for Reversed-Phase pH Peptide Fractionation (label-free)

Fraction no.	Final ACN (%)	Stock ACN (μL)	0.1% triethylamine (μL)
1	2	20	980
2	4	40	960
3	6	60	940
4	8	80	920
5	10	100	900
6	12	120	880
7	14	140	860
8	16	160	840
9	18	180	820
10	20	200	800
11	80	800	200

Table 3
Preparation of elution solutions for Reversed-Phase pH Peptide Fractionation (TMT labeled)

Fraction no.	ACN %	100% ACN (μL)	0.1% triethylamine (μL)
1	10.0	100	900
2	12.5	125	875
3	15.0	150	850
4	17.5	175	825
5	20.0	200	800
6	22.5	225	775
7	25.0	250	750
8	80.0	800	200

6. Place the column in a new 2.0 mL tube for the washing step. Add 300 μL of HPLC water onto the column and centrifuge 3000 rcf for 2 min. Repeat this step once. Combine wash 1 and wash 2 and dry the wash fraction by SpeedVac.
7. Place the column in a new 2.0 mL tube and add 300 μL of fraction 1 elution solution to the column. Centrifuge the column at 3000 rcf for 2 min and collect the eluent. Place the

fraction 1 eluent on dry ice. Perform fraction 2–11 using the same procedures as fraction 1 and place all collected fractions on dry ice. Dry peptide fractions by SpeedVac and proceed to MS analysis. Alternatively, fractionated peptides can be stored at −80 °C temporarily.

4 Notes

1. Urea interferes with BCA assay. A fluorescence-based protein quantification assay such as the Qubit Protein Assay Kit (Life Technologies, Inc.) can be useful for accurate protein quantification.
2. Use precast gels to save time, simplify preparation, and obtain high-quality uniform results.
3. Wash the cassette wells with 1× running buffer. Invert the gel and shake gently to remove excessive buffer. Be sure to displace all air bubbles from the cassette wells as they will affect sample running.
4. Load 1/8 volume of exosome sample per well and 1 μL of the flow-through fraction (equal amount). Load a positive control in one of the wells. It will indicate the procedure is optimized and working. We recommend you check the antibody datasheet, which often provides a suggested positive control.
5. Detailed instructions for the transfer process can be found on the websites of the manufacturers of transfer apparatus and will vary depending on the system. The principle is the same in each case.
6. Stain the SDS gel with Coomassie blue after transferring to check the efficiency of blotting proteins to the membrane.
7. Alternatively, the membrane can be incubated with the primary antibody overnight in a cold room on an orbital shaker.
8. Make sure that the membrane does not dry during the process. If the membrane is dry, the background will be stronger in the image.
9. Out of the 120 μg sample, 5 μg of the digested peptides is used for peptide quantification assay, and 100 μg of the peptide is used for high pH reversed-phase fractionation. The excess peptide can be stored in −80 °C freezer.
10. Calculate first the volume of water and other reagents that need to be added in the sample. Add water first to the sample before adding other reagents.
11. Put a beaker of water inside the air-circulating incubator. This will prevent the excessive evaporation of the sample. In addition, Parafilm can be used to seal the tube.

12. Do NOT premix Fluorometric Peptide Assay Buffer and Fluorometric Peptide Assay Reagent.
13. The reaction is completed within 5 min, but the signal is stable up to 30 min.
14. For methanol-chloroform-water precipitation, add 3× sample volume of methanol, 1× sample volume of chloroform, and 3× sample volume of water to each sample. Vortex the sample for 10 s and then centrifuge the sample at 13,000 rcf for 5 min. Remove the top aqueous layer without disturbing the interface. The interface contains the protein layer that is visible as a thin white wafer. Add 3× sample volume of methanol and centrifuge the sample again for 13,000 rcf for 5 min. Pipette out as much methanol as possible from the tube without disturbing the pellet. Dry the sample using the SpeedVac. Be careful not to overdry the sample.
15. The TMT reagents are amine-reactive and modify lysine residues and peptide N-termini. All amine-containing buffers and additives must be removed before digestion and labeling. The methanol-chloroform-water precipitation helps remove all the amine-containing buffers and additives.
16. The TMT reagents are moisture sensitive. To avoid moisture condensation of the product, the vial must be equilibrated to room temperature before opening.
17. Always add samples to the tag and not the other way around. The acetonitrile in the tags make it difficult to pipette as ACN is not viscous enough which will make it drip from the pipet tip.
18. 100 mM TEAB is added to all samples to maintain equal volumes for all samples for the ease of sample handling. This volume can be adjusted after calculating the volume of peptide samples and the TMT tags.

References

1. Carrera M, Gallardo JM, Pascual S, Gonzalez AF, Medina I (2016) Protein biomarker discovery and fast monitoring for the identification and detection of Anisakids by parallel reaction monitoring (PRM) mass spectrometry. J Proteomics 142:130–137.doi:10.1016/j.jprot.2016.05.012
2. Cominetti O et al (2016) Proteomic biomarker discovery in 1000 human plasma samples with mass spectrometry. J Proteome Res 15:389–399. doi:10.1021/acs.jproteome.5b00901
3. Crutchfield CA, Thomas SN, Sokoll LJ, Chan DW (2016) Advances in mass spectrometry-based clinical biomarker discovery. Clin Proteomics 13:1. doi:10.1186/s12014-015-9102-9
4. Ma H, Chen G, Guo M (2016) Mass spectrometry based translational proteomics for biomarker discovery and application in colorectal cancer. Proteomics Clin Appl 10:503–515. doi:10.1002/prca.201500082
5. Rossing K et al (2016) Urinary proteomics pilot study for biomarker discovery and diagnosis in heart failure with reduced ejection fraction. PLoS One 11:e0157167. doi:10.1371/journal.pone.0157167
6. Thomas S, Hao L, Ricke WA, Li L (2016) Biomarker discovery in mass spectrometry-based urinary proteomics. Proteomics Clin Appl 10:358–370. doi:10.1002/prca.201500102

7. Baker ES et al (2012) Mass spectrometry for translational proteomics: progress and clinical implications. Genome Med 4:63. doi:10.1186/gm364
8. Boukouris S, Mathivanan S (2015) Exosomes in bodily fluids are a highly stable resource of disease biomarkers. Proteomics Clin Appl 9:358–367. doi:10.1002/prca.201400114
9. Thery C (2011) Exosomes: secreted vesicles and intercellular communications. F1000 Biol Rep 3:15. doi:10.3410/B3-15
10. Thery C et al (2001) Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles. J Immunol 166:7309–7318
11. Chen EI et al (2015) Identifying predictors of taxane-induced peripheral neuropathy using mass spectrometry-based proteomics technology. PLoS One 10:e0145816. doi:10.1371/journal.pone.0145816
12. Kulkarni S et al (2016) Identifying urinary and serum exosome biomarkers for radiation exposure using a data dependent acquisition and SWATH-MS combined workflow. Int J Radiat Oncol Biol Phys 96(3):566–577. doi:10.1016/j.ijrobp.2016.06.008

Part IV

Antibody-based Discovery and Targeted Proteomics

Chapter 18

High-Density Serum/Plasma Reverse Phase Protein Arrays

Cecilia Hellström, Tea Dodig-Crnković, Mun-Gwan Hong, Jochen M. Schwenk, Peter Nilsson, and Ronald Sjöberg

Abstract

In-depth exploration and characterization of human serum and plasma proteomes is an attractive strategy for the identification of potential prognostic or diagnostic biomarkers. The possibility of analyzing larger numbers of samples in a high-throughput fashion has markedly increased with affinity-based microarrays, thus providing higher statistical power to these biomarker studies. Here, we describe a protocol for high-density serum and plasma reverse phase protein arrays (RPPAs). We demonstrate how a biobank of 12,392 samples was immobilized and analyzed on a single microarray slide, allowing high-quality profiling of abundant target proteins across all samples in one assay.

Key words Reverse phase protein array, RPPA, Serum, Plasma, Affinity proteomics, Noncontact inkjet printer, Protein profiling, Fluorescent detection

1 Introduction

Traditionally there are two main formats of affinity microarrays, forward phase arrays, and reverse phase arrays. In the former, the capture reagent is immobilized on the microarray surface, while in the latter the target analyte is immobilized. In what is now denoted reverse phase protein arrays (RPPAs), the spotted analyte is part of a complex biological sample [1].

RPPAs were first described in the context of printing lysates acquired from laser capture microdissection [2]. Some efforts have been made to apply RPPAs on other sample materials, such as cerebrospinal fluid, serum, and plasma, but the technology is still relatively unexplored in regard to these biofluids [3–8]. The main challenge of analyzing serum or plasma is the complexity of the sample matrix and the dynamic range [9]. Although the sensitivity of RPPAs is limited by the picoliter volumes used for spotting, profiling medium to highly abundant proteins in serum or

Authors Cecilia Hellström and Tea Dodig-Crnković have equally contributed to this work.

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_18,

plasma is feasible [10]. Additionally, RPPAs allow high sample throughput by simultaneous analysis of an analyte across thousands of samples [11].

Similarly to all other affinity-based methods, RPPA is dependent on validated high-quality affinity reagents. Today, there is no standardized approach to assessing reagent validity, and it is known that reagent performance is method dependent. However, functional assays such as ELISA, Western blotting, immunohistochemistry, and immunofluorescence are commonly used for determining the specificity, selectivity, and reproducibility of an affinity reagent. Inter- and intra-method reproducibility may also assist in the validation [12].

The most commonly used substratum for RPPAs is nitrocellulose, mainly due to its high binding capacity compared to other substrata [11, 13]. When selecting a reporter molecule, it is important to be aware that nitrocellulose autofluorescence overlap with emission wavelengths of some commonly used fluorescent detection molecules [14].

Different microarray printers are available on the market, and both contact and noncontact printers are suitable for printing RPPAs. Here, we present a protocol in which a noncontact printer is used. The printer uses the piezoelectric effect to eject droplets of 100 pl onto the slides at a distance of 1–5 mm, hence limiting the risk of disturbing the membrane surface [15, 16].

In this protocol, we describe the RPPA technology applied on serum and plasma samples and demonstrate its scalability to thousands of samples within one array. To our knowledge, this is the largest serum RPPA produced to date.

The protocol is divided into the following sections: sample preparation, printing of arrays, assay procedure, and image analysis (*see* Fig. 1). The protocol has been applied to a cohort of 12,392 serum samples (*see* Fig. 2a), which were collected within the TwinGene cohort (2004–2008, Sweden) [17]. The cohort comprises of samples from monozygotic and dizygotic twins, both paired (4851) and individual (2690). An even distribution of females (6764) and males (5628) were included, with an age range of 47–94 and a mean age of 64.9. The array was created with the purpose of studying proteins related to aging and twinness. The primary antibodies applied on the arrays have been produced and validated within the Human Protein Atlas (www.proteinatlas.org) [18] (*see* Fig. 2b, c).

2 Materials

2.1 Sample Preparation

1. Printing buffer: phosphate-buffered saline buffer pH 7.4 (1×PBS), supplemented with 0.1% (v/v) Tween 20 (1×PBS-T), and 50% (v/v) glycerol, stored at +4 °C (*see* **Notes 1** and **2**).
2. Benchtop semiautomatic pipettor: CyBi-SELMA (CyBio).

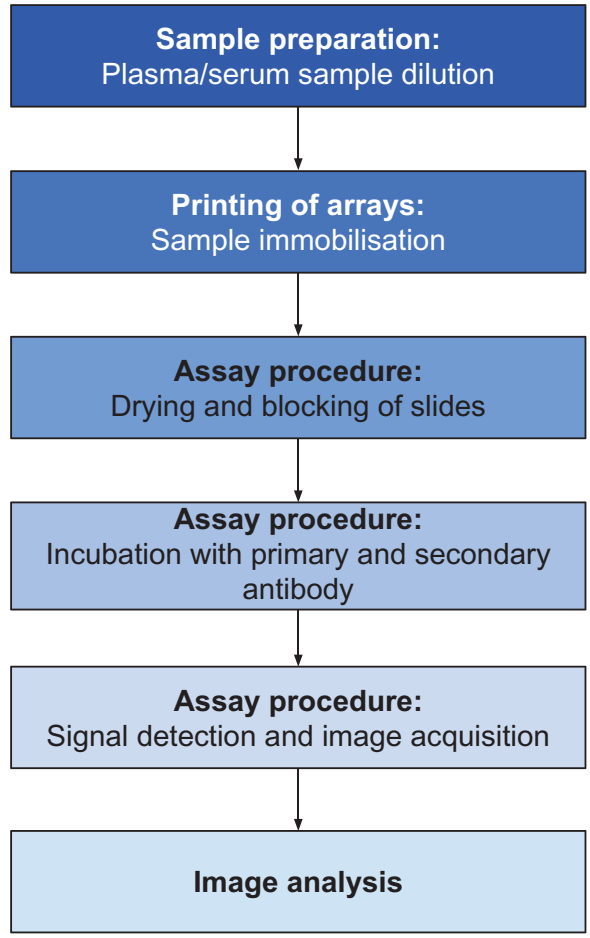


Fig. 1 A flowchart describing the major steps in the described protocol

2.2 Printing of Arrays

1. Microarray printer: Arrayjet Marathon Inkjet Microarrayer with a 12-sample JetSpyder (Arrayjet Ltd.).
2. Plate lids: JetGuard Probe Protector (Arrayjet Ltd.).
3. Slides: nitrocellulose-coated slides, 1 mm × 75.6 mm × 25.0 mm (UniSart 3D nitro, Sartorius Stedim).
4. System buffer: 47% (v/v) glycerol, 0.06% (v/v) Triton X-100 (*see* **Notes 1–3**).
5. Oven: hybridization oven/shaker.

2.3 Assay Procedure

1. Blocking buffer: 1×PBS-T supplemented with 3% (w/v) bovine serum albumin (BSA).
2. Wash buffer 1, WB1: 1×PBS-T.
3. Wash buffer 2, WB2: 1×PBS.
4. Antibody dilution buffer: 1×PBS-T.

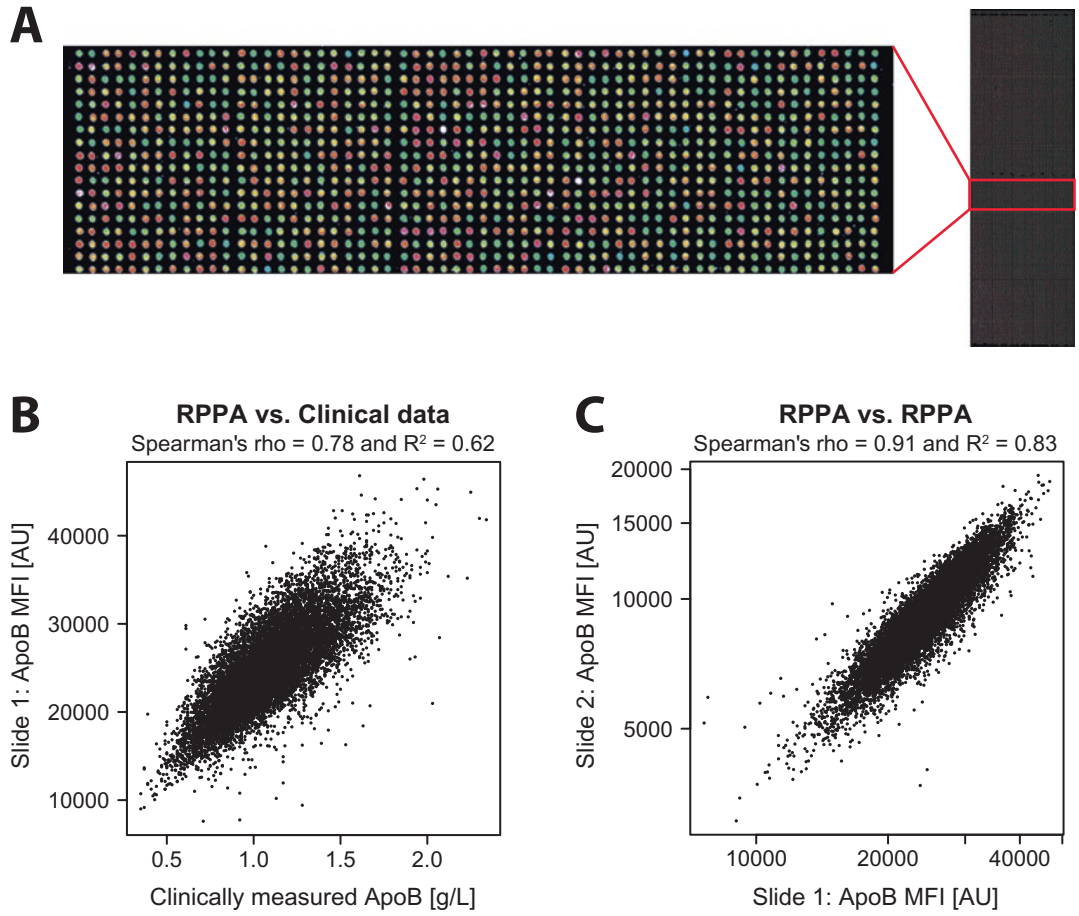


Fig. 2 Example data measuring apolipoprotein B (ApoB) in 12,648 features, consisting of 12,392 unique samples and 256 technical controls, using the serum RPPA arrays. Correlations are calculated using Spearman's rho and the square of Pearson's correlation coefficient. (**a**) Zoom in and full image of array, spots colored by signal intensity (*low-high: blue-green-yellow-red-white*), (**b**) RPPA signals correlated to clinically measured ApoB, and (**c**) correlation between two slides incubated with the same anti-ApoB antibody

5. Primary antibody: anti-ApoB (0.1167 mg/ml, HPA049793, Human Protein Atlas).
6. Secondary antibody: goat anti-rabbit IgG Alexa Fluor 647 (Invitrogen).
7. Slide tray: quadriPERM 4 × 12 (Sarstedt).
8. Orbital shaker: Sky Line DOS-10L (ELMI).
9. Microarray scanner: LuxScan HT24 (CapitalBio Corp.).

2.4 Image Analysis

1. Image analysis software: GenePix Pro 5.1 or later versions (Molecular Devices).
2. Statistical software: Microsoft Excel or R.

3 Methods

3.1 Sample Preparation

In the first part of this protocol, we describe the preparation of crude samples, applicable to both plasma and sera. We will not discuss array design or sample positioning in detail; however, it is important to plan for this prior to sample dilution and transfer to print plates (*see* **Note 4**). We recommend a printing design where samples are randomized within the array and supplemented with controls such as positive controls, negative controls, dilution series, and replicates. To account for local variations in the substratum, the controls should be arrayed in multiple locations and replicates should not be printed next to each other.

1. Place frozen samples at +4 °C to thaw overnight (*see* **Note 5**).
2. Ensure that the thawed samples are free from air bubbles by vortexing and centrifuging them using a benchtop centrifuge (2000 × *g*, 2 min).
3. Distribute printing buffer into print plates using an automated multichannel pipette, 20 μl printing buffer per sample well. Using a benchtop semiautomatic pipettor, dilute samples 1/5 by adding 5 μl sample to the print plates.
4. Prepare controls to be included on the array (*see* **Note 6**). Negative controls: printing buffer. Positive controls: species-specific antibodies (rabbit IgG, mouse IgG, goat IgG, donkey IgG, chicken IgY, and human IgG). Quality control: replicates of individual samples, pools, and dilution series of both individual samples and pools. Transfer each control to its designated position in the print plate.
5. Keep print plates at +4 °C if they are to be used immediately; otherwise, store at −20 °C.

3.2 Printing of Arrays

A range of different printers are available that may be suitable for printing serum and plasma samples on nitrocellulose slides. This protocol describes the printing procedure using the noncontact printer Marathon Inkjet Microarrayer from Arrayjet.

1. Start the printer and perform the instrument-specific start-up routine according to the printer's manual. Ensure that the humidity and temperature stabilize at 50% and +20 °C, respectively, before starting the printing process. Empty waste and refill system buffer bottles (*see* **Note 3**).
2. Thaw printing plates and let them reach room temperature. Vortex and centrifuge the plates until no air bubbles are present (2000 × *g*, 5 min). Attach plate lids to limit evaporation and contamination of wells during printing. Mount ready plates in the printer and fill up the plate tray with up to six print plates at a time.
3. Load the printer with unused slides, up to 100 slides per run (*see* **Note 7**).

4. Array the samples according to the desired printing design (*see* **Note 4**). Spot three drops per sample, which gives a final spot volume of 300 pl and a diameter of 200 μm.
5. During print run: Keep a log over the humidity, temperature, and instrument pressure. Empty waste and refill the system buffer bottles as needed. For consecutive loadings, thaw and prepare print plates as in **step 2**.
6. When the printing is finished, remove all slides from the printer and dry them in an oven at +37 °C for 16 h or until dry. If slides are not dried properly, tailing or bleed off may occur in the consecutive steps (*see* **Note 8**).
7. Store printed slides in air and light tight boxes at +4 °C.

3.3 Assay Procedure

Slides with printed samples are incubated with an antibody targeting the analyte of interest, and detection is enabled using a far-red-fluorescent dye antibody (*see* **Note 9**). The following instructions describe the usage of 1-pad slides for the screening of one analyte across thousands of samples. The protocol can be adjusted to also suit slides with multiple pads (e.g., 8- or 16-pad slides) (*see* **Note 10**).

1. Take out slides from +4 °C and let them reach room temperature.
2. Submerge each slide in 15 ml blocking buffer and incubate on an orbital shaker (100 rpm, 1 h).
3. For each slide, prepare a slide tray filled with WB1 and using tweezers quickly dip the slides into the buffer. Proceed by submerging each slide in a new slide tray filled with ca 15 ml WB1 and let wash on an orbital shaker (100 rpm, 5 min). Continue by repeating the dip and washing step in consecutive order, three times (*see* **Note 11**).
4. Dilute primary antibody 1/2350: 1.7 μl anti-ApoB in 4 ml antibody dilution buffer (*see* **Note 12**).
5. Incubate slides with 4 ml diluted primary antibody on an orbital shaker (65 rpm, 1 h). Cover the slides to avoid dust contamination.
6. Wash slides as described in **step 3**.
7. Dilute secondary antibody 1/60,000: 0.25 μl goat anti-rabbit IgG Alexa Fluor 647 in 15 ml antibody dilution buffer (*see* **Note 12**).
8. Incubate slides with 15 ml diluted secondary antibody on an orbital shaker (90 rpm, 1 h). Cover slides from light to protect fluorophores.
9. Wash slides as described in **step 3**. Keep slides covered from light.
10. In order to wash away any residual Tween 20, wash slides with WB2 on an orbital shaker (100 rpm, 5 min). Keep slides covered from light.

11. Remove potential residual salts by carefully rinsing the slides in deionized water.
12. Spin-dry each slide before loading them into a microarray scanner (*see* **Note 13**).

3.4 Image Analysis

In order to interpret the assay, numerical data has to be extracted from the scanned images. Different software products are available and different approaches can be used. Here, we describe the first steps in one approach.

1. Import the acquired gray-scale image into the image analysis software GenePix Pro (5.1 or later versions) and assign excitation wavelengths to the image.
2. Create an array list, such as a GenePix Array List file (GAL file), that contains information of which sample is located in which spot. This may be done in the printer software, the GenePix software, or manually in, for example, Microsoft Excel (*see* **Note 14**).
3. Load the array list, manually align the grid roughly, and then let the software automatically align the grid features to the spots. Use the setting "Find irregular features" to better fit spot morphology. Assess the alignment, adjust grid features if necessary, and flag spots that will require attention during data analysis. Analyze the array and save the result file.
4. The result file contains a numerical matrix with spot features per row and measurement parameters per column. The data can be analyzed using software products such as Microsoft Excel or R. The latter may be preferred for big data sets.
5. Before interpreting the data, it is important to assess its quality. Check if the printed controls give expected signals. Adjust for varying background signals and their potential influence on the spots by subtracting the local background of each spot. Exclude data points from spots that are flagged or give a signal below three standard deviations of the buffer signals plus buffer mean. If possible, assess the quality of the data by correlating it with measurements from a different method and between replicate slides.

4 Notes

1. Prepare buffers in ultrapure water, such as Milli-Q water (Merck Millipore).
2. Adjust the glycerol level in the printing buffer in order to accommodate for different dilution factors. Aim for a final glycerol percentage of 40–50%, and make sure it is the same for all wells in each print plate. If using a different printer than

a Marathon Inkjet Microarrayer, check instrument requirements regarding buffers before preparing them.

3. Handle Triton X-100 and waste with care since these are harmful to the environment and may carry biological hazards.
4. If unsure of printing design, do a test print with different distances between spots to make sure all samples will fit and still be distinguishable. Different membrane pad sizes are available, e.g., 1, 8, or 16 pads per slide. The choice of slide format depends on the number of samples, inclusion of replicates, and the number of analytes to be profiled. The uptake and printing order of samples depend on printer type; therefore, ensure that the printing plates are filled appropriately.
5. If a large number of samples are to be printed, thaw and dilute them in batches. When returning samples to freezer storage, take out the next batch to thaw. Do not thaw or freeze samples unless necessary. Keep samples on ice when working at room temperature and use appropriate protective wear (gloves, laboratory coat, and safety goggles).
6. To allow flexible usage of secondary antibodies from different species, print various species-specific antibodies to ensure the inclusion of a positive control on the array. Replicates should be present in a print plate as well as printed several times and distributed throughout large arrays for printing quality control.
7. Always handle slides with gloves and/or tweezers. Slides are preferably held by the edges or the barcode area, if present. Make sure not to touch the membrane surface or to damage the barcode if present, especially if using tweezers. Note the positions of all slides in the printer so that potential technical variability during printing can be tracked and accounted for.
8. An initial quality control can be done by observing the spots in a regular benchtop microscope. Spots are mainly visible before they have dried. Drying status, alignment, and morphology can be assessed in this way. Before assay usage, dried slides may also be roughly assessed with the naked eye in regard to alignment. Non-dried content may spread outside spot area, especially when submerged into blocking buffer, causing tailing of spots that complicates grid alignment during image analysis or bleed off that might cause increased local background.
9. Use fluorophore at red, far-red, or infrared wavelengths since the nitrocellulose membrane has a natural high autofluorescence at green wavelengths. Use a maximum of one fluorophore per detection channel in the scanner you will use.
10. If using 1-pad slides: Perform washes and incubations in slide trays, one tray per slide. Optimize buffer and antibody volumes

depending on the tray size. If using slides with multiple pads: block slides as described in the protocol (**steps 1–3**). Using a slide holder with a silicone mask, create a separate chamber for each pad during primary antibody incubation. Wash pads by first rinsing with WB1 by pipetting in and out five times and then adding WB1 for 5 min wash incubation; repeat four times. Take out slides from the holder, quickly dip in WB1, and submerge in secondary antibody for incubation and follow detailed protocol (from **step 7**). Optimize buffer and antibody volumes depending on the incubation chamber size.

11. Blocking buffer and antibody residues maintained within the nitrocellulose membrane can contribute to uneven background levels. This can be alleviated by varying the position of the slides on the shake table to ensure an even washing of the membranes.
12. A small-scale antibody dilution optimization and validation test is recommended to be performed beforehand, in order to determine a suitable working concentration of each antibody. Slide area and slide tray volume should be taken into consideration. If a large array is to be analyzed, it is recommended to also print a subset of the samples onto smaller pads to be used for the antibody dilution optimization.
13. Adjust laser power and photomultiplier-tube (PMT) settings in scanner to avoid saturated signals and for the acquisition of a suitable signal dynamic range. Use appropriate channel(s) for the fluorophores you have used. We recommend also including the channel for green wavelengths during scanning if available, since the autofluorescence may serve as support for grid alignment as well as identification of contaminations or membrane and printing irregularities.
14. For array list formatting, see the manual of the printer software, the GenePix software or Molecular Devices Knowledge Base (http:// mdc.custhelp.com).

Acknowledgment

We thank the Swedish Twin Registry for providing the serum samples. We also thank all members of the Plasma Profiling and the Auoimmunity Profiling groups at SciLifeLab, as well as the entire staff of the Human Protein Atlas. This study was funded by grants from Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, and the KTH Center for Applied Proteomics (KCAP) funded by the Erling-Persson Family Foundation. The authors declare no conflict of interest.

References

1. Aguilar-Mahecha A, Hassan S, Ferrario C, Basik M (2006) Microarrays as validation strategies in clinical samples: tissue and protein microarrays. OMICS 10:311–326. doi:10.1089/omi.2006.10.311
2. Paweletz CP, Charboneau L, Bichsel VE et al (2001) Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. Oncogene 20:1981–1989. doi:10.1038/sj.onc.1204265
3. Caiazzo RJ, Maher AJ, Drummond MP et al (2009) Protein microarrays as an application for disease biomarkers. Proteomics Clin Appl 3:138–147. doi:10.1002/prca.200800149
4. Ahmed F, Gyorgy A, Kamnaksh A et al (2012) Time-dependent changes of protein biomarker levels in the cerebrospinal fluid after blast traumatic brain injury. Electrophoresis 33:3705–3711. doi:10.1002/elps.201200299
5. Janzi M, Odling J, Pan-Hammarström Q et al (2005) Serum microarrays for large scale screening of protein levels. Mol Cell Proteomics 4:1942–1947. doi:10.1074/mcp.M500213-MCP200
6. Janzi M, Sjöberg R, Wan J et al (2009) Screening for C3 deficiency in newborns using microarrays. PLoS One 4:e5321. doi:10.1371/journal.pone.0005321
7. Aguilar-Mahecha A, Cantin C, O'Connor-McCourt M et al (2009) Development of reverse phase protein microarrays for the validation of clusterin, a mid-abundant blood biomarker. Proteome Sci 7:15. doi:10.1186/1477-5956-7-15
8. Kobayashi M, Nagashio R, Jiang SX et al (2015) Calnexin is a novel sero-diagnostic marker for lung cancer. Lung Cancer 90:342–345. doi:10.1016/j.lungcan.2015.08.015
9. Solier C, Langen H (2014) Antibody-based proteomics and biomarker research—current status and limitations. Proteomics 14:774–783. doi:10.1002/pmic.201300334
10. Ayoglu B, Häggmark A, Neiman M et al (2011) Systematic antibody and antigen-based proteomic profiling with microarrays. Expert Rev Mol Diagn 11:219–234. doi:10.1586/erm.10.110
11. Akbani R, Carragher N, Goldstein T et al (2014) Realizing the promise of reverse phase protein arrays for clinical, translational and basic research: a workshop report. Mol Cell Proteomics 13:1625–1643
12. Bordeaux J, Welsh AW, Agarwal S et al (2010) Antibody validation. BioTechniques 48:197–209. doi:10.2144/000113382
13. Sauer U (2011) Impact of substrates for probe immobilization. In: Korf U (ed) Protein microarrays: methods and protocols, Methods in molecular biology, vol 785. Springer Science+Business Media, LLC: Humana Press, Heidelberg, Germany, pp 363–378
14. Tighe P, Negm O, Todd I, Fairclough L (2013) Utility, reliability and reproducibility of immunoassay multiplex kits. Methods 61:23–29. doi:10.1016/j.ymeth.2013.01.003
15. Mcwilliam I, Kwan MC, Hall D (2011) Inkjet printing for the production of protein microarrays. In: Korf U (ed) Protein microarrays: methods and protocols, Methods in molecular biology, vol 785. Springer Science+Business Media, LLC: Humana Press, Heidelberg, Germany, pp 345–361
16. Tisone TC, Tonkinson JL (2005) Non-contact dispensing for protein microarrays. In: Schena M (ed) Protein microarrays. Jones and Bartlett Publishers, Sudbury, MA, pp 169–185
17. Magnusson PKE, Almqvist C, Rahman I et al (2013) The Swedish twin registry: establishment of a biobank and other recent developments. Twin Res Hum Genet 16:317–329. doi:10.1017/thg.2012.104
18. Uhlén M, Björling E, Agaton C et al (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol Cell Proteomics 4:1920–1932. doi:10.1074/mcp.M500279-MCP200

Chapter 19

Antibody Colocalization Microarray for Cross-Reactivity-Free Multiplexed Protein Analysis

Véronique Laforte, Pik-Shan Lo, Huiyan Li, and David Juncker

Abstract

Measuring many proteins at once is of great importance to the idea of personalized medicine, in order to get a snapshot of a person's health status. We describe the antibody colocalization microarray (ACM), a variant of antibody microarrays which avoids reagent-induced cross-reactivity by printing individual detection antibodies atop their corresponding capture antibodies. We discuss experimental parameters that are critical for the success of ACM experiments, namely, the printing positional accuracy needed for the two printing rounds and the need for protecting dried spots during the second printing round. Using small sample volumes (less than 30 μL) and small quantities of reagents, up to 108 different targets can be measured in hundreds of samples with great specificity and sensitivity.

Key words Microarray, Antibody, Sandwich immunoassay, Multiplexed, Fluorescence

1 Introduction

Immunoassays are currently used in the clinic to quantify specific proteins in the blood and plasma of patients to give clues about their health status. Proteomics, the measurement of tens or hundreds of proteins in a single sample, has the potential to empower diagnostics and patient monitoring by providing a more complete snapshot picture of the health status of a person using very little sample. To achieve this multiplexed measurement goal in the future, technologies that measure multiple proteins simultaneously with high sensitivity, precision, and reproducibility are required.

Sandwich immunoassays consist in capturing a target in a sample to the surface using a surface-bound capture antibody, followed by the binding of a detection antibody which recognizes the same target (e.g., a protein) but at a different epitope as the capture antibody. Sandwich immunoassays offer high sensitivity due to the high affinity of antibodies to their target and high specificity, thanks to the double recognition of different epitopes on that target. Antibody microarrays can measure multiple targets at once using the same

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_19,

amount of sample as a classical ELISA and minimal amounts of costly antibodies. However, when multiple detection antibodies are mixed, specificity is often lost due to cross-reactivity between reagents [1] which can be mitigated by extensive selection, optimization of the reagents, and limiting the number of targets measured simultaneously. Cross-reactivity leads to significant false-positive signals, which can mask significant binding or conversely give the appearance of target binding when none occurred [2].

The antibody colocalization microarray (ACM) was developed to avoid cross-reactivity in multiplexed measurements by physically separating individual detection antibody solutions and printing them directly atop their corresponding capture antibodies. Because detection antibodies are not mixed, the same high level of specificity as ELISA is reached with the ACM. Microscope slides are printed with capture antibodies using a microarray printer with silicon quill pins that are fabricated in-house [3]. After blocking and incubating samples on the microarray slides, slides are dried and moved back to the microarray printer where detection antibodies are then spotted over their respective capture antibodies with great positional accuracy. Microarray slides are incubated with a reporter molecule (streptavidin conjugated to Alexa Fluor 647 (AF647)) and scanned with a fluorescence scanner (*see* Fig. 1). Capture and detection antibodies are printed in different low-evaporation buffers which are suited for each step. These low-evaporation printing buffers allow for long printings of several hours without changing the composition of the printed solutions, allowing for better printing reproducibility [4].

Many other methods have been devised to circumvent cross-reactivity in multiplexed measurements. Similar to the ACM, a system was developed with two spotting rounds and an aqueous two-phase system to separate individual detection antibodies. The caveat of this system is the size of spots which limits the density of targets that can be measured in one sample [5]. Proximity extension assays (PEA) [6] and proximity ligation assays (PLA) [7, 8] make use of matched antibody or aptamer pairs conjugated to corresponding short DNA fragments and real-time PCR to quantify the amount of antigen bound. These methods have been shown to accurately measure up to 96 targets in as little as 1 μL samples; however, it requires the labeling of each individual antibody and the use of a separate microfluidic platform. The Simple Plex [9] is a simple polymer chip that uses microfluidics to separate the flow of individual detection antibodies over separate capture areas. The method can very quickly detect proteins in samples; however, the multiplexing capabilities are currently limited to detecting four targets in a same sample. Microarray chips that use force-based discrimination only leave tightly bound antibody on the surface, while cross-reacting antibodies that are expected to be weakly bound are removed [10]. This method was shown for eight

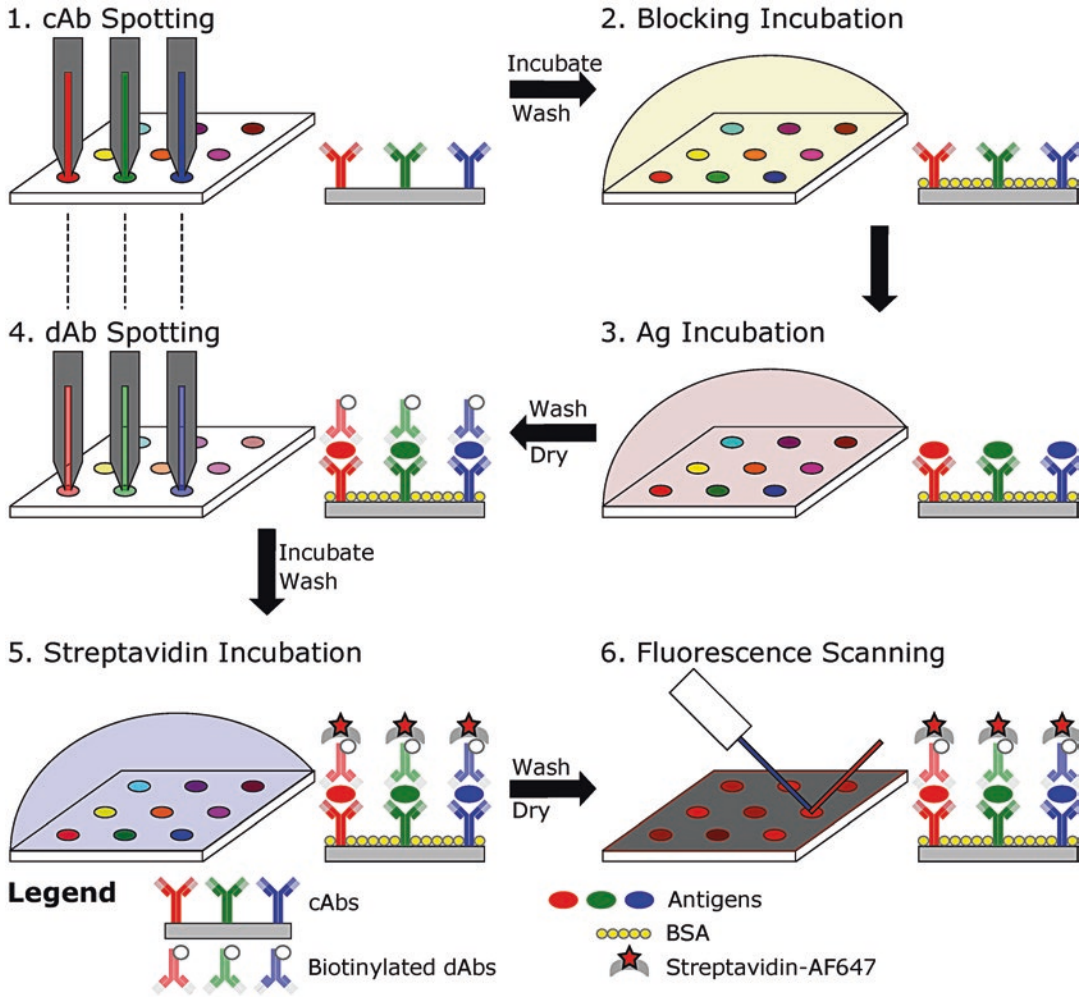


Fig. 1 Schematic of the antibody colocalization microarray. Capture antibodies are printed onto functionalized glass microarray slides (*1*) using silicon quill pins and incubated for 24 h. Microarray slides are washed and blocked (*2*), followed by an overnight incubation of diluted samples and antigen standard overnight at 4 °C (*3*). Microarray slides are then washed and dried before printing cognate biotinylated detection antibodies at the exact same location as previously printed capture antibodies (*4*). Microarray slides are incubated for 16–24 h before washing, followed by the incubation of fluorescent streptavidin (*5*). After washing and drying, microarray slides are scanned with a fluorescent scanner (*6*)

targets. In short, while PEA and PLA have good performance and multiplexing capabilities, these assays are very complex. On the other hand, Simple Plex and the aqueous two-phase system platform are simpler but have limited multiplexing capabilities. In comparison, ACM is simple and allows for more than one hundred targets to be measured.

Several aspects of the procedure described in this chapter are critical to the success of ACM experiments. These are printing positional accuracy, spots protection with trehalose during detection

antibody printing, duration of streptavidin incubation, features of the working environment, and experimental design. Each of these aspects contributes to the high accuracy, sensitivity, and reproducibility of the ACM platform, and they are described below.

Because detection antibodies are printed directly atop their corresponding capture antibodies, the microarray printer used, regardless of whether it is a contact printer or inkjet, should have excellent positional accuracy (10 μm) when a microarray slide is removed and handled between two spotting rounds on the same slide. In order to reach this level of performance, we found it important to avoid re-initializing the printer between printing rounds, as well as establishing a method for calibrating the printer head position. A microarray slide deck that is equipped with spring-loaded slots was used for accurate positioning of slides. Slides were pushed against a corner and the two adjoining sides. Precise and consistent alignment is also dependent on good manual dexterity, and slides were positioned at the same location on the deck.

Since spots containing the capture antibodies and targets are dried before printing the detection antibodies, it is important to coat the microarray slides with a protectant to prevent the degradation of antibodies and proteins at the surface [11]. Trehalose [12] was dried on the surface, forming a protective coating, without the presence of salts or buffer, which denature proteins because of the high salt concentration at the dry state. Detection antibodies are printed with a detection buffer containing glycerol and bovine serum albumin (BSA) that help protect the proteins at the surface during the following incubation. Printing many microarray slides (>10) takes several hours with our setup with four silicon pins used in parallel. In the absence of a trehalose coating just before detection antibody printing, we observed a slow degradation of capture antibodies and bound antigens at the surface that reduced assay reproducibility.

Following sample incubation which is done for ~18 h to allow the capture antibody to antigen binding to reach equilibrium, incubation times for the following steps are crucial. Because the detection antibody spots have very small volumes and high viscosity, which decreases the off-rate of the antibody-antigen complex, the quantity of bound antigen to the capture antibodies is not significantly decreased in spite of the long incubation time. However, during the following washes and streptavidin-AF647 incubation, a trade-off must be found between minimizing incubation times to limit off-rate unbinding and providing sufficient time for the streptavidin to bind in order to give a strong signal. We found that 20–30 min of streptavidin incubation is sufficient in our experiments. After the final washing and drying, all microarray slides were scanned at once unless they are kept in the dark and in vacuum. Fluorophores used are sensitive to ozone below levels that can normally be detected, and keeping them in the presence of

air or light for several hours leads to significant degradation to affect reproducibility and sensitivity.

A dedicated room was used for carrying out the procedures described below with lights turned off, ozone removal, and HEPA-filtered air. Ambient light and the presence of even small amounts (10 ppb) of ozone in the air during incubations, washes, and scanning can lead to fluorophore photobleaching and gradual degradation during scanning that can affect reproducibility of the assay. Alexa Fluor dyes are less sensitive to the effect of light than Cy dyes [13]; however, they are more sensitive to ozone [14]. Their use is still warranted by the fact that AF dyes have higher signals and decreased quenching compared to Cy5 and Cy3 [15, 16]. Their absorption and emission spectra also do not change when they are conjugated to antibodies [17]. Dust particles that are bigger than a spot size (~100 μm) can lead to one or several missing spots and, hence, missing data points in an experiment. Moreover, many dust particles are autofluorescent. The use of cellulose-based cotton from lab coats or paper to blot liquids in the work environment should be avoided to prevent contamination by dust particles. Microfiber cloths were used, along with clean-room quality lab coats. Generally following guidelines for a dust-free room (such as a clean room) is helpful to obtain high-quality defect-free microarrays.

If a microarray printer is not readily available, pairs of microarray slides containing preprinted, mirrored capture antibodies and detection antibodies, respectively, can be purchased from Parallex Bioassays (http://www.parallexbio.com, Montreal, Canada) along with a snap-chip device. The ACM assay can then be conveniently performed by using the snap chip to precisely transfer the detection antibodies to the spots with the corresponding capture antibody spots following sample incubation [18, 19]. The snap-chip procedure (*see* Fig. 2) corresponds to the steps described below involving sample incubation, washing, streptavidin-AF647 incubation, and scanning. Steps that are specific to the ACM are the printing steps and the protection of microarray slides with trehalose. This is not necessary in the snap-chip procedure because all detection antibodies are applied at once in parallel. Overall, an experiment using the snap chip is shorter in time than with home-printed microarray slides.

Experimental design is important to the success of ACM experiments (*see* Fig. 3). Two complete standard curves containing a mixture of known quantities of recombinant antigen that are serially diluted (1:2–1:4) with a minimum of seven points (but ideally 15 points to obtain an accurate curve fit) and a blank are included in the layout [20, 21]. Samples are measured at two different dilutions (e.g., 1:3 and 1:50) to allow the quantification of low- and high-abundance proteins. The well position of all samples is randomized to avoid measurement bias. Several blanks and normal replicate samples, for example, from a pooled normal serum or

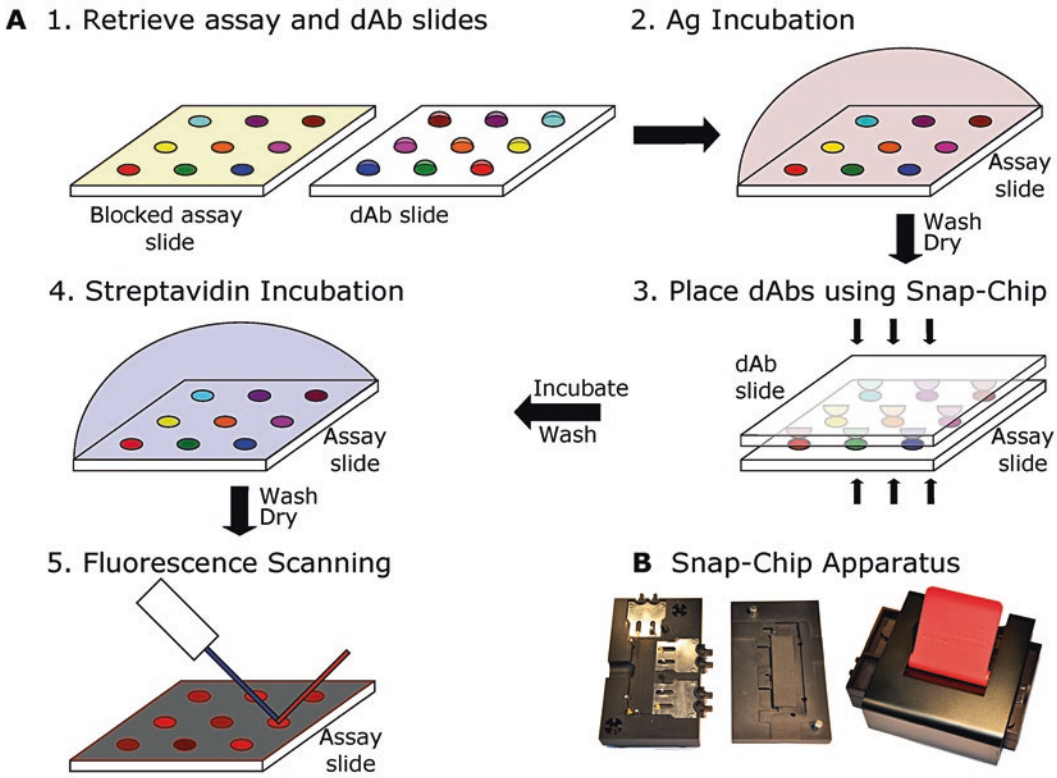


Fig. 2 Schematic of the ACM using a snap-chip apparatus. (**a**) The same number of assay (*capture*) and detection slides are pre-spotted with antibodies (*1*), and the assay slides are pre-rinsed, blocked, and dried prior to shipping. Assay slides are retrieved from storage by the user and incubated with diluted samples and antigen standards overnight at 4 °C (*2*). Assay slides are then washed and dried. After being retrieved from storage, detection slides are brought into contact with their respective assay slides using the snap-chip apparatus (*3*) which aligns the assay and detection slides and ensures incubation of each spot with a detection antibody solution for 1 h. Slides are then separated, and the assay slides are incubated with fluorescent streptavidin (*4*), washed, dried, and scanned (*5*) using a fluorescent scanner. (**b**) The snap-chip apparatus mechanically brings an assay slide and a detection slide in contact with precise force and alignment over the whole surface (*pictures from Parallex Bioassays*)

plasma sample, are measured at regular position intervals (e.g., once per microarray slide) in order to properly measure the limit of detection (LOD) and the reproducibility (coefficient of variation %CV) for each target measured. All values of samples and replicates are quantified by interpolating the log-transformed raw fluorescence value in a log-log curve fit using the standard curve values (without the blank).

The ACM has been used to measure up to 50 targets in 55 samples [1, 22]. Recently, the measurement of up to 108 targets with triplicate replicate spots per target and per sample and for upward of 300 samples with two dilutions per sample has been analyzed. The ACM has been used to measure human serum, plasma with different anticoagulants (EDTA, heparin, CTAD, and

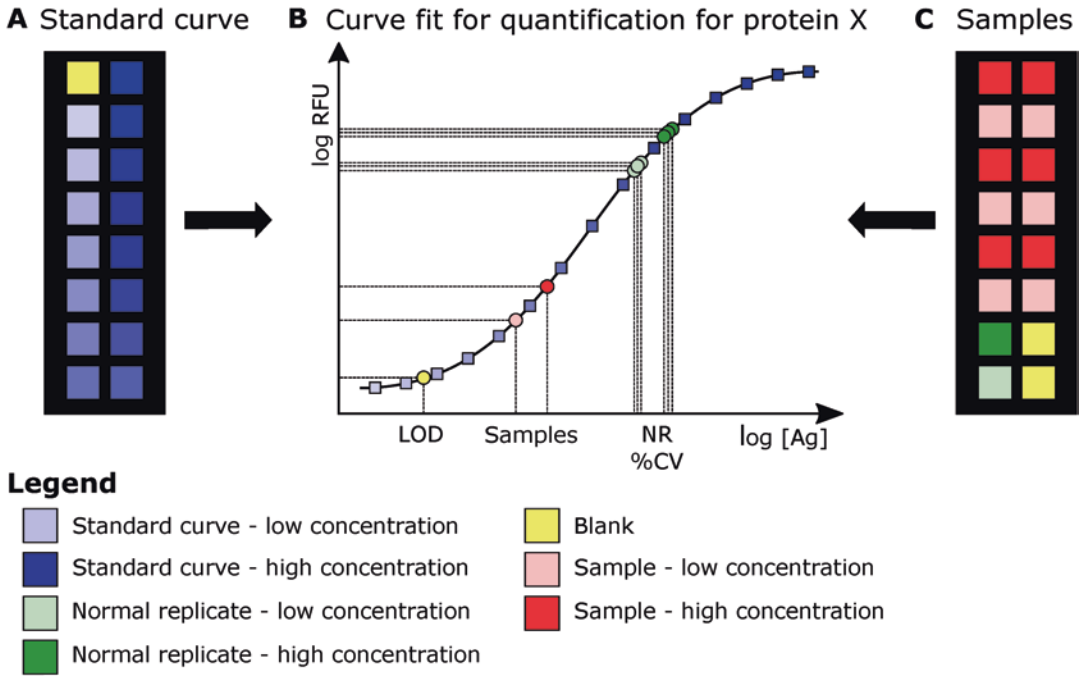


Fig. 3 Experimental design for ACM experiments. A slide containing 16 identical subarrays are incubated with (**a**) a serial dilution of recombinant antigen mixtures (*varying shades of blue*) and a blank (*yellow*). This standard curve is plotted on a log-log graph (**b**) and fitted with a four-parameter logistical curve. Other slides (**c**) are incubated with samples at different dilutions (*red*), blanks (*yellow*), and normal replicates (*pooled samples from healthy individuals*) at different dilutions (*green*). The concentration of each protein is derived by interpolating the values with the binding curves derived in (**a**). The limit of detection (LOD) is calculated as the mean + three times the standard deviation of blanks, whereas the assay coefficient of variation (%CV) is calculated from the interpolation of each normal replicate where the standard deviation is divided by the mean for each target

citrate), as well as other human fluids such as cerebrospinal fluid and urine. The sensitivity of assays using the ACM is comparable to that of ELISA, with LODs ranging from 0.1 to 300 pg/mL depending on the antibody pair used. Reproducibility of assays also varies depending on the antibody pair and can be as good as 10% variability over a large (>200 samples) experiment (unpublished results). Because of long duration of printing for large experiments, reproducibility is better when printing up to six to eight microarray slides with a 16-well gasket (*see* Fig. 2) which leads to printing rounds that are less than 3 h. Reproducibility can be further improved by normalizing the data [23].

2 Materials

2.1 Buffers and Materials

All buffers are prepared using ultrapure water which has a resistance of at least 18 MΩ·cm at 25 °C. All reagents are analytical grade and stored at room temperature unless otherwise indicated.

Follow local waste disposal regulations and MSDS recommendations for chemicals.

1. Wash buffer: 1× phosphate-buffered saline (PBS) containing 0.1% Tween 20. Mix 100 mL of 10× PBS stock to 900 mL of ultrapure water. Add 1 mL of Tween 20 using a viscous liquid pipette (*see* **Note 1**). 1× PBS can be prepared by other methods as long as it is free of small particles (*see* **Note 2**). This buffer can be stored in a squeeze bottle at 4 °C for 1 year.
2. Dilution buffer: 1× PBS containing 0.05% Tween 20. Mix 100 mL of 10× PBS stock to 900 mL of ultrapure water. Add 0.5 mL of Tween 20 using a viscous liquid pipette (*see* **Note 1**). This buffer can be stored at 4 °C for 1 year.
3. Blocking buffer: 3% protease-free bovine serum albumin (BSA) kept at 4 °C, 1× PBS, and 0.05% Tween 20. Mix 0.3 g BSA in 10 mL of dilution buffer. This solution is made fresh when needed.
4. Capture antibody printing buffer: make a 2.9 M betaine, 35.7% 2,3-butanediol in 1× PBS by mixing 1.6987 g of betaine to 1.785 mL 2,3-butanediol using the viscous liquid pipette (*see* **Note 1**), and 1.97 mL of 1× PBS. Dilute this concentrated printing buffer to the required concentration by mixing with 1× PBS before adding to individual capture antibodies. The final concentration for printing is 2 M betaine and 25% 2,3-butanediol in 1× PBS when antibodies have been added (*see* **Note 3**). This solution can be kept at room temperature for up to a week. Do not store this buffer at 4 °C (*see* **Note 4**).
5. Detection antibody printing buffer: make a concentrated BSA-T20 solution by mixing 0.15 g of protease-free BSA kept at 4 °C, 150 µL of wash buffer, and 4.85 mL of 1× PBS. When preparing detection antibodies, mix the appropriate volume of this concentrated BSA-T20 (3% BSA, 0.003% Tween 20, 1× PBS) to pure glycerol using the viscous liquid pipette (*see* **Note 1**) and 1× PBS so that the final concentration of the additives is 1% BSA, 0.001% Tween 20, and 45% glycerol solution in 1× PBS after adding the detection antibody stock solution (*see* **Note 3**). However, before adding the individual detection antibodies, filter the solution containing BSA, Tween 20, and glycerol with a 0.45 µm sterile filter, a syringe, and a hypodermic needle (*see* **Note 5**). Dispose of the hypodermic needle in a sharps container according to local waste management regulations. This solution is prepared fresh daily.
6. Slide rinsing solution: make 50 mL or more of 5% trehalose solution in water by mixing 2.5 g of trehalose to 50 mL of ultrapure water. This solution is kept at 4 °C in a squeeze bottle.

7. Microarray slides: Xenobind slides (Xenopore Corp.) are standard size glass microarray slides with a proprietary reactive aldehyde surface. PolyAn's 2D aldehyde microarray slides also work well with this protocol, as it has the right surface-binding capacity and chemistry (*see* **Note 6**). Slides should be clean and free of dust or visible smudges. Their surface coating should be homogeneous and can be verified by scanning slides at a high gain before using. If slides are not clean, an in-house cleaning and quality control step can be performed (*see* **Note 7**) but should be validated for each slide type.

2.2 Equipment

The ACM protocol requires a number of specialized equipment. Below is the list of equipment and their characteristics or performance parameters that are critical to the success of ACM assays.

1. Microarray printing: printing is done using a contact microarray printer using custom-made silicon quill pins with high liquid capacity [3]. The printer has a spot positioning accuracy of 10 μm or less when microarray slides are taken out of the slide deck and replaced in the same position after further processing. The capture and detection printing buffers are compatible with silicon quill pins. Quill pins are treated once with the flame from an ordinary kitchen torch which forms a plasma that makes the pin channels hydrophilic. During normal operation, pins are washed with a soap solution, followed by distilled water, and then dried using a vacuum pump or absorbent paper. Neither the source plate nor the slide deck is cooled. An inkjet-type microarray printer can be used if it meets minimum performance requirements (*see* **Note 8**). The printing chamber is kept free from dust by filtering the incoming air to the humidifier and minimizing manipulations with hands inside the chamber. Gloves and dust-free (clean room) lab coats are worn at all times in the room where the printer is located, and hair is tied or covered. A HEPA filter and dust-minimizing practices are also recommended for this room. If using a 1536-well plate to dispense liquids, make sure that the plate fits tightly in its enclosure and is well aligned. To facilitate loading the plates, they can be treated with a (gas) plasma. We found that 10 s at 100 W (PlasmaEtch PE-50) worked well for 1536-well plate for detection antibody solutions. The 1536-well plate for capture antibody solutions containing additives is not plasma-treated (*see* **Note 9**).
2. Rotary shaker: a type of flat rotary shaker that has a large surface area and small radius of rotation is used in order to maximize mixing within the 7 mm^2 wells. Moreover, it is compatible with temperatures down to 4 °C. In order to improve the adhesion between gaskets and the rotary shaker, a large, flat layer of polydimethylsiloxane (PDMS) is placed on top of the surface (*see* **Note 10**).

3. Fluorescent scanner: a microarray scanner which scans microarray slides in less than 15 min per slide at a resolution of 5 μm or less is recommended. At least one of the laser and filter combination should be compatible with the AF647 dye.

2.3 Antibody Pairs and Antigens

Assays should be performed with high-quality reagents that have undergone rigorous quality control. Only antibody pairs that have been validated for use as a pair in ELISA or on a microarray are used, and binding curves established using full-length proteins whenever available. All reagents are verified for binding against the species of the samples to be measured.

1. Capture antibodies: these antibodies are typically monoclonal antibodies, purified and unlabeled. If the stock concentration is not at least 0.25 mg/mL, a concentration and quantification step is required prior to aliquoting in working volumes and storing at the appropriate temperature as specified by the manufacturer. Capture antibodies should be free of carrier proteins (e.g., BSA) and contain less than 5% glycerol (*see* **Note 11**).
2. Detection antibodies: these antibodies can be monoclonal or polyclonal antibodies and are purified and labeled with biotin. Other labels are possible (*see* **Note 12**) but should be compatible with the labels of all other assays within a microarray slide's subarray. Their concentration is at least 0.1 mg/mL and they are aliquoted in working volumes and stored according to the manufacturer's instructions. If the concentration is too low, a concentration step, for example, using spin columns, may be used before aliquoting and storing.
3. Antigens: antigens are typically recombinant and purified. Antigens that are obtained from animals or humans and purified can lead to significant cross-reactivity with other targets within the subarray (*see* **Note 13**). Antigen stock solutions should have as high concentration as possible for better storage stability and can be stored in the presence of carrier protein such as BSA in order to improve their shelf life.

2.4 Other Materials and Equipment

1. Microarray gaskets: the microarray gaskets used are Grace Bio-Labs ProPlate® Multi-Array Slide System, more specifically the gaskets with metal clips and 16 wells (2 × 8) that allow us to incubate 16 different samples per microarray slide. Printing of microarray slides is designed to print 16 identical subarrays that correspond to the 16 wells of the gasket. Different gasket layouts can be used in order to have more or less targets and samples on each slide as long as the microarray printing layout is adjusted.

2. Compressed nitrogen stream: many fluorescent dyes are sensitive to heat, oxygen, ozone, and light. When drying microarray slides, it is very important to use a stream of compressed gas that is free of oxygen in order to avoid degradation of the AF647 reporter dye.
3. Square 24.5 cm^2 petri dishes (optional, *see* **Note 10**).
4. Solvent- and water-resistant permanent marker (*see* **Note 14**).
5. Microfiber cloth.
6. 1536-well plate, not plasma-treated (*see* **Note 9**).
7. 1536-well plate, plasma-treated (*see* **Note 9**).
8. Pure ethanol, for washing.
9. 8-channel multichannel pipette.
10. Dust-free, ozone-free sealed room with HEPA filter.
11. 4 °C storage, ideally a cold room.

3 Methods

3.1 Sample Storage

Samples of serum, plasma (EDTA, heparin, citrate, or CTAD), or other biological fluids such as urine, cell culture supernatant, saliva, or cerebrospinal fluid can be collected using standard procedures. Serum samples are allowed to coagulate for 30–60 min before centrifugation. All samples are centrifuged for 10 min at 1500 × *g* at room temperature to remove cells and other particles. The supernatant can be aliquoted in working volumes and kept at −80 °C for short periods of time (less than 1 year). For longer storage, samples are kept in the dry phase of a liquid nitrogen storage system.

3.2 Capture Antibody Printing

1. Mark clean slides with a small black line at the top right corner to identify the printed side, using a permanent marker that does not dissolve in water nor ethanol (*see* **Note 14**). In addition, identify each individual slide in the bottom corners using the same marker. Take care to always handle microarray slides by the sides, never touching the top or bottom with your hands or gloves.
2. Prepare the pin printer for printing by turning on the humidity to 65% at room temperature and cleaning the pins according to the manufacturer's instructions. Make sure all wash buffer containers are full and waste recipients are empty. The humidity in the printing chamber should be stable before capture antibody solutions are placed in the printer. Verify printer alignment (*see* **Note 15**).
3. Prepare capture antibody solutions. Capture antibodies are printed at a concentration of 0.1 mg/mL in the capture antibody printing buffer, and the final concentration of additives is

exactly 2 M betaine and 25% 2,3-butanediol. For example, mix 2 μL of an antibody that has a stock concentration of 1.0 mg/mL to 8 μL of a capture antibody printing solution that contains 2.5 M betaine and 31.3% of 2,3-butanediol. Any significant increase or decrease of the concentration of additives can prevent reproducible printing across all slides due to evaporation or swelling of solutions while printing. If loading a 1536-well plate, preparing 10 μL is sufficient for each individual antibody. If fluorescence signal is too low for all spots for a specific antibody solution, or if spots have streaks in their near vicinity, slightly increase or decrease the concentration accordingly until no streaking is seen and sufficient fluorescence is obtained for all spots (*see* **Note 16**).

4. Load 8 μL of each capture antibody solution into the 1536-well plate using pipette tips that are long and thin (e.g., as is used for loading polyacrylamide gels). Place the bottom of the tip at one of the bottom corner of a well before dispensing the solution slowly; this will ensure that no bubble is formed at the bottom of the well. The source well plate used for the capture antibody solutions is not plasma-treated (*see* **Note 9**).
5. Quickly remove dust from microarray slides using a stream of compressed nitrogen before loading them, as well as the 1536-well plate, into the printer when the relative humidity has reached 65% in the printing chamber. Set up the printing program to print at least three technical replicate spots of each solution per subarray, and 16 identical subarrays on each microarray slide, in locations that fit exactly the 16-well gaskets used. The spacing between spots is at least 200 μm to prevent spots from merging during printing or the subsequent incubation. Make sure that the microarray slides are secured in their position in a reproducible manner (*see* **Note 17**).
6. Print all microarray slides and incubate them for 24 h after printing is finished to allow capture antibodies to fully bind to the surface. Printed spots should be visible on the microarray slides, immediately after printing and also after the 24 h incubation.

3.3 Slide Blocking

1. Clean and assemble gaskets to make them free of dust or chemical residues (*see* **Note 18**).
2. Apply incubated microarray slides onto the gaskets. Mark the top right corner of the slide on the gasket using a small tape, along with the microarray slide number.
3. Wash slides with wash buffer in a squeeze bottle (*see* **Note 19**) by filling the gasket wells halfway with wash buffer, dumping the wash buffer, repeating twice, followed by filling the gasket wells halfway with wash buffer, and leaving on the rotary shaker for 5 min at room temperature and 450 rotations per min (rpm).

4. An entire wash cycle consists in repeating the above **step 3** times in total.
5. After microarray slides are washed, load 80 µL of blocking buffer into each well using an 8-tip multichannel pipette, and incubate on the rotary shaker for 3 h at room temperature, 450 rpm.

3.4 Antigens and Sample Incubation

1. Prepare the antigen standard curve. Antigens are added to dilution buffer as a mixture of low volumes of stock for each antigen before diluting. If the antigen stock concentration is too high to measure at least 1 µL of antigen stock solution into the mixture, then pre-dilute with dilution buffer as necessary. The final mixture containing all antigens to be assayed is diluted 1:2.5 15 times by mixing 66.6 µL of the previous concentration to 100 µL of dilution buffer. Each antigen starting concentration (i.e., its individual concentration in the original mixture) is picked such that 15 serial dilutions cover the entire s-shape of the assay standard curve (*see* Fig. 2). Increase or decrease the starting concentration of an individual antigen in the mixture as needed to shift the resulting s-shaped curve.
2. Retrieve samples from storage and let them thaw for at least 10 min at room temperature or longer at 4 °C. Mix individual samples by pipetting up and down before diluting them. Prepare the samples by diluting them at 1:3 and 1:50 in dilution buffer. For example, mix 31.8 µL of pure serum or plasma to 63.6 µL of dilution buffer to make the 1:3 dilution. Mix 5.4 µL of the 1:3 dilution to 84.6 µL of dilution buffer to make the 1:50 dilution for a sample. Use multiples of these quantities for replicate samples.
3. Dump the blocking buffer from the microarray slides. Knock the remaining liquid from the gaskets by hitting a dust-free surface. Use a dry microfiber cloth (*see* **Note 20**) to wipe the top of the gasket to prevent well-to-well contamination (*see* **Note 21**). Do not leave the microarray slide to dry. Immediately load the samples for one microarray slide before dumping the blocking buffer for the next microarray slide.
4. Place all the microarray slides in a square 24.5 cm^2 petri dish with PDMS at the bottom (*see* **Note 10**). Close the petri dish with its lid, and seal it with two full layers of paraffin film.
5. Incubate overnight (minimum 16 h) at 4 °C on the rotary shaker, at 450 rpm.

3.5 Detection Antibody Printing

1. Prepare the pin printer for printing by turning on the humidity to 65% at room temperature and cleaning the pins according to the manufacturer's instructions. Make sure all wash buffer containers are full and waste recipients are empty. The humidity in

the printing chamber should be stable before both the detection antibody solutions and the microarray slides are placed in the printer. Verify alignment of the printer head (*see* **Notes 15** and **17**).

2. Prepare detection antibody solutions. Detection antibodies are printed at a concentration of 0.01 mg/mL in the detection antibody printing buffer, and the final concentration of additives should be exactly 45% glycerol, 1% BSA, and 0.001% Tween 20. For example, mix 1 μL of an antibody that has a stock concentration of 0.2 mg/mL to 19 μL of a detection antibody solution that contains 47.4% glycerol, 1.053% BSA, and 0.001053% Tween 20. Because the detection antibody solutions are prone to making bubbles when mixing the components by pipetting up and down, make 20 μL even though only 8 μL is loaded onto a 1536-well plate.
3. Load 8 μL of each detection antibody solution into the 1536-well plate using the long and thin pipette tips. Place the bottom of the tip at one of the bottom corner of a well before dispensing the solution; this will ensure that no bubble is formed at the bottom of the well. The source well plate used for the detection antibody solutions is plasma-treated for proper loading without bubbles (*see* **Note 9**).
4. Incubate the slides at room temperature for 30 min before removing the paraffin film from the large square petri dish. This allows the slides to be at room temperature for further processing.
5. Wash microarray slides twice with washing buffer by performing two times Subheading 3.3, **step 3**.
6. Rinse microarray slides rapidly in their gasket three times with PBS without Tween 20. Dump the PBS from the gaskets. Add 80 μL of slide rinsing solution to the gaskets and incubate 5 min at room temperature on the rotary shaker at 450 rpm.
7. Dump the slide rinsing solution from the gaskets. Remove the gasket from the microarray slide and rinse the top side of the slide with more slide rinsing solution with a squeeze bottle.
8. Immediately dry the slide under a forceful, perpendicular stream of compressed nitrogen. This step is done to ensure that a small consistent film of trehalose is left on the surface to protect the complexed capture antibodies and antigens spots during the detection printing step (*see* **Note 22**). Wash and dry a single microarray slide at a time.
9. Print all slides in the same order that capture antibody solutions were printed. Detection antibodies are printed directly atop their corresponding capture antibody solutions. Incubate the slides for 24 h after printing is finished to allow detection antibodies to fully bind to the targets. Printed spots should be

visible on the microarray slides, immediately after printing and also after the incubation.

10. Clean and assemble gaskets (*see* **Note 18**).

3.6 Streptavidin Incubation

1. Apply incubated microarray slides onto clean gaskets. Mark the top right corner of the slide on the gasket using a small tape, along with the microarray slide number.
2. Wash microarray slides by performing Subheading 3.3, **step 4**.
3. Prepare a solution of 0.5 μg/mL of streptavidin-AF647 in blocking buffer and apply 80 μL in each well using a multichannel pipette. To apply to several slides, apply in the same order that washes will be performed, waiting 15–20 s in between each slide (*see* **Note 23**).
4. Incubate for 30 min at room temperature on the rotary shaker at 450 rpm. The microarray slides are kept in the dark at this point because of the presence of a fluorescence marker.
5. Wash microarray slides again by performing Subheading 3.3, **step 4**.
6. Remove the gasket and rinse both sides of the microarray slide with distilled water from a squeeze bottle or a gentle flow from a distilled water tap.
7. Immediately dry the microarray slide under a forceful stream of compressed nitrogen that is parallel to the small axis of the microarray slide, in order to remove all droplets of water. Rinse and dry a single microarray slide at a time.

3.7 Fluorescence Scanning

1. If using a fluorescence scanner that scans through the back of the slide, polish the back of all microarray slides using a dry microfiber cloth (*see* **Note 20**). Remove the microfiber dust particles using a stream of compressed nitrogen.
2. Turn on the scanner and allow enough time for the lasers to warm up. Refer to the manufacturer's instructions.
3. Set the photomultiplier gain to an appropriate number. This gain should be the highest that leads to no saturated pixel (*see* **Note 24**).
4. Scan all the microarray slides as quickly as possible after the experiment is done, and in as little time as possible (*see* **Note 25**).
5. Save all images as TIFF images. If compressing images, make sure that the compression algorithm is loss-less (*see* **Note 26**).

3.8 Data Extraction and Analysis

1. Verify that none of the microarray slide pictures have saturated pixels (*see* **Note 24**).
2. Align grids onto the TIFF images to extract all technical replicate spots on all subarrays and on all microarray slides. Grid spot size should be at least half of the size of the actual spots.

For spots of approximately 100 μm in size, we use a grid spot size of 60 μm diameter.

3. Extract the data by outputting the raw fluorescent intensity. There should be no negative value in this data.
4. Log-transform the data by calculating the log10 of each individual outputted value.
5. Perform outlier removal using Peirce's criterion or Grubbs' test (*see* **Note 27**). Calculate the mean and standard deviation of technical replicates to obtain a log-transformed raw fluorescence intensity value for a specific assay in a given sample, sample replicate, blank, or standard curve dilution. Note that the value of the concentration of antigen should also be log-10 transformed.
6. Perform a four-parameter logistical curve fit on each individual assay's standard curve using the values obtained in Subheading 3.8, **step 5**. It is important to use the standard deviations obtained for the standard curve dilutions to obtain a more accurate curve fit.
7. In order to calculate the limit of detection for a curve fit, calculate the mean and standard deviation for all blank values (obtained in Subheading 3.8 **step 5**) for a given assay within the experiment. Interpolate the value calculated by taking the mean + three times the standard deviation into the curve fit for the given assay. The concentration obtained is the log10 value of the lowest quantity of antigen that can be quantified using the assay.
8. In order to calculate the reproducibility for a given assay, interpolate all individual values from the replicate sample. Divide the standard deviation of all the quantities obtained by the mean of all those quantities. The value obtained is the coefficient of variation (%CV) for this assay.
9. Interpolate the values obtained in Subheading 3.8 **step 5** for all samples in the curve fit for a given assay to obtain the quantities measured in that sample. If the values obtained at dilution 1:3 cannot be quantified because they are above the maximum value from the standard curve, quantify the samples in the 1:50 dilution. Multiply the quantities obtained by the dilution in order to infer the concentration in the original sample (*see* **Note 28**).

4 Notes

1. The volume and concentration of viscous liquids to be measured are critical for this application. Therefore, in order to measure viscous liquids such as glycerol, Tween 20, or 2,3-butanediol, we recommend using a viscous liquid pipette which uses a piston to displace the viscous liquid rather than air.

2. The 1× PBS solution used throughout the protocol should be free of small particles that are often autofluorescent and can easily bind to the microarray surface or sometimes lead to missing spots if particles clog the silicon quill pins. For this reason we recommend buying a 10× PBS stock solution that has been prefiltered by the manufacturer.
3. Antibodies are normally supplied as liquid in a PBS buffer base, or freeze-dried, in which case they are reconstituted in PBS. The presence of up to 5% of glycerol or other cryopreservative chemicals did not affect our experiments. Antibodies supplied in a different buffer than PBS might also be used but should be individually tested. For calculations, we made the assumption that antibody stock solutions are the equivalent of 1× PBS.
4. At a concentration of 2 M betaine and 25% 2,3-butanediol, the capture printing buffer can safely be kept at 4 °C; however, at higher concentrations of betaine, this chemical can precipitate out of solution when kept at 4 °C. For this reason, it is best to keep the capture printing buffer at room temperature.
5. For the same reasons as listed (*see* **Note 2**) and because BSA has particles that can lead to missing spots, the detection printing buffers are filtered with a sterile 0.45 μm filter prior to mixing with individual stock detection antibody solutions. We do not recommend filtering the solutions once antibodies have been added, because the volumes are too small, in the range of 10–20 μL. However, 1 mL of the detection printing buffer containing glycerol, BSA, and Tween 20 can be filtered using a 3 mm luer-lock filter and a 1 mL sterile luer-lock syringe. A 1.5″ 18 G hypodermic luer-lock needle can be fitted to the 1 mL syringe to pick up the detection printing buffer before filtering. Dispose of the needle in a sharps container according to local waste management regulations.
6. The capture antibody printing buffer was optimized to work well on a 2D reactive aldehyde surface. While other surfaces with higher antibody-binding capacity were identified (e.g., high-capacity epoxy surfaces), antibody-target binding was strongest on the reactive aldehyde surface suggesting that antibodies were less denatured or less crowded.
7. Xenobind microarray slides can be cleaned by sonicating ten slides at a time in distilled water, followed by a quick rinse with distilled water and then pure ethanol using squeeze bottles, immediately followed by drying with a stream of compressed nitrogen. It's very important not to let the ethanol air-dry on the microarray slides, as it can leave visible chemical smudges.
8. The main requirements for the microarray printer are spot absolute positional accuracy in both the printing and in positioning

slides on a deck, the compatibility with the high-viscosity printing buffers, and printing speed. Slower (several hours) printing leads to degradation of assay signal for spots that are printed earlier and therefore gives a worse reproducibility performance for the assay.

9. Excessive plasma treatment of microplates can lead to cross-contamination between wells as liquid films form on the surface, and the optimal plasma processing time should be carefully verified. We observed that when loading printing buffers containing betaine and 2,3-butanediol on a 1536-well plate, contamination between adjacent wells occurred readily. Therefore, the 1536-well plate used for capture antibody solutions is not plasma-treated. The additives present in the detection printing buffer (glycerol and BSA) do not however flow easily onto the plastic surface and therefore require a short 10 s plasma treatment in order to easily be loaded into the 1536-well plate.
10. We prepared a flat PDMS surface that we laid on top of the rotary shaker surface by curing approximately 200 mL of PDMS in a 24.5 cm^2 square petri dish normally used for cell culture. Cured PDMS (Sylgard® 184, Corning) is prepared by mixing a ratio of 1:10 of curing agent to the polymer base, mixing thoroughly by hand before pouring into the petri dish. The dish is allowed to stand for 30 min to allow bubbles to escape and is then cured in a 60 °C oven for a minimum of 8 h. The PDMS can then be removed from the petri dish if needed. This step is optional if multiple microarray slide gaskets can be secured at once to the rotary shaker surface.
11. The presence of more than 5% glycerol in the capture antibody printing solution can significantly hinder binding of the antibodies to the surface.
12. If all detection antibodies within an assay are biotinylated, then all can be detected using a fluorescently labeled streptavidin. Alternatively, detection antibodies can be directly labeled with a fluorescent molecule. If detection antibodies are not labeled, they should (1) have been made in an animal species different than that of the capture antibody and (2) be probed with a secondary antibody made in the same species as that of the capture antibody in order to avoid cross-reactivity to the capture antibody. For example, if the capture antibody of an antibody pair is a mouse IgG and the detection antibody is a goat IgG, then the labeled secondary antibody should be a mouse anti-goat IgG. However all matched antibody pairs should be compatible with the labeled secondary antibody within a subarray. Any other combination (such as a capture antibody that is a goat IgG) will lead to false-positive signals due to the labeling of the capture antibody rather than the detection antibody for this assay.

13. Proteins that are obtained by purification from animal or human samples (e.g., cancer related and other proteins that have particular glycosylation patterns that cannot readily be reproduced by recombinant protein synthesis methods) often contain impurities in the form of unrelated proteins, some of which may be targets in the same subarray. This can lead to significant assay signal and cross-reactivity in the standard curves of other targets even in the absence of those targets' specific recombinant antigens [24].
14. Marker pens that dissolve in either water or organic solvents readily leak onto the slide and may smear the surface with highly fluorescent chemicals. We suggest using the solvent-resistant permanent marker from Thermo Fisher Scientific (laboratory marking pen, ref. #2000) which doesn't leak or smear using this protocol.
15. Alignment of printing between the capture and the detection steps is critical for the success of this assay. For some printers, it is as simple as avoiding re-initialization of the printer between the two printing rounds. For others which suffer from printing position drift with time, a method for calibrating the printing head position is necessary. A microscope calibration slide can be used for this in the following way. First, wash the microscope calibration slide with pure ethanol using a squeeze bottle, and dry it with a stream of compressed nitrogen. Before the capture antibody printing step, quickly print a subarray on top of the calibration area, using a solution of 50% glycerol in PBS. View the printed calibration spots under a microscope and make note of the position of a specific spot which lies within the calibration area, compared to the center of the slide. Before performing the detection antibody printing, redo this procedure. Adjust printing margins to compensate for any misalignment that occurred between the two printing rounds.
16. Individual capture antibodies will bind to the surface with different affinities and at different rates. To increase the amount of capture antibody bound on the surface, increase the concentration of that antibody in its printing solution while keeping the concentration of additives (betaine, 2,3-butanediol) the same. Conversely, if fluorescence signals are too strong or streaking is observed around a specific capture antibody, decrease its concentration in the printing solution while keeping the concentration of additives the same. The streaking is due to the surface being saturated by the capture antibody solution and the presence of unbound capture antibodies which quickly bind to the surface during the initial washing, prior to blocking the microarray slide surface.
17. The proper alignment of detection antibody spots onto their corresponding capture antibody spots requires a very precise

positioning of the microarray slides within a slide deck and the accurate alignment of the printing head prior to both printing rounds (*see* **Note 15**). Therefore it is important that slides can be positioned accurately and reproducibly within the slide deck (with less than 10 μm variability). A spring-clamped slide deck and a proper manual technique for loading slides are essential and can achieve this reproducibility.

18. Gaskets that have dust or a very hydrophilic surface can lead to contamination between wells because of an incomplete seal. In order to thoroughly wash gaskets, our method is to first rinse all parts in a mild soap solution, then rinse many times with distilled water, and finally dip for 10–15 s with shaking in pure ethanol. The gasket parts are then dried with a stream of compressed nitrogen, before being assembled and ready for use. They are then stored in a closed petri dish that prevents dust from falling on them.
19. It is important to wash the microarray slides with wash buffer with enough force, which is why the wash buffer is placed in a squeeze bottle. This helps prevent fluorescent streaks on the microarray surface. If streaking of certain spots is still seen in spite of proper washing, slightly decrease the concentration of the capture antibody in the solution for the streaking spots (*see* **Note 16**).
20. Because paper generates a lot of fluorescent dust particles, even low-dust clean-room paper is avoided when wiping microarray slides or surfaces within the working environment. Instead, we use a microfiber cloth of the like that is used to clean lenses or eyeglasses. Particles generated by microfiber cloth are not fluorescent, and they can easily be removed with a stream of compressed nitrogen.
21. If a small film of liquid is present at the top of the gasket, it can lead to contamination between wells when samples and antigens are shaken. Do not cover the gasket unless this cover has a liquid-tight seal. Any seal that is not tight will also lead to contamination between wells and falsify results.
22. Most antibodies and antigens will degrade at varying speeds after they are dried on the surface. Therefore, spots should be protected with 5% trehalose which was found to slow down degradation and in many cases even prevents it. The actual quantity of trehalose left on the surface after drying with the stream of compressed nitrogen is dependent on the drying method. A stream that is strong, and head-on, was found best to achieve an even surface. If the detection spots spread and lose their shape on the surface during detection antibody printing, then there is too much trehalose on the surface. It is then recommended to use a stronger stream of nitrogen and to hold

the nozzle close to the microarray slide (2–3 cm), although using a slightly less concentrated solution of trehalose is also possible. However, a lower concentration of trehalose can lead to more degradation of antibodies and antigens.

23. Binding of streptavidin to the detection antibodies is not limited by the concentration of streptavidin used in these experiments. However, the streptavidin signal will increase with an increase in incubation time. In order to maximize reproducibility across slides, it is important to incubate the streptavidin solution for exactly the same amount of time per slide. Therefore, if it takes 15 s to wash a single slide, apply the streptavidin solution to each slide with a 15 s delay between each slide. This will ensure that all slides have exactly the same streptavidin incubation time.
24. Saturated pixels are pixels that have the maximum value (or very close to the maximum value of 65,535 in a 16-bit system) and are in fact too high to be recorded by the scanner at the gain used. This leads to a loss of data and falsification of results. If a significant number of pixels are saturated for a given gain, all slides of the experiment should be re-scanned with a lower gain.
25. Fluorophores in a dry state slowly degrade in the presence of air (humidity, oxygen, heat), even in a sealed room where ozone is actively removed. For this reason it is important to scan all the slides in an experiment as quickly as possible to minimize the effect of this degradation on reproducibility. If it is necessary to scan slides multiples times at different gains, then first scan all slides at an initial gain, and then scan all of them at a second higher or lower gain depending on the results, rather than scanning each slide at multiple gains. This will maximize reproducibility between microarray slides.
26. When saving fluorescence images of microarray slides, it is very important to save the image data with high dynamic range (16 bit or 20 bit) while avoiding image compression. Formats such as GIF or JPEG may only accommodate 8-bit images and compress the data with information loss that will likely lead to false results. Image formats such as TIFF accommodate 16-bit and 20-bit images and also allow compression using loss-less LZW algorithms for example.
27. In instances where the number n of technical replicates is very low, the mean and standard deviation of the group of technical replicates is very sensitive to the presence of outliers. Therefore a test or method for removing outliers that is efficient at low n is required. A minimum of three technical replicates is required for proper statistics. Grubbs' test performs well with $n = 3$ or more, while Peirce's criterion only works with $n = 4$ or more.

28. Protein quantification is subject to matrix effects that limit the comparison of quantities of a target obtained in samples within a single dilution. Because of the matrix effects, the concentration values are not considered to be absolute, and a target concentration inferred from the 1:3 dilution may be lower than the one inferred from the 1:50 dilution. Different targets are subject to different matrix effects depending on the sample type and the dilution.

Acknowledgments

We wish to thank the National Science and Engineering Research Council (NSERC) of Canada, the Canadian Cancer Society (CCS), and the Canadian Institute of Health Research (CIHR) for funding. D.J. holds a Canada Research Chair (CRC) in Micro- and Nanobioengineering.

References

1. Pla-Roca M, Leulmi RF, Tourekhanova S, Bergeron S, Laforte V, Moreau E, Gosline SJ, Bertos N, Hallett M, Park M, Juncker D (2012) Antibody colocalization microarray: a scalable technology for multiplex protein analysis in complex samples. Mol Cell Proteomics 11(4):M111 011460. doi:10.1074/mcp.M111.011460
2. Juncker D, Bergeron S, Laforte V, Li H (2014) Cross-reactivity in antibody microarrays and multiplexed sandwich assays: shedding light on the dark side of multiplexing. Curr Opin Chem Biol 18:29–37. doi:10.1016/j.cbpa.2013.11.012
3. Laforte V, Olanrewaju A, Juncker D (2013) Low-cost, high liquid volume silicon quill pins for robust and reproducible printing of antibody microarrays. In: MicroTAS: miniaturized systems for chemistry and life sciences, Freiburg, Germany, 27–31 October 2013. Chemical and Biological Microsystems Society (CBMS), pp. 485–487
4. Bergeron S, Laforte V, Lo PS, Li H, Juncker D (2015) Evaluating mixtures of 14 hygroscopic additives to improve antibody microarray performance. Anal Bioanal Chem 407(28):8451–8462. doi:10.1007/s00216-015-8992-8
5. Frampton JP, White JB, Simon AB, Tsuei M, Paczesny S, Takayama S (2014) Aqueous two-phase system patterning of detection antibody solutions for cross-reaction-free multiplex ELISA. Sci Rep 4:4878. doi:10.1038/srep04878
6. Assarsson E, Lundberg M, Holmquist G, Bjorkesten J, Thorsen SB, Ekman D, Eriksson A, Rennel Dickens E, Ohlsson S, Edfeldt G, Andersson AC, Lindstedt P, Stenvang J, Gullberg M, Fredriksson S (2014) Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PLoS One 9(4):e95192. doi:10.1371/journal.pone.0095192
7. Fredriksson S, Dixon W, Ji H, Koong AC, Mindrinos M, Davis RW (2007) Multiplexed protein detection by proximity ligation for cancer biomarker validation. Nat Methods 4(4):327–329. doi:10.1038/nmeth1020
8. Fredriksson S, Gullberg M, Jarvius J, Olsson C, Pietras K, Gústafsdóttir SM, Östman A, Landegren U (2002) Protein detection using proximity-dependent DNA ligation assays. Nat Biotechnol 20:473–477
9. Aldo P, Marusov G, Svancara D, David J, Mor G (2016) Simple plex(TM): a novel multi-analyte, automated microfluidic immunoassay platform for the detection of human and mouse cytokines and chemokines. Am J Reprod Immunol 75(6):678–693. doi:10.1111/aji.12512
10. Blank K, Lankenau A, Mai T, Schiffmann S, Gilbert I, Hirler S, Albrecht C, Benoit M, Gaub HE, Clausen-Schaumann H (2004) Double-chip protein arrays: force-based multiplex sandwich immunoassays with increased specificity. Anal Bioanal Chem 379(7–8):974–981. doi:10.1007/s00216-004-2607-0

11. Christendat D, Yee A, Dharamsi A, Kluger Y, Savchenko A, Cort JR, Booth V, Mackereth CD, Saridakis V, Ekiel I, Kozlov G, Maxwell KL, Wu N, McIntosh LP, Gehring K, Kennedy MA, Davidson AR, Pai EF, Gerstein M, Edwards AM, Arrowsmith CH (2000) Structural proteomics of an archeon. Nat Struct Biol 7(10):903–909
12. Kusnezow W, Jacob A, Walijew A, Diehl F, Hoheisel JD (2003) Antibody microarrays: an evaluation of production parameters. Proteomics 3:254–264
13. Berlier JE, Rothe A, Buller G, Bradford J, Gray DR, Filanoski BJ, Telford WG, Yue S, Liu J, Cheung C-Y, Chang W, Hirsch JD, Beechem JM, Haugland RP, Haugland RP (2003) Quantitative comparison of long-wavelength alexa fluor dyes to Cy dyes: fluorescence of the dyes and their bioconjugates. J Histochem Cytochem 51(12):1699–1712
14. Byerly S, Sundin K, Raja R, Stanchfield J, Bejjani BA, Shaffer LG (2009) Effects of ozone exposure during microarray posthybridization washes and scanning. J Mol Diagn 11(6):590–597. doi:10.2353/jmoldx.2009.090009
15. Cox WG, Beaudet MP, Agnew JY, Ruth JL (2004) Possible sources of dye-related signal correlation bias in two-color DNA microarray assays. Anal Biochem 331(2):243–254. doi:10.1016/j.ab.2004.05.010
16. Anderson GP, Nerurkar NL (2002) Improved fluoroimmunoassays using the dye alexa fluor 647 with the RAPTOR, a fiber optic biosensor. J Immunol Methods 271:17–24
17. Becker W (2012) Fluorescence lifetime imaging–techniquesandapplications.JMicrosc247(2):119–136. doi:10.1111/j.1365-2818.2012.03618.x
18. Li H, Bergeron S, Juncker D (2012) Microarray-to-microarray transfer of reagents by snapping of two chips for cross-reactivity-free multiplex immunoassays. Anal Chem 84(11):4776–4783. doi:10.1021/ac3003177
19. Li H, Munzar JD, Ng A, Juncker D (2015) A versatile snap chip for high-density sub-nanoliter chip-to-chip reagent transfer. Sci Rep 5:11688. doi:10.1038/srep11688
20. Sittampalam GS, Coussens NP, Nelson H, Arkin M, Auld D, Bejcek B, Glicksman M, Inglese J, Iversen PW, Li Z, McGee J, McManus O, Minor L, Napper A, Peltier JM, Riss T, Trask OJ, Weidner J (2016) Assay guidance manual. Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda, MD
21. Wilson JJ, Burgess R, Mao YQ, Luo S, Tang H, Jones VS, Weisheng B, Huang RY, Chen X, Huang RP (2015) Antibody arrays in biomarker discovery. Adv Clin Chem 69:255–324. doi:10.1016/bs.acc.2015.01.002
22. Li H, Bergeron S, Annis MG, Siegel PM, Juncker D (2015) Serial analysis of 38 proteins during the progression of human breast cancer tumor in mice using an antibody colocalization microarray. Mol Cell Proteomics 14(4):1024–1037. 10.1074/
23. Wingren C, Borrebaeck CA (2009) Antibody-based microarrays. Methods Mol Biol 509:57–84. doi:10.1007/978-1-59745-372-1_5
24. Gonzalez RM, Seurynck-Servoss SL, Crowley SA, Brown M, Ommen GS, Hayes DF, Zangar RC (2008) Development and validation of sandwhich ELISA microarrays with minimal assay interference. J Proteome Res 7:2406–2414

Chapter 20

Surface Profiling of Extracellular Vesicles from Plasma or Ascites Fluid Using DotScan Antibody Microarrays

Larissa Belov, Susannah Hallal, Kieran Matic, Jerry Zhou, Sandra Wissmueller, Nuzhat Ahmed, Sumaiya Tanjil, Stephen P. Mulligan, O. Giles Best, Richard J. Simpson, and Richard I. Christopherson

Abstract

DotScan antibody microarrays were initially developed for the extensive surface profiling of live leukemia and lymphoma cells. DotScan's diagnostic capability was validated with an extensive clinical trial using mononuclear cells from the blood or bone marrow of leukemia or lymphoma patients. DotScan has also been used for the profiling of surface proteins on peripheral blood mononuclear cells (PBMC) from patients with HIV, liver disease, and stable and progressive B-cell chronic lymphocytic leukemia (CLL). Fluorescence multiplexing allowed the simultaneous profiling of cancer cells and leukocytes from disaggregated colorectal and melanoma tumor biopsies after capture on DotScan. In this chapter, we have used DotScan for the surface profiling of extracellular vesicles (EV) recovered from conditioned growth medium of cancer cell lines and the blood of patients with CLL. The detection of captured EV was performed by enhanced chemiluminescence (ECL) using biotinylated antibodies that recognized antigens expressed on the surface of the EV subset of interest. DotScan was also used to profile EV from the blood of healthy individuals and the ascites fluid of ovarian cancer patients. DotScan binding patterns of EV from human plasma and other body fluids may yield diagnostic or prognostic signatures for monitoring the incidence, treatment, and progression of cancers.

Key words Exosomes, Shed microvesicles, Chemiluminescence, Chronic lymphocytic leukemia, Ovarian cancer, CD antigen

1 Introduction

Cells in the body secrete extracellular vesicles (EV) that mediate intercellular communication by transporting proteins, RNA, microRNA, and DNA fragments from their cells of origin to other parts of the body via blood and other body fluids [1–5]. Two subsets of EV are distinguished on the basis of size and biogenesis: exosomes (30–100 nm; formed by inward budding of the endosomal membrane) and shed microvesicles (sMV; 100–1000 nm; outward budding from plasma membranes; also known as microparticles,

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_20,

ectosomes, microvesicles, or shed vesicles/particles). Both appear to play important roles in cancer progression, chemoresistance, and immune escape [6–11]. Although their relative contributions have not yet been established, exosomes may be more oncogenic than sMV [12]. Interestingly, however, sMV (30–1300 nm diameter; buoyant density 1.18–1.19 g/mL) purified from the human LIM1863 colorectal cancer (CRC) cell line by sequential centrifugal ultrafiltration [13] appeared to be more invasive than exosomes (30–100 nm; 1.10–1.11 g/mL) from the same cell line.

Cancer cells secrete significantly more EV than noncancerous cells [14]. However, in clinical fluid biopsies, cancer-derived EV can be outnumbered by EV from normal cells by several orders of magnitude, limiting the sensitivity of EV biomarker detection in liquid biopsies [15]. Although disease-specific EV have been identified in blood [16], urine [17, 18], ascites fluid [19], and saliva [20], their detection and proteomic characterization have been challenging [21]. Although advances in flow cytometry have enabled flow cytometric proteomic profiling of EV [22, 23], the number of antigens that can be analyzed simultaneously by this method is limited. The use of mass spectrometry requires purification of the EV of interest from soluble proteins, protein aggregates, and other EV subsets.

In this methods-based protocol, we describe an antibody microarray (DotScan) that was developed for the capture and surface profiling of leukemia cells. The diagnostic capability for analysis of cells from human blood or bone marrow has been validated with a clinical trial that demonstrated >95% correspondence between the diagnoses made using DotScan alone and conventional diagnoses from the multiple criteria routinely used by pathology laboratories [24]. DotScan has also been used to profile surface proteins on live cells recovered from disaggregated colorectal and melanoma tumors [25–29], analysis of cell surface antigens on peripheral blood mononuclear cells (PBMC) from HIV^+ individuals [30–32] and patients with liver disease [33, 34], and stable and progressive disease B-cell chronic lymphocytic leukemia (CLL) [35]. The applicability of DotScan for proteomic characterization and diagnosis of human disease has been reviewed [36].

DotScan has now been adapted for determining the surface profiles of EV recovered from the conditioned growth medium of human cancer cell lines or human plasma or ascites fluid. It also allows the direct comparison of the surface profiles of EV with those of the cells from which they were derived. Cell capture on antibody microarrays can be quantified by optical scanning with a DotScan DotReader and data analysis software (Medsaic Pty Ltd., Darlington, NSW, Australia), while the captured EV require detection by fluorescence or luminescence. Optimization of DotScan methodology was carried out using sMV-depleted EV (consisting largely of exosomes), purified by differential centrifugation from the conditioned growth media of a range of cancer cell lines. The final optimized

method was applied to the profiling of EV from blood to provide "proof of concept." B-cell CLL, the most common leukemia in the Western world [37], was used as a model to study cancer-derived EV that accumulate in the blood [38]. CLL is characterized by the progressive accumulation of mature, monoclonal $CD19^+/CD5^+$ B cells in the peripheral blood, bone marrow, lymph nodes, and spleen [39, 40]. Biotinylated CD19 antibody was used to detect EV captured on DotScan by luminescence. To minimize interference from abundant platelet-derived EV ($CD61^+$) in plasma, CLL-derived EV were depleted of $CD61^+$ particles using Miltenyi CD61 antibody-coated magnetic microbeads.

EV from the ascites fluid of ovarian cancer patients with advanced disease were also profiled on DotScan, using luminescence detection with biotinylated CD326 (EpCAM) antibody alone or in combination with biotinylated CD9 antibody to also detect CD326-negative exosomes in the ascites fluid. CD326 is often overexpressed on ovarian cancer cells and has prognostic significance [41]. The progression of ovarian cancer is characterized by the rapid growth and spread of peritoneal tumors and in most cases is accompanied by the accumulation of ascites within the peritoneum, which either at diagnosis or recurrence carries a bad prognosis [42]. More than one third of ovarian cancer patients at diagnosis, and nearly all patients at recurrence, present with malignant ascites [42]. Hence, ascites is a rich source of tumor material, from which valuable information can be obtained to understand the pathophysiology of ovarian cancer progression and for the development of prognostic and predictive markers [42].

The protocols and results presented here for the profiling of EV from liquid biopsies are based on preliminary findings, some of which are published [43]. Differential centrifugation was used for purification of EV for DotScan analysis, but alternative methods can be used, e.g., ExoQuick precipitation (System Biosciences, Palo Alto, CA, USA), membrane affinity columns, or exoRNeasy Serum/Plasma Kits (Qiagen, Hilden, Germany). Direct DotScan analysis of plasma has yielded unsatisfactory results, presumably due to interference from high levels of plasma proteins and platelet-derived EV.

EV from liquid biopsies were tested without further separation into EV subsets. The strict separation of exosomes and sMV is difficult, as their size distributions can overlap [44]. Separation based on differential protein expression can also be problematic as different subsets of secreted EV may express many common markers [45]. Also, levels of the tetraspanins CD9, CD63, or CD81, often used as exosome markers, may vary or be undetectable on some exosomes [46–49]. Also, tetraspanins have been detected on vesicles that have features of exosomes but originate through budding from the plasma membrane [50, 51]. Tetraspanins have also been detected on the sMV of several cell lines [46, 52]. Differential profiling of disease-specific exosomes and sMV on DotScan awaits

reliable purification methods and discriminatory markers for each subset.

To illustrate the versatility of DotScan, we also compare the binding patterns of EV isolated from the conditioned medium of several cell lines with those of the cells of origin and with cells prepared from disaggregated tumor biopsies from cancer patients. Further enhancements of sensitivity and versatility of DotScan are proposed for the profiling and monitoring of EV from fluid biopsies of patients with solid tumors.

2 Materials

2.1 Preparation of DotScan Antibody Microarrays

1. Oncyte nitrocellulose-coated slides (Grace Bio-Labs, Bend, OR, USA).
2. A panel of antibodies (50 μL each; *see* Table 1; *see* **Note 1**) selected for their ability to recognize extracellular epitopes of proteins on human cells and EV, isotype control antibodies, and a cocktail of antibodies for alignment dots, e.g., CD29, CD44, or other antibodies expected to bind EV of interest (*see* **Note 1**). Antibody solutions can be aliquoted and frozen or stored at 4 °C, as recommended by the manufacturer.
3. PixSys 3200 Aspirate and Dispense System (BioDot, Irvine, CA, USA) (*see* **Note 2**).
4. Phosphate-buffered saline, pH 7.3 (PBS): 150 mM NaCl, 2.7 mM KCl, 1.4 mM Na_2HPO_4, 4.3 mM KH_2PO_4 in Milli-Q water, pH 7.3.
5. Blocking buffer A: 5% (w/v) Diploma skim milk (Fonterra, Macquarie Park, NSW, Australia) in PBS.
6. Refrigerated bench centrifuge: Eppendorf 5810R with a four-place microplate swing-bucket rotor (Eppendorf AG, Hamburg, Germany).
7. Greiner 96-well polypropylene V-well plates (#651201, from Sigma-Aldrich, Castle Hill, NSW, Australia) to hold antibody solutions for application to Oncyte nitrocellulose-coated slides.
8. Greiner EASYseal™ Plate Sealer (#676001, Sigma-Aldrich).
9. AcroPrep 96-well filter plates with 0.45 μm GHP membranes (#PN5030, Pall Life Sciences, Cheltenham VIC, Australia).
10. Drying cabinet (set to 22 °C) for drying slides after antibody dotting, blocking, and washing.
11. Slide staining racks and dishes.
12. Sealable slide storage boxes (e.g., boxes in which Oncyte slides are packaged).
13. Silica gel sachets (#S002, Süd Chemie, JMP Holdings Pty Ltd., Mordialloc VIC, Australia).

Table 1
Antibodies used to make DotScan antibody microarrays

Protein	Antibody clone	Antibody isotype	Antibody source	Catalogue no.	Concentration (μg/mL)	BSA (0.1–0.2%)
TCRα/β	BMA031	IgG2b	BC	IM1466	500	+
TCRγ/δ	Immu510	IgG1	BC	IM1349	100	+
CD1a	BL6	IgG1	BC	IM0130	200	+
CD2	RPA-2.10	IgG1	BD	555324	500	–
CD3	UCHT1	IgG1	BC	IM1304	200	+
CD4	13B8.2	IgG1	BC	IM0398	200	+
CD5	BL1a	IgG2a	BioDesign	P42179M	200	+
CD7	8H8.1	IgG2a	BioDesign	P42179M	200	+
CD8	B9.11	IgG1	BC	IMBULK1	200	+
CD9	ALB6	IgG1	BC	IM0117	200	+
CD10	ALB1	IgG2a	BC	IMBULK2	200	–
CD11a	25.3.1	IgG1	BC	IM0157	200	+
CD11b	BEAR1	IgG1	BC	IM0190	200	+
CD11c	BU 15	IgG1	BC	IM0712	200	+
CD13	WM15	IgG1	BioLegend	301708	500	–
CD14	RM052	IgG2a	BC	IM0643	200	+
CD15	HI98	IgM	BD	555400	500	–
CD15s	CSLEX1	IgM	BD	551344	500	–
CD16	3G8	IgG1	BC	IM0813	200	+
CD19	J3-119	IgG1	BC	IM1313	200	+
CD20	H299 (B1)	IgG2a	BC	6602140	200	+
CD21	BL13	IgG1	BC	IM0111	200	+
CD22	HIB22	IgG1	BD	555423	500	–
CD23	9P.25	IgG1	BC	IMBULK3	200	+
CD24	ALB9	IgG1	BC	IM0118	200	+
CD25	B1.49.9	IgG2a	BC	IM0119	200	+
CD26	M-A261	IgG1	BD	555435	500	–
CD28	CD28.8	IgG1	BC	IM1376	200	+
CD29	K20	IgG2a	BC	IMBULK4	200	–
CD31	1F11	IgG1	BC	IM2052	200	+

(continued)

Table 1
(continued)

Protein	Antibody clone	Antibody isotype	Antibody source	Catalogue no.	Concentration (μg/mL)	BSA (0.1–0.2%)
CD32	2E1	IgG2a	BC	IM0417	200	+
CD33	WM53	IgG1	BD	555449	500	–
CD34	QBEND 10	IgG1	BC	IM0786	200	+
CD36	FA6-152	IgG1	BC	IM0765	200	+
CD37	M-B371	IgG1	BD	555456	500	–
CD38	T16	IgG1	BC	IM0366	200	+
CD40	MAB89	IgG1	BC	IM1374	200	+
CD41	P2	IgG1	BC	IM0145	200	+
CD42a	SZ1	IgG2a	BC	IM0538	200	+
CD43	DFT1	IgG1	BC	IM1843	200	+
CD44	J.173	IgG1	BC	IM0845	200	+
CD44v6	2F10	IgG1	R&D Systems	BBA13	200	–
CD45	HI30	IgG1	BD	555480	500	–
CD45RA	ALB11	IgG1	BC	IM0537	200	+
CD45RO	UCHL1	IgG2a	BD	555491	500	–
CD47	B6H12	IgG1	BD	556044	500	–
CD49b	AK-7	IgG1	BioLegend	314304	500	–
CD49c	ASC-1	IgG1	BioLegend	34801	500	–
CD49d	HP2/1	IgG1	BC	IM0764	200	+
CD49e	SAM1	IgG2b	BC	IM0771	200	+
CD49f	GoH3	IgG2a	BD	555734	500	–
CD51	NKI-M9	IgG2a	BioLegend	327902	500	+
CD52	YTH66.9HL	IgM	AbDSerotec	MCA349	200	+
CD54	84H10	IgG1	BC	IM0544	200	+
CD55	IA10	IgG2a	BD	555691	500	–
CD56	C218	IgG1	BC	IM1844	200	+
CD57	NK-1	IgM	BD	555618	500	–
CD58	HCD58	IgG1	BioLegend	322502	500	+
CD59	p282	IgG2a	BD	555761	500	–
CD60	M-T6004	IgM	AbDSerotec	MCA1314	500	–

(continued)

Table 1
(continued)

Protein	Antibody clone	Antibody isotype	Antibody source	Catalogue no.	Concentration (μg/mL)	BSA (0.1–0.2%)
CD61	SZ21	IgG1	BC	IM0540	200	+
CD62L	DREG-56	IgG1	BD	555542	500	–
CD62E	1.2B6	IgG1	BC	IM1243	200	+
CD62P	CLB-Thromb/6	IgG1	BC	IM1315	200	+
CD63	H5C6	IgG1	BD	556019	500	–
CD64	10.1	IgG1	BD	555525	500	–
CD66a	29H2	IgG1	abcam	ab49510	NA	–
CD66b	G10F5	IgM	BioLegend	305102	500	–
CD66c	KOR-SA3544	IgG1	MBL International	D028-3	500	+
CD66e (CEA)	C365D3(NCRC23)	IgG1	AbDSerotec	MCA1744	200	–
CD69	FN50	IgG1	BD	555529	500	–
CD71	M-A712	IgG2a	BD	555534	500	–
CD77	38-13	IgM	BC	IM0175	150	+
CD79a	HM47	IgG1	BD	555934	500	–
CD79b	CB3.1	IgG1	BD	555678	500	–
CD80	MAB104	IgG1	BC	IM1449	200	+
CD82	ASL-24	IgG1	BioLegend	342102	500	–
CD86	HA5.2B7	IgG2b	BC	IM2728	500	+
CD87	VIM5	IgG1	BD	555767	500	–
CD88	D53-1473	IgG1	BD	550493	500	–
CD95	UB2	IgG1	BC	IM1505	250	+
CD98	UM7F8	IgG1	BD	556074	500	–
CD102	B-T1	IgG1	AbDSerotec	MCA1140	200	+
CD103	2G5	IgG2a	BC	IM0318	200	+
CD104	450-9D	IgG1	BD	555721	500	–
CD117	YB5.B8	IgG1	BD	555713	500	+
CD120a	H398	IgG2	AbDSerotec	MCA1340	200	+
CD122	MIK-beta 1	IgG2a	AbDSerotec	MCA1941	500	–

(continued)

Table 1
(continued)

Protein	Antibody clone	Antibody isotype	Antibody source	Catalogue no.	Concentration (μg/mL)	BSA (0.1–0.2%)
CD126	M5	IgG1	BD	551462	500	–
CD128	5A12	IgG2b	BD	555937	500	–
CD130	AM64	IgG1	BD	555756	500	–
CD133	EMK08	IgG2b	eBioscience	14-1339	500	–
CD134	ACT35	IgG1	BD	555836	500	–
CD135	SF1-340	IgG1	BC	IM2036	200	+
CD138	DL-101	IgG1	BD	550804	500	+
CD151	14A2.H1	IgG1	BD	556056	500	–
CD154	TRAP1	IgG1	BC	IM1842	200	+
CD166	3A6	IgG1	BD	559260	500	–
CD175s	3H1951	IgG1	Santa Cruz	Sc-70558	200	–
CD177	MEM-166	IgG1	BioLegend	315802	500	–
CD184	1D9	IgG2a	BD	551413	500	+
CD227	HMPV	IgG	BD	555925	500	–
CD235a	11E4B7.6	IgG1	BC	IM2210	200	+
CD244	2-69	IgG2a	BD	550814	500	–
CD255	CARL-1	IgG3	BioLegend	308302	500	–
CD261	DJR1	IgG1	BioLegend	307202	500	–
CD262 (DR5)	DJR2-4	IgG1	BioLegend	307402	500	–
CD324	67A4	IgG1	BioLegend	324102	500	–
CD326 (EpCAM)	158206	IgG2b	R&D Systems	MAB9601	500	–
CD340	24D2	IgG1	BioLegend	324402	500	–
Annexin II	5/Annexin II	IgG1	BD	610069	250	+
A33	402104	IgG2a	R&D Systems	MAB3080	200	–
β-Catenin	polyclonal	IgG	R&D Systems	AF1329	1000	–
CA-125	4H9	IgG1	Lifespan Biosciences	LS-C53346	500	+
CA 19-9	2 clones	IgG1	Abnova	MAB1399	1000	+
Claudin-4	382321	IgG2a	R&D Systems	MAB4219	200	–

(continued)

Table 1
(continued)

Protein	Antibody clone	Antibody isotype	Antibody source	Catalogue no.	Concentration (μg/mL)	BSA (0.1–0.2%)
DCC	G92-13	IgG1	BD	554222	500	–
EGFR	EGFR.1	IgG2b	BD	555996	500	–
ErbB3/HER-3	1B4C3	IgG2a	BioLegend	324702	500	–
FAP	F11-24	IgG1	Calbiochem	OP188	100	–
FMC7	FMC7	IgM	Millipore	MAB1217	250	+
Galectin 3	A3A12	IgG1	abcam	ab27850	500	–
Galectin 4	198616	IgG2a	R&D Systems	MAB1227	1000	–
Galectin 8	210608	IgG2a	R&D Systems	MAB1305	1000	+
HLA-A,B,C	G46-2.6	IgG1	BD	555551	500	–
HLA-DR	B8.12.2	IgG2b	BC	IM0108	200	+
HLA-G	4H84	IgG1	BD	557577	500	–
Hsp27	3 Clones	IgG2a/2b	abcam	ab78307	1000	–
Hsp70	520608	IgG1	R&D Systems	MAB6010	500	–
Hsp90	68/Hsp90	IgG1	BD	610419	500	–
κ light chain	6E1	IgG1	BC	IM0173	500	+
λ light chain	C4	IgG1	BC	IM0174	500	+
MAGE-1	polyclonal	IgG	GeneTex	GTX16031	200	+
MICA	159227	IgG2b	R&D Systems	MAB1300	500	+
MMP-14	128527	IgG2b	R&D Systems	MAB9181	500	–
pIgR	polyclonal	IgG	R&D Systems	AF2717	200	–
sIg (IgA, G, M)	polyclonal	IgG	Chemicon	982320020	250	+
TSP-1	46.4	IgG1	Calbiochem	BA18	500	+
mIgG1	MOPC-21	IgG1	BD	554121	500–50	+ or -
mIgG2a	G155-178	IgG2a	BD	555571	500–50	+ or -
mIgG2b	27–35	IgG2b	BD	555740	500–200	+ or -
mIgG3	J606	IgG3	BD	555577	500	
mIgM	G155-228	IgM	BD	555581	500–50	+ or -
Mabthera	Chimeric murine/human	IgG1	Roche	R 60318	200	+

NA not available, *BSA* bovine serum albumin, *BC* Beckman Coulter, *BD* Becton Dickinson

2.2 Purification of EV and Cells from Cell-Line Growth Medium

1. Incomplete RPMI-1640 growth medium: RPMI-1640 medium containing 2 mM glutamine, and 100 U/mL penicillin/streptomycin, all from Thermo Fisher Scientific (Waltham, Massachusetts, USA).
2. Complete RPMI-1640 growth medium: incomplete RPMI-1640 growth medium with 10% (v/v) fetal calf serum (FCS; In Vitro Technologies, Noble Park North, NSW, Australia).
3. EV-depleted FCS (*see* **Note 3**): prepared by removing bovine EV from FCS by ultracentrifugation (100,000 × *g*, 16–18 h, 4 °C) [53] and filtering the supernatant through a 0.22 μm filter (Merck Millipore, VIC, Australia). Exo-FBS™ exosome-depleted FBS is also available commercially (System Biosciences, Palo Alto, CA, USA).
4. 0.05% trypsin-EDTA (Thermo Fisher Scientific).
5. Refrigerated bench centrifuge: Eppendorf 5810R with swing-bucket rotor and conical tube adaptors (Eppendorf).
6. Polypropylene centrifuge tubes.
7. Ultracentrifuge: Beckman Optima L-100 XP with SW32.1 Ti rotor (Beckman Coulter, Gladesville, NSW, Australia).
8. Polyallomer ultracentrifugation tubes (#337986) and quick-seal polypropylene centrifuge tubes (#356562) from Beckman Coulter.
9. Ultrafiltration Discs, Ultracel regenerated cellulose, 100 kDa NMWL, 63.5 mm (#PLHK 06210; Millipore, North Ryde, VIC, Australia).
10. Amicon® stirred cell, 200 mL capacity, with magnetic stirrer and nitrogen gas pressure source.

2.3 Purification of EV and PBMC from Blood

1. Blood collected into Vacuette® 9NC sodium citrate collection tubes (Greiner Bio-One, Frickenhausen, Germany) (*see* **Notes 4** and **5**).
2. Histopaque®-1077 (#10771; Sigma-Aldrich).
3. Protease inhibitor cocktail stock solution (×10): dissolve a cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail tablets (#11873580001; Sigma-Aldrich) in PBS (1 tablet/mL) and store in 0.5 mL aliquots at −20 °C.
4. Plasma dilution buffer: 5 mM EDTA in PBS, pH 7.3 (*see* **Note 6**).
5. Magnetic cell sorting (MACS) buffer stock solution (×20): 10% (v/v) bovine serum albumin (BSA; #A2058; Sigma-Aldrich), 2 mM EDTA, PBS, pH 7.2. Store frozen at −20 °C.
6. MACS buffer (0.5% (w/v) BSA, 2 mM EDTA in PBS, pH 7.2): prepared by adding 5 mL MACS Buffer Stock Solution (×20) to 95 mL PBS containing 2 mM EDTA, pH 7.2. Degas before use.

7. CD61 antibody-coated magnetic microbeads (CD61 MicroBeads; #130-051-101; Miltenyi Biotec, Macquarie Park, NSW, Australia).
8. Eppendorf® LoBind microcentrifuge tube (#Z666513).
9. LS microcolumn (#130-042-401; Miltenyi Biotec).
10. QuadroMACS separator (#130-090-976; Miltenyi Biotec).
11. Ultrafiltration Discs, Ultracel regenerated cellulose, 100 kDa NMWL, 25 mm (#PLHK 02510; Millipore, North Ryde, VIC, Australia).
12. Amicon® stirred cell, 3 mL capacity (#5125; Millipore), with magnetic stirrer and nitrogen gas pressure source.

2.4 Purification of EV from Ascites Fluid

1. Ascites (peritoneal fluid, 50 mL).
2. Bench centrifuge (*see* Subheading 2.2, **item 5**).
3. Ultracentrifuge (*see* Subheading 2.2, **item 7**).
4. 100 kDa Amicon ultra-15 centrifugal filter unit (# UFC910008; Millipore).

2.5 Preparation of a Viable Cell Suspension from Disaggregated CRC Tissue

1. Surgically resected patient CRC tissue samples in Hanks' balanced salt solution, pH 7.3 (HBSS; #H6136, Sigma-Aldrich).
2. Surgical blades (#090609, Livingstone International, Rosebery, NSW, Australia).
3. Tissue disaggregation buffer: RPMI-1640 medium containing 2% (v/v) collagenase type 4 (Worthington, Lakewood, NJ, USA) and 0.1% (w/v) deoxyribonuclease I from bovine pancreas (DNase I; Sigma-Aldrich).
4. Fine wire mesh strainer and plunger from a 10 mL syringe.
5. 200 and 50 μm Filcon filters (BD Biosciences, Franklin Lakes, NJ, USA).

2.6 DotScan Analysis of Live Cells

1. DotScan antibody microarrays (*see* Subheading 2.1).
2. Viable cell suspension at a density of 10–17 × 10^6 cell/mL in incomplete RPMI-1640 medium.
3. PBS (*see* Subheading 2.1, **item 4**).
4. Fixative: 3.7% (w/v) formaldehyde solution—prepared by diluting 37% (w/v) formaldehyde solution (#252549; Sigma-Aldrich) 1/10 in PBS.
5. Coplin jars (26 mm × 26 mm × 90 mm) for fixation and washing of slides.
6. DotScan DotReader and data analysis software (Medsaic Pty Ltd., Darlington, NSW, Australia) (*see* **Note 7**).

2.7 DotScan Analysis of EV Using Enhanced Chemiluminescence (ECL)

1. DotScan antibody microarrays (*see* Subheading 2.1).
2. Hydrophobic pen (Vector Laboratories, Burlington, CA, USA).
3. PBS (*see* Subheading 2.1, **item 4**).
4. EV resuspension buffer: incomplete RPMI-1640 medium containing 2% heat-inactivated human AB serum (Sigma-Aldrich) (*see* **Note 8**).
5. Fixative: 3.7% (w/v) formaldehyde solution (*see* Subheading 2.6, **item 4**).
6. Blocking buffer B: 1% BSA (w/v) in PBS, pH 7.3 (store frozen at −20 °C).
7. Detection antibodies: dilute in blocking buffer B (*see* **item 6**) immediately before use, as follows:
   - 1/10,000 dilution (final concentration 0.05 μg/mL) of biotinylated CD326 antibody (EpCAM; clone 9C4; #324216; BioLegend, San Diego, California, USA) for detection of $EpCAM^+$cancer-derived EV.
   - 1/200 dilution (2.5 μg/mL) of biotinylated CD19 antibody (clone HIB19; #302203; BioLegend) for detection of B cells, including CLL.
   - 1/200 dilution (5 μg/mL) of biotinylated CD9 antibody (clone MEM-61; #ab28094; Abcam, Cambridge, UK) for detection of $CD9^+$exosomes.
   - 1/200 dilution of biotinylated CD45 antibody (clone HI130: #304003; BioLegend) for detection of leukocyte-derived EV.
8. Pierce Streptavidin Poly-HRP (0.5 mg/mL; Thermo Fisher Scientific): dilute 1/40,000 in blocking buffer B (*see* **item 6**) immediately before use.
9. SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) – prepared according to manufacturer's instructions.
10. Coplin jars (*see* Subheading 2.6, **item 5**).
11. Clear plastic sheets (e.g., overhead projector sheets).
12. Amersham hyperfilm (GE Healthcare Life Sciences, Parramatta, NSW, Australia).
13. X-ray cassette.

2.8 DotScan Analysis of CRC Cells with Fluorescence Detection

The following are required in addition to the materials listed in **items 1–4** under Subheading 2.6:

1. Blocking buffer C: 2% BSA (w/v), 2% heat-inactivated AB serum, PBS, pH 7.3 (store frozen at −20 °C).

2. Detection antibody: 1/15 dilution of Alexa Fluor 647-conjugated EpCAM antibody (clone 9C4; #324212; BioLegend) in blocking buffer C (*see* **item 1**).
3. PBS (*see* Subheading 2.1, **item 4**).
4. Coplin jars (*see* Subheading 2.6, **item 5**).
5. Typhoon FLA 9000 scanner (GE Healthcare, Rydalmere, NSW, Australia).

2.9 DotScan Analysis of CRC-Derived EV with Fluorescence Detection

The following are required in addition to the materials listed in **items 1–5** under Subheading 2.7:

1. Blocking buffer C: (*see* Subheading 2.8, **item 1**).
2. Detection antibody: 1/100 dilution of Alexa Fluor 647-conjugated EpCAM antibody (clone 9C4; #324212; BioLegend) in blocking buffer C.
3. PBS (*see* Subheading 2.1, **item 4**).
4. Coplin jars (*see* Subheading 2.6, **item 5**).
5. Typhoon FLA 9000 scanner (GE Healthcare).

2.10 Quantification of EV/Cell Binding on DotScan

ImageQuant (version 7; GE Healthcare).

2.11 Nanoparticle Tracking Analysis (NanoSight)

1. NanoSight LM10-HS system (NanoSight Ltd., Amesbury, UK) with a 405 nm laser and nanoparticle tracking analysis software (NTA version 2.3; Malvern Instruments, Malvern, UK).
2. PBS (*see* Subheading 2.1, **item 4**).

3 Methods

The principles of DotScan analysis are shown in Fig. 1.

3.1 Preparation of DotScan Antibody Microarray

1. To remove precipitated material from antibody solutions, filter 50 µL of each antibody through the moistened wells of an AcroPrep 96-well filter plate with 0.45 µm GHP membranes by centrifugation into polypropylene V-well plates (*see* Subheading 2.1, **items 8** and **9**), using a swing-out rotor and plate adaptors (275 × *g*, 3 min).
2. Place the Oncyte slides on the platform of the arrayer, and program the arrayer to deliver 10 nL volumes of each antibody per dot, 800 µm apart, creating duplicate rectangular microarrays per slide, 7 dots/row/array with 1.6 mm between arrays, surrounded by alignment dots (*see* Subheading 2.1, **item 2**).

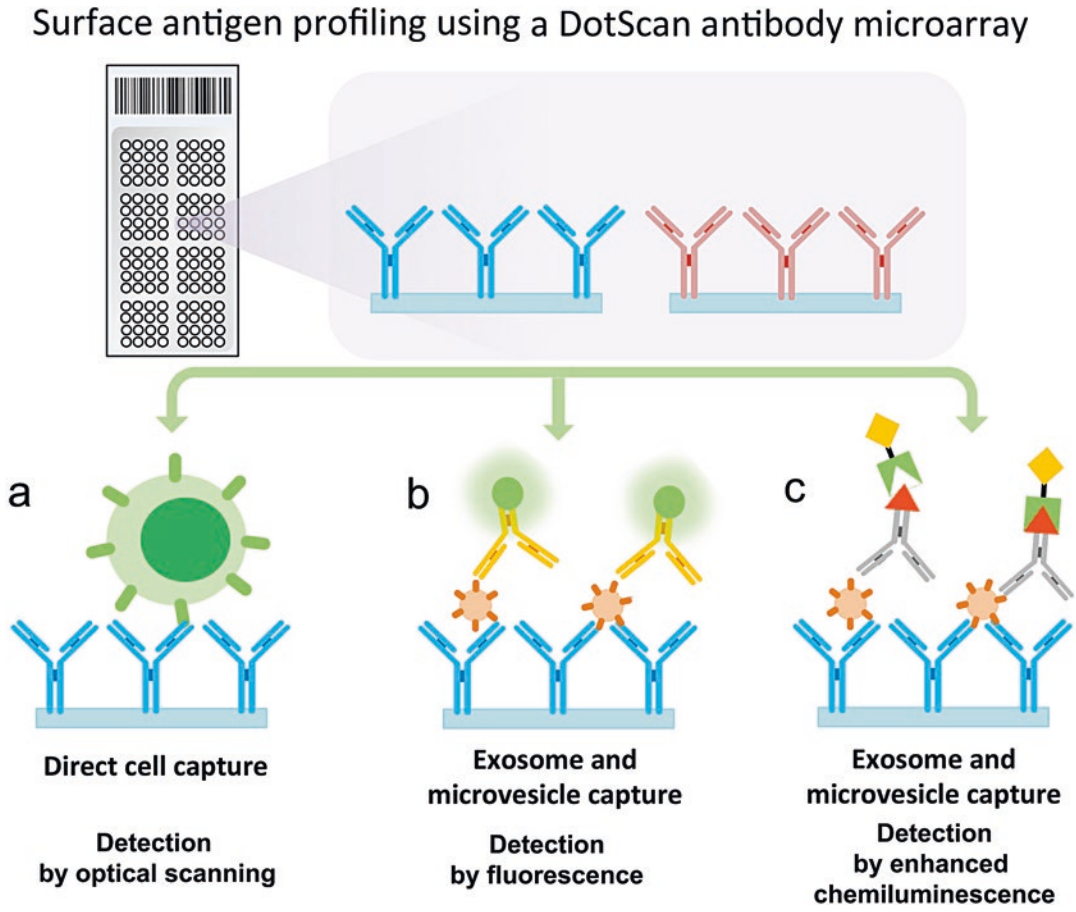


Fig. 1 Schematic representation of capture and detection of cells and EV on immobilized antibody dots of DotScan (*see* **Note 16**): (*a*) Captured cells are usually detected by optical scanning; however, fluorescent antibodies can be used for fluorescence multiplexing of cells [28, 29]. (*b*) EV are detected with fluorescent antibodies (*b*), or (*c*) by enhanced chemiluminescence (ECL), with biotinylated antibodies and Streptavidin Poly-HRP

3. Allow antibody microarrays to dry on the platform for 30 min.
4. Arrange the slides into slide staining racks and immerse in a dish of blocking buffer A (*see* Subheading 2.1, **item 5**), cover, and leave overnight (17–18 h) at 4 °C to block protein binding sites.
5. After blocking, wash the slides by immersion in two changes of Milli-Q water (900 mL, 20 s each).
6. Dry the washed racks of slides in a drying cabinet (25 °C, ~1 h).
7. Pack dry arrays into a slide box; add a sachet of silica gel desiccant, seal with tape, and store at 4 °C.

3.2 Purification of EV and Cells from Cell-Line Growth Medium

For optimization of EV analysis by DotScan, exosome-enriched EV were purified from the conditioned medium of a range of leukemia and cancer cell lines using a method based on previous protocols [53, 54]. It is important to use gentle procedures during purification and handling of EV, as recommended by the International Society for Extracellular Vesicles (ISEV) [45]. Intact and structurally damaged EV will both be captured and detected on DotScan, provided they express the antigens recognized by both the immobilized and detection antibodies. However, damage to EV membranes may result in the exposure and detection of intracellular proteins, in addition to external surface proteins.

1. Grow cells to late exponential phase in 175 cm^2 flasks using the appropriate medium recommended for the cell line (*see* Subheading 2.2, **item 1**).
2. Wash cells twice with PBS and incubate (24 h, 37 °C) with 16–20 mL of growth medium containing 10% (v/v) EV-depleted FCS (*see* Subheading 2.2, **item 3**; *see* **Note 3**).
3. Collect conditioned medium after growth of cells. For suspension cultures, remove cells (400 × *g*, 5 min, 23 °C) and collect supernatant.
4. Collect cells for DotScan profile comparison with their EV profile. Harvest attached monolayers of cells by 5 min incubation at 37 °C with 0.05% trypsin-EDTA.
5. To prepare exosome-enriched EV, centrifuge conditioned medium (1500 × *g*, 10 min, 23 °C, followed by 10,000 × *g*, 20 min, 4 °C) to remove cell debris and large vesicles (apoptotic bodies and sMV).
6. Concentrate exosome-enriched EV ~ fourfold by ultrafiltration [55, 56] on Millipore 100 kDa filters in an Amicon® stirred cell, 200 mL capacity, with magnetic stirrer and nitrogen gas pressure source.
7. Pellet EV by ultracentrifugation (100,000 × *g*, 16 h, 4 °C).
8. EV pellets can be resuspended in 200 μL incomplete RPMI-1640 medium or EV resuspension buffer (*see* Subheading 2.7, **item 4**) and analyzed immediately or stored at −80 °C and thawed quickly in a 37 °C water bath before resuspension for DotScan and NanoSight analyses (*see* Subheadings 3.7 and 3.11).

3.3 Purification of EV and PBMC from Blood

1. Collect blood (10 mL) into citrate anticoagulant tubes (*see* **Notes 4** and **5**).
2. Transport blood at ambient temperature and process within 4 h from collection.
3. If peripheral blood mononuclear cells (PBMC) are to be collected in addition to plasma, layer blood carefully on top of

Histopaque in 15 mL polypropylene centrifuge tube (2 volumes blood to 1 volume Histopaque), and centrifuge in a swing-out rotor, without brake (400 × *g*, 30 min, 23 °C).

4. Transfer plasma into a polypropylene centrifuge tube without disturbing the PBMC layer at the interface between the Histopaque and plasma.
5. Collect the PBMC layer carefully into a separate tube, wash with PBS, and analyze by DotScan as described under Subheading 3.6 (*see* **Note 9**).
6. To the plasma, add protease inhibitor cocktail stock solution (×10) at 1 mL per 9 mL plasma. This step may not be necessary due to the stability of exosomes in plasma [57].
7. Centrifuge plasma (three times, 2500 × *g*, 20 min, 4 °C) to remove platelets and cell debris, according to Ghosh et al. [38] (*see* **Note 10** for alternative procedure).
8. Clarified plasma can be stored at −80 °C.
9. Thaw frozen clarified plasma (3.5–5 mL) quickly in a 37 °C water bath, and dilute to 17 mL with plasma dilution buffer (*see* Subheading 2.3, **item 4**; *see* **Note 6**) in a polyallomer ultracentrifugation tube (16 × 102 mm).
10. Pellet EV (100,000 × *g*, 16 h, 4 °C) and resuspend in 200 μL MACS buffer, with thorough disaggregation of EV pellets (*see* **Note 11**).
11. Remove platelet-derived EV and collect CD61-depleted EV as follows.
12. In an Eppendorf® LoBind microcentrifuge tube, add CD61 MicroBeads to the EV suspension in the ratio of 18 μL beads per mL of original plasma.
13. Rotate gently (10 rpm, 1 h, 4 °C).
14. Pass through an LS column in a strong magnetic field using a QuadroMACS separator.
15. Wash the column with 2 mL MACS buffer.
16. Collect the eluent (CD61-depleted EV) into a 16 × 38 mm, quick-seal polypropylene centrifuge tube.
17. Pellet CD61-depleted EV by ultracentrifugation (100,000 × *g*, 3 h, 4 °C).
18. EV pellets can be resuspended in 200 μL in EV resuspension buffer (*see* Subheading 2.7, **item 4**) and analyzed immediately or stored at −80 °C and thawed quickly in a 37 °C water bath before resuspension for DotScan and NanoSight analyses (*see* Subheadings 3.7 and 3.11).
19. EV captured on the CD61 MicroBeads can also be collected for DotScan analysis (*see* **Note 12**).

3.4 Purification of EV from Ascites Fluid

Ascites fluid (50 mL) was collected from patients with advanced-stage serous ovarian adenocarcinoma, with ethics approval from the Research and Human Ethics Committee of the Royal Women's Hospital (Melbourne, Australia). EV were purified in the laboratory of Dr. Nuzhat Ahmed (Department of Obstetrics & Gynaecology, Women's Cancer Research Center, Royal Women's Hospital, Parkville, VIC, Australia) as follows:

1. Centrifuge ascites fluid (300 × *g*, 10 min, RT) to remove cells.
2. Centrifuge (2000 × *g*, 20 min, RT, three times) to remove apoptotic bodies.
3. Concentrate to ~1 mL using a 100 kDa Amicon ultra-15 centrifugal filter unit.
4. Pellet EV from the concentrated fluid by ultracentrifugation (100,000 × *g*, 16 h, 4 °C).
5. Resuspend in 200 μL PBS and store at −80 °C.

3.5 Preparation of a Viable Cell Suspension from Disaggregated CRC Tissue

1. Collect patient tissue samples into Hanks' balanced salt solution (HBSS) at 4 °C. Process within 3 h to maintain cell viability.
2. Cut samples into 2 mm cubes and incubate with occasional gentle mixing for 60 min at 37 °C with an equal volume of tissue disaggregation buffer (*see* Subheading 2.5, **item 3**).
3. Gently force the semi-digested tissue through a fine wire mesh strainer using the plunger from a 10 mL syringe and wash through with 2 mL HBSS.
4. Pass the cell suspension sequentially through 200 and 50 μm Filcon filters to remove cell aggregates.
5. Cells can be analyzed immediately or stored at −80 °C in heat-inactivated FCS with 10% dimethyl sulfoxide (DMSO).
6. Before profiling on DotScan, thaw frozen disaggregated cells quickly in a 37 °C water bath.
7. Add 10 mL HBSS, centrifuge (410 × *g*, 20 °C, 5 min), and discard the supernatant.
8. Resuspend the cells in 500 μL HBSS and treat with 0.1% (w/v) DNase I (20 min, 23 °C).
9. Determine the cell viability by trypan blue exclusion.
10. Wash cells with HBSS as above (**step 7**).
11. Resuspend cells to ~5 × 10^6 viable cells per 300 μL in incomplete RPMI-1640 medium for DotScan analysis (*see* Subheading 3.6).

3.6 DotScan Analysis of Live Cells

The capture of cells on antibody dots of DotScan requires live cells, as the active "capping" of CD antigens to the interface between the cells and dots increases binding interactions [58]. In addition,

electron microscopy has revealed the production of fine cellular projections (filopodia) by captured leukocytes, extending from the cells into to the antibody-coated nitrocellulose [59]. Longer incubation times are required for the binding of large cells (e.g., CRC) than for small cells (e.g., leukemia).

1. Prepare washed, live cell suspensions, and resuspend 3–5 × 10^6 cells in 300 μL incomplete RPMI-1640 medium (*see* **Note 13**).
2. Moisten the nitrocellulose strip by dipping into PBS for 20 s; then carefully dry the glass slide around the nitrocellulose with a folded tissue.
3. Pipette the cell suspension evenly over the moist DotScan microarray; then place horizontally in a humidified chamber.
4. Incubate for 1 h at 37 °C for CRC cell lines and cells from disaggregated CRC biopsies or 30 min at room temperature (23 °C) for MEC1 cells (a $CD5^-/CD19^+$ cell line originating from a patient with CLL [60]). For leukemia cells from the blood of CLL patients, the incubation period should be limited to 12–15 min at room temperature. Longer incubation can result in non-specific binding of CLL cells to the surface of the slide.
5. Wash off unbound cells with a single vertical immersion in PBS (20 mL, 10–20 s).
6. Fix captured cells to the nitrocellulose (2 h, 23 °C) by immersing the slide in 15 mL fixative (*see* Subheading 2.6, **item 4**) or by gently pipetting fixative (1 mL) onto the slide placed horizontally in a humidified chamber.
7. Wash slides by vertical immersion in three changes of PBS (20 mL, 2 min each), and wipe the bottom of the slide dry.
8. While the nitrocellulose is still moist and translucent, record cell binding patterns by optical scanning, using DotScan DotReader and data analysis software that records digital images of cell binding patterns on microarrays. If a DotReader is not available and/or for increased sensitivity, cell binding can be visualized by fluorescence detection (*see* Subheading 3.8; *see* **Note 14**).
9. Quantify cell binding by ImageQuant (*see* Subheading 3.10).

3.7 DotScan Analysis of EV Using Enhanced Luminescence (ECL)

Due to their small size, EV are captured easily on immobilized antibodies, compared to live cells (*see* Subheading 3.6). Unlike cells, EV suspensions can be gently rocked on antibody microarrays to enhance capture of the particles.

1. Draw a border around the antibody microarray with a hydrophobic pen to restrict liquid samples to the microarray area.
2. Moisten the nitrocellulose strip by dipping into PBS for 20 s; then carefully dry the glass slide around the hydrophobic border.

3. Incubate EV (10^8–10^{11} particles/200 μL in EV resuspension buffer or incomplete RPMI-1640 medium; *see* **Note 15**) on DotScan at 4 °C for 16 h, with gentle rocking in a moist humidified chamber.
4. Wash off unbound EV with a single vertical immersion in PBS (20 mL, 10–20 s).
5. Fix captured EV to the nitrocellulose (2 h, 23 °C) by immersing the slide in 15 mL fixative (*see* Subheading 2.6, **item 4**) or by gently pipetting fixative (1 mL) onto the slide placed horizontally in a humidified chamber.
6. Wash slides by vertical immersion in three changes of PBS (20 mL, 2 min each).
7. Without allowing the slide to dry, add 200 μL blocking buffer B (*see* Subheading 2.7, **item 6**) and incubate (20 min, 23 °C).
8. Pour off the blocking buffer and add 200 μL detection antibody (*see* Subheading 2.7, **item 7**).
9. After 60 min incubation at 23 °C in a humidified chamber, wash microarrays by vertical immersion in three changes of PBS (20 mL, 2 min each).
10. Add 200 μL Streptavidin Poly-HRP (*see* Subheading 2.7, **item 8**) and incubate (30 min at 23 °C).
11. Wash by vertical immersion in four changes of PBS (20 mL, 2 min each).
12. Add 300 μL SuperSignal West Pico Chemiluminescent Substrate (*see* Subheading 2.7, **item 9**).
13. After 5 min at 23 °C, drain off excess reagent, and immediately place the moist slides between two pieces of clear plastic (e.g., overhead projector sheets) inside an X-ray cassette, avoiding air bubbles.
14. In a darkroom, carefully place Amersham Hyperfilm ECL on top of the plastic, and close the cassette. The exposure period depends on the number of captured EV that express the antigen recognized by the detection antibody. For example, strong luminescence on an antibody microarray incubated with 10^{11} human LIM1215 CRC-derived EV requires only 30 s exposure with EpCAM detection, while 7.8×10^8 particles require 10 min exposure. CD61-depleted EV isolated from 10 mL of blood from an advanced CLL patient require 30 min exposure with CD19 detection.
15. Develop the ECL film, and scan with a GS-900™ Calibrated Densitometer. Visible dots were considered to be positive.
16. Quantify dot luminescence intensities using ImageQuant (*see* Subheading 3.10).

3.8 DotScan Analysis of CRC Cells with Fluorescence Detection

Fluorescence multiplexing allows DotScan profiling of CRC cells and T cells in disaggregated tumor tissue from CRC patients [27–29].

1. Capture cells on DotScan (*see* Subheading 3.6, **steps 1–4**).
2. Wash slides by vertical immersion in three changes of PBS (20 mL, 20 s each).
3. Fix captured cells to the slides (20 min, 23 °C) by immersion in 15 mL fixative (*see* Subheading 2.6, **item 4**).
4. Wash slides as above (**step 2**).
5. Without allowing the slides to dry, add 200 μL blocking buffer C (*see* Subheading 2.8, **item 1**) and incubate (20 min, 23 °C).
6. Pour off blocking buffer and add 150 μL detection antibody (*see* Subheading 2.8, **item 2**).
7. Incubate in a humidified chamber in the dark (30 min, 23 °C).
8. Wash off unbound antibody with three changes of PBS (20 mL, 30 s each).
9. Allow the microarrays to dry in the dark (23 °C).
10. Scan for Alexa Fluor 647 with a Typhoon FLA 9000 scanner with a 633 nm excitation laser and a 670 BP30 emission filter, with resolution set to 50 μm.

3.9 DotScan Analysis of CRC-Derived EV with Fluorescence Detection

1. After capture and fixation of the CRC-derived EV on DotScan (*see* Subheading 3.7, **steps 1–5**), wash slides by vertical immersion in three changes of PBS (20 mL, 2 min each).
2. Without allowing the slide to dry, block (20 min, 23 °C) with 200 μL blocking buffer C (*see* Subheading 2.8, **item 1**).
3. Pour off blocking buffer and add 150 μL diluted Alexa Fluor 647-conjugated EpCAM antibody (*see* Subheading 2.9, **item 2**).
4. Incubate in a humidified chamber in the dark (30 min, 23 °C).
5. Wash off unbound antibody by vertical immersion in three changes of PBS (20 mL, 2 min each).
6. Allow the microarrays to dry in the dark (23 °C).
7. Scan for Alexa Fluor 647 with a Typhoon FLA 9000 scanner with a 633 nm excitation laser and a 670 BP30 emission filter, with resolution set to 50 μm.

3.10 Quantification of EV/Cell Binding on DotScan

For a direct comparison of binding patterns for cells with their EV, analyze binding intensities by ImageQuant (*see* Subheading 2.10). Intensity data are then subjected to background and isotype control subtraction and median centered normalization [61, 62], and duplicate results are averaged.

3.11 Nanoparticle Tracking Analysis (NanoSight):

1. Make dilutions of EV in PBS for the NTA 2.3 software to detect 10^8–10^9 particles/mL via the standard CCD camera of the microscope.
2. Record EV diameters and concentrations (particles/mL) over a period of 60 s at 25 frames/s with the temperature of the laser unit controlled to 22–24 °C.
3. Calculate the yield of purified EV, expressed as a percentage of the number of EV in the original volume of sample from which the EV were purified.

4 Notes

As discussed previously [24, 29, 43, 63], the dot intensities of DotScan data reflect the level of binding of cells (or EV) to the antibody dots and are semiquantitative. The binding patterns of cells (or EV) on DotScan correspond with the proteins on their surfaces at levels above a threshold for capture that may vary with the affinity of each antibody and its accessibility to the relevant antigen. The larger the cell or particle, the more antibody interactions are required for capture. The binding intensities also depend on the number of cells or particles expressing each antigen, until saturation is reached. The sensitivity of the EV assay (i.e., the minimum number of particles required for DotScan detection of a distinctive surface profile for an EV sample) depends on the proportion of EV-expressing antigens recognized by the immobilized antibodies, the level of target antigen detected by the biotinylated detection antibody, and the ECL exposure time. Although fluorescence detection can also be used for EV profiling, a higher sensitivity (>10-fold) is achieved with ECL by extending the exposure time from 30 s to 30 min (not shown).

1. Table 1 lists the antibodies used to prepare DotScan antibody microarrays for this study, showing clones, concentrations, and absence or presence of bovine serum albumin (BSA; 0.1% w/v). Our results have shown that the addition of 0.1% BSA (w/v) to antibody solutions reduces non-specific isotype control binding of EV.
2. The PixSys 3200 Aspirate and Dispense System (BioDot) is excellent for applying 10 nL volumes to Oncyte slides without damage to the soft nitrocellulose layer. However, contact printing [64] can be used with surfaces such as ArrayIt® SuperNitro Microarray Substrates slides, (Arrayit Corporation, Sunnyvale, CA, USA) or glass slides coated with aldehyde silane, poly-L-lysine, or aminosilane [65] and may be more economical of antibody solutions.

3. Medium supplemented with 10% (v/v) EV-depleted FCS was used, rather than serum-free medium, to minimize cellular stress effects on EV protein composition [66]. Overnight ultracentrifugation removes approximately 95% of exosomes from FCS [67]. Any remaining bovine exosomes in EV samples were not detected by DotScan and did not interfere with the assay of human EV. To our knowledge, no cross-reactivity with bovine antigens has been reported for biotinylated detection antibodies used in this study.
4. Vacuette® lithium heparin and Vacuette® K3EDTA blood collection tubes were also used in our study and compared to Vacuette® 9NC sodium citrate tubes. Heparin was found to give sticky EV pellets and less consistent DotScan results than either EDTA or citrate. EV purified from the blood of healthy individuals collected into EDTA or citrate gave comparable normalized DotScan results ($p > 0.05$ by student t-test; $n = 3$) for CD61-depleted EV with CD45 detection (unpublished). Citrate, however, gave significantly higher yields of CD61-depleted EV ($p < 0.05$) than EDTA and is therefore the preferred anticoagulant for future studies, as also recommended by György et al. [68].
5. Blood was collected from patients with progressive CLL [69] and normal donors, with ethics approval from the Human Research Ethics Committee of the University of Sydney, Darlington, NSW.
6. PBS with 5 mM EDTA was used for plasma dilution, regardless of the anticoagulant in the blood collection tube. With the use of citrate collection tubes, dilution of plasma in PBS containing 3.2% (w/v) sodium citrate [70] may be preferred.
7. DotScan DotReader and data analysis software are no longer available from Medsaic Pty Ltd. For enquiries, please contact Prof. Richard Christopherson (richard.christopherson@sydney.edu.au).
8. The human AB serum used at 2% in the EV resuspension buffer (*see* Subheading 2.7, **item 4**) was heated to 56 °C for 30 min to inactivate complement. Although no EV profiles were detectable by DotScan in this buffer, EV-depleted AB serum is recommended for future work.
9. PBMC can be stored by freezing at −80 °C (in heat-inactivated FCS with 10% (v/v) DMSO) before DotScan analysis, provided high viability (>90%) is maintained after thawing and washing. Dead or damaged cells are not captured firmly on DotScan and are lost during the washing steps.
10. Centrifugation (2500 × *g*, 20 min, 4 °C, three times) removes platelets and cell debris, but it also depletes EV of 100–300 nm in diameter, as demonstrated by NanoSight analysis ($p < 0.05$;

unpublished data). To retain EV, this centrifugation step should be replaced by centrifugation twice at 1500 × *g*, (20 min, 23 °C).

11. To avoid EV damage, resuspend the EV pellet very gently by aspiration and scraping with a 200 μL pipette tip with the end cutoff, with intermittent low-speed vortexing. Avoid frothing.
12. EV captured on the CD61 MicroBeads can also be collected by removal of the column from the magnetic field and elution with 2 mL of incomplete RPMI-1640. This EV/microbead suspension can be concentrated to 300 μL using an Amicon stirred cell ultrafiltration method. Do not centrifuge, as this leads to irreversible aggregation of the EV/beads.
13. A cell suspension of 3–5 × 10^6 cells/300 μL is required for a duplicate DotScan antibody microarray on a 20 mm × 28 mm nitrocellulose surface. The moistened nitrocellulose strip holds 300 μL of cell suspension without the need for additional containment such as the application of a hydrophobic pen around the microarray. The volume could be adjusted for larger or smaller arrays.
14. Dots visible above background in DotSan images are considered to be positive. The limit for optical detection is approximately 50–100 cells per antibody dot, depending on size and refractive index of each of the cells. If a DotReader is not available and/or for increased sensitivity, cell binding can be visualized by fluorescence detection (*see* Subheading 3.8), using a fluorescent antibody that recognizes a protein on the cell surface of captured cells. However, fluorescence sometimes shows strong positive results for antibody dots where no bound cells are visible microscopically, suggesting the capture of EV secreted by the cells during their incubation on DotScan. Some proteins that are enriched on EV are not detectable on the cells of origin, presumably due to selective recruitment of these proteins into the EV [43].
15. For DotScan analysis of EV from plasma, 2% heat-inactivated human AB serum is included in the EV resuspension buffer to block isotype control binding [71], as most sMV from blood bind Fc fragments [72]. The choice of appropriate isotype control antibodies is also critical, as different isotype control antibodies can have different non-specific binding with sMV, depending on the origin of the sMV [73].
16. Figure 1 shows a schematic representation of capture and detection of cells and EV on immobilized antibody dots of DotScan. For simplicity, antibodies are shown attached to the nitrocellulose via their Fc portions, but in fact the orientation is random. Antibody microarrays should be prepared in a

humid atmosphere at temperatures ≤18 °C. If allowed to dry for >5 h before blocking with blocking buffer A (*see* Subheading 2.1, **item 5**), they gradually lose their binding activity. Note also that antibodies without 0.1% BSA (w/v) may bind EV non-specifically.

17. Figure 2 shows the steps in the purification of CD61-depleted EV from blood for DotScan analysis. Dotscan profiling of CD61-depleted EV (3×10^9) from the plasma of a healthy donor, using CD45 detection, is shown in Fig. 2a. CD45 is a pan-leukocyte antigen. Although $CD4^+$ T cells are more numerous than $CD8^+$ cells in the blood [74], CD8 was more strongly detected than CD4 on EV from normal plasma samples. EV derived from natural killer (NK) cells or monocytes [75, 76] may contribute to this strong CD8 binding. The monocyte marker, CD15, was also detected, but the NK cell

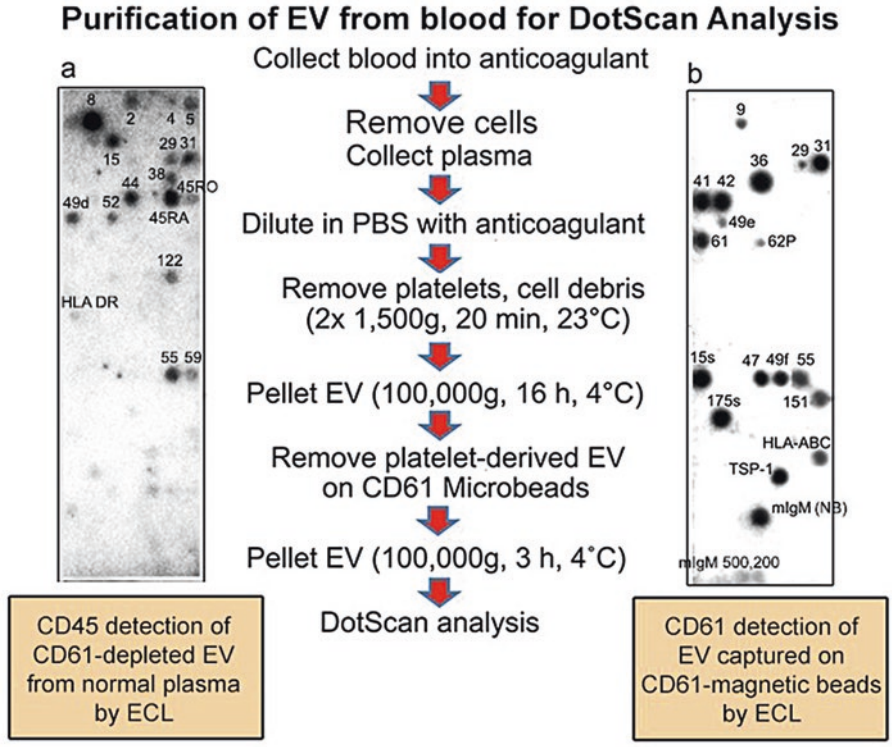


Fig. 2 Purification and profiling of EV from blood: DotScan profiles are shown for: (*a*) CD61-depleted EV (3×10^9) from the plasma of a healthy donor (*see* **Note 17**) and (*b*) platelet-derived EV captured on CD61 MicroBeads (*see* **Notes 18–20**). Numbers near the *dots* indicate CD antigens on EV that were captured by the immobilized antibodies. EV were incubated on DotScan (*a*) with rocking (16 h, 4 °C) and with 2% heat-inactivated AB serum or (*b*) without rocking (1 h at 37 °C) and without 2% heat-inactivated AB serum. Detection was by ECL using biotinylated (*a*) CD45 or (*b*) CD61 antibody. mIgM (NB) is murine IgM isotype control antibody at 500 μg/mL (no BSA); mIgM 500 and mIgM 200 refer to mIgM isotype control antibody at 500 and 200 μg/mL, respectively, with 0.1% (w/v) BSA

marker, CD56, was not. Tetraspanins CD9, CD63, and CD151 were not detected on the surface of CD45$^+$ EV from plasma. In contrast, these tetraspanins were strongly detected on CD61-depleted EV from plasma using biotinylated CD9 antibody (unpublished results). In Fig. 2a, most of the very weak dots are due to non-specific binding that sometimes occurred on antibodies without BSA, even in the presence of 2% AB serum, but were not seen on antibodies with 0.1% BSA.

18. EV captured on Miltenyi magnetic microbeads (50 nm) can be profiled on DotScan directly (Fig. 2b). They were incubated on DotScan for 1 h at 37 °C, without rocking. Longer incubation with rocking has not been tested.
19. In Fig. 2b, EV captured on CD61 MicroBeads showed typical platelet markers (CD9, CD36, CD41, CD42a, CD42b, and CD61) on DotScan; however, they lacked CD63, a marker of platelet activation [77].
20. The reason for the strong non-specific binding of microbead-captured CD61$^+$ EV (Fig. 2b) to murine IgM isotype control antibody without BSA (mIgM NB) is not understood, as Fcμ receptors are not detected on platelets [78], and only one of five other IgM antibodies (CD15s) showed strong binding at 500 μg/mL (without BSA). When the mIgM isotype control antibody contained BSA, the non-specific binding was largely blocked (bottom row, Fig. 2b). The strong binding on CD175s (IgG1) was also unexpected. To our knowledge, CD15s and CD175s are not expressed on platelets. However, these sialic acid-containing proteins are known to bind to the selectin CD62P [79], which is expressed on the platelet-derived EV (Fig. 2b and [80]). IgM antibodies also contain a range of sialylated *N*-linked glycans [81, 82] and are notorious for non-specific binding [83].
21. The incubation of EV on DotScan (16 h at 4 °C) with rocking increases the capture of EV > tenfold, compared to incubation (37 °C, 1 h) without rocking (Fig. 3a, b).
22. Figure 4 shows a comparison of cells and their EV from a human CLL cell line (MEC1) and the blood of a patient with advanced CLL. NanoSight analysis of MEC1-derived EV (Fig. 4a) showed a single EV peak (mode size, 114 nm). NanoSight analysis (Fig. 4b) was also used to compare the size distribution and yield (per mL of plasma) of purified CD61-depleted EV (mode size 65 nm) from the plasma of an advanced CLL patient and total EV in the plasma sample (mode size 72 nm). The yield of CD61-depleted EV was ~10% of the total number of particles in the plasma sample. DotScan profiles are compared for MEC1 cells and their CD19$^+$ EV (Fig. 4d, e) and for patient CLL cells and purified CD19$^+$ EV from the patient's

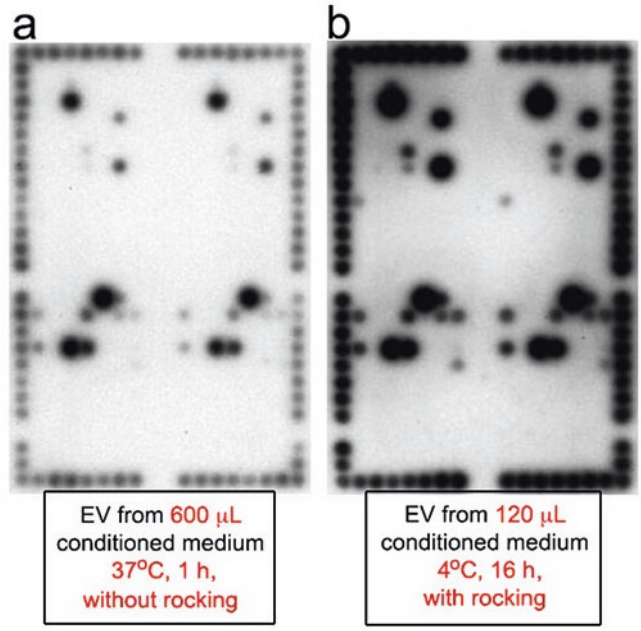
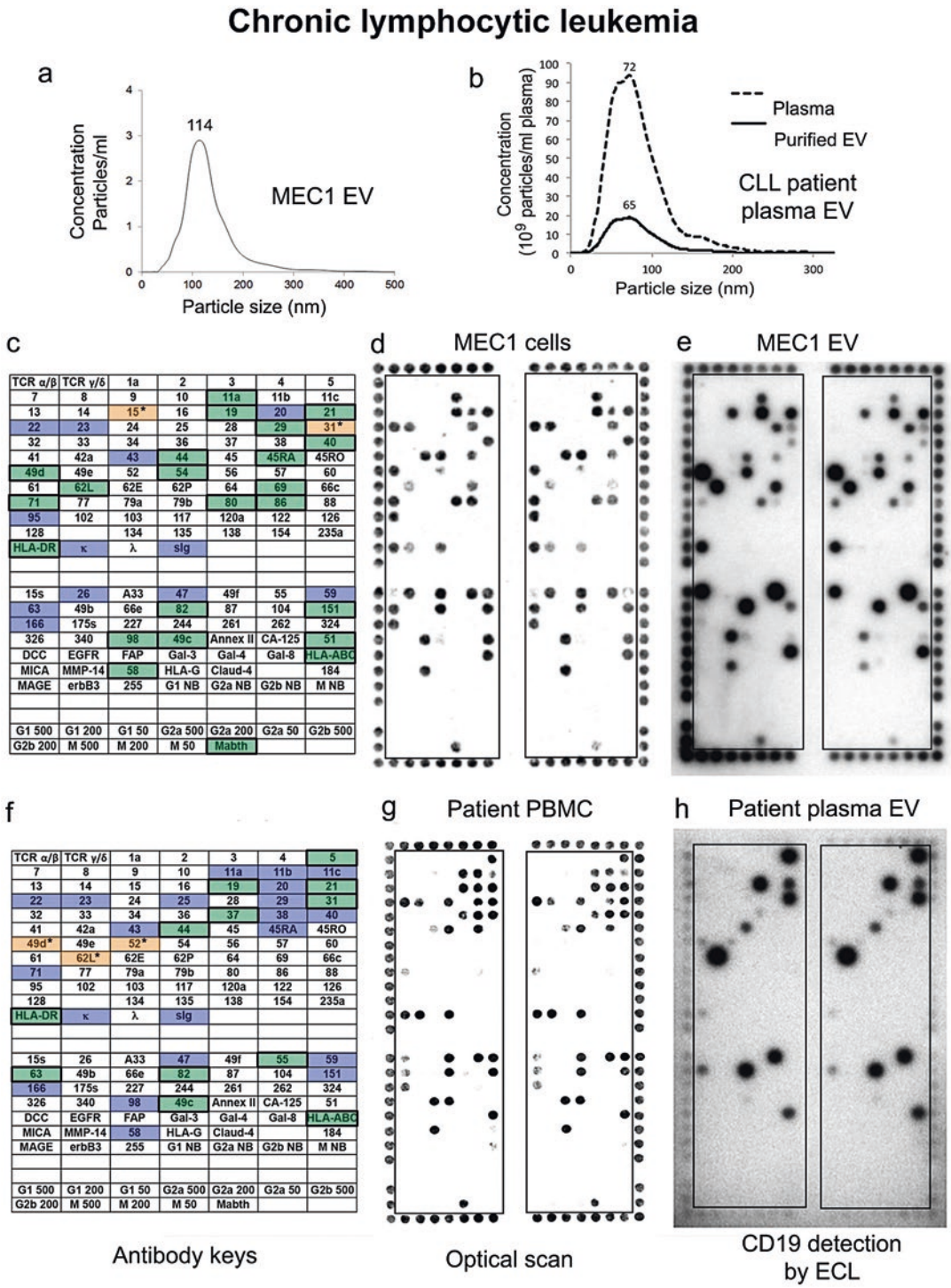


Fig. 3 Variation of incubation time for capture of EV from human SW480 colorectal cancer (CRC) cells. EV derived from the conditioned medium of SW480 cells were incubated on DotScan: (**a**) without rocking (37 °C, 1 h) or (**b**) with rocking (4 °C, 16 h). Five times more EV were used in (**a**) compared with (**b**). Detection was by ECL using biotinylated EpCAM antibody and Streptavidin Poly-HRP. *See* **Note 21** for discussion of results

plasma (Fig. 4g, h). Shading in the antibody keys (Figs. 4c, f) distinguishes antigens that are common to cells and EV (green; thick border) from those detected on cells only (purple) or EV only (orange; with asterisk). Average profiles for CLL-derived EV from the plasma of four CLL patients have been published

Fig. 4 (continued) recorded by optical scanning (DotScan DotReader) for (**d**) MEC1 and (**g**) patient PBMC. (**e**) Captured MEC1 EV (7×10^{10}), and (**h**) CD61-depleted EV from patient plasma ($\sim 5 \times 10^{10}$), were detected by ECL using biotinylated anti-CD19 antibody, with a 5 and 30 min exposure, respectively. Duplicate antibody arrays (outlined) are surrounded by a frame of alignment dots consisting of a mixture of CD44/CD29 antibodies. Antibody keys (**c**, **f**) show locations of antibodies, with *shading* indicating antigens common to cells and their EV (*green*, with *thick border*), on cells only (*purple*) and on EV only (*orange*, with *asterisk*). *See* **Note 22** for discussion of results. Abbreviations for antibodies: Annex II, annexin II; A33, glycoprotein A33, β-Cat, beta-catenin; Claud-4, claudin 4; DCC, deleted in colorectal cancer protein; EGFR, epidermal growth factor receptor; erbB3, erbB3/HER3 protein; FAP, fibroblast activation protein; Gal-3, Gal-4, Gal-8, Galectin-3, Galectin-4, Galectin-8; G1, G2a, G2b, G3, M, murine isotype control antibodies IgG1, IgG2a, IgG2b, IgG3, IgM; 500, 200, 50 refer to concentrations (μg/mL); G1a NB, G2a NB, G2b NB, G3 NB, M NB, murine isotype control antibodies (500 μg/mL) with no BSA; HLA-ABC, HLA-DR, HLA-G, human leukocyte antigens A + B + C, DR, G; κ, λ, immunoglobulin light chains kappa, lambda; Mabthera, chimeric mouse/human anti-CD20; MAGE, melanoma-associated antigen 1; MICA, MHC class I chain-related protein A; MMP-14, matrix metallopeptidase 14; PIGR, polymeric immunoglobulin receptor; slg, surface immunoglobulin; TCR, T-cell receptor; TSP-1, thrombospondin-1; 44v6, CD44 variant exon 6



TCR α/β	TCR γ/δ	1a	2	3	4	5
7	8	9	10	11a	11b	11c
13	14	15*	16	19	20	21
22	23	24	25	28	29	31*
32	33	34	36	37	38	40
41	42a	43	44	45	45RA	45RO
49d	49e	52	54	56	57	60
61	62L	62E	62P	64	69	66c
71	77	79a	79b	80	86	88
95	102	103	117	120a	122	126
128		134	135	138	154	235a
HLA-DR	κ	λ	sIg			
15s	26	A33	47	49f	55	59
63	49b	66e	82	87	104	151
166	175s	227	244	261	262	324
326	340	98	49c	Annex II	CA-125	51
DCC	EGFR	FAP	Gal-3	Gal-4	Gal-8	HLA-ABC
MICA	MMP-14	58	HLA-G	Claud-4		184
MAGE	erbB3	255	G1 NB	G2a NB	G2b NB	M NB
G1 500	G1 200	G1 50	G2a 500	G2a 200	G2a 50	G2b 500
G2b 200	M 500	M 200	M 50	Mabth		

TCR α/β	TCR γ/δ	1a	2	3	4	5
7	8	9	10	11a	11b	11c
13	14	15	16	19	20	21
22	23	24	25	28	29	31
32	33	34	36	37	38	40
41	42a	43	44	45	45RA	45RO
49d*	49e	52*	54	56	57	60
61	62L*	62E	62P	64	69	66c
71	77	79a	79b	80	86	88
95	102	103	117	120a	122	126
128		134	135	138	154	235a
HLA-DR	κ	λ	sIg			
15s	26	A33	47	49f	55	59
63	49b	66e	82	87	104	151
166	175s	227	244	261	262	324
326	340	98	49c	Annex II	CA-125	51
DCC	EGFR	FAP	Gal-3	Gal-4	Gal-8	HLA-ABC
MICA	MMP-14	58	HLA-G	Claud-4		184
MAGE	erbB3	255	G1 NB	G2a NB	G2b NB	M NB
G1 500	G1 200	G1 50	G2a 500	G2a 200	G2a 50	G2b 500
G2b 200	M 500	M 200	M 50	Mabth		

Fig. 4 DotScan comparison of CLL-derived EV with their cells of origin: (**a**) NanoSight analysis of purified MEC1-derived EV and (**b**) NanoSight comparison of the size distribution and yield (per mL of plasma) of purified CD61-depleted EV from the plasma of an advanced CLL patient and total EV in the plasma sample. DotScan profiles for: (**d**) MEC1 cells and (**e**) their EV and (**g**) CLL cells from a patient with a total blood leukocyte count of 45.3×10^9/L and (**h**) purified CD61-depleted $CD19^+$ EV from the patient's plasma. Cell binding profiles were

[43], with the following antigens detected on the EV of at least two of four CLL patients: CD5, CD19, CD21, CD31, CD44, CD55, CD62L, CD82, HLA-ABC, and HLA-DR. Co-expression of CD5 and CD19 is diagnostic of CLL [84]. Biotinylated CD19 antibody detected $CD5^+$/$CD19^+$ EV in CD61-depleted EV from the plasma of advanced CLL patients, but did not detect any $CD19^+$ EV from healthy individuals (not shown). CD9 was not detected on MEC1 cells (Fig. 4d), CLL cells from plasma (Fig. 4g), or their EV (Fig. 4e, h). CD63 was weakly detected on EV from CLL patients but not on MEC1 EV. Tetraspanin CD151 was detected on MEC1 EV but not on patient CLL EV. As discussed previously [43], the detection of very high levels of the homing receptor, CD62L, on CLL EV suggests that this molecule is selectively recruited into the outer membranes of EV and may play an important role in the homing of EV to areas of the body where they may offload their protein and miRNA cargo to promote disease progression and suppress immune responses.

23. DotScan profiling with CD45 detection may be useful for monitoring disease- or treatment-induced changes in the composition of leukocyte-derived EV in plasma or other body fluids. Figure 5 shows a comparison of CD61-depleted $CD45^+$ EV from the blood (10 mL) of healthy donors (Fig. 5b, c) and

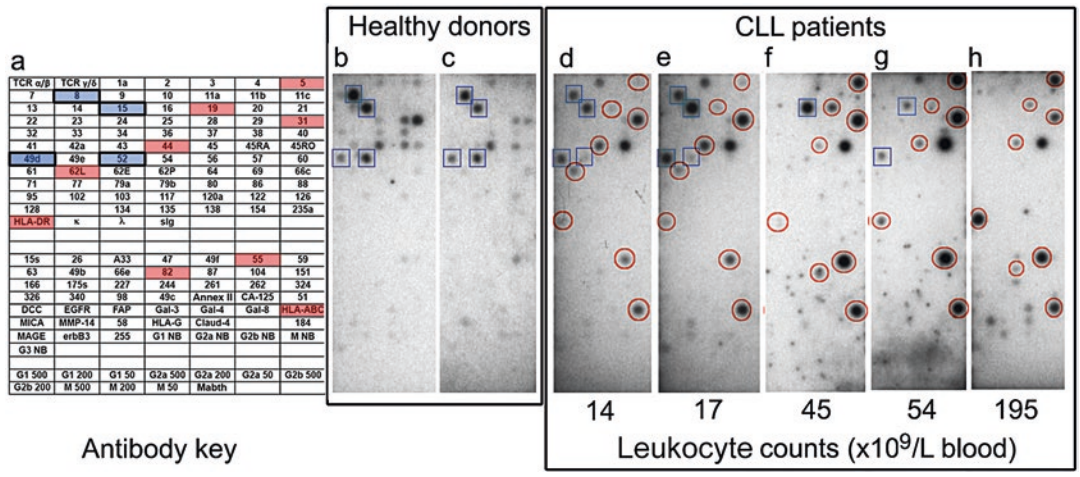
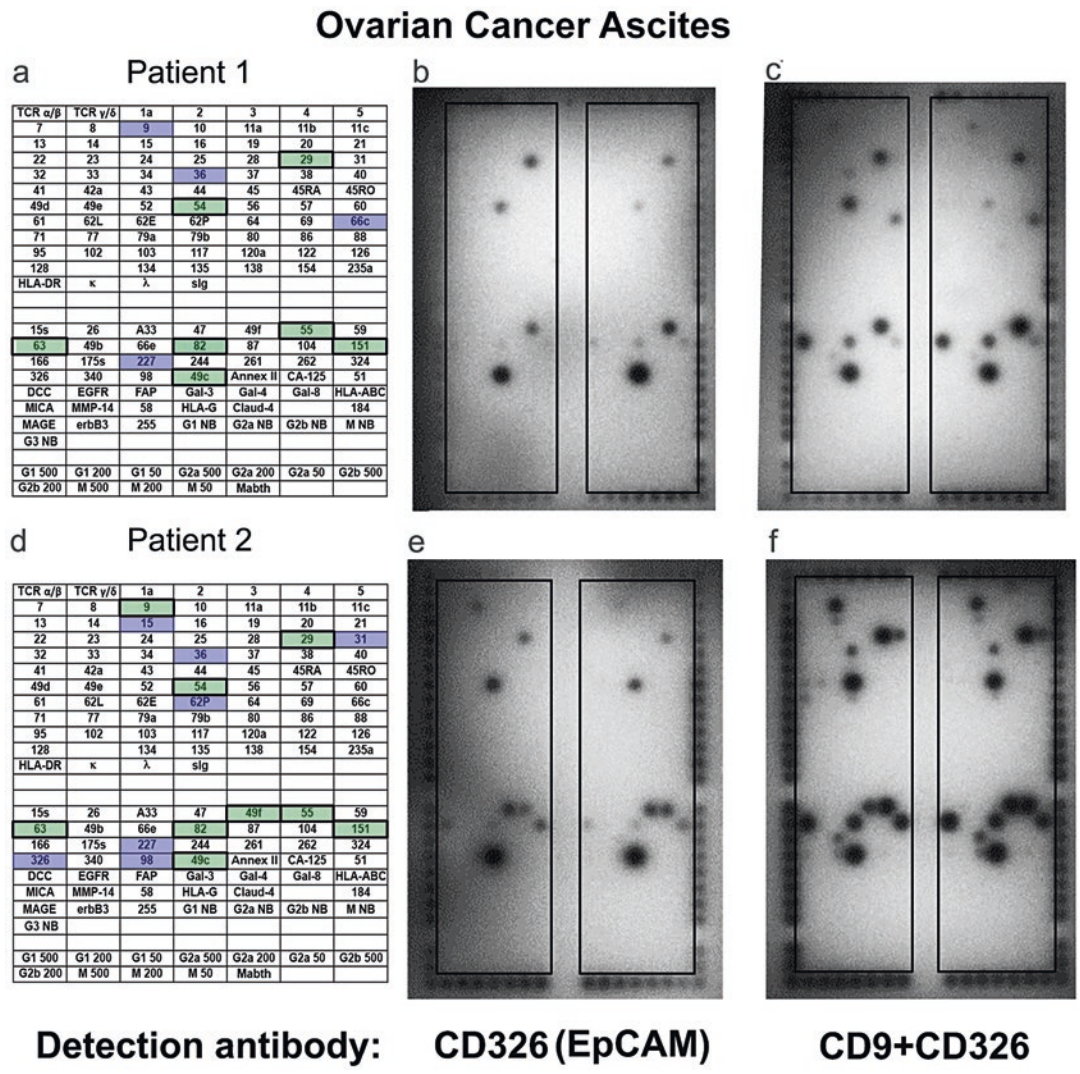


Fig. 5 DotScan comparisons of CD61-depleted EV from the plasma of (*b*, *c*) two healthy donors (average EV/array: 2×10^9) and (*d*–*h*) five CLL patients (average EV/array 6×10^{10}), with ECL detection using biotinylated CD45 antibody. Only half of each duplicate antibody microarray is shown. Blood leukocyte counts are shown for each CLL patient. *Shaded* antibodies (*red*, with *normal border*) in the key (*a*) and the corresponding *dots* (*circled*) in the CLL profiles (*d*–*h*) indicate proteins that were also identified with CD19 detection on CLL-derived EV (Fig. 4h and Belov et al. [43]; *see* **Note 22**). Antibodies *shaded* in *blue* (with *thick border*) in the key (*a*) and the corresponding *dots* (outlined with *squares*) indicate proteins that were strongly expressed on $CD45^+$ EV from healthy individuals (*b*, *c*) but gradually decreased and eventually disappeared in CLL patients with increasing blood leukocyte counts (*d*–*h*). *See* **Note 23** for discussion of results. *See* Fig. 4 for antibody key abbreviations

CLL patients with blood leukocyte counts 14–195 × 10^9/L (Fig. 5d–h). Higher numbers of CD61-depleted EV were isolated from the plasma of CLL patients than from healthy controls (average: ×30); hence EV from CLL patients were profiled on DotScan at a higher density. Several proteins strongly detected on EV from normal plasma (dots outlined with blue squares) were lost in CLL patients with high leukocyte counts (e.g., CD8, CD15, CD49d, CD52; Figs. 2a and 5b, c), probably reflecting the immune-compromised state of advanced CLL patients [85] that results from the impairment of cellular and humoral immunity, with qualitative and quantitative defects in B cells, T cells, NK cells, neutrophils, and the monocyte/macrophage lineage [86]. $CD8^+/CD45^+$ EV were detected at high levels in healthy donors (Figs. 2a and 5b, c) and in CLL patients with relatively low blood leukocyte counts (≤17 × 10^9/L; Fig. 5d, e), but not in CLL patients with ≥45 × 10^9 leukocytes/L (Fig. 5f–h). Levels of $CD15^+/CD45^+$ EV remained high in CLL patients with ≤45 × 10^9 leukocytes/L (Fig. 5d–f), but decreased in a patient with 54 × 10^9 leukocytes/L (Fig. 5g), and were not detectable in a patient with 195 × 10^9 leukocytes/L (Fig. 5h). CD49d was low or undetectable on $CD45^+$ EV from patients with ≥45 × 10^9 leukocytes/L (Fig. 5f–h); CD52 was low (Fig. 5d, e) or undetectable (Fig. 5f–h), while CD45RA was high in all five patients. Failure to detect luminescence on CD45 antibody dots with CD45 detection antibody (Figs. 6b, e) suggests that most CD45 antigenic sites on the EV were bound to the immobilized CD45 antibody and hence unavailable for binding to the detection antibody. Although many of the antigens detected with CD19 antibody (Fig. 4h and Belov et al. [43]) were also detected with CD45 antibody (circled in Fig. 5d–h: CD5, CD19, CD31, CD44, HLA-DR, CD55, and HLA-ABC), there were notable differences. While high levels of the homing receptor, CD62L, were detected on EV from CLL patients using CD19 antibody (Fig. 4h and Belov et al. [43]), CD62L was low (Fig. 5d, e) or undetectable (Fig. 5f–h) with CD45 detection. While CD19 dots were strong with CD19 detection (Fig. 4 and Belov et al. [43]), they were relatively weak with CD45 detection (Fig. 5d–h). These results suggest that CD19 and CD45 antibodies may detect different subsets of CLL-derived EV. This needs further investigation. Interestingly, Pugholm et al. (2016) reported that CD19 was not detected on small EV (<150 nm) from primary cultures of B cells purified from normal human plasma, when tested on their EV Array with CD9, CD63, and CD81 detection [87]. This suggests that the $CD19^+$ EV detected in our Figs. 4 and 5 may be sMV, rather than exosomes. It is also interesting to note that we did not detect CD3 (a T-cell lineage-specific protein) on

EV from the plasma of healthy individuals with CD45 detection (Figs. 2a and 5b, c), despite the detection of T-cell antigens CD2, CD4, CD5, and CD8 in these EV samples. Similarly, $CD3^+$ EV were not detected by Pugholm et al. in the culture supernatants of $CD4^+$ and $CD8^+$ T cells using detection with CD9, CD63, and CD81 antibodies [87].

24. DotScan was used to profile EV from the peritoneal ascites fluid of two patients with advanced ovarian cancer, using biotinylated CD326 antibody (Fig. 6b, e) or a cocktail of biotinylated antibodies (CD9 and CD326) (Fig. 6c, f) for the detection of captured EV. CD326 antibody detected positive results for CD29, CD49c, CD54, CD55, CD63, CD82, and CD151 for both patients, but CD9 and CD49f for Patient 2 only. The CD9/CD326 cocktail of antibodies detected additional antigens (CD9, CD36, and CD227) on the EV of both patients, as well as several differentially expressed antigens (CD15, CD31, CD62P, CD66c, and CD98). Failure to detect luminescence on CD326 antibody dots with biotinylated CD326 antibody (Figs. 6b, e) indicates that most CD326 antigenic sites on the EV were bound to the immobilized CD326 antibody and hence unavailable for binding to the biotinylated CD326 detection antibody. The negative or weak luminescence of CD326 dots with the CD9/CD326 antibody cocktail (Fig. 6c, f), together with the weak CD9 dots with CD326 detection (Fig. 6b, e), may suggest that only a small proportion of EV co-express CD9 and CD326 in these patients.
25. The LIM1863 exosomes (8 μg) were a kind gift from Prof. Richard Simpson (Ludwig Institute for Cancer Research and the Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3050, Australia). They were purified from the conditioned medium of the human LIM1863 CRC cell line by a published method [88], involving the following steps: removal of cell debris (480 × *g*, 5 min; then 1900 × *g*, 10 min), followed by filtration through a VacuCap 60 filter unit fitted with a 0.1-μm Supor membrane (Pall Life Sciences). Exosomes were concentrated using an Amicon Ultracel-5K (5000) molecular weight cutoff centrifugal filter device (Millipore) and centrifugation through a discontinuous iodixanol (OptiPrep) gradient. A comparison of DotScan binding patterns for human LIM1863 CRC exosomes (Fig. 7b) and the tumor cells from a patient with Australian Clinicopathological (ACP; [89, 90]) stage 3 CRC (Fig. 7c) showed many similarities (Fig. 7d). Interestingly, a number of antigens (CD15, CD66c, CD66e, CD104, claudin-4) that were strongly positive when tumor cells from CRC patients were analyzed with Alexa Fluor 647-conjugated EpCAM antibody were only weakly positive, or negative, with optical scanning [29]. This



TCR α/β	TCR γ/δ	1a	2	3	4	5
7	8	9	10	11a	11b	11c
13	14	15	16	19	20	21
22	23	24	25	28	29	31
32	33	34	36	37	38	40
41	42a	43	44	45	45RA	45RO
49d	49e	52	54	56	57	60
61	62L	62E	62P	64	69	66c
71	77	79a	79b	80	86	88
95	102	103	117	120a	122	126
128		134	135	138	154	235a
HLA-DR	κ	λ	sIg			
15s	26	A33	47	49f	55	59
63	49b	66e	82	87	104	151
166	175s	227	244	261	262	324
326	340	98	49c	Annex II	CA-125	51
DCC	EGFR	FAP	Gal-3	Gal-4	Gal-8	HLA-ABC
MICA	MMP-14	58	HLA-G	Claud-4		184
MAGE	erbB3	255	G1 NB	G2a NB	G2b NB	M NB
G3 NB						
G1 500	G1 200	G1 50	G2a 500	G2a 200	G2a 50	G2b 500
G2b 200	M 500	M 200	M 50	Mabth		

TCR α/β	TCR γ/δ	1a	2	3	4	5
7	8	9	10	11a	11b	11c
13	14	15	16	19	20	21
22	23	24	25	28	29	31
32	33	34	36	37	38	40
41	42a	43	44	45	45RA	45RO
49d	49e	52	54	56	57	60
61	62L	62E	62P	64	69	66c
71	77	79a	79b	80	86	88
95	102	103	117	120a	122	126
128		134	135	138	154	235a
HLA-DR	κ	λ	sIg			
15s	26	A33	47	49f	55	59
63	49b	66e	82	87	104	151
166	175s	227	244	261	262	324
326	340	98	49c	Annex II	CA-125	51
DCC	EGFR	FAP	Gal-3	Gal-4	Gal-8	HLA-ABC
MICA	MMP-14	58	HLA-G	Claud-4		184
MAGE	erbB3	255	G1 NB	G2a NB	G2b NB	M NB
G3 NB						
G1 500	G1 200	G1 50	G2a 500	G2a 200	G2a 50	G2b 500
G2b 200	M 500	M 200	M 50	Mabth		

Fig. 6 DotScan analysis of EV from the ascites fluid of two patients with advanced ovarian cancer. Patient 1 (**a**–**c**) is chemonaive; Patient 2 (**d**–**f**) is chemoresistant. The number of EV used per slide was (**b**, **c**) 1.6×10^8 and (**e**, **f**) 2.8×10^8. Captured EV were detected with (**b**, **e**) biotinylated CD326 (EpCAM) antibody or (**c**, **f**) a cocktail of biotinylated CD9 and CD326 antibodies. The location of the antibodies is shown in the keys (**a**, **d**). *Green shading* (with *thick border*) indicates the detection of captured EV by both methods; *purple shading* shows detection by the antibody cocktail (CD9 and CD326) only. *See* **Note 24** for discussion of results. *See* Fig. 4 for antibody key abbreviations

may be due to the fluorescence detection of CRC-derived EV, secreted and captured during the incubation (60 min, 37 °C) of CRC cells on DotScan (*see* **Note 14**).

26. In preliminary experiments (results not shown), CD61-depleted EV from the plasma of advanced CRC patients (n = 6) and healthy donors (n = 4) were profiled by DotScan, using incubation without rocking (1 h 37 °C) for the capture of EV,

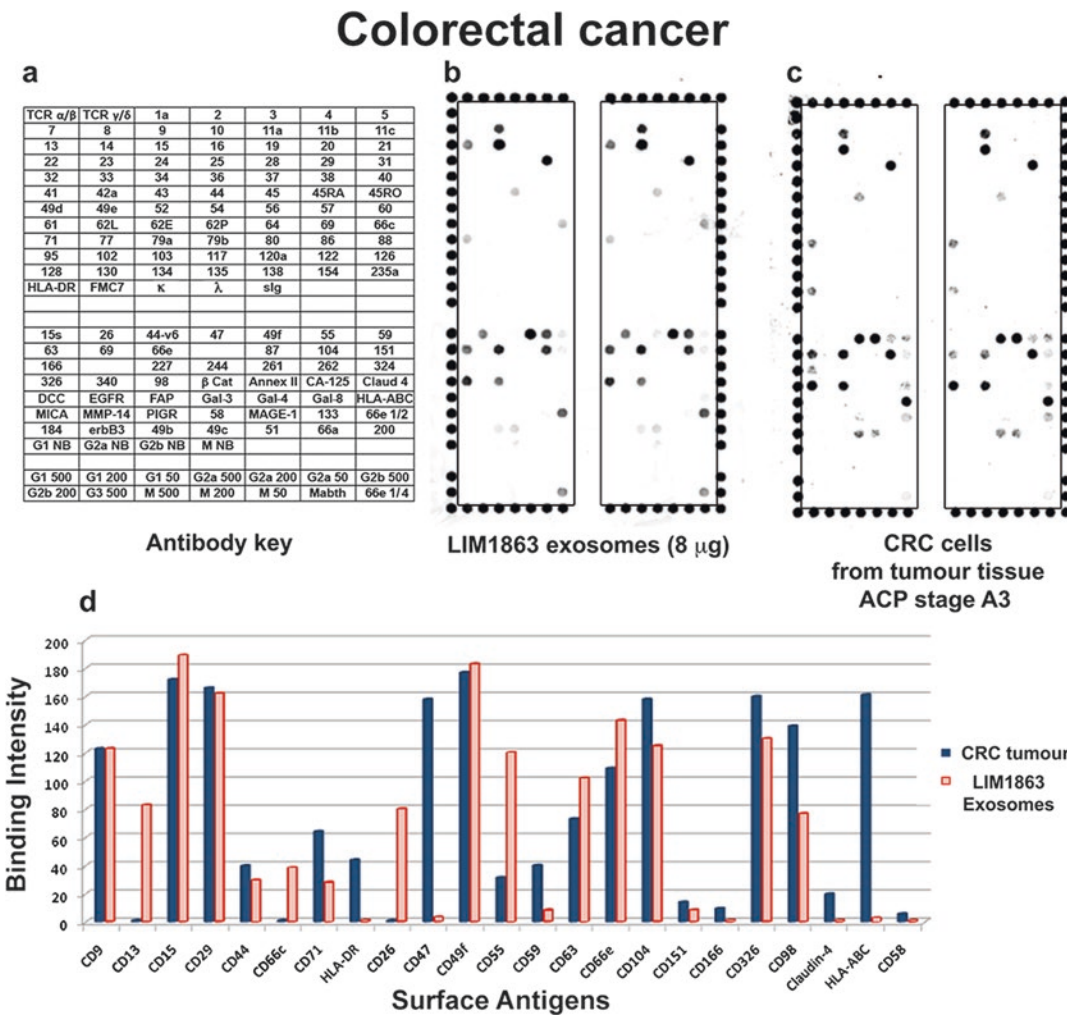


Fig. 7 DotScan comparison of human LIM1863 CRC exosomes and cells from a disaggregated tumor biopsy from a patient with ACP stage 3 CRC (*see* **Notes 25–27**). The key (**a**) shows antibody locations. Duplicate antibody arrays (outlined) are surrounded by a frame of alignment dots consisting of a mixture of CD44/CD29 antibodies. (**b**) LIM1863 exosomes and (**c**) CRC cells from disaggregated CRC tissue from a patient with ACP stage 3 CRC were captured on DotScan (1 h, 37 °C, without rocking) and detected with Alexa Fluor 647-conjugated EpCAM (Subheadings 3.8 and 3.9). (**d**) *Bar* chart compares levels of fluorescence for the indicated CD antigens, normalized against CD9. *See* Fig. 4 for antibody key abbreviations. CD66e antibody was tested at 3 different concentrations, 1000 μg/mL (66e), 500 μg/mL (66e 1/2) and 250 μg/mL (66e 1/4)

and biotinylated CD9 antibody for the detection of captured exosomes. High levels of $CD13^+$ EV were detected in four of six advanced CRC patients, but not in healthy donors. Interestingly, the presence of $CD13^+$ EV corresponded to high blood levels (>260 ng/mL) of carcinoembryonic antigen (CEA; CD66e; clinical pathology reports), an indicator of poor prognosis in CRC [91]. This may be due to a strong inflammatory response in the cancer tissues of these CRC

patients. Enhanced expression of CD13 in vessels of inflammatory and neoplastic tissues has been reported [92], and elevated CEA has been associated with acute and chronic inflammations [93]. DotScan also sometimes detected $CD26^+$, $CD49f^+$, and $CD66c^+$ EV in the plasma of CRC patients but not healthy donors (not shown). This warrants further investigation, employing overnight incubation with rocking at 4 °C (*see* **Note 21**) and additional suggested approaches (*see* **Note 27**) to further increase sensitivity.

27. With increased sensitivity, DotScan analysis of EV in plasma and other noninvasive liquid biopsies may enable early detection of small primary solid tumors, minimal residual disease or relapse, and determination of the effects of drug treatments. DotScan sensitivity could be further enhanced as follows:
    (a) Reduce background noise/luminescence by replacing the translucent Oncyte nitrocellulose microarray slides with transparent ArrayIt® SuperNitro Microarray Substrates slides, (Arrayit Corporation, Sunnyvale, CA, USA) or glass slides coated with aldehyde silane, poly-L-lysine, or aminosilane [65].
    (b) Limit the size of the antibody panel to increase the number of EV per antibody dot.
    (c) Reduce the number of alignment dots.
    (d) To avoid loss of sMV during platelet removal, centrifuge plasma only twice at 1500 × *g* (20 min, 23 °C) instead of three times at 2500 × *g* (20 min, 4 °C).
    (e) Use direct enrichment for disease-specific EV with Miltenyi microbeads instead of depleting platelet-derived EV; this will also avoid inadvertent loss of cancer-derived EV that have bound to, or fused with, platelet-derived EV [94] or are derived from CD61-expressing cancer cells [95].
    (f) Use a cocktail of detection antibodies to profile disease-specific EV, e.g., an antibody cocktail against CD326, CD66e, and A33 antigen may efficiently detect CRC-derived EV in plasma, as these antigens were strongly detected on human LIM1215 CRC EV [43].
    (g) Avoid the use of heparin anticoagulant that makes EV "sticky."

    It is important to note that although transparent slides may provide a more suitable surface for EV analysis than Oncyte nitrocellulose-coated slides (*see* "a" above), the latter are preferred for the profiling of leukocytes, especially CLL cells, as these tend to adhere non-specifically to the other smooth/shiny surfaces (unpublished data).
28. Rituximab is a chimeric murine monoclonal IgG1κ antibody against human CD20, a surface glycoprotein expressed on

normal and malignant mature B lymphocytes. To determine whether DotScan could monitor the effect of rituximab treatment on the secretion and surface protein expression of CLL-derived EV, DotScan was used to compare the profiles (n = 3) of EV secreted by MEC1 cells before and after treatment with 10 or 100 μg/mL rituximab (Fig. 8). This image demonstrates the reproducibility of the results when the method is carried out with care. Dot intensities were increased after rituximab treatment, as a result of increased EV secretion (1.5- and 2-fold after treatment with 10 and 100 μg/mL rituximab, respectively, as determined by NanoSight analysis). The EV binding profiles were similar for all treatments (Fig. 8a–c). However, after quantification and normalization of dot intensities (*see* Subheading 3.10), the averaged results showed that CD15,

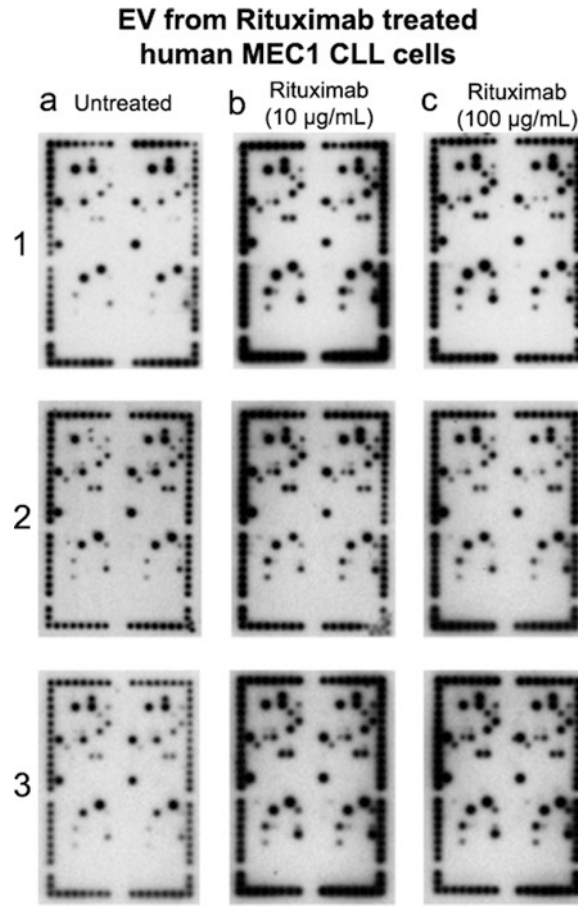


Fig. 8 DotScan binding patterns (in triplicate; 1–3) of EV derived from the conditioned medium of MEC1 cells after 24 h incubation: (**a**) without treatment, (**b**) with 10 μg/mL rituximab, or (**c**) with 100 μg/mL of rituximab (*see* **Note 28**). Captured EV were detected by ECL using biotinylated CD19 antibody. Cell viability was 96% before and after treatment

CD49d, CD54, CD55, and HLA-DR were reduced after rituximab treatment, while CD29 and CD98 were increased ($p < 0.05$ by two-tailed, paired t-test). The downregulation of CD55 (decay-accelerating factor splicing variant 1) by rituximab has also been demonstrated for MEC1 cells in our laboratory [96]. CD55 is an inhibitor of complement-dependent cell lysis, a pro-survival mechanism in CLL [97]. A reduction in CD55 pro-survival signals in CLL cells may increase their susceptibility to rituximab-mediated complement-dependent cell lysis. In contrast, the detected increase in CD98 may support clonal expansion by amplifying integrin signals that enable proliferation and prevent apoptosis [98].

Acknowledgments

We wish to thank Dr. Pauline Huang (School of Life and Environmental Sciences, University of Sydney, Australia) for making the DotScan antibody microarrays and A/Prof. Philip Beale (Concord Repatriation General Hospital, Concord, NSW, Australia) for supplying blood samples from CRC patients.

References

1. D'Souza-Schorey C, Clancy JW (2012) Tumor-derived microvesicles: shedding light on novel microenvironment modulators and prospective cancer biomarkers. Genes Dev 26(12):1287–1299
2. Taylor DD, Gercel-Taylor C (2005) Tumour-derived exosomes and their role in cancer-associated T-cell signalling defects. Br J Cancer 92(2):305–311
3. Cai J, Han Y, Ren H et al (2013) Extracellular vesicle-mediated transfer of donor genomic DNA to recipient cells is a novel mechanism for genetic influence between cells. J Mol Cell Biol 5(4):227–238
4. Cocucci E, Meldolesi J (2015) Ectosomes and exosomes: shedding the confusion between extracellular vesicles. Trends Cell Biol 25(6):364–372
5. Shifrin DA, Beckler MD, Coffey RJ et al (2013) Extracellular vesicles: communication, coercion, and conditioning. Mol Biol Cell 24(9):1253–1259
6. Camussi G, Deregibus MC, Tetta C (2013) Tumor-derived microvesicles and the cancer microenvironment. Curr Mol Med 13(1):58–67
7. Kucharzewska P, Belting M (2013) Emerging roles of extracellular vesicles in the adaptive response of tumour cells to microenvironmental stress. J Extracell Vesicles. doi:10.3402/jev.v2i0.20304
8. An T, Qin S, Xu Y et al (2015) Exosomes serve as tumour markers for personalized diagnostics owing to their important role in cancer metastasis. J Extracell Vesicles. doi:10.3402/jev.v4.27522
9. Tatischeff I (2015) Cell-derived extracellular vesicles open new perspectives for cancer research. Cancer Res Front 1(2):208–224
10. Webber J, Yeung V, Clayton A (2015) Extracellular vesicles as modulators of the cancer microenvironment. Semin Cell Dev Biol 40:27–34
11. Fujita Y, Yoshioka Y, Ochiya T (2016) Extracellular vesicle transfer of cancer pathogenic components. Cancer Sci 107(4):385–390
12. Keerthikumar S, Gangoda L, Liem M et al (2015) Proteogenomic analysis reveals exosomes are more oncogenic than ectosomes. Oncotarget 6(17):15375–15396. doi:10.18632/oncotarget.3801
13. Xu R, Greening DW, Rai A et al (2015) Highly-purified exosomes and shed microvesicles isolated from the human colon cancer cell line LIM1863 by sequential centrifugal ultrafiltration are biochemically and functionally distinct. Methods 87:11–25

14. Green TM, Alpaugh ML, Barsky SH et al (2015) Breast cancer-derived extracellular vesicles: characterization and contribution to the metastatic phenotype. Biomed Res Int. doi:10.1155/2015/634865
15. Santiago-Dieppa DR, Steinberg J, Gonda D et al (2014) Extracellular vesicles as a platform for 'liquid biopsy'in glioblastoma patients. Expert Rev Mol Diagn 14(7):819–825
16. Mizutani K, Terazawa R, Kameyama K et al (2014) Isolation of prostate cancer-related exosomes. Anticancer Res 34(7):3419–3423
17. Royo F, Zuñiga-Garcia P, Torrano V et al (2016) Transcriptomic profiling of urine extracellular vesicles reveals alterations of CDH3 in prostate cancer. Oncotarget 7(6):6835–6846
18. Gámez-Valero A, Lozano-Ramos SI, Bancu I et al (2015) Urinary extracellular vesicles as source of biomarkers in kidney diseases. Front Immunol 6:6. doi:10.3389/fimmu.2015.00006
19. Runz S, Keller S, Rupp C et al (2007) Malignant ascites-derived exosomes of ovarian carcinoma patients contain CD24 and EpCAM. Gynecol Oncol 107(3):563–571
20. Elashoff D, Zhou H, Reiss J et al (2012) Prevalidation of salivary biomarkers for oral cancer detection. Cancer Epidemiol Biomark Prev 21(4):664–672
21. Nicholas J (2013) A new diagnostic tool with the potential to predict tumor metastasis. J Natl Cancer Inst 105(6):371-372
22. Kelleher RJ, Balu-Iyer S, Loyall J et al (2015) Extracellular vesicles present in human ovarian tumor microenvironments induce a phosphatidylserine-dependent arrest in the T-cell signaling cascade. Cancer Immunol Res 3(11):1269–1278
23. Momen-Heravi F, Balaj L, Alian S, et al. 2012 Alternative methods for characterization of extracellular vesicles Front Physiol 3(354. 103389
24. Belov L, Mulligan SP, Barber N et al (1999) Analysis of human leukaemias and lymphomas using extensive immunophenotypes from an antibody microarray. Br J Haematol 135(2):184–197
25. Ellmark P, Belov L, Huang P et al (2006) Multiplex detection of surface molecules on colorectal cancers. Proteomics 6(6):1791–1802
26. Kaufman KL, Belov L, Huang P et al (2010) An extended antibody microarray for surface profiling metastatic melanoma. J Immunol Methods 358(1-2):23–34
27. Zhou J, Belov L, Huang PY et al (2010) Surface antigen profiling of colorectal cancer using antibody microarrays with fluorescence multiplexing. J Immunol Methods 355(1-2):40–51
28. Zhou J, Belov L, Chapuis P et al (2015) Surface profiles of live colorectal cancer cells and tumor infiltrating lymphocytes from surgical samples correspond to prognostic categories. J Immunol Methods 416:59–68
29. Zhou J, Belov L, Solomon MJ et al (2011) Colorectal cancer cell surface protein profiling using an antibody microarray and fluorescence multiplexing. J Vis Exp 55:3322. doi:10.3791/3322
30. Wu JQ, Dyer WB, Chrisp J et al (2008) Longitudinal microarray analysis of cell surface antigens on peripheral blood mononuclear cells from HIV+ individuals on highly active antiretroviral therapy. Retrovirology 5(1):24–35. doi:10.1186/1742-4690-5-24
31. Wu JQ, Wang B, Belov L et al (2007) Antibody microarray analysis of cell surface antigens on CD4+ and CD8+ T cells from HIV+ individuals correlates with disease stages. Retrovirology 4(1):83–95. doi:10.1186/1742-4690-4-83
32. Woolfson A, Stebbing J, Tom BD et al (2005) Conservation of unique cell-surface CD antigen mosaics in HIV-1-infected individuals. Blood 106(3):1003–1007
33. Rahman W, Huang P, Belov L et al (2012) Analysis of human liver disease using a cluster of differentiation (CD) antibody microarray. Liver Int 32(10):1527–1534
34. Rahman W, Tu T, Budzinska M et al (2015) Analysis of post-liver transplant hepatitis C virus recurrence using serial cluster of differentiation antibody microarrays. Transplantation 99(9):e120–e126
35. Huang PY, Kohnke P, Belov L et al (2013) Profiles of surface mosaics on chronic lymphocytic leukemias distinguish stable and progressive subtypes. J Pharm Pharm Sci 16(2):231–237
36. Ellmark P, Högerkorp C-M, Ek S et al (2008) Phenotypic protein profiling of different B cell sub-populations using antibody CD-microarrays. Cancer Lett 265(1):98–106
37. Siegel R, Ma J, Zou Z et al (2014) Cancer statistics, 2014. CA Cancer J Clin 64(1):9–29
38. Ghosh AK, Secreto CR, Knox TR et al (2010) Circulating microvesicles in B-cell chronic lymphocytic leukemia can stimulate marrow stromal cells: implications for disease progression. Blood 115(9):1755–1764
39. Chiorazzi N, Ferrarini M (2003) B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. Annu Rev Immunol 21(1):841–894

40. Seifert M, Sellmann L, Bloehdorn J et al (2012) Cellular origin and pathophysiology of chronic lymphocytic leukemia. J Exp Med 209(12):2183–2198
41. Lee M (2014) Prognostic impact of epithelial cell adhesion molecule in ovarian cancer patients. J Gynecol Oncol 25(4):352–354
42. Ahmed N, Stenvers KL (2013) Getting to know ovarian cancer ascites: opportunities for targeted therapy-based translational research. Front Oncol 3:256. doi:10.3389/fonc.2013.00256
43. Belov L, Matic KJ, Hallal S et al (2016) Extensive surface protein profiles of extracellular vesicles from cancer cells may provide diagnostic signatures from blood samples. J Extracell Vesicles 5:25355. doi:10.3402/jev.v5.25355
44. Kanada M, Bachmann MH, Hardy JW et al (2015) Differential fates of biomolecules delivered to target cells via extracellular vesicles. Proc Natl Acad Sci U S A 112(12):E1433–E1442
45. Lötvall J, Hill AF, Hochberg F et al (2014) Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. J Extracell Vesicles 3:26913. doi:10.3402/jev.v3.26913
46. Crescitelli R, Lässer C, Szabo TG et al (2013) Distinct RNA profiles in subpopulations of extracellular vesicles: apoptotic bodies, microvesicles and exosomes. J Extracell Vesicles 2:20677. doi:10.3402/jev.v2i0
47. Luo X, Fan Y, Park I-W et al (2015) Exosomes are unlikely involved in intercellular nef transfer. PLoS One 10(4):e0124436. doi:10.1371/journal.pone.0124436
48. Oksvold MP, Kullmann A, Forfang L et al (2014) Expression of B-cell surface antigens in subpopulations of exosomes released from B-cell lymphoma cells. Clin Ther 36(6):847–862
49. Saunderson SC, Schuberth PC, Dunn AC et al (2008) Induction of exosome release in primary B cells stimulated via CD40 and the IL-4 receptor. J Immunol 180(12):8146–8152
50. Booth AM, Fang Y, Fallon JK et al (2006) Exosomes and HIV gag bud from endosome-like domains of the T cell plasma membrane. J Cell Biol 172(6):923–935
51. Fang Y, Wu N, Gan X et al (2007) Higher-order oligomerization targets plasma membrane proteins and HIV gag to exosomes. PLoS Biol 5(6):e158. doi:10.1371/journal.pbio.0050158
52. Andreu Z, Yáñez-Mó M (2014) Tetraspanins in extracellular vesicle formation and function. Front Immunol 5:442. doi:10.3389/fimmu.2014.00442
53. Thery C, Amigorena S, Raposo G et al (2006) Isolation and characterization of exosomes from cell culture supernatants and biological fluids. Curr Protoc Cell Biol Chapter 3:Unit 22. doi:10.1002/0471143030.cb0322s30
54. Logozzi M, De Milito A, Lugini L et al (2009) High levels of exosomes expressing CD63 and caveolin-1 in plasma of melanoma patients. PLoS One 4(4):e5219. doi:10.1371/journal.pone.0005219
55. Colino J, Snapper CM (2007) Dendritic cell-derived exosomes express a Streptococcus Pneumoniae capsular polysaccharide type 14 cross-reactive antigen that induces protective immunoglobulin responses against pneumococcal infection in mice. Infect Immun 75(1):220–230
56. Lai RC, Tan SS, Teh BJ et al (2012) Proteolytic potential of the MSC exosome proteome: implications for an exosome-mediated delivery of therapeutic proteasome. Int J Proteomics. doi:10.1155/2012/971907
57. Kalra H, Adda CG, Liem M et al (2013) Comparative proteomics evaluation of plasma exosome isolation techniques and assessment of the stability of exosomes in normal human blood plasma. Proteomics 13(22):3354–3364
58. Ellmark P, Woolfson A, Belov L et al (2008) The applicability of a cluster of differentiation monoclonal antibody microarray to the diagnosis of human disease. Methods Mol Biol 439:199–209. doi:10.1007/978-1-59745-188-8_14
59. Brown A, Lattimore JD, McGrady M et al (2008) Stable and unstable angina: identifying novel markers on circulating leukocytes. Proteomics Clin Appl 2(1):90–98
60. Stacchini A, Aragno M, Vallario A et al (1999) MEC1 and MEC2: two new cell lines derived from B-chronic lymphocytic leukaemia in prolymphocytoid transformation. Leuk Res 23(2):127–136
61. Hamelinck D, Zhou H, Li L et al (2005) Optimized normalization for antibody microarrays and application to serum-protein profiling. Mol Cell Proteomics 4(6):773–784
62. Zhou J, Belov L, Armstrong N et al (2013) Antibody microarrays and multiplexing. In: Wang X (ed) Bioinformatics of human proteomics. Springer, New York, NY, pp 331–359. doi:10.1007/978-94-007-5811-7_15
63. Belov L, de la Vega O, dos Remedios CG et al (2001) Immunophenotyping of leukemias

using a cluster of differentiation antibody microarray. Cancer Res 61(11):4483–4489

64. Austin J, Holway AH (2011) Contact printing of protein microarrays. Methods Mol Biol 785:379–394. doi:10.1007/978-1-61779-286-1_25
65. Seurynck-Servoss SL, White AM, Baird CL et al (2007) Evaluation of surface chemistries for antibody microarrays. Anal Biochem 371(1):105–115
66. Li J, Lee Y, Johansson HJ et al (2015) Serum-free culture alters the quantity and protein composition of neuroblastoma-derived extracellular vesicles. J Extracell Vesicles 4:26883. doi:10.3402/jev.v4.26883
67. Shelke GV, Lässer C, Gho YS et al (2014) Importance of exosome depletion protocols to eliminate functional and RNA-containing extracellular vesicles from fetal bovine serum. J Extracell Vesicles 3:24783. doi:10.3402/jev.v3
68. György B, Pálóczi K, Kovács A et al (2014) Improved circulating microparticle analysis in acid-citrate dextrose (ACD) anticoagulant tube. Thromb Res 133(2):285–292
69. Hallek M, Cheson BD, Catovsky D et al (2008) Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the international workshop on chronic lymphocytic leukemia updating the National Cancer Institute–working group 1996 guidelines. Blood 111(12):5446–5456
70. Adcock DM, Kressin DC, Marlar RA et al (1997) Effect of 3.2% vs 3.8% sodium citrate concentration on routine coagulation testing. Am J Clin Pathol 107(1):105–110
71. Sedlmayr P, Leitner V, Pilz S et al (2001) Species-specific blocking of fc-receptors in indirect immunofluorescence assays. Lab Hematol 7:81–84
72. Gardiner C, Ferreira YJ, Dragovic RA (2013) Extracellular vesicle sizing and enumeration by nanoparticle tracking analysis. J Extracell Vesicles 2:19671. doi:10.3402/jev.v2i0.19671
73. Trummer A, De Rop C, Tiede A et al (2008) Isotype controls in phenotyping and quantification of microparticles: a major source of error and how to evade it. Thromb Res 122(5):691–700
74. Amadori A, Zamarchi R, De Silvestro G et al (1995) Genetic control of the CD4/CD8 T-cell ratio in humans. Nat Med 1(12): 1279–1283
75. Gibbings DJ, Marcet-Palacios M, Sekar Y et al (2007) CD8α is expressed by human monocytes and enhances FcγR-dependent responses. BMC Immunol 8(1):12. doi:10.1186/1471-2172-8-12
76. Campbell JP, Guy K, Cosgrove C et al (2008) Total lymphocyte CD8 expression is not a reliable marker of cytotoxic T-cell populations in human peripheral blood following an acute bout of high-intensity exercise. Brain Behav Immun 22(3):375–380
77. van Velzen JF, Laros-van Gorkom BA, Pop GA (2012) Multicolor flow cytometry for evaluation of platelet surface antigens and activation markers. Thromb Res 130(1):92–98
78. Kubagawa H, Oka S, Kubagawa Y et al (2009) Identity of the elusive IgM fc receptor (FcμR) in humans. J Exp Med 206(12):2779–2793
79. Varki A, Gagneux P (2012) Multifarious roles of sialic acids in immunity. Ann N Y Acad Sci 1253(1):16–36
80. Heijnen HF, Schiel AE, Fijnheer R et al (1999) Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and alpha-granules. Blood 94(11):3791–3799
81. Colucci M, Stöckmann H, Butera A et al (2015) Sialylation of N-linked glycans influences the immunomodulatory effects of IgM on T cells. J Immunol 194(1):151–157
82. Arnold JN, Wormald MR, Suter DM et al (2005) Human serum IgM glycosylation: identification of glycoforms that can bind to mannan-binding lectin. J Biol Chem 280(32):29080–29087
83. Creasey AM, Staalsoe T, Raza A et al (2003) Nonspecific immunoglobulin M binding and chondroitin sulfate a binding are linked phenotypes of plasmodium falciparum isolates implicated in malaria during pregnancy. Infect Immun 71(8):4767–4771
84. Molica S, Mauro FR, Giannarelli D et al (2011) Differentiating chronic lymphocytic leukemia from monoclonal B-lymphocytosis according to clinical outcome: on behalf of the GIMEMA chronic lymphoproliferative diseases working group. Haematologica 96(2):277–283
85. Shanafelt T (2013) Treatment of older patients with chronic lymphocytic leukemia: key questions and current answers. ASH Educ Program Book 2013(1):158–167
86. Dearden C (2008) Disease-specific complications of chronic lymphocytic leukemia. ASH Educ Program Book 2008(1):450–456
87. Pugholm LH, Bæk R, Søndergaard EKL et al (2016) Phenotyping of leukocytes and leukocyte-derived extracellular vesicles. J Immunol Res 2016:6391264. doi:10.1155/2016/6391264

88. Mathivanan S, Lim JW, Tauro BJ et al (2010) Proteomics analysis of A33 immunoaffinity-purified exosomes released from the human colon tumor cell line LIM1215 reveals a tissue-specific protein signature. Mol Cell Proteomics 9(2):197–208
89. Newland RC, Chapuis PH, Pheils MT et al (1981) The relationship of survival to staging and grading of colorectal carcinoma: a prospective study of 503 cases. Cancer 47(6): 1424–1429
90. Davis NC, Evans EB, Cohen JR et al (1984) Staging of colorectal cancer. The Australian clinico-pathological staging (ACPS) system compared with Dukes' system. Dis Colon Rectum 27(11):707–713
91. Tarantino I, Warschkow R, Worni M et al (2012) Elevated preoperative CEA is associated with worse survival in stage I–III rectal cancer patients. Br J Cancer 107(2):266–274
92. Di Matteo P, Arrigoni GL, Alberici L et al (2011) Enhanced expression of CD13 in vessels of inflammatory and neoplastic tissues. J Histochem Cytochem 59(1):47–59
93. Ruibal MA (1991) CEA serum levels in non-neoplastic disease. Int J Biol Markers 7(3):160–166
94. Tesselaar ME, Romijn FP, Van Der Linden IK et al (2007) Microparticle-associated tissue factor activity: a link between cancer and thrombosis? J Thromb Haemost 5(3):520–527
95. Moreno A, Lucena C, Lopez A et al (2002) Immunohistochemical analysis of ß3 integrin (CD61): expression in pig tissues and human tumors. Histol Histopathol 17:347–352
96. Alomari M (2014) Proteomic characterisation of chronic lymphocytic leukaemia cells treated with rituximab. Thesis- http://hdl.handle.net/2123/12407
97. Golay J, Lazzari M, Facchinetti V et al (2001) CD20 levels determine the in vitro susceptibility to rituximab and complement of B-cell chronic lymphocytic leukemia: further regulation by CD55 and CD59. Blood 98(12): 3383–3389
98. Cantor JM, Ginsberg MH (2012) CD98 at the crossroads of adaptive immunity and cancer. J Cell Sci 125(6):1373–1382

Chapter 21

Serum Profiling for Identification of Autoantibody Signatures in Diseases Using Protein Microarrays

Shabarni Gupta, K.P. Manubhai, Shuvolina Mukherjee, and Sanjeeva Srivastava

Abstract

Protein microarrays are platforms for studying protein-protein interactions and identifying disease-related self-antigens/autoantigens, which elicit an immune response in a high-throughput format. Protein arrays have been extensively used over the past two decades for several clinical applications. By using this platform, serum containing autoantibodies against potential self-antigens can be screened on proteome-wide arrays, harboring a large repertoire of full-length human proteins. Identification of such autoantigens can help deducing early diagnostic, as well as, prognostic markers in case of malignancies, autoimmune disorders, and other systemic diseases. Here, we provide an overview of the protein microarray technology along with details of an established method to study autoantibody profiles from patient sera.

Key words Serum profiling, Autoantibody, Protein microarrays, Biomarker discovery, Diagnostics

1 Introduction

Protein microarrays are powerful tools in the interactomics and proteomics arena [1, 2]. In protein arrays, features are printed on glass slides harboring several immobilized proteins, which serve as a platform to study protein interaction with query molecules like proteins, peptides, ligands, or small molecules [1–3]. The inception of this high-throughput technology has its foundations laid in DNA microarray platforms. While DNA microarrays became very popular for its ability to comprehensively perform gene expression profiling, protein microarrays were designed to provide insights on the interactome, the set of molecules that interact with each other in the cell [2, 4]. Since proteins are the "work horses" or the key effector molecules in a cell, a high-throughput platform like protein microarray is indispensable for screening of protein interactors, inhibitors of potential drug molecules, novel biomarkers in certain disease biospecimens, etc. [5–7].

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_21,

Proteins are highly dynamic and are much more labile as compared to DNA. This poses tremendous logistic challenges in fabricating these protein arrays [4]. With the primary goal of increasing the density of proteins printed on the chip, it is also important to maintain its tertiary structure, retain its functionality in its native state, account for any posttranslational modifications that may be present, as well as avoid steric hindrance resulting due to chip printing. Advancements in this field have provided answers to combat many of these hurdles, and the process of technological innovations continues to push the boundaries of conventional protein arrays [4].

Depending on the application or assay to be performed, protein microarrays are primarily divided into three types: (1) functional arrays, where a vast range of query molecules can be probed against printed proteins or peptides on a chip; (2) analytical arrays, where query proteins are probed against antibodies, aptamers, ligands, or affibodies are printed on the chip; and (3) reverse phase protein arrays, where cell lysates are spotted on a chip for detecting expression of a query protein in each lysate using specific antibodies against it [2, 4]. On the basis of contents fabricated on chip, protein arrays can be classified as protein/peptide/analyte-based arrays, cell-free expression (CFE)-based arrays, and reverse phase arrays (RPA) [4]. Protein/peptide/analyte-based arrays involve printing of purified proteins/peptides of analytes like antibodies/affibodies/ligands on a chip for probing against query molecules. CFE-based arrays involve cDNA or plasmid DNA containing cloned gene printing on the chip. The proteins are expressed in vitro on the chip using cell-free expression lysates and are captured using an immobilized ligand/capture antibody on the chip. RPA, as has been described previously, are distinctive in their fabrication as they contain cell of tissue lysates spotted on the chip in a high-throughput manner, rather than a single purified analyte. Thus, protein microarrays are dynamic and complex, and allow a variety of applications, depending on how the chip is designed [6, 8].

An assay involving protein arrays primarily involves four steps: chip printing, assay, scanning, and data analysis (Fig. 1) [4, 9]. The chip printing step is cumbersome and is therefore often circumvented by utilizing the vast number of protein array chips that are commercially available. There are also options for users to customize their own chip depending on the scope of their experiments, commercially. However, researchers with adequate infrastructure opt to print proteins/DNA on their slides and customize their printing layout as per their experimental design. There are several types of assays that can be performed on protein array chips. The principle of the assay resembles a western blotting experiment. The assay described in this chapter involves one of the most sought-after applications using protein arrays, viz., autoantibody screening

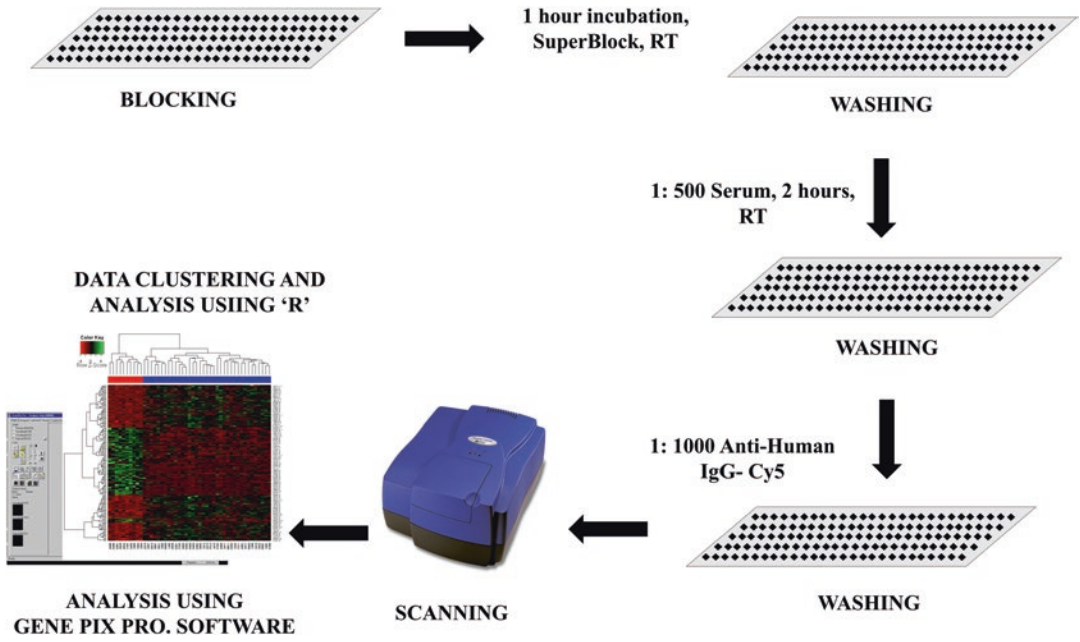


Fig. 1 An overview of protocol for autoantibody screening using high-throughput protein microarrays

from patient sera [6, 10]. Autoantibodies are antibodies against self-proteins [11], which are produced as an aberration of the immune system, where the immune system fails to distinguish between self and nonself proteins and begins attacking "self" cells that result in diseases like rheumatoid arthritis, lupus, celiac disease, etc. [12]. However, production of autoantibodies has also been observed in cancer [10, 12]. It is believed that aberrations in native structure of a mutated protein or overproduction of a given protein, aberrant expression of a protein in a tissue type where it is usually not present, etc. could also trigger an autoimmune response where the body would identify these aberrant expression patterns as foreign [13]. Such proteins are called tumor-associated antigens (TAAs). Immune system being the primary defense system in the body, autoantibodies are produced at the very onset of malignancies when TAAs are produced. This makes autoantibodies a very relevant early diagnostic, minimally invasive marker [10, 13]. Thus, if a protein microarray slide harboring representative proteins from the human proteome is screened with patient sera, autoantibodies from the sera would bind to the printed proteins [6]. These autoantibodies can be detected using a Cy-labeled antihuman IgG [10]. Using a control cohort of healthy individuals, a set of differentially regulated responses can be studied and statistically analyzed as significant and pose as a panel of putative biomarkers which can be validated further using immunochemistry [10]. This involves the downstream steps of scanning and data analysis. Array

design and printing control features become very crucial to distinguish nonspecific signals and eliminate background issues. An assay with this principle has been described in detail in this chapter (Fig. 1) [10].

2 Materials

2.1 Reagents

1. Anti-human IgG antibody (Jackson ImmunoResearch).
2. Anti-GST rabbit antibody (Millipore, Cat. No. AB3282).
3. Anti-rabbit antibody, Alexa Fluor conjugate (Invitrogen, Cat. No. A21429).
4. Blocking buffer: Prepare 10 ml working volume of Super block blocking buffer solution (SuperBlock, blocking buffer, Pierce) in 3% BSA.
5. Primary antibody: 1:500 dilution of serum and 1:5000 dilution of anti-GST in 10 ml 1× TBST containing 2% BSA.
6. Secondary antibody: 1:1000 dilution of antihuman IgG (Cy5) and 1:5000 dilution of anti-rabbit antibody (Cy3) in 10 ml 1× TBST containing 2% BSA.
7. Tween 20.
8. Washing buffer: make 1 Litre lX of TBST containing Tris base, 0.1% Tween 20, KCl, and NaCl.
9. Bovine serum albumin.
10. 10× TBS pH 7.4.
11. Reagents for coating the microarray chip surface for printing, if they are printed in-house (this is customizable and one can use a vast range of coating interfaces, which can be referred to in Gupta et al. [4]). In this chapter, we have described a protocol, where we have used commercially available HuProt™ chips for the experiment.

2.2 Instruments and Software

1. GenePix 4000B microarray scanner (Molecular Devices).
2. Sorvall Legend X1R centrifuge (Thermo Fisher Scientific, Cat. No. 75004260).
3. Rockymax Rocking Shaker (Tarsons, Cat. No. 4080).
4. GenePix Pro software (Molecular Devices).
5. Accessories and Replacement Parts for 20-Slide Glass Staining Dish (Wheaton, Cat. No. 08-812).
6. Robotic arrayer (required if microarrays are printed in-house.

2.3 Materials

1. Fine needle tweezers.
2. Protein microarrays (either printed in-house or commercially available chips can be used. In this chapter, we have described a protocol, where we have used the commercially available HuProt™ chips for the experiment).
3. Serum samples from cancer patients and healthy individuals obtained after informed consent and ethical clearance under an institutional review board.

3 Methods

3.1 Protocol for Printing

Protein microarrays are dynamic and are categorized into several categories such as functional, analytical, and reverse phase arrays [4]. Microarray printing has thereby evolved extensively over time, which has allowed users to achieve such customizability through a software interface. Contact printing and noncontact printing are the two types of printing platforms [4]. Parameters like source and consistency of samples, pin cleaning protocol, and type of plate where sample must be printed are defined by users depending on the nature of the experiment. The quality of printing depends on factors such as quality of the pin used (e.g., noncontact or piezo contact printing allows accurate amount of proteins to be printed each time and provides more reliable quantification than noncontact printing where amount of protein printed cannot be controlled across the samples [14]), humidity of the printing chamber, washing of pins between the sample transfer, and nature of sample (consistency, viscosity, and surface tension [4, 14]) (*see* **Note 1**).

1. The first step in printing protein arrays is defining its surface chemistry. Surface chemistry of the slide depends on the nature of sample to be printed. For example, if a user aims to print purified protein, epoxy-coated slides are a popular choice; however, if users use cell-free expression systems to express protein by printing DNA on the slide, aminosilane coating could be used (*see* **Note 2**). Several other surface chemistry platforms could be used depending on the downstream objective and requirements of the end user [4].
2. Next, the array design must be established where information on the number of pins and position for sample printing and control features is fed into a software interface programming the arrayer (Fig. 2a) (*see* **Note 1**).
3. Dipping time of the pins in the sample must be optimized. This depends on the nature of the samples and the type of pins being used (Fig. 2b).

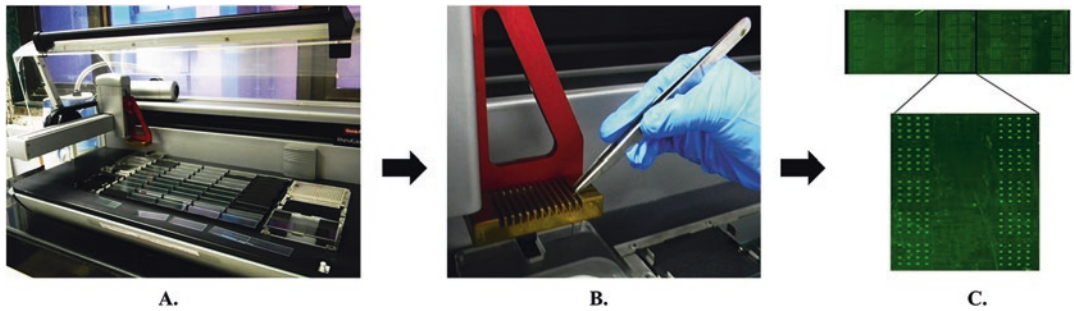


Fig. 2 Steps involved in printing the microarray slides. Panel (**a**) shows a representative arrayer, Microarrayer (OmniGrid Ascent 100, Digilab, Inc.), which can be used for printing features. Panel (**b**) demonstrates the needle insertion step. Panel (**c**) represents a printed slide with distinct features

4. Parameters like duration and number of cleaning incidences of pins during the printing of arrays must be optimized. This helps avoid contamination or carry-over samples between features during printing.
5. Number of blotting incidences of pins to remove the extra samples must be optimized.
6. The .gal file containing information on the positions of samples to be printed is then fed to the system.
7. Depending on the above optimizable parameters (Fig. 3), the array printing is subject to high levels of customization, so as to allow users to achieve desired configuration of arrays (Fig. 2c). The above steps provide an overview of the various factors that must be considered in protein array fabrication. However, our assay is optimized to screen serum autoantibodies on the commercially available HuProt™ chips, which has been described in detail below.

3.2 Assay

1. The microarray experiment is carried out at room temperature.
2. Remove the microarray chips from ultra-cold temperatures (−80 ° C), and allow it to thaw on ice prior to the assay for few minutes in a plastic box (Fig. 4).
3. Add 10 ml of blocking solution (2% BSA in 10 ml SuperBlock) in a plastic box.
4. Then carefully remove the microarray slide using tweezers and place it in the blocking solution by keeping the active surface of the chip faced up and submerged evenly in the solution. While performing assay, it is important that the user wear gloves throughout the experiment. The chips must be strictly handled using tweezers near the bar code-printed area where no protein is printed (*see* **Note 2**).

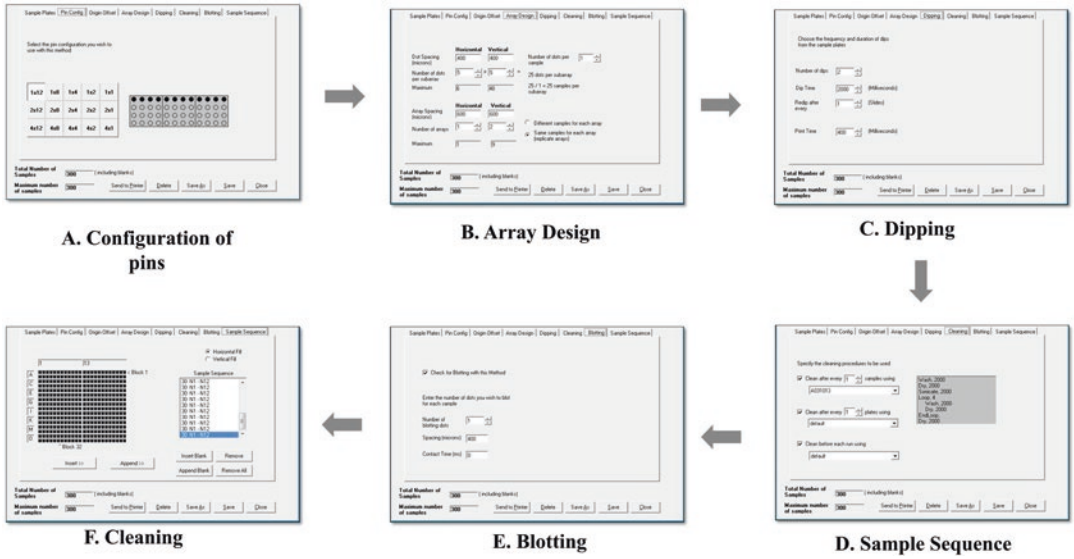


Fig. 3 Software parameters in printing using Microarrayer (OmniGrid Ascent 100, Digilab, Inc.). Panels (**a**–**d**) show the software interface for feeding information on configuration of the pins, array design, dipping, sample sequence, blotting, and cleaning steps, respectively

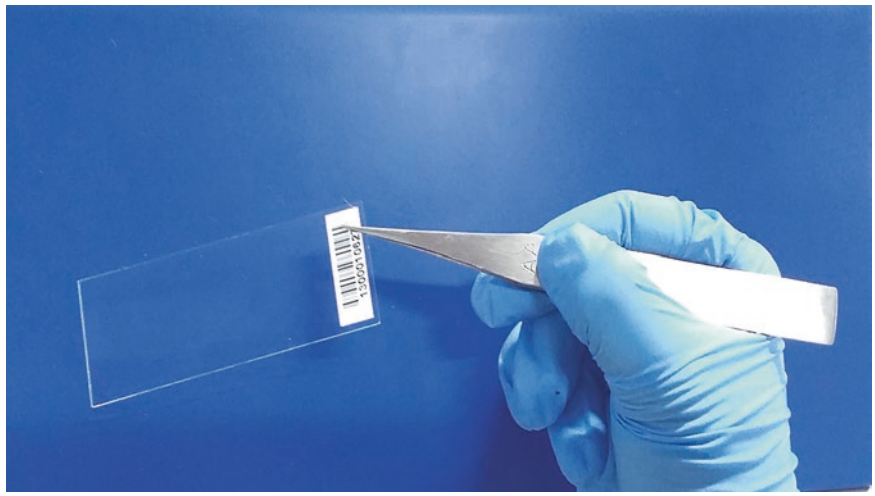


Fig. 4 The protein chips (eg: HuProt chips) is handled using a fine needle tweezer and is allowed to thaw prior to the assay

5. Incubate the chips in blocking solution for 2 h with gentle shaking, at room temperature (*see* **Note 2**).
6. Remove the blocking buffer using a pipette and rinse it with 1× TBST three times.
7. Using fine needle tweezers, remove the slides and keep them in Wheaton glass chamber. Wash it with TBST 4–5 min using a

Fig. 5 The slide is placed in the Wheaton glass chamber and rinsed with 1× TBST using a magnetic stirrer

magnetic bead in the chamber. Be careful to not let the magnetic bead touch the chips; it may lead to scratches (Fig. 5).

8. Next, rinse the slide with distilled water to remove TBST and centrifuge at 900 rpm for 2 min.
9. Immediately, add 10 ml of primary antibody containing diluted serum (*see* **Note 3**) and anti-GST (*see* **Note 1**) antibody on to the chip. Incubate the chips with 2 h of gentle shaking at room temperature (*see* **Note 2**).
10. Remove the primary antibody solution by aspiration using a pipette, and rinse it with 1× TBST three times.
11. Repeat **steps 7** and **8**.
12. Add 10 ml of secondary antibody containing antihuman IgG and anti-rabbit antibody.
13. Incubate the chips in the dark for 2 h of gentle shaking at room temperature.
14. Remove the secondary antibody solution by aspiration using a pipette, and rinse it with 1× TBST, three times.
15. Repeat the **steps 7** and **8**.

3.3 Scanning

Preview scan (40 μm) is used to select the area of the slide, which is to be scanned at a higher resolution for acquiring the image.

3.3.1 Preview Scan

3.3.2 Data Scan

1. The following parameters can be followed where pixel size = 10 μm, laser power = 100%, red channel PMT = 525 nm, and green channel PMT = 450 nm. These settings could change

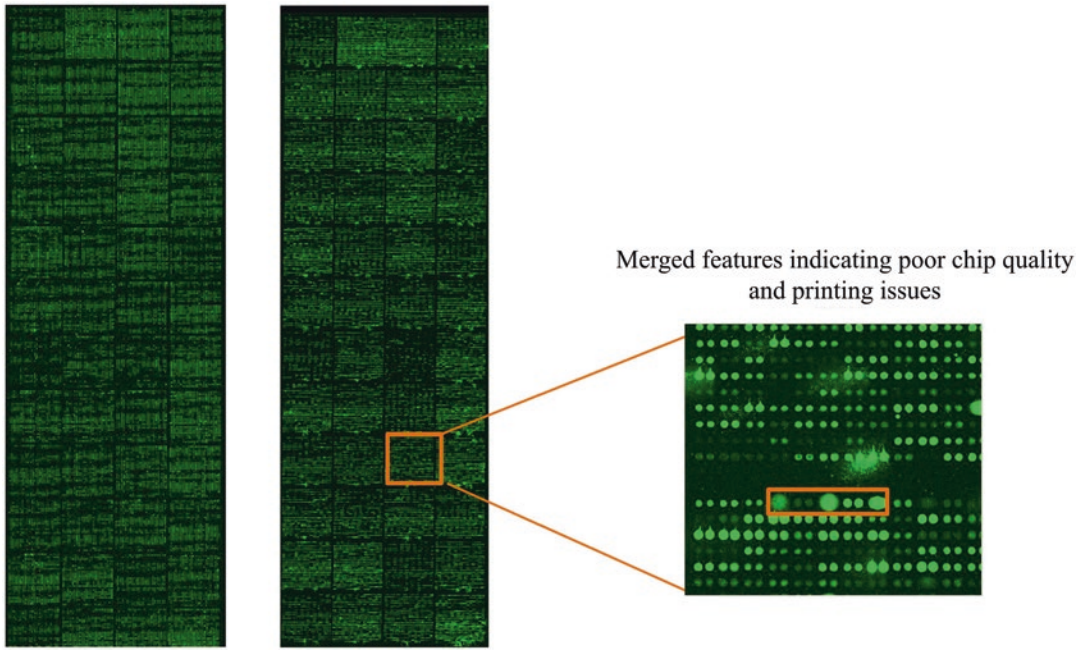


Fig. 6 An overview of processing and data analysis and steps involved in microarray data preprocessing and analysis. Panel (**a**) shows the insertion of the slide into the scanner. Panel (**b**) shows the setting of scanning parameters. Panel (**c**) shows the interface for laying the grid for analysis. Panel (**d**) shows the software data output. Panel (**e**) represents the use of R programming for analysis

depending on the experiment to be performed. For the given protocol, the above settings are optimal (*see* **Note 2**; Fig. 6b).

2. The data scan is used to acquire image for the analysis, and captured image could be saved in .tiff format. The .gal file is used to identify the correct features and their location.
3. Open the .tiff image in GenePix Pro analysis software.
4. Open the .gal file (it appears in the form of a grid) and correctly overlay and align the features according to their correct identity. When all features are arranged according to the .gal file, save the settings as .gps file.
5. The image intensities should be saved in .gpr format. ".gpr" indicates GenePix Results.

3.3.3 Storage

The processed slides can be stored in the dark at −20° C inside a dark airtight box.

3.4 Data Analysis

1. The scanned slides (Fig. 6a) are analysed using GenePix Pro software to create result files (.gpr files). These result files include the cumulative intensities of each spot, converting the pixel intensity values into numerical values and are subject to customizable scanning parameters (Fig. 6b–d).

2. Using statistical parameters, these values are normalized based on intensities from control spots and background noise. Normalization is performed to reduce technical day-to-day variations (*see* **Note 4**) (Fig. 6e).
3. These normalized values are subjected to further statistical analysis, and thresholds for significance can be defined on the basis of *p*-value, fold changes, or a combination of two or more such parameters.
4. Significant proteins can be further subjected to a recursive feature elimination model, which could yield a panel of classifier proteins distinguishing diseased from healthy cohorts.
5. Data analysis processes are also widely dynamic, and the end user must customize the data analysis strategy in consultation with a statistician prior to the study design, which may improve the stringency of an experiment by employing adequate sample size or appropriate statistical tests.

4 Notes

1. *Errors arising from chip printing*. Accurate chip printing is one of the critical determinants of a well-performed protein microarray experiment (Fig. 7a, b). It is often seen that features on commercial chips may be misaligned, which may not exactly

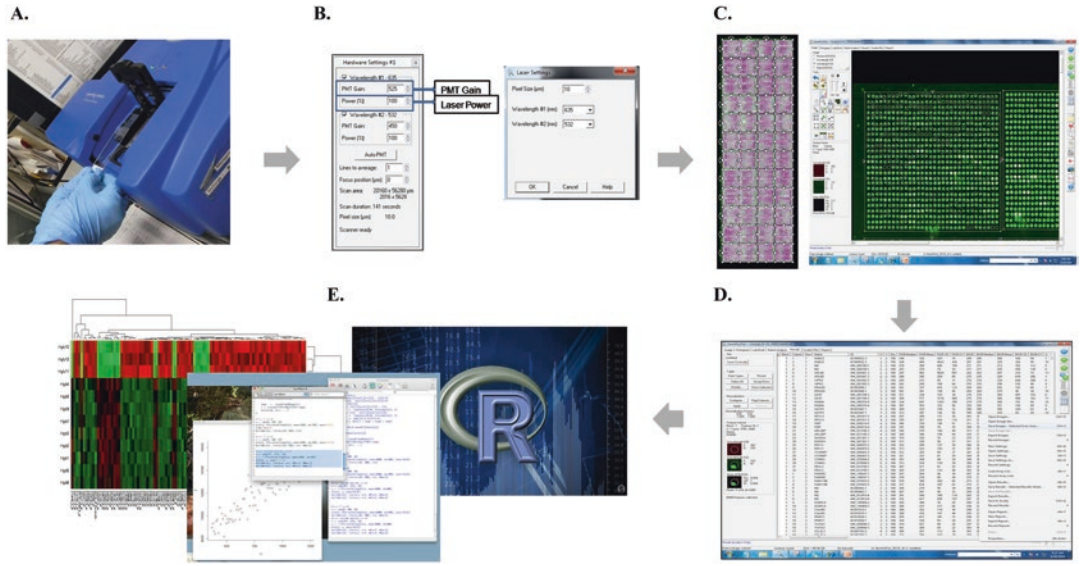


Fig. 7 Illustration of images captured by GenePix Pro 4000B. Panel (**a**) represents an image of a good slide with distinct features. Panel (**b**) is a representative image of a bad slide having merged features. Panel (**c**) is a zoomed-in panel depicting merged spots

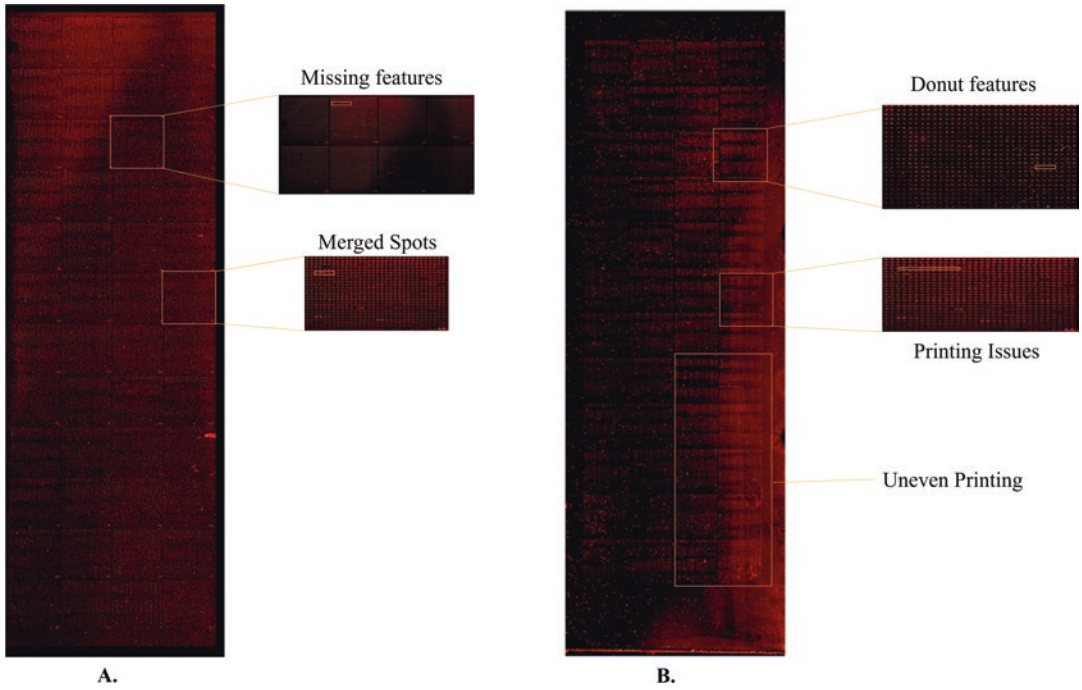


Fig. 8 Illustration of slides with issues that can affect the downstream analysis. Panel (**a**) is a slide showing missing features, merged spots, and donut features. Panel (**b**) is a slide depicting uneven background and printing issues

follow a grid layout. While aligning the .gal file, one can manually align them to extract the true signal intensity from a given spot. On other occasions, users may observe merged spots (Figs. 7c and 8a). This arises when proteins printed on the chip have high fluid content, which may allow merging of two spots on the chip or increased humidity in the array printing robot. At some other instances, users may observe "missing spots" (Fig. 8a). This may arise if the printing pin has not made adequate contact or the pin does not have enough protein to print on that feature. An outer ring on a spot with a hollow interior referred to as the donut structure (Fig. 8a), which also indicates chip printing issue where either the humidity conditions in the arrayer may not have been optimized leading to the drying of spots or a printing pin, may have been worn out. For combatting chip printing issues, a chip/few chips must be spared for Quality Control (QC) experiments. In these experiments, the protein printed on the chips must be checked for their printing profile using the antibody against the tag, e.g., GST-tagged antibody must be checked using an anti-GST antibody, which can be detected using a fluorophore. If there are errors in printing, the chips must not be used on more precious biological samples. While undertaking biological experiments

as well, these QC steps should be performed as has been demonstrated in the protocol described here. If a user is utilizing CFE-based chips, then quality of DNA printed must be checked on a chip from a batch with pico-green staining. If the result from this QC is satisfactory, the expression of protein can be next checked on another chip using a similar anti-GST/tag-based screening.

2. *Combatting background issues.* There are multiple troubleshooting strategies that can be applied to combat this. These may include increasing blocking time, increasing washing time/steps, reducing secondary antibody concentration, etc. Depending on the extent of the noise and nature of localized or overall background (Fig. 8b), one can make a decision on which of the above or all must be utilized. If the noise is localized, increasing blocking time or washing time could help. One must also ensure that the chip is not allowed to stand dry after the blocking step. This could lead to uneven backgrounds. However, if the global profile of the chip shows high background, then one could reduce the secondary antibody concentration. The signal to noise ratio can be minimized by adjusting the PMT so as to allow control spots to reach near saturation. Additionally, one can minimize the concentration of anti-GST antibody in the primary antibody solution, which can result in background noise. The surface chemistry of the chip could also be a factor contributing to global backgrounds. e.g., epoxy-coated slides are known to result in more background as compared to nitrocellulose backgrounds. In such cases, one may try to alter/optimize the blocking buffer composition or try a different surface chemistry for fabricating their chips.
3. *Aggregates on chip.* Aggregates of precipitates on chips indicate a flaw in the buffer preparation where the components may not have dissolved completely or particulate impurities leading to the accumulation of matter (Fig. 8b). Alternatively, while using biospecimens like serum or urine, one must ensure that there are no floccules and the specimen used is clear. Scratches or smudges on the chip can arise if the user has not used tweezers on the edges or has held the chip with bare hands where they may have accidentally touched the printed surface.
4. *Quality check of antibodies being used.* Protein microarrays are a technique that relies immensely on antibodies. A major challenge with antibody-based assays is the amount of variability in terms of efficacy they pose with each batch that is commercially generated. Protein microarrays are extremely sensitive to these, and it is possible that in spite of using an antibody with the same catalogue number, the assay may not work as one batch may not be the same as the one used for optimizing the assay.

This is a technical limitation, and therefore the only way to troubleshoot this is to optimize the assay with every change of antibodies that may be incorporated during the course of an experiment. It is preferable that once an antibody is optimized, one keeps stock of that batch of antibody specifically for the assay which could last the user for an entire set of experiment which may involve cross comparisons to avoid technical variations.

Acknowledgments

This work has been supported from the grant BT/PR4599/BRB/10/1042/2012, 6242-P28/RGCB/PMD/DBT/SNJS/2015 Department of Biotechnology, India. The authors also acknowledge the help of Shaikh Khaja for editing the figures in this chapter.

Conflict of interest
The authors have declared no conflict of interest.

References

1. MacBeath G (2002) Protein microarrays and proteomics. Nat Genet 32:526–532. doi:10.1038/ng1037
2. Reymond Sutandy F, Qian J, Chen C-S, Zhu H (2013) Overview of protein microarrays. Curr Protoc protein Sci 27:Unit 27.1. doi:10.1002/0471140864.ps2701s72
3. Hu S, Xie Z, Qian J et al (2011) Functional protein microarray technology. Wiley Interdiscip Rev Syst Biol Med 3:255–268. doi:10.1002/wsbm.118
4. Gupta S, Manubhai KP, Kulkarni V, Srivastava S (2016) An overview of innovations and industrial solutions in protein microarray technology. Proteomics 16:1297–1308. doi:10.1002/pmic.201500429
5. MacBeath G, Schreiber SL (2000) Printing proteins as microarrays for high-throughput function determination. Science 289:1760–1763
6. Ramachandran N, Srivastava S, LaBaer J (2008) Applications of protein microarrays for biomarker discovery. Proteomics Clin Appl 2:1444–1459. doi:10.1002/prca.200800032
7. Liotta LA, Espina V, Mehta AI et al (2003) Protein microarrays: meeting analytical challenges for clinical applications. Cancer Cell 3:317–325
8. Matson RS (2009) Microarray methods and protocols. CRC Press, New York
9. Syed P, Gupta S, Choudhary S et al (2015) Autoantibody profiling of glioma serum samples to identify biomarkers using human proteome arrays. Sci Rep 5:13895. doi:10.1038/srep13895
10. Owen J, Punt J, Stranford S (2012) Kuby immunology, 7th edn. Macmillan Learning, London
11. Lleo A, Invernizzi P, Gao B et al (2010) Definition of human autoimmunity—autoantibodies versus autoimmune disease. Autoimmun Rev 9:A259–A266. doi:10.1016/j.autrev.2009.12.002
12. Zaenker P, Gray ES, Ziman MR (2016) Autoantibody production in cancer—the humoral immune response toward autologous antigens in cancer patients. Autoimmun Rev 15:477–483. doi:10.1016/j.autrev.2016.01.017
13. Wellhausen R, Seitz H, Wellhausen R, Seitz H (2012) Facing current quantification challenges in protein microarrays, facing current quantification challenges in protein microarrays. Biomed Res Int 2012:e831347. doi:10.1155/2012/831347
14. Kambhampati D (2003) Protein Microarray Technology. Wiley VCH Verlag GmbH & Co. KGaA, Weinheim, FRG. doi:10.1002/3527601554.fmatter

Part V

Developments in Discovery and Targeted Proteomics

Chapter 22

Quantitative Comparisons of Large Numbers of Human Plasma Samples Using TMT10plex Labeling

Pengyuan Liu, Lynn A. Beer, Bonnie Ky, Kurt T. Barnhart, and David W. Speicher

Abstract

One strategy for improving the throughput of human plasma proteomic discovery analysis while maintaining good depth of analysis is to multiplex using isobaric tags. At present, the greatest multiplexing that is commercially available uses the TMT10plex kit. As an example of this approach, we describe efficient shotgun discovery proteomics of large numbers of human plasma to identify potential biomarkers. In the analysis strategy, a common pooled reference was used to enable comparisons across multiple experiments. Duplicate samples showed excellent overall reproducibility across different TMT experiments. Data filters that improved the quality of individual peptide and protein quantitation included using a filter for purity of the targeted precursor ion in the isolation window and using only unique peptides.

Key words Isobaric tag quantitation, TMT10plex, Plasma biomarkers, Proteomics

1 Introduction

Quantitative comparisons of plasma or serum proteomes for discovery of potential clinical biomarkers continue to be of great interest despite substantial challenges in achieving in-depth analysis of samples with adequate throughput. Despite impressive improvements in mass spectrometer performance over the past decade, most plasma discovery strategies require substantial fractionation prior to LC-MS/MS analysis in order to effectively detect low-abundance proteins (<100 ng/mL), which is the concentration range of most clinical biomarkers. The major quantification methods applied in shotgun proteomics can be categorized as label-free, metabolic labeling, and isobaric chemical labeling [1]. In comparison with label-free quantification, stable isotopic labeling approaches make it possible to multiplex samples, that is, to analyze multiple samples in the same LC-MS/MS run to provide direct comparisons. Metabolic labeling methods, such as stable isotope labeling by amino acids in cell culture (SILAC) [2], are very

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_22,

tolerant of variations in any processing steps because samples to be compared can be mixed immediately upon sample collection. A moderate level of multiplexing can be achieved by combining two or three differentially labeled samples, but this also increases the total peptide sample complexity by two- or threefold, respectively, which decreases overall depth of analysis. Regardless, such methods are not feasible for analysis of human plasma. In contrast, chemical labeling methods are sensitive to any variations that may occur prior to and during the labeling step, which is typically performed after protease digestion. The chemical labeling approach that is most widely used is isobaric tags because they have the dual advantage that fractionation and LC-MS/MS analysis can be substantially multiplexed without greatly increasing peptide complexity.

A number of different isobaric labeling reagents have been introduced over the past decade with isobaric tag for relative and absolute quantification (iTRAQ) [3] and tandem mass tag (TMT) [4] being the most popular ones. These two isobaric labeling reagents have very similar molecular structures which consist of an amine-specific reactive group, a mass reporter group for quantification, and a mass normalizer group to link the reactive and reporter groups and balance the total masses prior to fragmentation. The reactive group employed in these reagents is an *N*-hydroxysuccinimide ester which reacts with primary amines, i.e., unblocked *N*-terminals and lysine side chains. The reporter groups are partially fragmented from the peptide during precursor fragmentation in the mass spectrometer. Because each reagent in a multiplex kit has a reporter with a different mass, peptides from different biological samples are readily quantified according to the reporter ion intensities. The mass normalizer group ensures that the peptide complexity in the MS^1 spectra does not increase with multiplexing.

Since the initial introduction of TMT reagents, this labeling strategy has undergone modifications to improve accurate quantification and extent of multiplexing capacity. For example, Dayton et al. [5] expanded the number of quantification channels to make a 6plex version by incorporating different numbers of ^{13}C atoms in the reporter ion group. Specifically, the 6plex TMT reagent produced a series of six different reporter ions with nominal masses from 126 to 131 Da at 1 Da intervals. Subsequently, McAlister et al. [6] and Werner et al. [7] both expanded the reagents to 8plex. In their design, they made very similar reporter ions that differed by 0.0063 Da by replacing one ^{13}C with a ^{15}N on the TMT-127 and TMT-129 Da reporter groups. This took advantage of capacities of current high-end mass spectrometers that have sufficient resolution and mass accuracy to resolve such small mass differences. Viner et al. [8] applied the same ^{15}N replacement strategy to the 128 and 130 Da channels to extend the TMT multiplexing capacity to its current 10plex version and made this a reliable

commercial kit. In addition, Everley et al. [9] extended the reagents to 18plex based on the 6plex version. The 18plex reagents included the original 6plex reagents, 6plex medium TMT reagents by inserting an *N*-ethylformamide group in the mass normalizer group, and 6plex heavy TMT reagents by inserting an *N*-propylformamide group in the mass normalizer group. Another type of 18plex method was also proposed by Everley et al. [9] by combining TMT6plex with triple labeling SILAC. Furthermore, a 54plex TMT method was designed and demonstrated by combining the proposed two types of 18plex methods.

The method described in this protocol uses commercial TMT10plex reagent kits to compare a relatively large number of plasma samples that require multiple sets of 10plex samples with quantitative comparisons across 10plex experiments. As an example, shotgun proteome analysis was conducted of plasma samples from breast cancer patients in efforts to identify potential cardiotoxicity biomarkers caused by therapeutic treatment. Currently, a commonly used and highly effective breast cancer treatment combines doxorubicin and trastuzumab (Herceptin®) [10]; however, the adverse effects of cardiotoxicity become a major issue as up to 18% of patients develop cardiac dysfunction [11]. Since current cardiovascular biomarkers lack sufficient specificity and sensitivity for detection of onset of cardiotoxicity in cancer patients receiving these therapies, it is important to discover better markers both for doctors to make decisions and for researchers to uncover the disease mechanisms.

Our protocol starts with depleting 20 abundant human plasma proteins on an immunoaffinity depletion column. The depleted plasma are then reduced by dithiothreitol, alkylated by iodoacetamide, and in-gel digested by trypsin. The digested peptides are labeled with TMT10plex reagents and combined. In order to increase the depth of analysis, the combined peptides are fractionated by high pH HPLC into 20 fractions. The fractions are analyzed by LC-MS/MS. The raw data files are searched with MaxQuant software.

2 Materials

2.1 FPLC Affinity Depletion of Plasma

1. Human plasma.
2. Microcentrifuge tube with 0.22 μm filter.
3. HPLC or FPLC system capable of operating at low pressure (<30 psi) with automatic sample collector.
4. Equilibration buffer: 1× phosphate-buffered saline (PBS).
5. Elution buffer: 0.1 M glycine and 0.1% (w/v) octyl-β-glucopyranoside (OGP) adjusted to pH 2.5 with HCl.
6. ProteoPrep® 20 Immunodepletion Column (Sigma-Aldrich) .

2.2 Ethanol Precipitation

1. Unbound fraction from ProteoPrep® 20 LC depletion.
2. Ethanol (200 proof, −20 °C).
3. SpeedVac® centrifuge (Thermo Scientific).

2.3 Reduction and Alkylation of Samples Prior to 1D SDS-PAGE

1. Depleted and ethanol-precipitated pellet of human plasma.
2. Protein resuspension buffer: 1% (w/v) SDS buffer solution containing 50 mM Tris–HCl, pH 8.0.
3. 1 M aqueous dithiothreitol.
4. 0.5 M iodoacetamide in 50 mM Tris–HCl, pH 8.0.
5. 37 °C thermostatically controlled incubator/shaker.

2.4 1D SDS-PAGE

1. Reduced and alkylated protein sample.
2. 2× Protein solubilizing buffer: 0.4 M sucrose, 6% (w/v) SDS, 125 mM Tris–HCl, 4 mM Na_2EDTA, 2% (v/v) 2-mercaptoethanol, and 2% (v/v) saturated bromophenol blue solution, pH 8.0.
3. 1-D SDS-PAGE gel (e.g., NuPAGE® Bis-Tris Mini Gels, 1 mm, 10 wells).
4. Running buffer: 3-(*N*-morpholino)propanesulfonic acid (MOPS, 50 mM) SDS.
5. XCell SureLock™ Mini-Cell (Life Technologies).
6. Heat block set to 90 °C.
7. BenchMark™ (Life Technologies) molecular-weight marker.
8. Novex® colloidal blue staining kit (Life Technologies) containing Stainer A and Stainer B.
9. Fixing solution: 50% (v/v) methanol, with 10% (v/v) acetic acid in water.
10. Staining solution: 20% (v/v) methanol, with 20% (v/v) Stainer A in water.
11. Staining trays.

2.5 In-Gel Trypsin Digestion

1. PCR laminar flow hood with a HEPA filter and a light box.
2. 0.1% (v/v) trifluoroacetic acid (TFA), 50% (v/v) methanol in water.
3. 96-Well V-bottomed pierced plate and storage plates with polystyrene plate covers.
4. Gel-cutting device, e.g., MEG-1.5 Gel Cutter (The Gel Company) or stainless steel razor blades.
5. SpeedVac centrifuge equipped with 96-well plate centrifuge rotor.
6. 37 °C thermostatically controlled incubator/shaker.

7. Destain solution: 50% (v/v) acetonitrile in 50 mM aqueous 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 8.5.
8. Sequencing grade-modified trypsin (Promega)
9. Trypsin working solution: 0.02 μg/μL trypsin in 50 mM aqueous HEPES, pH 8.5.
10. Trypsin wash buffer: 50 mM aqueous HEPES.

2.6 Tryptic Peptide Desalting

1. SpeedVac® centrifuge.
2. Spectrafuge™ 16 M microcentrifuge (Labnet).
3. MacroSpin™ column (30–300 μg sample capacity, 50–150 μL elution volume).
4. Conditioning solvent: acetonitrile.
5. Loading buffer: 1% (v/v) aqueous TFA.
6. Equilibration buffer: 5% (v/v) acetonitrile in 1% (v/v) aqueous TFA.
7. Releasing buffer A: 50% (v/v) acetonitrile in 0.1% (v/v) aqueous formic acid.
8. Releasing buffer B: 80% (v/v) acetonitrile in 0.1% (v/v) aqueous formic acid.

2.7 Peptide TMT10plex Labeling

1. TMT10plex™ isobaric label reagent set (available in either 0.8 mg or 0.2 mg aliquots per label).
2. Resuspend buffer: 200 mM aqueous HEPES.
3. Anhydrous acetonitrile.
4. Quench buffer: 5% (w/v) hydroxylamine.

2.8 Labeled Peptide Mixing Ratio Checking

2.9 High pH Reverse Phase HPLC Fractionation

1. 1100 HPLC platform (Agilent).
2. 2.1 × 10 mm Xbridge™ C18 guard column (Waters).
3. 2.1 × 250 mm Xbridge™ BEH300 C18 column (Waters).
4. Buffer A: 10 mM ammonium formate aqueous solution (pH 10).
5. Buffer B: 10 mM ammonium formate in 80% (v/v) acetonitrile (pH 10).
6. 1.5 mL microcentrifuge tube with push cap.

2.10 LC-MS/MS

1. Q Exactive™ Plus mass spectrometer (Thermo Scientific) coupled to a nanoACQUITY UPLC system (Waters) using a nanospray ion source.
2. Symmetry trap column (180 μm i.d. × 2 cm packed with 5 μm C18 resin; Waters).

3. BEH C18 nanocapillary analytical column (75 μm i.d. × 25 cm, 1.7 μm particle size, Waters).
4. Solvent A: 0.1% (v/v) aqueous formic acid.
5. Solvent B: acetonitrile containing 0.1% (v/v) formic acid.

2.11 Data Processing and Analysis

1. MaxQuant software.

3 Methods

3.1 FPLC Affinity Depletion of Plasma or Serum

Plasma depletion using a ProteoPrep® 20 column that removes 20 abundant proteins (~97% of total plasma protein) is described. The ProteoPrep® 20 LC column requires a low-pressure HPLC or FPLC system because its packing has a pressure limit of 30 psi, which is much lower than the minimum operating pressure for most regular HPLC systems (typical 100 psi or higher). More or less plasma proteins can be depleted by using alternative immunoaffinity columns (*see*, for example, Chapter 23 by Beer et al.).

1. Thaw and filter plasma samples through a 0.22 μm microcentrifuge tube. Keep the filtered plasma on ice before injection onto the depletion column.
2. Before connecting the depletion column, flush the system with equilibration buffer for 15 min at 3 mL/min to remove any trapped air.
3. Immediately before usage, take the ProteoPrep® 20 LC column from storage at 2–8 °C to equilibrate at room temperature for 15 min.
4. Connect the column to the system while the system flushing with equilibration buffer at 0.5 mL/min. Avoid introducing gas into the column.
5. It is recommended to run a blank before injection of the first sample of the day.
6. Inject 80–100 μL filtered plasma sample at 0.3 mL/min. The 60 min HPLC gradient is as follows:

   100% equilibration buffer for 30 min at 0.3 mL/min.

   100% elution buffer for 15 min at 3 mL/min.

   100% equilibration buffer for 15 min at 0.3 mL/min.
7. Collect the eluent into pre-cleaned 10 mL polystyrene test tubes using the fraction collector. Switch test tubes every 7.5 min for the first 30 min and then every 2 min thereafter.
8. Unbound proteins are collected in three test tubes from 7.5 to 30 min. Combine the unbound proteins into a pre-cleaned 50 mL centrifuge tube. Keep the proteins on ice.

3.2 Ethanol Precipitation

Ethanol precipitation of proteins is an easy and efficient way to remove salts and detergent. The ethanol used in the precipitation should be of high quality, precooled at −20 °C, and added quickly for efficient precipitation.

1. Quickly add ~9-fold volumes of ethanol (200 proof, −20 °C) relative to unbound fraction volume, and vortex thoroughly.
2. Incubate at 0 °C overnight.
3. Centrifuge for 25 min at 3480 × *g* at 4 °C.
4. Remove the ethanol supernatants carefully and leave the intact pellet. Dry the pellet carefully by blowing a gentle stream of argon across the surface.
5. Store the pellet at −20 °C for future use or immediately precede with next steps.

3.3 Reduction and Alkylation of Samples Prior to 1D SDS-PAGE

1. Thaw ethanol-precipitated depleted plasma pellet.
2. Resuspend pellet in 100 μL resuspension buffer.
3. Reduce the proteins by adding dithiothreitol to a final concentration of 20 mM, and incubate for 1 h at 37 °C with shaking.
4. Alkylate the proteins by adding iodoacetamide to final concentration of 60 mM, and incubate for 1 h at 37 °C in dark with shaking.
5. Quench the alkylation by adding additional dithiothreitol to the sample solutions with the final concentration reaching 50 mM. Incubate for 15 min at 37 °C with shaking.

3.4 1D SDS-PAGE

This protocol uses a short SDS gel to clean up and digest samples. An alternative approach is to ethanol precipitate the alkylated protein and perform a solution trypsin digestion (*see* **Note 1**).

1. Mix the reduced and alkylated plasma proteins with 2× solubilizing buffer.
2. Heat the mixtures at 90 °C for 2 min.
3. Assemble the gel in the XCell SureLock™ Mini-Cell unit and fill the chambers with running buffer. Load samples into each lane with BenchMark™ molecular-weight marker in the first lane.
4. Run gels at constant 200 V. Stop when the dye front has migrated ~0.5 cm.
5. Disassemble the gel unit and transfer the gel to a plastic container.
6. Fix the gel by adding 100 mL fixing solution. Shake gently for 10 min. Discard the fixing solution carefully.
7. Stain the gel by adding 95 mL staining solution. Shake gently for 10 min. Add 5 mL Stainer B to the staining solution. Shake

gently for another 3–12 h. Discard the staining solution carefully.

8. Destain the gel with water till the gel shows a clear background.

3.5 In-Gel Trypsin Digestion

1. Turn on the fan in the PCR hood at least 15 min before doing any experiment to achieve optimal flow of dust-free air.
2. Excise the entire stained area (~0.5 cm) from each gel lane of interest, and cut it into six vertical slices. Transfer the slices into two wells of a pre-cleaned, pierced, 96-well plate (three slices each well to avoid excessive gel volume per reaction).
3. Destain the gel slices by adding 50 μL destaining solution per well (*see* **Note 2**). Incubate for 15 min at 37 °C with shaking. Centrifuge the plates for 1 min to remove the buffer. Repeat the destaining step until the gels appear light blue and white.
4. Dry the gel slices using a SpeedVac® evaporator for at least 30 min.
5. Rehydrate gel slices by adding 45 μL trypsin working solution. Incubate at 37 °C for 16–18 h in a thermostatically controlled incubator and another 15 min at room temperature. Collect the digested protein extract into a clean, 96-well collecting plate by centrifuging for 1 min.
6. Add 25 μL trypsin wash buffer per well to the 96-well pierced plate. Incubate at 37 °C for 30 min and another 15 min at room temperature. Collect the second extract into the same 96-well collecting plate by centrifuging for 1 min.
7. Transfer digested extracts into pre-cleaned 0.5 mL centrifuge tubes and combine two separately digested sample halves.

3.6 Tryptic Peptide Desalting

1. Lyophilize the tryptic peptides using a SpeedVac® evaporator.
2. Resuspend the dried tryptic peptides in 100 μL loading buffer.
3. Condition the MacroSpin™ Column by pipetting 400 μL conditioning solvent into the column and centrifuging it for 1 min at ~110 × *g*. Flush the column by pipetting 400 μL water into the column and centrifuging it for 1 min. Repeat the flush once.
4. Load the resolubilized peptides onto the column. Centrifuge it for 1 min at ~110 × *g*. Pipette 200 μL loading buffer into the column and centrifuge it for 1 min. Repeat once (*see* **Note 3**).
5. Pipette 200 μL equilibration buffer into the column and centrifuge it for 1 min. Repeat once (*see* **Note 3**).
6. Replace the collecting tube with another pre-cleaned tube. Release the peptides by pipetting 100 μL releasing buffer A into the column and centrifuge it for 1 min. Pipette 100 μL

releasing buffer B into the column and centrifuge it for 1 min. Combine the two elutes.

7. Lyophilize the combined elutes using a SpeedVac® evaporator.

3.7 Peptide TMT10plex Labeling

Studies that involve analysis of more than ten samples require the use of a reference sample to compare peptide yields across multiple TMT10plex experimental sets. In the example described here, the reference is a pool of all plasma samples in the study, which is assigned to the same reporter ion channel (e.g., 126 Da) in all the experimental sets. It is recommended that at least some samples be replicated across multiple TMT10plex experimental sets to evaluate reproducibility. In this example, we analyzed 41 different plasma samples that were assigned to six TMT10plex experimental sets with 13 duplicated samples. Each experimental set consists of the reference and nine different plasma samples. Optimal distribution of samples within and across experiments depends upon goals of the study (*see* **Note 4**).

1. Resuspend desalted peptides with resuspension buffer at an estimated concentration of ~1 μg/μL (*see* **Note 5**).
2. Pool 3 μL of peptides from all 41 plasma samples to form a reference. Adjust the reference volume to 200 μL.
3. Equilibrate the TMT label reagents at room temperature for 10–15 min with the lid sealed. Resuspend each 0.8 mg TMT label reagent with 42 μL anhydrous acetonitrile (*see* **Note 6**). Vortex briefly to make sure the reagents are fully dissolved.
4. Label peptides by adding 42 μL TMT label reagent to every 100 μL resuspended peptides. Incubate the reaction for 1 h at room temperature (*see* **Note 7**).
5. Quench the reaction by adding 8 μL quench buffer to the reaction. Incubate for 15 min at room temperature.

3.8 Labeled Peptide Mixing Ratio Checking

After isobaric labeling, similar levels of total peptide per sample should be combined. However, recoveries can be variable when equal volumes of plasma are processed as described above. Typically, protein or peptide assays are used to check yields and to ensure mixing of similar amounts of peptides [12]. An alternative quantification method is to perform a pilot mixing experiment followed by a single LC-MS/MS run (without peptide fractionation) to check the ratios of total reporter ion intensity in each channel as described below.

1. Combine 2 μL of each labeled peptide sample containing different tags to be compared in each experimental set.
2. Desalt the pooled peptides following the same desalting method described in Subheading 3.6.

3. Resuspend the desalted peptides with 40 μL 0.1% (v/v) aqueous FA.
4. Inject 4 μL resuspended peptide sample (estimated ~0.9 μg, *see* **Note 5**) into the LC-MS/MS system, and run a 2 h gradient (*see* Subheading 3.10 for LC-MS/MS method details).
5. Search the resulting LC-MS/MS raw file using MaxQuant (*see* Subheading 3.11 for MaxQuant method details). Sum the total reporter ion intensity per channel, which represents the total amounts of identified peptides in each channel (*see* **Note 8**).
6. Calculate adjusted volumes of labeled peptides by dividing each reporter ion intensity by the average intensity, and use these correction factors for a second pilot experiment. Combine 1–5 μL of each labeled sample to achieve equal total reporter ion intensities. Repeat the LC-MS/MS analysis following Subheading 3.8, **steps 2–5**. If all reporter ion channels show similar total intensities in the second check, apply these adjusted mixing ratios to the bulk samples. If reporter ion channels continue to still show large variations (>±30%), repeat this step (*see* **Note 9**) prior to preparing the bulk pooled multiplexed sample.

3.9 High pH Reverse Phase HPLC Fractionation

In order to obtain in-depth analysis for the plasma proteome, it is necessary to apply two-dimensional HPLC separation. In addition to the low pH reversed-phase LC-MS/MS analysis, digested peptides are usually fractionated by another peptide separation method, such as high pH reversed-phase HPLC, strong cation exchange (SCX) chromatography, electrostatic repulsion hydrophilic interaction chromatography (ERLIC), etc. We chose high pH reversed-phase HPLC as it yields the highest-resolution separation of tryptic peptides [13], is easy to perform, and uses MS-friendly volatile solutions that eliminate the need for an extra desalting step prior to LC-MS/MS analysis.

1. Desalt the bulk pooled sample following the desalting method described in Subheading 3.6 (*see* **Note 10**).
2. Resuspend the peptides with 100 μL buffer A.
3. Set the flow rate at 0.2 mL/min and equilibrate the column with 5% buffer B.
4. Inject and separate the resuspended 100 μL sample using an HPLC gradient is as follows:

   Hold at 5% buffer B for 8 min.

   5–24% buffer B over 7 min.

   24–50% buffer B over 52 min.

   50–55% buffer B over 7 min.

   55–60% buffer B over 5 min.

60–90% buffer B over 1 min.

Hold at 90% buffer B for 15 min.

Return to 5% buffer B over 0.5 min.

Re-equilibrate at 5% buffer B for 10 min.

5. Collect HPLC elutes into 1.5 mL microcentrifuge tubes from 8 to 95 min. Switch collection tube every 1 min. After collection, consolidate the 1 min aliquots into 20 fractions in a checkerboard manner by pooling every 20th fraction, e.g., 1 + 21 + 41 + 61 + 81, 2 + 22 + 42 + 62 + 82, etc.
6. Acidify pooled fractions using FA to a final pH = 3. Lyophilize the fractions using a SpeedVac® evaporator.

3.10 LC-MS/MS

Each pooled fraction from the high pH reversed-phase HPLC separation is analyzed individually with a 150 min LC-MS/MS gradient. In order to obtain good quantification results from the reporter ions, high-resolution mass spectrometers are necessary to clearly resolve the reporter ions with very similar masses. A fast scan speed is also essential in order to achieve a good depth of analysis. The study described here was performed using a Thermo Q Exactive™ Plus mass spectrometer.

1. Set the flow rate at 250 nL/min and equilibrate the column with 5% solvent B.
2. Resuspend the dried fractions with solvent A. Inject 4–8 μL (*see* **Note 5**) of each fraction into the LC-MS/MS system. Trap the loaded fractions with the trapping column for 5 min at isocratic 0% solvent B with 6 μL/min flow rate.
3. The 150 min UPLC gradient is as follows:

   5–28% solvent B over 120 min.

   28–40% solvent B over 5 min.

   40–90% solvent B over 10 min.

   Hold at 90% solvent B for 10 min.

   Return to 5% solvent B over 2 min.

   Re-equilibrate at 5% solvent B for 5 min (*see* **Note 11**).
4. The following parameters are used for MS/MS data acquisition and have been optimized for downstream analysis.
   (a) Nanospray ion source: 2.5 kV spray voltage and 300 °C lens temperature.
   (b) Scan mode: full MS/dd-MS2 (TopN).
   (c) Full scan: 400–2000 m/z range with 70,000 resolution, 3×10^6 automatic gain control (AGC) target, and 50 ms maximum injection time (IT).
   (d) MS/MS scan: top 20 (+1 charge and unassigned ions excluded) selection mode, high-energy collisional dissoci-

ation, 1.2 *m/z* isolation window (*see* **Note 12**), 32 normalized collision energy (NCE), first mass fixed at 115 *m/z* scan range (*see* **Note 13**), 35,000 resolution (*see* **Note 14**), 1×10^6 AGC target, 120 ms maximum IT, 5% underfill ratio, and 30 s dynamic exclusion.

5. Monitor the instrument performance by analyzing standard yeast digests before and after the experiment. For long-term experiments, also perform a yeast digest QC run approximately every 48 h.

3.11 Data Processing and Analysis

A number of different software tools can be used to analyze TMT data. Two programs tested in our lab are MaxQuant [14, 15] and Proteome Discoverer (Thermo Scientific). Proteome Discoverer has a user-friendly visual interface, and the latest version of Proteome Discoverer (v2.1) added a number of functions to enhance the TMT data analysis. MaxQuant has the advantage that it is freely available and frequently updated and can use many processors in parallel for fast data processing. In the example described here, we used MaxQuant for data processing and analysis.

1. Import all raw mass spectrometric data files into the same MaxQuant session. Define each experimental set and fractions and process them together.
2. Select the 10plex TMT reporter ion MS^2 mode as the searching mode. Set reporter mass tolerance small enough to be able to distinguish the closest reporter ions (*see* **Note 14**). Check Filter by PIF and set min. Reporter PIF as 0.75 (*see* **Note 12**).
3. The following parameters are used for the MaxQuant database search:
   (a) Carbamidomethyl group on cysteine as fixed modification.
   (b) Acetyl group on protein N-terminal and oxidation on methionine as variable modification.
   (c) Trypsin/P as digestion mode with the maximum missed cleavages at 2.
   (d) Uniprot human database appended with common expected contaminants including keratins and trypsin.
   (e) False discovery rate (FDR) set to 0.01 for proteins and peptides.
4. After the search is complete, filter out contaminants, reverse hits, and proteins only identified by site (*see* **Note 15**).
5. Based on the assumption that each sample should have equal amounts of total protein, the data is then normalized based on the total reporter ion intensity in each channel of each experimental set to correct for variations in total yield. For each channel, the total reporter ion intensities are divided by the

total number of identified proteins to get average intensities for that channel. To normalize, each protein reporter ion intensity is divided by the average protein intensity of that channel (*see* **Note 16**).

6. Individual protein values are then normalized across experimental sets by dividing each protein reporter ion intensity by the reference reporter intensity in that experimental set (*see* **Note 17**).

4 Notes

1. To perform the in-solution digestion, depleted plasma is resuspended in 120 μL 8 M urea aqueous solution containing 50 mM HEPES, pH 8.5. Add 6.3 μL 1 M DTT aqueous solution and incubate 30 min at 37 °C for reduction. Add 40 μL 0.5 M iodoacetamide aqueous solution containing 50 mM HEPES, and incubate 1 h at 37 °C in dark for alkylation. Add 2 μL 1 M DTT aqueous solution and incubate 15 min at 37 °C to quench the alkylation. Add 800 μL 50 mM HEPES aqueous solution to dilute the urea to ~1 M concentration. Add 1:50 (w/w, E:S) trypsin and incubate overnight to digest the sample. After digestion, the samples are desalted, TMT labeled, combined, fractionated, and LC-MS analyzed using the same procedure as described.
2. In TMT labeling experiments, 50 mM HEPES is used as the digestion buffer rather than the commonly used ammonium bicarbonate because the TMT reagent reacts with primary amines such as ammonium ion. Although a desalting step is performed before the TMT labeling, the buffer is not typically completely removed. HEPES is also preferred over the triethylammonium bicarbonate buffer recommended by the vendor because the residual HEPES causes less interference in the subsequent high pH reverse phase separation than the residual triethylammonium bicarbonate.
3. Collect and save the eluent and wash fractions during loading and desalting of peptides, so that the peptides can be recovered if they are not well retained on the column for any reason.
4. A good experiment design is to randomly assign samples to reporter ion channels and to use different reporter ions for duplicates in different TMT experimental sample sets. Also, if some direct comparisons are more important than other, e.g., a case and matched control, these should be placed in the same experimental set. This is because low-abundance proteins may be inconsistently detected and quantitated across different TMT experimental sets due to the somewhat stochastic detection of low-abundance peptides in very complex

samples. Another consideration is to place similar numbers of cases and controls in each experimental set.

5. An optimal peptide load on 75 μm columns should be about 1–2 μg. For depleted plasma, we estimate the recovery of tryptic peptides as follows: [1] a BCA protein assay is used to quantitate the unbound fraction of a ProteoPrep® 20 LC column (typically approximately 400 μg recovered protein per 80 μL human plasma) [2]; a 50% recovery from the in-gel digest is assumed (this is based on recoveries we typically observe in digestions of standard proteins) [3]; losses during TMT labeling, sample cleanup, and high pH reverse phase separation are assumed to be low and are not corrected; and [4] the distribution of total peptides among fractions from the high pH separation is assumed to be equal.
6. TMT reagents are water-sensitive and have limited stability in solution. Therefore it is important to dissolve the reagents with dehydrated acetonitrile to avoid degrading the reagent before labeling. Because of the water sensitivity, it is also better to use the entire vial of reagent at one time. If the entire vial is not needed, lyophilize the rest of the reagent and store at −20 °C. For smaller-scale TMT experiments (6–25 μg peptides per reaction), one can now purchase the reagents in 0.2 mg quantities.
7. The TMT reagent from a 0.8 mg vial (42 μL of solution in acetonitrile) can label up to 100 μg of total peptides (100 μL at about 1 μg/μL). Because current mass spectrometers are very sensitive and only about 1 μg of multiplexed peptides is injected per LC-MS/MS run, a 0.8 mg vial can be used to label two, three, or four different samples. In the experiments described here, we divided the 42 μL reagent solutions into 3 × 14 μL and labeled three different plasma samples (33 μg each). For the pooled reference, two vials for a total of 82 μL of TMT-126 label reagent were used to label 200 μL pooled reference solution. In future experiments we will label 20–25 μg per sample using the new smaller aliquots of reagent (*see* **Note 5**).
8. When a study involves a large number of samples and multiple TMT10plex experimental sets, the first ratio check and adjustment usually does not fully correct for variable yields within and across experimental sets. Figure 1 shows the first ratio check results for the illustrated experiment using six TMT10plex experimental sets. It clearly shows large reporter ion intensity variations between different reporter ion channels in the same experimental set and the reference channel (yellow bars) varies substantially across experiments. This illustrates the importance of performing a pilot mixing experiment before committing the bulk samples.

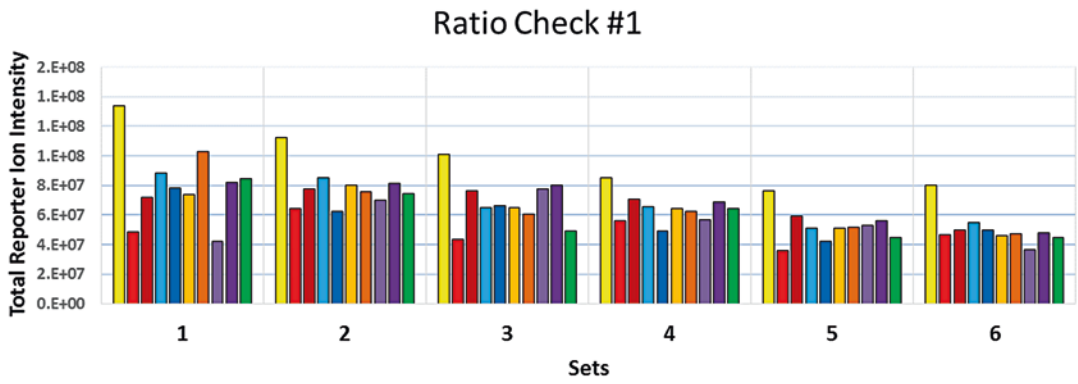


Fig. 1 Results of the first pilot ratio check for a study involving six TMT10plex experimental sets. The total intensity of each reporter ion channel is shown after 2 μL of each differentially tagged sample were combined and analyzed in a single LC-MS/MS run. Reporter ions are color-coded from lowest (*yellow*, 126 Da) to highest mass from *left* to *right*, respectively, for each experimental set

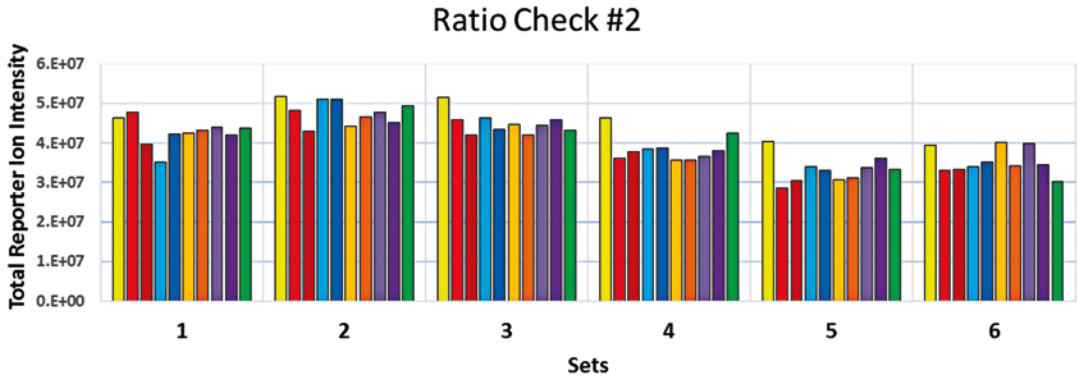


Fig. 2 Results of the second pilot ratio check for a study involving six TMT10plex experimental sets. The total intensity of each reporter ion channel is shown after between 1 and 5 μL of each differentially tagged sample (volumes based on yields observed in the first pilot ratio check as shown in Fig. 1) were combined and analyzed in a single LC-MS/MS run. Reporter ions are color-coded from lowest (*yellow*, 126 Da) to highest mass from left to right, respectively, for each experimental set

9. The second ratio check (Fig. 2) showed similar report intensities across all samples, so no further pilots were performed. Specifically, in each experiment set, the lowest-yielding reporter ion channel and highest-yielding reporter ion channel differ by less than 30%. However, the volumes used to pool the bulk samples were further adjusted based on the relative yields observed in the second ratio check.
10. It is recommended that samples should always be desalted after combining the TMT-tagged samples. Although high pH reversed-phase HPLC should be able to separate the excess TMT reagents and salts/buffers from the labeled peptides, we

have observed interference from reagents and buffers with the peptide separation and the subsequent LC-MS/MS runs.

11. It is recommended that a blank be run between each sample to minimize carryover between samples. We use a rapid blank gradient that takes a total of ~25 min.

12. In large-scale or highly complex shotgun proteomic analysis, a general challenge is that the co-elution of peptides with similar mass is co-isolated and co-fragmented [16]. This type of co-elution is most detrimental for isobaric tag quantitations that rely on reporter ions detected in MS^2 spectra because most peptides in an experiment will be present at similar levels in all samples. Hence, most commonly, a peptide that is differentially abundant in different samples will be co-isolated with a peptide that is similar across all samples. This skews and compresses the calculated ratios on which quantitations are based. In order to reduce the co-isolation interference, Ting et al. recommended that an additional isolation and fragmentation of a selected high-intensity MS^2 fragment ion be performed. This will usually circumvent the co-isolation problem in the MS^3 spectra, and therefore quantitation is based on the ratios of reporter ions in this spectra which efficiently ignores the co-fragmented peaks in the MS^2 [17]. Alternatively, Wenger et al. recommended using proton-transfer ion-ion reactions (PTRs) to reduce the charge states of the isolated ions, which might separate the precursor species from other co-isolated peptides within a fairly large m/z range and provide higher purity isolation and cleaner fragmentation in further MS^3 analysis [18]. Although these reported novel methods usually efficiently reduce the co-isolation interference, they require a mass spectrometer capable of performing high-resolution MS^3 measurements. Many currently available instruments, including the Q Exactive Plus mass spectrometer that was used in our study, cannot perform MS^2 isolation and fragmentation. Hence, in our case, two other approaches were applied to reduce the co-isolation interference. First, a narrow isolation window of 1.2 m/z was set in MS/MS experiments compared with the 2.0 m/z isolation window that we would typically use. It has also been reported that further narrowing the isolation window from 2.0 to 0.5 m/z could further reduce the interference effect [17]; however, the narrow isolation would also result in fewer identifications. Therefore, the 1.2 m/z isolation window is a reasonable compromise. Second, a post-acquisition filter of precursor ion fraction (PIF) of 75% was set, as has been previously recommended [18]. PIF is defined as the fraction of the total ion intensity within the isolation window that is contributed by the targeted precursor ion

with the range from 0 to 1. In MaxQuant search reports, the PIF was determined based on the peak list in the isolation window from the closest full scan of the tandem mass spectrum. It was previously reported that setting the PIF filter at 75% could efficiently improve overall quantification [18].

13. Due to the fact that TMT reporter ions fall in the range from 126 to 131 *m/z*, it is acceptable as long as the MS/MS scan range covers the TMT reporter ion range. Therefore, it is a good strategy to fix the low end of the scan range and make it smaller than *m/z* 126 so the scan range can cover all TMT reporter ions. We chose *m/z* 115 in our experiment. Other setting should be acceptable as well provided that they are less than 126 Da.
14. In TMT10plex experiment, four pairs of reporter ions are very similar in mass, which are TMT-127 N at 127.124760 Da and TMT-127C at 127.131079 Da, TMT-128 N at 128.128114 Da and TMT-128C at 128.134433 Da, TMT-129 N at 129.131468 Da and TMT-129C at 129.137787 Da, and TMT-130 N at 130.134822 Da and TMT-130C at 130.141141 Da. For these four pairs of reporter ions, the mass difference is only the difference between a ^{13}C atom and a ^{15}N atom, which is 0.006319 Da (~50 ppm). Therefore, it is very important for the instrument to resolve the close reporter ion peaks to achieve good quantification. A minimum resolution of at least 30,000 in the MS^2 scan has been reported [6] to be required for adequate separation of these similar reporter ions. Although higher resolution provides better quantification, this takes more transient time when using an orbitrap. More transient time for a single MS/MS acquisition reduces the total number of peptides that can be analyzed for a given LC gradient length. To balance resolution and transient time, a resolution of 35,000 is a relatively optimized parameter, which could adequately separate the similar reporter ion peaks while maintaining a reasonable duty cycle. The similar reporter ions also affect data processing. In the database search setup, the reporter ion mass tolerance has to be set to <0.006319 Da. In our experiment, we set the reporter mass tolerance to be 0.003 Da.
15. The proteins only identified by site are these proteins identified only by a modification site. In general, these protein identifications or lower confidence assignments should be eliminated from quantification.
16. TMT reporter ion intensities quantify the relative intensities of a given protein across the different samples analyzed in a single multiplexed experiment. In our study, we used relative quantifications based on the assumption that every plasma

sample should have the same overall protein amounts, which means the total protein intensities for every TMT label channel should be the same in theory. Although the protein intensities were similar among different channels as a result of our pilot LC-MS/MS runs to check reporter ion yield across samples, they were adjusted to exactly the same level. An alternative strategy which is not shown here is to base relative protein amounts across samples on the same plasma volume. For this approach, if equal volumes of all samples are processed identically and recoveries prior to mixing the TMT-tagged samples are expected to be constant, the pilot mixing experiments and this normalization should not be used. Instead, equal volumes of each tagged sample should be mixed, and only the internal normalization to the common reference should be used (*see* **Note 17**).

17. In order to compare the plasma proteins across different experimental sets, the protein intensities need to be further normalized according to the reference channel that is common to all experiments as it is a single pooled plasma sample that has been tagged in a single modification experiment with the 126 reporter ion. This normalization was performed by dividing every individual protein intensity by the corresponding protein intensity in the TMT reference channel (TMT-126 channel in our study) from the same experiment. By performing scaling and normalization, the protein intensities from different TMT label channels were able to be directly compared across the entire study.

Acknowledgments

This work was supported by NIH Grants RO1HD076279, RO1CA131582, and WW Smith Charitable Trust Grants H1205 and H1305 (D.W. Speicher), PA Department of Health Commonwealth Universal Research Enhancement (CURE) Program Grant (B. Ky) as well as CA10815 (NCI core grant to the Wistar Institute).

References

1. Li Z, Adams RM, Chourey K, Hurst GB, Hettich RL, Pan C (2012) Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. J Proteome Res 11(3):1582–1590
2. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1(5):376–386
3. Ross PL, Huang YN, Marchese JN, 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, Pappin DJ (2004) Multiplexed protein quantitation in

Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3(12):1154–1169

4. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 75(8):1895–1904
5. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, Burkhard PR, Sanchez J-C (2008) Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem 80(8): 2921–2931
6. McAlister GC, Huttlin EL, Haas W, Ting L, Jedrychowski MP, Rogers JC, Kuhn K, Pike I, Grothe RA, Blethrow JD, Gygi SP (2012) Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. Anal Chem 84(17):7469–7478
7. Werner T, Becher I, Sweetman G, Doce C, Savitski MM, Bantscheff M (2012) High-resolution enabled TMT 8-plexing. Anal Chem 84(16):7188–7194
8. Viner R, Bomgarden R, Blank M, Rogers J (2013) Increasing the multiplexing of protein quantitation from 6- to 10-plex with reporter ion isotopologues. PN_AMAS_W617_RViner_R1
9. Everley RA, Kunz RC, McAllister FE, Gygi SP (2013) Increasing throughput in targeted proteomics assays: 54-plex quantitation in a single mass spectrometry run. Anal Chem 85(11): 5340–5346
10. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 344(11):783–792
11. Telli ML, Hunt SA, Carlson RW, Guardino AE (2007) Trastuzumab-related cardiotoxicity: calling into question the concept of reversibility. J Clin Oncol 25(23):3525–3533
12. Villen J, Gygi SP (2008) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc 3(10):1630–1638
13. Cao Z, Tang H-Y, Wang H, Liu Q, Speicher DW (2012) Systematic comparison of fractionation methods for in-depth analysis of plasma proteomes. J Proteome Res 11(6):3090–3100
14. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13(9):2513–2526
15. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372
16. Michalski A, Cox J, Mann M (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC–MS/MS. J Proteome Res 10(4):1785–1793
17. Ting L, Rad R, Gygi SP, Haas W (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat Methods 8(11):937–940
18. Wenger CD, Lee MV, Hebert AS, McAlister GC, Phanstiel DH, Westphall MS, Coon JJ (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. Nat Methods 8(11):933–935

Chapter 23

Efficient Quantitative Comparisons of Plasma Proteomes Using Label-Free Analysis with MaxQuant

Lynn A. Beer, Pengyuan Liu, Bonnie Ky, Kurt T. Barnhart, and David W. Speicher

Abstract

Mass spectrometry (MS)-based quantitation of plasma proteomes is challenging due to the extremely wide dynamic range and molecular heterogeneity of plasma samples. However, recent advances in technology, MS instrumentation, and bioinformatics have enabled in-depth quantitative analyses of very complex proteomes, including plasma. Specifically, recent improvements in both label-based and label-free quantitation strategies have allowed highly accurate quantitative comparisons of expansive proteome datasets. Here we present a method for in-depth label-free analysis of human plasma samples using MaxQuant.

Key words Label-free quantitation, Plasma biomarkers, Proteomics, MaxQuant

1 Introduction

MS-based proteomics is an important tool for biomarker discovery using patient plasma samples. However, in-depth analysis of human plasma proteomes is challenging due to a wide dynamic range of protein concentrations, substantial patient-to-patient variability, and the fact that most specific biomarkers are present at very low concentrations [1, 2]. Hence, extensive depletion of high-abundance proteins and fractionation of large numbers of samples is generally needed to effectively identify low-abundance biomarkers. The need for extensive fractionation greatly restricts sample throughput and makes accurate and reproducible quantitation across many samples more challenging [3, 4]. Nonetheless, interest in plasma proteomics has strengthened due to recent advances in both MS instrumentation and MS-based quantitation strategies which have greatly increased the depth of analysis and reliability of quantitative comparisons across plasma proteomes [5, 6].

There are several general strategies currently used for MS-based quantitation of plasma proteomes. Two popular label-based strategies

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_23,

involve chemical modification of proteins and peptides with isobaric tags for relative and absolute quantification (iTRAQ) or tandem mass tags (TMT) [7–9]. Both of these technologies target primary amines and rely on measurement of reporter ion intensities detected at the MS^2 level after fragmentation [10] and allow multiplexing of up to eight or ten samples to increase throughput. However, these commercial labeling reagents are relatively expensive, require complex and careful sample preparation, and incomplete labeling of the proteome can be observed, which can reduce depth of analysis somewhat. Also, isobaric labeling requires homogeneous precursor ion selection in the full scan (MS^1) mode. Ideally this process is highly selective when narrow precursor isolation windows are used; however, in practice a common occurrence is that unrelated ions are also isolated within the specified m/z window and are therefore co-fragmented with the targeted precursor ion. This co-isolation can result in inaccurate quantitation and underestimation of changes in the ratios of protein levels across samples. This can be especially problematic in very complex samples such as plasma [6, 11].

Label-free quantitation (LFQ) of plasma proteomes is a promising alternative approach to isobaric tags based on the observation that ion peak intensity in ESI-MS is generally proportional to the concentration of a peptide in a sample [12, 13]. By comparing ion intensities between LC and MS runs of multiple samples, comprehensive quantitation of each peptide between samples is achievable [13, 14]. However, one limitation of the label-free method is that in order to achieve accurate, relative quantitation of peak intensities in multiple LC-MS datasets, it is necessary to have high-resolution and high-mass-accuracy mass spectrometers (e.g., LTQ Orbitrap or Q Exactive Series, Thermo Fisher Scientific, Waltham, MA). In addition, specialized data analysis software packages are necessary for precise data extraction, alignment of corresponding signals across runs, and processing to achieve accurate quantitation. Nonetheless, label-free quantitation has become an economical and attractive alternative to stable isotope labeling, especially when large numbers of samples are to be analyzed, because label-free is not limited by predefined numbers of isotope labels [15].

Recently, there have been several in-depth comparisons of isobaric labeling with label-free quantitation, with varying results [6, 16–18]. One analysis shows overall good agreement between the two methods and particularly superior quantification with a label-free approach when two or more peptides are required for protein identifications [18], while another study found the label-free method to be less accurate than TMT labeling [17]. However, both studies agree that label-free quantitation gives superior results in terms of protein coverage and increased protein identifications [17, 18]. A third group observed equally linear quantitation down to 1 fmol for iTRAQ, TMT, and label-free methods. While protein identifications

were increased in the label-based methods, they also determined that quantitative accuracy for TMT-labeled samples was affected by precursor mixing [6]. Finally, in another direct comparison of label-free, iTRAQ, and TMT labeling, the label-free method provided the best proteome coverage for identification, but reproducible quantitation was worse for label-free than the label-based methods [16]. However, it should be noted that in this study, MS/MS spectral counts were used for label-free quantitations rather than MS peak intensities, and spectral counts are generally considered to be less accurate than summed MS intensities for quantitation, especially for low-abundance proteins [15, 19].

When we compared IgY14- and Supermix-depleted plasma using either 10-plex TMT labeling or label-free quantitation using a single 4 h run per proteome, approximately 850 proteins were detected by label-free, and approximately 690 proteins were identified by TMT (two or more peptides, protein and peptide false discovery rate of 1%). However, we found that the TMT label had two limitations. First, when multiple TMT experiments were used to compare more than ten samples, the total number of proteins that could be compared across multiple TMT experiments dropped markedly due primarily to stochastic detection of TMT reporter ions for low-abundance peptides. For example, in one large-scale proteome analysis experiment of 36 plasma samples that we recently conducted using TMT labeling and fractionation of the multiplexed proteomes into 20 fractions, approximately 1200 total proteins were detected, but only about 600 proteins were consistently detected across all samples, despite the presence of the same reference sample in all experiments (data not shown). If the approximately 600 proteins that are inconsistently detected are discarded, we will discard most low-abundance proteins, which are the most likely biomarkers. Another limitation of this TMT data was that for low-abundance peptides, noise in the reporter ion region and frequent co-isolation of abundant unchanged peptides in the isolation window of targeted peptides resulted in ratio suppression for peptide and protein fold changes, as has been previously reported [20–22]. The detrimental consequence of this effect on identifying low-abundance biomarkers in the single TMT10-plex experiment mentioned above is illustrated for ADAM12 quantitation (Fig. 1), a previously identified plasma biomarker that was shown to be high in normal intrauterine pregnancy (IUP) and very low to undetectable in ectopic pregnancy (EP) [23]. This expected difference is very obvious in the case of the label-free comparison but much less apparent for the TMT data. Figure 2 compares the coefficients of variation (CVs) of protein identifications of the same plasma samples analyzed by TMT (Fig. 2a) and label-free quantitation (Fig. 2b). In this comparison, ~162 more proteins are identified by the label-free method. This reduced depth of analysis in TMT-labeled samples is most likely due to partial modification of some peptides. While overall TMT modification was estimated to

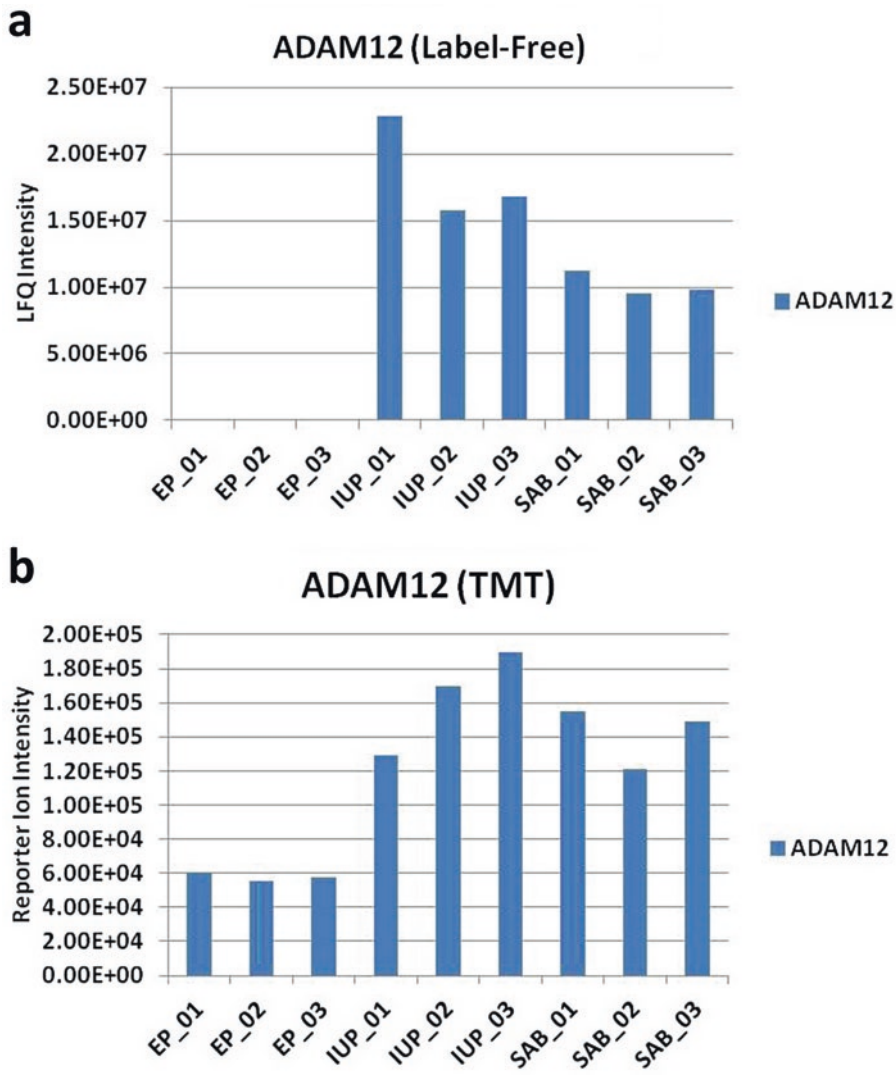


Fig. 1 Comparison of label-free and TMT protein quantitation for ADAM12, a known ectopic pregnancy biomarker. (**a**) Label-free quantitation of a 4 h LC-MS/MS run for triplicate depletions using IgY14-Supermix columns of pooled plasma from individuals having an ectopic pregnancy (EP), normal intrauterine pregnancy (IUP), or spontaneous abortion (SAB). Normalized LFQ protein intensity for ADAM12 is shown. (**b**) 10-plex TMT labeling of the same triplicate pooled plasma samples. Summed intensities of measured peak *m/z* values for the TMT reporter ions representing ADAM12 are shown

be >98% complete, partially modified high-abundance peptides could be readily detected. This significantly increases the complexity of the sample making it more difficult to identify low-abundance peptides. On the other hand, while the label-free method had increased total protein identifications, these samples also have a substantial number of protein quantitations with high CVs (Fig. 2b), due primarily to the stochastic nature of identification of these low-abundance proteins that result in missing values in some replicates.

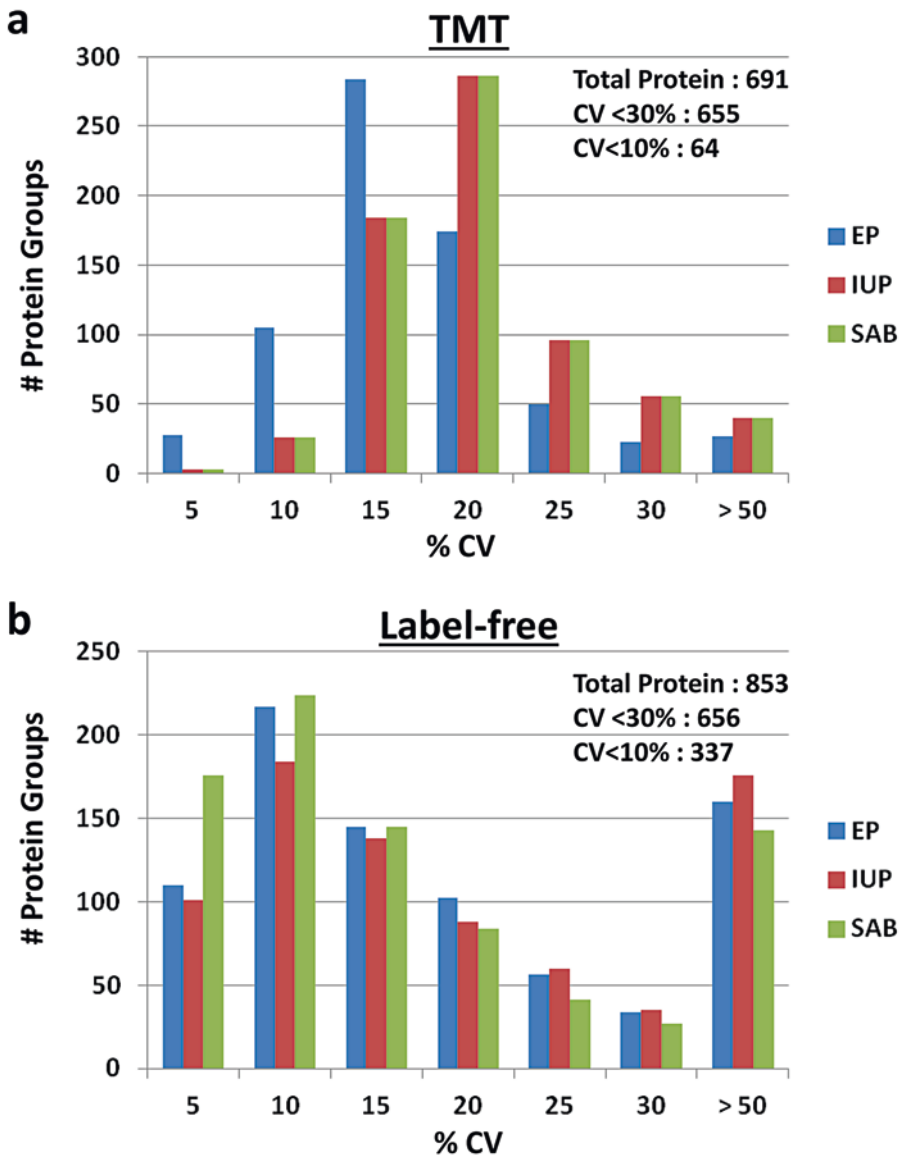


Fig. 2 Coefficients of variation (CV) of protein identifications for a comparison of label-free and TMT quantitation. (**a**) 10-plex TMT labeling of pools of EP, IUP, and SAB plasma that were depleted in triplicate using IgY14-Supermix columns. (**b**) Label-free quantitation of the same triplicate pooled plasma samples. "Total Protein" values indicated represent numbers of protein groups across all samples identified by two or more peptides and with peptide and protein FDR of 1%. "CV <30% or <10%" values represent averages of the three groups of samples (EP, IUP, SAB)

It is therefore interesting that for both quantitative methods, equal numbers of proteins were quantitated with CVs <30%, although the label-free method had the highest number of identifications with CVs <10%. Also, if more stringent filtering criteria are applied to the label-free quantitation, the total number of proteins decreases to

about the level of the TMT experiment and with similar CVs. This comparison of TMT and label-free quantitation is consistent with the other studies conducting similar comparisons summarized above, as it illustrates that the two methods can either yield similar results or one method can outperform the alternative method depending upon the filtering criteria used.

While single TMT experiments are similar to label-free quantitation methods, the number of proteins consistently quantitated across multiple TMT experiments is greatly reduced as noted above. Due to this limitation, we further explored the feasibility of using a modified label-free proteome analysis method to increase depth of analysis in plasma. Although use of isobaric tags can greatly reduce mass spectrometer time for any given level of fractionation compared with label-free quantitation, the missing values across TMT experiments and reduced reliability of quantitation for low-abundance proteins at least partially offset the advantages of multiplexing.

The following protocol describes an in-depth label-free analysis of human plasma samples with MaxQuant. MaxQuant is a freely available software program which uses its own search engine, Andromeda, to search and analyze high-resolution LC-MS/MS data [24, 25]. While the software was primarily developed for analysis of stable isotope labeling with amino acids in cell culture (SILAC) data, MaxQuant also employs the MaxLFQ algorithm for label-free quantitation [26]. In the analysis described herein, three different plasma pools were depleted in triplicate with the IgY14/Supermix tandem immunoaffinity columns followed by 1D–SDS PAGE, trypsin digestion (for complete protocol, see Chapter 7 in this book [27]), and label-free quantitation with MaxQuant. For a similar detailed protocol for MaxQuant quantitation of SILAC datasets, see References [28, 29].

2 Materials

2.1 Hardware Requirements

1. Intel Pentium III/800 MHz or higher (or compatible); a dual-core processor is recommended.
2. 2 GB RAM per thread that is executed in parallel is required.
3. There is no upper limit on the number of cores; however, a multi-core processor operating on a shared memory machine will maximize the parallelization capabilities of the software [28].

2.2 Software and Other Requirements

1. .NET framework 4.5 or higher (downloadable from Microsoft).
2. MSFileReader (downloadable from Thermo Fisher Scientific).
3. 64-bit Windows operating system (Supported versions: Vista SP2, Windows 7, Windows 8, Windows Server 2008, and Windows Server 2012).

4. MaxQuant Software, version 1.5.2.8 or higher (freely available at www.maxquant.org).
5. FASTA Sequence Database (e.g., human Uniprot database, downloadable from www.uniprot.org).

3 Methods

Unless otherwise noted, the following parameters used for label-free quantitation rely on the default settings in MaxQuant which have been optimized by the developers and are appropriate for most label-free experiments. A more detailed description of these and other parameters, including the Andromeda search configuration, can be found in Reference [28].

3.1 MaxQuant: Raw Files Tab

1. Load .raw files. (All .raw files to be analyzed should be stored in the same folder).
2. Define Experiments (*see* **Note 1**).
3. Define Fractions (*see* **Note 2**).
4. Define Parameter Groups (*see* **Note 3**).

3.2 MaxQuant: Group-Specific Parameters Tab

1. Select "General" tab.
2. Set Type: select Standard for label-free quantitation (*see* **Note 4**).
3. Set Multiplicity: select 1 for label-free quantitation (*see* **Note 5**).
4. Set Variable modifications (e.g., methionine oxidation and protein N-terminal acetylation).
5. Specify Digestion mode and Enzyme used (e.g., "Specific" and "Trypsin/P" for full-tryptic cleavage constraints).
6. Set max. # of missed cleavages: default = 2.
7. Set Match type: Match from and to.
8. Select "Instrument" tab.
9. Choose instrument used (e.g., Thermo Fisher Orbitrap, Bruker Q-TOf, AB Sciex Q-TOF). The default parameters will change accordingly based on the instrument selected.
10. Select "Label-free quantification" tab.
11. Set Label-free quantitation: select LFQ.
12. Set the LFQ min. Ratio count to one (*see* **Note 6**).
13. Keep "Fast LFQ" enabled (*see* **Note 7**).
14. Select "Advanced" tab.
15. Set the max. Number of modifications per peptide: default = 5.

3.3 MaxQuant: Global Parameters Tab

1. Select "General" tab.
2. Load a FASTA database file, which has been previously configured in Andromeda (*see* **Note 8**).
3. Set "Fixed modifications" (e.g., carbamidomethyl cysteine).
4. If appropriate, enable "Match between runs" (*see* **Note 9**).
5. Set the Match time window and Alignment time windows to 0.7 min and 10 min, respectively (*see* **Note 10**).
6. Select "Sequences" tab. Parameters include:
   (a) Decoy mode, e.g., revert (reversed sequences).
   (b) Special AAs, e.g., KR.
   (c) Include contaminants: enable (*see* **Note 11**).
7. Select "Identification" tab: keep default settings for a protein and peptide false discovery rate (FDR) of 1% (*see* **Note 12**).
8. Select "Protein quantification" tab. Parameters include:
   (a) Min. ratio count: 2 (Only applies to SILAC labeling).
   (b) Peptides for quantification: Unique + razor (*see* **Note 13**).
   (c) Enable "Use only unmodified peptides" (e.g., methionine oxidation and N-terminal acetylation) and "Discard unmodified counterpart peptide" features.
9. Select "Label-free quantification" tab.
10. If more than one parameter group is used, and LFQ normalizations are to be kept separate (*see* **Note 3**), enable "Separate LFQ in parameter group."
11. Keep default "Stabilize large LFQ ratios," "Require MS/MS for LFQ comparisons," and "Advanced site intensities" settings enabled.
12. Other global parameter tabs (default settings are suitable for most experiments).
    (a) Tables.
    (b) AIF.
    (c) MS/MS-FTMS.
    (d) MS/MS-ITMS.
    (e) MS/MS-TOF.
    (f) MS/MS-Unknown.
    (g) Advanced.

3.4 MaxQuant: Performance Tab

1. In the footer of the program, set the number of parallel threads (physical cores) to be used (*see* **Note 14**).
2. Press "Start." The completed and ongoing processes can be monitored on the Performance page by selecting the "Show all activities" tab.

3. When the analysis is complete, several output files, including the ProteinGroups.txt and Peptided.txt files, will be available within the "combined/txt" folder located in the same folder where the .raw files are stored (*see* **Note 15**).

4 Notes

1. The Experiment column allows you to specify which raw files should be quantified together and which should be kept separate. For example, in an experiment with conditions A, B, and C, where each condition has three replicate runs, by designating the experiments as "A," "B," and "C," all intensities for a given protein will be summed for each condition, and three protein intensities will be reported. On the other hand, replicates may be analyzed separately by giving them nine different experiment names, i.e., A1, A2, A3, B1, B2, etc., and nine protein intensities will be reported. Protein intensities can then be summed for each condition manually or within Perseus software post-analysis. We prefer to treat every sample separately and then perform post-analysis processing such as averaging replicates in Excel.
2. This column is annotated when pre-fractionation (e.g., 1-D PAGE, SCX, high pH separations, etc.) and multiple LC-MS/MS analyses per proteome have been done. Specifying fractions is important if you use the "match between runs" feature (*see* **Note 9**). Identifications will be matched between all raw files with the same or adjacent fraction numbers.
3. Parameter groups are specified when fundamentally different parameters are to be used, e.g., some files may have been from a phosphoproteome analysis and others may have been from a standard proteome analysis. By setting different parameter groups, all data can be analyzed together, choosing appropriate settings for each condition. Parameter groups can be applied to LFQ analyses where a number of files are to be processed together but the experiments are very different from one another, and therefore the underlying assumption of the LFQ algorithm (that for the most part, all samples are similar) does not hold true. When activated, all LFQ normalizations and quantitations will only be done within each parameter group, but the same protein groups will be present across all samples.
4. "Standard" is appropriate for label-free quantitation. Other options include "Reporter ion MS2" or "Reporter ion MS3" for TMT- or iTRAQ-labeled experiments.
5. "Multiplicity 1" indicates that no label was used. Multiplicity 2 or 3 is to be selected when two or three light, medium, or heavy labels are used, such as with SILAC experiments.

6. MaxQuant quantitates protein levels by computing pair-wise ratios of all common peptides occurring in any two samples within an experiment and then using the median of peptide ratios as the pair-wise protein ratio. All pair-wise protein ratios are calculated between any two samples within an experiment to achieve the maximum possible protein quantitation information, and resulting protein ratios are ultimately used to determine LFQ intensity profiles [26]. The LFQ minimum ratio count allows the user to define the minimum number of peptides that has to be available in pair-wise comparisons between two samples. The default setting requires two common peptides to be present between any two samples in an experiment. If less peptides are present, the ratio between these two samples is not used for the determination of the LFQ intensities. However, this does not necessarily mean that the LFQ intensities for these two samples are not calculated since they might be inferred from other pair-wise sample comparisons. We generally change this setting to one to preserve quantitation between samples for proteins that have been identified by two or more peptides somewhere in the dataset, but where individual peptides were identified with low intensity and stochastically. After the MaxQuant search is complete, we typically remove low confidence identifications where all samples are identified by only a single peptide.
7. "Fast LFQ" is the high-speed version of MaxQuant that uses a meaningful subset of comparisons to determine the normalization factors for each LC-MS run and hence significantly reduces computational time [26]. This feature is recommended when LFQ is being applied to large numbers of samples (i.e., more than ten different "experiment" names). If less than ten experiments are being compared, the standard LFQ normalization algorithm is automatically used.
8. For more details regarding Andromeda configuration, *see* [28].
9. When "match between runs" is enabled, identifications are transferred to non-sequenced or non-identified MS features in other LC-MS runs that match identified peptides in one or more runs [30]. The prerequisite for matching identifications is that the peptides have the same mass within the mass tolerance (ppm) of the identified partner, and the peptides elute at the same point in the gradient within the 0.7 min match time window tolerance (*see* **Note 10**).
10. The "alignment time window" is the time window that is used in retention time alignment to search for the best alignment function. The default setting is 20 min; however, we typically reduce this value to 10 min for analyses where mass spectrometer, HPLC, and autoinjector performances were carefully

monitored to maintain consistent retention times between all runs. The "match time window" is the time window allowed during "match between runs" for the transfer of peptide identifications and accounts for potential retention time shifts after the retention time alignment has been performed; hence, a narrow window (<1 min, typically 0.7 min) is optimal to minimize false matches of peptides across runs.

11. MaxQuant provides a contaminants.fasta database file within the software that is automatically added to the list of proteins for the in-silico digestion when this feature is enabled. The contaminants database is located in the conf/contaminants.fasta file within the folder containing the MaxQuant executable files. It is recommended to closely look at the contaminants file before performing database searches. We have found that this file contains many bovine serum proteins which are appropriate contaminants for cell lysate experiments using fetal calf serum. However, a few of these bovine proteins can have high homology to human plasma proteins and can result in true identifications being flagged as contaminants when analyzing plasma proteomes, e.g., actin. Therefore, we replace the MaxQuant contaminants.fasta file with our own list of commonly observed keratin and trypsin contaminants and in the case of cell culture, with a more restricted bovine serum list.
12. FDR is specified at the peptide spectrum match (PSM) and protein level, as determined by the target-decoy approach [31]. The default values are 0.01 (1%); however, they may be increased to relax stringency.
13. This selection will calculate protein ratios using both unique peptides and razor peptide intensities; razor peptides are non-unique peptides, and these are assigned to the protein group containing the largest number of other peptides, according to Occam's razor principle.
14. The number of threads refers to the number of physical computer processing cores. Each raw file will be analyzed by one core and the use of multiple cores will considerably reduce analysis times. However, if all available cores are being used for a MaxQuant analysis, any other computational processes running on the computer will be slowed down. Therefore, for large datasets, a dedicated multi-core processor operating on a shared memory machine is recommended [29].
15. The ProteinGroups.txt file includes information on the identified protein groups in the processed raw files. Each row contains the group of proteins that could be reconstructed from a set of peptides. Raw and normalized intensities can be observed in the "Intensity" and "LFQ Intensity" columns, respectively (Fig. 3).

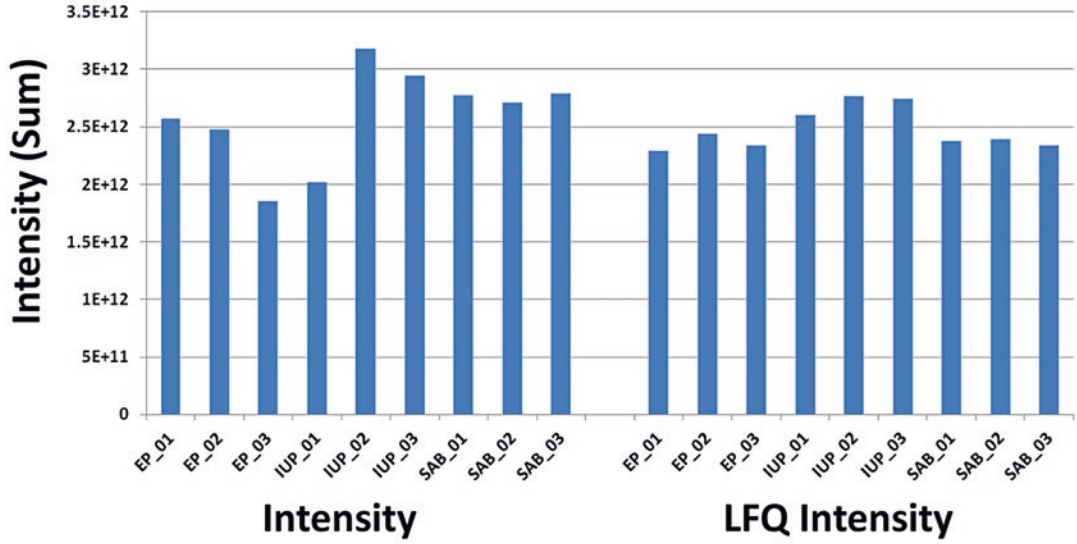


Fig. 3 Summed protein intensities for all protein groups before (Intensity, *left*) and after (LFQ Intensity, *right*) normalization. Single 4 h LC-MS runs of triplicate pools of IgY14-Supermix-depleted EP, IUP, and SAB plasma were quantitated in a label-free analysis using MaxQuant, with "match between runs" enabled

Acknowledgments

This work was supported by NIH Grants RO1HD076279, RO1CA131582, and WW Smith Charitable Trust Grants H1205 and H1305 (D.W. Speicher), PA Department of Health Commonwealth Universal Research Enhancement (CURE) Program Grant (B. Ky), as well as CA10815 (NCI core grant to the Wistar Institute).

References

1. Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1(11):845–867
2. Jacobs JM, Adkins JN, Qian WJ, Liu T, Shen Y, Camp DG 2nd, Smith RD (2005) Utilizing human blood plasma for proteomic biomarker discovery. J Proteome Res 4(4):1073–1085. doi:10.1021/pr0500657
3. Hoffman SA, Joo WA, Echan LA, Speicher DW (2007) Higher dimensional (Hi-D) separation strategies dramatically improve the potential for cancer biomarker detection in serum and plasma. J Chromatogr B Analyt Technol Biomed Life Sci 849(1–2):43–52. doi:10.1016/j.jchromb.2006.10.069. S1570-0232(06)00885-3 [pii]
4. Gulcicek EE, Colangelo CM, McMurray W, Stone K, Williams K, Wu T, Zhao H, Spratt H, Kurosky A, Wu B (2005) Proteomics and the analysis of proteomic data: an overview of current protein-profiling technologies. Curr Protoc Bioinformatics. Chapter 13:Unit 13.11. doi:10.1002/0471250953.bi1301s10
5. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M (2016) Plasma proteome profiling to assess human health and disease. Cell Syst 2(3):185–195. doi:10.1016/j.cels.2016.02.015. S2405-4712(16)30072-2 [pii]
6. Sandberg A, Branca RM, Lehtio J, Forshed J (2014) Quantitative accuracy in mass spectrometry based proteomics of complex samples: the impact of labeling and precursor interference. J Proteome 96:133–144. doi:10.1016/j.jprot.2013.10.035. S1874-3919(13)00550-2 [pii]
7. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, Hamon C (2003) Tandem mass tags: a novel quantifica-

tion strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 75(8):1895–1904

8. Ross PL, Huang YN, Marchese JN, 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, Pappin DJ (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3(12):1154–1169. doi:10.1074/mcp.M400129-MCP200. M400129-MCP200 [pii].
9. Liu P, Beer LA, Ky B, Barnhart KT, Speicher DW (2017) Quantitative comparisons of large numbers of human plasma samples using TMT10plex labeling. Methods Mol Biol 1619
10. Wuhr M, Haas W, GC MA, Peshkin L, Rad R, Kirschner MW, Gygi SP (2012) Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster. Anal Chem 84(21):9214–9221. doi:10.1021/ac301962s
11. Houel S, Abernathy R, Renganathan K, Meyer-Arendt K, Ahn NG, Old WM (2010) Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies. J Proteome Res 9(8):4152–4160. doi:10.1021/pr1003856
12. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics 4(10):1487–1502. doi:10.1074/mcp.M500084-MCP200. M500084-MCP200 [pii]
13. America AH, Cordewener JH (2008) Comparative LC-MS: a landscape of peaks and valleys. Proteomics 8(4):731–749. doi:10.1002/pmic.200700694
14. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B (2007) Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem 389(4):1017–1031. doi:10.1007/s00216-007-1486-6
15. Bantscheff M, Lemeer S, Savitski MM, Kuster B (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem 404(4):939–965. doi:10.1007/s00216-012-6203-4
16. Li Z, Adams RM, Chourey K, Hurst GB, Hettich RL, Pan C (2012) Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. J Proteome Res 11(3):1582–1590. doi:10.1021/pr200748h
17. Megger DA, Pott LL, Ahrens M, Padden J, Bracht T, Kuhlmann K, Eisenacher M, Meyer HE, Sitek B (2014) Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines. Biochim Biophys Acta 1844(5):967–976. doi:10.1016/j.bbapap.2013.07.017. S1570-9639(13)00289-6 [pii]
18. Patel VJ, Thalassinos K, Slade SE, Connolly JB, Crombie A, Murrell JC, Scrivens JH (2009) A comparison of labeling and label-free mass spectrometry-based proteomics approaches. J Proteome Res 8(7):3752–3759. doi:10.1021/pr900080y
19. Schulze WX, Usadel B (2010) Quantitation in mass-spectrometry-based proteomics. Annu Rev Plant Biol 61:491–516. doi:10.1146/annurev-arplant-042809-112132
20. Karp NA, Huber W, Sadowski PG, Charles PD, Hester SV, Lilley KS (2010) Addressing accuracy and precision issues in iTRAQ quantitation. Mol Cell Proteomics 9(9):1885–1897. doi:10.1074/mcp.M900628-MCP200. M900628-MCP200 [pii]
21. Ow SY, Salim M, Noirel J, Evans C, Rehman I, Wright PC (2009) iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". J Proteome Res 8(11):5347–5355. doi:10.1021/pr900634c
22. Ting L, Rad R, Gygi SP, Haas W (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat Methods 8(11):937–940.doi:10.1038/nmeth.1714.nmeth.714 [pii]
23. Beer LA, Tang HY, Sriswasdi S, Barnhart KT, Speicher DW (2011) Systematic discovery of ectopic pregnancy serum biomarkers using 3-D protein profiling coupled with label-free quantitation. J Proteome Res 10(3):1126–1138. doi:10.1021/pr1008866
24. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372. doi:10.1038/nbt.1511. nbt.1511 [pii]
25. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 10(4):1794–1805. doi:10.1021/pr101065j
26. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13(9):2513–2526. doi:10.1074/mcp.M113.031591. M113.031591 [pii].

27. Beer LA, Ky B, Barnhart KT, Speicher DW (2017) In-depth, reproducible analysis of human plasma using IgY 14 and supermix immunodepletion. Methods Mol Biol 1619.
28. Tyanova S, Mann M, Cox J (2014) MaxQuant for in-depth analysis of large SILAC datasets. Methods Mol Biol 1188:351–364. doi:10.1007/978-1-4939-1142-4_24
29. Cox J, Matic I, Hilger M, Nagaraj N, Selbach M, Olsen JV, Mann M (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nat Protoc 4(5):698–705. doi:10.1038/nprot.2009.36. nprot.2009.36 [pii]
30. Geiger T, Wehner A, Schaab C, Cox J, Mann M (2012) Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Mol Cell Proteomics 11(3):M111.014050. doi:10.1074/mcp.M111.014050. M111.014050 [pii]
31. Elias JE, Gygi SP (2010) Target-decoy search strategy for mass spectrometry-based proteomics. Methods Mol Biol 604:55–71. doi:10.1007/978-1-60761-444-9_5

Chapter 24

Blood and Plasma Proteomics: Targeted Quantitation and Posttranslational Redox Modifications

Julie A. Reisz, Katelyn M. Chessler, Monika Dzieciatkowska, Angelo D'Alessandro, and Kirk C. Hansen

Abstract

Proteome profiling using mass spectrometry is extensively utilized to understand the physiological characteristics of cells, tissues, fluids, and many other biological matrices. From the earliest days of the proteomics era, exploratory analyses of the blood protein complement have attracted a great deal of interest, owing to the pivotal importance of blood cells and biofluids (serum, plasma) for research and biomedical purposes. Once challenged by the high dynamic range of protein concentrations, low sensitivity of mass spectrometers, and poor annotation of proteomics databases, the techniques in this field have quickly evolved in recent years, particularly in the areas of absolute quantification of proteins and in mapping of posttranslational modifications. Here we describe (a) the design and production of heavy isotope-labeled peptides used as reporter internal standards for absolute protein quantification and (b) a redox proteomics approach to optimize sample preparation and database searching to elucidate oxidative modifications to protein amino acids. The two methods achieve complimentary goals in the field of blood research and pave the way for future translation of next-generation proteomics technologies into clinical practice.

Key words Quantitative proteomics, Cysteine oxidation, Isotope labeling, Protein expression

1 Introduction

Mass spectrometry-based proteomics provides a powerful analytical approach to identify and robustly quantify proteomic changes, including expression levels, binding partners, regulatory posttranslational modifications (PTMs), and localization [1–5]. Recent strides in proteomics technologies have enabled investigators to provide extensive maps of the human proteome complement to the genome [6–9]. These advances in our understanding of the human proteome have been fostered by the introduction of novel high-resolution/fast-scanning mass spectrometers, improved database annotation, and implementation of next-generation database searching algorithms [10]. Despite such advances, blood-derived cells and biofluids have been underrepresented in the recent

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_24,

updated drafts of the human proteome [6–9]. On the other hand, ante litteram proteomics applications were performed mostly on blood-derived cells and biofluids during the early days of electrophoresis [11], followed by electrophoretic studies in the late 1970s and early 1980s (reviewed in [12]). Owing to their accessibility and relative abundance, in addition to their clinical diagnostic value [13], blood and blood-derived cells have been extensively explored over the past four decades through different proteomics approaches, an effort that culminates in the identification of 3784 serum/plasma proteins [14]; 2838 and 1989 immature and red blood cell proteins, respectively [15, 16]; and almost 4000 platelet proteins [17].

The initial tide of exploratory discovery-mode proteomics investigations was aimed at qualitatively describing the proteomes of blood fluids and cellular components. These early studies provided key information for the closely related fields of hematology [18] and immunology [19], along with regenerative [20] and transfusion medicine [21, 22]. Ultimately, proteomics tools have been exploited in a wide array of functional studies on blood and blood-derived matrices to understand protein interactions [22, 23], metabolic regulation [24], and molecular signaling in health and disease [25–28]. Blood and its derived products are attractive for clinical/medical study as they are easy to obtain from donors in adequate amounts over multiple time points (longitudinal studies). In addition, the blood proteome composition is largely reflective of the overall physiological status rather than focused on one particular organ or tissue. However, a significant challenge exists in processing blood samples in the stringent and time-consuming manners required due to the high sensitivity of mass spectrometers, where impurities and background signals undoubtedly interfere with, and sometimes eliminate, the detection of analytes of interest. This is particularly true for plasma, red blood cell, and platelet proteomics, where protein concentrations span up to 10 orders of magnitude and the proteome composition is largely quantitatively biased toward extremes in abundances for a handful of proteins (e.g., albumin, immunoglobulins, hemoglobins, actin [12]). Of note, disproportionate abundance of proteins such as hemoglobins—accounting for ~90% of the dry weight of mature erythrocytes—hampers the detection of low abundance proteins and sub-stoichiometric PTMs, such as redox modifications of residues altering protein activity and/or function [2].

In this article we describe two approaches to cope with these issues by performing targeted quantitative and redox proteomics approaches in order to facilitate the routine detection of low abundance or redox-modified blood proteins. These methods are based on our previous optimization of routine workflows applicable to diverse biological matrices, including, but not limited to, plasma, platelets, and red blood cells [29–36]. The first approach to be

covered is targeted quantitation of proteins through the use of stable heavy labeled internal peptide standards [29, 30, 37–39] (Fig. 1), and the second is a switch-tag redox proteomics approach to investigate modifications associated with oxidative stress and/or signaling [31–33] (Fig. 2). These strategies are highly complementary, and though we describe their use in parallel, they may be used separately if needed.

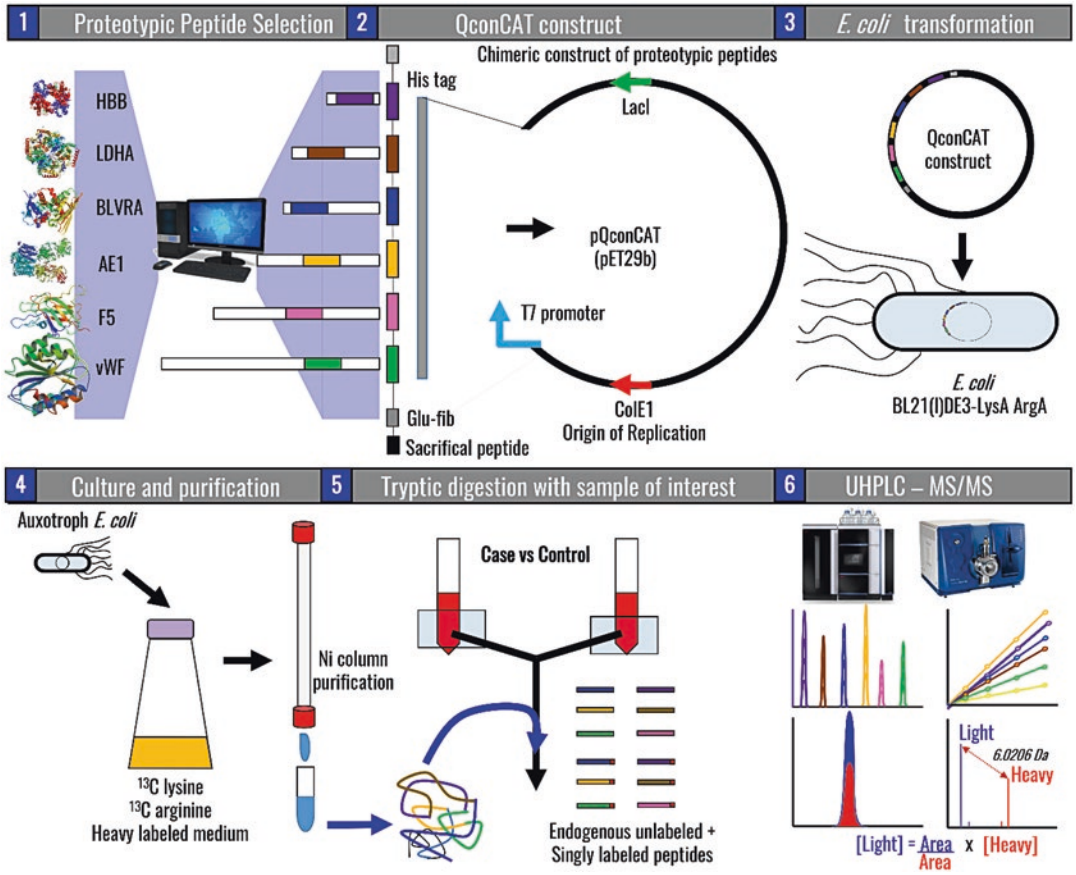


Fig. 1 The QconCAT approach for absolute quantitation of blood proteins. An overview of the QconCAT workflow. Proteotypic peptides are selected for proteins of interest to ensure uniqueness and specificity of the peptide sequence on the basis of organism and protein isoform characteristics [1]. A gene construct is designed as to code for a chimeric protein that includes all the proteotypic peptides. The construct is loaded onto an expression vector [2] before transformation of an *E. coli* strain that is auxotroph for arginine and lysine [3]. Bacteria are grown in media supplemented with lysine and arginine isotopologues that are stably labeled with ^{13}C at all six carbon atom positions [4]. Recombinant protein is thus expressed in the recombinant system prior to purification and quantitation [4]. The purified chimeric protein is used as internal standard for analytical variables of the sample of interest [5], from the digestion efficiency to retention time reproducibility and absolute quantification through direct ratios of peak areas for the light endogenous peptide vs the known amounts of the spiked in proteotypic peptide (labeled at C-term arginine or lysine—6). The figure is adapted from Ref. 29, upon the introduction of significant modifications consistently with the application to quantitation of blood proteins described here

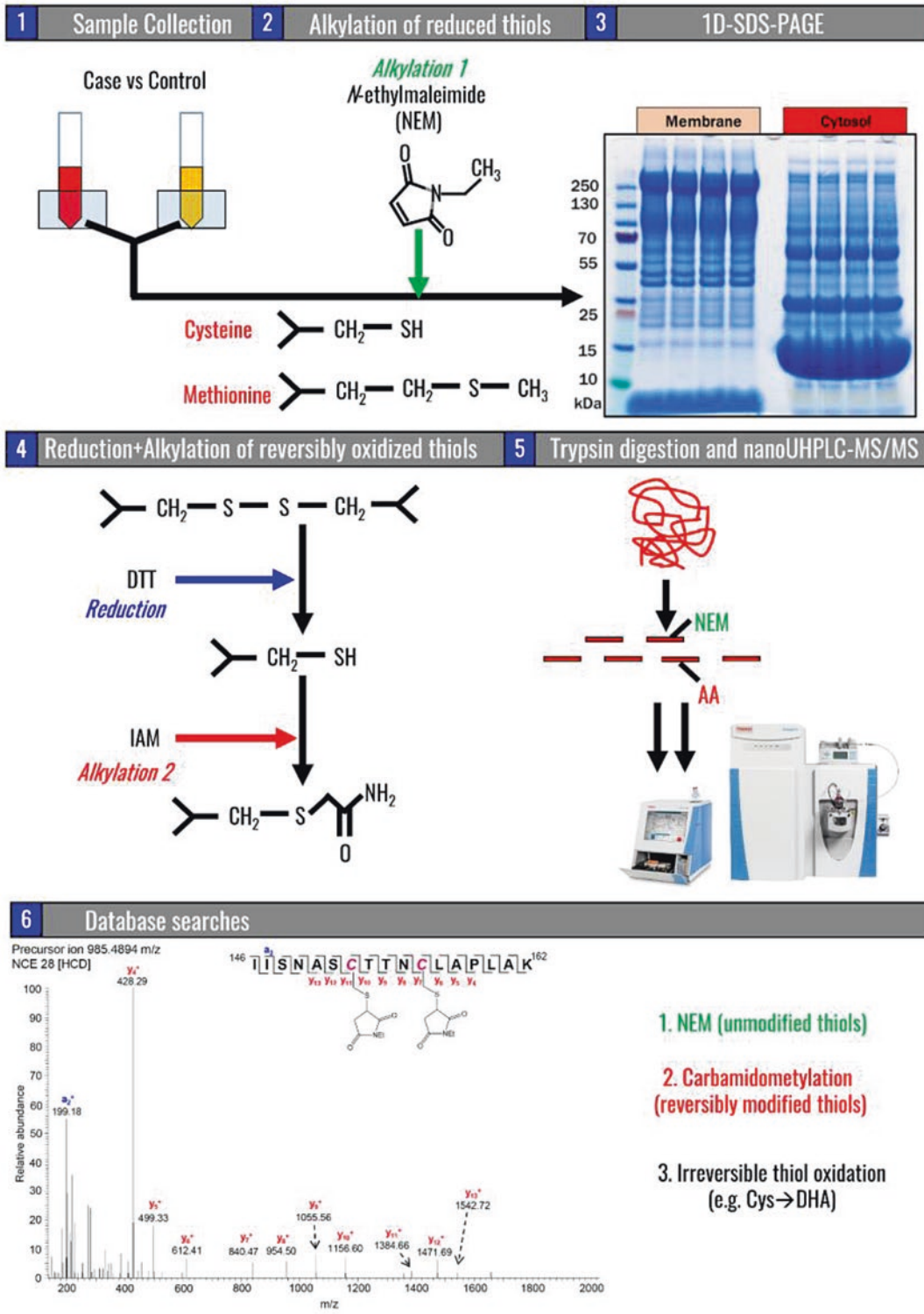


Fig. 2 Switch-tag approach for redox proteomics application in blood research. From 1 to 6, an overview of the workflow for redox proteomics application through the switch-tag approach is described. Steps include preliminary alkylation through N-ethylmaleimide (NEM) of reduced thiols during sample extraction. 1D–SDS-PAGE separation of proteins, prior to reduction (dithiothreitol—DTT) and alkylation (e.g., with iodoacetamide) of reversible oxidized thiols (e.g., disulfide bonds). Finally, trypsin digestion and MS/MS analysis are performed, prior to extensive database searches as detailed in the main body of the article

Developments in protein quantification by mass spectrometry have evolved by necessity due to the highly differential ionization properties of peptides and technical reproducibility issues over the analysis of large sample sets (oftentimes due to the stability of ion spray) [40, 41]. Enzymatic digestion of intact proteins in bottom-up proteomic approaches (the current standard in the vast majority of proteomics laboratories) results in the generation of peptides characterized by diverse chemical-physical properties, such as varying length of sequence, charge, differential polarities, and presence of PTMs. This diversity influences peptide ionizability to such an extent that no two peptides can be expected to ionize identically. Therefore relative quantitation can be inaccurate when determined using ratios of signal intensities or areas. To address this challenge, a multitude of MS-based quantification strategies have been developed, the majority of which rely on stable isotope labeling (^{13}C and ^{15}N) either endogenously, using standards, or by tagging during sample preparation [41–43]. A streamlined quantitative method has been optimized in our laboratory that exploits the QconCAT (quantitative concatemers) approach, in which stable isotope-labeled peptides are produced to be used as reporters for proteins of interest in analytical samples [29, 30, 36–39]. This strategy provides reproducible absolute quantification, typically achieved through a targeted MS method, but requires a significant time investment early in the workflow as peptide sequence selection, protein expression and purification, and assay validation are all upstream of sample analysis. In order to ease reproducibility of the method in other laboratories, a step-by-step description of the protocol to generate and exploit chimeric QconCAT proteins is detailed in the sections below.

Targeted methods, where data is acquired for only the peptide and fragment ions of interest, offer tremendous sensitivity and robustness of quantification, but are not useful for discovery purposes. Many targeted approaches are achieved using selected reaction monitoring (SRM) on a triple quadrupole MS, where the quadrupoles are used chronologically to select a peptide ion (by m/z ratio), induce fragmentation, and detect one or more fragment ions, respectively. In contrast, discovery proteomics is based on the ability of high-resolution, rapid-scanning MS instruments to monitor all ions entering the detector and reporting them as a function of time throughout the analysis. Because general discovery methods, also termed “global” or “untargeted” proteomics, have been extensively detailed by our group and others [6, 7, 44], the second focus of this chapter is to describe discovery approaches aimed at uncovering sub-stoichiometric redox modifications to protein amino acids, with a special emphasis on the sulfur-containing cysteine and methionine residues [31]. Technical considerations in terms of data analysis follow standard protocols; distinct changes exist here in both sample preparation and data analysis with

extended database searching, for example, care to preserve physiological redox modifications while minimizing artifact introduction along with database searches to targeting specific irreversible modifications to redox-sensitive residues. The method detailed herein covers design and synthesis of QconCAT proteins, SRM method validation, sample preparation with depletion of highly abundant proteins, mass spectrometry data acquisition (either in a data-dependent or data-independent acquisition—DDA or DIA MS^2 approach [45, 46]), and post hoc analysis, including interpretation of quantification data and strategies for redox-focused database searching. Finally, even though not directly described in this manuscript, it is possible to combine the discovery-mode approach aimed at identifying redox modifications and the targeted quantification of the identified residues through in silico prediction (e.g., through the freely available software Skyline—https://skyline.gs.washington.edu/labkey/project/home/software/Skyline/begin.view [47]) of the expected m/z for the parent ion and expected fragment ion series.

2 Materials

Prepare all solutions using mQ H_2O (18 MΩ cm) and analytical grade reagents. Solutions can be stored at room temperature unless otherwise stated. When indicated, volumes are sufficient for six 1 L cell growths and the subsequent purifications.

2.1 General

1. Benchtop centrifuge capable of 16,000 × *g* with refrigeration to 4 °C.
2. Vortex.
3. Speed vacuum.
4. Probe sonicator with pulse feature.
5. Standard spectrophotometer (for absorbance readings in the visible range, including 600 nm).
6. Benchtop plate reader.

2.2 Protein Expression and Inclusion Body Preparation

1. 50 mg/mL ampicillin solution: Prepare in sterile distilled water. Store at −20 °C in 1 mL aliquots for up to several months.
2. 1 M isopropyl β-D-1-thiogalactopyranoside (IPTG) solution: Prepare in sterile distilled water and store at 4 °C in the dark until use.
3. Plasmid vector pET21a with desired QconCAT sequences.
4. Competent *E. coli* BL21(DE3) LysA, ArgA. Frozen competent cells of BL21(DE3) can be obtained from many suppliers, including Stratagene, Promega, and Genlanatis.

5. Luria broth (LB): Dissolve 25 g of LB powder in 1 L of distilled water. Sterilize by autoclaving for 15 min at 121 °C.
6. LB-ampicillin agar plates.
7. Incubator capable of shaking at 250 rpm.
8. 10× M9 Salt solution: Dissolve 60 g of NaH_2PO_4, 30 g of KH_2PO_4, 5 g of NH_4Cl, and 2.5 g of NaCl in 1 L H_2O. Confirm pH is 7.3–7.5, and then filter sterilize.
9. 84 μg/μL of $^{13}C_6$ L-arginine HCl stock in water, consistent with SILAC media formulation, as optimized for the *E. coli* strain used in the workflow here described.
10. 180 μg/μL of $^{13}C_6$ L-lysine HCl stock in water, consistent with SILAC media formulation, as optimized for the *E. coli* strain used in the workflow here described. Prepare fresh.
11. M9 minimal media: To 800 mL H_2O, add 100 mL of 10× M9 salts, 2 mL of 1 M $MgSO_4$, 10 mL of 20% glucose or 2 g of glucose, 1 mL of 100 mM $CaCl_2$, 1 mL of 100 mM $FeSO_4 \cdot 7H_2O$, 200 μL of 0.5% vitamin B_1, 1 mL of 50 mg of ampicillin, 595 μL of $^{13}C_6$ L-arginine HCl stock*, and 280 μL of $^{13}C_6$ L-lysine HCl stock*. Bring up to 1 L with H_2O. Prepare 1 L per construct, plus an additional 100 mL for each starter culture. Filter sterilize all solutions prior to addition. *Add amino acids immediately before use, after filter sterilization.
12. Glass homogenizers, 50 mL capacity. Have one clean homogenizer available for each QconCAT construct prepared in parallel.
13. Break buffer: 100 mM Tris–HCl, 100 mM NaCl, 10% glycerol, 1 mM EDTA, pH 8.0. Store at 4 °C.
14. Wash buffer 1: 25 mM Tris–HCl, 100 mM NaCl, 1% Triton-X100, 1 mM EDTA, pH 8.0. Store at 4 °C.
15. Wash buffer 2: 25 mM Tris–HCl, 100 mM NaCl, 0.5% Triton-X100, 1 mM EDTA, pH 8.0. Store at 4 °C.
16. Wash buffer 3: 25 mM Tris–HCl, 50 mM NaCl, pH 8.0. Store at 4 °C.
17. Lysozyme: 10 mg/mL in break buffer. Store at −20 °C in 1 mL aliquots.

2.3 Nickel Affinity Purification

1. Bio-Rad Econo-Column chromatography columns, 50 mL capacity with stopper.
2. 6–8 kDa MWCO dialysis membrane.
3. High Affinity Ni-Charged nitrilotriacetic acid (NTA) Resin (GenScript).
4. Loading buffer: 6 M guanidinium chloride, 500 mM NaCl, 50 mM NaH_2PO_4, 20 mM imidazole, 0.05% w/v sodium azide, pH 7.4.

5. Elution buffer: 6 M guanidinium chloride, 500 mM NaCl, 50 mM NaH_2PO_4, 400 mM imidazole, 0.05% w/v sodium azide, pH 7.4.
6. Dialysis buffer: 10 mM ammonium bicarbonate (ABC), pH 8.5, the pH is adjusted with ammonium hydroxide.

2.4 SDS-PAGE

1. Ammonium persulfate (APS): 10% w/v in water. Store at −20 °C.
2. Tetramethylethylenediamine (TEMED): Store at 4 °C. Very odorous. Open in fume hood.
3. Polyacrylamide gel: 12 mL Amresco Next Gel 10% acrylamide. Add 72 μL of 10% APS and 7.2 μL of TEMED to catalyze polymerization. Makes two gels.
4. Gel running buffer: SDS-PAGE running buffer (20×) from commercial source, diluted to 1× with water.
5. Staining buffer: Dissolve 0.625 g of Coomassie Brilliant Blue R 250 in 500 mL of 50% methanol, 10% acetic acid in water.
6. Destaining buffer: 50% methanol, 10% acetic acid in water.
7. Gel loading buffer for reducing gel: 2× Laemmli buffer with 5% β-mercaptoethanol.
8. Gel loading buffer for nonreducing gel: 2× Laemmli buffer.
9. Cold ethanol for ethanol precipitation.
10. Prestained molecular weight standards.

2.5 Filter-Aided Sample Preparation (FASP)

1. Sartorius VIVACON 500 10 or 30 kDa MWCO filters.
2. 8 M urea buffer in 100 mM ABC, pH 8.0. Freshly prepare 1 mL per sample.
3. 50 mM IAM solution: 50 mM iodoacetamide in urea buffer. Freshly prepare 100 μL per sample and store in dark.
4. 10 mM DTT solution: 10 mM dithiothreitol in urea buffer. Freshly prepare 100 μL per sample and store at 4 °C until used.
5. 50 mM ammonium bicarbonate (ABC) in water. Prepare 250 μL per sample.
6. 25 mM ammonium bicarbonate (ABC) in water. Prepare 250 μL per sample.
7. Sequencing grade modified trypsin. (cat #V5113, Promega).
8. ProteaseMAX (0.02% w/v) in 50 mM ABC: Prepare 100 μL per sample.
9. 10 mM ABC/0.2% FA in water. Prepare 0.2 mL per sample.

2.6 Immunodepletion of Plasma Samples

1. Blood/plasma samples of interest.
2. Immunodepletion resin (R&D Systems, Inc., Minneapolis, MN, or alternative products from other brands, e.g., Agilent).

2.7 Analysis of Posttranslational Redox Modifications

1. *N*-ethylmaleimide (NEM). Store at 4 °C until needed.
2. Bicinchoninic acid (BCA) assay or Bradford assay kit.
3. Nonreducing Laemmli buffer (2× or 4×).
4. Dithiothreitol (DTT). Store at 4 °C until needed.
5. Iodoacetamide (IAM). Store at 4 °C in the dark until needed.
6. 4–12% Bis-Tris gradient precast gel. Store at 4 °C until needed.
7. SDS-PAGE running buffer (10–20×) from commercial source, diluted to 1× with water.
8. Coomassie Blue stain from commercial source.

2.8 Mass Spectrometry Data Acquisition and Analysis

1. Mass spectrometer capable of data-dependent or other untargeted analysis with MS^1 and MS^2 acquisition. Instrument should be interfaced with a LC system, preferably nanoLC.
2. Mass spectrometer capable of selected reaction monitoring or parallel reaction monitoring, for example, a triple quadrupole or quadrupole-Orbitrap. Instrument should be interfaced with a LC system.
3. Highest-quality water and acetonitrile, each containing 0.1% v/v formic acid. We currently use Optima grade from Fisher Scientific.
4. Skyline software (freely available at https://skyline.gs.washington.edu/).

3 Methods

3.1 QconCAT Sequence Design

1. Select the proteins of interest to measure. Species and protein isoforms should be carefully selected owing to their unique amino acid sequence.
2. Select tryptic peptides from previous experimental proteomics data (untargeted) and publicly accessible databases, including SRM Atlas (http://www.srmatlas.org/), Global Proteome Database (http://gpmdb.thegpm.org/), and Passport (http://passport.maccosslab.org/). Ideally, two to four peptides are chosen per target protein isoform to ensure specificity and accurate quantification.

   We recommend the following criteria for tryptic peptide selection:

   (a) Sequence length of 7–18 residues.

   (b) Sequence unique to the protein of interest and previously observed by MS (preferably on the same instrument type in terms of ionization source and fragmentation technique).

   (c) Avoid residues that are prone to modifications: M, C, N-terminal Q or E, or containing known posttranslational

modification sites in the native protein (e.g., phosphorylation, glycosylation) (*see* **Notes 1** and **2**).

(d) Avoid sequences that commonly result in missed cleavages (e.g., Lys-Lys and Arg-Arg).

3. When designing more than one QconCAT construct, group peptides from proteins with similar endogenous abundances in the samples of interest. This allows for adjustment of each QconCAT construct concentration in samples after a pilot run has been performed to optimize the heavy to light ratio (as close to 1:1 for as many of the proteins covered by the construct as possible). The natural flanking sequences can be included to better mimic the local sequence of the native peptides which can be particularly important for trypsin digestion kinetics.
4. Add a minimum of two internal standard peptides for which pure peptides or purified peptide mixes are commercially available. The concentration of the resulting QconCAT construct will be quantified by reference to the accurately quantified standard peptides. We use one alcohol dehydrogenase (ADH, UniProt P00330) peptide and one β-galactosidase from *Escherichia coli* per QconCAT protein for this purpose.
5. Add a hexahistidine purification tag (His-tag) at the C-terminus for purification of the expressed QconCAT.
6. Clone the gene product into the NdeI and BamHI sites of the pET-21a vector.

3.2 Protein Expression

1. Transform 50 μL of competent BL21(DE3) LysA, ArgA with 10–100 ng plasma QconCAT plasmid; incubate on ice for 20 min. This strain is an auxotroph for arginine and lysine, i.e., it grows only in the presence of exogenously supplemented K or R in culture media.
2. Heat shock cells in a 42 °C water bath for 30 s and return to ice for 2 min.
3. Add 200 μL of LB to cells and incubate for 1 h at 37 °C with shaking at 150 rpm.
4. Spread 100 μL of transformed cells on a LB agar plate containing 100 mg/mL ampicillin. Incubate the plate overnight at 37 °C.
5. Inoculate 100 mL of M9 medium (supplemented with 100 μL ampicillin stock) with multiple colonies, and incubate at 37 °C with shaking overnight at 150 rpm.
6. Inoculate 1 L of M9 medium (supplemented with 1 μL ampicillin stock) in a 4 L baffled-bottomed shake flask with 50 mL of the small overnight growth from **step 5**. Add the $^{13}C_6$ L-arginine HCl and $^{13}C_6$ L-lysine HCl stocks to the media just before inoculation.

7. Shake the flasks at 250 rpm and 37 °C until the cells reach mid-log phase ($OD_{600\ nm}$ approximately 0.6–0.8). Remove 250 μL of sample of un-induced cells for SDS-PAGE analysis.
8. Add 1 mL of 1 M IPTG stock to each flask for induction, and resume shaking for 4–5 h. To monitor protein induction, collect 250 μL of samples hourly for SDS-PAGE analysis. For low expressing contracts, it may be necessary to immunoblot using an anti-polyHis antibody.
9. Harvest cells from liquid culture by centrifugation for 15 min at 6000 × *g* and 4 °C. Remove 250 μL of sample of induced cells for SDS-PAGE analysis.
10. Transfer cells into a pre-weighed 50 mL centrifuge tube, determine the wet weight of the cell pellets, and store cell pellets at −20 °C. An average yield should be 2–3 g wet pellet per 1 L growth.

3.3 Inclusion Body Preparation

1. Thaw frozen cells at room temperature for 10 min. Add 5 mL of break buffer per gram of cell pellet, with a 20 mL minimum so the volume is adequate for sonication. Allow cell pellet to resuspend for 30 min on a rotator at 4 °C, and then pipet up and down with a serological pipette to complete resuspension.
2. Transfer cell suspension to glass beaker, and then add 100 μL of 10 mg/mL of lysozyme stock solution. Stir for 30 min at 4 °C.
3. Sonicate the cells on ice at 60% power in cycles of 30 s on and 60 s off, for a total of five cycles. Cool probe on ice for 1 min prior to sonication. Remove 20 μL of homogenate and spin at 15,000 × *g* for 10 min at 4 °C.
4. Spin the sonicated cell lysate at 15,000 × *g* and 4 °C for 15 min.
5. Wash pellet with 25 mL of wash buffer 1. Resuspend pellet with a serological pipette and transfer to a cold glass homogenizer. Pump pestle up and down slowly until no clumps remain and homogenate is evenly distributed. Remove 20 μL for SDS-PAGE analysis, and then centrifuge the homogenate at 15,000 × *g* for 15 min at 4 °C.
6. Repeat **step 5** first with wash buffer 2 and then with wash buffer 3 for a total of three washes.
7. Resuspend the pellet in 25 mL of loading buffer for nickel purification.
8. Prechill samples on ice for 15–20 min. Prechill sonication probe for 1 min.
9. Sonicate solution on ice at 90% power for three cycles of 20 s on, 60 s off. Centrifuge solubilized inclusion bodies at 15,000 × *g* for 30 min at 4 °C. Supernatant may be directly loaded onto the nickel column. Retain 20 μL of supernatant for SDS-PAGE.

10. Prior to SDS-PAGE analysis, remove guanidine from the inclusion body fractions by precipitation with nine volumes of cold ethanol. Verify the presence of the desired plasma QconCAT in the inclusion body via SDS-PAGE and subsequent Coomassie Blue staining.

3.4 Nickel Affinity Chromatography Purification

1. Gently resuspend the resin by inverting the bottle several times.
2. Transfer 10 mL of resin slurry into the gravity flow column with stopper in place. The resin has a binding capacity of about 3 mg of protein/mL of resin. Because the resin is supplied as 50% slurry, transferring 10 mL results in 5 mL of settled resin.
3. Allow the resin to settle by draining the storage buffer until it reaches the top of the resin bed. Do not let the resin bed dry out.
4. Wash the resin with two bed volumes of DI water and then three bed volumes of loading buffer.
5. Apply the clarified crude extract onto the column and incubate at room temperature for 60 min with rocking.
6. Allow resin to settle and drain flow through until it reaches the top of the resin bed. Collect 20 μL of the unbound sample for SDS-PAGE analysis.
7. Wash the resin with five bed volumes of washing buffer, and collect the wash fractions. The column should be extensively washed until the A_{280} of the eluate is stable and near that of the wash buffer.
8. Elute the bound proteins by adding five bed volumes of elution buffer and collect the elutions in several small fractions of about 10 mL each. Collect 20 μL of each fraction for SDS-PAGE analysis (*see* **Note 3**).
9. Wash column with one bed volume of 1 M imidazole. Collect 20 μL of this fraction for SDS-PAGE analysis and retain eluate, but do not pool with other elution fractions.
10. Pool fractions containing desired plasma QconCAT, usually about 50 mL, and dialyze against 5 L of 10 mM ABC pH 8.5, to remove contaminating reagents such as imidazole and guanidine. Dialyze at 4 °C overnight.
11. If the protein precipitates during dialysis, pellet the slurry of insoluble protein by centrifugation at 16,000 × *g* for 20 min at 4 °C.
12. If the protein does not precipitate during dialysis, concentrate the protein using the Amicon stirred cell fitted with a YM10 membrane according to the manufacturer's instructions. Once protein is visible, pellet the slurry of insoluble protein by centrifugation at 16,000 × *g* for 20 min at 4 °C.

13. The purified QconCAT protein pellet can be stored at −80 °C directly or can be solubilized in 8 M urea and aliquoted before storage at −80 °C.
14. BCA or Bradford assay can be used to determine the protein concentration. However, more accurate assessment of protein concentration can be achieved by MS analysis (XIC or SRM) of the ADH or β-galactosidase peptides included into the QconCAT design with reference to a known amount of unlabeled ADH or β-galactosidase.

3.5 SDS-PAGE Analysis

Before analysis by SDS-PAGE, guanidine must be removed by cold ethanol precipitation (**steps 1**–7).

1. Add nine volumes of 100% cold ethanol to each protein sample.
2. Store at −80 °C for 2 h or −20 °C overnight.
3. Spin at 15,000 × *g* at 4 °C for 10 min and discard supernatant.
4. Wash pellet with 100 μL of 90% cold ethanol by resuspending and vortexing briefly.
5. Spin at 15,000 *g* at 4 °C for 10 min and discard supernatant.
6. Repeat **steps 4**–**5** once more.
7. Dry pellet using speed vac, taking care not to overdry.
8. Resuspend pellets in 50 μL of SDS-PAGE loading buffer. Gel samples not containing guanidine can be resuspended in gel loading buffer directly.
9. Boil samples at 95 °C for 8 min, and load appropriate volumes (~10 μL) onto a prepared SDS-PAGE gel.
10. Run gel at 160 V until the dye front reaches ~3 cm from the bottom of the gel. After running, stain gel with staining buffer for 1 h, and then destain using destaining buffer until only bound protein is visible.

3.6 Plasma QconCAT Characterization

1. Evaluate purity of the QconCAT protein by SDS-PAGE and Coomassie Blue staining.
2. Perform in-gel digestion on a protein band corresponding to expected molecular mass of the plasma QconCAT followed by LC-MS/MS analysis to confirm the sequence of the expressed QconCAT protein [29, 30, 36, 39].

3.7 Digestion Optimization

1. Optimize digestion conditions such as enzyme concentration, and length of digestion incubation time, starting from previously optimized conditions.

3.7.1 LC-SRM Method Development

1. Alternatively, parallel reaction monitoring (PRM) may be utilized here if a quadrupole-Orbitrap instrument is available.

2. Generate tandem mass spectra for the plasma QconCAT peptides of interest by infusion or by LC-MS/MS.
3. Determine the charge state distribution (MS^1 level) for each peptide and select the most abundant charge state(s) for further monitoring (*see* **Note 4**).
4. Determine reproducible retention times for each peptide using LC information in **step 1** or by running in triplicate the QconCAT peptide mixture.
5. Optimize the MS^2 collision energy to establish the most favorable fragmentation conditions for each peptide. Use instrument vendor-specific linear equations based on m/z values to establish the most sensitive collision energies for your peptide(s) (further details about this step are provided at these references [48, 49]).
6. Determine the limit of quantitation (LOQ), limit of detection (LOD), and linear dynamic range of the assay by a selected reaction monitoring (SRM) experiment.
7. Determine $^{13}C_6$ arginine and $^{13}C_6$ lysine incorporation (%) into the QconCAT at the peptide level using a SRM experiment. This information will provide a correction factor, if applicable, when quantifying experimental samples.
8. Develop a corresponding LC-SRM method for the m/z values of the endogenous (light) peptides, keeping the method parameters developed for the heavy plasma QconCAT peptides consistent between the two.

3.8 Absolute Quantification Using a Targeted Proteomics Analysis with QconCAT Peptides

1. Plasma sample preparation: Immunodepletion and normalization. Remove albumin and IgG from plasma or blood samples using serum protein immunodepletion resins (Proteome Purify 2, R&D Systems, Inc.) according to the manufacturer's protocol, as extensively reported [30, 44, 50]. We recommend mixing 10 μL of plasma samples with 1 mL of suspended immunodepletion resin and incubating end-over-end at room temperature for 30–60 min.
2. Transfer the resin into the upper chamber of a Spin-X 0.22-μm filter unit, and spin for 2 min at 2000 × *g*. Depleted samples may then be concentrated using spin concentrators (3000 MWCO, EMD Millipore).
3. Measure protein concentrations and normalize using 25 mM ABC, pH 7.5.
4. For each sample, add a known amount of desired plasma QconCAT construct(s). A general rule of thumb is to have ~100 fmol of each QconCAT protein present per injection, but the actual amount should be tailored to the endogenous levels of the proteins of interest in the particular sample set.

5. Digest samples using the filter-aided sample preparation (FASP) protocol [51].
6. Perform LC-SRM with the optimized assay.
7. Determine the peak area ratio for targeted peptide (heavy/light isotopologue ratio—extensive details about this step are provided here [52] and further described with in-depth video tutorials at this link:

   https://skyline.gs.washington.edu/labkey/project/home/software/Skyline/begin.view).

3.9 Identification of Posttranslational Redox Modifications Using Untargeted Proteomics Analysis

1. Add NEM (~10 mM in 50 mM ABC) to normalized, immunodepleted plasma samples, approximately 0.2 mg of protein each. Incubate at room temperature for 30 min to alkylate reduced cysteine thiols. SDS (1% w/v final concentration) or other surfactants may be added to denature proteins if desired.
2. Remove an aliquot (~40 μg) from the reaction mixture, and mix with an equal volume of nonreducing Laemmli buffer (2×) to prepare sample for SDS-PAGE. The remainder of the reaction mixture may be stored at −20 °C.
3. Load gel and run at 180 V for ~45 min. Time and voltage may be adjusted for compatibility with gel manufacturer's recommendations (*see* **Note 5**).
4. Stain gel in clean box with staining buffer for 1 h, and then destain using water until protein bands are clearly visible. Keep box covered during incubation periods to minimize the introduction of dust or contaminants.
5. Cut desired molecular weight regions from each lane, taking into consideration that disulfide-containing protein hetero- or homodimers likely exist, resulting in molecular weight increases.
6. Perform in-gel digestion of bands as previously described [53]. Include reduction and alkylation steps such that DTT-reducible cysteine redox forms become iodoacetyl modified.
7. Resuspend dried peptide extracts in 20 μL of 0.1% formic acid, and inject 10 μL for analysis by LC-MS/MS. Utilize a 60–80 min gradient of water (0.1% formic acid) with increasing acetonitrile (0.1% formic acid) content along with a C18 column for separation of peptides. Operate the mass spectrometer in data-dependent acquisition mode, where the n (generally 8–12) most abundant precursor ions are sequentially selected for MS^2 fragmentation.

   Alternatively, data-independent acquisition (DIA) can be performed, as detailed by Egertson and colleagues [46]. In DIA, MS^2 scans are collected systematically and independently of precursor information. Select a mass window range (e.g., 500–900 m/z), and identify multiple isolation windows per cycle (e.g., 100 windows, each 4 m/z wide) on a Q Exactive series mass spectrometer (Thermo Scientific).

8. When acquisition is completed, search data files against proteomics databases using the Mascot search engine or other preferred search method. For Mascot searches, files should first be converted to Mascot generic format (.mgf) using MassMatrix or ProteoWizard. We recommend searching together all peptide samples originating from the same gel lane. This may be accomplished by merging .mgf files to one comprehensive peak list using the command line (*see* **Note 6**). Merging .mgf files holds advantages such as increasing confidence in the assignment of identified proteins and sub-stoichiometric PTMs. When merging is performed at the post search stage (e.g., Scaffold elaboration), it also allows to retain the information related to the original band when a protein and its modification are identified, easing follow-up preparative or analytical investigations.
9. Perform one or more database searches with the following redox modifications listed as variable (or dynamic) modifications: Cys dioxidation (+32 Da), Cys to dehydroalanine conversion (−34 Da), and Met, His, Trp, and/or Tyr monooxidation (+16 Da). Include Cys carbamidomethylation (+57 Da) and NEM-labeled Cys (+125 Da) as variable modifications to account for reversibly oxidized thiols (e.g., disulfide, sulfenic acid) or unmodified as well (*see* **Notes 7–9**).

4 Notes

1. Efforts to quantify a protein of interest and identify new redox PTMs are best achieved by performing these two approaches in parallel on the same samples. For example, one may determine Cys and Met oxidation status through an untargeted redox experiment *and* quantify the overall protein content by monitoring non-redox-sensitive peptides using the QconCAT approach.
2. If the QconCAT must contain cysteinyl residues, it is also necessary to reduce and alkylate cysteinyl residues. Reducing agents such as TCEP and DTT are used to reduce disulfide bonds. TCEP offers several advantages including greater resistance to oxidation; it is a stronger reductant and is not prone to side reactions with peptide functional groups. Check the pH of your TCEP stock solution as it may be acidic when brought up in solution depending on buffering conditions and needs to be brought to neutral pH prior to addition to the protein sample.
3. High concentrations of guanidine interfere with SDS-PAGE and must be removed before analysis; remove guanidine from SDS-PAGE aliquots by precipitation with nine volumes of cold EtOH as described in Subheading 3.5; treat with StrataClean resin.
4. In general, the most abundant precursor ions are the +2 ions, though peptides containing histidine residues are more often

observed as +3 ions. For SRM assays, selection of specific fragment ions (also called reporter ions or transitions) is required for development of the instrument method. For PRM assays, all fragment ions generated within the wide scan range of the method are observed, and selection of specific fragments used for quantification is done during post hoc data analysis.

5. For SDS-PAGE utilized with in-gel digestion, we recommend commercial gels, running buffer, and staining and destaining solutions to minimize contamination by keratins and other proteins originating from the skin, hair, and dust.
6. The majority of proteomic database search algorithms are currently limited to files acquired using data-dependent acquisition. DIA methods offer the distinct advantage of acquiring MS^2 data for all observed precursors, rather than only the most abundant, but data processing is more complex and time consuming. Skyline is a recommended tool for interpretation of DIA data, which can be streamlined using publicly available or in-house spectral libraries.
7. Mass spectrometers equipped with high-resolution detectors are well suited for the identification of PTMs, particularly those involving sulfur and oxygen, as one sulfur and two oxygens have the same nominal mass. Mass tolerances for database searches using high-resolution detection are typically set at 10–15 ppm.
8. The switch-tagging approach described here results in carbamidomethylation of DTT-reducible cysteines, including those previously in states including disulfide (RSSR or RSSR′), glutathionyl (RSSG), perthiol (RSSH), persulfide ($RSSS_nR$), sulfenic acid (RSOH), and nitrosothiol (RSNO). The lability of these cysteine oxoforms renders them extremely difficult to directly detect by MS, and DTT reducibility is used as a proxy for the in vivo reductase activity of thioredoxins, glutaredoxins, glutathione, and others on these oxidized cysteine targets. Notably, many irreversibly modified cysteine forms, like sulfinic acid (RSO_2H), sulfonic acid (RSO_3H), and the beta-elimination product dehydroalanine, are readily detected using MS techniques.
9. Artifactual oxidation of thiol residues may result during sample handling [54]. Efforts to minimize the generation of these artifacts include, for example, initial NEM alkylation of unmodified thiol residues and the incorporation of the appropriate controls.

Conflict of Interest

A.D. received funds from the National Blood Foundation. Though unrelated to the contents of the manuscript, the authors disclose that A.D. and K.C.H. are part of Endura LLC, and A.D. is a consultant for New Health Sciences.

References

1. Bantscheff M, Schirle M, Sweetman G et al (2007) Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem 389:1017–1031
2. Bachi A, Dalle-Donne I, Scaloni A (2013) Redox proteomics: chemical principles, methodological approaches and biological/biomedical promises. Chem Rev 113:596–698. doi:10.1021/cr300073p
3. Stelzl U, Worm U, Lalowski M et al (2005) A human protein-protein interaction network: a resource for annotating the proteome. Cell 122:957–968
4. Dengjel J, Kratchmarova I, Blagoev B (2010) Mapping protein-protein interactions by quantitative proteomics. Methods Mol Biol Clifton NJ 658:267–278
5. D'Alessandro A, Rinalducci S, Zolla L (2011) Redox proteomics and drug development. J Proteomics 74:2575–2595
6. Uhlén M, Fagerberg L, Hallström BM et al (2015) Tissue-based map of the human proteome. Science 347:1260419
7. Kim M-S, Pinto SM, Getnet D et al (2014) A draft map of the human proteome. Nature 509:575–581
8. Rual J-F, Venkatesan K, Hao T et al (2005) Towards a proteome-scale map of the human protein–protein interaction network. Nature 437:1173–1178
9. Wilhelm M, Schlegl J, Hahne H et al (2014) Mass-spectrometry-based draft of the human proteome. Nature 509:582–587
10. Yates JR, Ruse CI, Nakorchevsky A (2009) Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng 11:49–79
11. Tiselius A (1937) Electrophoresis of serum globulin. I. Biochem J 31:313–317
12. Liumbruno G, D'Alessandro A, Grazzini G, Zolla L (2010) Blood-related proteomics. J Proteomics 73:483–507
13. Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1:845–867
14. Nanjappa V, Thomas JK, Marimuthu A et al (2014) Plasma Proteome Database as a resource for proteomics research: 2014 update. Nucleic Acids Res 42:D959–D965
15. Wilson MC, Trakarnsanga K, Heesom KJ et al (2016) Comparison of the proteome of adult and cord erythroid cells, and changes in the proteome following reticulocyte maturation. Mol Cell Proteomics. doi:10.1074/mcp.M115.057315
16. D'Alessandro A, Righetti PG, Zolla L (2010) The red blood cell proteome and interactome: an update. J Proteome Res 9:144–163
17. Burkhart JM, Vaudel M, Gambaryan S et al (2012) The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways. Blood 120:e73–e82
18. Cristea IM, Gaskell SJ, Whetton AD (2004) Proteomics techniques and their application to hematology. Blood 103:3624–3634
19. Lea P, Keystone E, Mudumba S et al (2011) Advantages of multiplex proteomics in clinical immunology: the case of rheumatoid arthritis: novel IgXPLEX™: planar microarray diagnosis. Clin Rev Allergy Immunol 41:20–35
20. D'Alessandro A, Liumbruno G, Grazzini G et al (2010) Umbilical cord blood stem cells: towards a proteomic approach. J Proteomics 73:468–482
21. Liumbruno G, D'Alessandro A, Grazzini G et al (2010) How has proteomics informed transfusion biology so far? Crit Rev Oncol Hematol 76:153–172
22. D'Alessandro A, Kriebardis AG, Rinalducci S et al (2015) An update on red blood cell storage lesions, as gleaned through biochemistry and omics technologies. Transfusion 55:205–219
23. Goodman SR, Kurdia A, Ammann L et al (2007) The human red blood cell proteome and interactome. Exp Biol Med Maywood NJ 232:1391–1408
24. Guest PC, Guest FL, Martins-de Souza D (2015) Making sense of blood-based proteomics and metabolomics in psychiatric research. Int J Neuropsychopharmacol Off Sci J Coll Int Neuropsychopharmacol CINP. doi:10.1093/ijnp/pyv138
25. Macaulay IC, Carr P, Gusnanto A et al (2005) Platelet genomics and proteomics in human health and disease. J Clin Invest 115:3370–3377
26. Geyer PE, Kulak NA, Pichler G (2016) Plasma proteome profiling to assess human health and disease. Cell Syst 2:185–195
27. Liotta LA, Ferrari M, Petricoin E (2003) Clinical proteomics: written in blood. Nature 425:905
28. Bosman GJCGM (2016) The involvement of erythrocyte metabolism in organismal homeostasis in health and disease. Proteomics Clin Appl. doi:10.1002/prca.201500129
29. D'Alessandro A, Dzieciatkowska M, Hill RC et al (2016) Supernatant protein biomarkers of

red blood cell storage hemolysis as determined through an absolute quantification proteomics technology. Transfusion. doi:10.1111/trf.13483

30. Dzieciatkowska M, D'Alessandro A et al (2015) Plasma QconCATs reveal a gender-specific proteomic signature in apheresis platelet plasma supernatants. J Proteomics 120:1–6
31. Baez NOD, Reisz JA, Furdui CM (2015) Mass spectrometry in studies of protein thiol chemistry and signaling: opportunities and caveats. Free Radic Biol Med 80:191–211
32. Wood ST, Long DL, Reisz JA et al (2016) Cysteine-mediated redox regulation of cell signaling in chondrocytes stimulated with fibronectin fragments. Arthritis Rheumatol Hoboken NJ 68:117–126
33. Wither M, Dzieciatkowska M, Nemkov T et al (2016) Hemoglobin oxidation at functional amino acid residues during routine storage of red blood cells. Transfusion 56:421–426
34. Julian CG, Subudhi AW et al (2014) Exploratory proteomic analysis of hypobaric hypoxia and acute mountain sickness in humans. J Appl Physiol (1985) 116:937–944
35. Vosseller K, Hansen KC, Chalkley RJ et al (2005) Quantitative analysis of both protein expression and serine/threonine post-translational modifications through stable isotope labeling with dithiothreitol. Proteomics 5:388–398
36. Hill RC, Calle EA, Dzieciatkowska M et al (2015) Quantification of extracellular matrix proteins from a rat lung scaffold to provide a molecular readout for tissue engineering. Mol Cell Proteomics 14:961–973
37. Brownridge PJ, Harman VM, Simpson DM et al (2012) Absolute multiplexed protein quantification using QconCAT technology. Methods Mol Biol Clifton 893:267–293
38. Johnson TD, Hill RC, Dzieciatkowska M et al (2016) Quantification of decellularized human myocardial matrix: a comparison of six patients. Proteomics Clin Appl 10:75–83
39. Pratt JM, Simpson DM, Doherty MK et al (2006) Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. Nat Protoc 1:1029–1043
40. Boja ES, Rodriguez H (2012) Mass spectrometry-based targeted quantitative proteomics: achieving sensitive and reproducible detection of proteins. Proteomics 12:1093–1110
41. Aebersold R (2003) Quantitative proteome analysis: methods and applications. J Infect Dis 187:S315–S320
42. Geiger T, Wisniewski JR, Cox J et al (2011) Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. Nat Protoc 6:147–157
43. Krüger M, Moser M, Ussar S et al (2008) SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell 134:353–364
44. Dzieciatkowska M, D'Alessandro A, Moore EE et al (2014) Lymph is not a plasma ultrafiltrate: a proteomic analysis of injured patients. Shock 42:485–498
45. Doerr A (2015) DIA mass spectrometry. Nat Methods 12:35–35. doi:10.1038/nmeth.3234
46. Egertson JD, Kuehn A, Merrihew GE et al (2013) Multiplexed MS/MS for improved data-independent acquisition. Nat Methods 10:744–746
47. MacLean B, Tomazela DM, Shulman N et al (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26:966–968
48. Holstein Sherwood CA, Gafken PR et al (2011) Collision energy optimization of b- and y-ions for multiple reaction monitoring mass spectrometry. J Proteome Res 10:231–240
49. Wühr M, Haas W, McAlister GC et al (2012) Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster. Anal Chem 84:9214–9221
50. Dzieciatkowska M, D'Alessandro A, Burke TA et al (2015) Proteomics of apheresis platelet supernatants during routine storage: Gender-related differences. J Proteomics 112:190–209
51. Wiśniewski JR, Zougman A, Nagaraj N et al (2009) Universal sample preparation method for proteome analysis. Nat Methods 6:359–362
52. MacCoss MJ, Wu CC, Matthews DE et al (2005) Measurement of the isotope enrichment of stable isotope-labeled proteins using high-resolution mass spectra of peptides. Anal Chem 77:7646–7653
53. Dzieciatkowska M, Hill R, Hansen KC (2014) GeLC-MS/MS analysis of complex protein mixtures. Methods Mol Biol Clifton 1156:53–66
54. Herbert B, Hopwood F, Oxley D et al (2003) Beta-elimination: an unexpected artefact in proteome analysis. Proteomics 3:826–831

Chapter 25

SWATH Mass Spectrometry for Proteomics of Non-Depleted Plasma

Christoph Krisp and Mark P. Molloy

Abstract

The limitations commonly observed in data-dependent acquisition (DDA) mass spectrometric investigation of non-depleted human plasma are mainly due to the large dynamic concentration range of protein expression. Less abundant proteins are usually masked by highly abundant proteins and are therefore difficult to reliably detect. Sequential window acquisition of all theoretical fragment-ion spectra (SWATH) mass spectrometry (MS), as a representative of data-independent acquisition (DIA) approaches, provides an opportunity to improve plasma-based biomarker discovery studies because this approach does not rely on precursor intensity for fragmentation selection but rather analyzes all precursors in specified mass ranges. Here, we describe a workflow for SWATH-MS-based analysis of non-depleted plasma including sample preparation, data acquisition, and statistical analysis.

Key words Plasma, SWATH, Label-free quantitation, Data-independent acquisition, Mass spectrometry

1 Introduction

In recent years there has been a strong proliferation in mass spectrometry (MS)-based plasma proteome investigations with objectives to identify disease-related marker proteins especially for the early detection of cancer and life-threatening cardiovascular diseases [1–6]. However, the vast concentration ranges of plasma proteins, spanning from mg/mL to sub ng/mL concentrations, ensure that disease-specific biomarker discovery remains challenging. Indeed, many of the proposed biomarker proteins are liver-derived acute-phase proteins which are commonly associated with an inflammatory response but may not be specific enough for us to understand their link with disease processes. It is widely accepted that deeper proteome profiling will enable greater marker protein specificity, and considerable research efforts have been directed toward this task. Much effort has been devoted to eliminate the most highly concentrated proteins from plasma, as this provides more ready access to proteins in lower concentrations [7].

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_25,

However, those depletion techniques commonly introduce sample-sample variability due to incomplete depletion and co-depletion of bound proteins, which are likely to be those of lower abundance. The bias introduced by those techniques may outweigh biological variance, causing misinterpretation of findings [8].

Novel developments in data-independent acquisition (DIA) mass spectrometry may address some of the limitations of data-dependent acquisition (DDA) mass spectrometry in plasma. DIA or SWATH (sequential window acquisition of all theoretical fragment-ion spectra) does not rely on precursor ion intensity to initiate MS/MS; therefore, more lowly abundant proteins will be fragmented with this approach, providing an opportunity to use them for biomarker discovery. This is achieved by more thoroughly dissecting the MS1 precursor space into small m/z windows and generating mixed MS/MS spectra of all precursors present in a given MS1 m/z window at a given chromatographic retention time [9]. SWATH acquisition has advanced from using 32 × 25 m/z overlapping windows to cover the 400–1200 m/z MS1 window to utilizing variable windows (vW) of narrower or wider width dependent upon the predicted density of peptide precursors in a given MS1 m/z range [10]. This resulted in deeper proteome coverage, especially due to improvements in the tryptic peptide-rich region between 600 and 800 m/z. Using such an approach, it is possible to reliably quantitate more than 250 proteins from non-depleted plasma in a systematic and reproducible manner [11].

2 Materials

All solutions are prepared with ultrapure water (purification of deionized water to achieve a sensitivity of 18 MΩ/cm at 25 °C) and HPLC or analytical grade reagents. All solutions are stored at room temperature unless indicated otherwise.

2.1 Sample Preparation

1. Denaturing buffer: 100 mM triethylammonium bicarbonate (TEAB) and 1% (w/w) Sodium deoxycholate (DOC), pH 7.8–8.2; transfer 1 mL of a 1 M stock solution TEAB and 1 g DOC in a 100 mL shot bottle and make up to 100 mL with water.
2. Reducing stock solution: For a 1 M dithiothreitol (DTT) solution, transfer 154.2 mg (1 mmol) into a 1.5 mL sample tube and make up to 1 mL with DOC buffer. Prepare 30–50 μL aliquots and store for up to 6 month at −20 °C. Alkylating stock solution: For a 0.5 M iodoacetamide (IAA) solution, take 92.5 mg (0.5 mmol) IAA and adjust to 1 mL with DOC buffer. Prepare 20–50 μL aliquots and store at −20 °C for up to 2 months when kept in the dark (IAA is light sensitive).

2.2 LC-MS/MS

1. Equilibrating buffer: 2% acetonitrile (ACN) and 0.1% formic acid (FA). Add to 979 mL water 20 mL ACN and 1 mL FA.
2. Eluting buffer: 99.9% ACN and 0.1% FA. Add to 999 mL ACN 1 mL FA.

2.3 Software

1. MS control software: Analyst version 1.7.
2. Data base searches with ProteinPilot version 5.0 for assay library generation.
3. SWATH data extraction with PeakView version 2.1 with SWATH MicroApp version 2.0.

3 Methods

3.1 Plasma Storage

After plasma is obtained from whole blood by removal of platelets, cell debris, and hemoglobin, store samples at −80 °C until used.

3.2 Protein Assay

1. Thaw plasma on wet ice and divide into 50–100 µL aliquots. Take one aliquot and store the remaining at −80 °C
2. Take 20 µL of plasma from the aliquot. At this stage a protein precipitation can be performed to purify proteins and remove potentially interfering macromolecules (*see* **Note 1**). Depending on method either dilute crude plasma sample 1:10 with DOC buffer or resuspend protein pellet in 200 µL DOC buffer.
3. Perform an additional 1:10 dilution with DOC buffer (10 µL plasma and 90 µL DOC buffer) and follow protein assay kit manual to estimate protein concentration. Commonly, plasma protein concentrations range between 50 and 100 mg/mL (undiluted).

3.3 Protein Digestion

1. Take desired protein concentration from the first 1:10 dilution, here 100 µg, and adjust the sample volume to 100 µL with DOC buffer.
2. Add 1 µL of 1 M DTT stock and incubate for 30 min at 60 °C. After samples are cooled down, add 4 µL of 0.5 M IAA stock and incubate for 30 min at 37 °C in the dark.
3. Add trypsin in a protein to enzyme ratio of 20–50:1 and incubate over night at 37 °C.
4. After enzymatic digestion, acidify the sample by adding 1 µL FA to precipitate DOC. Spin at 14,000 × *g* for 5 min and transfer supernatant into a new sample tube. Lyophilize the sample in a vacuum concentrator and store at −20 °C until further use (*see* **Note 2**).

3.4 Data-Dependent Acquisition for Assay Library Generation

DDA can be performed at any given high-resolution mass spectrometer; however, sample condition and LC conditions should be comparable to avoid retention time variation.

1. Perform conventional top N data-dependent acquisition of the samples or representatives of the samples (pool or a subset) which will be later investigated by SWATH acquisition. Repetitive analysis can help improve peptide detection and increase final library size, but other strategies can also be used (*see* **Note 3**).
2. Perform data base search combining all DDA LC-MS/MS acquisitions to build a comprehensive assay library.
3. If using the Paragon algorithm in ProteinPilot, the resulting group file can readily be used as an assay library for PeakView. If using other search engines or other SWATH processing software, follow instruction for import usually available on tutorial websites for specific software.

3.5 Data-Independent Mass Spectrometry

The following procedures are specific for SWATH acquisition using SCIEX 5600 and 6600 TripleTOF mass spectrometers, but with adjustment to the sample acquisition methods, data-independent mass spectrometry can also be performed on other time of flight (ToF) or high-resolution ion-trapping instruments capable of data-independent data acquisition (*see* **Note 4**).

3.5.1 Variable Window SWATH-MS

1. Open a DDA file representing precursor elution profile for sample of interest in PeakView. Generate a summed MS1 of the region in which proteins are eluting and export the summed MS1 spectrum as "Data as text."
2. Open the Excel template "SWATH® Variable Window Calculator 1.0" (downloadable here: http:// sciex.com/ software-downloads-x2110). Specify number of windows, MS1 mass range, and collision energy (CE) spread. Here, 100 variable windows from 400 to 1250 m/z with a CE spread of 5 V are used (*see* **Note 5**). Paste m/z and intensity information from PeakView export into the Input tab, wait until processing has been completed, and save output as a text file (tab delimited). *See* Fig. 1 for typical variable window distribution for a plasma sample.
3. Open Analyst software on the computer attached to the mass spectrometer and create a new SWATH variable window method by importing the text file in the manual options in the "create SWATH Exp" dialoged box. Specify method acquisition length and select LC methods (*see* **Note 6**).
4. Acquire SWATH data of all samples (randomized) either as digestion replicates (at least three) or biological replicates only, depending on sample cohort.

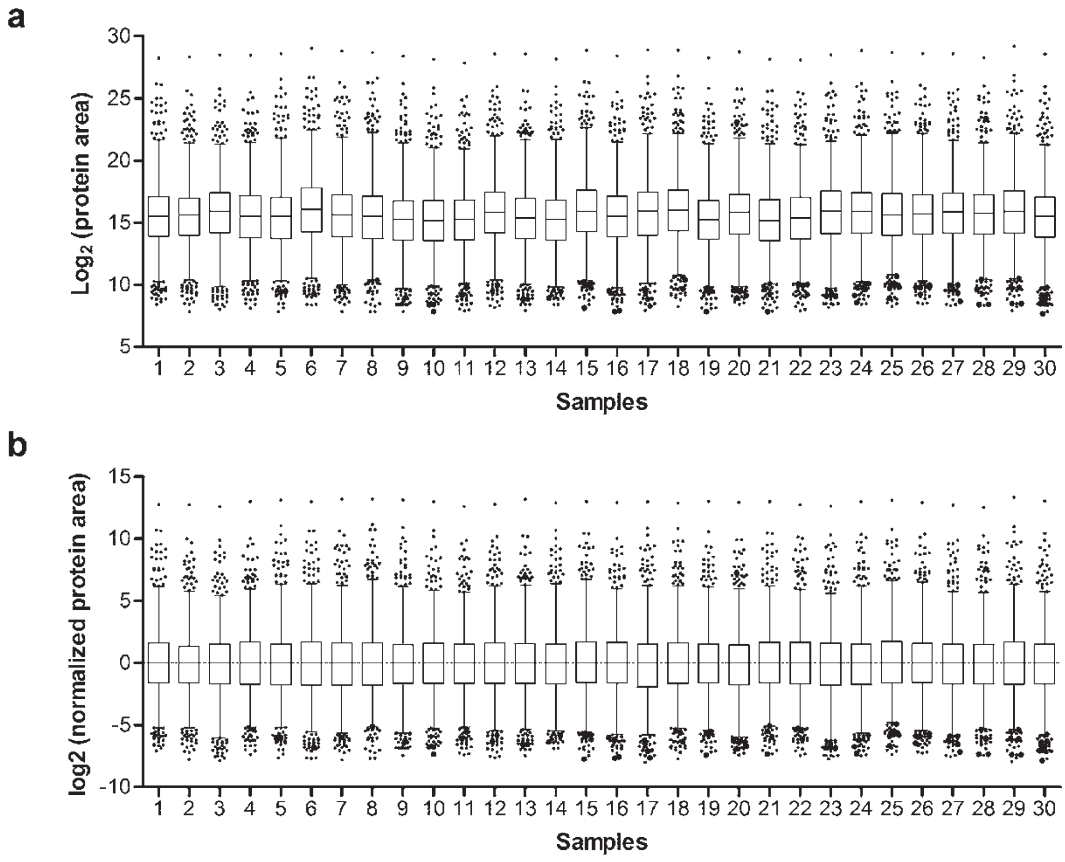


Fig. 1 Box plots of typical log2 transformed protein area distribution of SWATH-MS analysis of plasma samples prior to normalization (**a**) and after median log2 protein area normalization (**b**)

3.5.2 Data Extraction

Multiple software are available to extract quantitative information from SWATH data sets (e.g., PeakView (license required), OpenSWATH (freeware), Skyline (freeware), or Spectronaut (license required)). Here, methods for PeakView are described.

1. Open PeakView and, if importing assay library for the first time, select a .group file generated as described in Subheading 3.4. Specify the number of proteins to import. Exclude shared peptides to avoid false interpretation from differentially regulated proteins with shared peptides.
2. Select SWATH data files for processing. Under processing settings, select maximum number of peptides 100 and number of transitions per peptide 6 with peptide confidence ≥0.99 and select 75 ppm as fragment mass tolerance. Allow an extraction FDR of less than 1% and an extraction window of 5–10 min depending on retention time stability.
3. Check expected retention time from assay library with actual retention time in SWATH experiments and adjust if necessary

with the PeakView built-in retention time alignment tool by selecting endogenous peptides from various time points across the elution time (at least three peptides).

4. The imported library can now be saved as a text file (optional) and if desired optimized (*see* **Note 7**).
5. Process data and export results.

3.6 Data Analysis and Statistics

For data analysis and statistics, the software tool Perseus (freeware) is used. For a detailed description of the software, visit http://www.coxdocs.org/doku.php?id=perseus:start. To identify protein expression differences, the protein area (summed area of peptides with ≤1% extraction FDR) is used.

1. Generate a text file listing all proteins with quantitative information for all samples (tab delimited). At this point peptides which do not have extraction FDR ≤ 1% in all samples or peptides with a score below a certain cutoff can be removed (*see* **Note 8**).
2. Open Perseus and import the text file and remove proteins with missing values (usually less than two to three proteins depending on library size).
3. Transform protein area raw data to log2 data and perform normalization by subtracting sample median (*see* **Note 9**). Check if median log2 areas align before continuing with further analysis (Fig. 2).
4. Perform pair-wise Student's *t*-test between the different sample types and filter for proteins with *p*-values ≤0.05 and depending on sample size (statistical power) appropriate fold

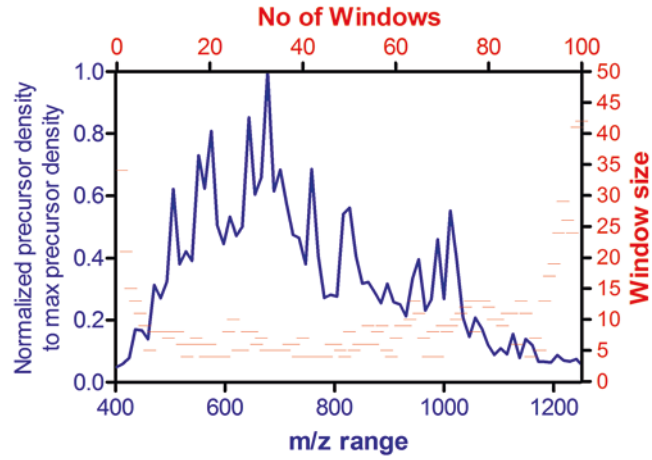


Fig. 2 Precursor ion density distribution of trypsin-digested human plasma (*blue*) over the mass range of 400–1250 *m/z* and the variable window size (*red*) calculated by the variable Window calculator

changes (e.g., > ±2-fold or > ±1.5-fold). Proteins which remain significant can then be assessed for biological relevance and used for further validation studies.

4 Notes

1. Plasma is a highly complex biological matrix containing proteins but also a variety of other biomolecules including small metabolites, hormones, vitamins, lipids, and large macromolecules such as circulating DNA/RNA. These biomolecules usually interfere with protein digestion efficiencies by masking trypsin cleavage sites, cysteine alkylation efficiencies caused by redox active substances, chromatographic retention, and peptide ionization. To reduce these effects and to improve mass spectrometric peptide detection, a chloroform/methanol precipitation following the protocol by Butler et al. [12] is effective in our hands.
2. As all label and label-free quantitation strategies, SWATH relies highly on reproducible sample preparation. Alkylation and digestion efficiency are crucial for reproducible quantitation. Therefore, prior to the acquisition of large data sets, the sample preparation protocol should be assessed. The full FDR report generated by ProteinPilot 5.0 when using the Paragon algorithm can be helpful in identifying sources of undesired peptide modification and allows for estimation of alkylation and digestion efficiency. If other search engines are used, peptide N-terminal as well as lysine, aspartic acid, and glutamic acid carbamidomethylation should be allowed as variable modification. As a guide, >85% digestion efficiency and >99% cysteine alkylation with <10% alkylation of other amino acids are acceptable.
3. Assay library generation for plasma samples can be quite challenging. Common 1D RP LC-MS/MS DDA approaches analyzing non-depleted human plasma soon reach the limit of peptide detectability and usually only succeed in identifying 200–300 proteins. Immunodepletion of the most abundant proteins is commonly performed and might improve assay library generation; however, because those highly abundant proteins frequently undergo interactions with less abundant proteins, co-depletion of those proteins is very likely, which could have dramatic consequences for SWATH quantitation. Additional chromatographic sample fractionation by the classical strong cation/anion exchange [13] or more recent high pH [14] RP prior to LC-MS/MS can be of great value. Especially high pH RP which delivers a somewhat orthogonal separation to low pH RP due to peptide hydrophobicity changes depending on the pH. If those techniques are not

available because the sample quantity is to low, the assay library can also be extended with data sets available in online repositories (e.g., Farrah et al. [15]) using an approach similar to the one described here [16].

4. Data-independent acquisition can be performed on several different mass spectrometers; however, some crucial factors should be considered. Important for quantitation is the cycle time; this includes the number of MS1 windows, time spent on each MS/MS collection per window, and if not using a reflectron ToF. The cycle time should not be longer than 1/10th of the average analyte peak width, for example, 3 s if average peak width is 30 s, to guarantee reliable quantitation. Using trap instruments, e.g., QExactive Plus or Orbitrap Fusion, fixed fill times for selected MS1 windows might result in severe over- or under-filling of the trap; therefore, MS1 window MS/MS analysis should be based on a fixed number of ions in the trap for accurate comparability between different samples. However, this will result in different time spent on the analysis of certain window and could negatively influence quantitation using instruments with slower duty cycle times.
5. Variable window SWATH acquisitions have been shown to be more effective than fixed window SWATH acquisitions when spanning a mass from 400 to 1250 m/z [10]. Variable windows allow for narrower investigation of m/z areas with dense precursor occurrence and broader windows in areas with only a few precursors, e.g., $m/z > 1000$. Smaller windows in highly populated precursor m/z ranges lead to less interfering signals and improved quantitation; however, variable window sizes must not be smaller than the lower limit defined in the software used for data extraction.
6. The gradient length for SWATH acquisition does not have to match the gradient length used for library generation. Since SWATH acquisition does not require the best possible peptide separation but rather higher precursor intensities, it is recommendable to use shorter gradients for SWATH and long gradients for assay library generation. However, the gradient development should be scalable to allow retention time adjustments by linear regression.
7. If using PeakView for SWATH data extraction, be cautious about product ion selection. The current algorithm aims for high specificity in preference to sensitivity; therefore, the algorithm favors product ions with larger m/z precursors. For peptides with 400–600 m/z, this does not have a large consequence, but for larger precursors, this could prevent detection as product ions <1000 m/z are commonly more intense in TOF-based MS/MS. Therefore, selecting peptides based on fragment intensity improves scores and FDRs during data extraction (Fig. 3). This selection, however, cannot be made easily in

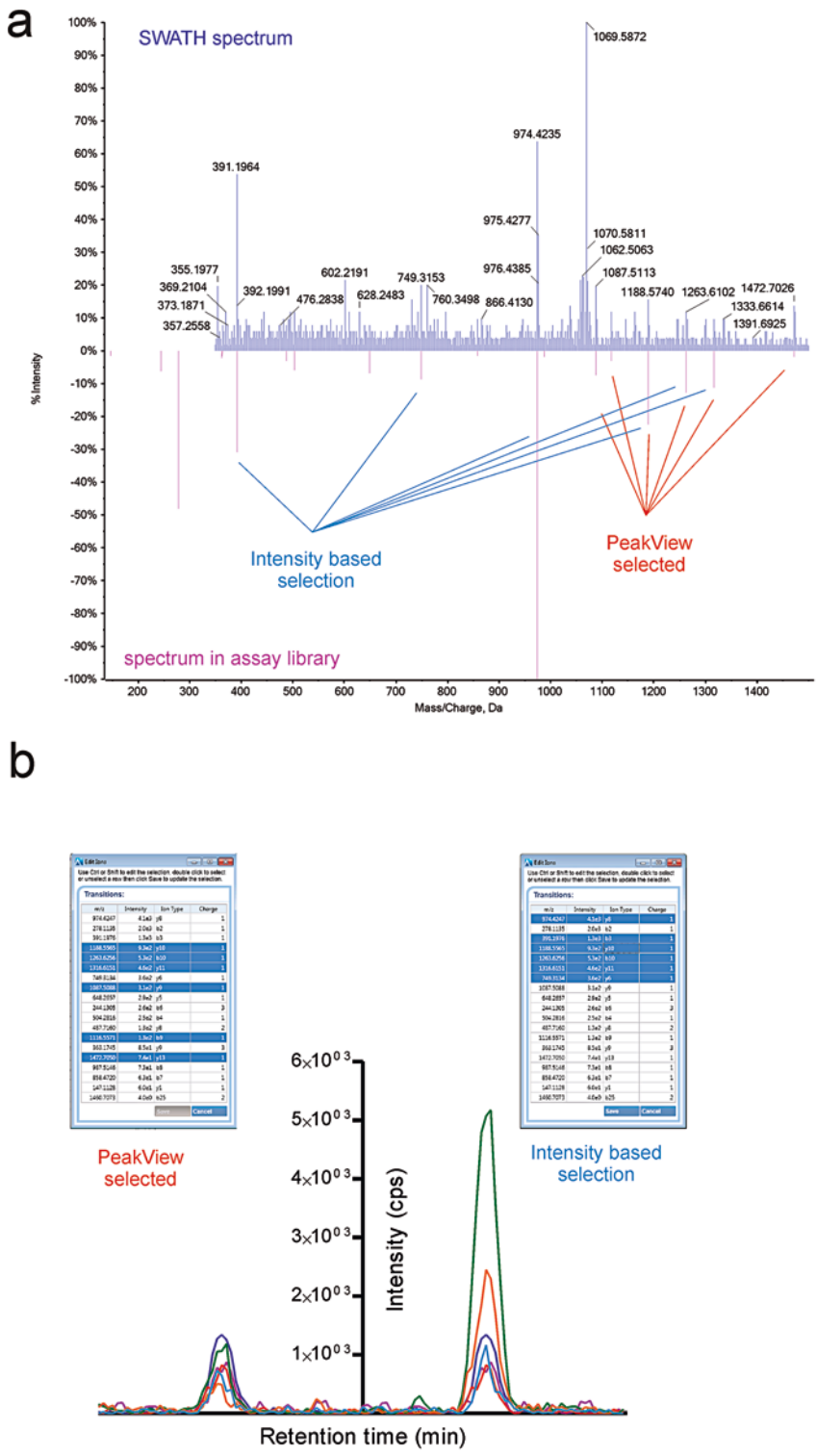


Fig. 3 Contrast of PeakView precursor (*red*) versus intensity-based (*light blue*) fragment ion selection for peptides with precursor $m/z > 900$ using the triply charged alpha-2-macroglobulin peptide YNILPEKEEFPFALGVQTLPQTCDEPK with 1055.19 m/z. Matching the fragments in the assay library (*pink*) with the SWATH spectrum (*blue*) demonstrate more intense fragment ions with m/z less than the precursor m/z are missed (**a**). Selecting the most intense ions from the assay library, the area under the curve for this peptide will increase and will improve the score and FDR calculated for this peptide (**b**)

PeakView software but requires either a script or manual curation of the assay library. If using Skyline this decision making is not necessary, since it always selects the most intense ions from the assay library.

8. To improve protein area reproducibility across biological replicates, low scoring peptides (negative score assigned by PeakView) and peptides with a certain number of samples in which it received an extraction FDR > 1% can be removed. Those peptides are commonly of low intensity or have interferences in one or more transitions. Especially in plasma, where the dynamic range spreads over several orders of magnitude, a low abundant protein may be readily interfered by a transition from a highly abundant protein, and this will have severe consequences and lead to false results.
9. For most SWATH data sets, it is recommended to perform post-processing normalization to correct for sample concentration differences. Several strategies have been used to normalize SWATH data, since SWATH data reflect the relative protein abundance level quite well. Normalizations which only adjust mean or median protein level are preferred to preserve the dynamic range within a sample. Therefore, valid strategies may be mean or median protein area, total area sum, or most likely ratio normalization.

Acknowledgments

This research was carried out at the Australian Proteome Analysis Facility supported by the Australian Government's National Collaborative Research Infrastructure Scheme (NCRIS).

References

1. Liu Y, Hüttenhain R, Collins B, Aebersold R (2013) Mass spectrometric protein maps for biomarker discovery and clinical research. Expert Rev Mol Diagn 13:811–825
2. Parker CE, Borchers CH (2014) Mass spectrometry based biomarker discovery, verification, and validation – quality assurance and control of protein biomarker assays. Mol Oncol 8:840–858
3. Pernemalm M, Lehtiö J (2014) Mass spectrometry-based plasma proteomics: state of the art and future outlook. Expert Rev Proteomics 11:431–448
4. Mesmin C, van Oostrum J, Domon B (2016) Complexity reduction of clinical samples for routine mass spectrometric analysis. Proteomics Clin Appl 10:315–322
5. Ray S, Reddy PJ, Jain R, Gollapalli K et al (2011) Proteomic technologies for the identification of disease biomarkers in serum: advances and challenges ahead. Proteomics 11:2139–2161
6. Ossola R, Schiess R, Picotti P, Rinner O et al (2011) Biomarker validation in blood specimens by selected reaction monitoring mass spectrometry of N-glycosites. In: Simpson JR, Greening WD (eds) Serum/plasma proteomics: methods and protocols. Humana Press, Totowa, NJ, pp 179–194
7. Polaskova V, Kapur A, Khan A, Molloy MP, Baker MS (2010) High-abundance protein depletion: comparison of methods for human plasma biomarker discovery. Electrophoresis 31:471–482

8. Patel BB, Barrero CA, Braverman A, Kim PD et al (2012) Assessment of two immunodepletion methods: off-target effects and variations in Immunodepletion efficiency may confound plasma proteomics. J Proteome Res 11:5947–5958
9. Gillet LC, Navarro P, Tate S, Röst H et al (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics 11:O111.016717
10. Zhang Y, Bilbao A, Bruderer T, Luban J et al (2015) The use of variable Q1 isolation windows improves selectivity in LC–SWATH–MS acquisition. J Proteome Res 14:4359–4371
11. Liu Y, Buil A, Collins BC, Gillet LC et al (2015) Quantitative variability of 342 plasma proteins in a human twin population. Mol Syst Biol 11:786
12. Butler AM, Blatt H, Southgate H (1935) The solubility of the plasma proteins: II. Dependence on pH, temperature, and lipid content in concnetrated solutions of potassium phosphate and application to their separate precipitation. J Biol Chem 109:755–767
13. Chan KC, Issaq HJ (2013) Fractionation of peptides by strong cation-exchange liquid chromatography. In: Zhou M, Veenstra T (eds) Proteomics for biomarker discovery. Humana Press, Totowa, NJ, pp 311–315
14. Batth TS, Francavilla C, Olsen JV (2014) Offline high-pH reversed-phase fractionation for in-depth phosphoproteomics. J Proteome Res 13:6176–6186
15. Farrah T, Deutsch EW, Aebersold R (2011) Using the Human Plasma PeptideAtlas to study human plasma proteins. In: Simpson JR, Greening WD (eds) Serum/plasma proteomics: methods and protocols. Humana Press, Totowa, NJ, pp 349–374
16. Wu JX, Song X, Pascovici D, Zaw T et al (2016) SWATH mass spectrometry performance using extended peptide MS/MS assay libraries. Mol Cell Proteomics 15(7):2501–2514

Chapter 26

Shotgun and Targeted Plasma Proteomics to Predict Prognosis of Non-Small Cell Lung Cancer

Qing-Run Li, Yan-Sheng Liu, and Rong Zeng

Abstract

Lung cancer is the leading cause of cancer deaths worldwide. Clinically, the treatment of non-small cell lung cancer (NSCLC) can be improved by the early detection and risk screening among population. To meet this need, here we describe in detail a shotgun following the targeted proteomics workflow that we previously applied for human plasma analysis, which involves (1) the application of extensive peptide-level fractionation coupled with label-free quantitative proteomics for the discovery of plasma biomarker candidates for lung cancer and (2) the usage of the multiple reaction monitoring (MRM) assays for the follow-up validations in the verification phase. The workflow features simplicity, low cost, high transferability, high robustness, and flexibility with specific instrumental settings.

Key words Non-small cell lung cancer, Plasma proteomics, Shotgun proteomics, Targeted proteomics, Diagnostic biomarker

1 Introduction

Lung cancer is the most frequent cancer in the world, in terms of both incidence and mortality. Non-small cell lung cancer (NSCLC) accounts for 80–85% of lung cancer with an overall 5-year survival rate less than 14% [1]. Specifically, the 5-year survival rate is barely 3–7% for stage IIIB and is less than 1% for stage IV disease [2]. Therefore, new diagnostics are urgently needed to detect early stage lung cancer.

Currently, the disease-driven proteomics based on mass spectrometry has been introduced to the discovery of both histological and serological biomarkers [3]. We previously endeavored to use shotgun proteomics to directly discern the differential serum proteome patterns associated with NSCLC diseases, to understand the regulated serum proteins in terms of their biological relevance [4]. In blood-based proteomics, the major difficulty in validating biomarkers is the limited availability of sandwich enzyme-linked immunosorbent assay (ELISA) kits for novel candidates and the restricted

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_26,

possibility to multiplex assays. Additionally, the shotgun MS analysis is not always reproducible to multiple clinical samples. These factors have led to the development of targeted proteomics, which is based on the technology of multiple reaction monitoring (MRM). Better sensitivity and dynamic range can be achieved by MRM than shotgun proteomics, which are crucial for clinical validation [3]. Herein, we described in detail this integrative proteomic method, i.e., an extensive fractionation on peptide level to profile the albumin-depleted plasma proteome for discovery proteomics, following the MRM to targeted measure the protein concentrations (0.09–0.39 μg/mL range) in clinical plasma samples with better selectivity, sensitivity, and dynamic range for clinical validation.

2 Materials

Prepare all solutions using ultrapure water with 18 MΩ cm resistance at 25 °C (Milli-Q, Millipore), unless specified otherwise. All reagents should be of analytical grade or higher.

2.1 Sample Preparation and Protein Digestion

1. Equipment: microtube centrifuge, and 1.5 mL Eppendorf tube.
2. Human blood sample (50 μL).
3. Delipidation: 0.22 μm filter to remove lipids.
4. Albumin depletion: Multiple Affinity Removal System spin cartridge (e.g., #5188–8825, Agilent Technologies following the instruction of the vender).
5. Lysis buffer: 8 M urea, 4% CHAPS, 40 mM Tris-base, 65 mM DTT.
6. Protease inhibitor cocktail.
7. 10 mM dithiothreitol (DTT).
8. 50 mM iodoacetamide (IAA).
9. 50 mM ammonium bicarbonate buffer (pH 8.5).
10. Resolving buffer: 100 mM NaCl, 10 mM HEPES, pH 7.4.
11. Pre-chilled ethanol.
12. Pre-cold acetone/ethanol (1:1, v/v) with 0.1% acetic acid.
13. Sequencing-grade modified trypsin.

2.2 Chromatography Fractionation

1. Equipment: high-performance liquid chromatography system.
2. Bi-phase integrated column: a strong cation column (SCX, 320 μm i.d., 50 mm length, Column Technology Inc., CA) connected with a reversed-phase (RP) chromatography column (150 μm i.d., 100 mm length, Column Technology Inc.).

3. Eleven pH steps: pH 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 7.0, 8.0, 8.5 (configured by 5 mM citric acid adjust by NH_4OH (ACS Reagent grade, MW = 35.04, d = 0.89 g/mL, and 28–30%)).
4. 0.1% formic acid (v/v) aqueous (A).
5. 0.1% formic acid (v/v) acetonitrile (B).

2.3 Synthetic Peptides

1. Stable isotopically labeled amino acids: [$^{13}C_6$] leucine and [$^{13}C_3$] alanine (>98 atom percent isotopic enrichment).
2. Six isotopic peptides derived from the target proteins were synthesized with uniform [$^{13}C_6$] leucine and [$^{13}C_3$] alanine using standard Fmoc chemistry. Unlabeled [^{12}C] forms of each peptide were also synthesized. All synthetic peptides were purified by manufacturer to >95% purity with given amounts by the vendor.
3. Lyophilized peptides of known amount were resolubilized in 15% acetonitrile/0.1% formic acid before used.

2.4 Mass Spectrometry

1. Mass spectrometer with tandem MS capabilities (e.g., LTQ-Orbitrap Velos, Thermo Fisher Scientific, and TripleTOF 5600 system, SCIEX).
2. Triple quadrupole instruments (e.g., 6410 QQQ mass spectrometer, Agilent Technologies, and TSQ Vantage, Thermo Fisher Scientific).
3. MS/MS data analysis software (e.g., SEQUEST [5], Mascot [6], or MaxQuant [7]).
4. MRM data processing software (e.g., MassHunter Qualitative software (Agilent)).
5. Software for differential analysis (spreadsheet program with statistical analysis package or dedicated programs for this purpose, e.g., Excel or R software (http://www.r-project.org/) and SPSS.

3 Methods

3.1 Collection of Plasma Samples

1. Mix 50 μL human blood sample (*see* **Note 1**), 250 μL resolving buffer (100 mM NaCl, 10 mM HEPES, pH 7.4) in 1.5 mL Eppendorf tube.
2. Immediate centrifuge the diluted blood sample at 3000 × *g* for 20 min (4 °C).
3. Collect the supernatant, transfer to a polypropylene capped tube in 200 μL aliquots (*see* **Note 2**).

3.2 Delipidation and Albumin Depletion

1. Dilute 50 μL human blood sample with 250 μL dilution buffer (100 mM NaCl, 10 mM HEPES, pH 7.4).

2. Centrifuge at 10,000 × *g* for 30 min (4 °C) through a 0.22 μm filter to remove lipids.
3. For shotgun proteomics, samples were depleted of human albumin and IgG using a commercial column following manufacturer's protocol (e.g., MARS, Agilent Technologies).
4. Alternatively, to increase the sample throughput and reduce the experimental cost, precipitate 260 μL delipidated plasma with 180 μL pre-chilled ethanol, incubate for 1 h at 4 °C with gentle mixing, and centrifuge at 16,000 × *g* for 45 min.
5. Collect, lyophilize, and resuspend the albumin-depleted pellets in 100 μL lysis buffer with protease inhibitor cocktail.

3.3 In-Solution Protein Digestion

1. Incubate protein mixtures with 10 mM DTT at 37 °C for 2.5 h, and carbamidomethylated with 20 mM IAA for 45 min at room temperature in darkness.
2. Precipitate protein solutions by adding five volumes of pre-cold acetone/ethanol (1:1, v/v) with 0.1% acetic acid at −20 °C for 12 h. [4].
3. Resuspend the protein pellets in 50 mM ammonium bicarbonate buffer, and incubate with trypsin (50–25:1) for 16–20 h at 37 °C.
4. Lyophilize the digested peptides and store at −80 °C for mass spectrometry analysis.
5. For MRM analysis, resuspend the crude plasma (without depletion) in lysis buffer (1:50, v/v) with protease inhibitor cocktail. Perform protein digestion as described **steps 1–4**.

3.4 Online Two-Dimensional MS/MS Analysis on LTQ-Orbitrap

1. Separate the peptide mixtures from each sample on a fully automated two-dimensional liquid chromatography (2-D LC) system [8], which including a strong cation exchange fractionation approach utilizing 11 online continuous pH gradients from pH 2.5 to 8.5, followed by reversed-phase (RP) chromatography.
2. The nano-RP-HPLC solvents used were (A) 0.1% formic acid (v/v) aqueous and (B) 0.1% formic acid (v/v) acetonitrile with a gradient of 5–35% of mobile phase B for 135 min at 300 nL/min.
3. All the data were acquired on a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific) using the "high-low" mode, which involves a full scan at high resolution (60,000 at m/z 400) by the Orbitrap mass analyzer following a data-dependent "top ten" ms/ms scan by the ion trap mass analyzer. The normalized collision energy (NCE) was 35.0. Dynamic Exclusion settings were as follows: repeat count, 1; repeat duration, 30 s; exclusion duration, 180 s.

3.5 Protein Identification

1. The BioWorks™ 3.2 software suite was used to generate the peak lists of all acquired MS/MS spectra (default parameters). They were then automatically searched against the Human International Protein Index protein sequence database (version 3.73, containing 89,652 proteins) using the SEQUEST version 2.7 (University of Washington, licensed to Thermo Fisher Scientific) searching program (*see* **Note 3**).
2. Search engineer parameters were used as follows: trypsin was designated as the protease; only one missed cleavage was allowed. Maximal mass tolerance was set as 500 p.p.m. (finally filtered by 10 p.p.m.) for the precursor ion and 0.5 Da for fragment ions. Carbamidomethylation (+57.0125 Da) was searched as a fixed modification on cysteine, and oxidation (+15.99492 Da) was set as a variable modification on methionine.
3. An in-house software BuildSummary was used to delete the redundant data [9, 10]. Proteins were identified with stringent criteria: thresholds for xcorr according to preliminary peptide-spectrum match (PSM) FDR < 1% and the protein FDR < 1% were applied for each pair of the shotgun samples. All accepted SEQUEST results were required a ΔCn score of at least 0.1 regardless of the charge state.
4. Apply statistical analysis methods to determine significantly regulated proteins (*see* **Note 4**).

3.6 MRM Assays Configuration (Fig. 1)

1. Resuspend the peptide digests from 0.1 μL crude plasma (for every sample) in 0.1% formic acid, spike with internal standard peptides, and separate by reversed-phase micro high-performance liquid chromatography on a 1200 HPLC system (Agilent Technologies).
2. Separation conditions were as follows (*see* **Note 5**): mobile phase flow rate of 1.5 μL/min with buffer A (0.1% formic acid) and buffer B (90% acetonitrile in 0.1% formic acid), on one C18 trap column (300 μm × 5 mm, Agilent Technologies) followed by an analytical C18 column (150 μm × 100 mm, Column Technology Inc., CA).
3. The dynamic MRM analysis was performed on, e.g., a 6410 QQQ mass spectrometer (Agilent Technologies). Data acquisition parameters were used as follows: the capillary voltage of 4000 V, drying gas of 300 °C at 3.0 L/min, and nebulizer gas of 18 psi. Fragmentor voltage and collision energy (CE) were optimized with infusion of each peptide standard following the MassHunter Optimizer (Agilent) protocol (*see* **Note 6**).
4. Monitor 2–4 most responsive transitions per peptide and acquired at unit resolution both in the first and third quadrupole (*see* **Note 7**).

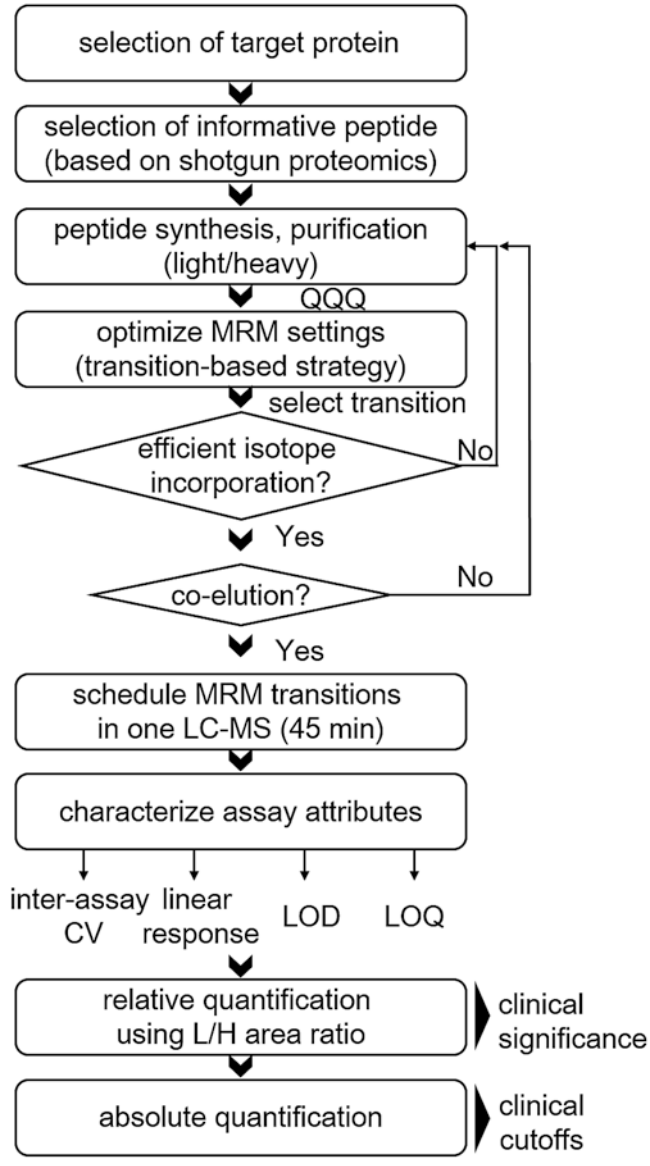


Fig. 1 MRM assays configuration. Fragmentor voltage and collision energy (CE) were optimized with infusion of each peptide standard following the MassHunter Optimizer protocol (Agilent). Specifically, we first iteratively changed the fragmentor voltage to maximize the precursor ion transmission at the highest SIM signal intensity. Once fragmentor voltage was decided and fixed, a MRM survey scan was executed in which all the appointed transitions were tested by ramping CE (5–50 V) with minimal default dwell time each (5 ms). The signals within each tracing channel were compared to determine the best CE. By this "transition-based strategy," best charge state, fragmentor value, CE, and dynamic dwell time were chosen and identically used in each [^{12}C]/[^{13}C] pair

5. In total, 42 scheduled MRM transitions were adopted for the three target proteins, with dynamic dwell time from 28 to 296 ms (dependent on the number of concurrent transitions). The dwell time of MRM was 50 ms, in which only the best transition pair for each peptide were monitored. In all experiments, cycle times did not exceed 1.2 s, and a minimum of 20 data points were collected per peak.
6. MRM data analysis was performed using MassHunter Qualitative software (Agilent). $[^{12}C]/[^{13}C]$ peak area was recorded and manually inspected for the strict co-eluting behavior, taking all the $[^{13}C]$ transitions as reference (*see* **Note 8**).
7. For each protein, the verifications of relative abundances in different samples were achieved by all the transition pairs. Choose the best transition which harbored best linearity response and also high abundance for absolute quantification (*see* **Note 9**) and thus increased the quantitative accuracy.
8. Area ratios were used to calculate endogenous concentrations of target proteins in sera, by formula fitted the linearity curve. Inter-assay coefficient of variation (CV) was evaluated by five separated processing and MRM replicates of a mixed sample, and limit of detection (LOD) was defined as the concentrations at which the S/N of the analyte is equal to 3 [11] (*see* **Note 10**).

3.7 Statistics and Bioinformatics

1. Hierarchical clustering analysis (HCA) was performed with cluster 3.0 and was visualized with TreeView using quantile normalized values of logarithmic transformed spectral counts, i.e., by transformation of the $\log_2$ (spectral counts +1) for each plasma protein in every sample.
2. In the receiver operating characteristic (ROC) analysis, a logistic regression model was used to assess the outcome of the biomarker panel (*see* **Note 11**). This modeling was performed in SPSS (v13.0).
3. Setting up the cutoffs. The predicted probabilities of occurrence of events (e.g., good or poor prognosis) were incorporated as predictors in the ROC plot, and area under the ROC curve (AUC) was calculated to estimate the power of the panel [12]. In the ROC curve, the distance from ideal (DFI) is defined as the distance from the ideal point (100% sensitivity, 100% specificity) and is calculated as $((1\text{-sensitivity})^2 + (1\text{-specificity})^2)^{1/2}$. The best cutoff can be determined from the point at which the DFI is minimal [13]. Kaplan-Meier survival analysis was used to compare the cumulative probability of recurrence or survival at any specific time.
4. Classification and regression tree (CART) analysis [14] was used to establish the decision rules and to classify SCC patients with a poor or good prognosis using the programming in R

tree package. The CART used a Gini splitting algorithm that favored even splits [15]. As a pruning rule, the splitting continues until the terminal nodes are fewer than six observations.

4 Notes

1. Blood samples from newly diagnosed patients were obtained before any treatment in the Department of Thoracic Surgery at Shanghai Cancer Hospital. The inclusion criteria of subjects consisted of confirmed diagnosis of NSCLC, no distant metastasis, or other surgical contraindications. The written informed consents should be provided by all the individuals involved in the study. Anonymous samples from the patients were randomly selected.
2. Time interval between processing and freezing should be finished in no more than 2 h for each sample. Samples which was thawed more than twice before analysis need to be prepared freshly. It is important to minimize freeze-thaw cycles because this could be detrimental to many serum components and lead to protein degradation [16–18].
3. Freely available software like MaxQuant [7] (www.maxquant.org) and Pattern Lab for proteomics package [19] (http://patternlabforproteomics.org/) are also suggested.
4. In our previous work [20], we used the following stringent criteria: (1) an LSPAD (localized statistics of protein abundance distribution [21]), $p < 0.05$ in all the paired samples; (2) an average of $p < 0.01$; (3) a fold change of >1.3 between the inner groups (paired); and (4) at least 20 MS/MS spectra had to be assigned to the candidate due to the sensitivity issue of spectral counting.
5. The fast, binary gradient was optimized according to the separation performance observed from synthesized peptides, consisting of 3–17% B in 2.5 min, 17–23% B in 25 min, 23–40% B in 5 min, 40–100% B in 5 min, and at 100% B for 2.5 min.
6. Specifically, we first iteratively changed the fragmentor voltage to maximize the precursor ion transmission at the highest SIM signal intensity. Once fragmentor voltage was decided and fixed, a MRM survey scan was executed in which all the appointed transitions were tested by ramping CE (5–50 V) with minimal default dwell time each (5 ms). The signals within each tracing channel were compared to determine the best CE. By this "transition-based optimization strategy," best charge state, fragmentor value, CE, and dynamic dwell time were chosen and identically used in each [^{12}C]/[^{13}C] pair.

7. In general, transitions were chosen based upon their signal abundance and relative higher mass-to-charge ratio (m/z).
8. The amounts of spiked [^{13}C] peptides for each peptide were different, so that most [^{13}C]/[^{12}C] area ratios could be kept below 20-fold change.
9. The assay linearity was characterized by diluting a tryptic digest of one plasma sample which was incorporated with light peptides to generate a range of endogenous analyte concentrations spanning 3^8-fold (i.e., a threefold dilution was used for the standard curve, while up to eight dilution steps were adopted) concentration range (modified from Michael et al. [22]). Spiked with [^{13}C] peptide mixture, each dilution was analyzed by LC-MRM four times, with blank solvent injections between samples. The amount of [^{12}C] peptide was measured by the concentration point at which the area ratios of [^{12}C]/[^{13}C] were closest to one (1.329 for C4BP, 0.922 for LRG1, and 1.213 for SAA).
10. Since background noise is extremely low in MRM mode and the accurate measurement of co-eluting noise in transitions was not feasible, the limit of quantitation (LOQ) was empirically determined as the lowest analyte concentration which can keep the acceptable linearity (Pearson correlation, $R > 0.99$) and can be measured with <20% CV [22].
11. Incorporate all of the candidates in the panel along with probabilities, i.e., using block entry of variables.

References

1. Jemal A, Siegel R, Ward E et al (2009) Cancer statistics, 2009. CA Cancer J Clin 59(4):225–249
2. O'Byrne KJ, Danson S, Dunlop D et al (2007) Combination therapy with gefitinib and rofecoxib in patients with platinum-pretreated relapsed non small-cell lung cancer. J Clin Oncol 25(22):3266–3273
3. Huttenhain R, Malmstrom J, Picotti P et al (2009) Perspectives of targeted mass spectrometry for protein biomarker verification. Curr Opin Chem Biol 13(5–6):518–525
4. Liu YS, Li C, Xing Z et al (2010) Proteomic mining in the dysplastic liver of WHV/c-myc mice—insights and indicators for early hepatocarcinogenesis. FEBS J 277(19):4039–4053
5. Eng J, McCormack A, Yates J (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(11):976–989
6. Perkins DN, Pappin DJ, Creasy DM et al (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20(18): 3551–3567
7. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372
8. Zhou H, Dai J, Sheng QH et al (2007) A fully automated 2-D LC-MS method utilizing online continuous pH and RP gradients for global proteome analysis. Electrophoresis 28(23):4311–4319
9. Tang LY, Deng N, Wang LS et al (2007) Quantitative phosphoproteome profiling of Wnt3a-mediated signaling network: indicating the involvement of ribonucleoside-diphosphate reductase M2 subunit phosphorylation at residue serine 20 in canonical Wnt signal transduction. Mol Cell Proteomics 6(11):1952–1967
10. Deng WJ, Nie S, Dai J et al (2010) Proteome, phosphoproteome, and hydroxyproteome of liver mitochondria in diabetic rats at early pathogenic stages. Mol Cell Proteomics 9(1):100–116

11. Keshishian H, Addona T, Burgess M et al (2007) Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. Mol Cell Proteomics 6(12):2212–2229
12. Taylor IW, Linding R, Warde-Farley D et al (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 27(2):199–204
13. Peat JK, Barton B (2005) Medical statistics : a guide to data analysis and critical appraisal, 1st edn. Blackwell Publishers, Malden, MA. xii, 324p
14. Breiman L, Friedman RA, Olshen RA (1984) In: Group WI (ed) Classification and regression trees. Belmont, Chapman and Hall
15. Patz EF Jr, Campa MJ, Gottlin EB et al (2007) Panel of serum biomarkers for the diagnosis of lung cancer. J Clin Oncol 25(35):5578–5583
16. O'Shaughnessy DF, Atterbury C, Bolton Maggs P et al (2004) Guidelines for the use of fresh-frozen plasma, cryoprecipitate and cryosupernatant. Br J Haematol 126(1):11–28
17. https://www.thermofisher.com/cn/en/home/references/protocols/cell-and-tissue-analysis/elisa-protocol/elisa-sample-preparation-protocols/plasma-and-serum-preparation.html
18. Mitchell BL, Yasui Y, Li CI et al (2005) Impact of freeze-thaw cycles and storage time on plasma samples used in mass spectrometry based biomarker discovery projects. Cancer Inform 1:98–104
19. Carvalho PC, Lima DB, Leprevost FV et al (2016) Integrated analysis of shotgun proteomic data with PatternLab for proteomics 4.0. Nat Protoc 11(1):102–117
20. Liu YS, Luo XY, Li QR et al (2012) Shotgun and targeted proteomics reveal that pre-surgery serum levels of LRG1, SAA, and C4BP may refine prognosis of resected squamous cell lung cancer. J Mol Cell Biol 4(5):344–347
21. Li RX, Chen HB, Tu K et al (2008) Localized-statistical quantification of human serum proteome associated with type 2 diabetes. PLoS One 3(9):e3224
22. Kuzyk MA, Smith D, Yang J et al (2009) Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Mol Cell Proteomics 8(8):1860–1877

Chapter 27

High-Throughput Parallel Proteomic Sample Preparation Using 96-Well Polyvinylidene Fluoride (PVDF) Membranes and C18 Purification Plates

Tue Bjerg Bennike and Hanno Steen

Abstract

Meaningful proteomic-based biomarker discovery projects using primary human-derived specimens require the analysis of hundreds of samples in order to address the issue of interpersonal variability. Thus, robust high-throughput methods for the digestion of plasma samples are a prerequisite for such large clinical proteomic studies with hundreds of samples. Commonly used sample preparation methods are often difficult to parallelize and/or automate. Herein we describe a method for parallel 96-well plate-based sample preparation. Protein digestion is performed in 96-well polyvinylidene fluoride (PVDF) membrane plates and the subsequent purification in 96-well reversed phase C18 purification plates, enabling the usage of multichannel pipettes in all steps. The protocol can be applied using neat or depleted plasma/serum samples, but has also proven effective with other sample types.

Key words Multichannel pipette, Protein digestion, Plasma, Serum, Method, Depleted, Sample processing, Automatization

1 Introduction

Instrumental and methodological advances over the past two decades have allowed the field of mass spectrometry (MS)-based proteomics to enter the clinical field, which is characterized by the need for large-scale proteomic studies with hundreds of samples [1]. The increasing number of samples being processed requires high-throughput sample preparation techniques, i.e., the ability to process many samples quickly and reproducibly. Commonly, samples are prepared using individual tubes/filters/SDS gel lanes, which are proven and robust [2, 3]. However, such methods are often incompatible with multi pipettes and liquid handling automation. Herein, we describe a protocol for preparing up to 96 samples in parallel for subsequent LC-MS-based proteome mapping using 96-well polyvinylidene fluoride (PVDF) membrane plates for protein digestion and 96-well reversed phase C18 plates for purification. PVDF

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_27,

membranes adsorb and retain proteins through hydrophobic interactions. These PVDF membranes are widely used for Western blotting, i.e., their protein-binding capacity is well established, and the membranes have in addition proven to be compatible with protein digestion protocols [4]. The starting material for the protocol is 1 µL plasma or serum which is diluted 160-fold; only 19% of the resulting solution is trypsinized leaving plenty of protein material for other analyses. For increasing the plasma proteome coverage, the protocol can also be performed with depleted plasma. The protocol includes an optional 96-well reversed phase C18 purification step, which can be performed prior to LC-MS analysis.

2 Materials

Prepare all solution using ultrapure water, HPLC-grade solvents, and analytical-grade reagents. The following reagents are needed for digestion of a full plate (96 samples). The additional amount of solvent needed to compensate for the losses in disposable reservoirs for the multichannel pipettes is accounted for.

2.1 Reagents for Digestion

1. 70% ethanol in water. Mix 14 mL ethanol and 6 mL water in a glass beaker.
2. ABC buffer: 50 mM ammonium bicarbonate (ABC). Weigh 0.80 g of ABC and add to a glass beaker. Add water to a final volume of 200 mL and mix. Store at 4 °C.
3. Urea sample solution: 8 M urea in ABC buffer. Weigh 48.04 g and add to a glass beaker. Add ABC buffer to a volume of 100 mL and mix (*see* **Note 1**).
4. DTT stock: 1 M dithiothreitol (DTT) in water. Weigh 0.154 g DTT into a tube in a fume hood and add water to a volume of 1 mL Mix, and store at −20 °C.
5. DTT solution: 50 mM DTT in urea sample solution. Dilute 250 µL DTT stock in 4.75 mL urea sample solution.
6. IAA solution: 0.25 M iodoacetamide (IAA) in urea sample solution. Weigh 0.23 g IAA into a tube in a fume hood, and add urea sample solution to a volume of 5 mL (*see* **Note 2**). Mix until all IAA is dissolved.
7. Digestion solution: 5% acetonitrile (ACN), 5% trifluoroethanol (TFE), in 50 mM ABC. Add 550 µL ACN and 550 µL TFE to 9.9 mL ABC buffer and mix (*see* **Note 3**).
8. Sequencing grade modified trypsin: 40 µg (V5111, Promega).
9. 40% ACN, 0.1% formic acid (FA): Mix 16 mL ACN and 24 mL water. Add 40 µL FA in a fume hood.

2.2 Reagents for C18 Purification

1. 0.1% trifluoroacetic acid (TFA) in water: Add 60 µL TFA to 60 mL water. TFA should be handled in a fume hood.
2. 40% ACN, 0.1% TFA: Mix 10 mL ACN and 15 mL water. Add 15 µL TFA.
3. 70% ACN, 0.1% TFA: Mix 35 mL ACN and 15 mL water. Add 50 µL TFA.

2.3 Plates and Disposables for Digestion

1. PVDF membrane plate: Multiscreen HTS 0.45 µm hydrophobic high protein-binding membrane 96-well filtration plate (MSIPS4510, Millipore).
2. 96-Well microplate vacuum manifold (MAVM0960R, Millipore).
3. 2× 96-well plates with V-bottom which can contain >300 µL solvent.
4. 1× large 96-well plate with V-bottom which can contain >500 µL solvent.
5. Disposable reagent reservoirs.
6. Adhesive cover-slides with hairline crosscuts for the 96-well plates (X-Pierce Film 29997–0100, USA Scientific).

2.4 Plates and Disposables for C18 Purification

1. 96-Well MACROSpin plate TARGA C18 (#SNS SS18RL, Nest group) and the two enclosed large 96-well collection plates.
2. Disposable reagent reservoirs.

3 Methods

Carry out all procedures at room temperature unless specified otherwise (Fig. 1).

3.1 Protein Reduction and Alkylation

1. Randomize the sample list (*see* **Note 4**).
2. Label a disposable reagent reservoir and add 50 mL urea sample solution. Add 100 µL Urea sample solution to each well in a 96-well plate with a multichannel pipette.
3. Add 1 µL plasma/serum sample (50–70 µg protein) to each well (*see* **Note 5**).
4. Label a disposable reagent reservoir, and add 5 mL DTT solution in a fume hood. Reduce cysteine disulfide bonds by adding 30 µL DTT solution to each well with a multichannel pipette (final DTT concentration 11.5 mM).
5. Add a cover-slide to prevent contaminations and evaporation. Briefly shake the plate to mix and incubate for 20 min.

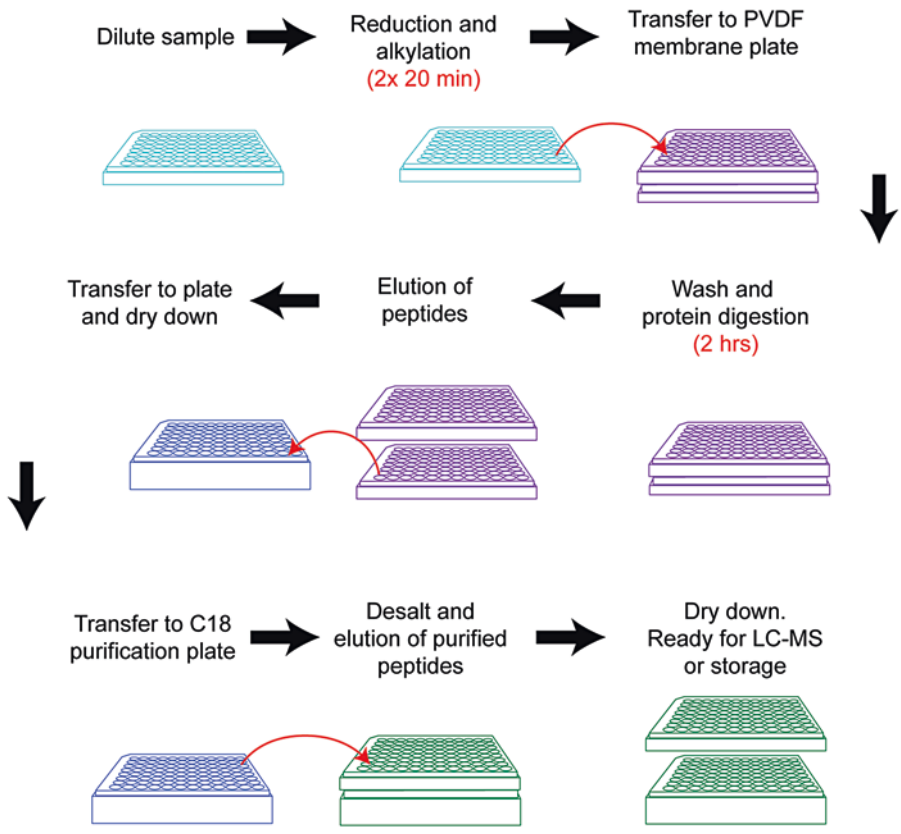


Fig. 1 Overview of the 96-well sample preparation protocol. The protocol consists of three main parts: (1) reduction and alkylation, (2) protein binding on the PVDF membrane and digestion, and (3) sample purification prior to analysis

6. Label a disposable reagent reservoir and add 5 mL IAA solution in a fume hood. Alkylate cysteine residues by adding 30 µL IAA solution to each well with a multichannel pipette (final IAA concentration 46.9 mM).
7. Briefly shake the plate to mix and incubate for 20 min in the dark.

3.2 Prepare the 96-Well PVDF Membrane Plate

1. While the alkylation reaction is ongoing, place the 96-well PVDF membrane plate on a 96-well (*see* **Note 6**). Label a disposable reagent reservoir, and add 20 mL 70% ethanol. Activate the PVDF membrane plate by adding 150 µL 70% ethanol to each well with a multichannel pipette.
2. Place the 96-well PVDF membrane plate with the 96-well collection plate underneath in the vacuum manifold. Slowly apply the house vacuum, and allow the liquid to pass through (*see* **Note 7**). Do this for all digestion solution-passing steps.
3. Discard the flow-through from the collection plate (*see* **Note 8**).

4. Prime the PVDF membrane by adding 200 µL Urea sample solution to each well with a multichannel pipette.
5. Pass the urea sample solution through the PVDF membrane, and discard the flow-through (*see* **Note 9**).

3.3 Protein Digestion in 96-Well PVDF Membrane

1. Add 30 µL reduced and alkylated protein solution (approx. 9–13 µg protein) to the middle of each well with a multichannel pipette, pass it through, and discard the flow-through (*see* **Note 10**).
2. Label a disposable reagent reservoir and add 25 mL 50 mM ABC buffer. Wash by adding 200 µL 50 mM ABC buffer with a multichannel pipette, pass it through, and discard the flow-through.
3. Replace the bottom 96-well collection plate to a new one, which will collect the peptide solution.
4. Dissolve 40 µg sequencing grade modified trypsin in 11 mL Digestion solution.
5. Label a disposable reagent reservoir and add 11 mL digestion solution with trypsin. Add 100 µL Digestion solution with trypsin to each well with a multichannel pipette.
6. Cover the membrane with a lid and incubate at 37 °C for 2 h.
7. Remove the plate from the incubator and pass the digested peptides through. The flow-through contains peptides; do not discard.
8. Label a disposable reagent reservoir and add 40 mL 40% ACN, 0.1% FA. Elute remaining peptides by adding 150 µL 40% ACN, 0.1% FA to each well with a multichannel pipette and pass it through.
9. Transfer the eluent from the 96-well collection plate to a large 96-well plate with a multichannel pipette.
10. Repeat the 40% ACN, 0.1% FA elution and transfer step once.
11. Cover the plate with a cover-slide, and dry down the samples in a vacuum centrifuge (*see* **Note 11**).

3.4 C18 Purification in 96-Well Plates

1. Label a disposable reagent reservoir and add 60 mL 0.1% TFA. Resuspend the peptides in the large 96-well collection plate by transferring 100 µL 0.1% TFA to each well with a multichannel pipette. Briefly shake the plate to mix and then sonicate for 2 min.
2. Place a 96-well C18 plate in the enclosed 96-well collection plate (*see* **Note 12**).
3. Label a disposable reagent reservoir and add 50 mL 70% ACN, 0.1% TFA. Wash the 96-well C18 plate by adding 150 µL 70% ACN, 0.1% TFA to each well with a multichannel pipette.

4. Pass the solution through by centrifuging the plate at 2000 × *g* for 2 min. Do this for all C18 purification solution-passing steps. Discard the flow-through.
5. Equilibrate the 96-well C18 plate by adding 200 μL 0.1% TFA with a multichannel pipette. Pass the solution through and discard the flow-through.
6. Transfer the resuspended peptide samples to the 96-well C18 plate with a multichannel pipette. Pass the solution through and discard the flow-through.
7. Wash by adding 200 μL 0.1% TFA to each well in the 96-well C18 plate with a multichannel pipette. Pass the solution through and discard the flow-through.
8. Place the 96-well C18 plate in the second enclosed 96-well collection plate.
9. Label a disposable reagent reservoir and add 25 mL 40% ACN, 0.1% TFA. Elute peptides by adding 200 μL 40% ACN, 0.1% TFA to each well with a multichannel pipette and pass the solution through. The flow-through contains peptides; do not discard!
10. Re-elute peptides by adding 200 μL 70% ACN, 0.1% TFA to each well with a multichannel pipette and pass the solution through.
11. Place the 96-well collection plate in a vacuum centrifuge, and dry down the peptides. Store the purified dry peptide product at −20 °C until time of LC-MS analysis.

4 Notes

1. The urea sample solution should be freshly prepared. The dissolving of urea can take time. The reaction is endothermic and can be accelerated by gentle heating to 20 °C. However, care should be taken not to heat the urea solution above room temperature, which accelerates the breakdown of urea into ammonium isocyanate, which can result in carbamylation of primary amino groups such as lysine side chains.
2. Iodoacetamide is light sensitive. Avoid direct sunlight and keep in the dark when possible.
3. Digestion in the presence of organic solvents (ACN and TFE) can increase the proteolysis efficiency, thereby lowering the abundance of peptides with several tryptic missed cleavages.
4. Randomization of samples is extremely important to avoid batch effects. This is especially true for large proteomic studies, which can be compromised if not randomized. We recommend assigning each sample a random number and sorting the samples by these. Because humans are surprisingly bad at producing ran-

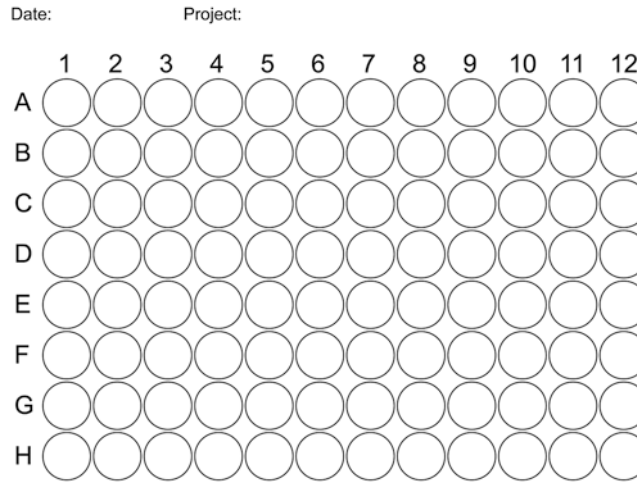


Fig. 2 Overview of the 96-well plates

dom numbers, we recommend using an online tool, e.g., Research Randomizer [5].

5. To easily keep track of the samples, one can preferably download and print the 96-well plate overview (Fig. 2).
6. In all transfer steps, ensure that the A1-well on the 96-well PVDF membrane plate is positioned on the A1-well on the 96-well collection plate. The plate design might be different, so marking the A1 position of the plates can help keep the orientation.
7. To avoid damaging the PVDF membranes, care should be taken when applying the vacuum, which should not exceed −0.5 psi (−1 inHg, −0.034 bar, −3.4 kPa). The liquid should pass through within 10–30 s. Also, the membranes should never be left with high organic solvents for prolonged time.
8. Droplets often remain underneath the 96-well PVDF membrane plate following liquid extraction. If this is the case, hold the vacuum manifold with plates inside firmly, and tap it once to release the droplets.
9. It is important that the PVDF membrane does not dry out. Therefore, add the urea sample solution, but do not pass it through before you are ready to load the reduced and alkylated protein sample.
10. If loading more than 15 μg protein, significant amounts of protein will not be bound by the PVDF membrane [4].
11. The dry peptide product can be stored at −20 °C, analyzed directly, or C18 purified. In our experience, C18 cleaning the samples reduces issues with overpressure on the LC-MS columns.

12. Tap the 96-well C18 plate prior to removing the protective foil to ensure that all C18 powder is at the bottom of the wells. Always add solutions slowly and dropwise to the middle of the wells.

References

1. Aebersold R, Mann M (2016) Mass-spectrometric exploration of proteome structure and function. Nature 537:347–355
2. Bennike TB, Kastaniegaard K, Padurariu S et al (2016) Comparing the proteome of snap frozen, RNAlater preserved, and formalin-fixed paraffin-embedded human tissue samples. EuPA Open Proteomics 10:9–18
3. Bennike T, Ayturk U, Haslauer CM et al (2014) A normative study of the synovial fluid proteome from healthy porcine knee joints. J Proteome Res 13:4377–4387
4. Berger ST, Ahmed S, Muntel J et al (2015) MStern blotting–high throughput polyvinylidene fluoride (PVDF) membrane-based proteomic sample preparation for 96-well plates. Mol Cell Proteomics 14:2814–2823
5. G.C. Urbaniak and S. Plous Research Randomizer (version 4.0) [Computer software], http://www.randomizer.org/

Chapter 28

Targeted Quantification of the Glycated Peptides of Human Serum Albumin

Garikapati Vannuruswamy, Arvind M. Korwar, Mashanipalya G. Jagadeeshaprasad, and Mahesh J. Kulkarni

Abstract

Glycated human serum albumin (HSA) serves as an important marker for monitoring the glycemic status. Developing methods for unambiguous identification and quantification of glycated peptides of HSA using high-throughput technologies such as mass spectrometry has a great clinical significance. The following protocol describes the construction of reference spectral libraries for Amadori-modified lysine (AML), N(ε)-(carboxymethyl) lysine (CML)-, and N(ε)-(carboxyethyl)lysine (CEL)-modified peptides of synthetically modified HSA using high-resolution mass spectrometers. The protocol also describes work flows, for unambiguous identification and quantification of glycated modified peptides of HSA in clinical plasma using standard spectral libraries by various mass spectrometry approaches such as parallel reaction monitoring (PRM), sequential window acquisition of all theoretical fragment ion spectra (SWATH), and MS^E.

Key words Targeted quantification, PRM, SWATH, Human serum albumin, Diabetes, Glycation, Post-translation modifications

1 Introduction

The inevitable consequence of hyperglycemic condition in diabetes is increased rate of nonenzymatic glycation between glucose and plasma proteins. The epsilon amino group of lysine and arginine is modified by glucose to form a relatively stable Amadori modification, which undergoes subsequent series of reactions involving oxidation, dehydration, condensation, fragmentation, or cyclization leading to the formation of advanced glycation end products (AGEs) (Fig. 1). Carboxymethyl lysine and carboxyethyl lysine are the predominant AGEs [1, 2].

Human serum albumin (HSA) is the principal target of glycation in the plasma, as it is a most abundant plasma protein with large number of lysine and arginine residues and relatively longer half-life. Glycated albumin interacts with receptor for advanced glycation end

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_28,

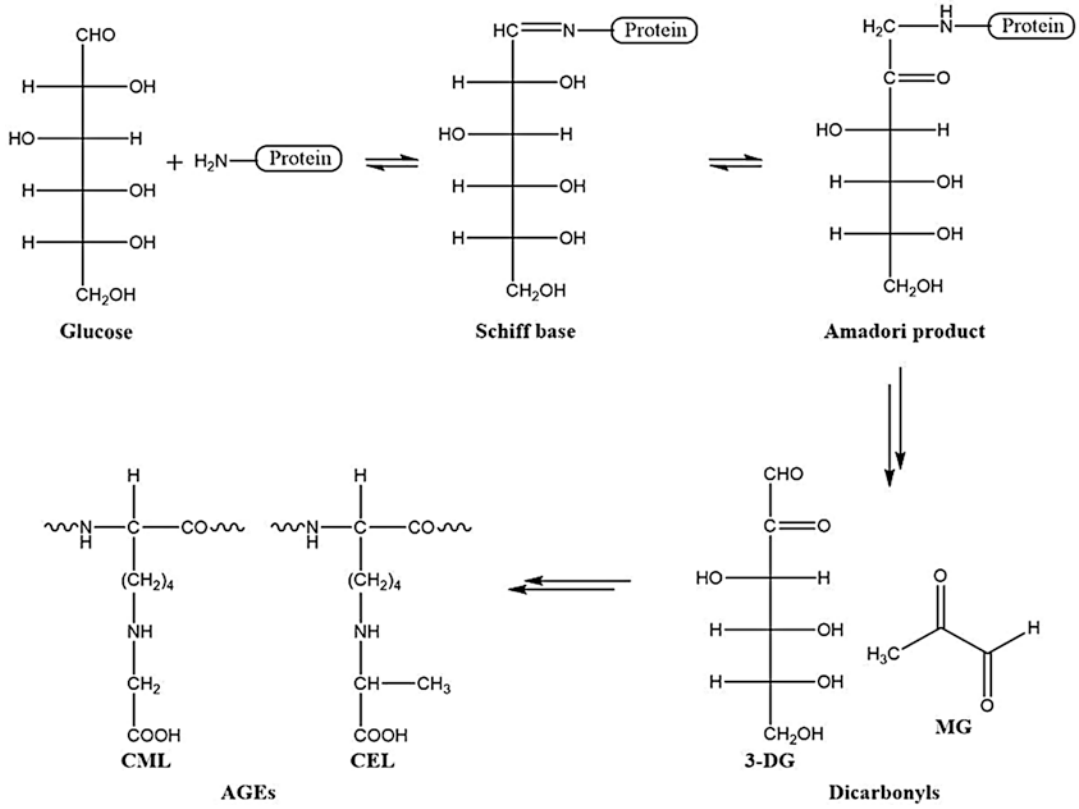


Fig. 1 Formation of advanced glycation end products (AGEs)

product (RAGE) activating NADPH oxidase and transcription factor NF-kB leading to oxidative stress and inflammation respectively. The AGE-RAGE axis is implicated in the pathogenesis of diabetes and its complications. Since glycated albumin constitutes the predominant AGEs, and considering the limitations of HbA1c in conditions like anemia, splenomegaly, and gestational diabetes, quantification of glycated albumin has a great clinical significance. Mass spectrometry-based quantification approaches such as multiple reaction monitoring (MRM), parallel reaction monitoring (PRM), and sequential window acquisition of all theoretical fragment ion spectra (SWATH) heavily rely on fragment ion library [3, 4]. In this context, we constructed a fragment ion library for the synthetically Amadori-modified lysine (AML), carboxymethyl lysine (CML), and carboxyethyl lysine (CEL)-modified peptides of albumin by using a high-resolution accurate mass spectrometer followed by rigorous inspection and validation of MS/MS spectra. Furthermore, using the ion library, AML-, CML-, and CEL-modified albumin peptides were quantified by targeted SWATH analysis in the clinical plasma. The detail work flow for the construction of diagnostic fragment ion spectral libraries and quantification of glycated HSA using these libraries is depicted in Fig. 2.

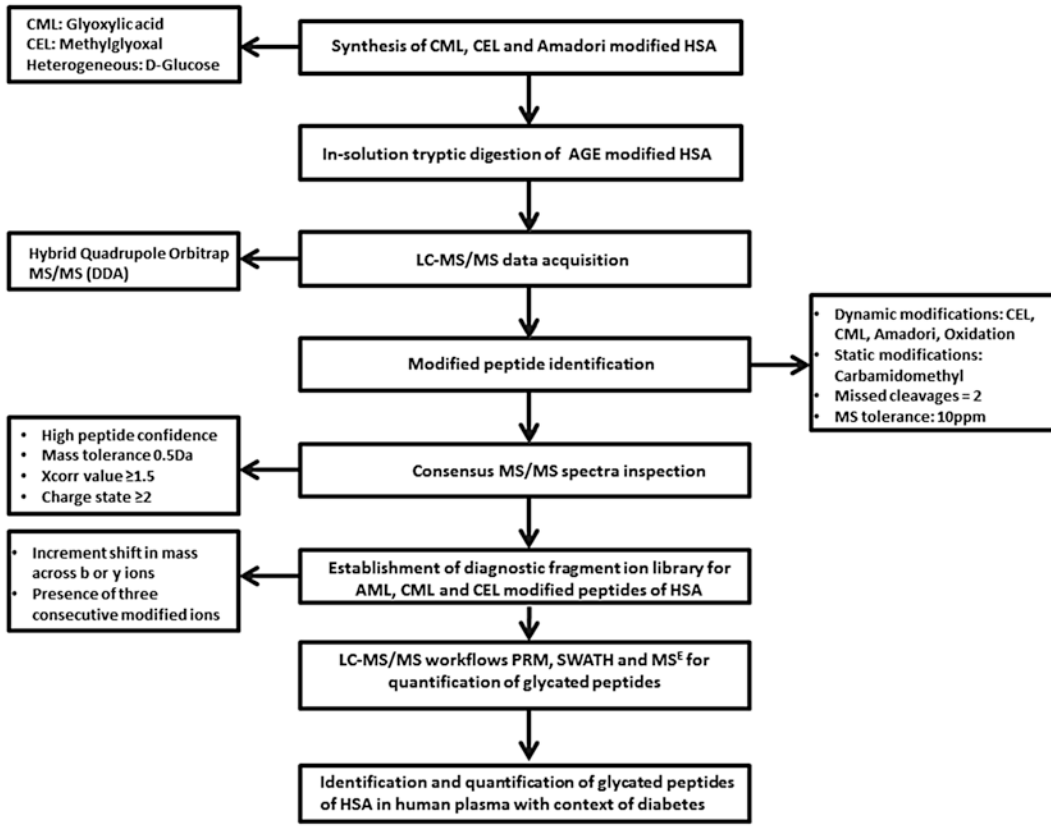


Fig. 2 Schematics of sample preparation and pipeline for identification and quantification of glycated peptides of HSA in human plasma

2 Materials

All the chemicals and reagents should be highly pure. The solvents and additives (acetonitrile, water, methanol, formic acid, hydrochloric acid) are MS grade. The detergents used for the proteome solubilization should be MS compatible. Proteomic grade trypsin as protease (Sigma-Aldrich).

2.1 Reagents Preparation

1. 0.2 M sodium phosphate buffer (pH 7.4):
   Dissolve 35.61 g of disodium hydrogen phosphate dihydrate ($Na_2HPO_4 \bullet 2H_2O$) and 27.6 g of sodium phosphate monobasic monohydrate ($NaH_2PO_4 \bullet H_2O$) separately in water and make up the volume to 1000 mL. To prepare 1000 mL of 0.2 M sodium phosphate buffer, add 770 mL of $Na_2HPO_4 \bullet 2H_2O$ and 230 mL of $NaH_2PO_4 \bullet H_2O$ and adjust the pH to 7.4 if necessary. Store the buffer solution at 4 °C.
2. 50 mM ammonium bicarbonate (NH_4HCO_3) buffer:
   Dissolve 35 mg of NH_4HCO_3 in 10 mL of water.

3. 0.1% RapiGest (Waters):
   Dissolve 1 mg of RapiGest SF powder (vial) in 1 mL of 50 mM of NH_4HCO_3 buffer. The aliquots can be stored at 2–8 °C for 1 week. Long-term storage of frozen aliquots is not recommended.
4. 100 mM dithiothreitol (DTT):
   Dissolve 15.42 mg of DTT in 1 mL of 50 mM NH_4HCO_3 buffer which acts as reducing agent.
5. 200 mM iodoacetamide (IAA):
   Dissolve 36.99 mg of IAA in 1 mL of 50 mM NH_4HCO_3 buffer which acts as alkylating agent. IAA is unstable and light-sensitive, so we recommend to cover the IAA solution with aluminum foil.
6. Trypsin solution:
   Dissolve 20 μg (1 vial) of proteomic grade trypsin in 20 μL of 50 mM NH_4HCO_3 buffer, vortex followed by centrifuge. Make the aliquots of trypsin and store at −20 °C. In general, we recommend 2 μg of trypsin for 100 μg of protein (1:50) ratio.

2.2 Chromatography Buffers

Mobile Phase A (aqueous): 0.1% formic acid in 100% water.

Mobile Phase B (organic): 0.1% formic acid in 100% acetonitrile.

3 Methods

3.1 Synthesis of In Vitro Glycated HSA

In vitro glycated modified HSA synthesis was described in detail in the following sections, and schematic representation is depicted in Fig. 3.

3.1.1 In Vitro AGE-Modified HSA

The in vitro heterogeneous AGE-modified HSA was synthesized according to the earlier report with slight modifications [5].

1. Prepare 0.2 M sodium phosphate buffer (pH 7.4) freshly.
2. Dissolve 500 mg of HSA, 900.8 mg of glucose, and 5 mg of sodium azide in 10 mL of sodium phosphate buffer (pH 7.4) and sterilize by ultrafiltration using 0.22 μm filters.
3. Incubate the solution at 37 °C for 7 days.
4. After incubation, wash the sample with buffer using 30 kDa cutoff filters.
5. Measure the AGE fluorescence at 370 nm (excitation)/440 nm (emission) for the confirmation of AGE modification.

3.1.2 In Vitro CML- and CEL-Modified HSA

The in vitro CML- and CEL-modified HSA were synthesized according to earlier reports with slight modifications [6].

1. Prepare 0.2 M sodium phosphate buffer (pH 7.4) freshly.
2. Dissolve 500 mg of HSA, 94.26 mg of sodium cyanoborohydride, and 33.31 mg of glyoxylic acid (GA) or 36.03 mg of

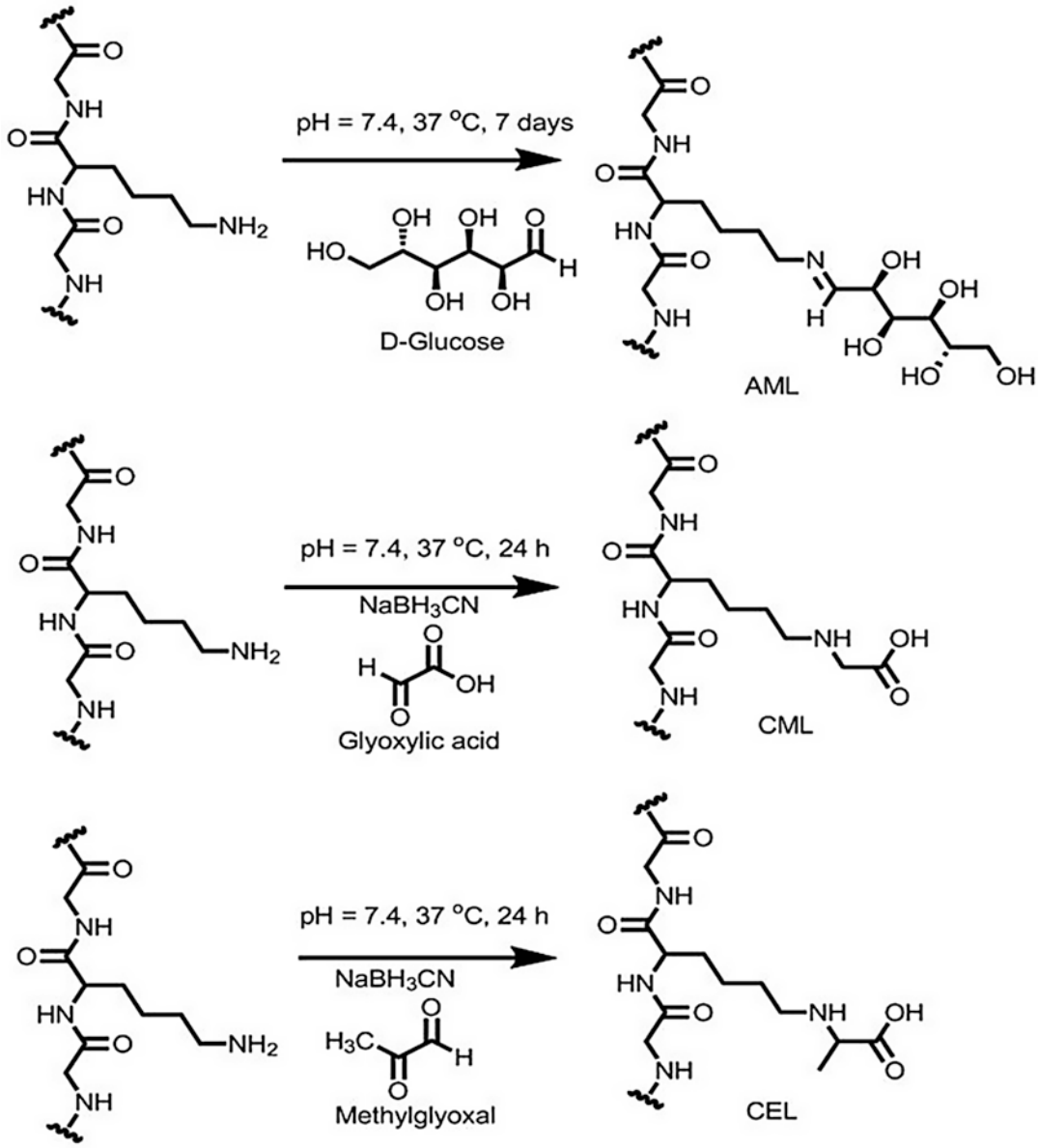


Fig. 3 Schematic representation of synthesis of AGE-modified HSA

methylglyoxal (MG) in 10 mL of sodium phosphate buffer (pH 7.4) and sterilize by ultrafiltration using 0.22 μm filters.

3. Incubate the solutions at 37 °C for 24 h.
4. Glyoxylic acid and methylglyoxal (*see* **Note 5a**) will induce the carboxymethyl (CML) and carboxyethyl (CEL) modifications at the lysine residues respectively.
5. Perform the **steps 4** and **5** as mentioned in Subheading 3.1.1.

3.2 Human Plasma Preparation

1. Collect the peripheral blood in anticoagulant (EDTA) vacutainers from clinical subjects approved by Institutional Human Ethics Committee.
2. Remove the blood cells by centrifugation for 15 min at 1500 × *g* using refrigerated centrifuge.
3. Collect the supernatant (plasma) into a clean polypropylene tubes and store at −80 °C till further use.

3.3 In-Solution Digestion

1. Estimate the protein concentration using Bradford assay (*see* **Note 1a**). And use 100 μg of protein for the digestion.
2. Adjust the volume to 100 μL with 50 mM NH_4HCO_3 buffer containing 0.1% RapiGest SF (*see* **Note 1b**). After addition of RapiGest SF, incubate the protein at 80 °C for 15 min for complete proteome solubilization.
3. Add 5 μL of reducing agent (100 mM DTT), vortex and incubate at 60 °C for 15 min.
4. Cool down the solution to room temperature. And add 5 μL of alkylating agent (200 mM IAA), vortex and incubate for 30 min in dark.
5. Add 2 μg of trypsin (1:50), vortex and incubate overnight at 37 °C (16–18 h) (*see* **Notes 1c** and **d**).
6. After overnight incubation, digestion reaction was stopped by adding 2 μL concentrated HCl (*see* **Note 5b**) and incubate for 10 min at 37 °C before centrifugation.

3.4 Peptide Cleanup (Desalting)

Perform the peptide cleanup with Ziptip and reconstitute the dried tryptic digest peptides with mobile phase A or 3% acetonitrile with 0.1% FA.

1. Prepare the solutions freshly, set the pipettor (P10) to 10 μL and place Ziptip on P10 pipettor.
2. Equilibrate the Ziptip by aspirating with 100% ACN (4–5×) followed by 0.1% FA (4–5×).
3. Slowly aspirate the peptide sample and expel (*see* **Notes 2a** and **b**) the liquid in to the tube (10×).
4. Wash the Ziptip with 0.1% FA (1–2×).
5. Elute the peptides by pipetting the Ziptip up and down with 50% ACN with 0.1% FA.

3.5 Concentrate the Sample

1. Speed vac the eluate from Ziptip to dry the peptide digest (*see* **Note 2c**).
2. Reconstitute the dried peptide digest with mobile phase A or 3% ACN with 0.1% FA.

3.6 LC-MS/MS

We always recommend to use the high mass accuracy and high-resolution mass spectrometers for the precise identification and quantification of glycation or any other post-translational

modification of proteins. The following sections describe about accurate site-specific identification and quantification of glycated modified peptides of albumin. First, we have explained the construction of standard diagnostic fragment ion library using high-resolution accurate mass spectrometers (e.g., orbitrap), which serves as a reference spectral library. And later, we have explained the work flows for the quantification of the modified peptides using different mass spectrometry approaches such as PRM, SWATH, and MS^E.

3.6.1 Construction of Diagnostic Fragment Ion Library for Glycated HSA

Data Acquisition

1. Calibrate, the obritrap mass spectrometer with the external standard calibration solution (<3 ppm RMS) in a positive ion mode. We recommend to use internal mass calibration (lock mass, <1 ppm RMS) to maintain the high mass accuracy of instrument. Users should tune the mass spectrometer to obtain the high-quality/acceptable mass spectrum.
2. Maintain the temperature of the column and auto sampler at 40 °C and 4 °C respectively.
3. Chromatographic separation gradient from 2% to 40% of mobile phase B in 45 min for synthetic modified peptide digest. In case of plasma samples, the LC method was extended to 120 min with a linear gradient of 2–50% of mobile phase B (*see* **Note 3a**).
4. The samples need to acquire in at least technical triplicates for the reliability. Short gradient blank injections are recommended in between of samples to avoid carry-over from column.
5. To perform data-dependent acquisition (DDA) for typical proteomics (single protein) experiments with orbitrap MS, we suggest to use following parameters. Positive ion mode, HCD with top five intense ions in the mass range of 350–1800 $m/z (Z \geq 2)$, orbitrap precursor (MS) ion resolution at 70,000 (m/z at 200) with the target (AGC) of $1xe^6$ ions and maximum injection time of 120 ms. For fragment ions, resolution at 17,500 (m/z at 200) with AGC of $1xe^5$ ions, dynamic exclusion time of 15 s and underfill ratio of 0.3%. In case of complex proteome 10 or 15 intense ions would be recommended for MS/MS experiments.
6. We advise to optimize the instrumental parameters to obtain highest quality of data (*see* **Note 3b**).

Data Analysis

The following section describes the standard work flow for the identification and construction of diagnostic fragment ion libraries of glycated modified peptides of HSA using Proteome Discoverer (V 1.4) software with SEQUEST HT as a search engine (*see* **Note 4a**).

1. Use the human serum albumin protein database (P02768-UniProt) available from different database sources.

2. Search parameters: Carbamidomethylation (C) as fixed (static) modification, oxidation (M), Amadori (162.0528 Da at K, R), carboxymethyl (58.0055 Da at K), and carboxyethyl (72.0211 Da at K) as variable (dynamic) modifications, with 10 ppm (MS), 0.5 Da (MS/MS) mass tolerance. The additional search parameters include two missed cleavages, ESI ionization, HCD fragmentation, and 1% false discovery rate (FDR) with percolator.
3. After completion of search, open the report in Proteome Discoverer using high peptide confidence as a filter and export the peptide group results (.xls file).
4. The glycated peptides were identified and library of diagnostic fragment ions for modified peptides of HSA was constructed based on manual inspection described in earlier reports [7, 8]. Main criteria are described below:
   (a) Presence of missed cleavage at the site of modification.
   (b) Presence of unmodified peptide precursor for the corresponding AGE-modified peptide.

   For example, the Amadori (162.052)-modified peptide of **KAML**QTALVELVK (m/z = 645.8, MH^+ = 1290.74) at lysine position K549 was manually inspected for the presence of unmodified KQTALVELVK (m/z = 564.85, MH^+ = 1128.69) peptide. Similarly, all the modified peptides were validated manually for increment in mass of 162.052 (AML), 58.005 (CML) and 72.021 (CEL) Da at peptide precursor level.
   (c) Depending upon the site of modification, fragment ions (either b or y) should retain the mass of increment.
      - If the modification is at N-terminal lysine, b ions (b1, b2, b3, …) should bear the increment of mass shift.
      - If the modification is at C-terminal lysine, y ions (y2, y3, y4, etc.) should bear the increment of mass shift. Readers should remember that, it is unlikely to observe the modification on y1 ion at C-terminus.
      - If the modification is at the middle of the peptide, both b and y ions bear the increment of mass shift.
   (d) Low quality MS/MS spectra may result in false identification of glycated peptides. To rule out false positive identifications, we advise to monitor the three consecutive fragment ions bearing mass shift.
   (e) For instance, MS/MS annotation of CML- and CEL-modified peptide (**K**QTALVELVK) is depicted in Figs. 4 and 5, respectively.

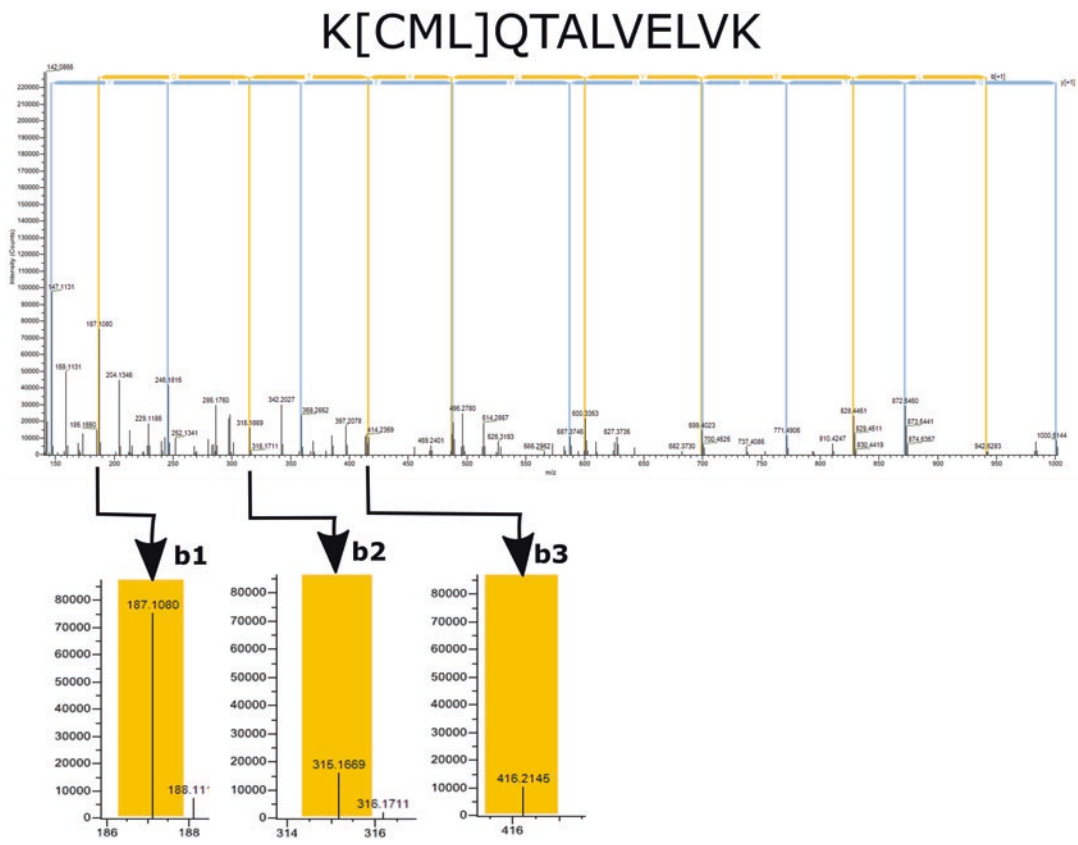


Fig. 4 MS/MS annotation of CML-modified (**KCML**QTALVELVK) peptide of HSA

5. Based on the above criteria, 50 glycated modified peptides of HSA which represents 23 lysine-modified sites were identified and fragment ion library was constructed (*see* **Note 4b**). Out of these 20, 17, and 13 modified peptides were induced by glucose, GA and MG respectively.
6. Based on the results, the lysine sites K549, K438, K183, K375, and K490 of HSA are more sensitive to glycation modifications.
7. These diagnostic fragment ion libraries can serve as a reference, for the identification and quantification of glycated peptides of HSA at fragment ion (MS/MS) level with any type of mass spectrometer (*see* **Note 4b**).

3.6.2 Work Flow for the Identification and Quantification of Glycated Modified Peptides of HSA in Human Plasma

Sample (Plasma) Processing for MS Analysis

Wash the plasma sample with PBS using 30 kDa filters, followed by protein estimation and in-solution tryptic digestion and peptide cleanup as mentioned above (Subheadings 3.3–3.5).

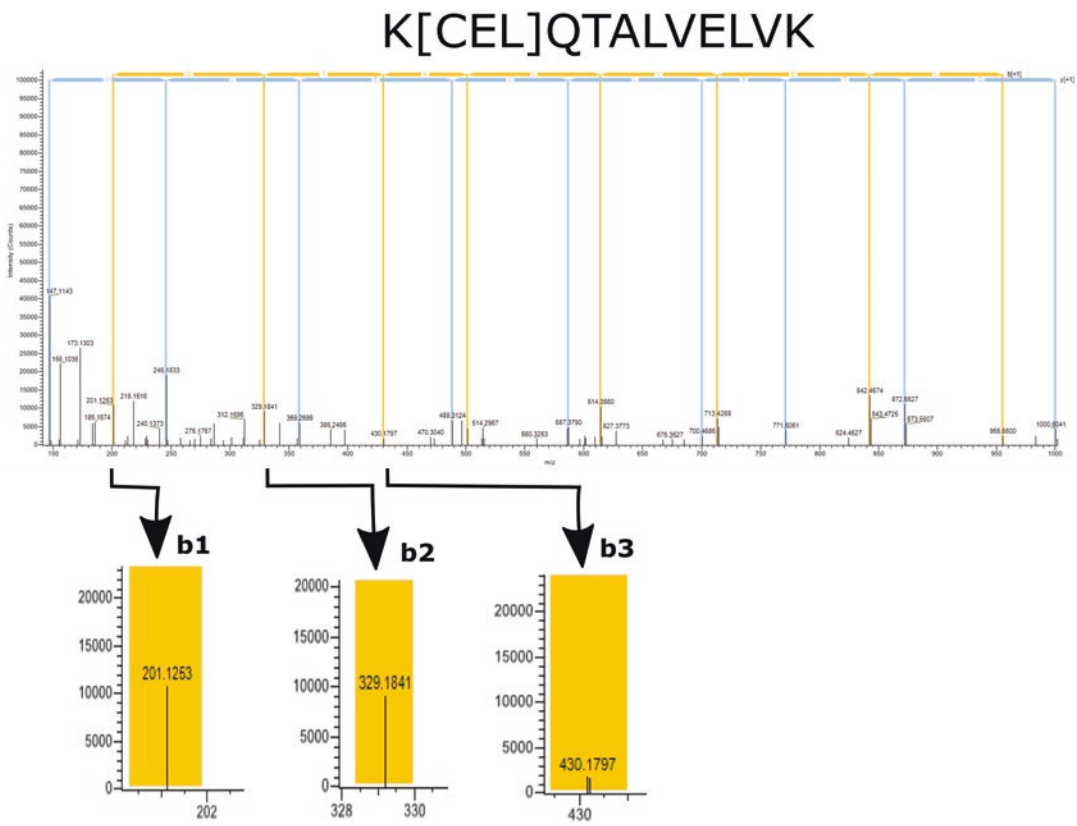


Fig. 5 MS/MS annotation of CEL-modified (**KCEL**QTALVELVK) peptide of HSA

Parallel Reaction Monitoring (PRM) Work Flow

SRM or MRM are considered as gold standards for the targeted quantitative proteomics approach, since they provide high selectivity with two stage mass filtering (MS and MS/MS). Normally these experiments can be performed on low-resolution mass spectrometers (triple-quadrupole), resulting difficulties in removal of interferences. This will be problematic when working with complex biological fluids. To overcome these problems, PRM technology was developed. It is also known as HR-MRM where MRM-like experiments will be performed using high-resolution mass analyzers such as orbitrap or TOFs [9].

Based on in vitro modified HSA spectral library, a list of glycated modified peptides of HSA was generated in excel sheet, and readers can access the inclusion list of peptides (m/z, charge state, retention time) for PRM-based experiments from the supplementary Table 2 from earlier report [8].

The PRM data can be acquired with the inclusion list of peptides using orbitrap mass spectrometers. Glycated peptides can be inspected for the quality of MS/MS spectrum based on above

mentioned criteria. Quantitative information for the glycated peptides of HSA among the different samples can be obtained either manually (XIC using XCalibur) or by using quantitative proteomic software like Skyline or Pinpoint etc. PRM or MRM data will be a good choice for validation of spectral libraries.

SWATH Work Flow

SWATH is a novel data-independent mass spectrometry acquisition method, coupled with peptide spectral library. In general peptide spectral libraries will be generated by using DDA method. Development of high-quality spectral library is a primary step in SWATH-based quantification approach [3].

1. Calibrate and tune the mass spectrometer (Triple TOF MS coupled to LC) in both MS and MS/MS in high sensitive mode (*see* **Note 3b**).
2. Separation gradient from 3% to 50% of mobile phase B in 120 min. Same chromatographic conditions need to maintain for both DDA and SWATH experiments (*see* **Note 3a**).
3. Construct the spectral library by analyzing the in vitro modified peptide digest in positive, high-sensitive mode using data-dependent acquisition with the following parameters. Mass range from 350 to 1800 *m/z* with the accumulation time of 250 ms.
4. For the fragmentation, ions were selected with more than 120 counts per second, charge stage in between +2 to +5, mass tolerance of 50 mDa using rolling collision energy (CE ± 15) as source for fragmentation.
5. For the SWATH-based experiments, tune the quadrupole for the selection of precursor ion window of 25 *m/z*. With an isolation width of 26 *m/z* (1 *m/z* for the window overlap), there is a set of 34 sequential windows in the mass range of 400–1250 *m/z*.

SWATH Data Analysis

1. Search synthetic glycated peptide DDA files (.wiff) against the human serum albumin protein database using Protein Pilot (Paragon algorithm) software with 1% FDR. The resulted (.group) file serves as standard peptide spectral library for the quantification of glycated HSA peptides using SWATH.
2. We advise to acquire the high-quality technical triplicate SWATH acquisition of plasma samples. Spectral alignment and targeted data extraction of SWATH data can be performed by using PeakView software. Quantification, normalization, and other statistical data analyses using MarkerView software (*see* **Note 4a**).
3. We advise to use OpenSWATH or SWATHProphet software for data analysis in addition to the above mentioned.

MS^E Work Flow

1. Peform MS^E experiments with Synapt HDMS mass spectrometer coupled to LC.
2. Tune the instrument in positive V-mode and calibrate using glu-fibrinopeptide infusion. Separation gradient from 2% to 40% of mobile phase B (*see* **Note 3a**).
3. Acquire the data in positive V-mode in a mass range of 50–2000 *m/z* with 0.7 s scan time with alternative low (4 eV) and high (15–40 eV) collision energy.
4. Analyse the data using Protein Lynx Global Server (PLGS) and Mass Lynx software with manual inspection for glycated HSA peptides.

Statistical Analysis

For the significant quantification data analysis, we advise to perform various statistical tools like T-test, two-way ANOVA, etc. across the various samples.

Standard Spectral Libraries for Glycated HSA

In total 50 glycated modified peptides of HSA reference spectral library constructed from in vitro synthesized AGE-modified HSA (*see* **Note 4b**). Out of these K549, K438, K183, K375, and K490 sites are more prone to glycation.

4 Notes

1. Digestion:
   (a) Quantification of protein concentration is a necessary step, which helps to calculate the required amount of reducing, alkylating agent and protein to protease enzyme ratio. We advise to perform protein estimation using simple techniques such as Bradford or NanoDrop before tryptic digestion every time.
   (b) We always recommend to use MS-compatible detergents with very less concentration for proteome solubilization. If not, results in improper digestion and interference in MS spectrum. Long-term storage of RapiGest frozen aliquots is possible but not recommended because of solubilization issues.
   (c) DTT is susceptible to oxidation and should be prepared freshly every time. Iodoacetamide is unstable and light sensitive; perform the experiments in dark conditions while working with IAA. Try to avoid many freeze thaw cycles of frozen trypsin aliquots.
   (d) Perform the protein digestion in clean low bind tubes.
2. Peptide cleanup (desalting):
   (a) During aspiration and dispensing, take precaution to prevent introducing air bubbles in to Ziptip.
   (b) Need to consider the Ziptip capacity (saturating limit).

ZiptipC18: greater than or equal to 1 μg. Typically 5.0 μg.

(c) Concentrate the peptide digest, at a temperature not more than 40 °C.

3. LC-MS/MS:

(a) Use MS grade solvents for LC-MS/MS experiments. Prepare mobile phase buffers freshly in clean solvent bottles with proper sonication and degassing steps. Chromatographic conditions (total run time, flow rate, etc.) need to optimize, depending on LC and column availability.

(b) Monitor both chromatography (peak shape, high column back pressures, RT shift, etc.) and mass spectrometry (mass accuracy, signal to noise ratio, etc.) parameters to obtain high-quality data.

4. Data analysis:

(a) We advise to use different algorithms (SEQUEST, SEQUEST HT, Paragon, Mascot, etc.), alternative software tools (MaxQuant, Progenesis, OpenSWATH, SWATHProphet etc.), and different statistical tools depending on user accessibility.

(b) Reader can access the complete list of modified peptide information (peptide sequence, site of modification, start-end site of peptide, diagnostic fragment ions, MS/MS annotation of each peptide, XIC, and additional information) from our earlier report [8].

5. General considerations:

(a) Use personal protective equipment (e.g., lab coat, gloves, goggles) while handling toxic chemicals (e.g., sodium azide, methylglyoxal) and to avoid contamination of keratin in proteomics experiments.

(b) Formic acid (FA), hydrochloric acid (HCl) solutions, and vapors are toxic; use fume hood while handling with these solutions.

References

1. Reddy S, Bichler J, Wells-Knecht KJ, Thorpe SR, Baynes JW (1995) N-epsilon-(carboxymethyl)lysine is a dominant advanced glycation end-product (AGE) antigen in tissue proteins. Biochemistry 34:10872–10878
2. Thornalley PJ, Battah S, Ahmed N, Karachalias N, Agalou S, Babaei-Jadidi R, Dawnay A (2003) Quantitative screening of advanced glycation end products in cellular and extracellular proteins by tandem mass spectrometry. Biochem J 375:581–592
3. Schubert OT, Gillet LC, Collins BC, Navarro P, Rosenberger G, Wolski WE, Lam H, Amodei D, Mallick P, MacLean B, Aebersold R (2015) Building high-quality assay libraries for targeted analysis of SWATH MS data. Nat Protoc 10:426–441
4. Dong Q, Yan X, Kilpatrick LE, Liang Y, Mirokhin YA, Roth JS, Rudnick PA, Stein SE (2014) Tandem mass spectral libraries of peptides in digests of individual proteins: human serum albumin (HSA). Mol Cell Proteomics 13:2435–2449

5. Bhonsle HS, Singh SK, Srivastava G, Boppana R, Kulkarni MJ (2008) Albumin competitively inhibits glycation of less abundant proteins. Protein Pept Lett 15:663–667
6. Ikeda K, Higashi T, Sano H, Jinnouchi Y, Yoshida M, Araki T, Ueda S, Horiuchi S (1996) N-epsilon-(carboxymethyl)lysine protein adduct is a major immunological epitope in proteins modified with advanced glycation end products of the Maillard reaction. Biochemistry 35:8075–8083
7. Bhonsle HS, Korwar AM, Kesavan SK, Bhosale SD, Bansode SB, Kulkarni MJ (2012) "Zoom--In"—a targeted database search for identification of glycation modifications analyzed by untargeted tandem mass spectrometry. Eur J Mass Spectrom (Chichester) 18:475–481
8. Korwar AM, Vannuruswamy G, Jagadeeshaprasad MG, Jayaramaiah RH, Bhat S, Regin BS, Ramaswamy S, Giri AP, Mohan V, Balasubramanyam M, Kulkarni MJ (2015) Development of diagnostic fragment ion library for glycated peptides of human serum albumin: targeted quantification in prediabetic, diabetic, and microalbuminuria plasma by parallel reaction monitoring, SWATH, and MSE. Mol Cell Proteomics 14(8):2150–21599
9. Jagadeeshaprasad MG, Batkulwar KB, Meshram NN, Tiwari S, Korwar AM, Unnikrishnan AG, Kulkarni MJ (2016) Targeted quantification of N-1-(carboxymethyl) valine and N-1-(carboxyethyl) valine peptides of β-haemoglobin for better diagnostics in diabetes. Clin Proteomics 13(1):1

Chapter 29

Absolute Quantification of Middle- to High-Abundant Plasma Proteins via Targeted Proteomics

Julia Dittrich and Uta Ceglarek

Abstract

The increasing number of peptide and protein biomarker candidates requires expeditious and reliable quantification strategies. The utilization of liquid chromatography coupled to quadrupole tandem mass spectrometry (LC-MS/MS) for the absolute quantitation of plasma proteins and peptides facilitates the multiplexed verification of tens to hundreds of biomarkers from smallest sample quantities. Targeted proteomics assays derived from bottom-up proteomics principles rely on the identification and analysis of proteotypic peptides formed in an enzymatic digestion of the target protein. This protocol proposes a procedure for the establishment of a targeted absolute quantitation method for middle- to high-abundant plasma proteins waiving depletion or enrichment steps. Essential topics as proteotypic peptide identification and LC-MS/MS method development as well as sample preparation and calibration strategies are described in detail.

Key words Targeted proteomics, Multiple reaction monitoring, Absolute quantification, Proteotypic peptide, Biomarker validation

1 Introduction

Reliable quantitation is a basic requirement for protein and peptide biomarker verification studies in human body fluids. To date, immunoassays are considered the gold standard for quantification of proteins [1]. Nevertheless, in addition to disadvantages like cross-reactivity and poor inter-laboratory comparability, major drawbacks of immunoassays are long, costly development times as well as a limitation to single-parameter analysis [2]. In recent years, especially latter issues were addressed by coupling liquid chromatography and quadrupole tandem mass spectrometry (LC-MS/MS) aiming multiparametric protein quantification independent from antibodies. Due to the restricted mass range of typically used triple quadrupole mass spectrometers, proteins cannot be analyzed in their native form. Instead the so-called proteotypic peptides formed by an enzymatic digestion of the target protein are used as

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_29,

surrogates in the LC-MS/MS analysis performed in multiple reaction monitoring (MRM) mode [3]. In a first dimension, the precursor peptide ion of a specific mass-to-charge ratio (*m/z*) is isolated and subsequently fragmented via collision-induced dissociation. In a second dimension, a limited number of sequence-specific fragment ions are analyzed. The monitoring of such peptide-specific mass transitions provides the basis for the high specificity of targeted proteomics assays in complex sample matrices. However, the choice of suitable peptides, which have to be unique to the protein of interest in the investigated proteome, is one of the most important steps in the development of such assays that can be ready for use within a few weeks. Thereby, absolute protein quantification is dependent on the implementation of internal standards which can compensate for sample losses during sample preparation and matrix effects in the ionization process. For this purpose, mostly stable isotope labeled (SIL) synthesized analogs of the proteotypic peptides are used in combination with an external calibration to ensure a solid quantitation.

The presented protocol to targeted protein quantification by LC-MS/MS can be easily implemented for middle- to high-abundant plasma proteins without the need for enrichment or depletion steps requiring only smallest sample quantities [4].

2 Materials

2.1 Reagents and Consumables

1. Solvents.
   (a) ULC-MS grade methanol, acetonitrile, and 2-propanol.
   (b) LC grade ethanol.
   (c) Ultrapure deionized water.
2. Tryptic digestion.
   (a) 100 mM ammonium bicarbonate in water.
   (b) 2,2,2-trifluoroethanol.
   (c) 20 mM *tris*(2-carboxyethyl)phosphine hydrochloride in 100 mM ammonium bicarbonate.
   (d) 50 mM *N*-Ethylmaleimide in ethanol (*see* **Note 1**).
   (e) Sequencing grade modified trypsin (Promega, #V5111, #V5117, #V5113) (*see* **Note 2**).

   Dissolve lyophilized trypsin (#V5111, #V5117) or dilute trypsin solution (#V5113) with accompanying trypsin resuspension buffer to a final concentration of $0.1\overline{3}$ g/L.
   (f) 2% formic acid in water.
   (g) 0.5 mL safe-lock tube.

3. Solid phase extraction
   (a) Oasis HLB 1 cc flangeless vac cartridge, 10 mg sorbent per cartridge, 30 μm particle size (Waters).
   (b) 1.5 mL micro tube.
   (c) Buffer A: 0.1% trifluoroacetic acid in water.
   Buffer B: 0.1% trifluoroacetic acid in acetonitrile/water (80/20, v/v).
4. LC-MS/MS analysis.
   Short thread vial including slotted cap (WICOM).

2.2 HPLC and Mobile Phases

1. HPLC instrument: microLC system (e.g., Ultimate 3000 RSLCnano System, Thermo Scientific Dionex).
2. Separation column: ZORBAX 300SB-C18 (150 × 1.0 mm id, 3.5 μm particle size) with corresponding guard column.
3. Mobile phases.
   (a) Eluent A: 0.1% formic acid in methanol/water (10/90, v/v).
   (b) Eluent B: 0.1% formic acid in methanol/water (90/10, v/v).

2.3 Mass Spectrometer

1. Mass spectrometer: hybrid triple quadrupole/linear ion trap mass analyzer (e.g., QTRAP® 5500, SCIEX).
2. Ionization source: electrospray ionization.
3. Software: Analyst (SCIEX).

2.4 Data Processing and Analysis

1. Analyst (SCIEX).
2. Skyline (MacCoss Lab).
3. MultiQuant (SCIEX).

2.5 Synthetic Peptide Standards

1. Proteotypic peptide standards of interest.
2. SIL (^{13}C, ^{15}N) analogs of the proteotypic peptides of interest (*see* **Note 3**).

2.6 Human Specimen

For method development, a human specimen of the desired matrix is required. Serum or EDTA plasma is possible.

3 Methods

The following protocol summarizes basic steps in the method development of a quantitative proteomics assay. Starting from an in silico identification of proteotypic peptides and their subsequent verification in human specimens to data processing, all mandatory requirements for an absolute quantification of middle- to high-abundant plasma proteins by LC-MS/MS are described.

3.1 In Silico Identification of Proteotypic Peptides

The purpose of this section is the identification of proteotypic peptides. Basically, this can be performed in silico or empirically by the analysis of digested recombinant proteins of interest. Below a detailed procedure for an in silico proteotypic peptide identification is listed (*see* **Note 4**).

1. Search the protein of interest in a universal protein database (www.uniprot.org). Specify the concerned organism (human).
   (a) Identify the chain of the protein of interest and sites of modified amino acids in paragraph "PTM/Processing" (*see* **Note 5**). Determine natural variants and polymorphisms in paragraph "Sequence." Consider isoforms of the target protein.
   (b) Choose "Peptide cutter" in paragraph "Sequence" to perform an in silico digestion. Choose your enzyme of choice from the given list. Trypsin is used most commonly (*see* **Note 6**). Select "Table of sites, sorted sequentially by amino acid number" to generate a list of formed peptides.
   (c) Remove peptides with modified amino acids or known natural variants and polymorphisms from the resulting list. Proceed in the same way with peptides which are less than seven amino acids long and include methionine residues. Cysteine residues also prone to oxidation can be part of a proteotypic peptide if a complete alkylation can be confirmed during sample preparation. Disease-related peptide modifications and variants might be an option for monitored peptides. Additional guidelines on the selection of proteotypic peptides were described before [5].
2. Perform a protein BLAST (basic local alignment search tool) on http:// blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins.
   (a) Enter potential proteotypic peptide sequences one after another in single-letter amino acid code. Choose "Non-redundant protein sequences (nr)" as database and specify "human (taxid:9606)" as investigated organism.
   (b) Significant alignments of the input query and similar database sequences are provided in the list of results. In case of a proteotypic peptide, both query cover and amino acid identity have to be at 100% exclusively for the expected protein (*see* **Note 7**). Further, gaps must not be detected. The E-value, describing the number of BLAST alignments one would expect to see by chance with the observed score or higher, has to be considered with regard to the length of the investigated peptide sequence (*see* **Note 8**).

3.2 Verification of Proteotypic Peptides in Human Specimens

A verification of proteotypic peptides identified *in silico* is performed by the use of digested human specimens, preferably in combination with a hybrid triple quadrupole/linear ion trap mass spectrometer as described below (*see* **Note 9**). Several commercial

as well as open-source software solutions for the simulation of multiple reaction monitoring (MRM) mass transitions (e.g., Skyline, MacCoss Lab) are available. Additionally, to the conformation of proteotypic peptides, this procedure enables a first assessment of the analyzability of the target protein in human blood applying the chosen method parameters.

1. SCIEX instrument settings for Skyline can be downloaded from MacCoss lab. Modulate Skyline transition settings and choose the most suitable setup for your instrumentation:
   (a) Collision energy: ABI 5500 QTrap.
   (b) Declustering potential: ABI.
   (c) Optimize by: Transition.
   (d) Precursor charges: 2, 3.
   (e) Ion charges: 1, 2.
   (f) Filter product ions from "*m/z* > precursor" to "3 ions" (*see* **Note 10**).
2. Copy potential proteotypic peptides identified in Subheading 3.1 to Skyline.
3. Export the Skyline transition list as a single method without optimization options.
4. Open the IDA Method Wizard in the Analyst software to generate an information dependent acquisition method for the acquisition of enhanced product ion (EPI) scans. Use the following options:
   (a) Select "Type of IDA Experiment: MRM >> Enhanced Product."
   (b) Select "Positive Mode."
   (c) Declustering Potential (DP): 100.
   (d) Resolution Q1/Q3: UNIT.
   (e) Select "Manually Enter MRM Transitions."
   (f) Select "Apply CE to all MRM."
   (g) Enter the MRMs exported from Skyline.
   (h) Select "Use Rolling Collision Energy."
   (i) Consider signals for fragmentation, if the peak intensity exceeds 2000 cps.
   (j) Exclude former target ions for 30 s if they have occurred for at least three times previously.
5. Create an HPLC gradient and use gas and temperature settings according to Subheading 3.3, **step 13**.
6. Prepare a human plasma or serum sample according to Subheading 3.5. Use 100 mM ammonium bicarbonate instead of the internal standard mix in **step 2**.

7. Depending on the abundance of the protein of interest, inject a reasonable amount of sample (1–20 μL) and run the created method.
8. Search the generated raw data with an algorithm like MASCOT MS/MS Ions Search (http://www.Matrixscience.Com/Cgi/Search_form.Pl?FORMVER=2&SEARCH=MIS; Matrix Sciences) using appropriate search parameter.
   (a) Database: Swiss-Prot.
   (b) Select the enzyme used for digestion.
   (c) Do not allow missed cleavages.
   (d) Taxonomy: *Homo sapiens* (human)
   (e) Select "Display all modifications" and choose "*N*-Ethylmaleimide (C)" as fixed modification (*see* **Notes 1** and **11**).
   (f) Peptide tol. ±: 0.5 Da.
   (g) MS/MS tol. ±: 0.3 Da.
   (h) Peptide charge: 2+ and 3+.
   (i) Select the appropriate instrument, e.g., ESI-TRAP.
9. Examine the list of protein hits and identified peptides.
   (a) Compare detected peptide sequences with the list of potential proteotypic peptides identified in silico.
   (b) Assess the score of each detected peptide with regard to defaults for the indication of significant homology or extensive homology according to applied MASCOT search parameters.
   (c) Check if detected peptides are defined as unique.
10. Explore the chromatogram for one to three proteotypic peptides with high peak intensities and signal-to-noise ratios as well as good chromatographic peak shape and proceed with the steps described below.

3.3 Selection of Mass Transitions and HPLC Optimization

After the selection of appropriate proteotypic peptides, synthetic standards of those as well as SIL analogs have to be prepared which are used for method development and calibration (*see* **Note 3**). Numerous suppliers are available. After MS tuning and selection of the best mass transitions, the HPLC part of the method has to be optimized. Below a detailed procedure applying a hybrid triple quadrupole-linear ion trap instrument is described (*see* **Note 9**):

1. Calculate potential fragment ions of the proteotypic peptides, especially y- and b-ions, e.g., by the use of the free web tool fragment ion calculator of the Institute for Systems Biology (http://Db.Systemsbiology.Net:8080/proteomicsToolkit/FragIonServlet.Html).

2. Dissolve peptide standards in the recommended solvent, e.g., water/2-propanol (1:1, v/v), to yield stock solutions of 1–10 mmol/L. Apply ultrasound or add trifluoroacetic acid to a final concentration of 1%, if standards are less soluble.
3. Prepare working standards in 100 mM ammonium bicarbonate.
4. Prepare tuning solutions of 200 nmol/L or lower in methanol/water (1:1, v/v) + 0.1% FA.
5. Activate the tune mode and directly infuse the tuning solution at 10 µL/min using the following parameters:
   (a) Curtain gas: 20 psi.
   (b) Collision gas: High.
   (c) Nebulizer gas: 20 psi.
   (d) Heater gas: 0 psi.
   (e) Ion spray voltage: 5500 V.
   (f) Source temperature: 0 °C.
   (g) Polarity: Positive.
6. Perform a Q1 scan to identify the precursor ion.
7. Check the assumed charge state of the precursor ion in an enhanced resolution scan (*see* **Note 12**).
8. Tune the declustering potential in a Q1 multiple ions scan.
9. Identify fragment ions in an EPI scan performed with ramped collision energy in multichannel analysis (MCA) mode. Compare detected ions with the list of expected fragment ions identified before.
10. Select five to eight of the most intense fragment ions to obtain a sensitive assay. Fragment ions with greater *m/z* than the *m/z* of the precursor ion should be preferred (*see* **Note 10**).
11. Tune the collision energy and the collision cell exit potential in a MRM scan.
12. Optimize curtain gas, nebulizer gas, heater gas, ion spray voltage, source temperature, and collision gas by infusion of the peptide standard solution in a constant flow of eluent operated under HPLC conditions via a tee. Use a composition of eluents which corresponds to the conditions at the time of elution of the proteotypic peptide. Choose a flow rate which will be applied in the final method. If necessary, repeat this step during HPLC optimization.
13. Create a MRM method containing the optimized mass transitions and MS parameters or use the method applied for the verification of proteotypic peptides (Subheading 3.2). A typical microLC gradient as well as standard gas and temperature settings are given below (*see* **Note 13**).

Linear increase from 20% to 100% B in 3.5 min, 3 min 100% B, 1 min re-equilibration with initial conditions applying a flow rate of 50 µL/min at 40 °C.

(a) Curtain gas: 35 psi.

(b) Collision gas: High.

(c) Nebulizer gas: 20 psi.

(d) Heater gas: 50 psi.

(e) Ion spray voltage: 5500 V.

(f) Source temperature: 400 °C.

14. Run a digested human specimen. Identify two to three mass transitions per peptide which have high signal intensities and high signal-to-noise ratios at the same time. Signals of the proteotypic peptide should have the same retention time for all monitored mass transitions. Prefer transitions with greater *m/z* of the fragment ion compared to the *m/z* of the precursor ion (*see* **Note 10**). Use the best transition for quantitation. Use the same transitions for analyte and corresponding internal standard regarding precursor ion charge and monitored fragment ion.

15. Optimize HPLC parameters and settings including analytical column, eluents, modifier, gradient, flow rate, column temperature, and sample volume. Refer to basic literature on HPLC method development for guidance [6].

3.4 Calibration, Internal Standardization, and Quality Controls

One of the simplest and most quickly implementable approaches for the absolute quantification of proteins by means of LC-MS/MS is the combined application of peptide calibration standards and SIL proteotypic peptide analogs for internal standardization (*see* **Note 14**). Avoid multiple freeze-thaw cycles and store calibrators, internal standard mix, and quality controls aliquoted, ready for use.

1. Calibration.

   (a) Search the literature for reference values or published concentration levels of the target protein in human plasma or serum.

   (b) Based on the literature research, prepare a standard calibration curve of at least four calibration points. The averaged expected concentration in human specimens should be in the middle of the chosen calibration points. Blank matrix should be used. Alternatively, aqueous calibration standards can be prepared.

   (c) Perform a complete sample preparation procedure for the calibration standards and a human specimen. Run the samples using the developed MRM method.

   (d) Test the linearity of the calibration curve and compare peak areas of calibration standards and the human specimen. Adapt the calibration standards if necessary.

2. Internal standard mix.

   Prepare an internal standard mix in 100 mM ammonium bicarbonate with concentrations similar to expected concentrations of the target protein in human specimens. Note that signal intensity and peak area of the internal standard should be comparable to the ones of the monitored proteotypic peptide.

3. Quality controls.

   For quality assessment and guarantee of batch comparability, quality controls should be carried along in the analytical process. Commercially available quality control materials for proteins are rare. Due to this fact, it might be necessary to prepare in-house controls. Preferably, two to three concentration levels should be investigated (low, middle, high). Similar to the preparation of calibration standards, blank matrix should be preferred over aqueous solutions like ammonium bicarbonate.

   (a) Peptide standards can be used for the preparation of quality controls if protein standards are unavailable or too expensive. It is important to note that the digestion process cannot be controlled by such controls.

   (b) Human specimens, which should be stored as aliquots to avoid freeze-thaw cycles, can be used as quality control. Ring trial materials or the like with known target values can also be used.

3.5 Preparation of Human Specimens, Calibrators, and Control Material

1. Add 3 µL plasma/serum or calibrator or control material to a safe-lock tube.
2. Add 3 µL internal standard mix (*see* **Note 15**).
3. Slowly add 6 µL 2,2,2-trifluoroethanol.
4. Add 4 µL of 20 mM *tris*(2-carboxyethyl)phosphine and close the tube tightly. Incubate at 60 °C for 30 min with moderate shaking.
5. After a cooling phase of 5 min, spin down the condensate. Subsequently, add 4 µL of 50 mM *N*-Ethylmaleimide and incubate at room temperature for 30 min with moderate shaking.
6. Add 180 µL of 100 mM ammonium bicarbonate.
7. Add 60 µL of 0.13 g/L trypsin solution (8 µg, enzyme-to-protein ratio ~ 1:30), close the tube tightly, and incubate at 37 °C for 16 h with moderate shaking (*see* **Note 16**).
8. Spin down the condensate and acidify the sample with 14 µL of 2% formic acid to stop digestion.
9. Solid phase extraction (*see* **Note 17**).

   (a) Place an Oasis HLB 1 cc flangeless vac cartridge in a centrifuge tube.

   (b) Add 500 µL of Buffer B to the cartridge and centrifuge for 1 min at 270 × *g*. Repeat this procedure for another time.

(c) Wash the cartridge with 500 µL of Buffer A and centrifuge for 1 min at 270 × *g*. Retry this procedure two more times.

(d) Transfer the complete sample volume to the cartridge. Centrifuge for 1 min at 270 × *g*.

(e) Rinse the sample vessel with 200 µL Buffer A and repeat the preceding step.

(f) Wash the cartridge with 500 µL Buffer A and centrifuge for 1 min at 270 × *g*. Repeat this procedure for another time and centrifuge for 2 min.

(g) Place the cartridge in a 1.5 mL micro tube located in a centrifuge tube and add 200 µL Buffer B and centrifuge for 2 min at 270 × *g*.

(h) Add 200 µL Buffer B and centrifuge for 3 min at 3220 × *g*.

(i) Dry the sample under a nitrogen stream.

(j) Reconstitute the sample in 1 mL Eluent A and seal the tube tightly. Place the sample in an ultrasonic bath for 2 min. Subsequently, centrifuge for 5 min at 13,000 × *g*.

(k) Transfer the supernatant to a clean short thread vial.

3.6 Sample Analysis

1. Run the developed LC-MS/MS method (Subheading 3.3).
2. Run the external standard calibration set including a blank at the very beginning of each batch. Several aliquots of a quality control should be carried along in the sample preparation procedure and be measured accordingly. The measurement of the external standard calibration set should be repeated at the end of each batch for quality assessment of the analysis.

3.7 Data Analysis

Use data processing software like MultiQuant (SCIEX) with standardized peak finding algorithms (*see* **Note 18**).

1. Manually verify that the transitions of a proteotypic peptide and its corresponding SIL internal standard elute at the same retention time.
2. Check the peak integration. Peaks should only be considered for integration if the signal-to-noise ratio is ≥3.
3. Generate a linear calibration curve, where the peak area ratio of analyte to internal standard is plotted against the concentration ratio of analyte to internal standard. Use a weighting factor of $1/x$. Ensure the linearity of the calibration curve and check the accuracy of the calibrators (*see* **Note 18**).
4. Ensure that the peak area of the internal standard is constant throughout the complete batch of samples.

3.8 Validation of the Analytical Process

Assess the validity of the complete method including sample preparation with the following experiments:

1. Determine the limit of detection (LOD) and lower the limit of quantification (LLOQ) in a serial dilution or more appropriately by spiking blank matrix according to concentration levels. By the use of a serial dilution, the method's linearity can be assessed simultaneously.
   (a) A signal-to-noise ratio of 3 is commonly accepted for LOD estimation.
   (b) LLOQ is defined as the analyte concentration which can be measured with a coefficient of variation <20%.
2. Assess within-day and between-day precision at two to three concentration levels.
   (a) Within-day precision: Prepare ten replicates per sample and measure them in one run.
   (b) Between-day precision: Prepare and measure one replicate per sample in a single run on ten consecutive working days.
3. Determine the recovery rate by spiking specified amounts of protein standard in blank matrix or human material. Analyze these samples as well as the starting material. Compare the difference in calculated concentration of spiked sample and concentration of starting material with the amount of analyte added.

4 Notes

1. The presented protocol uses *N*-Ethylmaleimide for alkylation, while in proteomics research most commonly photosensitive iodoacetamide is used as alkylating agent. Both agents can be used at a final concentration of 10 mM [4, 7].
2. The origin of trypsin influences the variability of tryptic digestions. An increased formation of miscleaved peptides was observed for the application of bovine trypsin. The usage of porcine trypsin might result in a higher percentage of semi-tryptic peptides. The introduction of modifications suppresses trypsin autolysis while a tosyl phenylalanyl chloromethyl ketone (TPCK) treatment inactivates chymotrypsin. Furthermore, trypsin performance varies between manufacturers [8].
3. The incorporation of stable isotopes should result in a mass shift of at least three to four Dalton [9]. The application of ^{13}C and ^{15}N should be preferred over ^{2}H since deuterated peptides can separate in most commonly used reversed-phase chromatography [10].
4. Literature on the target protein should be reviewed first to find proteotypic peptides used and published before. Furthermore,

auxiliary web tools like the MRM peptide picking tool (http://mrmpeptidepicker.proteincentre.com/peptidepicker9/; University of Victoria Genome British Columbia Proteomics Centre) can be used for proteotypic peptide identification [11].

5. The signal peptide should not be considered for a search for proteotypic peptides of plasma proteins since this signal sequence is cleaved prior to protein secretion. Consequently, the molecular mass of the secreted target protein differs from the one given in databases which consider the complete protein sequence.
6. The protease trypsin offers several advantages which explain its popularity in quantitative proteomics applications:
   (a) Trypsin has a highly specific cleavage pattern with cleavage sites lysine and arginine which are abundant in the human proteome and well distributed throughout the proteins resulting in averaged peptide lengths of nine amino acids [12, 13].
   (b) Tryptic peptides preferentially generate doubly charged precursor ions in electrospray ionization well suited for detection in triple quadrupole mass spectrometers [14].
   (c) Trypsin is available in various levels of purity and price categories.
7. In case a protein other than the targeted one results with a 100% query cover, it should be checked if the peptide can be formed during an enzymatic digestion of the second protein in question.
8. In a BLAST search, low E-values indicate that the results obtained are not due to chance but rather due to a biologically meaningful correlation of query and database sequence. However, the E-value depends on query and database length. An E-value of 10^{-6} is used by NCBI for internal processes which might be too restrictive for short peptide sequences.
9. All MS experiments can be performed in a modified way on other MS platforms, e.g., triple quadrupole mass spectrometers or quadrupole time of flight mass spectrometers. However, specific experiments like enhanced product ion scans using the ion trap function for ion accumulation can only be performed by the use of hybrid triple quadrupole/linear ion trap mass spectrometers.
10. Due to the fact that tryptic peptides most commonly generate multiple-charged precursor ions, less-charged fragment ions can have higher m/z than the corresponding precursor. The usage of such transitions is favored since background noise is decreased.

11. If iodoacetamide is used as alkylating agent, "Carbamidomethyl (C)" has to be selected as fixed modification in a MASCOT search.
12. The charge state of a peptide can be determined from the isotope ratio. The difference of the monoisotopic peak and its following peak is $1/z$.
13. According to our experience, microLC coupled to MS/MS is a robust system for the peptide quantitation. In contrast, nanoLC applications are frequently prone to failure, e.g., due to undetected volume leaks or column clogging.
14. Instead of SIL peptides, more expensive labeled recombinant proteins can be used for a more accurate internal standardization. Such SIL proteins as well as "winged" peptides which include at least two endoproteolytic cleavage sites reproduce variabilities in the enzymatic digestion that are not accounted for by the use of standard SIL proteotypic peptide analogs [15]. However, such recombinant proteins might not behave identical to the endogenous protein due to missing posttranslational modifications or glycosylation sites [16].
15. The time of SIL peptide addition is of high importance for an accurate protein quantitation. If SIL peptides are used, from internal standard addition to enzymatic formation of proteotypic peptides, unspecific cleavages or adsorption can alter the ratio of analyte to internal standard. Nevertheless, an early addition of SIL peptides prior to digestion should be preferred over a concurrent or post-digest addition to ensure accurate protein quantification [17].
16. The time needed for a complete digestion of the target protein has to be identified empirically. Digest a human specimen for at least 24 h, whereby digestion is stopped after specified periods, e.g. 30 min, 1 h, 2 h, 4 h, 8 h, 12 h, 16 h and 20 h, to determine digestion kinetics. A complete digestion has to be ensured if SIL peptides are used for internal standardization which do not account for digestion variabilities [7].
17. By the use of a second column, the solid-phase extraction can be performed online to increase automation.
18. Data processing can also be performed by the use of simple calculation software like excel:
    (a) Export integrated peak areas of the proteotypic peptide and its SIL analog using the most intense mass transition.
    (b) Calculate the area ratio of proteotypic peptide to SIL analog for all calibrators, unknowns, and controls.
    (c) Calculate the concentration ratio of proteotypic peptide to SIL analog for all calibrators.
    (d) Plot the calculated area ratio against the concentration ratio for all calibrators.

(e) Create a linear trend line according to the equation $y = mx + n$.

(f) Solve the resulting equation for x and substitute calculated area ratios of unknowns and controls (y).

Acknowledgment

This publication is supported by LIFE—Leipzig Research Center for Civilization Diseases, Universität Leipzig. LIFE is funded by means of the European Union, by the European Regional Development Fund (ERDF), and by the Free State of Saxony within the framework of the excellence initiative.

References

1. Kuzyk MA, Smith D, Yang J et al (2009) Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Mol Cell Proteomics 8:1860–1877
2. Hoofnagle AN, Wener MH (2009) The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. J Immunol Methods 347:3–11
3. Mallick P, Schirle M, Chen SS et al (2007) Computational prediction of proteotypic peptides for quantitative proteomics. Nat Biotechnol 25:125–131
4. Ceglarek U, Dittrich J, Becker S et al (2013) Quantification of seven apolipoproteins in human plasma by proteotypic peptides using fast LC-MS/MS. Proteomics Clin Appl 7:794–801
5. Lange V, Picotti P, Domon B et al (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol 4:222
6. Meyer VR (2010) Practical high-performance liquid chromatography. John Wiley & Sons, Ltd, Chichester
7. Dittrich J, Becker S, Hecht M et al (2015) Sample preparation strategies for targeted proteomics via proteotypic peptides in human blood using liquid chromatography tandem mass spectrometry. Proteomics Clin Appl 9:5–16
8. Walmsley SJ, Rudnick PA, Liang Y et al (2013) Comprehensive analysis of protein digestion using six trypsins reveals the origin of trypsin as a significant source of variability in proteomics. J Proteome Res 12:5666–5680
9. Ong S-E, Mann M (2005) Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol 1:252–262
10. Zhang R, Regnier FE (2002) Minimizing resolution of isotopically coded peptides in comparative proteomics. J Proteome Res 1:139–147
11. Mohammed Y, Domański D, Jackson AM et al (2014) PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. J Proteome 106:151–161
12. Brownridge P, Beynon RJ (2011) The importance of the digest: proteolysis and absolute quantification in proteomics. Methods 54:351–360
13. Vandermarliere E, Mueller M, Martens L (2013) Getting intimate with trypsin, the leading protease in proteomics. Mass Spectrom Rev 32:453–465
14. Burkhart JM, Schumbrutzki C, Wortelkamp S et al (2012) Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. J Proteome 75:1454–1462
15. Carr SA, Abbatiello SE, Ackermann BL et al (2014) Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. Mol Cell Proteomics 13:907–917
16. Huillet C, Adrait A, Lebert D et al (2012) Accurate quantification of cardiovascular biomarkers in serum using protein standard absolute quantification (PSAQ) and selected reaction monitoring. Mol Cell Proteomics 11:M111.008235–M111.008235
17. Shuford CM, Sederoff RR, Chiang VL et al (2012) Peptide production and decay rates affect the quantitative accuracy of protein cleavage isotope dilution mass spectrometry (PC-IDMS). Mol Cell Proteomics 11:814–823

Part VI

Developments in Biomarker Discovery

Chapter 30

A Highly Automated Shotgun Proteomic Workflow: Clinical Scale and Robustness for Biomarker Discovery in Blood

Loïc Dayon, Antonio Núñez Galindo, Ornella Cominetti, John Corthésy, and Martin Kussmann

Abstract

With recent technological developments, protein biomarker discoveries directly from blood have regained interest due to elevated feasibility. Mass spectrometry (MS)-based proteomics can now characterize human plasma proteomes to a greater extent than has ever been possible before. Such deep proteome coverage comes, however, with important limitations in terms of analysis time which is a critical factor in the case of clinical studies. As a consequence, compromises still need to be made to balance the proteome coverage with realistic analysis time frame in clinical research. The analysis of a sufficient number of samples is compulsory to empower statistically robust candidate biomarker findings. We have, therefore, recently developed a scalable automated proteomic pipeline (ASAP^2) to enable the proteomic analysis of large numbers of plasma and cerebrospinal fluid (CSF) samples, from dozens to a thousand of samples, with the latter number being currently processed in 15 weeks. A distinct characteristic of ASAP^2 relies on the possibility to prepare samples in a highly automated way, mostly using 96-well plates. We describe herein a sample preparation procedure for human plasma that includes internal standard spiking, abundant protein removal, buffer exchange, reduction, alkylation, tryptic digestion, isobaric labeling, pooling, and sample purification. Other key elements of the pipeline (i.e., study design, sample tracking, liquid chromatography (LC) tandem MS (MS/MS), data processing, and data analysis) are also highlighted.

Key words Mass spectrometry, Plasma, Human, Clinical research, Large scale, Biomarker, Isobaric tagging, Depletion, Automation

1 Introduction

Human plasma is the preferred body fluid for clinical chemistry investigation because of its minimally invasive collection and relatively simple treatment procedure. Molecules measured in plasma are, for instance, universally used to support and take informed decisions in terms of diagnosis and prognosis as well as intervention monitoring [1]. With the advent of deep molecular phenotyping, human plasma constitutes a rich source of potentially novel biomarkers.

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_30,

In the past 20 years, proteomic analyses have generated a plethora of putative biomarker candidates [2]. Unfortunately, very few markers discovered via proteomic routes are used in today's clinical practice [3]. Several limitations have been argued to explain the pitfalls of proteomic pipelines in delivering usable biomarkers [4]. Among those, the limited sample size of the initial discovery studies pertains to the generation of false-positive and false-negative hits. Under such circumstances, verification and validation of discovery findings often become challenging, highly time-consuming, and expensive, especially if biomarkers are eventually discarded.

By allowing discoveries in large cohorts of individuals, as usually performed with genomic and increasingly with metabolomic technologies, proteomics may also reach its full potential in clinical research. Analytical throughput is a compulsory enabler, but mass spectrometry (MS)-based proteomic workflows present today various limitations in that regard [5], despite being able to deeply characterize human plasma proteomes [6]. Sample preparation, analyte separation, MS acquisition, data processing, and data analysis are all key components of the MS-based proteomic workflow that should be improved toward higher throughput, automation, and integration. In this chapter, we focus mainly on one of these modules, i.e., proteomic sample preparation of human plasma in a scalable and highly automated way.

In MS-based shotgun proteomics, blood plasma analysis requires typical sample preparation steps, including (1) spiking of standard proteins for quality control (QC) and normalization purpose, (2) removal of highly abundant proteins that otherwise mask the less abundant ones during MS analysis, (3) protein digestion into peptides, and (4) purification of the samples to remove the excess of liquid chromatography (LC) MS interfering reagents (*see* Fig. 1). The discovery of biomarkers is performed by quantitative comparisons of samples (e.g., from different patients, groups thereof, from different time points, etc.). We have therefore incorporated isobaric labeling (i.e., tandem mass tag (TMT) technology [7, 8]) in the procedure to enable relative protein quantification between conditions (*see* Fig. 1). Herein, we describe a method where all those elements have been integrated into a highly automated workflow, the so-called ASAP^2 (standing for a scalable automated proteomic pipeline) [9], that allows blood plasma proteomics at clinical research scale and enables high robustness [10].

2 Materials

2.1 Depletion

1. Tubes of 1.4 and 1.0 mL, screw caps, and 2D barcode reader.
2. Water (18.2 MΩ cm at 25 °C) obtained with a Milli-Q apparatus (Millipore).

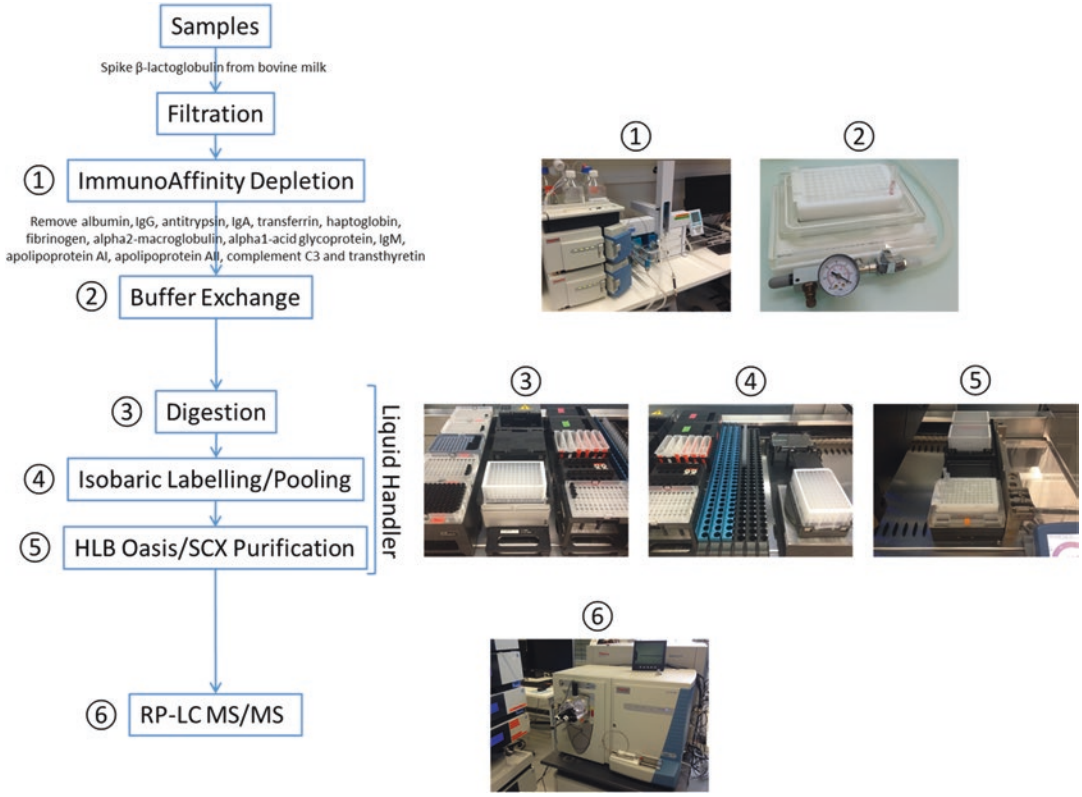


Fig. 1 General scheme of the ASAP^2 workflow

3. Sterile, clear 96-well filter plate with 0.22 μm pore size PVDF membrane.
4. *Vacuum* manifold (Millipore).
5. Polypropylene 96-well plate (V-bottom).
6. Adhesive mat.
7. Multiple affinity removal system (MARS) column Human 14, Buffer A, and Buffer B (Agilent Technologies).
8. β-Lactoglobulin (LACB) from bovine milk prepared at 0.0134 mg/mL in Buffer A.
9. High-performance LC (HPLC) system (Thermo Scientific).
10. HTC-PAL equipped with matrix-assisted laser desorption/ionization (MALDI) option fraction collection system (CTC Analytics AG).
11. Liquid nitrogen.
12. Vibrating platform (Heidolph Instruments).
13. Twelve-channel 10–100 μL multi-pipette.
14. Freezer at −80 °C.

2.2 Buffer Exchange

1. 2D barcode reader and tube capper/decapper (FluidX).
2. Strata-X 33u polymeric reversed-phase (RP) (30 mg/1 mL) cartridges (Phenomenex).
3. *Vacuum* manifold and 96-hole holder (Phenomenex).
4. Water (18.2 MΩ cm at 25 °C).
5. Solution of CH_3CN/0.08% trifluoroacetic acid (TFA).
6. Solution of water/0.1% TFA.
7. Solution of 30% water/70% CH_3CN/0.08% TFA.
8. Vibrating platform (Heidolph Instruments).
9. Twelve-channel 50–1200 µL multi-pipette.
10. Two milliliters 96-DeepWell plate and polypropylene mat.
11. Centrifuge (Eppendorf, Hamburg, Germany), SpeedVac system (Thermo Scientific), or equivalent units.
12. Freezer at −80 °C.

2.3 Reduction, Alkylation, and Proteolytic Digestion of Proteins

1. Water (18.2 MΩ cm at 25 °C).
2. Solution of 100 mM triethylammonium hydrogen carbonate buffer (TEAB) pH 8.5 in water.
3. Solution of 2% sodium dodecyl sulfate (SDS) in water (*w*/*V*).
4. Solution of 20 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP) in water.
5. Solution of 150 mM iodoacetamide (IAA, ≥99%) in CH_3CN.
6. Solution of 0.25 µg/µL sequencing grade modified trypsin/Lys-C (Promega) in 100 mM TEAB.
7. One milliliter polypropylene tubes.
8. Four-channel Microlab Star liquid handler (Hamilton).
9. Temperature-controlled mixer, temperature-controlled dark chamber, reservoirs, and tube racks mounted on the liquid handler deck (*see* Fig. 2).

2.4 TMT Labeling

1. TMT reagents (Thermo Scientific).
2. Water (18.2 MΩ cm at 25 °C).
3. Solution of 5% hydroxylamine (*w*/*V*) in water from hydroxylamine solution 50 wt. % in water (99.999%).
4. Solution of 95% water/5% CH_3CN/0.1% TFA (RP loading buffer).
5. One and 0.5 mL polypropylene tubes.
6. Five milliliters polypropylene tubes.
7. Four-channel Microlab Star liquid handler.

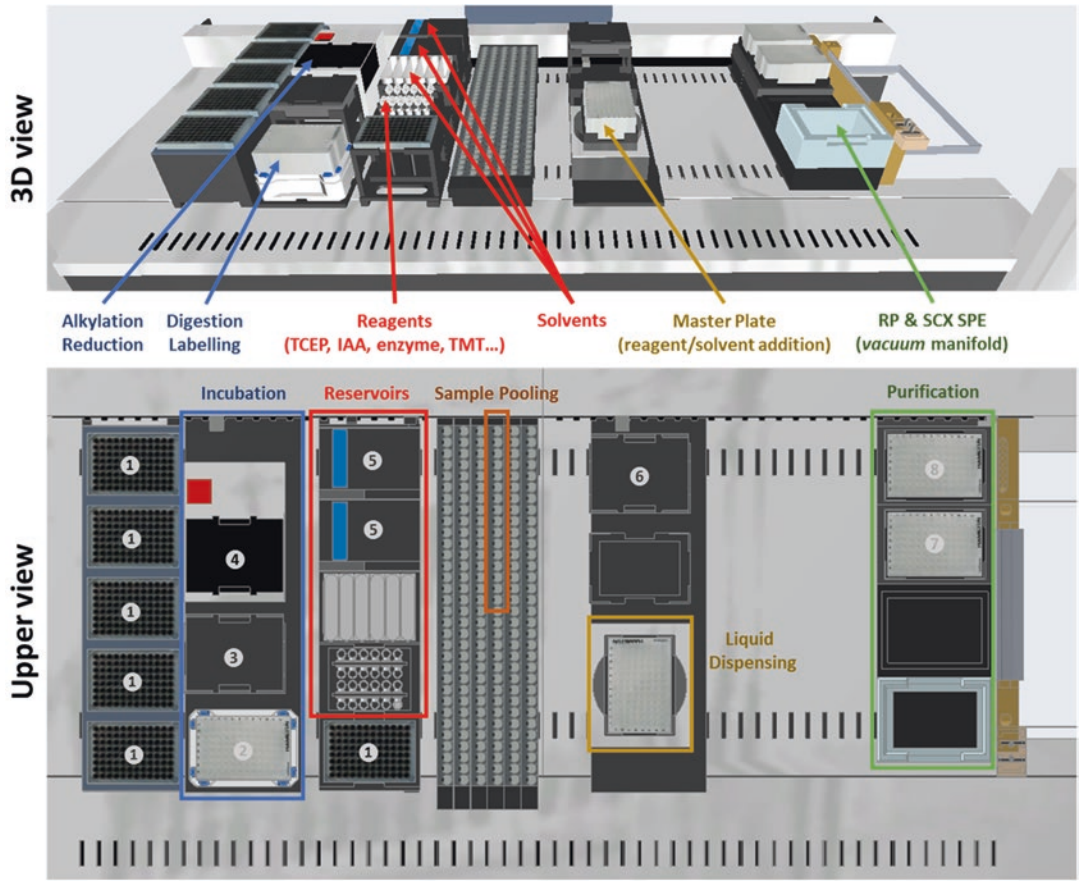


Fig. 2 Representations of the liquid handler deck. The main sample preparation step locations are indicated in the 3D view. In the upper view is displayed the general location of the functional blocks: tips are positioned in ①, plate heater and shaker in ②, cover for the dark chamber in ③, dark chamber in ④, reservoirs fed with stock bottles in ⑤, plastic cover for the 96-DeepWell plate in ⑥, RP elution 96-DeepWell plate in ⑦, and SCX elution 96-DeepWell plate in ⑧. Under the liquid handler deck, a pump and a *vacuum* pump are placed to allow, respectively, the filling of the reservoirs ⑤ and operating of the manifold. The liquid handler is controlled by a computer

8. Temperature-controlled mixer and tube racks mounted on the liquid handler deck (*see* Fig. 2).
9. Vortex mixer (Thermo Scientific), centrifuge, or equivalent units.

2.5 Sample Purifications

1. Oasis HLB 1 cc (1 cc, 30 mg) extraction cartridges (Waters) or an alternative phase support.
2. Strata-X-C 33u Polymeric Strong Cation (30 mg/1 mL) cartridges (Phenomenex).
3. Two milliliters 96-DeepWell plate.
4. Water (18.2 MΩ cm at 25 °C).

5. Solution of 5% water/95% CH_3CN/0.1% TFA (RP conditioning buffer).
6. Solution of 95% water/5% CH_3CN/0.1% TFA (RP loading buffer).
7. Solution of 50% water/50% CH_3CN/0.1% TFA (RP elution buffer).
8. Solution of 75% water/25% CH_3CN/0.1% TFA (SCX loading buffer).
9. Solution of 75% water/25% CH_3CN with 400 mM ammonium acetate (NH_4OAc) (SCX elution buffer).
10. Four-channel Microlab Star liquid handler from Hamilton.
11. *Vacuum* manifold mounted on liquid handler deck (*see* Fig. 2).
12. SpeedVac system.

2.6 RP-LC Tandem MS (MS/MS) Analysis

1. UltiMate 3000 RSLC nano-system (Thermo Scientific) or equivalent systems.
2. Acclaim PepMap 75 μm × 2 cm (C18, 3 μm, 100 Å) pre-column and Acclaim PepMap RSLC 75 μm × 50 cm (C18, 2 μm, 100 Å) analytical column (Thermo Scientific).
3. Stainless steel nanobore emitter (40 mm, OD 1/32″) mounted on a Nanospray Flex Ion Source (Thermo Scientific).
4. Water (18.2 MΩ cm at 25 °C) obtained with a Milli-Q apparatus (Millipore).
5. CH_3CN.
6. Solution of water 98%/CH_3CN 2%/formic acid (FA, 99%) 0.1% (LC solvent A).
7. Solution of water 20%/CH_3CN 80%/FA 0.08% (LC solvent B).
8. Solution of water 98%/CH_3CN 2%/TFA 0.05% (LC loading solvent).
9. LC vials (Infochroma, Zug, Switzerland).
10. Hybrid linear ion trap-Orbitrap (LTQ-OT) Elite (Thermo Scientific) or equivalent mass spectrometer.

2.7 Data Analysis

1. Proteome Discoverer (version 1.4) (Thermo Scientific).
2. Mascot (version 2.4.2) (Matrix Science, London, UK).
3. Human UniProtKB/Swiss-Prot database.
4. Scaffold Q+ 4.3.2 (graphical user interface (GUI) and/or ScaffoldBatch) (Proteome Software).
5. Microsoft Excel.
6. R version 3.1.1 (http://www.r-project.org/) and R package mixOmics version 5.0-3.
7. Prism (GraphPad Software).

3 Methods

3.1 General Practice

ASAP^2 was firstly developed with the aim of (1) discovering protein biomarkers directly in plasma (*see* **Note 1**) and (2) analyzing cohorts of 100s to 1000s of subjects/samples. It is composed of independent analytical steps, used by default or optionally in a modular fashion (*see* Fig. 1), according to the sample and/or experimental design. First, abundant protein removal with immunoaffinity depletion is achieved with antibody-based columns and LC systems equipped with refrigerated autosamplers and fraction collectors (*see* **Note 2**). Depleted samples are frozen and stored until the next step. This step is linked to and followed by buffer exchange performed in 96-well plates. The rest of the workflow is automated and includes (1) reduction, alkylation, and enzymatic digestion; (2) TMT labeling and pooling; (3) RP solid-phase extraction (SPE); and (4) strong cation-exchange (SCX) SPE. A liquid handling platform dispenses predefined volumes of reagents with a four-channel pipetting arm, which acts also as a gripper to move plates within the deck (Fig. 2). The robot incorporates a 96-well plate shaker and a heater. A dark chamber is installed for the alkylation step. The liquid handler aspirates samples and dispenses them into new laboratory wares. A *vacuum* manifold is installed for automated SPE of up to 96 samples, with reservoirs for stocking solvents. RP-LC MS/MS is then performed to analyze the samples.

3.2 Depletion

1. Check the HPLC system for proper functioning and refill Buffer A and B. Mount the MARS depletion column. Equilibrate the system with 100% Buffer A at a flow rate of 125 µL/min.
2. Place the 1.4 mL collection tubes in the autosampler to collect 1000 µL of depleted plasma. Scan the barcodes of the collection tubes (*see* **Note 3**).
3. Place the 1.0 mL tubes in the autosampler and fill them with 900 µL water for tool cleaning (*see* **Note 4**).
4. Place a vial filled with Buffer A for blank injections.
5. Prepare the sample injection sequence (*see* **Note 4**).
6. Check the LC pressure values, their stability, and control for leaks.
7. Take plasma samples from the −80 °C freezer (*see* **Note 5**). Let the samples thaw at room temperature for 15–30 min. Scan and record the barcodes of the sample tubes.
8. Shake the plasma samples for 30 s on the vibrating platform.
9. Using the multi-pipette, place 90 µL of Buffer A containing 0.0134 mg/mL of LACB (i.e., the internal standard) in the wells of a filter plate mounted on a polypropylene 96-well plate (i.e., reception plate).

Table 1
HPLC depletion chromatographic method

Time (min)	Buffer A (%)	Buffer B (%)	Flow rate (µL/min)
0.00	100	0	125
18.00	100	0	125
18.10	100	0	1000
20.00	100	0	1000
20.01	0	100	1000
27.00	0	100	1000
27.01	100	0	1000
36.70	100	0	1000
36.80	100	0	800
38.00	100	0	800

10. Using the multi-pipette, add 30 µL of the plasma samples in the previously filled wells.
11. Shake the diluted plasma samples for 30 s on the vibrating platform.
12. Filter the samples using a *vacuum* manifold during 1 min.
13. Shortly centrifuge the reception plate to spin down the diluted/filtered samples and put an adhesive mat to cover it.
14. Place the polypropylene 96-well plate containing the diluted/filtered plasma samples in the autosampler of the HPLC system, at 6 °C.
15. Perform the depletion of the samples using the HPLC method indicated in Table 1 (*see* **Note 5**) by systematically collecting the unbound fractions that elute first from the MARS column. Bound fractions containing the abundant proteins are eluted with Buffer B and discarded.
16. After the depletion of samples, recover the tubes containing the depleted samples from the autosampler. Close the tubes with screw caps.
17. Snap-freeze the samples with liquid nitrogen.
18. Store the tube at −80 °C.

3.3 Buffer Exchange

1. Take 96 samples from the −80 °C freezer to compose a full 96-well rack (*see* **Note 5**).
2. Let the samples thaw at room temperature for 60 min. If necessary, thaw further the samples in a container filled with cold water and agitate with a vibrating platform to accelerate the process.

3. Record the samples' barcodes, using the 2D barcode reader (*see* **Note 6**).
4. Shortly centrifuge the tubes/96-well rack to spin down the liquid. Uncap the tubes with the capper/decapper.
5. Use Strata-X 33u Polymeric RP cartridges mounted on a 96-hole holder and a *vacuum* manifold.
6. Condition the cartridges with 1 mL CH_3CN/0.08% TFA.
7. Equilibrate with 1 mL water/0.1% TFA.
8. Slowly load the samples on the cartridges using the multipipette equipped with long tips.
9. Wash the emptied sample tubes with 300 µL of water/0.1% TFA and slowly load those washing solutions on the cartridges.
10. Elute with 1 mL of 30% water/70% CH_3CN/0.08% TFA in a 2 mL 96-DeepWell plate previously barcoded (*see* **Note 3**).
11. At the end of the elution, slowly increase the *vacuum* and wait for 1 min.
12. Dry the samples of the 96-DeepWell plate with a SpeedVac overnight.
13. Recover the 96-DeepWell plate and cover it with a polypropylene mat before storage at −80 °C.

3.4 Reduction, Alkylation, and Proteolytic Digestion of Proteins

1. Place the source 96-DeepWell plate containing the 96 samples (typically previously depleted and exchanged from the buffer (*see* **Notes 3** and 7)) on the robot deck (*see* Fig. 2, "Master Plate").
2. Completely fill the 100 mM TEAB and 2% SDS reactant reservoir and tube (*see* **Note 8**) (*see* liquid handler deck layout in Fig. 2, "Solvents" and "Reagents", respectively). Fill the 20 mM TCEP reactant tube with 1 mL of solution (Fig. 2, "Reagents") (*see* **Note 9**).
3. Start the robotic run. The robot adds 5 µL of 2% SDS and 95 µL of 100 mM TEAB to each well. After shaking the 96-DeepWell plate for 30 s, the robot adds 5.3 µL 20 mM TCEP to each well. When the solutions are dispensed, the robot shakes the 96-DeepWell plate for 30 s and incubates for 1 h at 55 °C to reduce disulfide bridges (*see* Fig. 1) (*see* **Note 10**); the incubation step happens in the dark chamber for practical reasons (*see* Fig. 2).
4. Fill with 1 mL 150 mM IAA the dedicated reactant tube (*see* Fig. 2, "Reagents") shortly before the end of the previous incubation.
5. After waiting for the samples to cool down at room temperature for 2 min, the robot adds 5.5 µL 150 mM IAA to each well,

shakes the 96-DeepWell plate for 30 s, and incubates the samples at room temperature for 1 h in the dark chamber (*see* Fig. 2) (*see* **Note 10**).

6. Fill the trypsin/Lys-C reactant tube (*see* Fig. 2, "Reagents"). The robot adds 10 µL of 0.25 µg/µL trypsin/Lys-C in 100 mM TEAB to each well, places a plastic cover on the 96-DeepWell plate, and incubates overnight at 37 °C with gentle shaking (*see* **Note 10**).
7. Stop the robotic program and the incubation.
8. Trash used and/or empty laboratory wares.
9. Shortly centrifuge the 96-DeepWell plate containing the samples to spin down liquid.

3.5 TMT Labeling

1. Position the 96-DeepWell plate containing the samples on the liquid handler deck (*see* Fig. 2, "Master Plate").
2. Spin down the tubes containing the dried TMT reagents (*see* **Note 11**). Dissolve lyophilized TMT reagents (13 mg in 700 µL µL CH_3CN). Vortex 30 s and spin down the liquid.
3. Transfer the liquids containing the TMT reagents in cleaned 500 µL tubes labeled with the TMT reagent type (i.e., 126, 127, 138, 129, 139, and 131) (*see* **Note 12**). Place the tubes on the liquid handler deck (Fig. 2, "Reagents").
4. Start the robotic run. The robot adds 41 µL of TMT solution to the samples in 96-DeepWell plate according to the design of the TMT experiments and incubates for 1 h at room temperature while shaking.
5. Fill the hydroxylamine reactant tube with 1 mL (Fig. 2, "Reagents") before the end of the previous incubation.
6. The robot adds 8 µL 5% hydroxylamine in water and incubates at room temperature during 15 min under shaking to quench the TMT reaction and reverse occasional labeling of tyrosine, serine, and threonine residues.
7. Next, the robot mixes/pools samples (6 by 6 in case of TMT 6-plex) in a clean 5 mL tube (Fig. 2, "Sample Pooling"). From 96 initial samples, 16 TMT experiments are obtained at this stage.
8. Fill the dedicated reservoir with RP loading buffer (Fig. 2, "Solvents").
9. The robot washes each well of the 96-DeepWell plate with 120 µL of RP loading buffer and adds the washing liquid to the dedicated sample pools.
10. Trash used and/or empty laboratory wares.
11. Proceed to sample purification.

3.6 Sample Purifications

1. Fill the stock bottles containing the RP loading buffer and the RP elution buffer. Fill the RP conditioning buffer reservoirs (Fig. 2, "Solvents").
2. Place Oasis HLB cartridges (1 cc, 30 mg) on the holder of the *vacuum* manifold and a clean 96-DeepWell plate previously barcoded (*see* Fig. 2, "RP & SCX SPE").
3. Start the robotic run. The liquid handler dilutes the peptide samples with 2.7 mL RP loading buffer. Then the robot prepares the cartridge with 2 × 0.95 mL of RP conditioning buffer and 4 × 0.95 mL of RP loading buffer. It loads the samples (aspirating 5 × 0.86 mL from the 5 mL tubes containing typically the pooled TMT-labeled samples (Fig. 2, "Sample Pooling")) and washes the samples with 4 × 0.9 mL RP loading buffer.
4. The peptides are eluted with 2 × 0.8 mL RP elution buffer in the clean 96-DeepWell plate (Fig. 2).
5. Trash empty laboratory wares.
6. Proceed to SCX purification.
7. Fill the stock bottles containing the SCX loading buffer and the SCX elution buffer.
8. Place SCX cartridges on the holder of the *vacuum* manifold and a clean 96-DeepWell plate previously barcoded (*see* Fig. 2, "RP & SCX SPE").
9. Start the robotic run. The robot equilibrates the cartridge with 4 × 0.93 mL of SCX loading buffer. The robot loads the samples (aspirating 2 × 0.95 mL from the 96-DeepWell plate containing typically the RP-purified samples (Fig. 2)) and washes the samples with 4 × 0.9 mL SCX loading buffer.
10. The peptides are eluted with 2 × 0.8 mL RP elution buffer in the clean 96-DeepWell plate (Fig. 2).
11. Trash empty laboratory wares.
12. Lyophilize the samples with a SpeedVac overnight.
13. Dissolve the samples with 1.6 mL SCX loading buffer, and lyophilize again.
14. Cover the 96-DeepWell plate with a polypropylene mat and keep the dried peptide samples at −80 °C before RP-LC MS/MS analysis.

3.7 RP-LC MS/MS Analysis

1. Dissolve the samples in 500 µL of LC solvent A. Shake the 96-DeepWell plate for 5 min. Take 50 µL and transfer the volumes into pre-labeled LC vials. Place the LC vials at 8 °C in the LC autosampler. Store the remaining sample volumes at −80 °C.
2. Prepare the sample sequence. Blank (i.e., LC solvent A) and wash (i.e., CH_3CN) samples are inserted every four plasma injections.

3. Inject 5 μL of sample per analysis.
4. Run the RP-LC for 150 min using a gradient of LC solvent A and LC solvent B (as indicated below) at a flow rate of 220 nL/min. After 10 min loading and washing of the sample on the pre-column using LC loading solvent, analytical separation uses gradient typically as follows: 0–1 min 93.7% LC solvent A and 6.3% LC solvent B, then to 62.5% LC solvent A and 37.5% LC solvent B at 140 min, and 2% LC solvent A and 98% LC solvent B at 150 min, followed by re-equilibration of the analytical column.
5. Perform RP-LC MS/MS analysis using data-dependent acquisition (*see* **Note 13**). With the LTQ-OT Elite, MS survey scans in the OT are recorded with resolution set to 120,000, and the ion population set to 1×10^6 with an m/z window from 300 to 1500. A maximum of ten precursors is selected for higher-energy collisional dissociation (HCD) with analysis in the OT. For MS/MS in the OT, the ion population is 1×10^5 (isolation width of 2 m/z), with resolution of 15,000, first mass at $m/z = 100$, and maximum injection time of 250 ms. The normalized collision energy is 35% for HCD. Ions with 1+ and unassigned charge states are rejected from MS/MS analysis. The dynamic exclusion is set for 60 s within a ±5 ppm window. A lock mass of $m/z = 445.1200$ is used.

3.8 Data Analysis

1. Convert raw files to peak lists using Proteome Discoverer (*see* **Note 14**). Analysis with Mascot can be directly triggered from Proteome Discoverer.
2. Identification is performed against the human UniProtKB/Swiss-Prot database including the LACB sequence. Mascot is used as search engine. Variable amino acid modifications are oxidized methionine, deamidated asparagine/glutamine, and 6-plex TMT-labeled peptide amino terminus (+229.163 Da). 6-plex TMT-labeled lysine (+229.163 Da) is set as fixed modification as well as carbamidomethylation of cysteine. Trypsin is selected as the proteolytic enzyme, with a maximum of two potential missed cleavages. Peptide and fragment ion tolerance are set to, respectively, 10 ppm and 0.02 Da.
3. Mascot result files (.dat files) are loaded into Scaffold Q+ to be further searched with X! Tandem. Both peptide and protein FDRs are fixed at 1%, with a two unique peptide criterion to report protein identification.
4. Relative quantitative protein values are exported from Scaffold Q+ as Log 2 of the protein ratio fold changes with respect to their measurements in the reference TMT channel (Fig. 3a) (*see* **Notes 15** and **16**), i.e., mean Log 2 values after isotopic purity correction but without normalization applied between samples and experiments.

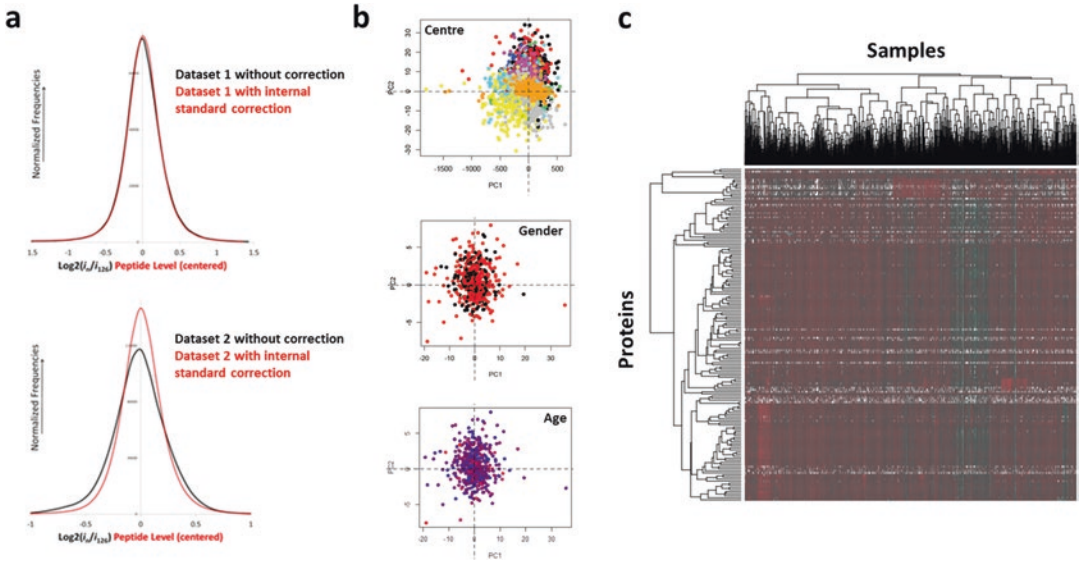


Fig. 3 Examples of data analysis performed on proteomic datasets obtained with ASAP^2. Correction by internal standard can increase the accuracy of the data, as illustrated on the lower graph by the sharper distribution of the quantitative values obtained after correction (**a**). Principal component score plots can be used to evaluate center, gender, or age effects (**b**). Heat map can be used to have a general view on a large dataset (**c**)

5. Perform quality check of each TMT experiment, controlling, for example, for the numbers of identified and quantified proteins, standard deviations, and errors of the quantitative values of the spiked internal standard LACB (*see* **Note 15**).
6. Preprocessing steps before statistical analysis include the removal of samples with over 70% of missing quantitative values for proteins, or with more than half of their protein values belonging to the top ten percentile of the protein ranges.
7. Replicate values per sample are averaged when two replicates are available. For more than two replicates, using the median value is recommended to reduce the impact of outliers. Otherwise, the only non-missing value between the replicates is kept.
8. In order to identify outliers and data structure (such as identifying the presence of clusters) and to assess any potential effects from analytical batch, collection center, gender, or age, principal component analyses (using R package mixOmics) are performed over the full dataset (Fig. 3b) (*see* **Note 17**).
9. Identify proteins that are quantitatively increased or decreased in one sample group versus others (Fig. 3c) (*see* **Note 18**).

4 Notes

1. The ASAP² workflow was developed and is characterized for the analysis of plasma samples, collected in tubes containing EDTA as anticoagulant. EDTA has been previously identified by others as the preferred anticoagulant for plasma collection in proteomic analysis [11]. The use of plasma has been recommended over that of serum because of possible irreproducibility issues of the clotting during serum generation [11]. ASAP² was also successfully applied to the analysis of CSF samples [12]. Because CSF presents lower protein concentrations than plasma, a liquid evaporation step (from 400 µL of CSF sample) was incorporated at the beginning of the workflow [12].
2. In the described workflow, the depletion procedure is the rate-limiting step as it requires 4 days to be sequentially completed for 96 samples using one HPLC apparatus. In our laboratory, we use two identical HPLC systems to be able to deplete 192 samples per week.
3. Robust and simple tracking of samples is mandatory to enable and maintain an efficient work stream. Barcoded laboratory wares are used at every step. We use laboratory information management systems (LIMS) to support the sample tracking process.
4. It is important to maintain the system clean and avoid any contamination or carry-over after multiple injections of plasma samples. Typically, a blank sample containing only Buffer A is analyzed every four plasma samples to allow the lines and column to stay clean and check for eventual contaminations. For this, we record chromatograms using ultraviolet detection at 254 nm.
5. For the depletion step, we recommend to process 24 samples per day, typically starting in the morning the HPLC depletion sequence that finishes the next day. Depletion can therefore be usually performed on Monday, Thursday, Wednesday, and Thursday. The depletion columns have proven stability and accuracy for more than 360 samples, and we suggest to replace them after about 300 sample injections.
6. The study and sample plate template needs to be carefully designed before starting the study. This prevents or minimizes the introduction of experimental batch effects.
7. Non-depleted plasma samples could also be prepared from this step. This could eventually apply to CSF samples.
8. It is recommended to prepare fresh TCEP, IAA, and trypsin/Lys-C solutions prior to each experiment. Two percent SDS is prepared only once a week.

9. With the robotic system we use in our laboratory, it is worth mentioning that, in case a tube of reagent is not present, the liquid handler stops and waits for the operator to place a solution in the reagent tubes or reservoirs. It is recommended to always overfill the reagent tubes (and reservoirs) with a minimal excess volume of 150–200 μL.
10. The robotic platform includes a shaker/heating unit (Fig. 2, *see* ②). The shaker speed is static and defined *a priory* for all processes. The shaker/heating unit can be switched on or off and is temperature-controlled. The dark chamber (Fig. 2, *see* ④) is temperature-controlled.
11. The amount of TMT reagent is custom-made and specially ordered to the manufacturer to allow the labeling of 96 samples on the plate. TMT reagents are moisture sensitive. Reagent stocks should be allowed to reach room temperature before opening to avoid moisture condensation.
12. TMT 6-plex is used in the procedure as 96 is a multiple of 6. A 96-well plate can therefore be entirely filled with TMT experiments.
13. Other mass spectrometers can be used such as tandem time-of-flight (TOF–TOF) and quadrupole (Q)-TOF allowing MS/MS and detection of reporter ions in the low mass range. Other MS acquisition parameters than the ones described in this chapter will be required. We have recently used an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific) to analyze plasma samples prepared with ASAP^2; with rather equivalent RP-LC conditions, a very significant increase of plasma proteome coverage (>60%) was observed with respect to the coverage typically obtained with an LTQ-OT Elite [9]. The mass spectrometer is calibrated every week, following standard operating procedures and well-defined quality checks. We recommend to use a complex standard peptide mix (e.g., obtained from the digestion of a complex protein sample), week after week, to control the performance of the RP-LC MS/MS instrumentation and fine-tune it. The number of MS/MS scans, number of identified proteins and peptides, reference peptide retention time, their elution peak width and intensity, ion injection time, mass accuracy, and multiplier voltage values are examples of readouts we record and track when analyzing this complex standard sample.
14. Alternatives to Proteome Discoverer to convert raw MS files into peak lists exist. The msconvert freeware which is incorporated within the ProteoWizard software is one solution, although not supporting vendor conversion for Linux-based operating system. We usually select mzML data format output and generate it using the following parameters: 32-bit preci-

sion, no zlib conversion (to allow compatibility with X! Tandem software), HCD activation, MS levels 1 and 2 as well as zero sample filter level 1 and 2.

15. An internal protein standard (e.g., LACB) is spiked in the same amount in each sample for later correction of bias occurring during experimental handling (Fig. 3a). This internal protein standard should neither be initially contained in the studied sample nor have tryptic peptides with similar sequences in the human proteome. Correction factors could be determined and applied to correct for manipulation bias. Since spiked in equal amounts in individual samples, the internal protein standard also provides a quality control of the quantitative experiments.
16. To link TMT experiments between each other in a large study, we use a biological reference (i.e., a pool of plasma samples) in each of the experiments. This biological reference is used for the calculation of relative protein differences between all analyzed samples.
17. We recommend to also study principal components with lower variances than the first two or three principal components commonly used to obtain a better picture of the data and its potential effects (Fig. 3b). Additional visual inspection steps of the data include box plots of proteins per analytical batch, collection center, gender, and age and evaluation of similarity and differences. Plotting heat maps of the values also allows to identify those samples with consistently high or low values for a large percentage of proteins (Fig. 3c). These samples could be cross-checked with the laboratory notes to see if they have been flagged as problematic (e.g., too diluted, hemolyzed, low volume, etc.). In such cases, they can be excluded from further statistical analyses.
18. Supervised machine learning techniques are well-suited for biomarker identification in large clinical proteomic settings when the classes are known a priori (e.g., disease versus control, disease progression states, response to treatment/intervention). Examples of commonly used supervised machine learning tools are support vector machines (SVM), decision trees, random forests, rule-based classifiers, and naïve Bayes, among many others. While traditional tests such as *t*-tests are often not recommended for large datasets with few replicates, in particular for datasets with a large number of variables, these tests can still be useful as preliminary screening tools, but it is key to adjust for multiplicity testing to increase the confidence in the results. Additional adjustments must be taken into account such as confounding factors like age, gender, and any other known or suspected effect. Validation of the candidate biomarkers in an independent cohort is essential to add

confidence in the generalization of the proteins identified. After these potential biomarkers are identified and validated, additional works will have to be performed to assess their practical suitability as tools for the clinics.

Acknowledgements

We thank Guillaume Gesquiere and Gianluca Carboni from Hamilton Robotics S.A.R.L. and Martin Jech from Thermo Scientific for their support and productive collaboration.

References

1. Rifai N, Gillette MA, Carr SA (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol 24:971–983
2. Polanski M, Anderson NL (2007) A list of candidate cancer biomarkers for targeted proteomics. Biomark Insights 1:1–48
3. Anderson NL (2010) The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. Clin Chem 56:177–185
4. Hernández B, Parnell A, Pennington SR (2014) Why have so few proteomic biomarkers "survived" validation? (Sample size and independent validation considerations). Proteomics 14:1587–1592
5. Dayon L, Kussmann M (2013) Proteomics of human plasma: a critical comparison of analytical workflows in terms of effort, throughput and outcome. EuPA Open Proteom 1:8–16
6. Keshishian H, Burgess MW, Gillette MA et al (2015) Multiplexed, quantitative workflow for sensitive biomarker discovery in plasma yields novel candidates for early myocardial injury. Mol Cell Proteomics 14:2375–2393
7. Dayon L, Hainard A, Licker V et al (2008) Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem 80:2921–2931
8. Dayon L, Sanchez JC (2012) Relative protein quantification by MS/MS using the tandem mass tag technology. In: Marcus K (ed) Quantitative methods in proteomics. Methods in molecular biology, vol 893. Humana Press, Totowa, NJ, pp 115–127
9. Dayon L, Núñez Galindo A, Corthésy J et al (2014) Comprehensive and scalable highly automated MS-based proteomic workflow for clinical biomarker discovery in human plasma. J Proteome Res 13:3837–3845
10. Cominetti O, Núñez Galindo A, Corthésy J et al (2016) Proteomic biomarker discovery in 1000 human plasma samples with mass apectrometry. J Proteome Res 15:389–399
11. Surinova S, Schiess R, Hüttenhain R et al (2011) On the development of plasma protein biomarkers. J Proteome Res 10:5–16
12. Núñez Galindo A, Kussmann M, Dayon L (2015) Proteomics of cerebrospinal fluid: throughput and robustness using a scalable automated analysis pipeline for biomarker discovery. Anal Chem 87:10755–10761

Chapter 31

Mass Spectrometry-Based Serum Proteomics for Biomarker Discovery and Validation

Santosh D. Bhosale, Robert Moulder, Petri Kouvonen, Riitta Lahesmaa, and David R. Goodlett

Abstract

Blood protein measurements are used frequently in the clinic in the assessment of patient health. Nevertheless, there remains the need for new biomarkers with better diagnostic specificities. With the advent of improved technology for bioanalysis and the growth of biobanks including collections from specific disease risk cohorts, the plasma proteome has remained a target of proteomics research toward the characterization of disease-related biomarkers. The following protocol presents a workflow for serum/plasma proteomics including details of sample preparation both with and without immunoaffinity depletion of the most abundant plasma proteins and methodology for selected reaction monitoring mass spectrometry validation.

Key words Serum/plasma, Label-free quantification, Selected reaction monitoring, Proteomics, Mass spectrometry

1 Introduction

Mass spectrometry-based proteomics has remained in focus as a method of biomarker discovery from clinical samples, in particular serum or plasma [1]. Ideally, quantitative data of specific marker panels could enable the clinician to predict the subclinical status and decide upon the therapeutic management of a disease process [2, 3]. Currently, with the maturation of omics platforms and systems biology approaches, it has become possible to improve sample throughput to a level suitable for biomarker discovery and validation on moderate size [4].

Although proteomics literature includes many examples where putative biomarkers have been identified by mass spectrometry, the list of approved and implemented markers still remains marginal. Following candidate identification, it is crucial that the plausibility of the identified but putative marker is established in order to propagate their movement from the discovery to the translational

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_31,

pipeline [5]. ELISA assays are sensitive and easy to implement and are accepted as the standard approach for validation. However, in recent years there has been a growing use of selected reaction monitoring (SRM) mass spectrometry for validation of putative markers discovered by proteomics [6]. Specifically, SRM is used for the quantification of selected peptides in a targeted manner from a complex sample matrix. The SRM technique, protocol, and its application have previously been reviewed in detail, for which the reader is referred to reviews by the Aebersold group [7, 8]. In comparison to ELISA, SRM provides the advantages of facilitating the monitoring of multiple targets in a single analysis, and the targets can be directly interpolated from the discovery data. As SRM is a peptide-centric approach, it is less susceptible to the influence of sample aging, whereas with ELISA conformational changes in the proteins may influence the interaction with the epitope.

To enable unambiguous SRM validation of the differentially abundant proteins detected in a discovery experiment, it is essential to target peptides that uniquely identify the target protein (i.e., proteotypic peptides) that are preferably stable (i.e., not prone to oxidation or other variable modification), neither too long nor short, and frequently observed in the discovery data with a reproducible retention time. Peptide tandem mass spectra are used to create SRM transitions lists, which are derived from the parent and fragment ion *m/z* values, that are unique to a given peptide sequence and retention time. To realize the full capabilities of SRM, it is beneficial to establish retention indices for the targeted peptides. Spiking with retention time standards, even in the discovery phase, can be thus used and enable the scheduling of a series of targeted measurements [9]. In order to authenticate the peptide targets, isotopically labelled synthetic peptides should be obtained (e.g., incorporating heavy lysine and arginine containing the following heavy isotopes, $^{13}C_6{}^{15}N_2$ and $^{13}C_6{}^{15}N_4$, respectively).

Blood serum is an easily accessible biofluid, which in essence carries a biochemical record of an individual's health status. Its proteomic analysis, however, is challenged by the wide range of protein abundances that are characteristic of its composition. Albumin alone, for instance, constitutes around half of serum protein composition by weight, and the next 12 proteins account for another 45%. The dominance of albumin therefore limits sample loading and the depth of profiling [10]. A frequently necessary, but debatable, consideration is whether to deplete the abundant proteins or not. Targeted removal of the more abundant proteins can be used to alter the range of protein abundance, although the depletion step can be influenced by non-targeted and thus non-specific interactions [11]. Analysis of the undepleted sample, on the other hand, provides a better record of the key serum proteins, although is less likely to provide useful quantitative data for the lower abundance proteins [12]. Several manufacturers provide antibody affinity media that may be used to target the removal of

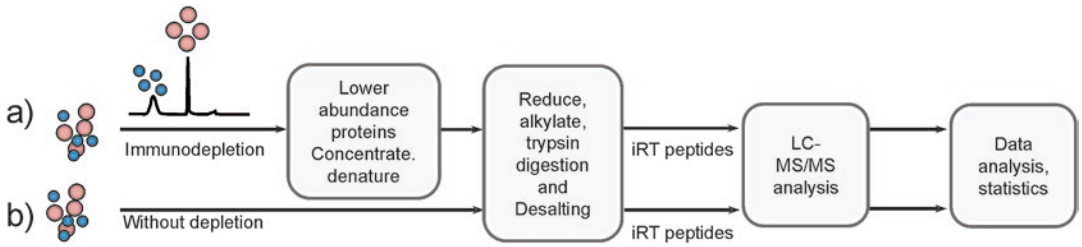


Fig. 1 Sample preparation for LC-MS/MS analysis. (**a**) From serum samples, the high-abundance proteins are removed by immunoaffinity depletion and the fraction containing the low abundant proteins concentrated and denatured prior to enzymatic digestion and desalting. (**b**) Alternatively, for discovery or targeted validation, the serum is directly digested without depletion. Indexed retention time peptides are added to the samples prior to LC-MS/MS to assist in the development of targeted assays by SRM. Although both approaches (**a**, **b**) can be used in the SRM validation phase, the additional sensitivity provided frequently facilitates directed analysis from the undepleted sera

the most abundant proteins. These can be used in a chromatographic or spin-cartridge format. Resins are available for albumin alone, albumin and the IgGs, as well as the 6, 7, 12, 14, and 20 most abundant proteins. Frequently, researchers have removed 12–14 proteins using such resins, which are commercially available from companies such as Agilent and Sigma [13, 14]. The chromatographic approach has been a popular choice, although the recent availability of single use cartridges has provided new opportunities for throughput and scalability [15].

Here we present a pipeline for serum proteomics biomarker discovery and validation, indicating workflows that use immunoaffinity depletion as well as one that is not shown in Figs. 1 and 2.

2 Materials

2.1 Equipment/ Instrument

1. For LC-based immunoaffinity depletion with chromatographic columns of 4–6.6 mm i.d., use a HPLC system capable of delivering a flow rate between 0.1 and 1.0 ml together with a UV detector (280 nm) and fraction collector. *See* **Note 1** for other technical specifications.
2. 0.22 μm Eppendorf spin filters.
3. A depletion column, e.g., MARS Hu-14 column (4.6 mm i.d. × 50 mm; capacity 20 μl, Agilent Technologies, Santa Clara, California, USA) and operating buffers A and B (*see* **Note 2**).
4. Spin concentrators with 5 kDa MW cutoff (Sartorius Stedim, Vivaspin, 4 ml, 5 kDa cutoff).
5. For peptide concentration estimation, a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) or similar instrument.
6. A centrifugal evaporator, e.g., a SpeedVac™ (Thermo Fisher Scientific).

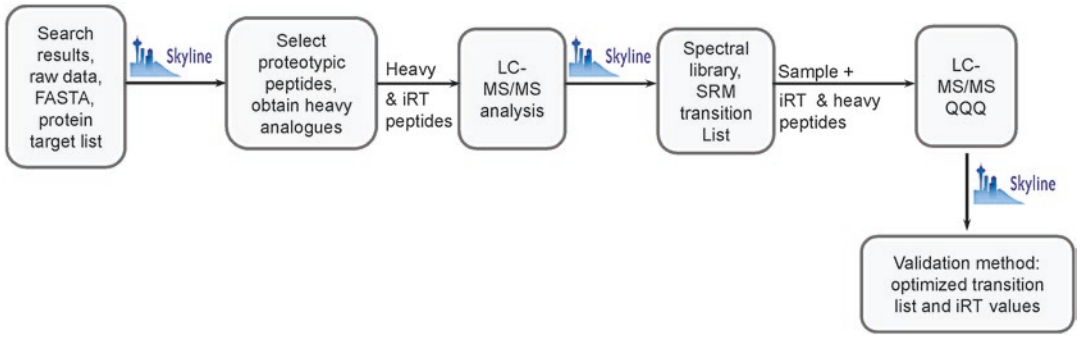


Fig. 2 Development of an SRM targeted assay. Skyline is used to identify the proteotypic peptides associated with the differentially abundant proteins found in the discovery data. To enable unambiguous confirmation of the tandem mass spectral identifications, isotopically labelled synthetic analogues are obtained. LC-MS/MS analysis is made of the heavy peptides together with indexed retention time (iRT) peptides, for example using a Q Exactive, and the data used to develop a spectral library of the targets. Unscheduled/scheduled analysis of the heavy peptides together with a serum sample using a triple quadrupole mass spectrometer (QQQ) is used to confirm the co-elution of the targets and their isotopes and choose the best transitions for the SRM method. The iRT peptides are used to provide retention time indices for the targets in the sample matrix, which are subsequently used for scheduling the SRM method

7. Sep-Pak C18 cartridges, 50 mg (Waters).
8. LC-MS/MS system, i.e., for chromatography and electrospray tandem mass spectrometry with a nanospray interface, e.g. LTQ Orbitrap Velos Pro mass spectrometry coupled with Easy-nLC II (Thermo Fisher Scientific).

2.2 Chemicals

1. 50 mM dissolution buffer: Dissolve 395 mg ammonium bicarbonate (ABC, i.e., NH_4HCO_3) in 100 ml MilliQ water to give 50 mM ABC.
2. Denaturation buffer: 8 M urea; dissolve 24 g of urea in 20.0 ml of 50 mM ABC. When dissolved, make volume to 50 ml.
3. 1 M dithiothreitol (DTT) stock solution: Dissolve the weighed amount of dithiothreitol in 50 mM ABC to give 1.0 M final concentration.
4. 1 M iodoacetamide stock solution: Dissolve weighed iodoacetamide in 50 mM ABC to give 1.0 M final concentration.
5. Sequencing grade modified trypsin (Promega).
6. Indexed retention time (iRT) peptides (Biognosys).

3 Methods

A—Discovery Phase

3.1 Depleted Serum Sample Workflow

Proteins targeted for depletionusing an Agilent Hu14 MARS column:

α1-acid glycoprotein (P02763), fibrinogen (P02761), α1-antitrypsin (P01009).

Haptoglobin (P00738), α 2-macroglobulin (P01023), IgA (P01876), albumin (P02768).

IgG (P01857, P01859-61; all major subclasses of gamma globulin), apolipoprotein A-I.

(P02647), IgM (P01871), apolipoprotein A-II (P02652), transferrin (P02787), Complement 3 (P01024), and transthyretin (P02766).

3.1.1 LC-Based Immunodepletion of High-Abundance Proteins

1. To ensure sample homogeneity, vortex the sample and spin briefly.
2. Dilute the serum aliquot with three volumes of the dilution buffer A, e.g., 15 μl plus 45 μl.
3. Filter the sample through 0.22 μm spin filters to remove particulates.
4. The diluted sample is introduced onto the MARS column with a mobile phase composition of 100% buffer A, at a flow rate of 0.125 ml/min. A generic gradient program is indicated as follows (Table 1).
5. The autosampler method is set up to deliver the flow-through (depleted serum) to the collection tube with sufficient allowance for delay from detection to collection (*see* **Note 3**).
6. Elute the bound fraction; collect if required for further analyses (100% buffer B at flow rate of 1.0 ml/min for 6 min).
7. Column regenerated and equilibration in 100% buffer A for 7 min.
8. If there is concern about carryover, system blanks can be injected between runs.

Table 1
LC method for depletion with an Agilent 4.6 × 50 mm Hu14 column

Cycle	Time (min)	Dilution buffer—A %	Stripping buffer—B %	Flow rate (ml/min)
Injection	0	100	0	0.125
Wash	9.5	100	0	0.125
Wash	9.6	100	0	1.0
Wash	11.5	100	0	1.0
Wash	11.6	0	100	1.0
Stripping	16.00	0	100	1.0
Re-equilibration	16.1	100	0	1.0
Stop	25.00	100	0	1.0

3.1.2 Buffer Exchange

Following depletion, it will be necessary to concentrate the protein solution and change to a denaturing buffer. A method for buffer exchange is presented; *see* **Note 4** for other examples.

1. Pre-rinse the ultrafiltration spin columns for the samples with 1 ml of the dilution buffer: +4 °C, 3000 × *g*, ~20 min. At this stage, it is possible to identify any spin columns that vary in performance (i.e., slow filtration); include sufficient columns to substitute these.
2. Concentrate the collected depleted serum fractions to a volume of ~100 µl using the washed ultrafiltration spin columns (+4 °C, 3000 × *g*, ~20 min).
3. Perform buffer exchange with 8 M urea in 50 mM ABC.
   (a) 1200 µl of 8 M urea: +4 °C, 3000 × *g*, 35 min.
   (b) 500 µl of 8 M urea: +4 °C, 3000 × *g*, 35 min.
   (c) 500 µl of 8 M urea: +4 °C, 3000 × *g*, 30 min.
   Final volume of sample should be ~100 µl.
4. To ensure that the concentrated proteins are in solution and reduce losses to the filter, ultrasonicate the spin columns for 5 min on ice and withdraw the liquid with a pipette.
5. To further reduce the losses from transfer of the concentrate, wash the spin columns with 50 µl of 8 M urea by sonicating 5 min on ice. Combine this with the concentrated sample.

3.1.3 Trypsin Digestion

1. Add 1 µl 1 M dithiothreitol (DTT) and incubate for 1 h at 37 °C.
2. Add 1 µl 1 M iodoacetamide; maintain in darkness at room temperature for 30 min.
3. Dilute the samples to reduce the urea concentration <1 M.
4. Add 10 µl sequencing grade modified trypsin at 37 °C at a ratio of 1:30 (trypsin/protein) for overnight (16–18 h).

3.1.4 Desalting and Peptide Concentration

The digested samples are acidified using 10% trifluoroacetic acid (TFA), and then desalted using Sep-Pak cartridge.

1. Acidify the digested peptides with 50 µl of 10% TFA. Check the pH (<2.5).
2. "Wet" the Sep-Pak column with 1 ml of 100% methanol.
3. Equilibrate with 1 ml of 80% acetonitrile + 0.1% TFA.
4. Equilibration with 2 × 1 ml of 0.1% TFA.
5. Collect the flow-through as you pass the sample steadily through the column (not to dryness), and then pass it through once again.

6. Wash the cartridge with 3 × 1 ml of 2% ACN + 0.1% FA (notice the change to formic acid).
7. Elute the peptides with 1 ml of 80% acetonitrile + 0.1% FA (*see* **Note 5**).
8. SpeedVac to dry.
9. Reconstitute with 2% formic acid +2% acetonitrile (use 60 µl for a sample from 15 µl of serum depleted of the 14 most abundant proteins).

3.1.5 Dilution for Sample Injection

1. Estimate the peptide content using a NanoDrop spectrophotometer with the protein method at 280 nm. Use the 260:280 ratio as a sample integrity; values in the order of 0.6–1 should be expected for a typical sample from 15 µl of serum depleted of the 14 most abundant proteins dissolved in 60 µl.
2. In order to inject 200 ng from a volume of 5 µl injection, the sample should be further diluted to a concentration of 0.04 µg/µl.
3. To assist in retention time mapping and SRMassay development, include a spiked amount of synthetic retention time standards and indexed retention time (iRT) peptides (Biognosys). Spike 1 µl of iRT peptides (50 µl stock) per 20 µl of sample.
4. Perform LC-MS/MS analysis of the samples in triplicate as randomized batches.

3.2 Undepleted Serum Sample Workflow

1. Dilute a 2 µl raw serum sample with 100 µl denaturation buffer (*see* **Note 6**).
2. Reduce the protein disulfide bonds with DTT (final concentration 10 mM) for 1 h at 37 °C. Add 1 µl of DTT stock solution.
3. After 1 h incubation, alkylate the disulfide bridges with iodoacetamide (final concentration 13 mM); incubate for 30 min at room temperature in the dark. Add 1.4 µl of iodoacetamide stock solution.
4. Dilute the samples to 900 µl with 50 mM ABC (to reduce urea <1 M).
5. Trypsin reconstitution: 20 µg trypsin vial (Promega sequencing grade) in 70 µl MilliQ water (0.29 µg/µl).
6. Add in 1:30 ratio (trypsin/protein). For 2 µl undepleted serum (approx. conc ~65–70 µg/µl), add 15 µl of trypsin (4.35 µg). Incubate at 37 °C overnight.
7. Desalting as discussed in the depleted serum sample workflow.

3.3 Sample Analysis and Data Acquisition

LC-MS/MS Conditions

1. Easy nano-LC: Vented pre-column configuration, a 20 × 0.1 mm i.d. pre-column packed with ReproSil-Pur 5 μm 200 Å C18-AQ, connected by a new objective two-way union together with a 75 μm × 150 mm analytical column packed with same packing material.
2. A separation gradient from 2% to 35% B in 65 min at a flow rate of 300 nl/min.
3. Autosampler setup: 20 μl sample loop for 5 μl injections.
4. Sample randomization: The samples are randomized to remove the influence of injection order. Batches of single injections (each samples separated by a 15 min blank method), with three/four replicate injections in total (three or four batches) with the system performance monitored between batches using a lab standard. A pool of the samples in the batch is analyzed at the start and finish of each batch. The maintenance of constant/accountable instrument performance is essential for a successful LFQ experiment.
5. (a) *Orbitrap Velos Pro*: Data dependent MS/MS data acquisition, with ionization in positive ion mode with CID of the 15 most intense ions (m/z 300–2000, charge states >1+). Dynamic exclusion 60 s, Orbitrap precursor ion scan resolution 60,000 (at m/z 400), with a target value of 1,000,000 ions and a maximum injection time of 100 ms. For the ion trap, the target values and maximum injection time values are set to 50,000 and 50 ms. When making the selection of the top "n" most intense ions, it is important to consider the associated duty cycle and the width of the *chromatographic peak, i.e., ensuring that there are sufficient MS1 data points to describe the peak elution profile.

   (b) *Q Exactive*: The method employed data-dependent MS/MS data acquisition, with ionization in positive ion mode with HCD of the ten most intense ions (m/z 300–2000, charge states >1+). Dynamic exclusion 20 s, a resolution of 140,000 (at m/z 400), with a target automatic gain control (AGC) values of 3,000,000 ions and 100 ms maximum injection time. For the MS2 level, the target values and maximum injection time values are set to 50,000 and 250 ms, respectively.

3.4 Data Analysis

Currently, there are two popular software packages available for analyzing label-free data, the freeware, MaxQuant [16], and the commercial package Progenesis from Nonlinear [17]. A description of MaxQuant is presented. Details of the implementation of Progenesis for label-free quantitation are described elsewhere [15].

3.4.1 MaxQuant

MaxQuant is a proteomics mass spectrometry data analysis package that includes its own database search algorithm, Andromeda [18]. It can be used to perform the complete data analysis work flow on a single or groups of raw MS data files to determine peptide and protein identification along with estimations of abundance. For usage of the software and additional information concerning its installation, the reader is referred to www.maxquant.org. Here MaxQuant v.1.4.1.2 is described together with Perseus v.1.4.0.20, which is an open source platform that was developed to process MaxQuant data [19].

1. Open the MaxQuant.exe file.
2. Load the acquired LC-MS/MS raw files into the MaxQuant environment [20].
3. The key points in the context of label-free quantification [21] are as follows:
4. In the "Group-specific parameters" tab, specify following:
   (a) Type > Standard.
   (b) Multiplicity > 1.
   (c) Label-free quantification > LFQ.

   Other parameters like variable modifications, digestion mode, instrument type, and max missed cleavages are user specific.
5. In "Global parameters" tab, select the following:
   (a) FASTA files > Navigate to the FASTA file, uploaded into the Andromeda interface.
   (b) Match between runs > Select this to compare multiple files.
6. For general usage leave the other parameters with default settings.
   (a) Number of threads: This depends on your computer configuration, the number of core. The general recommendation is one thread/2GB derived from MaxQuant Google group discussions.
7. Select start:

   The progress can be monitored in the "Performance" tab, and once the analysis is completed, the status will be displayed in the notification window. The results can be found in the ~*combined*\ directory as the *proteinGroups.txt* file. *See* **Note 7**.
8. The reproducibility across the analytical replicates can be assessed by plotting the scatter plot using Perseus as shown in Fig. 3.
9. For downstream bioinformatics analysis Perseus, which was tailored for use with MaxQuant, provides many options. Additional analyses can be made with R [22] or statistical packages like SPSS [23].

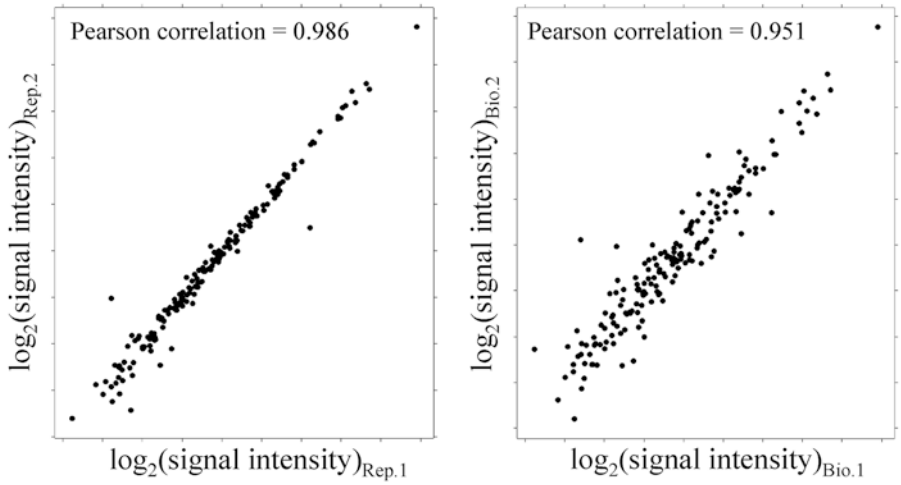


Fig. 3 Scatter plot depicting the correlation between technical replicate (**a**) and biological replicate (**b**). (**a**) and (**b**) show high Pearson correlation coefficient values for analytical replicates reflecting the reproducibility of label-free quantification

B—Validation Phase

In this protocol, the use of Skyline [24] is described. Skyline is a vendor-independent software developed for SRM data analysis (*see* **Note 8**). The discovery data and search results can be imported into Skyline and the availability of proteotypic peptides and stable interference-free SRM transitions established. Similarly, the data from the analysis of heavy-labelled synthetic peptides can be used.

In setting up the Skyline file and associated spectral library and creating peptide retention time indices, you will need:

(a) Representative *.raw files in which the targets were identified.

(b) The search results, e.g., from MaxQuant (msms.txt and modifications.xml) or Proteome Discoverer (*.msf).

(c) A FASTA file of the proteome.

(d) iRT peptides to spike the analyzed samples.

3.4.2 Workflow for SRM Analysis Using Skyline

1. Create a library in Skyline using the search results.
   (a) In the peptide settings, select a background proteome (e.g., HUMAN Swiss-Prot), the number of missed cleavages, and enzyme.
   (b) Follow the tabs through to select the filters for peptide length (e.g., 7–25) and unwanted modifications.
   (c) Build the library by importing the search results.
   (d) Define the permitted modifications.
   (e) Once the library is assigned you can query the library for spectra of the protein targets.

2. Use Skyline to analyze the discovery data and to choose proteotypic peptides identified in the discovery phase (*see* **Note 9**). At this stage, you should have a reasonable overview of the peptides useful for quantifying your protein targets. Transition list for the native peptides can be made although it will become important to validate these with heavy-labelled synthetic equivalents.
3. Synthetic peptide analysis:
   (a) Pool synthetic peptides for transition generation (*see* **Note 10**).
   (b) LC-MS/MS analysis with a Q Exactive: Analyze the pooled sample together with indexed retention time peptides (iRT) on a Q Exactive MS (*see* **Notes 11** and **12**).
   (c) Database search: For the database search, use a FASTA file where the synthetic peptides are concatenated as one protein; the sequence coverage instantly provides an indication of how successful the choice of peptides and their analysis was (*see* **Note 13**).
4. Use Skyline to establish the best co-eluting, interference free transitions for the assays using a triple quadrupole mass spectrometer.
   (a) From the synthetic peptide analysis results, the library can be recreated (*see* above).
   (b) Import the *.raw file and build an iRT retention time calculator using the iRT peptides as retention time standards.
   (c) Check the peak detection and suitability of transition and aim to include five good transitions per peptide, deselecting inappropriate choices and selecting the correct peak apex.
   (d) Export the transition list for the TSQ mass spectrometer. The *.csv output includes such parameters as the precursor and fragment mass, the collision energy, the retention time window (start and stop), the peptide sequence, the protein identity, and the fragment ion and its rank. With the current version of Skyline, manual editing (*see* Table 2) is required so that the table is in the appropriate format for the TSQ triple quadrupole; for some AB Sciex instruments, the output may directly be used.

      The polarity is 1 for positive electrospray, the trigger is an ion intensity threshold (100 by default) and "reference" is used to indicate a standard that can be used to modify the scheduling on the fly (a default value of 0 removes this option).

Table 2
For SRM analysis with a TSQ, the exported transition list (a) should be reordered (b)

(a) Transition list output: (note that the columns are not labelled in the created file)								
Precursor mass	Fragment mass	Collision energy	Start time	End time	Sequence	Protein ID	Ion	Ion rank
(b) Transition list edited for use with a Thermo Scientific TSQ								
Precursor mass	Fragment mass	Collision energy	Start time	End time	Polarity	Trigger	Reference	Name

The first five columns are identical, but the polarity, the intensity trigger for MS/MS, and reference status are added. The name can be derived from the protein ID, sequence, ion, and ion rank

5. SRM analysis:
   (a) Perform a nonscheduled run from your sample where only the iRT peptides are monitored. Measurement of the retention times together with known iRT values allows calculation of retention time indices for the target peptides.
   (b) Import the nonscheduled run*.raw file into Skyline interface for calculation of actual retention times for the target peptides. This is achieved by utilizing calculated iRT values for the target peptides and measured retention times for iRT peptides.
   (c) Assign the retention time indices and permitted tolerance for the SRM method together with all collision energies.
   (d) Perform a scheduled run where proteotypic peptides (protein target) transitions are monitored along with its heavy-labelled counterpart and iRT. Then analyze the scheduled *.raw file with Skyline, and evaluate the transitions for all targets. A few scheduled runs can be submitted for transition optimization and validation (*see* **Note 14**).
   (e) Finally, analyze the synthetic peptide pool together with iRT peptides in sample matrix using optimized, scheduled method to monitor the transitions (protein targets). Transitions affected by sample background interference can be used for identification but should not be used for quantification.
   (f) In order to maintain the mass spectrometry performance stable throughout the batch, the inclusion of short-run blank runs (i.e., no sample) in between samples is beneficial. Additionally, the inclusion of a pooled sample to the batch, for example, at the start and end of the batch provides a useful overview of the instrument performance (*see* **Note 15**).

4 Notes

1. As the recommended procedures required to the dilution of the serum, it is important that the sample loop volume used with the LC system is larger enough, e.g., in the order of three times the injected volume (200–300 µl depending on the processed amount and the column type) to accommodate for the diluted sample and reduce losses. Also the sampler should preferably be configured such that after loading the sample would be back-flushed directly onto the column, i.e., does not need to pass through the whole loop.
2. Buffers: The identity of the Agilent buffers remain proprietary information. However, for the use of the IgY media, as has been supplied by Sigma and others, the buffers may be prepared according to the details provided. Dilution buffer: Tris-buffered saline (TBS)—10 mM Tris–HCl with 0.15 M NaCl, pH 7.4; stripping buffer: 1 M Glycine, pH 2.5; neutralization buffer: 1 M Tris–HCl, pH 8.0.
3. As the method includes a flow increase, the volume of the collected fraction will change rapidly as its duration exceeds the flow change. Method development with test serum should be made to optimize the timing of the fraction collection.
4. In comparison to buffer exchange, commonly used actone or acetone/TCA precipitation approaches can be used to recover the proteins from depleted flow-through. Which approach to follow is based on the yield (buffer exchange vs protein precipitation) however should be tested.
5. Elution with a lower acetonitrile concentration, e.g., can be sufficient and may help to remove large problematic peptides/proteins and unwanted contaminants.
6. For quantitative analysis, it is important to aliquot reproducible quantities. Increasing the aliquot size could be used to improve the reproducibility, e.g., to 5 µl, although due to demands for larger quantities of trypsin, this has its drawbacks. Alternatively an aliquot of the diluted serum can be digested.
7. The processing time is dependent on the number and type of *.raw files as well as the configuration of processing computers. In our experience, processing 60 *.raw files obtained from an LTQ Orbitrap Velos Pro takes 1 day using a computer with the following specifications.
   (a) Processor—Intel(R)Xeon(R) CPU E5—26090 at 24 GHz, RAM—128 GB, and 64-bit operating system.
8. The development of transition lists can be made using different software dedicated to SRM data analysis [25–27]. Detailed step-by-step instructions for using Skyline are found on the

web page https:// skyline.gs.washington.edu/labkey/project/home/software/Skyline/begin.view.

9. Ideally synthetic peptides should be 7–25 amino acids in length and unique. The absence of methionine as well as missed cleavages and ragged ends is preferred. Synthetic peptides with carbamidomethylation modification can also be obtained as well as other variants. Shorter or more diverse combinations may be necessary when there is a shortage of unique peptides for the target. Additional proteotypic peptides that were not identified might be considered for synthesis and subsequence confirmation, particularly if only one to two peptides have been successfully identified for given protein.
10. Usually the concentrations reported by the manufacturer are based on total concentration from peptide synthesis, which also includes synthesis by-products, such as single amino acids and truncated peptides. Therefore, calculating equimolar amount of peptides into the pool is not feasible. In our case, 10 μl from each peptide vial was used for initial analysis.
11. For library creation HCD spectra from a Q Exactive, compare favorably with the CID spectra generated by triple quadrupole instruments, such as the TSQ Vantage. In the following workflow, using heavy-labelled synthetic peptides analyzed with a Q Exactive and TSQ Vantage is described.
12. At this stage, it is advantageous if you are using similar LC configuration and gradient as intended for the subsequent validation. Nevertheless, differences in gradient can be compensated by the iRT peptides if the column stationary phase/packing media is the same.
13. In the case of missing identifications, missing peptides can be pooled together or even injected individually to complement the results.
14. Validate the integrity of the transitions in a pooled sample spiked with heavy-labelled and iRT peptides. The pooled sample should be representative of the sample matrix to be analyzed in the validation step. (If possible, this sample should also be used in following step as a quality control sample, QC-sample. *See***step 5** point f.) The sample matrix will generate some interference, which should be taken into account for when choosing the stable transitions for detection and quantification. Ideally, the selected transitions should have clear and matching profiles, and simple ions that are likely to be common to many proteins, e.g., y1 and y2, should be avoided.
15. The sample order should be randomized for data acquisition. However, the QC-sample should be analyzed in the beginning, within and at the end of the batch. This can be to used to determine the stability of the system.

Acknowledgment

The National Technology Agency of Finland (Finland Distinguished Professor Programme, grant 40398/11), the Academy of Finland (Centre of Excellence in Molecular Systems Immunology and Physiology Research, grant 250114), JDRF, the Sigrid Jusélius Foundation, and Biocentre Finland (Turku Proteomics Facility) are thanked for their financial support.

References

1. Veenstra TD, Conrads TP, Hood BL, Avellino AM, Ellenbogen RG, Morrison RS (2005) Biomarkers: mining the biofluid proteome. Mol Cell Proteomics 4:409–418
2. Lyons TJ, Basu A (2012) Biomarkers in diabetes: hemoglobin A1c, vascular and tissue markers. Transl Res 159:303–312
3. Crutchfield CA, Thomas SN, Sokoll LJ, Chan DW (2016) Advances in mass spectrometry-based clinical biomarker discovery. Clin Proteomics 13:1
4. Dayon L, Kussmann M (2013) Proteomics of human plasma: a critical comparison of analytical workflows in terms of effort, throughput and outcome. EuPA Open Proteom 1:8–16
5. Rifai N, Gillette MA, Carr SA (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol 24:971–983
6. Surinova S, Schiess R, Hüttenhain R, Cerciello F, Wollscheid B, Aebersold R (2011) On the development of plasma protein biomarkers. J Proteome Res 10:5–16
7. Picotti P, Aebersold R (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods 9:555–566
8. Lange V, Picotti P, Domon B, Aebersold R (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol 4:222
9. Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, MacCoss MJ, Rinner O (2012) Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics 12(8):1111–1121
10. Issaq HJ, Xiao Z, Veenstra TD (2007) Serum and plasma proteomics. Chem Rev 107(8): 3601–3620
11. Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJ, Liebler DC (2010) Depletion of abundant plasma proteins and limitations of plasma proteomics. J Proteome Res 9(10):4982–4991
12. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M (2016) Plasma proteome profiling to assess human health and disease. Cell Syst 2(3):185–195
13. Ahmed N, Barker G, Oliva K, Garfin D, Talmadge K, Georgiou H, Quinn M, Rice G (2003) An approach to remove albumin for the proteomic analysis of low abundance biomarkers in human serum. Proteomics 3(10):1980–1987
14. Smith MP, Wood SL, Zougman A, Ho JT, Peng J, Jackson D, Cairns DA, Lewington AJ, Selby PJ, Banks RE (2011) A systematic analysis of the effects of increasing degrees of serum immunodepletion in terms of depth of coverage and other key aspects in top-down and bottom-up proteomic analyses. Proteomics 11(11):2222–2235
15. Moulder R, Goo YA, Goodlett DR (2016) Label-free quantitation for clinical proteomics. Methods Mol Biol 1410:65–76
16. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372
17. Megger DA, Bracht T, Kohl M, Ahrens M, Naboulsi W, Weber F, Hoffmann AC, Stephan C, Kuhlmann K, Eisenacher M, Schlaak JF, Baba HA, Meyer HE, Sitek B (2013) Proteomic differences between hepatocellular carcinoma and nontumorous liver tissue investigated by a combined gel-based and label-free quantitative proteomics study. Mol Cell Proteomics 12:2006–2020
18. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 10:794–1805
19. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 13(9):731–740

20. Tyanova S, Mann M, Cox J (2014) MaxQuant for in-depth analysis of large SILAC datasets. Methods Mol Biol 1188:351–364
21. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13:2513–2526
22. R Development Core Team (2011) R: a language and environment for statistical computing, Vienna, Austria
23. IBM Corp. Released 2013. IBM SPSS statistics for windows version 22.0. IBM Corp., Armonk, NY
24. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26:966–968
25. Brusniak MK, Kwok S, Christiansen M, Campbell D, Reiter L, Picotti P, Kusebauch U, Ramos H, Deutsch EW, Chen J, Moritz RL, Aebersold R (2011) ATAQS: A computational software tool for high throughput transition optimization and validation for selected reaction monitoring mass spectrometry. BMC Bioinformatics 12:78
26. Teleman J, Karlsson C, Waldemarson S, Hansson K, James P, Malmström J, Levander F (2012) Automated selected reaction monitoring software for accurate label-free protein quantification. J Proteome Res 11(7):3766–3773
27. Aiyetan P, Thomas SN, Zhang Z, Zhang H (2015) MRMPlus: an open source quality control and assessment tool for SRM/MRM assay development. BMC Bioinformatics 16:411

Chapter 32

Metabolomics Toward Biomarker Discovery

Peiyuan Yin and Guowang Xu

Abstract

Metabolomics has been used as practical tool in the discovery of novel biomarkers in a broad area in the clinic. The analytical platforms including nuclear magnetic resonance (NMR) and mass spectrometry (MS) can cover thousands of metabolites. With the help of multivariate data analysis, many potential biomarkers can be defined in the studies. Since metabolites stand at the end point of metabolism, it remains difficult to find novel biomarkers with good diagnostic or prognostic performance. In this chapter, we will introduce a general protocol for biomarker discovery within the scope of metabolomics using MS.

Key words Metabolomics, Biomarker, Mass spectrometry

1 Introduction

Metabolites have been used as molecular markers in the clinic for a long history [1]. With the development of modern techniques of analytical chemistry, it is now possible to measure thousands of metabolites in an efficient and sensitive way. And new computational algorithms also enable us to process complex data and discover important metabolites from numerous organic compounds [2]. Since the concept has been raised and defined in the late 1990s [3], metabolomics is now widely used in a broad area of life science, including biomarker discovery, classification of disease, drug development, etc. [4, 5].

A metabolomics study always means acquisition and processing of large volume of information. Now, about 8000 endogenous metabolites and about 40,000 related compounds could be searched online (www.hmdb.ca) [6]. Therefore, a well-designed plan and protocol in either preanalytical or analytical procedures is intensively required for any metabolomics study [7]. For example, the standard operation procedure (SOP) for sample collection, storage, and transportation has been a focus of concern recently, which shows great influence on the results of metabolomics [8]. Another important issue is the coverage of metabolic analysis.

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_32,

Analytical methods on nuclear magnetic resonance (NMR) and mass spectrometry (MS) could detect and measure most of these metabolites under special protocols. NMR, gas chromatography (GC) MS and liquid chromatography (LC) MS are commonly employed platforms which provides different coverage of metabolites. As for a single MS method, hundreds of metabolites could be identified out of thousands of detected ion features. However, it is still far from a "comprehensive"-omics analysis. Considering the advantages of different analytical platforms, a combination usage of these platforms is practical option for a metabolomics project [9]. The development of sample pretreatment also enlarges the coverage of metabolites, such as new materials and techniques for extraction, enrichment, and derivatization.

Data analysis is one of the key steps of metabolomics studies. Data acquired from MS always consists of thousands of ions. There are two challenges that have to be addressed before the data could be used [10]. The first question is "what are these ions?" and the other is "which one(s) is important?" Generally, it is neither possible nor necessary to know every ion in the dataset. A more realistic approach is to define those with importance in the projects, then identify their chemical structures. With the help of multivariate analysis and statistical methods, it is able to define those important metabolites as candidates for either diagnostic or other purpose. The last but not least, all those differential metabolites or important variables are only candidate biomarkers. Clinical or biochemical validations are indispensable procedures before a biomarker can be defined [11].

In this chapter, we will introduce a common protocol for biomarker discovery in a LC MS or GC MS metabolomics study including the following steps: design of the study, preanalytical aspects, instrument analysis, data analysis, and validations.

2 Materials

2.1 Reagents and Chemicals

1. Methanol, acetonitrile, dichloromethane (HPLC grade).
2. Ultrapure water: 18.2 MΩ cm, TOC = 6 ppb.
3. Chemicals: formic acid, ammonium bicarbonate, urease, pyridine methoxamine hydrochloride, and *N*-methyl-*N*-(trimethylsilyl)-trifluoroacetamide (MSTFA).
4. Internal standards:

   GC MS: tridecanoic acid in methanol (40 μg/mL).

   LC MS [12]: leucine-d3 (0.3 μg/mL), acetylcarnitine-d3 (0.05 μg/mL).

   Chenodeoxycholic acid glycine conjugate-d4 (0.006 μg/mL).

   Free fatty acid C16:0-d3 (0.5 μg/mL).

Lysophosphatidylcholine C19:0 (0.6 μg/mL).

Phosphatidylcholine C38:0 (1.6 μg/mL).

(*see* **Note 1**).

2.2 Instrument

1. High-speed freezing centrifuge.
2. Vacuum freeze dryer.
3. Ultrahigh-performance liquid chromatography, gas chromatography (with temperature-controlled autosampler).
4. MS: time of flight (TOF) MS or Orbitrap MS.
5. Automated tissue homogenization.
6. Incubator shaker (temperature-controlled).

2.3 Study Design and Samples

1. Criterion for the enrollment of patients (*see* **Note 2**).
2. Collection of clinical information, including confirmed diagnosis, laboratory test, histology, and physical examination.
3. Enrollment of matched control group.
4. Conformation of the sample qualities, including the collection procedures, tubes, and transportation of samples [12].

3 Methods

3.1 Sample Preparation

1. Thawing samples

   Serum or plasma (50–100 μL), urine (>50 μL), tissues (>20 mg) thawing on ice for 30–60 min (*see* **Note 3**).
2. Tissue homogenize

   Add 600 μL cold 80% methanol/water (v/v, with internal standards) to tissue samples, and vortex the mixture for 30 s. Then homogenize the mixtures at 25 Hz for 1 min twice. Then centrifuge the mixture at 15,000 × *g* for 10 min at 4 °C. Transfer 480 μL supernatant to a new tube.
3. Preparation of QC samples

   Pipette 5 μL (or more) of serum, plasma, or tissue extractions from each sample in the study, and mix together as pool QC samples.
4. Urine enzymolysis for GC MS

   Add 150 μL urease water solution (10 mg/mL) to 100 μL urine. Vortex for 10 s, and then put the mixture into water bath at 37 °C for 15 min (*see* **Note 4**).
5. Protein precipitation

   Add 300 μL methanol with IS to 100 μL serum/plasma or urine, and then centrifuge at 15,000 × *g* for 10 min at 4 °C. Transfer 300 μL supernatant to a new tube.

6. Freeze-drying

   Lyophilize the above supernatant from blood, urine, and tissue in a lyophilizer at about −50 °C for about 3–4 h. Take out the dry sample, and store the samples in −80 °C until analysis.

7. GC derivatization

   Add 100 μL methoxyamine solution (20 mg/mL in pyridine) to the dry samples, and mix by ultrasonic for 15 min at room temperature. Put the samples into water bath at 40 °C for 2 h. Add 80 μL MSTFA to the samples and stand in water bath at 40 °C for 1 h. Centrifuge the mixture at 15,000 × *g* for 10 min at 4 °C (*see* **Note 5**).

3.2 Instrument Analysis

1. Sample redissolving

   Redissolve samples with acetonitrile/water for reversed phase LC MS analysis (1/4, v/v) (*see* **Note 6**). Vortex the mixture for 60 s at room temperature (*see* **Note 7**).

2. MS preparation

   For LC MS, set the scan range from 50 to 1000 (*m/z*). A better resolution of MS should be larger than 10,000 (at 200 *m/z*). The scan speed should be faster than 1 spectrum/second. MS system should be calibrated to ensure the mass accuracy. The scan range of GC MS is set at 30–600 (*m/z*).

3. LC preparation

   Mobile phase for LC MS (reversed phase):

   Positive ion mode: A, water with 0.1% formic acid; B, acetonitrile with 0.1% formic acid.

   Negative ion mode: A, water with 5 mmol/L ammonium bicarbonate, PH 9.0; B, 95% methanol/water (v/v) with ammonium bicarbonate (*see* **Note 8**).

   Set column temperature at 50 °C and sample temperature at 8 °C (*see* **Note 9**). Rinse the LC column with 100% mobile phase B for 10 min at the flow rate of 0.35 mL/min, and then change to 50% B for 5 min.

   Then equilibrate the LC system with 98% mobile phase A.

4. Sequence managing

   Run five blank samples and five QC samples before the first injection of experimental samples. All experimental samples queue randomly in the batch. Insert QC samples every ten samples (*see* **Note 10**).

5. GC MS analysis

   Inject 1 μL samples with an autosampler.

   Set the flow rate of carrier gas (helium, 99.9995%) at 1.2 mL/min and the split ratio set at 10:1.

Oven temperature program: Firstly, keep the temperature at 70 °C for 3 min, rise to 220 °C at a step of 4 °C/min, and then heat to 300 °C at a step of 8 °C/min. Keep at 300 °C for 10 min.

6. LC MS analysis

   Inject 5 μL samples with an autosampler.

   The flow rate set at 0.35 mL/min.

   LC gradient elution program:

   Initial: 98% mobile phase A (2% B).

   Keep for 2 min.

   Change to 0% A (100% B) in 16 min.

   Keep for 4 min.

   Change to 98% A in 0.1 min.

   Equilibrate for 1.9 min.

3.3 Data Analysis

1. Raw data extraction

   Extract the ions and deconvolutions (GC MS) from raw data with the commercial software or free software such as XCMS.

   Align the ion peaks with a time window of 0.5 min and mass window of 10 ppm (*see* **Note 11**).

   Export the peak list to an Excel file.

2. Data pretreatment

   Calculate the RSD of every ion peaks among all QC samples.

   Remove the variables with missing values in 50% of the samples in one group (*see* **Note 12**).

   Replace the missing values with one tenth level of baseline noise.

   Normalize all peaks to the total intensity of ions (*see* **Note 13**).

3. Multivariable analysis

   Perform principal component analysis (PCA) using all the data.

   Mean center all variable and scaling with unit variance.

   Calculate principal component of the model.

   Show score plot of the PCA model (*see* **Note 14**). Perform partial least squares discriminant analysis (PLS-DA) (*see* **Note 15**).

   Mean center all variable and scaling with unit variance or Pareto variance (*see* **Note 16**).

   Calculate components of the model.

   The score plot of PLS-DA may indicate whether the experimental groups could be classified (*see* **Note 17**).

   Model parameters: R^2Y and Q^2 should be checked. R^2Y shows the fitness of the model and Q^2 shows the ability of prediction.

Permutation test: A permutation test should be performed to avoid the risk of overfitting (*see* **Note 18**).

The discovery of biomarkers:

Use variable importance in the project (VIP) to find potential biomarkers. Variable with VIP larger than 1 means it contributes to the classification (*see* **Note 19**).

Variables with higher VIP values can be included in the subsequent discovery of biomarkers.

4. Statistical analysis

   Proper statistical methods should be chosen according to the study groups.

   False discovery rate (FDR) correction of the statistical analysis should be involved to decrease the risk of false positive in biomarker discovery (*see* **Note 20**).

   Receiver operating characteristic (ROC) curve: Use ROC to perform a preliminary evaluation of the diagnostic performance of the metabolic biomarkers. Area under the curve (AUC) is a criterion of ROC (*see* **Note 21**).

5. Identification of candidate biomarkers

   Extract the accuracy mass of ion ready to be identified.

   Find out its related ions by their correlation coefficients.

   Designate quasi-molecular ions.

   Check the MS/MS spectrum of the ion.

   Search the online database of HMDB [6] or METLIN [13] using the accuracy mass of the quasi-molecular ions.

   Validate the results of identification with authentic samples (*see* **Note 22**).

3.4 Validation

1. Internal validation

   Use Monte Carlo cross validation to divide all samples into model group and test group randomly (*see* **Note 23**).

2. External validation

   Enrollment of validation samples (*see* **Note 24**).

   Establish target analysis methods for the quantification of candidate biomarkers.

   Set up new models using the data from quantification.

   Provide diagnostic performance of the potential biomarkers.

4 Notes

1. We provide a list of internal standards only for reference. We suggest about a dozen of stable isotopes included as internal standards to calibrate the signals of MS or the retention time [12].

2. To identify potential biomarkers for diseases, we suggest 20–30 samples enrolled in one group. For the diseased group, typical patients should be included with "golden standards" of diagnosis. In addition, similar genetic background and clinical subtypes should also be taken in consideration.
3. Thawing samples at room temperature should be avoided, which may introduce unexpected ions in the samples.
4. This enzymolysis step is only used for GC analysis of urine samples.
5. The derivatization includes two steps of oximation reaction and silylation reaction. Please note that there will be insoluble materials in the derivatization systems. Make sure to remove all the deposit after centrifugation [14].
6. The solvent constitution should be similar with the initial mobile phase of LC MS, so that the solvent effect can be minimized.
7. If there is precipitate in the tube, vortex the mixture for 30 s twice. Do not reconstruct samples on ice. If necessary, centrifuge the mixture at 15,000 × *g* for 10 min at 10 °C. Ultrafiltration membranes are not recommended in this step.
8. Ammonium bicarbonate should be dissolved in water as stock solution. Mobile phase A and B should be prepared freshly.
9. Low sample temperature may decrease the solubility of lipids and block the LC system.
10. QC samples should dissolve every day. The ion source needs to be cleaned every 1–2 days. After cleaning of the ion source, five QC samples should be injected before the sample sequence continues.
11. The time window should be set according to the MS instrument. High resolution MS with better mass accuracy may set the mass window less than 5 ppm or 0.001 Da.
12. It is also recommended to remove variables with missing values among 20% samples in one group.
13. Normalization is an essential step for urine to adjust the different volumes among individuals. Creatinine is not recommended to adjust all ions in the urine.
14. The cluster of all QC samples indicates the stability of the whole procedure of analysis. A wide distribution of QC samples on the score plot always means the instrument analysis is not stable.
15. For the purpose of biomarker discovery, data with two experimental groups are recommended using PLS-DA models.
16. Pareto variance retains part of the intensity information of each ions. When Pareto scaling is used, variables with higher intensities may have larger weight for the classification.

17. In the software of SIMCA-P, only those components with significance could be calculated. If there is no component calculated, it means the groups cannot be separated by PLS-DA.
18. For a good model, the intercept of R^2Y and Q^2 should be below 0.4 and 0.1, respectively.
19. VIP values have confidence intervals derived from jackknifing. The intervals should not exceed the according VIP values.
20. For multiple hypothesis testing of the metabolomics data, Benjamini-Hochberg method is recommended to control the FDR, with a significance level of 0.1.
21. For a single metabolite, the content could be used directly in ROC. When multiple metabolites are included, binary logistic is recommended to calculate the predictive probability. The value of predictive probability could be used to create ROC.
22. A detailed strategy for the identification of known ions could be seen in the literature [15]. High resolution MS such as Fourier transform ion cyclotron resonance (FT-ICR) can provide accurate measurement of the m/z of ions. For most ions, it is not easy to confirm the quasi-molecular ions. The database of my compound ID [16] (www.mycompoundid.org) is recommended to search such ions after one or two reactions [17].
23. Monte Carlo cross validation is a practical strategy to avoid the risk of overfitting when multiple biomarkers are used [18, 19].
24. Comparing with the samples for biomarker discovery, the scope of the including samples may be enlarged. In the meantime, a larger sample set is also recommended.

References

1. Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, Willett WC (2001) Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. N Engl J Med 345:790–797
2. van der Greef J, Stroobant P, van der Heijden R (2004) The role of analytical sciences medical systems biology. Curr Opin Chem Biol 8:559–565
3. Nicholson JK, Lindon JC, Holmes E (1999) 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29:1181–1189
4. Yang M, Soga T, Pollard PJ (2013) Oncometabolites: linking altered metabolism with cancer. J Clin Invest 123:3652–3658
5. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM, Wu Y, Schauer P, Smith JD, Allayee H, Tang WH, DiDonato JA, Lusis AJ, Hazen SL (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 472:57–63
6. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorndahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A (2013) HMDB 3.0—the human metabolome database in 2013. Nucleic Acids Res 41:D801–D807
7. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, Brown M, Knowles JD, Halsall A, Haselden JN, Nicholls AW, Wilson ID, Kell DB, Goodacre R, Human Serum Metabolome (HUSERMET) Consortium (2011) Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid

chromatography coupled to mass spectrometry. Nat Protoc 6:1060–1083

8. Yin PY, Peter A, Franken H, Zhao XJ, Neukamm SS, Rosenbaum L, Lucio M, Zell A, Haring HU, Xu GW, Lehmann R (2013) Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. Clin Chem 59:833–845
9. Yin PY, Xu GW (2014) Current state-of-the-art of nontargeted metabolomics based on liquid chromatography-mass spectrometry with special emphasis in clinical applications. J Chromatogr A 1374:1–13
10. Wagner S, Scholz K, Donegan M, Burton L, Wingate J, Volkel W (2006) Metabonomics and biomarker discovery: LC-MS metabolic profiling and constant neutral loss scanning combined with multivariate data analysis for mercapturic acid analysis. Anal Chem 78:1296–1305
11. Johnson CH, Ivanisevic J, Siuzdak G (2016) Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol 17(7):451–459
12. Yin P, Zhou L, Zhao X, Xu G (2015) Sample collection and preparation of biofluids and extracts for liquid chromatography-mass spectrometry. Methods Mol Biol 1277:51–59
13. Zhu ZJ, Schultz AW, Wang J, Johnson CH, Yannone SM, Patti GJ, Siuzdak G (2013) Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database. Nat Protoc 8:451–460
14. Ye G, Zhu B, Yao Z, Yin P, Lu X, Kong H, Fan F, Jiao B, Xu G (2012) Analysis of urinary metabolic signatures of early hepatocellular carcinoma recurrence after surgical removal using gas chromatography-mass spectrometry. J Proteome Res 11:4361–4372
15. Chen J, Zhao X, Fritsche J, Yin P, Schmitt-Kopplin P, Wang W, Lu X, Haring HU, Schleicher ED, Lehmann R, Xu G (2008) Practical approach for the identification and isomer elucidation of biomarkers detected in a metabonomic study for the discovery of individuals at risk for diabetes by integrating the chromatographic and mass spectrometric information. Anal Chem 80:1280–1289
16. Huan T, Tang C, Li R, Shi Y, Lin G, Li L (2015) MyCompoundID MS/MS search: metabolite identification using a library of predicted fragment-ion-spectra of 383,830 possible human metabolites. Anal Chem 87:10619–10626
17. Dai W, Yin P, Zeng Z, Kong H, Tong H, Xu Z, Lu X, Lehmann R, Xu G (2014) Nontargeted modification-specific metabolomics study based on liquid chromatography-high-resolution mass spectrometry. Anal Chem 86:9146–9153
18. Xu QS, Liang YZ (2001) Monte Carlo cross validation. Chemometr Intell Lab 56:1–11
19. Pihur V, Datta S, Datta S (2007) Weighted rank aggregation of cluster validation measures: a Monte Carlo cross-entropy approach. Bioinformatics 23:1607–1615

Chapter 33

Plasma Biomarker Identification and Quantification by Microparticle Proteomics

Michal Harel and Tamar Geiger

Abstract

Plasma biomarker discovery necessitates a method for deep proteomic profiling, as well as for highly accurate quantification of the proteins in the sample. Furthermore, to obtain strong candidates for potential biomarkers, the method should be high throughput to enable a large scale analysis. Here we describe in detail PROMIS-Quan (PROteomics of MIcroparticles using Super-SILAC Quantification), a method for a simple and robust fractionation of the plasma samples by extraction of plasma microparticles, followed by SILAC-based relative and absolute quantification.

Key words Super-SILAC quantification, Plasma fractionation, Microparticles, Extracellular vesicles, Plasma biomarker discovery, Absolute quantification, Relative quantification

1 Introduction

One of the holy grails of mass spectrometry (MS)-based proteomics is to identify plasma-based diagnostic biomarkers in an untargeted manner [1]. But despite tremendous progress in this field, the MS inherent tendency to preferentially identify highly abundant proteins still impedes the endeavors to obtain a deep coverage of the plasma proteome, let alone to find low abundant proteins in the dense haystack of the core plasma proteins. To overcome these challenges, much effort is thus needed to biochemically reduce the complexity of the plasma proteome, typically by depleting highly abundant proteins, followed by substantial biochemical fractionation of proteins and peptides [2]. Despite these efforts, most studies identify only several hundred of proteins and thus do not reach the necessary depth for untargeted biomarker discovery. Recently, Keshishian and colleagues published the most comprehensive plasma proteomics study, identifying 5340 proteins (4591 protein groups) after extensive sample fractionation [3]. However, this approach does not allow high-throughput analysis of multiple clinical samples, as needed when searching and establishing a bonafide biomarker. A distinct kind of

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_33,

fractionation strategy involves the isolation of plasma-derived extracellular vesicles that are shed from cells all over the body and can be released into the bloodstream [4, 5]. There are three main types of extracellular vesicles, which differ in their size, content, and secretion mechanism [6, 7] (Fig. 1). Exosomes (50–100 nm) are secreted through the multivesicular body system [8]; microparticles (100–1000 nm) are shed from plasma membrane invaginations [9]; and apoptotic bodies (1–4 μm) are formed during apoptotic cell death and often contain organelles [10]. Microparticles, which are the focus of our method, bud from the plasma membrane and have an average surface area of 1600 mm^2 and volume of 53 nl [11]. Such dimensions allow for a broad sampling of the cellular and membrane proteome of the cell of origin, thus implying their great potential in the plasma biomarker discovery field. While exosome extraction requires ultrahigh-speed centrifugation [12], often accompanied by density gradients or the use of commercial kits, microparticles can be easily isolated from the plasma by a simple high-speed centrifugation step. In recent years, there is a growing interest in proteomic profiling of extracellular vesicles as the source of biomarkers and as mediators of disease mechanisms [13, 14]. In this manuscript, we describe a recently developed workflow for simple plasma microparticle extraction with highly reproducible deep proteomics coverage of over 3200 proteins in a single shot and over 3600 proteins in triplicate analyses [15]. This protocol was developed for plasma samples, but can also be readily applied to other body fluids.

Beyond the high proteome coverage, biomarker discovery requires also accurate quantification. The method we describe here focuses on Stable Isotope Labeling with Amino acids in Cell culture (SILAC)-based quantification [16], which allows both relative

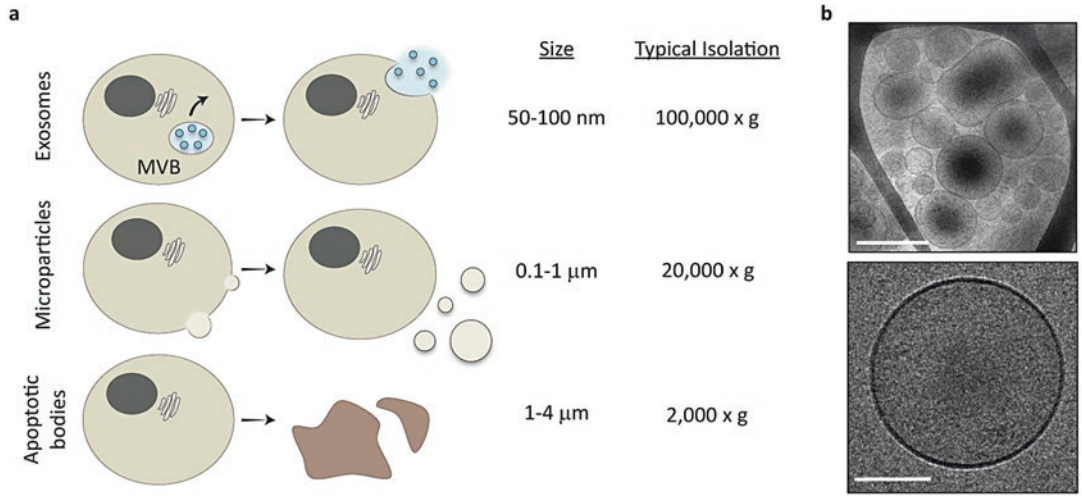


Fig. 1 The main types of extracellular vesicles. (**a**) The different types of vesicles differ in their size and secretion mechanism, as well as in the sedimentation protocol of these vesicles. *MVB* multivesicular body. (**b**) Cryo-electron microscopy images of plasma-derived microparticles. *Top*, scale bar is 500 nm; *bottom*, scale bar is 100 nm

and absolute quantification. However other quantification techniques, such as label-free analysis [17] and chemical labeling techniques, TMT for instance, can be applied as well (Fig. 2). In the SILAC-based approach, metabolically labeled cell lines serve as an internal standard, which is combined with each microparticle sample prior to trypsin digestion [18]. For absolute quantification of selected proteins (e.g., proteins that comprise a predictive signature that was identified by a preceding relative quantification

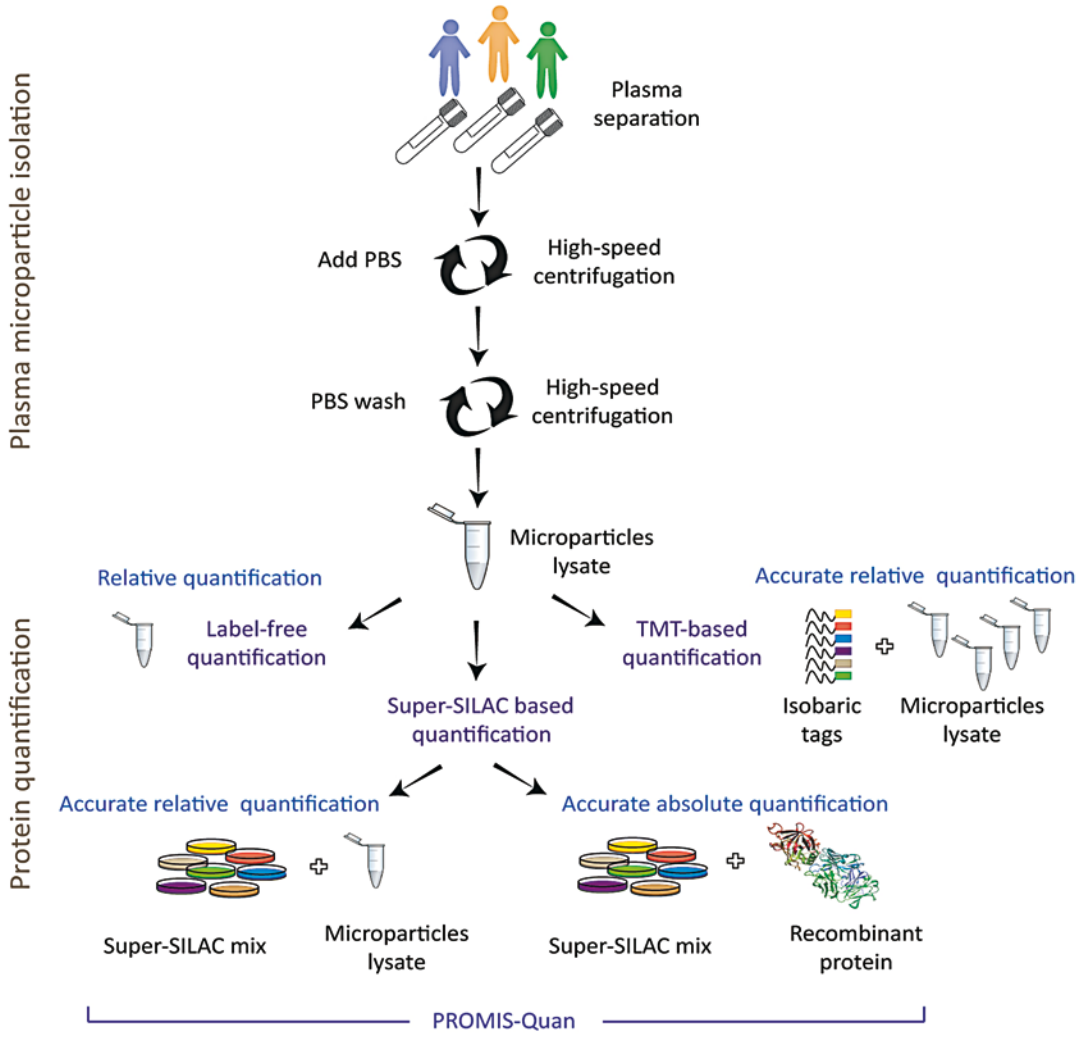


Fig. 2 Schematic representation of the method workflow. Plasma microparticle isolation begins with plasma separation, followed by a high-speed centrifugation step (20,000 × *g*) and an ice-cold PBS wash. Protein quantification can be performed using label-free algorithms, chemical labeling methods (e.g., TMT labeling), or SILAC-based quantification. The latter can be applied in a dual mode; first, highly accurate relative quantification can be achieved by adding a super-SILAC mix as an internal standard; next, the absolute concentration of selected proteins can be measured when combining an unlabeled recombinant protein of interest with the super-SILAC mix. This enables absolute quantification retrospectively to all samples

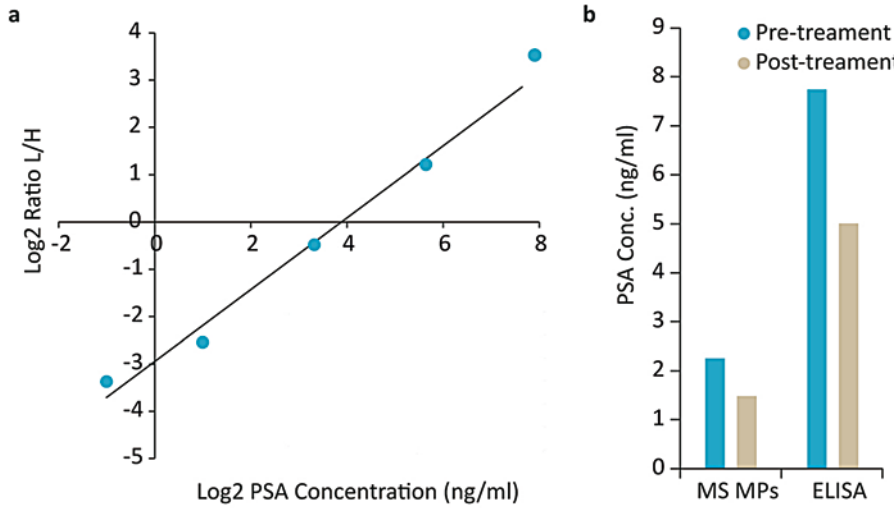


Fig. 3 Absolute quantification of a selected protein. (**a**) An example for absolute quantification. Increasing concentrations of prostate-specific antigen (PSA) were combined with the super-SILAC mix to create a calibration curve and subsequently extract the exact concentration of PSA in the standard. L, light; H, heavy. (**b**) Both soluble PSA, as measured using ELISA kit, and MS-based measurements of PSA levels in microparticles (MPs) show identical decrease of 35% upon treatment

analysis), absolute levels can be determined in the SILAC standard, and from these measurements, the absolute amounts can be extrapolated for each of the experimental samples (Fig. 3). The absolute quantification of the protein is thus obtained retrospectively for all of the samples, without having to rerun them again. This dual mode of SILAC-based relative and absolute quantification was introduced recently as PROMIS-quan (PROteomics of MIcroparticles using Super-SILAC Quantification).

2 Materials

All solutions must be prepared using Milli-Q ultrapurified water.

2.1 Microparticles Extraction

1. Phosphate-buffered saline (PBS) × 1.
2. 1 M Tris pH 8.5 (*see* **Note 1**).
3. Lysis buffer: 6 M urea, 2 M thiourea in 0.1 M Tris buffer pH 8.5 (*see* **Note 2**).
4. ABC buffer: 50 mM ammonium bicarbonate (ABC) in DDW (*see* **Note 3**).
5. Refrigerated benchtop microcentrifuge suitable for 20,000 × *g*.

2.2 In-Solution Trypsin Digestion, Stage Tipping, and Elution

1. DTT 1 M in 50 mM ABC buffer (*see* **Note 4**).
2. Iodoacetamide 0.5 M in 50 mM ABC buffer (*see* **Note 5**).
3. Trypsin/Lys-C Mix, Mass Spec Grade (Promega).

4. Sequencing grade modified trypsin (Promega).
5. Trifluoroacetic acid (TFA).
6. Methanol LC-MS grade.
7. Acetonitrile LC-MS grade.
8. LoBind eppendorf tubes.
9. Thermomixer.
10. Empore C18 disks for StageTips.

2.3 SILAC-Based Quantification

1. SILAC cell culture medium (without lysine and arginine).
2. Heavy-labeled arginine and lysine: $^{13}C_6{}^{15}N_2$-L-lysine (lysine 8) and $^{13}C_6{}^{15}N_4$-L-arginine (arginine 10).
3. Dialyzed serum.
4. Antibiotics.
5. Recombinant protein of interest (for absolute protein quantification).

2.4 MS Run and Bioinformatics Analysis

1. Buffer A: 0.1% formic acid in DDW (*see* **Note 6**).
2. Buffer B: 80% acetonitrile, 0.1% formic acid (*see* **Note 7**).
3. Buffer A*: 2% acetonitrile, 0.1% TFA (*see* **Note 8**).
4. High-resolution mass spectrometer (MS). Preferably Orbitrap-based instrument.
5. Nanoflow HPLC system.
6. Cappilary column 75 μm inner diameter, 50 cm long.
7. Vacuum concentrator (SpeedVac).
8. Data analysis software, preferably MaxQuant [19].
9. Fasta file of the examined organism.

3 Methods

3.1 Plasma Separation

1. Collect the blood sample into EDTA blood collection tubes (*see* **Note 9**). Plasma must be further separated within 1 h. Blood tubes should be shaken mildly at room temperature until the next step.
2. Centrifuge blood sample at 1500 × *g* for 10 min at room temperature. The blood will be separated into three layers: the bottom one includes the erythrocytes, middle one is the white blood cells (buffy coat), and the top layer contains the plasma.
3. Transfer the plasma to a new tube.

4. Centrifuge sample at 1500 × *g* for 10 min at room temperature.
5. Freeze 0.5 ml aliquots of plasma samples and store at −80 °C (*see* **Note 10**).

3.2 Isolation of Plasma Microparticles

1. Thaw plasma samples gradually on ice (*see* **Note 11**).
2. Centrifuge samples at 3300 × *g* for 20 min at 4 °C to remove potential particle contaminations. Transfer the supernatant to a new tube.
3. Dilute the supernatant two-fold in ice-cold PBS.
4. Centrifuge at 20,000 × *g* at 4 °C for 1 h, and use the slow break option.
5. Remove supernatant gently. Pellet is barely visible, if at all. Do not dry the pellet (*see* **Note 12**).
6. To further remove residual core plasma proteins add ice-cold PBS (the same volume as in **step 3**). Do not pipette. Vortex mildly (*see* **Note 13**).
7. Centrifuge at 20,000 × *g* at 4 °C for 1 h, slow brake. Pellet is barely visible.
8. Solubilize the pellet in 30 μl lysis buffer containing 6 M urea, 2 M thiourea in 50 mM ammonium bicarbonate (*see* **Note 14**).
9. Transfer the sample into a LoBind Eppendorf tube.

3.3 Super-SILAC Mix Preparation

The super-SILAC mix is a combination of cell lysates originating from sources that altogether represent the complexity of the examined proteome [18]. In our work, we selected the following cell lines: HeLa, MDA-MB-231, HepG2, Jurkat, RKO, LNCaP, and U2OS. The appropriate combination should be designed by the researcher, based on the question of interest, while the main guideline is to select cells that best represent the examined tissue. Super-SILAC design and preparation have been previously thoroughly described [20–22] and will be briefly described here:

1. Culture cells for ten doublings in SILAC cell culture medium with dialyzed serum and 1:2000–2500 dilution of lysine 8 and arginine 10. Amino acid concentration may require adjustment based on labeling checks.
2. Lyse cells with buffer containing 6 M urea, 2 M thiourea in 50 mM ammonium bicarbonate.
3. Perform labeling check for each labeled cell line individually to make sure that the proteins are indeed fully labeled and that there is a minor conversion of proline to arginine (to that end, follow the steps described below for in-solution trypsin digestion and stage tipping).

4. Mix equal protein amounts of the different cell line lysates. The lysate mix can be stored in aliquots in −80 °C and is stable for 1 year. It is recommended to prepare super-SILAC standard in excess, to ensure it is sufficient for the entire project and potentially additional future projects.

3.4 In-Solution Trypsin Digestion

1. The super-SILAC mix should be combined with each microparticle lysate at a 1:1 ratio (w/w). Standardization experiments showed a yield of 5 µg microparticle protein from 0.5 ml plasma. Accurate protein measurements can be determined with a variety of protein determination methods.
2. Reduce samples with 1 mM DTT and shake vigorously at room temperature for 30 min.
3. Alkylate samples with 5 mM iodoacetamide. Shake vigorously at room temperature for 30 min. Cover the tubes well, as the reaction is sensitive to light.
4. Add Lys-C/trypsin mix (1:100 enzyme to protein ratio).
5. Dilute lysates fourfold in 50 mM ammonium bicarbonate.
6. Add trypsin for overnight digestion (1:50 enzyme to protein ratio).
7. After complete digestion, add 0.1% TFA to the peptides.

3.5 Stage Tipping

1. Prepare StageTips by inserting three layers of Empore C18 disks into a 200 µl pipette tip at a distance of approximately 0.5 cm from the tip end [23]. Make sure that the surface area that points toward the inside on the tip is flat.
2. Activate 3 × C18 disk StageTips with 100 µl methanol. Centrifuge 1000–2000 × *g* for 2–3 min. Ensure that there is no liquid above the C18 disks; repeat centrifugation step if needed. Discard the flow-through.
3. Wash StageTips with 100 µl buffer B. Centrifuge as described in **step 2**.
4. Wash StageTips twice with 100 µl buffer A. Centrifuge as described in **step 2**.
5. Load samples on the StageTips. Centrifuge as described in **step 2**.
6. Wash StageTips twice with 100 µl buffer A. Centrifuge as described in **step 2**.
7. Store StageTips at 4 °C until elution.

3.6 Elution and MS Run

1. Elute peptides with 80 µl buffer B.
2. Evaporate acetonitrile using vacuum concentrator (preferably until 2–4 µl of total volume is left).
3. Complete to 6–7 µl total volume with buffer A*.
4. Inject 3 µl of peptides for each LC-MS/MS run.

5. Run gradient in nanoHPLC coupled to high-resolution mass spectrometer using buffer A and buffer B as the mobile phase.
6. Process raw data using suitable data processing software.

3.7 SILAC-Based Absolute Quantification of Selected Proteins

1. Prepare a standard curve of the recombinant protein of interest. Plan the concentration of the standards in the curve based on the presumed concentration of the protein in the plasma microparticles.
2. Add recombinant protein standard to the super-SILAC mix. Use the same super SILAC amounts as used in the previous steps.
3. Follow the instructions for in-solution trypsin digestion (Subheading 3.4) and stage tipping (Subheading 3.5).
4. Plot the standard curve using log2 ratio L/H of each concentration. Fit the trend line of the plot based on linear regression, and assign the normalized log2 ratio L/H to the linear equation to compute the final concentration of the examined protein in each sample (Fig. 3).

4 Notes

1. Dissolve 60.57 g Trizma base in 400 ml DDW. Titrate using HCl until pH reaches 8.5 at room temperature. Complete to 500 ml total volume and check pH again.
2. Dissolve 18.018 g urea and 7.612 g thiourea in 0.1 M Tris buffer and complete to 50 ml. The lysis buffer can be stored in −20 °C in aliquots.
3. Dissolve 1.5812 g ammonium bicarbonate in 400 ml DDW. Stock solution can be stored at room temperature.
4. Dissolve 1.542 g DTT in 10 ml of 50 mM ABC buffer. Stock solution can be stored in aliquots in −20 °C.
5. Dissolve 0.925 g iodoacetamide in 10 ml of 50 mM ABC buffer. Stock solution can be stored in aliquots in −20°C. Iodoacetamide is a teratogen. Use mask when weighing the powder.
6. Add 0.5 ml formic acid into 499.5 ml DDW.
7. Add 0.5 ml formic acid into 400 ml acetonitrile (LC-MS grade) and 99.5 ml DDW (HPLC/LC-MS grade).
8. Add 2 ml of acetonitrile and 0.1 ml TFA to 97.9 ml DDW. Stock solution can be stored at room temperature.
9. To minimize technical variability [24], collect all blood samples for a specific project using the same protocol and materials (anticoagulant in tube, syringe gauge, etc.).
10. To reach the indicated proteome depth, the lowest recommended plasma volume that should be used for this protocol is

0.5 ml. Since different samples have different microparticle counts (specifically, the number of microparticles may change upon disease, which is the basis for some diagnostic tests [25]), also the total protein amounts differ between different samples. Thus we strongly recommend using the same starting plasma volume for all samples in a given project.

11. To avoid microparticle lysis, it is critical to thaw frozen plasma samples gradually in ice. Avoid freeze-thaw cycles of the plasma samples.
12. The microparticle pellet is extremely delicate. Be very cautious when separating the pellet from the supernatant. Since the pellet is barely visible and very gentle, mark the tube orientation and the expected location of the pellet prior to centrifugation to avoid detaching it while removing the supernatant.
13. When adding PBS before the second centrifugation, do not resuspend the pellet, as a portion of the microparticles might cling to the tip. Instead, add the PBS and vortex it mildly. Additional PBS washes can be applied if the samples still contain high percentage of core plasma proteins.
14. Do not heat samples with urea-based lysis buffer above 30 °C all throughout the procedure to avoid protein modifications that would result in lower protein indentification rates.

References

1. Hanash SM, Pitteri SJ, Faca VM (2008) Mining the plasma proteome for cancer biomarkers. Nature 452:571–579
2. Baker ES, Liu T, Petyuk VA et al (2012) Mass spectrometry for translational proteomics: progress and clinical implications. Genome Med 4:63
3. Keshishian H, Burgess MW, Gillette MA et al (2015) Multiplexed, quantitative workflow for sensitive biomarker discovery in plasma yields novel candidates for early myocardial injury. Mol Cell Proteomics 14:2375–2393
4. Voloshin T, Fremder E, Shaked Y (2014) Small but mighty: microparticles as mediators of tumor progression. Cancer Microenviron 7(1–2):11–21
5. D'souza-Schorey C, Di Vizio D (2014) Biology and proteomics of extracellular vesicles: harnessing their clinical potential. Expert Rev Proteomics 11:251–253
6. Xu R, Greening DW, Zhu HJ et al (2016) Extracellular vesicle isolation and characterization: toward clinical application. J Clin Invest 126:1152–1162
7. Raposo G, Stoorvogel W (2013) Extracellular vesicles: exosomes, microvesicles, and friends. J Cell Biol 200:373–383
8. Greening DW, Gopal SK, Xu R et al (2015) Exosomes and their roles in immune regulation and cancer. Semin Cell Dev Biol 40:72–81
9. Yuana Y, Bertina RM, Osanto S (2011) Pre-analytical and analytical issues in the analysis of blood microparticles. Thromb Haemost 105:396–408
10. Taylor RC, Cullen SP, Martin SJ (2008) Apoptosis: controlled demolition at the cellular level. Nat Rev Mol Cell Biol 9:231–241
11. Van Der Pol E, Coumans F, Varga Z et al (2013) Innovation in detection of microparticles and exosomes. J Thromb Haemost 11(Suppl 1):36–45
12. Greening DW, Xu R, Ji H et al (2015) A protocol for exosome isolation and characterization: evaluation of ultracentrifugation, density-gradient separation, and immunoaffinity capture methods. In: Posch A (ed) Methods in molecular biology. Springer, New York, Heidelberg, Dordrecht, London
13. Minciacchi VR, Freeman MR, Di Vizio D (2015) Extracellular vesicles in cancer: exosomes, microvesicles and the emerging role of large oncosomes. Semin Cell Dev Biol 40:41–51

14. D'souza-Schorey C, Clancy JW (2012) Tumor-derived microvesicles: shedding light on novel microenvironment modulators and prospective cancer biomarkers. Genes Dev 26:1287–1299
15. Harel M, Oren-Giladi P, Kaidar-Person O et al (2015) Proteomics of microparticles with SILAC quantification (PROMIS-Quan): a novel proteomic method for plasma biomarker quantification. Mol Cell Proteomics 14:1127–1136
16. Ong SE, Mann M (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat Protoc 1:2650–2660
17. Cox J, Hein MY, Luber CA et al (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13:2513–2526
18. Geiger T, Cox J, Ostasiewicz P et al (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. Nat Methods 7:383–385
19. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372
20. Geiger T, Wisniewski JR, Cox J et al (2011) Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. Nat Protoc 6:147–157
21. Shenoy A, Geiger T (2015) Super-SILAC: current trends and future perspectives. Expert Rev Proteomics 12:13–19
22. Pozniak Y, Geiger T (2014) Design and application of super-SILAC for proteome quantification. In: Warscheid B (ed) Methods in molecular biology. Springer, New York, Heidelberg, Dordrecht, London
23. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2:1896–1906
24. Dinkla S, Brock R, Joosten I et al (2013) Gateway to understanding microparticles: standardized isolation and identification of plasma membrane-derived vesicles. Nanomedicine (Lond) 8:1657–1668
25. Guiducci S, Distler JH, Jungel A et al (2008) The relationship between plasma microparticles and disease manifestations in patients with systemic sclerosis. Arthritis Rheum 58:2845–2853

Chapter 34

Bronchoalveolar Lavage: Quantitative Mass Spectrometry-Based Proteomics Analysis in Lung Diseases

Ana Sofia Carvalho and Rune Matthiesen

Abstract

Bronchoalveolar lavage (BAL) fluid, obtained by a relatively noninvasive procedure, is used as a practice for diagnosis of various lung diseases as source of cells for cytology analysis. The acellular component of BAL potentially can complement and be a key for the establishment of diagnostic or as a prognostic indicator. This chapter discusses the aspects of standardization of BAL sample preparation and processing and its implications on the BAL fluid proteome quantitative analysis by high-throughput mass spectrometry. The detailed conditions for quantitative analysis of BAL proteome in the context of biomarker discovery are introduced.

Key words Bronchoalveolar lavage fluid, Proteome, Mass spectrometry, Diagnosis, Lung diseases, Biomarkers

1 Introduction

Bronchoalveolar lavage (BAL) fluid obtained by fiber-optic bronchoscopy is widely used for the clinical evaluation of patients with various forms of respiratory diseases (e.g., lung cancer and inflammatory diseases). BAL was first used in 1979 by Bell and Hook to study patients with pulmonary alveolar proteinosis [1]. BAL assures a unique representation of the molecular and cellular components from the peripheral airspaces and small airways (airway epithelial and immune cells) as well as the extracellular lining fluid (ELF) and consists of soluble molecules such as phospholipids, proteins, peptides, nucleic acids, and cells [2]. Exploring BAL fluid by mass spectrometry-based proteomics for clinical research of lung disease mechanisms has been carried out since the wide spreading of flexible fiber-optic bronchoscopy; however, we are still a long way from the bench-to-bedside research using BAL proteomics in lung disease diagnostics. Initially, protein identification and expression were studied by using two-dimensional (2D) gel electrophoresis [3]. These early studies led to the characterization of the BAL fluid proteome relevant to deepened the understanding of the pathogenesis

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_34,

of diseases such as asthma [4], cystic fibrosis [5], sarcoidosis [6, 7], chronic obstructive pulmonary disease [8], and idiopathic pulmonary fibrosis (IPF) [9, 10]. BAL proteome is enriched for plasma proteins (albumin and immunoglobulins), immune inflammatory mediator proteins, proteolytic factors, heat shock, and complement proteins mainly [11, 12]. Further, than the proteome landscape, BAL proteomics analysis showed specific disease markers and also differential protein expression between diseases such as sarcoidosis and IPF [13–15]. The stratification of a disease in clinical subgroups is a recurrent goal in clinical research [16–18]. BAL proteomics analysis can contribute to better define mixed phenotypes for diagnosis, prognosis, and therapy.

The development of high-throughput and quantitative MS proteomics enabled the detection of low abundant proteins in a complex mixture such as BAL fluid [19, 20]. In a recent study, the total number of identified proteins in BAL was 5779 isoforms and 2195 when collapsed into encoding genes [9]. However, the standardization of BAL procedure is still a critical issue, and routine processing of BAL is required to assure reproducibility, accuracy, and comparison of the results. Over the years, the European Respiratory Society (ERS) task forces have drafted documents on methods for performing BAL [21, 22]. Currently, the recommendations include the instillation by fiber-optic bronchoscope of a volume of sterile saline solution (0.9% NaCl) ≥100 mL in adults (240 mL is recommended) divided into four aliquots and the fluid recovered by gentle aspiration. The volume recovered is dependent on the region of the lung where the fluid is instilled and the patient lung condition. Regarding fluid recovery, there is a controversy if the first retrieved aliquot must be kept seperate from the sequential ones, and both approaches are currently being used. Despite the procedure, aliquots are collected and cleared from cells by filtration or centrifugation. The acellular component can be stored at −80 °C until further analysis. The BAL sample variability mainly causes a sample dilution problem. To overcome this problem ERS task force recommends that measurements of two or more components can be expressed as proportions relative to each other using the same principal as for differential counting of BAL cells. Moreover, it is also recommended that data on lavage fluid input, recovery volumes, and percent recoveries should be included in reports to confirm that the quantitative differences are not consequences of the sample variability [22].

Upon BAL analysis the cellular component is often used for cell count, culture, and cytology. The excess acellular part is typically discarded. Therefore, excess BAL specimens discarded after cell collection can be utilized and can add to the diagnosis of malignancies non-visible during the bronchoscopy procedure. Analysis of acellular (soluble) BAL proteome proved to be less cumbersome compared with other body fluids such as serum or plasma since depletion of the most abundant proteins is not a

requirement for LC-MS analysis. Furthermore, in patients with lung cancer, BAL is in most cases in direct contact with the tumor in contrast to other body fluids.

Research groups with experience in BAL proteomics would have an advantage by working in a global project in which guidelines could be defined that includes BAL collection procedures and information reports; BAL sample preparation methods for mass spectrometry-based proteomics, data analysis, and normalization could be standardized so data can be reproduced and comparable between different working groups. This initiative could contribute to develop and implement BAL proteomics, for example, in the diagnostic evaluation of patients with suspected lung cancer who lack an undefined diagnosis despite chest radiographic abnormality or to prioritize patients for follow-up and surgery.

The information obtained from sequencing the human genome and especially specific patients' tumors will enable us to improve the protein sequence databases leading to an increased number of identified proteins. Especially personalized RNA-seq of tumors combined with proteomics will allow identification of fusion proteins, mutated proteins, isoform distribution, and abnormal expression on the transcriptomics and proteomics level. Personalized RNA-seq combined with proteomics calls for development of novel experimental protocols and computational methodologies.

In summary, more efforts on defining international guidelines and promotion of the standards for operators in the sector must be prioritized. Personalized RNA-seq and proteomics will be expensive and therefore preferably used on an explorative level as an aid to defining a more cost-effective methodology for personalized diagnosis and patient management.

2 Materials

2.1 BAL Sampling

1. Sterile saline solution: 0.9% (w/v) NaCl.
2. Ultrapure water (double distilled, deionized, >18 Ω) is used for all reagent preparations.
3. Protease inhibitor cocktail tablets (Complete, Mini, Roche).
4. HPLC-grade ice-cold acetone.
5. Pierce BCA Protein Assay Kit (Thermo Fisher Scientific).
6. Low-speed centrifuge with swing-bucket rotor and appropriate tubes.
7. Ultracentrifuge with fixed-angle rotor and appropriate tubes.
8. Microplate Absorbance Spectrophotometer (Bio-Rad).

2.2 BAL Preparation for MS Analysis

1. 1 M Dithiothreitol (DTT). Add 1.54 g DTT to 10 mL deionized H_2O.
2. 8 M urea, 0.1 M HEPES buffer, and pH 8.5. Add 48.0 g urea and 2.38 g HEPES to 80 mL deionized H_2O. Adjust pH to 8.5. Bring final volume to 0.1 L.
3. 0.05 M iodoacetamide in 8 M urea, 0.1 M HEPES buffer, and pH 8.5. To 25 mL 8 M urea, 0.1 M HEPES buffer, and pH 8.5, add 2.3 g iodoacetamide. Protect from light.
4. 40 mM NH_4HCO_3 and 10 mM NH_4HCO_3. To prepare 0.1 M NH_4HCO_3 stock solution, pH 8.5. Add 0.79 g NH_4HCO_3 to 80 mL deionized H_2O. Adjust pH to 8.5. Bring final volume to 0.1 L. For 40 mM NH_4HCO_3, add 40 mL 0.1 M NH_4HCO_3, pH 8.5 to 60 mL deionized H_2O. For 10 mM NH_4HCO_3, add 10 mL 0.1 M NH_4HCO_3, pH 8.5 to 90 mL deionized H_2O.
5. 0.05 μg/μL trypsin in 40 mM NH_4HCO_3, pH 8.5. Reconstitute 100 μg of Trypsin Gold, Mass Spectrometry Grade (Promega) in 100 μL 50 mM acetic acid to 1 μg/μL. Add 50 μL 1 μg/μL of Trypsin Gold to 950 μL 40 mM NH_4HCO_3, pH 8.5.
6. Thermomixer.
7. Microcentrifuge with fixed-angle rotor.

3 Methods

3.1 BAL Sampling

BAL collection procedure here described was applied to patients with suspected lung cancer, and BAL was targeted toward affected lung segments. In most cases, the procedure was performed by wedging the bronchoscope in a subsegmental bronchus (*see* **Note 1**).

1. Three lavages were performed using approximately 50 mL of 0.9% saline solution per lavage.
2. The recovered fluid was placed at 4 °C immediately.
3. BAL was centrifuged at 320 × *g* for 10 min at 4 °C to remove the cellular fraction (*see* **Note 2**).
4. The resulting cell-free supernatant was immediately aliquoted (*see* **Note 3**).
5. 40 μL of 25× protease inhibitor cocktail stock solution was added to each aliquot of 1 mL BAL sample and frozen at −80 °C until further analysis (*see* **Note 4**).

3.2 BAL Preparation for MS Analysis

1. Thaw an aliquot of each BAL sample on ice (*see* **Note 5**).
2. Measure BAL sample protein concentration using Pierce BCA Protein Assay Kit according to the manufacturer's instructions.
3. Add six times the volume of ice-cold acetone to BAL sample corresponding to 100 μg, and incubate overnight at −20 °C (*see* **Note 6**).

4. Spin down the sample at 4 °C for 20 min at 13–15,000 × *g*.
5. Solubilize the proteins in 100 μL 8 M urea, 0.1 M HEPES, 2% SDS, and pH 8.5.
6. Mix 100 μg of total protein with 10 μL of 1.0 M DTT.
7. Incubate in a thermomixer for 3 min at 40 °C.
8. Centrifuge at 14,000 × *g* for 2 min and transfer the supernatant (a pellet was observed) to a Microcon YM-30 (Millipore).
9. Add 100 μL 8 M urea, 0.1 M HEPES, and pH 8.5 and centrifuge for 25 min at 14,000 × *g* at 20 °C. Discard the flow through from the collection tube. Repeat this step six times.
10. Add 100 μL of 0.05 M iodoacetamide in 8 M urea, 0.1 M HEPES, and pH 8.5, and mix in a thermomixer for 1 min. Incubate without mixing for 20 min in the dark.
11. Centrifuge for 35 min at 14,000 × *g* at 20 °C. Discard the flow through from the collection tube.
12. Add 100 μL of 8 M urea, 0.1 M HEPES, and pH 8.5. Centrifuge at 14,000 × *g* for 25 min at 20 °C. Discard the flow through from the collection tube. Repeat this step four times.
13. Add 100 μL of 40 mM NH_4HCO_3 and pH 8.5. Centrifuge at 14,000 × *g* for 25 min at 20 °C. Discard the flow through from the collection tube. Repeat this step four times.
14. Add 40 μL of 40 mM NH_4HCO_3 and pH 8.5 with trypsin (0.05 μg/μL) and mix for 1 min.
15. Incubate the filters at 37 °C O/N. Replace collection tube. Centrifuge filters for 20 min at 14,000 × *g* at 20 °C.
16. Collect the flow through.
17. Add 40 μL of 10 mM NH_4HCO_3 and centrifuge at 14,000 × *g* at 20 °C and collect the flow through. The total sample volume is approximately 80 μL in 25 mM NH_4HCO_3.

3.3 BAL Proteomics Data Analysis

The peptides obtained in **step 17** of Subheading 3.2, after desalting (ZipTip, Millipore) and vacuum centrifugation, are reconstituted in 10–20 μL of 5% formic acid and can be readily analyzed by LC-MS/MS. The following is a brief description on data analysis across BAL samples obtained by high-resolution mass spectrometry and the approach undertaken to overcome BAL sample variability.

The technical issues concerning database searching of MS-based proteomics data have been reviewed elsewhere [23, 24]. Furthermore, several programs have been proposed for identification and quantitation of proteins (e.g., MaxQuant [25], VEMS [26], and X!Tandem and related tools [27]). The output from these programs is typically a quantitative matrix where each row corresponds to a specific protein or a group of proteins that cannot be distinguished based on the MS data. Rows representing groups of proteins can also be unfolded

and assigned into evidence groups [28, 29]. Each column corresponds in general to a specific sample. The quantitative data which for clinical samples are typically based on label-free quantitation can be based in spectral counting, ion counts in survey scans, or both. We prefer to use the statistical programing language R for all subsequent steps such as data integration, statistical analysis, and data plotting (e.g., using the limma package and related packages [30]). In the early days of BAL proteomics, the discussion was focused on normalization procedures (e.g., total protein versus specific household protein). However, large-scale data sets obtained with today's MS instruments generate quantitative data of 3000 to 5000 proteins in BAL which allows the use of normalization procedures that standardize the overall distribution of the quantitative values across samples (e.g., methods such as RMA [31]). It is recommended to compare statistical analysis of normalized data with the original data points since we sometimes observe artifacts being introduced by the normalization procedure. This is a concern especially for data sets with many missing data points.

4 Notes

1. Human bronchoalveolar lavage requires topical lidocaine anesthesia and endotracheal tube manipulation and must thus be performed by a competent physician in an adequate environment. Studies using BAL must obtain informed consent from patients.
2. Mucus plugs can be noticed in BAL samples which can be removed by filtration through nylon mesh.
3. To avoid freeze-thaw cycles, BAL samples must be aliquoted before storage.
4. Protease inhibitor prevents protein degradation during sample handling.
5. BAL protein concentration can differ significantly between patients. Different volumes for each sample can be required to obtain 100 μg for MS analysis.
6. Precipitation overnight ensures a minimum of 6 h incubation time. To accelerate precipitation, it can be performed at −80 °C for a minimum of 4 h.

Acknowledgment

All experiments including MS analysis were supported by Fundação para a Ciência e Tecnologia project EXPL/DTP-PIC/0616/2013. RM is supported by FCT Investigator Program 2012 (IF/01002/2012). A.S.C. is supported by grant SFRH/BPD/85569/2012 funded by Fundação para a Ciência e Tecnologia.

References

1. Bell DY, Hook GE (1979) Pulmonary alveolar proteinosis: analysis of airway and alveolar proteins. Am Rev Respir Dis 119:979–990. doi:10.1164/arrd.1979.119.6.979
2. Reynolds HY (2000) Use of bronchoalveolar lavage in humans—past necessity and future imperative. Lung 178:271–293
3. Govender P, Dunn MJ, Donnelly SC (2009) Proteomics and the lung: analysis of bronchoalveolar lavage fluid. Proteomics Clin Appl 3:1044–1051. doi:10.1002/prca.200900032
4. Larsen K et al (2006) Specific haptoglobin expression in bronchoalveolar lavage during differentiation of circulating fibroblast progenitor cells in mild asthma. J Proteome Res 5:1479–1483. doi:10.1021/pr050462h
5. McMorran BJ et al (2007) Novel neutrophil-derived proteins in bronchoalveolar lavage fluid indicate an exaggerated inflammatory response in pediatric cystic fibrosis patients. Clin Chem 53:1782–1791. doi:10.1373/clinchem.2007.087650
6. Kriegova E et al (2006) Protein profiles of bronchoalveolar lavage fluid from patients with pulmonary sarcoidosis. Am J Respir Crit Care Med 173:1145–1154. doi:10.1164/rccm.200507-1126OC
7. Sabounchi-Schutt F, Astrom J, Hellman U, Eklund A, Grunewald J (2003) Changes in bronchoalveolar lavage fluid proteins in sarcoidosis: a proteomics approach. Eur Respir J 21:414–420
8. Plymoth A et al (2003) Human bronchoalveolar lavage: biofluid analysis with special emphasis on sample preparation. Proteomics 3:962–972. doi:10.1002/pmic.200300387
9. Rottoli P et al (2005) Carbonylated proteins in bronchoalveolar lavage of patients with sarcoidosis, pulmonary fibrosis associated with systemic sclerosis and idiopathic pulmonary fibrosis. Proteomics 5:2612–2618. doi:10.1002/pmic.200401206
10. Rottoli P et al (2005) Cytokine profile and proteome analysis in bronchoalveolar lavage of patients with sarcoidosis, pulmonary fibrosis associated with systemic sclerosis and idiopathic pulmonary fibrosis. Proteomics 5:1423–1430. doi:10.1002/pmic.200301007
11. Noel-Georis I, Bernard A, Falmagne P, Wattiez R (2002) Database of bronchoalveolar lavage fluid proteins. J Chromatogr B Analyt Technol Biomed Life Sci 771:221–236
12. Wattiez R, Falmagne P (2005) Proteomics of bronchoalveolar lavage fluid. J Chromatogr B Analyt Technol Biomed Life Sci 815:169–178. doi:10.1016/j.jchromb.2004.10.029
13. Magi B, Bargagli E, Bini L, Rottoli P (2006) Proteome analysis of bronchoalveolar lavage in lung diseases. Proteomics 6:6354–6369. doi:10.1002/pmic.200600303
14. Magi B et al (2002) Bronchoalveolar lavage fluid protein composition in patients with sarcoidosis and idiopathic pulmonary fibrosis: a two-dimensional electrophoretic study. Electrophoresis 23:3434–3444. doi:10.1002/1522-2683(200210)23:19<3434::AID-ELPS3434>3.0.CO;2-R
15. Wattiez R, Hermans C, Cruyt C, Bernard A, Falmagne P (2000) Human bronchoalveolar lavage fluid protein two-dimensional database: study of interstitial lung diseases. Electrophoresis 21:2703–2712. doi:10.1002/1522-2683(20000701)21:13<2703::AID-ELPS2703>3.0.CO;2-W
16. Zupa A et al (2012) A pilot characterization of human lung NSCLC by protein pathway activation mapping. J Thorac Oncol 7:1755–1766. doi:10.1097/JTO.0b013e3182725fc7
17. Hu J et al (2012) Expression patterns of USP22 and potential targets BMI-1, PTEN, p-AKT in non-small-cell lung cancer. Lung Cancer 77:593–599. doi:10.1016/j.lungcan.2012.05.112
18. Postma DS, Kerkhof M, Boezen HM, Koppelman GH (2011) Asthma and chronic obstructive pulmonary disease: common genes, common environments? Am J Respir Crit Care Med 183:1588–1594. doi:10.1164/rccm.201011-1796PP
19. Carvalho AS et al (2017) Bronchoalveolar lavage proteomics in patients with suspected lung cancer. Sci Rep 7:42190
20. Ortea I, Rodriguez-Ariza A, Chicano-Galvez E, Arenas Vacas MS, Jurado Gamez B (2016) Discovery of potential protein biomarkers of lung adenocarcinoma in bronchoalveolar lavage fluid by SWATH MS data-independent acquisition and targeted data extraction. J Proteome 138:106–114. doi:10.1016/j.jprot.2016.02.010
21. Haslam PL, Baughman RP (1999) Report of ERS Task Force: guidelines for measurement of acellular components and standardization of BAL. Eur Respir J 14:245–248
22. Haslam PL & Baughman RP (1999) Report of European Respiratory Society (ERS) Task Force: guidelines for measurement of acellular components and recommendations for standardization of bronchoalveolar lavage (BAL). Eur Respir Rev 9:25–157
23. Matthiesen R, Carvalho AS (2013) Methods and algorithms for quantitative proteomics by

mass spectrometry. Methods Mol Biol 1007: 183–217.doi:10.1007/978-1-62703-392-3_8

24. Matthiesen R, Azevedo L, Amorim A, Carvalho AS (2011) Discussion on common data analysis strategies used in MS-based proteomics. Proteomics 11:604–619. doi:10.1002/pmic.201000404
25. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372. doi:10.1038/nbt.1511
26. Carvalho AS et al (2014) Global mass spectrometry and transcriptomics array based drug profiling provides novel insight into glucosamine induced endoplasmic reticulum stress. Mol Cell Proteomics 13:3294–3307. doi:10.1074/mcp.M113.034363
27. Duncan DT, Craig R, Link AJ (2005) Parallel tandem: a program for parallel processing of tandem mass spectra using PVM or MPI and X!Tandem. J Proteome Res 4:1842–1847. doi:10.1021/pr050058i
28. Prieto G et al (2012) PAnalyzer: a software tool for protein inference in shotgun proteomics. BMC Bioinformatics 13:288. doi:10.1186/1471-2105-13-288
29. Matthiesen R et al (2012) SIR: deterministic protein inference from peptides assigned to MS data. J Proteome 75:4176–4183. doi:10.1016/j.jprot.2012.05.010
30. Law CW, Alhamdoosh M, Su S, Smyth GK, Ritchie ME (2016) RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Research 5:1408. doi:10.12688/f1000research.9005.1
31. Irizarry RA et al (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264. doi:10.1093/biostatistics/4.2.249

Chapter 35

Protein Multiplexed Immunoassay Analysis with R

Edmond J. Breen

Abstract

Plasma samples from 177 control and type 2 diabetes patients collected at three Australian hospitals are screened for 14 analytes using six custom-made multiplex kits across 60 96-well plates. In total 354 samples were collected from the patients, representing one baseline and one end point sample from each patient. R methods and source code for analyzing the analyte fluorescence response obtained from these samples by Luminex Bio-Plex® xMap multiplexed immunoassay technology are disclosed. Techniques and R procedures for reading Bio-Plex® result files for statistical analysis and data visualization are also presented. The need for technical replicates and the number of technical replicates are addressed as well as plate layout design strategies. Multinomial regression is used to determine plate to sample covariate balance. Methods for matching clinical covariate information to Bio-Plex® results and vice versa are given. As well as methods for measuring and inspecting the quality of the fluorescence responses are presented. Both fixed and mixed-effect approaches for immunoassay statistical differential analysis are presented and discussed. A random effect approach to outlier analysis and detection is also shown. The bioinformatics R methodology present here provides a foundation for rigorous and reproducible analysis of the fluorescence response obtained from multiplexed immunoassays.

Key words Multiplex immunoassay analysis, R programming language, Mixed-effects model, Proteomics, xMap, Quality, Bioinformatics

1 Introduction

xMap® bead-based multiplex immunoassays [1] are used in the life sciences, for high-throughput measurements and clinical screening of biological serum, plasma, and tissue samples for the concentrations of analytes such as cytokines, chemokines, and growth factors that are involved in cell signaling, cellular taxis, immune system responses, development, and cell death. Often, samples for test samples, standards, blanks, and controls are assayed in a 96-well plate format, 8 rows by 12 columns of reactions. Each of the 96 wells of

Electronic supplementary material: The online version of this chapter (doi:10.1007/978-1-4939-7057-5_35) contains supplementary material, which is available to authorized users.

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5_35,

assays contains one sample. The technology uses color-coded bead sets, one set per analyte, and each bead in a set is pre-coated with the same analyte-specific capture antibodies, and up to 100 different bead sets can be detected in theory [2], although in practice multiplex panels of less than 40 different analytes per assay are typically used. This number is often extended by using multiple assay types and different panels against the same samples.

In each assay, the bead-anchored analyte-specific antibodies capture the analyte of interest. Then typically biotinylated detection antibodies specific to the analyte of interest are added and complete the formation of a bead-antibody-antigen-antibody-biotin complex. Then phycoerythrin (PE)-conjugated streptavidin is added and used as the reporter. The optically encoded beads are separated from the assay matrix by particle-based flow cytometry. Dual lasers are used to interrogate each bead at the flow cytometer level. The first laser identifies the different bead sets, and the second laser excites PE, which is used for analyte quantification. The minimum number of beads used to count per analyte for statistical meaningful values is 25; however, in practice assays are set up, generally, to count at least 50 beads per analyte [3, 4].

Software, for example, the Bio-Plex Manager®, is used to control the Bio-Plex® xMAP flow cytometry instrument and for data collection and analyte concentration estimation. The software records the medium fluorescence response from the bead events for each analyte per sample. It can be set up to run in high and low sensitivity mode. However, for cytokine analysis low sensitivity is usually recommended. The Bio-Plex Manager® can export result data, per 96-well plate, as an Excel spreadsheet containing analyte fluorescence and concentration responses. Many options are available for the exported file format, but only the Bio-Plex multiple analyte layouts will be considered here.

2 Methods

There are software packages such as R's drLumi package [5] that allow uses to read and manipulate xPONENT®-derived multiplexed data and the Bio-Plex® analysis software, Data-Pro, for statistical analysis (*t*-tests and one-way ANOVA) and data visualization. Here only the Bio-Plex-exported result data and the analysis of its recorded fluorescence intensities (FI) using the R programming language [6], which is an interactive scripting environment for doing statistics, are explored.

2.1 The R Programming Language

R is a freely available programming environment, which provides a wide variety of statistical and graphical techniques: linear and nonlinear modeling, statistical tests, time series analysis, classification, clustering, heat maps, lattice plots, etc. You can get R, if you don't have it, from: http://www.r-project.org/.

CRAN (Comprehensive R Archive Network) is a worldwide network of FTP and web servers that store identical, up-to-date versions of code and documentation for R, for example, http://cran.us.R-project.org. R is extremely extendable via including R packages into your R session or sourcing R code files. Packages are generally user developed and contributed (free) to the R community via CRAN. Packages are formal implementation of R code and incorporate functions, data, and documentation. An R environment is created when the R interpreter is started. The top level environment is at the command prompt where the user inputs data and expressions, and this level is called the R global environment.

2.2 R Variable Types

R variables are created through assignment and are assigned values using one of three assignment operators: =, <−, and <<−. Here, only the equal symbol, **y** = **x**, is used to assign values to variables. Variables have scope and type: logical, numeric, complex, integer, character, vectors, matrices, arrays, data frames, lists, and factors. A vector is a 1D array of atomic values where each element is of the same basic type. Matrices are 2D vectors; arrays are generic matrices and vectors and can represent one- to N-dimensional data sets. A data frame is similar to a matrix in the sense that it has columns and rows, and its values can be obtained or assigned by indexing, that is: [**row**,**col**]. Data frame columns can also be specified and selected by name using the dollar sign; for example, **df**$**y** = **x**, where **df** is a data frame and **y** is the column name/header. While the contents of any given column must be of the same type, different columns can contain different data types.

2.3 R Factors

Categorical variables in R are known as factors, which are stored as a vector of integer values with a corresponding set of labels known as levels. Factors are used for summary statistics, plots, and statistical modeling. Factor levels by default are created from the unique factor values. For example, the immunoassay multiplex responses from each assay can be grouped using factor values defined for standards, controls, blank, and test samples. Patients and subjects identifiers can be defined as a patient or subject factor with factor values defined by the patient and sample identifier codes. If the levels of a categorical variable have a clear ordering other than lexicographical/alphanumerical, then that factor should be declared as an ordered factor. Ordinal variables in R are represented by ordered factors. For example, low, high, and medium can be seen as three levels of a dosage factor ordered such that low < medium < high. An ordered factor cannot be generated by simply specifying the levels of dosage such that low = 1, medium = 2, and high = 3 as R will treat this factor as a categorical variable rather than an ordinal value.

2.4 R Functions

Functions in R are similar to functions in other functional computer languages (C/C++, java, perl, etc). However, most conveniently, R uses named arguments which can be assigned default values. For example, **foo** = **function**(**x**,**y**=**x**) **print**(**c**(**x**,**y**)) defines **foo** to be a function that takes two arguments **x** and **y**, but **y** has a default value defined by **x**. This function concatenates the arguments together and prints them to the screen. You can call **foo** by **foo**(1), **foo**(**y**="world!", **x**="Hello"), or even with vectors, such as foo(1:3, 4:2), where **i:j** defines an increasing or a decreasing sequence of values from **i** to **j** in steps of 1. Therefore, R is highly vectorized—almost all operations work equally well on scalars and arrays. Function parameters are polymorphic as is the function return value. R functions can be nested within functions and parameter lists. To view the manual page for any R-distributed function, at the R command line prompt, enter **help(function.name)** or ?**function.name**.

2.5 R Formulas

The use of formulas in R is important to understand and the use of the tilde operator "~" that separates the left and right side of an R formula, such as **y ~ grp**, which specifies that **y** is modeled by the values or categories defined by **grp**. The number of elements in **y** and **grp** must be same, and **y** is the dependent variable, and **grp** is the independent variable, and often referred to as the explanatory variable or the predictor. The colon operator when used in a formula, such as **grp:treatment**, doesn't imply a sequence of values but rather generally the interaction or effect from two factors. Formulas are used to define relationships between the data components; they are not only used in statistics and regression analysis but also for creating plots and tables, that is, for specifying how to visualize data.

3 Methods

3.1 Plasma Samples and Patients

The details of the plasma samples used are given in [7]. The data analyzed represents the fluorescence responses from a multi-assay experiment on plasma samples collected from 177 patients at three Australian hospitals. The patients were classified as either control (**cntrl**) or type 2 diabetes (**T2D**) patients. Each patient supplied a baseline (**BL**) and end point (**EP**) samples. A total of 354 samples were analyzed in duplicate. End point samples were taken 6 months after the baseline sample.

3.1.1 Sample Metadata

The above experimental detail associated with each patient represents the sample metadata or the sample covariates for the experiment. This information should be collected as seen in Table 1. It's best represented by a table, a data frame, or as a CSV or Excel file, where each row is used to identify and represents each patient's

Table 1
Metadata file example for plasma samples

Hospital	Patient	Date	T.pt.	Cond	pid	Type
H1	SA	39979	EP	T2D	p021	X20
H2	AJ	40150	EP	T2D	p023	X28
H1	TT	40154	EP	cntrl	p024	X11
H1	PK	40423	BL	T2D	p054	X2
H2	DP	40641	BL	T2D	p054	X34
H1	SA	39783	BL	T2D	p055	X18
H2	VC	39798	BL	cntrl	p016	X1
H1	CD	39798	BL	cntrl	p036	X5
H1	HS	39499	BL	T2D	p008	X10
H1	MM	39456	BL	cntrl	p018	X25

Each row identifies each test sample whose immunoassay responses are located in one of 60 Bio-Plex-exported Excel files. The metadata Type field matches the Type field given in the Bio-Plex-Exported Excel file. The plate identifier column pid is used to identify the particular Excel file for each sample. The condition column, Cond, identifies each sample as coming from a control patient, cntrl, or a type 2 diabetes (T2D) patient. The T.pt. column is used to identify whether that sample is a baseline sample, BL, or an end point, EP, sample taken 6 months later. The metadata file also contains a date column which encodes the hospital's date of sample collection in integer form. The patient column encodes the patient identifiers, and the Hospital column identifies the three participating hospitals

sample analyzed. If a patient contributes several samples, then there should be several rows of information for that patient in the metadata file. Each row is associated to a patient using a unique patient ID code. Each column in the metadata file then defines the experimental factors of interest, such as the blocking and grouping factors and patient codes. Also the metadata assay file will have a column that links each sample's metadata to its associated results, which in turn will be contained in a separate assay result file. Figure 1 gives a schematic view of the nested relationship of the main experimental factors shown in Table 1 and the associated patient numbers.

The metadata table in Table 1 contains just seven columns of explanatory variables/covariates. The **Type** column is used to match the sample's metadata to the **Type** column given in each of the immunoassay result files. The plate identifier column **pid** is used to identify the associated result file. The other columns contain information, on the **Hospital** (**H1**, **H2**, or **H3**), where the actual plasma samples were collected; a patient identification code, **patient**; and time point readings (**T.pt**) that are associated with each sample, where **BL** represents baseline reading and **EP** represents an end point reading. All patients gave both samples. The metadata also contains each patient's condition (**Cond** column), and **cntrl** is used to identify the control patients, and **T2D** is used to identify the type 2 diabetes patients.

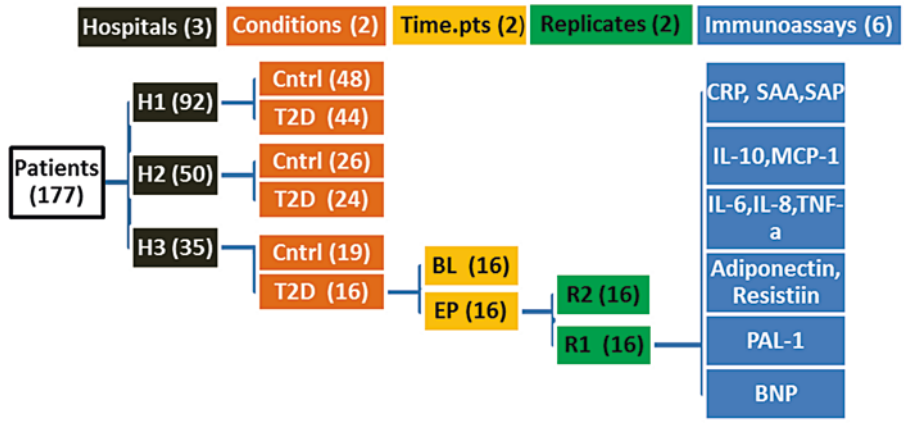


Fig. 1 Schematic tree representation of patient and experimental factors in a multi-assay immunoassay experiment. The number of items or patients is represented by the numbers in brackets. Notice that hospital H3 contributed plasma samples from 35 of the 177 patients. Of these 35 patients, 19 were controls and 16 were type 2 diabetes patients. Of the 16 type 2 diabetes patients, 16 baseline samples (BL) and 16 end point (EP) paired samples were obtained. From the 16 EP samples, 16 R1 and 16 R2 replicates were analyzed for 14 analytes using six immunoassay kits. From this graph it is seen that when taking the hospital into account, there are 12 ($3 \times 2 \times 2$) distinct test sample types

Table 2
Human analytes and kit details for each of the six immunoassay kits used here

Analytes	Kit	Plex	Standards	Blanks	Controls
CRP,SAA,SAP	k001	3	6	1	2
IL-10,MCP-1	k002	2	6	1	2
IL-6,IL-8,Leptin,TNF-a,VEGF	k003	5	7	1	2
Adiponectin, resistin	k004	2	7	1	2
PAI-1	k005	1	6	1	2
BNP	k006	1	7	1	2

3.1.2 Immunoassay Kits

Each plasma sample was analyzed for 14 analytes using 6 X10 immunoassays kits (*see* Table 2), that is, 60 plates, 10 plates per kit. As each sample was assayed in replicate, 9912 fluorescence responses were recorded. However, one analyte within one assay failed to give a test reading; therefore, 9911 fluorescence actual readings were obtained. Table 2 gives a summary of the immunoassays kits used here, and it is seen the number of analytes per kit varied from one to five (plex), and the number of standards per kit was either six or seven.

3.2 Assay Plate Layout Design

The 354 samples analyzed were divided into ten plates, and depending on the kit, 38 or 39 samples could be assayed in duplicate per plate. To minimize plate affects across assay plates, each plate contained similar number of subjects with respect to four test sample types defined from the combination of two conditions and two time point samples (*see* Table 1). The order of the samples and plates was kept the same for every kit. For the kits with 38 samples per plate (k003, k004, and k006), the samples for X39 were measured on the last tenth plate.

3.2.1 The Assay Plate

In multiple plate experiments, it is important to understand the plate-sample-layout pattern and how such designs affect the analysis, its relationship to the stored immunoassay result files, and to the metadata used to identify and characterize the test samples on each plate. Ideally in a multiple plate experiment, each plate should contain the same pattern and ratio of test sample types. That is, the distribution of fluorescence responses on each plate should be statistically identical. Although the data analyst or bioinformatician doesn't often get to determine the assignment of samples to plate and hence the plate layout, it is important to understand it and to look and correct for any biases in the plate assignment of test samples that can affect and confound resulting statistical analysis [8–10].

When filling in the assay plate, all samples are first placed in a sample plate, for example, a 96-well test tube plate. Generally, there is also a reference plate, which holds just the reference analyte samples (standards, controls, blanks). Schematically, here they are represented on one plate with the test samples (Fig. 2a). The samples in the sample plate are mapped and placed in an assay plate. During this mapping the association between the sample information and the sample assay wells need to be managed. The test sample information is entered via Bio-Plex Manager's Enter Sample Info dialog, which then is used to assign a description to each of the unknown wells that were defined during Bio-Plex Manager's plate formatting section.

The accepted mapping of samples to assays according to protocol replicates/duplicates every column in the sample plate to adjacent columns in the assay plate (Fig. 2b). This is achieved most efficiently using eight-channel pipetting either manually or via automation. However, column effects and data artifacts can get generated by this process [11] and as seen in Fig. 3. Figure 3 gives boxplot distributions of the pooled fluorescence responses for all analytes from each test sample within each column. Note this data is from another experiment and is not discussed here, but simply used to demonstrate a column effect brought about by eight-channel pipetting. Note from Fig. 3 that the even-numbered columns represent the technically replicated assays for the test assays

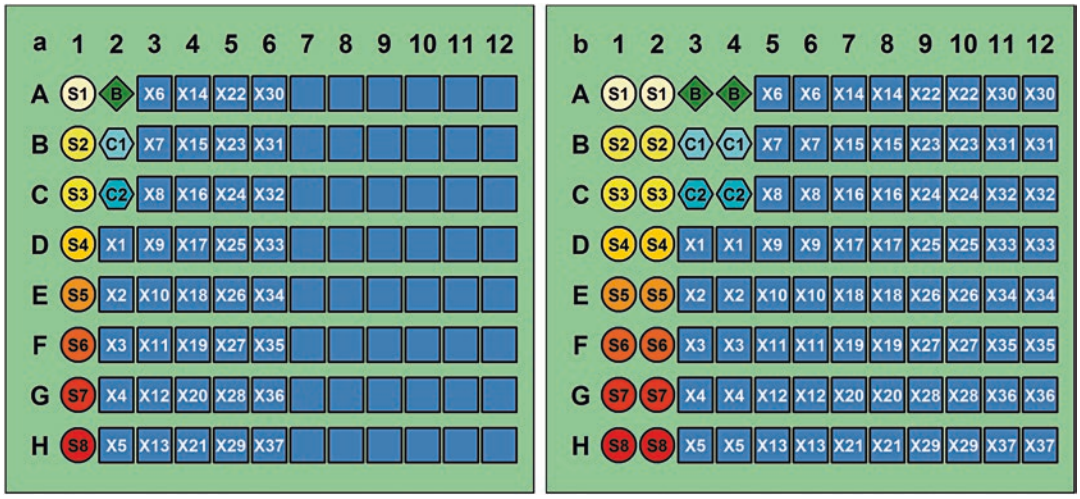


Fig. 2 Sample and assay plate layouts for replicate analysis. (**a**) The recommended sample plate layout for 8 standards (S), 1 blank (B), 2 controls, and 37 tests samples (X). (**b**) Assay plate layout constructed from the sample plate layout by replicating each sample column. The reference samples are represented by *circle* = standard (S), *diamond* = blank (B), and *hexagon* = control (C), while the test subject samples (X) are represented in *squares*. The number within each shape identifies the same sample for that type. Note that only 37 test samples can be analyzed per plate. The controls are used for measuring plate-to-plate variations and are not always included as the standards can also be used for measuring inter-plate variability if needed [7]. Standards are used to quantify analyte concentrations, and each plate includes wells containing a dilution series (S1 to Sn, where n is generally 6, 7, or 8 but can be larger) of known analyte concentrations

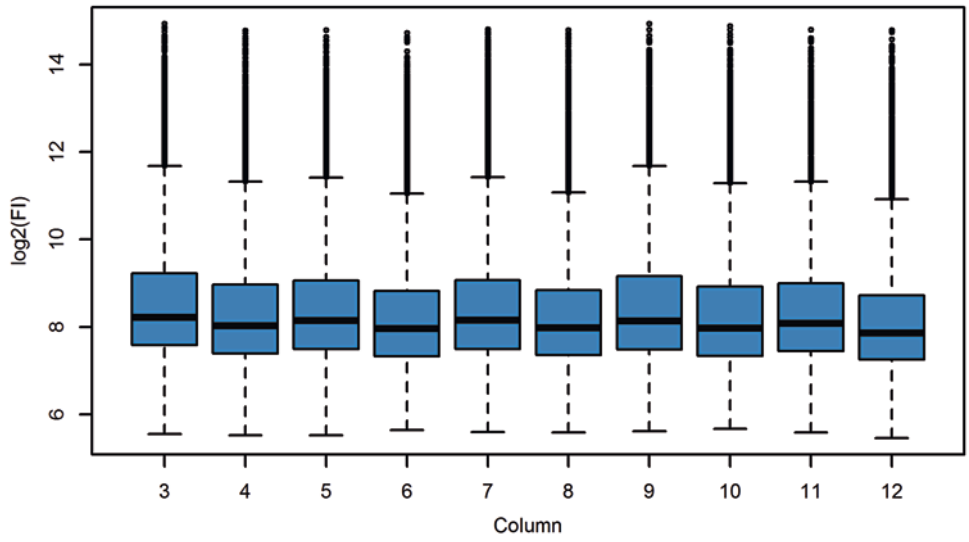


Fig. 3 Column effect seen in replicate columns. Results from the pooling of 21, 21-plex, immunoassay plates representing readings from 798 patients of effectively the same test sample type (a cohort of similar people). Note that each even-numbered column contains the same assays as the immediately previous odd-numbered column. Also note that each even-numbered column has a slightly but noticeable lower median fluorescence response than the previous odd-numbered column

in the previous odd-numbered column, Fig. 2b. From Fig. 3, it is easily noticed that the even-numbered columns are slightly lower in fluorescence intensity, on average, than the associated odd-numbered column. This is reflecting a systematic column effect, and while we don't understand why this is the case, we suspect it's due to a pipetting effect, caused maybe by some type of carryover of beads from the odd-numbered columns to the even-numbered columns [3]. This is based on the fact that the same set of tips were used to deliver each sample column and its replicate assay column, but a new set of tips was used with respect to each sample column. The advice for pipetting for immunoassay analysis is therefore to use a different tip for each sample delivered to the assay plate.

3.2.2 Determine the Number of Plates Required

Each plate of assays, Fig. 2b, can cost upward of several thousands of dollars; therefore, there is a need to maximize the statistical power of the experiment while reducing the number of plates used. This can be done by considering the number of biological and technical replicates used for any particular experiment.

Biological replicates are parallel measurements taken from different individuals or distinct biological samples. Technical replicates are repeated measurements on the same biological sample. The total variation in assay measurement is the sum of variances of each effect, biological and technical, weighted by the number of times each effect has been independently sampled [12]. Let the variance in the measurement of a subject's analyte response due to biological variance be vb and technical variance be vt. Then, the variance in response from a single assay is $vb + vt$, that is, the sum of the individual variances. The variance of each source is divided by the number of times that source is independently sampled. If biological replicates are assayed in replicate, then the single assay variance is $vb + vt/2$, and for n biological replicates analyzed in replicate, the assay variance becomes $vb/n + vt/2n$. Now by dropping the technical replicates and increasing the numbers of distinct biological replicates for analysis twofold (2X), the experimental assay variation becomes $vb/2n + vt/2n$. The effects on the observed assay variance depending on various immunoassay experimental designs are shown in Fig. 4a.

As seen in Fig. 4a, a singlet experimental design, where all assays come from unique biological replicates, the expected observed assay variance is not that different from the variances expected from a fully replicated experiment (each biological replicate is assayed twice) when $vb = 10$ and $vt = 3$; that is, the technical variance is less than or equal to 30% of the total variances. The difference between the variances between a singlet and replicate design gets smaller as n increases, after $n = 20$ the disparity is less than 1% that of the difference seen when $n = 1$, and after $n = 10$ it is less than 5%. The main difference, of course, between the curves red and blue lines (Fig. 4a) is that the replicated design requires

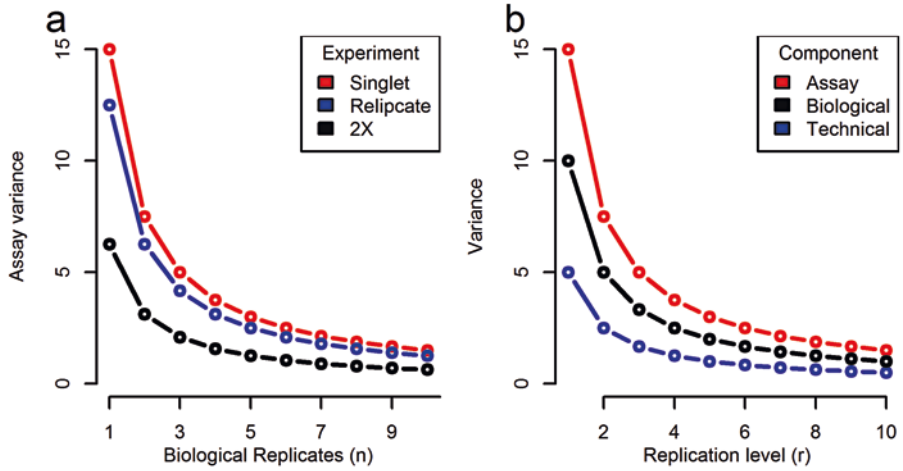


Fig. 4 Assay variance with respect to: (**a**) biological replicate numbers, *n*, and (**b**) level of replication. Plots are conditional on that $vb = 10$ and $vt = 3$. Note for the replicate experiment, when the number of biological replicates is *n*, there are $2n$ assays produced, likewise for the 2X experiment. However, for the 2X experiment, there are $2n$ distinct biological replicates used

twice the number of assays as used in a singlet experiment design. Alternatively, replacing the technical assays with assays for unique biological replicates creates a 2X design which results in the lowest expected assay variances per *n* assays produced as seen for the black line in Fig. 4a. Naturally, the 2X design is just a special case of a singlet design, but is worth considering because it makes it very clear that using *n* assays with *n* biological replicates gives almost a 50% reduction in assay variance than using a replicated experimental design, whereby half the number of assays are used for technical replicates and the other half for biological replicates. Figure 4b gives a decomposition of the assay variance according to the contributions from the biological and technical replicates with respect to replication levels.

In the example code given in the supplementary file **getBioPlexInput.R**, the function **bplxPlateNos** can be used to help determine the number of plates and number of samples to assign to each plate—given the number of test samples and the maximum number of test samples allowed per plate. For example, consider the problem of how many plates to use given 302 test samples to analyze, when the maximum number of samples per plate is 74 (no replication):

```
> bplxPlateNos(302, max.spp = 74)
      N No.plates    plate1    plate2    plate3    plate4    plate5
    302         5        60        61        60        61        60
```

The above output from **bplxPlateNos** suggests that for 302 test samples with a maximum of 74 test samples per plate, five

plates are required. If you want to allow for replication, then set **max.spp** = 37. The output above also suggests that 60 test samples should get placed on the first plate, 61 on the second plate, etc. This means that on each plate there will be essentially the same number of test samples and there will be 13 or 14 assay wells per plate that can be used for technical replication if desired.

3.3 Bio-Plex® Data Files

The analysis of immunoassay data has several distinct components. First, there is the experimental data, the immunoassay responses, which get collected by the Bio-Plex xMAP technology. This data represents the fluorescent responses measured from specific molecules bound to the surfaces of fluorescent microspheres and which gets stored into Bio-Plex result files, such as an **rbx** file. These files can contain the observed concentration estimates, bead event details, and read times for each analyte.

The Bio-Plex Manager produces for each assay plate two main types of files: (a) a protocol file which contains the analytes, settings, and system parameters for reading a 96-well plate and (b) a result **rbx** file. Protocol files can contain all the readings and bead event data, and it can be used to regenerate a result file if needed. The **rbx** file is generally considered the main result file generated by the Bio-Plex Manager. However, users that want to analyze their data outside the Bio-Plex platform need to export their results to one of several file formats: Bio-Plex XML, a CSV file format compatible with xPONENT software, Microsoft Excel files, and a 96-well ASCII file format, which is used to represent plate layouts.

3.3.1 Exporting Bio-Plex Data to Excel

The Bio-Plex Manager has many export data options, but for the methods presented here, the following shows how to export an **rbx** file to get the desired Excel file used as input to the procedures discussed below:

- Load an **rbx** file into the Bio-Plex Manager.
- With an **rbx** file loaded, select the Export to Excel button from the File menu.
- After the Excel view of your data is shown, click the Report Table Display button on the Report Table toolbar. This will launch the Report Table Display Options dialog, where you set:
  - Report Scheme: Quantitative
  - Select Columns to Display: Select All
  - Layout Table by: Multiple Analyte
  - Organize Samples by: Type
  - All else leave unchecked
- Once the data is in Excel, you need to use Excel to save it as an "xlsx" type file.

3.3.2 Bio-Plex Multi-Analyte Format

Each worksheet in an exported multi-analyte format Excel file, except for the standard curve sheet, is formatted the same but containing data for only one data type, such as fluorescence (Fl), background values (blanks), background-corrected values, etc. In fact up to 35 such data types can be exported:

```
> require(openxlsx)
> sheetCount(bpfiles[1])
[1] 35
> sheetNames(bpfiles[1])
 [1] "Outlier"            "FI"                 "FI - Bkgd"          "Std Dev"
 [5] "Std Err"            "%CV"                "Norm Ratio"         "Norm Ratio Std Dev"
 [9] "Norm Ratio Std Err" "Norm Ratio %CV"     "Norm FI"            "Norm FI Std Dev"
[13] "Norm FI Std Err"    "Norm FI %CV"        "Conc in Range"      "Obs Conc"
[17] "Obs Conc Std Dev"   "Obs Conc Std Err"   "Obs Conc %CV"       "Exp Conc"
[21] "(Obs Exp) X 100"    "Group"              "Ratio"              "Dilution"
[25] "Bead Count"         "Bead Mean"          "Bead Std Dev"       "Bead Std Err"
[29] "Bead %CV"           "Trimmed Mean"       "Trimmed Std Dev"    "Trimmed Std Err"
[33] "Trimmed %CV"        "Sampling Errors"    "Standard Curve"
```

where bpfiles [1] is a string holding the name and path to such an exported Bio-Plex file. An example worksheet in the desired format is given in Fig. 5. Note, other than the analyte columns, the worksheet shows only three other columns:

- **Type column**: this column labels each sample with a plate unique identifier and is also used to order the rows in the spreadsheet. The order of the samples is in divisions according to Bio-Plex(R) sample type starting with the blanks followed by the standards, then the controls if they're included, and, finally, the test or unknown samples. The number associated with each assay identifies the same assay sample with respect to that sample type. That is, every **S1** in a given plate is a replicate of the same standard assay, and every **X1** identifies the same test sample. Therefore, technical replicates for a given plate are identified by having the same **Type** field (*see* Fig. 2).
- **Well column**: this column contains the well locations of each sample. For replicate groups, the mean fluorescence values for the group analytes will be calculated and given in the result file. While all well assays in a replicate group have the same **Type** label, the replicate group **Well** field contains a comma-separated list of the well location used by the samples in the replicate group.
- **Description column:** this column is generally empty for reference samples. For test samples, it can also be empty, it can be a simple unique number or a sample id/bar-code, or it can contain the entire experiment's metadata associated with each test sample encoded into a single string for each sample.

	A	B	C	D	E	F	G	H	I	J
1	File Name:									
2	Acquisition Date: 14-Jun-2016, 05:45 PM									
3	Reader Serial Number: LX10011215402									
4	Plate ID: Plate 2 2nd reading									
5	RP1 PMT (Volts): 618.56									
6	RP1 Target: 16643									
7										
8				Hu IL-1b (39)	Hu IL-1ra (25)	Hu IL-2 (38)	Hu IL-4 (52)	Hu IL-5 (33)	Hu IL-6 (19)	Hu IL-7 (
9	Type	Well	Description	FI	FI	FI	FI	FI	FI	FI
10	B	H11,H12		58.3	79	884.3	74.8	50.5	139.8	
11	S1	A1,A2		---	---	---	---	---	32121.3	---
12	S2	B1,B2		---	---	30377.5	---	---	31659.5	315
13	S3	C1,C2		29283.8	30130.3	24993	29992.3	30595.8	29393.3	255
14	S4	D1,D2		12934.3	18785.8	8840.5	13758.5	12365.5	15815	59
15	S5	E1,E2		3115.8	4174.5	2948.8	2817.3	1598.3	4504	12
16	S6	F1,F2		733.5	865.5	1399	512	256.8	1208	32
57	X38	F11,F12	5077	126.5	158	166.5	230.8	72.3	178	13
58										
59										
60				Hu IL-1b (39)	Hu IL-1ra (25)	Hu IL-2 (38)	Hu IL-4 (52)	Hu IL-5 (33)	Hu IL-6 (19)	Hu IL-7 (
61	Type	Well	Description	FI	FI	FI	FI	FI	FI	FI
62	B	H11		63	81	915	78.5	50	144	
63	B	H12		53.5	77	853.5	71	51	135.5	
64	S1	A1		32160	31113.5	31334	31988.5	32107	32101.5	3189
65	S1	A2		32400	31677	31309	32189	32482	32141	32
66	S2	B1		31627.5	31155	30230	31669	32086	31454	31
67	S2	B2		31922	31377	30525	31802.5	32214.5	31865	31
159	Sampling Errors: 1 - Low bead #; 2 - Agg beads; 3 - Classify %; 4 - Region selection; 5 - Platform temperature									
160	*** = Value not available; --- = Designated as an outlier									
161	*Value = Value extrapolated beyond standard range									
162	OOR = Out of Range; OOR> = Out of Range Above; OOR< = Out of Range Below									
163	Exp Conc = Expected Concentration; Obs Conc = Observed Concentration									
164	Concentration Units = pg/ml									
165	Ratio = Member/Reference									

Fig. 5 Bio-Plex Manager Multi-Analyte Layout of the fluorescence (FI) sheet. For brevity rows 17–56 and rows 68–158 have been removed. Each sheet in an Excel file is laid out the same, except for the Standard Curve sheet. Each sheet generally has four section 1, 6 or more lines of header information, rows 1–6 above. This is then usually followed by 2, a data listing for each analyte from the replicate groups which represents the averaged data from the assays within each replicate group (rows 8–57). The replicated wells for each replicate group are shown together in the "Well" column. This section is followed by the raw data section 3, where the reading for each of the individual assay/well is given (rows 60–157). This "raw" data section is then followed by 4, a footer section, rows 159–165. Note that the footer and header section can be optional and can be excluded during the export process; also some users only keep the replicate group data and don't keep or don't export the raw individual readings

3.3.3 The Description Field

The description field is also, at times, used to identify test samples defined by the user as blanks and control samples. These can be in addition or as a replacement to the Bio-Plex-specified controls and blank samples in the immunoassay kit. Mostly, however, the description field generally just contains user-defined sample information or sample identification codes.

The description field entries, if used, should be kept as simple as possible. Extra information pertaining to each test sample should be contained in an associated metadata file. The link between the metadata file and the Excel file can be via the **Description** field or

the **Type** field with or without file identification. The entries in the **Description** fields or **Type** fields in both the Excel file and the metadata file are expected to be identical, although their column heading names may differ.

3.3.4 Labeling Biological and Replicate Samples

When using the description field to identify test sample types, the sample assays associated with replicates from the same biological sample should be identified using the same test sample identifier code, for example, **abc**, regardless of the sample's immunoassay plate (file) or well address. It doesn't really matter what the code is as long as it is unique and matches the description in the metadata file and consistent for each sample across the entire experiment. However, repeat samples taken from the same subject should be identified using their own related but unique identifier coding system such as **abc**.1, **abc**.2, etc.

3.4 Reading Bio-Plex® Excel Files

The supplementary file **getBioPlexInput.R** contains several useful functions for working with Bio-Plex result data. The R code in this supplementary is given free, but should not to be considered production quality code, and it may exit ungraciously leaving you with some cryptic error message, but used as directed here provides a useful platform for collecting and analyzing Bio-Plex result data.

The function **bplxGetData** is used to read Bio-Plex®-exported Excel files as outlined above. The arguments passed to **bplxGetData** are:

- **Files:** a filename or character vector of one or more Bio-Plex result Excel "xlsx" files.
- **sheetName:** a sheet name or character vector of one or more sheet names to read; by default it is set to "FI."
- **Exclude:** columns in the Excel file to ignore. It is a character vector of one or more columns. By default it is set to **NA** (not available). This parameter is at times needed because many exported Excel files actually get manually edited and have added columns inserted or not all the analytes or the **Description** field may be of interest.
- **FUN:** a function or function name used to generate a plate identifier for each plate read. During the reading of a batch files or even single file, this function will be called after reading each file, and will be passed the filename, and a count value in that order. The count value is the order number of each file in the file list. By default the **bplxGetData** uses an internal function **makePlateId(filename, cnt)**, in **getBioPlexInput.R**, and it will assign **p001** to first file read, **p002** to second, and so on up to **p999**.

The return value from **bplxGetData** is an R list object where each component in the list is an R data frame associated with reading the input Bio-Plex result file(s). The components returned are:

- **Header**: it contains read information collected from each file read.
- **bp:** the Bio-Plex result data, plate, and replicate information. It also contains a sample column to distinguish the sample types as either **test** (**X**), **standard** (**S1,…,Sn**), control (**C1,…,Cn)**, and blank **(B)**. It is similar to **Type**, but the test samples are identified with a single label (X). It contains a group factor (**grp**), similar to **sample**, to distinguish samples based on their Bio-Plex sample groups. It also contains a **Rep** column, used to identify the plate replicates as R1, R2, …, Rn. It contains the contents of the **Well** column split into its row (**row**) and column (**col**) parts, and it contains a **Type2** column, which groups the responses as either **raw** or **mfi**. M**fi** is used to identify the replicate groups, while **raw** is used to identify the individual well responses, that is, the individual assay responses. There is also a **plex** column to specify the number of analytes associated with each analyte **kit** identified.
- **Readings:** it contains a table for the number of missing, **NA**, fluorescence response value for each of Bio-Plex sample types (standards, controls, blanks, and test samples) read. Response values not available are automatically excluded after reading all the files.
- **Kits:** it tables the different sets of analytes detected per plate ID.

The next R session demonstrates the use of **bplxGetData**. The first input line, where input lines start with a ">" symbol and represents user-entered R commands, is used to source the **getBioPlexInput.R** file. This will make available the various functions in this file. There is also a number of R packages such as **openxlsx** [13], **phia** [14], **lme4** [15], **nnet** [16], **gdata** [17], **lattice** [18], and **reshape** [19] required for this demonstration and which can be all be installed by calling **bplxInstalls()**. If you already have all or some of these packages installed, in your own R distribution, calling **bplxInstalls** will not reinstall them, but it will cause them to be loaded, if they are not already loaded, into your R session. Calling **bplxInstalls** a second time in the same R session has no effect. Therefore, it is an idempotent function.

```
> source('../R functions/getBioPlexInput.R')
> bplxInstalls()
> rmd = bplxGetData(rmd_files, exclude='Description',
                    sheetName = c('FI', 'Dilution'),
                    makeIDfromBaseName)
> names(rmd)
[1] "Header"   "bp"       "Readings" "kits"

> head(rmd$Header)
        Reader  Read.Date Read.Time    Plate  Volts Target          file  pid  kit plex
1 LX10004298302 2012-07-09  04:54 PM  plate 1 581.20   3784  Plate 1.XLSX p001 k001    3
2 LX10004298302 2012-07-17  12:20 PM Plate 10 584.62   3784 Plate 10.XLSX p010 k001    3
3 LX10004298302 2012-07-18  02:28 PM Plate 11 585.35   3784 Plate 11.XLSX p011 k002    2
4 LX10004298302 2012-07-18  04:11 PM plate 12 585.35   3784 Plate 12.XLSX p012 k002    2
5 LX10004298302 2012-07-19  01:03 PM plate 13 584.86   3784 Plate 13.XLSX p013 k002    2
> some(rmd$bp)
      pid sample grp row   Rep col Type2 Type  Well plex    Cytokine      FI Dilution  kit
937   p012      X   X C,C R1,R2 5,6   mfi  X10 C5,C6    2       IL-10    34.3        1 k002
3823  p020      X   X A,A R1,R2 7,8   mfi  X16 A7,A8    2       MCP-1  7473.0        1 k002
4736  p022      X   X   B    R2   4   raw   X3    B4    5        IL-6    21.0        6 k003
4905  p022      X   X   B    R1   7   raw  X16    B7    5        IL-8  1091.0        6 k003
7891  p026     S6   S   E    R1   2   raw   S6    E2    5      Leptin 12485.0        1 k003
12535 p034      X   X   C    R2   8   raw  X17    C8    2 Adiponectin  3567.0      400 k004
12868 p035     S2   S E,F R1,R2 1,1   mfi   S2 E1,F1    2    Resistin    73.5        1 k004
12926 p035     S6   S   E    R1   2   raw   S6    E2    2    Resistin 22976.0        1 k004
15111 p042      X   X   C    R2   6   raw  X10    C6    1       PAI-1   470.0      100 k005
18557 p060      X   X   D    R2   4   raw   X4    D4    1         BNP    57.5        1 k006
> rmd$Readings
             C1  C2  S1  S2  S3  S4  S5  S6  S7   B     X
Readings    420 420 420 420 420 420 420 420 240 415 15906
Missing       0   0   0   0   0   0   0   0   0   0     1
For.Analysis 420 420 420 420 420 420 420 420 240 415 15905
> rmd$kits
  kit plex                  Analytes standards blanks controls
1 k001    3                CRP,SAA,SAP         6      1        2
2 k002    2                IL-10,MCP-1         6      1        2
3 k003    5 IL-6,IL-8,Leptin,TNF-a,VEGF         7      1        2
4 k004    2        Adiponectin,Resistin         7      1        2
5 k005    1                     PAI-1         6      1        2
6 k006    1                       BNP         7      1        2
```

After sourcing **getBioPlexInput.R** and calling **bplxInstalls**, the function call to **bplxGetData** directs **bplxGetData** to exclude the **Description** column. By default **bplxGetData** reads the **FI** sheets only; here we read both **FI** and **Dilution** sheets; however, if you want this plus the concentration values and the expected concentration responses, then set **sheetName** as **sheetName = c("FI," "Obs Conc," "Exp Conc," "Dilution")**. Where the **c()** command is used in R to concatenate comma-separated lists of elements into a single vector. The **bplxGetData** function call in this example is also passed the name of a user-defined function to make the plate IDs, **makeIDfromBaseName**, which is:

```
> makeIDfromBaseName
function(filename, ...)
{
  sprintf("p%03d", as.numeric(gsub("[^0-9]", '', basename(filename))));
}
```

The **makeIDfromBaseName** function simply extracts the numbers found in the **basename** of each file and appends it to "p" right justified in a field of size 3, padded with zeros. This only makes sense, as in this case, if the numbers in the filenames can be used to uniquely identify each file. For example:

```
> basename(rmd_files)
 [1] "Plate 1.XLSX"  "Plate 10.XLSX" "Plate 11.XLSX" "Plate 12.XLSX" "Plate 13.XLSX"
 [6] "Plate 14.XLSX" "Plate 15.XLSX" "Plate 16.XLSX" "Plate 17.XLSX" "Plate 18.XLSX"
…
[51] "Plate 55.XLSX" "Plate 56.XLSX" "Plate 57.XLSX" "Plate 58.XLSX" "Plate 59.XLSX"
[56] "Plate 6.XLSX"  "Plate 60.XLSX" "Plate 7.XLSX"  "Plate 8.XLSX"  "Plate 9.XLSX"
```

where **basename**(**x**) is an R function that removes all of a file's path up to and including the last path separator in **x**. Therefore, functions like **makeIDfromBaseName** are experiment specific, and this is the reason why **bplxGetData** allows, if needed, for the user to specify their own function for this purpose.

After the files have been read, the names of the components within the returned list object from **bplxGetData**, stored in **rmd**, are displayed using R's **names(rmd)** function call. This is used to display the names of the components held in **rmd**. After that, the few lines of each of the returned components in **rmd** are given using R's **head(x)** function. Note, there is also the corresponding **tail(x)** function that prints out the last few rows in **x**, and if **bplxInstalls** has been called, then there is also **some**(**x**), which can be used to display a random selection of rows.

3.4.1 Merging Metadata with Assay Data

For brevity, a copy of **bp** from **rmd** that only includes the test sample **raw** fluorescence responses is constructed. Then this copy is merged with the test sample metadata as shown in Table 1. The first line of code below uses R's **droplevels** function to remove empty factor levels that might be generated from the sub-setting of a data frame by rows. The second line reads in the experimental test sample metadata using **read.xlsx** from the R package **openxlsx.** Generally, you need to treat the column data as factors, so the first thing to be done with the return data frame from **read.xlsx** is to convert any character column types to factors using the **convertChar2Factor** routine, from **getBioPlexInput.R**. Next, **merge** is used to merge the two data frames, **md** and **bp**, and store the result back into **bp**. R's **merge** command does a database style join between **bp** and **md** by the columns identified in the **by** argument **c**("**Type**," "**pid**"). These columns must be in both data frames. The order of the two data frames passed to **merge** is identified by **merge** as **x** and **y**, respectively, so the argument **all.x** directs **merge** that extra rows will be added to the output, one for each row in **bp** that has no matching row in **md**. This is set as such to insure that the responses due to standards and blanks and controls are retained.

```
> bp = droplevels(rmd$bp[rmd$bp$Type2 == 'raw' & rmd$bp$sample == 'X',])
> md = read.xlsx('../rmdData/rmdMetaData.xlsx')
> md = convertChar2Factor(md)
> bp = merge(bp, md, by=c("Type", 'pid'), all.x=T, sort=F)
> bp = droplevels(bp[ !is.na(bp$patient),])
> some(bp, n=3)
     Type  pid sample grp row Rep col Type2 Well plex Cytokine    FI Dilution  kit
6133  X30 p033      X   X   H  R1   9   raw   H9    2 Resistin 600.0      400 k004
7992  X10 p049      X   X   C  R1   5   raw   C5    1    PAI-1 380.0      100 k005
8789  X19 p057      X   X   E  R1   7   raw   E7    1      BNP  19.5        1 k006
     Hospital patient           date T.pt  Cond
6133       H1      JH          40364   BL   T2D
7992       H3     AC1 001-059_10/9/08   BL   T2D
8789       H1      DS          40588   EP cntrl
```

After the merge any extra rows added to **bp** that doesn't have matching information in **md** will have **NAs** (not available) in those columns that are usually filled with values from **md**. This can occur, not only due to reference samples but, when the plates have more assays than identified in the metadata file. Therefore, as shown above, it's a good idea to filter out these extra rows using the R function **is.na()** which is a missing value indicator. After merging and the removal of any rows with empty patient fields, three random rows from **bp** are displayed, from which it is seen that **bp** has had the metadata defined by Table 1 assigned to each response.

3.4.2 Merging Assay Data with Metadata

In the example data given here, the plate IDs are also given in the metadata file. This isn't always the case, and for such instances you will need to merge the plate ID information found in **bp** or any descriptive statistics derived from **bp** with the metadata, **md**, even if it's just for display purposes.

This is done by essentially knowing how to reverse the roles of **bp** and **md** shown with the previous **merge** command. For example, to merge the **kit** information from **bp** into **md**, the R session below uses R's **unique(x)** function, which returns a vector, data frame, or array-like **x** but with duplicate elements/rows removed, thereby creating a data frame which has matched **Type** and **pid** fields to those in the metadata file, and as each **c("Type," "pid")** is only associated with one **kit** type, no duplicated rows from **md** are created:

```
> md = merge(md, unique(bp[, c('Type', 'pid', 'kit')]),
             by=c('Type', 'pid'))
> str(md)
'data.frame':      2124 obs. of  8 variables:
 $ Type    : Factor w/ 39 levels "X1","X10","X11",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ pid     : Factor w/ 60 levels "p001","p002",..: 1 2 3 4 5 6 7 8 9 10 ...
 $ Hospital: Factor w/ 3 levels "H1","H2","H3": 2 1 1 1 1 2 2 1 1 3 ...
 $ patient : Factor w/ 177 levels "AB","AC","AC1",..: 72 58 103 23 142 172 28 ...
 $ date    : Factor w/ 293 levels "001-002_14/9/07",..: 141 161 186 242 260  ...
 $ T.pt    : Factor w/ 2 levels "BL","EP": 1 1 1 1 1 1 1 1 1 1 ...
 $ Cond    : Factor w/ 2 levels "cntrl","T2D": 1 1 1 1 1 1 1 1 1 1 ...
 $ kit     : Factor w/ 6 levels "k001","k002",..: 1 1 1 1 1 1 1 1 1 1 ...
```

3.4.3 Identifying Replicates: Multi-Assay Plate IDs

Unique to multi-assay experiments is that not all analytes are found on the same plate. However, there is generally the same number of plates per assay type and the same patients/subjects per assay. For example, here there are ten assay plates per kit each containing assays for the same set of patients and each kit containing different analytes. In a multiple assay experiment, at times you will need to be able to specify a conceptual plate identifier so that each analyte can be conceptually treated as if they came from the same plate**:**

```
> a = unique(md[, c('pid', 'kit')])
> a$plate = bplxReplicates(a$kit)[,1]
> some(a)
    pid  kit plate
9  p009 k001     R9
36 p036 k004     R6
39 p039 k004     R9
50 p050 k005    R10
56 p056 k006     R6
60 p060 k006    R10
> bp = merge(bp, a, by=c('pid', "kit"))
```

The R session above first extracts the **unique** combinations of **pid** and **kit** from the metadata **md** and assigns the result to **a**. It then uses **bplxReplicates** from **getBioPlexInput.R**, to score the replicate kits for each assay. For example, if the kit types for five plates were defined as ("a," "b," "a," "a," "b"), then the replicate set produced would be (R1, R1, R2, R3, R2). The **bplxReplicates** function is used for identifying replicates that not only occur within a plate but also across plates, for example, identifying the replicates in the **Description** field. The **bplxReplicates** functions return a two-column data frame, where the first column, **[,1]**, is the replicate identification codes. The second column corresponds to the input vector. The replicate information for **kit** above is then assigned to **bp** using **merge** to merge the **plate** factor into **bp**.

3.5 Quality Analysis of the Data

The function **bplxFactorNos** in the next example is used to count how often a given factor level uniquely appears with other factor levels in a data frame. The **Freq** column given in the first output below shows the number of patients associated with each hospital. The second output gives the number of patients associated with **Hospital**, **Cond**, and **T.pt**. Note that these numbers are as expected and do concur with the patient numbers shown in Fig. 1 and concur with the results from **md** shown in the third output from **bplxFactorNos**:

```
> bplxFactorNos(bp, c('patient', 'Hospital'))
  Hospital Freq
1       H1   92
2       H2   50
3       H3   35
> bplxFactorNos(bp, c('patient', 'Hospital', 'Cond', 'T.pt'))
   Hospital  Cond T.pt Freq
1        H1 cntrl   BL   48
2        H2 cntrl   BL   26
3        H3 cntrl   BL   19
4        H1   T2D   BL   44
5        H2   T2D   BL   24
6        H3   T2D   BL   16
7        H1 cntrl   EP   48
8        H2 cntrl   EP   26
9        H3 cntrl   EP   19
10       H1   T2D   EP   44
11       H2   T2D   EP   24
12       H3   T2D   EP   16
> bplxFactorNos(md, c('patient', 'Hospital'))
  Hospital Freq
1       H1   92
2       H2   50
3       H3   35
```

3.5.1 Check for Hierarchical Relationships

Hierarchical structures such as nesting for statistical models in R are arranged from the broadest to narrowest (or higher to lower). For example, to specify that toes are nested in feet, you specify "foot/toe," where the forward "/" slash is used to specify nesting. In a nested design, each level of the nested factor is uniquely associated with only one level of the higher-level factor. Otherwise the factors are considered crossed; that is, the levels of both factors are associated with more than one level of the other factor. For example, to check if **pid** is nested in **kit,** and **patient** is nested in **Hospital**, use R's **xtabs** function:

```
> xtabs(~kit+pid, bp)[,sample(1:60, 10)]
     pid
kit   p008 p040 p026 p010 p047 p027 p030 p035 p014 p029
  k001  222    0    0   54    0    0    0    0    0    0
  k002    0    0    0    0    0    0    0    0  156    0
  k003    0    0  380    0    0  370  180    0    0  360
  k004    0   72    0    0    0    0    0  148    0    0
  k005    0    0    0    0   76    0    0    0    0    0
  k006    0    0    0    0    0    0    0    0    0    0

> xtabs(~Hospital+patient,bp,sparse=T)[,sample(1:177, 20)]
3 x 20 sparse Matrix of class "dgCMatrix"
   [[ suppressing 20 column names ... ]]

H1 56 56  . 56 56 56  .  . 56  .  . 56 56  .  .  . 56 56  .  .
H2  .  . 56  .  .  .  . 56  .  . 56  .  . 56 56 56  .  . 56 56
H3  .  .  .  .  .  . 56  .  . 56  .  .  .  .  .  .  .  .  .  .
```

With matrices and data frames, the **xtabs()** function creates a cross tabulations of the data. The result is a contingency table in array format, whose dimensions are determined by the number of terms on the right side of the ~ tilde; therefore, the above results from **xtabs** are two-way contingency tables for the frequency, or count, of the levels of the categorical variables involved. The above outputs make it clear that **pid** is explicitly nested in **kit** and that **patient** is nested in **Hospital**, since each **pid** and **patient** only appears once in each column of their respective tables. Note the indexing **[,sample(1:60,10)**] used to index the columns in the return value from the first call to **xtabs** above. It specifies that ten random columns are to be selected from 60, using R's **sample(x,size)** function, which by default takes a random sample, without replacement, of the specified **size** from the elements in **x**. The second output from **xtabs** above is also showing for larger matrices that the **sparse** option can be used to suppress printing of column names and replaces sole zeros by ".", that is, produce a slightly more compact form of display.

3.5.2 Factor Level Coding

Whether you explicitly specify a factor as nested or not depends (in part) on the way the levels of the factor are coded. It is considered best practice to code the nested levels uniquely so that confusion between nested and crossed effects is less likely. For example, from the experimental data alone, it's not clear that the patient conditions, **Cond**, are actually nested in **Hospital** or are in fact crossed factors. This is because the coding for the nesting between **Cond**

and **Hospital** is *implicit*, that is, the levels in the lower factor are repeated with respect to the levels of the higher factor:

```
> xtabs(~Cond+Hospital, bp)
      Hospital
Cond      H1   H2   H3
  cntrl 2688 1456 1064
  T2D   2464 1343  896

> bp$hcond = factor(bp$Hospital:bp$Cond)
> xtabs(~hcond+Hospital,bp)
         Hospital
hcond       H1   H2   H3
  H1:cntrl 2688    0    0
  H1:T2D   2464    0    0
  H2:cntrl    0 1456    0
  H2:T2D      0 1343    0
  H3:cntrl    0    0 1064
  H3:T2D      0    0  896
```

The method in R to create explicitly nested factors from implicitly nested factors is given above. From the first output from **xtabs** above, it appears that **Cond** and **Hospital** are crossed factors. In the second input line above, the pairing of these factors is used to create the new factor **hcond**., which is then seen to be nested in Hospital. However, what is important also is the associated information such as that shown in Fig. 1 which is provided, to the analyst, so explicit nesting can be specified in statistical models as required.

Crossed factors can also be complete or partial. For example, **Hospital** and **pid** are only partially crossed:

```
> xtabs(~Hospital+pid,md)[,1:10]
        pid
Hospital p001 p002 p003 p004 p005 p006 p007 p008 p009 p010
      H1   23   21   17   28   20   18   24   24    7    2
      H2   13   14   19    7   15   17   11    4    0    0
      H3    3    4    3    4    3    4    3    9   30    7
```

The first ten **pids** are all associated with kit **k001**. From the output it is seen that hospital **H2** has no corresponding readings on plates **p009** and **p010**. This pattern is repeated in each of the ten plates for each of the five other kits. It means that **Hospital** and **pid** cannot be used to form a fixed-effect interaction in any later statistical analysis performed unless plates 9 and 10 from each kit or hospital H2 are removed from the analysis.

3.5.3 Check for Metadata Plate Balance

To check for metadata plate balance means to check if all test sample types have been randomly assigned to plate locations. For this, the metadata needs to have in each row of its data set the plate IDs (**pid**) obtained either explicitly or merged from **bp** as discussed above. Then a test for metadata balance across the plates can be performed using multinomial regression [16], followed by an ANOVA to assign *p-values* and significances to these effects, where the NULL hypothesis is that the probabilities of assignment of test sample types to plates are equal:

```
> mmod = multinom(pid~Hospital*Cond*T.pt, data=md)
> Anova(mmod)
Analysis of Deviance Table (Type II tests)

Response: pid
                  LR Chisq  Df Pr(>Chisq)
Hospital            862.67 118  < 2.2e-16 ***
Cond                 12.46  59     1.0000
T.pt                  3.97  59     1.0000
Hospital:Cond       238.74 118  3.386e-10 ***
Hospital:T.pt        79.15 118     0.9977
Cond:T.pt             2.45  59     1.0000
Hospital:Cond:T.pt   85.75 118     0.9888
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The first line, in the previous session above, uses **multinom** from the R package **nnet** [20] to model the plate ID (**pid**) as a function of **Hospital**, **Cond**, and **T.pt.**, from the metadata, **md**. The model term X * Y used implies X + Y + X:Y, where X:Y represents the interaction between X and Y. After modeling, ANOVA from R's **car** package [21] is used to test the significance of the model effects, and its output confirms that **Hospital** patient numbers are not equal on all plates (*p-value* < 2.2e−16) and that **Cond** and **T.pt.** patient numbers are in balance across the plates (*p-value* = 1). There is also a plate imbalance with respect to the interaction between **Hospital** and **Cond**. This interaction shows the ratio of patient conditions also changes against hospital and plate. Interactions exist when a change in the level of one factor has different effects on the response variable, depending on the value of the other factor [14]. This may be of no consequence if the analyte responses associated with these effects are similar, but unfortunately they're not, as shown in Fig. 6. Therefore, conclude that **Hospital** maybe associated with confounding variances in the response values.

3.5.4 Check Fluorescence Response Intensities

A simple way to look at the data is to use boxplots [22]. R's **boxplot** function like many of R's plotting routines can take a formula as an argument, such as **y ~ grp**, where **y** is a numeric vector to be split into groups according to the grouping variable **grp.** In R, for single figures, it's straightforward to produce a boxplot, for example, the first plot in Fig. 7 was produced by **boxplot(log2(FI) ~ Rep:sample,rmd$bp)**, with the exception of added color. For multiple plots, it requires setting up the display to combine multiple figures using R's **par()**, **layout()**, or via one of the many plotting functions available in the **lattice** package as seen in Table 3.

Figure 7 gives the boxplot distributions of the pooled analyte fluorescence distributions with respect to the raw sample replicates (**Rep**:**Sample**), the plate rows containing only test samples (**test rows**), and the test sample columns (**test columns**). It can be seen that each replicate for each Bio-Plex sample type is essentially identical, although the replicates for control **C1** appear slightly different. It is also noticed that the majority of the test samples responses (**X**) appear above the blank distributions. The test rows show some row-to-row variations as does the test

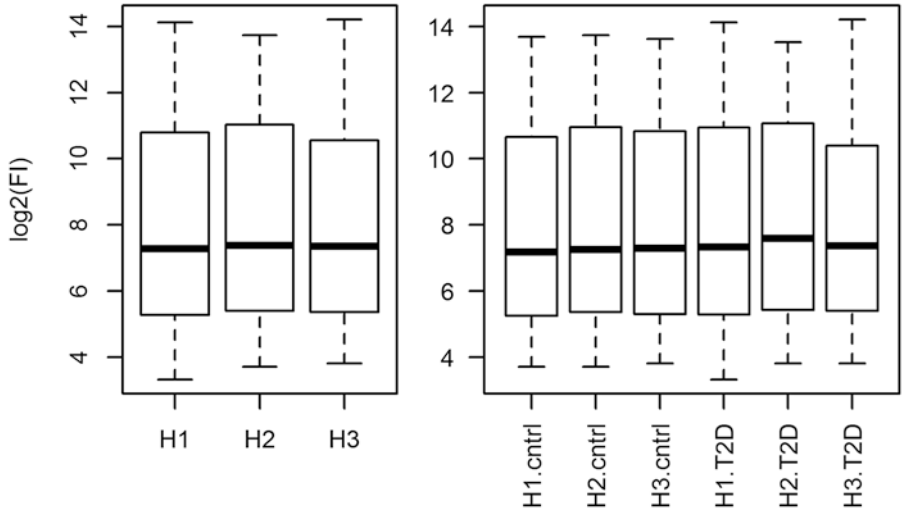


Fig. 6 Boxplots of the log2 of the fluorescence responses with respect to Hospital (H1, H2, and H3) and Cond (cntrl, T2D). The R commands used to generate these figures were: > layout(matrix(c(1,2), 1,2,byrow=T), widths=c(3,4))> par(mar=c(4,4,1,0), oma=c(0,0,0,0))> boxplot(log2(FI)~Hospital, bp, ylab='log2(FI)')> boxplot(log2(FI)~Hospital:Cond, bp,las=2)

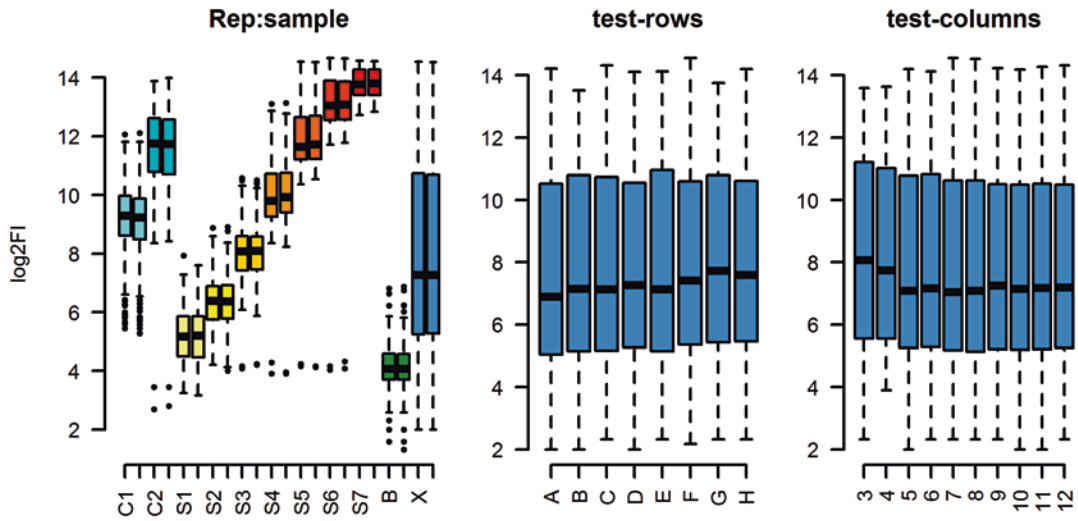


Fig. 7 Boxplot distributions for sample replicates (Rep:sample), test sample rows (test rows), and columns (test columns). Each boxplot represented the pooled analyte log2 fluorescence responses for all analytes with respect to its grouping

columns show column-to-column variation, especially with respect to columns 3 and 4.

Figure 7 (left plot) gives the dynamic range of the test samples with respect to the dynamic range of the standards, controls, and blanks. Figure 8 gives the boxplot distributions for the pooled

Table 3
Example lattice plot functions available in R

Function	Description	Formula example
histogram()	Histogram	~x
densityplot()	Kernel density plot	~x|A*B
qqmath()	Theoretical quantile plot	~x|A
stripplot()	1D scatter plots	A ~ x or x ~ A
bwplot()	Comparative box-and-whisker plots	x ~ A or A ~ x
dotplot()	Cleveland dot plot, caterpillar plot	~x|A
barchart()	Bar plot	x ~ A or A ~ x
xyplot()	Scatter plot	y ~ x|A
splom()	Scatter plot matrix	data frame
contourplot()	Contour plot of surfaces	z ~ x * y
levelplot()	False color level plot of surfaces	z ~ y * x
wireframe()	3D wireframe graph	z ~ y * x

analyte log2 responses on each plate (**p001** to **p060**). From this plot, it is noted there is a large between kit variance and a reasonable sized within kit variance:

```
> VarCorr(lmer(log2(FI)~(1|kit),bp))
 Groups   Name        Std.Dev.
 kit      (Intercept) 2.1024
 Residual             2.4557
```

Here **VarCorr** is used to estimate the standard deviations (std. dev) between the random effect terms within a **lmer** mixed-effects model. Above the between kit std.dev is seen to be 2.1, which is lower than the within kit variance as measured by the residuals with a std.dev of 2.5. The between kit variance is of no real interest, and the larger within kit variance can be accounted for because within different kits, there can be different sets and numbers of analytes.

In summary, taking the results shown in Figs. 7 and 8 together, it can be concluded that the test sample responses appear real and are not just noise because the overall test samples median response is closest to the medium response of standard S3 than to the blank.

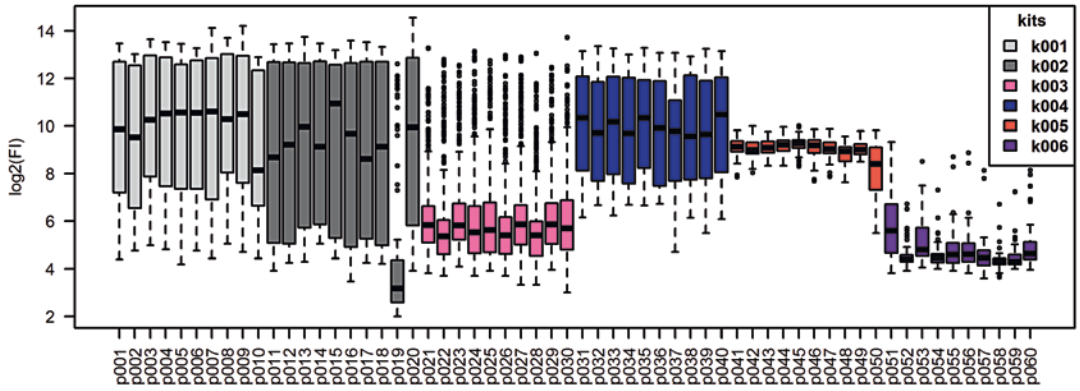


Fig. 8 Variation in plate and kit fluorescence responses. Boxplots of the log2 of the raw fluorescence responses grouped according to plate and colored according to kit type

3.5.5 *Check for Outliers*

The results shown in Fig. 8 suggest that plate **p019**, which is the plate with the lowest overall signal, could be considered for exclusion since its median log2 response is approx. 3, which is less than the blank's medium response of 3.5. For quality control it is important to identify plates, subjects, analytes, and plate effects that did not perform as expected, e.g., if the sample preparation for one reason or another failed. For a continuous variable, outliers are sometimes identified as observations that lie outside say 2X IQR, where IQR is the inter-quartile range of the variable in question. For the multivariate linear case 4X, the mean of the cook's distance [23] is sometimes used to threshold observation using the model's residuals. Although discussed later in more detail is mixed-effect analysis for multiplexed immunoassay responses, here the random effects from a mixed-effect model are used for outlier detection. Random effects are ideal for outlier detection because they provide within group and between group means and variances in terms of population estimates rather than that of the observed sample means and variances [24].

In the following code example, the **lmer** function from the **lme4** package is used to perform a mixed-effect analysis, whereby the log2 of the fluorescence responses is modeled using four independent scalar random effects terms in brackets. Each bracketed expression is of the form **(1|x)** or **(1|y/x)**, where 1 is used to indicate an intercept (mean) model matrix. In the first form **x** is nested in the intercept, in the second form **x** is nested within **y**, and this grouping is nested within the intercept as well as **y**. The return value from **lmer** is stored in **bp.mer** after which the standard deviations of the random effects are displayed:

```
> bp.mer=lmer(log2(FI)~(1|Hospital/patient) + (1|kit/pid) +
                (1|row) + (1|col), bp)
> VarCorr(bp.mer)
 Groups           Name        Std.Dev.
 patient:Hospital (Intercept) 0.102227
 pid:kit          (Intercept) 0.659254
 col              (Intercept) 0.087361
 row              (Intercept) 0.125563
 kit              (Intercept) 2.089250
 Hospital         (Intercept) 0.045222
 Residual                     2.377775

 > df=bplxMER(bp,bp.mer)

 > names(df)
[1] "remIdx" "df"     "remove"

> df$remove
     Effect     level    Zscore      p.value sig. Action
196 pid:kit p019:k002 -6.118149 9.466832e-10  *** REMOVE

> dim(bp)
[1] 9911   20

> bp = droplevels(bp[ !df$remIdx, ])
> dim(bp)
[1] 9763   20
```

Above the function, **bplxMER = function(bp, bp.mer, alpha = 0.01,cex = 1, adjust = FALSE, doOutlier = T)** from **getBioPlexInput.R** is used for selecting possible outliers. The first two arguments passed to **bplxMER** are the data, **bp**, and mixed-effect model, **bp.mer**, as created above. Note, while the random effects above in **bp.mer** are created from factors within **bp**, the **bplxMER** function expects the effects in the model handed to be simple additive scalar effects (intercept models only), which includes nested effects but not random slope models. At this stage random slope models are simply ignored. The next parameter passed is an **alpha** level parameter; by default it is 0.01, that is, representing a 1% criterion. This is used to select the levels of a random effect that have a probability of less than **alpha** as being part of that random effect as possible outliers. The test is a two-tailed test; therefore, half of the alpha is used for testing the statistical significance in one direction and half for testing statistical significance in the other direction. All levels with respect to their effect that have a two-tailed **p.value** of less than **alpha** are marked for removal. The parameter **cex** is used to scale the size of any text drawn in the output diagnostic plots (Fig. 9) produced by **bplxMER**. The **adjust** parameter, if set to true, is used to perform multiple test correction adjustment, via **fdr** [25], on the *p-values* prior to thresholding. The last parameter is **doOutlier** which, when set to TRUE, directs **bplxMER** to actually perform outlier analysis; otherwise it will produce sets of random effect diagnostic plots (see below). Above the return value from **bplxMER** is assigned to the variable **df**. The names of the components in **df** are then displayed and are explained as:

- **remIdx** is a logical vector of TRUEs and FALSEs, where the TRUE elements are those rows in **bp** that should be considered for removal, based on the given alpha level.

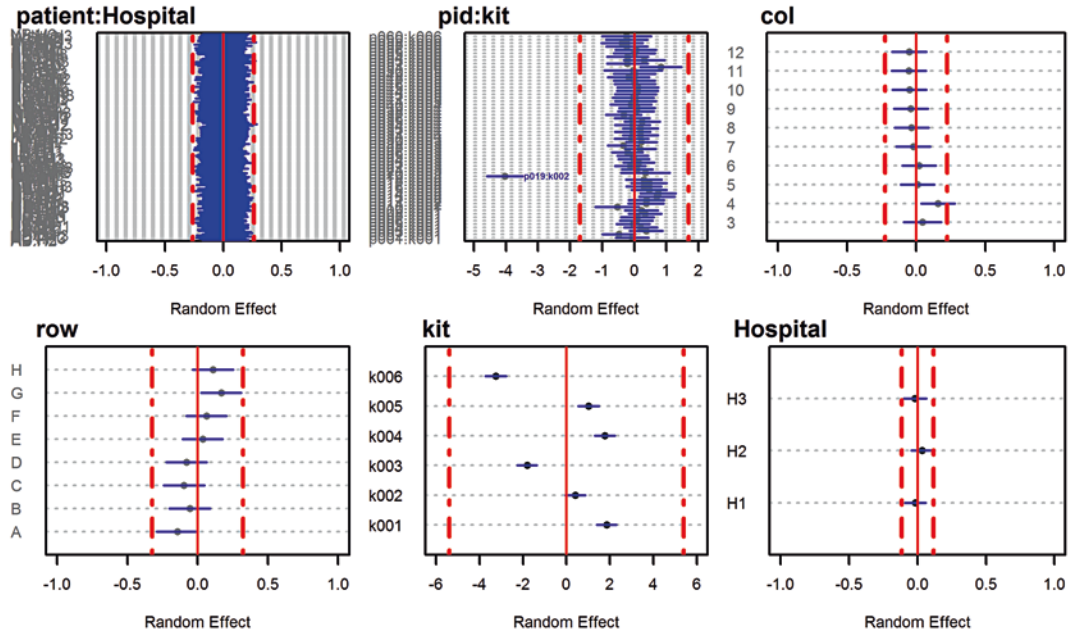


Fig. 9 Random effects test sample outlier analysis plots. The *red dashed vertical lines* give the alpha (default 99%) confidence interval for each random effect, which should be centered on zero (*red solid line*). The name above each plot is the random effect for that plot. Each row in each plot gives each level's estimated mean and 95% confidence interval for that mean. Ideally all random effects should be centered on zero and have similar variances within an effect

- **df** is a data frame, summarizing the analysis of each random effect in the model. It gives the effect, the level within the effect, and their **Zscore**, **p.value**, significance code, and **Action** information. The **Action** column informs which elements are either ok or should be removed.
- **remove**, for convenience, is the subset from **df** for those elements that should be considered for removal.

The output above from **df$remove** shows that **bplxMER** is recommending that only level **p019:k002** from effect **pid:kit** be removed. After which the responses associated with the recommendations were removed as seen in the above routine that also shows that 148 responses were removed from the **bp** data set by this step. Notice that there was a not "!" operator applied to the **remIdx** values in the sub-setting procedure, and if desired the analytes could have been just as easily included in the outlier analysis done here, but for brevity the analysis was kept to six random effects only.

To help decisions on whether to accept **bplxMER** recommendation, **bplxMER** creates a visual display of the random effects found in the outlier model and as shown in Fig. 9. The diagnostics plots show each random effect and the levels for that effect. For each effect it shows the confidence levels, red dashed line, and the mean

effect, which should be zero, red solid line. Note that if you want to run the **bplxMER** a second time, on the data subset, then you need to rebuild **bp.mer** to reflect the actual data set being considered.

3.5.6 Check Dilution Levels

With plasma or serum samples, it is known that sample dilution has a nonlinear effect on the immunoassay analyte fluorescence responses [26]. Generally the recommended dilution level is often 4X but does vary with experiment and kit [7]. To check to see if the test sample assays have all the same dilution levels or at least for the case here are held constant within each assay kit, the following procedure is used:

```
xtabs(~Dilution+kit, bp)
          kit
Dilution k001 k002 k003 k004 k005 k006
    1       0 1268    0    0    0  706
    2.5     0    0    0    0    0    2
    6       0    0 3540    0    0    0
    100     0    0    0    0  707    0
    400     0    0    0 1416    0    0
    2000 2124    0    0    0    0    0
> dim(bp)
[1] 9763   20
> bp = droplevels(bp[ bp$Dilution != 2.5, ])
> dim(bp)
[1] 9761   20
```

The output above shows that with respect to the test assays, six different dilution levels exist. However, all kits are associated with just one dilution level except for kit **k006**, which has just two readings associated with a second dilution (2.5X). Therefore these readings were removed as shown in the last line of the above R session.

3.6 Descriptive Statistics: Intra- and Inter-Assay CVs

To express the precision, or repeatability, of immunoassay results, two measures of the coefficient of variability (CV) are typically used, and these are the intra- and inter-assay CV. The CV is simply the standard deviation/mean. While the intra-assay CV is given by the Bio-Plex Manager for each analyte and sample type, it doesn't provide the inter-assay CV. Therefore, both are calculated here for the reference samples. The intra-assay CV in an experiment is calculated as the arithmetic mean of all intra-assay CV:

```
> bp2 = droplevels(rmd$bp[ rmd$bp$Type2 == 'raw' & rmd$bp$sample != 'X',])
> summaryStats = function(x) c(mean=mean(x), sd=sd(x), N = length(x),
                               sem = sd(x)/sqrt(length(x)),
                               CV = sd(x)/mean(x))
> a = aggregate(FI~Cytokine + pid + sample, data=bp2, summaryStats )
> a = cbind(a[, 1:3], a$FI)
> some(a)
       Cytokine  pid sample  mean       sd N     sem    CV
62         IL-8 p023     C1   403    17.32 2   12.25 0.043
142         SAA p001     C2   342     0.35 2    0.25 0.001
301         SAP p007     S1    60     0.71 2    0.50 0.012
447         SAP p009     S2   220    21.21 2   15.00 0.096
463       IL-10 p017     S2    96     3.54 2    2.50 0.037
693         BNP p053     S3  1100    68.24 2   48.25 0.062
817 Adiponectin p039     S4   739    26.52 2   18.75 0.036
1045       VEGF p023     S6 11894   974.39 2  689.00 0.082
1070       VEGF p028     S6  9128  1938.18 2 1370.50 0.212
1191        BNP p051     S7 18168   454.67 2  321.50 0.025
```

The CV determined from the replicates forms the intra-assay CV [27]. This needs to be calculated with respect to each Bio-Plex sample type per plate and per analyte. The first input line in the previous R session creates a second **bp** object called **bp2**, which contains only the Bio-Plex reference samples types: standards, controls, and blanks. The second input line creates summary statistic functions which takes a vector **x** as input and returns a vector containing the input's, **mean**, standard deviation (**sd**), size (**N**), standard error of the mean (**sem**), and **CV** (coefficient of variance). The third input line uses R's **aggregate** function, to determine the summary statistics from the groups defined by **~Cytokine + pid + sample**, with respect to the **FI** values. The fourth input uses **cbind** (column bind) to combine two data frames, column wise, so that the summary statistics form individual columns within **a**. Note that there is also a **rbind** function, which combines/appends data frames row wise. Next a random sample of the rows from the resulting aggregation, **a**, is displayed.

3.6.1 Displaying CV Values

The below R sessions use the **lattice** routine **barchart** to display the mean intra-assay CVs obtained from **a**, with 95% CI (confidence interval) error bars:

```
> b = aggregate(CV~Cytokine+sample, data=a, summaryStats)
> b = cbind(b[, 1:2], b$CV)
> scales=list(alternating=FALSE, cex=0.8, tck=c(0.5,0))
> barchart(mean~sample|Cytokine, data=b,
              as.table=T, scales=scales,
              ux  = b$mean + b$sem * 1.96,
              lx = b$mean,
              panel=panelfun, main="Average plate CV + 95% CI",
              ylab = "CV", xlab= "Sample Types")
```

The resulting lattice plot is given in Fig. 10a. In the above R session, aggregate is used again to collect summary statistics for the CV values stored in **a** and with respect to the grouping defined by **~Cytokine + sample**. The scale variable defines a list object used to control how the **x** and **y** axis labels and ticks are drawn, plus a text scaling variable, **cex**, that affects axis tick-label text size. The **barchart** function from the **lattice** package is then used to produce a set of conditioning plots for all analytes. Conditioning sets the view relationships across "panels" to a common scale. The vertical bar "|" is used to specify the conditioning variable, **Cytokine**. Like most R plotting routines, **lattice** functions use the formula notation of statistical models to describe the desired plot. By default each panel is displayed using a default panel function, but the call to **barchart** above uses the panel function defined as:

```
> panelfun =
function(x, y, ux, lx, subscripts, ...) {
  panel.barchart(x, y, ...);
  panel.abline(h=0.2, col='red', lwd=1.5, lty=4)
  panel.arrows( x, lx[subscripts], x, ux[subscripts], col = 'black',
                length = 0.25, unit = "native",
                angle=90 ,code = 2)
}
```

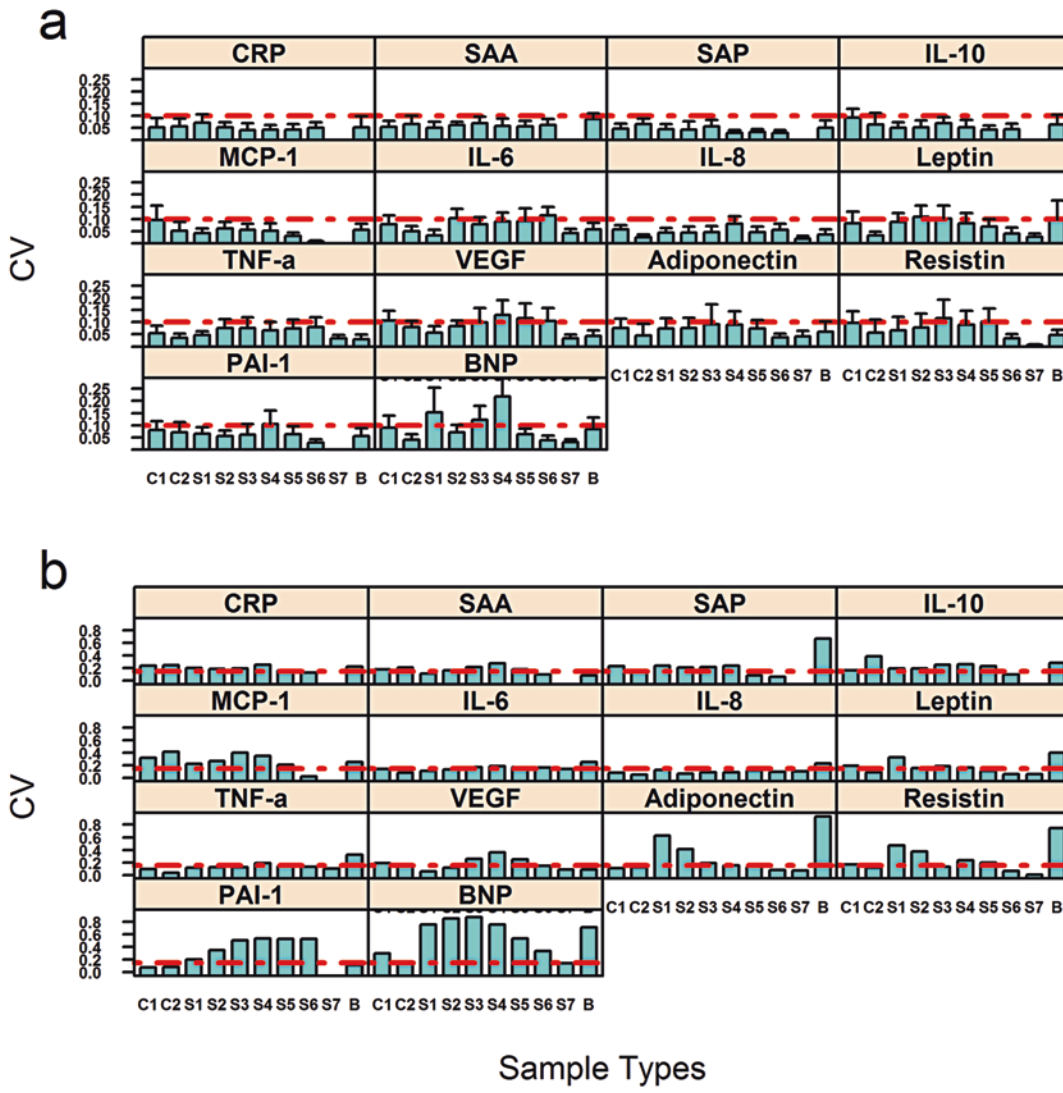


Fig. 10 Intra- and inter-assay fluorescence CV. (**a**) Intra-assay CVs: bars represent the Bio-Plex reference sample type average CV per analyte across ten plates. *Red dashed line* represents the CV value of 0.1, below which is generally considered good. Error bars represent the 95% confidence level. (**b**) Inter-assay: bars represent CV values across all plates with respect to each analyte. *Red dashed line* represents a CV of 0.15 below which is considered good

The inter-assay CV is calculated from the mean of the replicates or assay values across all plates:

```
> b2 = aggregate(mean~Cytokine+sample, data=a, summaryStats)
> b2 = cbind(b2[, 1:2], b2$mean)
```

The **b2** variable above is then plotted in a similar fashion as done for **b** previously; except the formula passed is **CV ~ sample**

and no error bars are possible. Figure 10b shows the inter-assay CVs for the Bio-Plex reference samples. Note that the results collected in **a** were used for both inter- and intra-assay analysis.

3.7 Mixed-Effect Modeling

In research, analysis of variance (ANOVA) is often used to analyze the difference between two or more means related to experimental factors of interest. To perform an ANOVA, the data must contain a continuous response variable and at least one categorical factor with two or more levels. The NULL hypothesis being tested is that the factor level means are equal. While there are many forms of ANOVA such as one-way, two-way, nested, repeat value, hierarchical, etc., all are based on an approach in which the procedure uses variances to determine whether the means are different, by comparing the variance between group means versus the variance within group values.

ANOVA can be treated as a special case of general linear regression where independent/predicator variables are factors. The ANOVA's model coefficients, or "effects," associated with predictor variables can be formed from either fixed or random factors, where fixed factors are typically treatments and factors of interest and the levels of the factor are consider fixed and measured without error, such as gender, dosage, genotype, etc. Random factors are variables considered to represent a random sample of possible values: such as patients. A statistical model that incorporates both fixed and random factors is called a mixed-effect model [15], and in R there are two main functions for performing linear mixed-effect analysis, and these are **lme** from the R package **nlme** [28] and **lmer** from the R package **lme4** [15]. However, here we are only considering **lmer**.

3.7.1 Fixed Effects

As fixed effects are the experimental effects we are interested in, what helps in determining which factors form the fixed factors is a list of the statistical objectives for the experiment. With respect to the **rmd** data, the objectives are:

1. Control group baseline compared to T2D group baseline for each analyte.
2. Control group 6 months compared to T2D group 6 months for each analyte.
3. Control group baseline compared to control group 6 months for each analyte.
4. T2D group baseline compared to T2D group 6 months for each analyte.

From the data given in **rmd**, the **Cond** factor will allow comparisons between control and T2D patients; the **T.pt.** factor allows for comparisons between end point and baseline readings. As the objectives include finding individual analyte differences, the

analytes will be used to form a fixed effect. Then the significances of the interactions between **Cond** and **T.pt.** obtained from a pairwise within group analysis of the interaction with respect to fixed **Cytokine**, **Cond**, or **T.pt.** are used to answer the objectives.

For example, with respect to objectives 1 and 2, it is relatively easy to get a preliminary feel for the results using a lattice boxplot (**bwplot**):

```
> scales[['x']] = list(rot = 45, cex = 0.5, font=2)
> nms = levels(bp$Cytokine)
> levels(bp$Cytokine) = abbreviate(nms, 5)
> stripInfo = strip.custom(bg="grey90",
                           par.strip.text=list(col="black",
                                               cex=.5, font=2))
> bwplot(log2(FI)~Cond|Cytokine+T.pt, bp, pch='|',
       par.settings = list(box.rectangle =
                             list(fill= rep(c('violetred','skyblue3'),2)),
                                  lwd = 2),
       scales = scales,
       strip = stripInfo,
       key=list(text=list(levels(bp$Cond)),
                points=list(pch=c(15,15), col=c('violetred','skyblue3'))
                ,columns=2))
> scales[['x']] = NULL
> levels(bp$Cytokine) = nms
```

Approximately 90% of the above code is just for altering the default display style of a normal lattice style plot (Table 3). This code shows how to **rotate** the *x*-axis labels by 45°, to save the analyte names into a variable, **nms**, as obtained from the **levels** of their factor. It shows how to **abbreviate** analyte names to five characters, to change the levels of a factor, to set the color of the boxplot boxes, to change the strip font and background color, and to add a legend to the plot using a **key**. Otherwise, **bwplot(log2(FI) ~ Cond|Cytokine + T.pt., bp)** would have been suffice for personal viewing. Note above that the **scale** variable is assigned a new component "**x**" (a list element) to hold *x*-axis information, but after plotting it is removed by setting it to **NULL**. In R setting components of a data frame or a list to **NULL** removes that component from the object.

The resulting plot from the above code is shown in Fig. 11. Simply, by exchanging the roles of **T.pt.** and **Cond** in the above routine, it is possible to evaluate visually the data with respect to objectives 3 and 4. From Fig. 11, it is seen that there doesn't appear to be great deal of difference between the analyte expression levels as judged by the difference in the adjacent boxplot distributions, other than that seen (maybe) for the end point (**EP**) readings between **cntrl** and **T2D** patients with respect to analyte **CRP** (top left panel).

Visual displays for confirmatory analysis when the dynamic range of the analytes is large compared to the analyte differences make it difficult to discern real but small differences; therefore, it is possible to regress the analyte responses against **T.pt.** and **Cond**, with respect to each analyte using R's **lmList** function from the **lme4** package for a more detailed analysis. **lmList** is used to fit

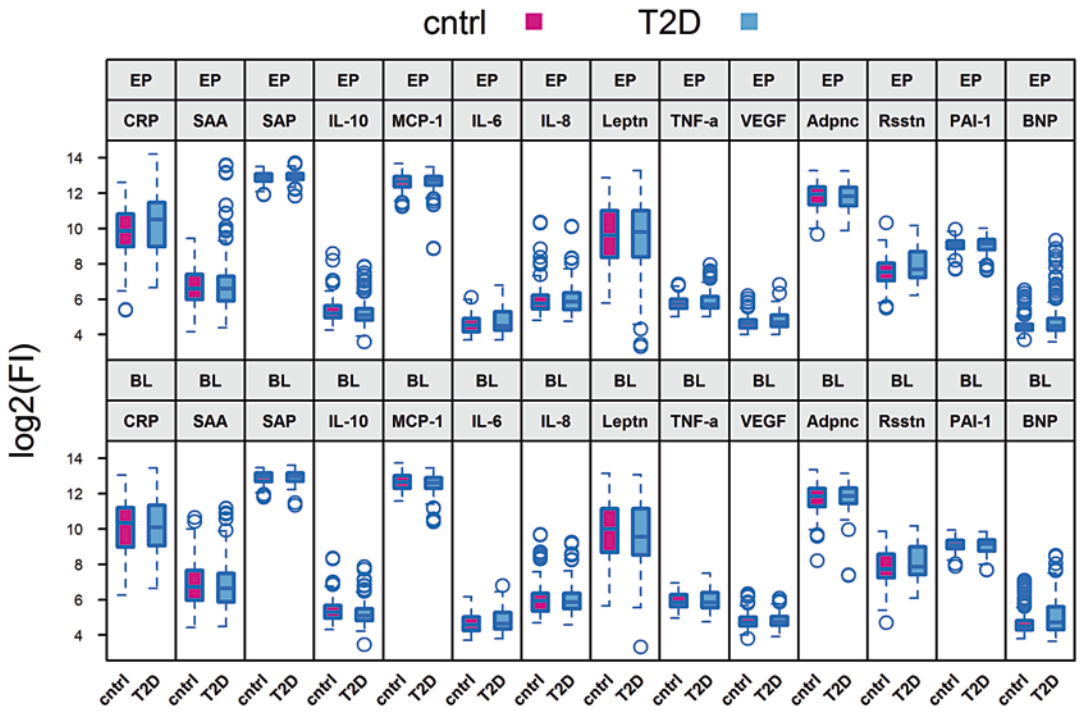


Fig. 11 Log2 of the fluorescence values versus condition (cntrl, T2D) by analyte and T.pt. (BL, EP). Abbreviations: leptn = Leptin, Adpnc = Adiponectin, and Rsstn = Resistin

linear statistical models to subgroups within data. Here I'm only interested in showing the use of **lmList** and its resulting confidence plots, before considering more advanced strategies of including random effects:

```
> stripInfo = strip.custom(bg="grey90",
+                          par.strip.text=list(col="black", cex=1, font=2) )
>
> lst = lmList(log2(FI)~T.pt + T.pt:Cond | Cytokine, bp)
> p = plot(confint(lst), strip = stripInfo)
> p + layer(panel.abline(v=0, col='red', lwd=2, lty=4))
```

The result of the above R session is given in Fig. 12, which also gives the 95% confidence intervals for each analyte difference and the intercept values. Note, this model formula (**T.pt. + T.pt.:Cond**) lacks the main effect from **Cond**. This is not a problem, as the model still uses 3° of freedom, just as a full two-way model **T.pt. + Cond + T.pt.:Cond** would, but rather than just providing a single interaction effect between **T.pt.:Cond**, this model splits the interaction into two separate effects: one for each level in **T.pt.** as seen in the top row of plots given in Fig. 12.

From Fig. 12 the top left plot gives the analyte mean differences with 95% confidence between patient conditions at baseline. The top right plot gives the analyte difference between patient

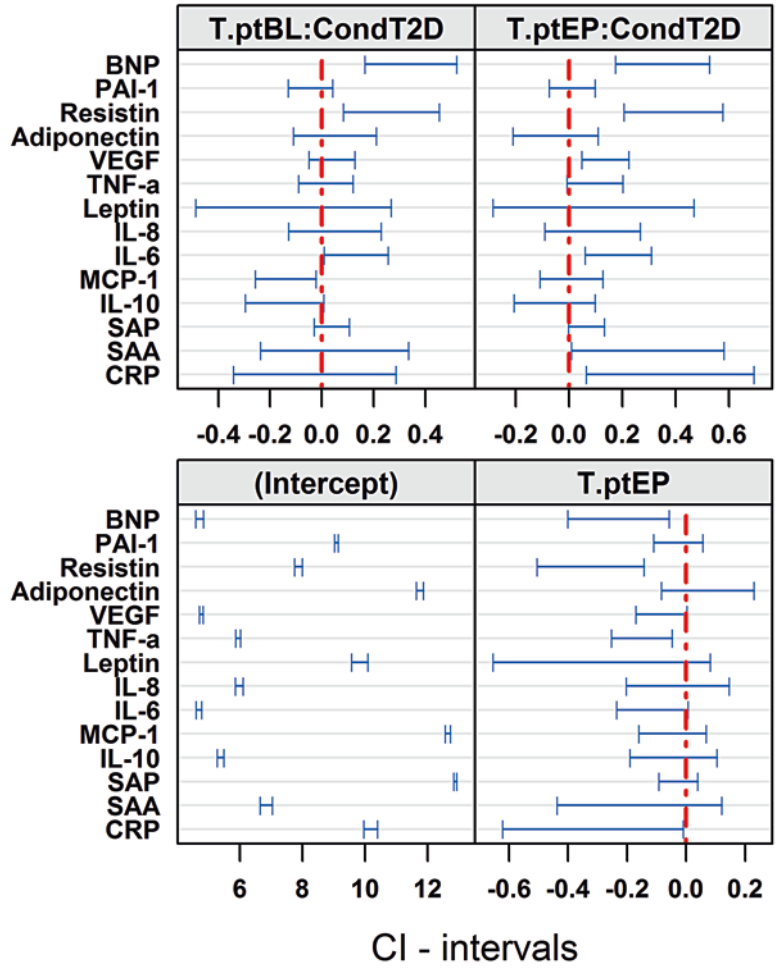


Fig. 12 Analyte 95% confidence interval. *Bottom left*: the estimates of the analyte log2 fluorescence responses intercept. *Bottom right*: difference between baseline (BL) and end point (EP) readings adjusted for patient condition. *Top left*: the baseline differences between control and T2D patients. *Top right*: the end point differences between control and T2D patients. Confidence intervals not crossing zero (*red dashed line*) are highlighting significant (p-*value* ≤ 0.05) analyte differences

conditions at end point. The bottom left plot gives the intercept values, which is actually the mean responses for each analyte at baseline. The bottom right gives the differences between analyte responses with respect to time points and shows that for Resistin and BNP their response values are significantly (p-value < 0.05) lower at end point than observed at base line. The top left plot shows that for BNP there is a significant difference at baseline where its response is higher in T2D patients than in the control group patients. The exact but unadjusted *p-values*, for the differences shown in Fig. 12, are retrieved using **summary**(**lst**). For example:

```
> summary(lst)$coefficients[,,3]
             Estimate Std. Error t value Pr(>|t|)
CRP           -0.0265     0.0991  -0.267 0.789107
SAA            0.0496     0.0991   0.500 0.616969
SAP            0.0390     0.0991   0.393 0.694178
IL-10         -0.1437     0.1046  -1.374 0.169547
MCP-1         -0.1392     0.1046  -1.331 0.183372
IL-6           0.1336     0.0991   1.347 0.177970
IL-8           0.0508     0.0991   0.513 0.608139
Leptin        -0.1090     0.0991  -1.100 0.271375
TNF-a          0.0162     0.0991   0.163 0.870303
VEGF           0.0399     0.0991   0.402 0.687491
Adiponectin    0.0518     0.0991   0.522 0.601631
Resistin       0.2697     0.0991   2.721 0.006525
PAI-1         -0.0430     0.0993  -0.433 0.665215
BNP            0.3446     0.0995   3.465 0.000532
```

where the index term [,,3] above is used to restrict the output to just the *p-values* associated with the third plot in Fig. 12 (top left), given that the bottom left plot is 1 ([,,1]). These results go directly to satisfying the statistical objectives 1 and 2 given above. To satisfy objectives 3 and 4, just reverse roles of **T.pt.** and **Cond** in the above formula passed to **lmList.** While **lmList** is very useful and powerful, the statistical correctness of this approach other than possible confirmatory type analysis through visualizations is questionable, because there are repeat values, nested relationships, and technical replicates which haven't been taken into consideration, and because of these relationships the data violates the independence assumption required for this type of analysis.

3.7.2 Random Effects

Random effects are the effects used to normalize the data against. While fixed effects influence the mean value of the responses, random effects influence the variances in the response. Fixed effects are estimated using least squares (maximum likelihood), and random effects are estimated with shrinkage (best linear unbiased predictors, BLUPs) [15, 29].

Assumptions:

- Fixed effects assume that the individual specific effect is correlated to the independent variable.
- Random effects assume the individual specific effects are uncorrelated with the independent variables.

All experimental factors that aren't considered as fixed effects can be treated by default as random effects if needed. However, this dichotomy isn't black and white, since the same factor can actually be used for both effects in one equation. There are also situations in which calling an effect fixed or random depends solely on the objective of the experiment and not on the variables themselves.

3.7.3 Repeat Measures

Many mixed-effect models which include repeated samples taken from the same patient or subject incorporate random effects associated with a single grouping factor such as patient or subject and as shown below for the model **mod.lmer**, albeit patient is nested in **Hospital**. Most of the time however, controlling this way for each subject/patient is enough to deal with all the nonindependence of

the residuals for each subject. If there is extra nonindependence (or even nonconstant variance) among the residuals when examined, you can add in the effect for the repeat factor. The below R session shows how to conduct a mixed-effect model in R using **lmer**. It implements a full three-way fixed effects ANOVA model with four scalar random effect terms and shows how to specify repeat values and nested relationships. The model also allows for possible random plate row and column effects.

Below the output from **lmer** is stored in the R object **mod.lmer**. The ANOVA model display of **mod.lmer** gives the significances of model's fixed effects, and it shows that there is a significant (*p-value* < 0.05) interaction between **Cytokine** and **Cond** and **T.pt.** and a significant interaction effect between **Cond** and **T.pt.**, but not with respect to the highest term, the three-way interaction. To inspect the model's random effects and coefficients, use **summary**(**mod.lmer**), **ranef**(**mod.lmer**), and/or **coef**(**mod.lmer**).

```
mod.lmer = lmer(log2(FI)~Cytokine*Cond*T.pt - 1 + (T.pt|Hospital/patient)
                + (1|kit/pid) +(1|row) + (1|col), bp)

> Anova(mod.lmer)
Analysis of Deviance Table (Type II Wald chisquare tests)

Response: log2(FI)
                  Chisq Df Pr(>Chisq)
Cytokine        68756.87 14    < 2e-16 ***
Cond                1.46  1     0.2270
T.pt                4.09  1     0.0431 *
Cytokine:Cond      53.43 13    7.6e-07 ***
Cytokine:T.pt      32.97 13     0.0017 **
Cond:T.pt          24.39  1    7.9e-07 ***
Cytokine:Cond:T.pt 15.21 13     0.2946
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3.7.4 Mixed-Effect Analysis Using Lmer

With **lmer** the fixed effects are specified first, to the right of the tilde, followed by one or more random terms enclosed in brackets, for example, (**1|x**). The fixed effects specified here are typical for a three-way analysis between three main effects (**Cytokine**, **Cond**, and **T.ptr**); however, note the −1 which specifies no intercept for the fixed effects. This only changes the main fixed effects but not the higher order effects. The no fixed effect intercept model as specified here has a scaling effect during model estimation preventing the generation of large eigenvalue ratios. For the random effects, the vertical bar in a random effect term is read as given or conditional on the grouping factor **x**. The form (1|x) is an intercept model of scalar random effects, where each level within the grouping factor **x** generates a random effect**.** For the above model, there are 177 **patient:Hospital** levels as patients are nested in hospital, 3 **Hospital** levels, 59 **pid:kit** levels, 6 **kit** levels, 10 **column**, and 8 **row** levels forming 263 scalar random effects in total. Note the nested structures in term such as **(1|x/y)** which are used to explicitly specify to **lmer** that **pid** is nested in **kit** and that **patient** is nested in **Hospital.** Generally, crossed random effects take the form **(1|x)** + **(1|y),** while nested random effects take the form

(1|x/y). For repeat measures such as for samples taken over a time course experiment or as is the case here with baseline and end point samples (**T.pt**), you specify the repeat factor, if required, using a random effect term of the form (**T.pt.|Hospital/patient**), which defines an intercept and slope model matrix for that term. This means that **Hospital** and **patient**:**Hospital** random effects will have both an intercept and a slope effect at each level.

3.7.5 Post hoc Analysis

Often, after an ANOVA, you know that the response variable differs significantly across the factors of interest, but you do not know which pairs of the factor levels are significantly different from each other. At this point, you then conduct post hoc pairwise comparisons. In R there are several packages for doing post hoc analysis, but here only **testInteractions** from the **phia** package [14] is considered. By using **testInteractions** the first two statistical objectives, given above, can be easily obtained simply by considering **Cytokine** and **T.pt.** as fixed parameters with respect to the pairwise contrast between patient conditions as shown below. Note also that the fourth argument passed to **testInteractions** is an adjustment parameter, to correct the *p-values* for multiple comparison correction, which is specified here as "fdr" to control for the false discovery rate [25, 30]:

```
> testInteractions(mod.lmer, fixed=c("Cytokine", "T.pt"),
                   pairwise='Cond', adjust='fdr')
Chisq Test:
P-value adjustment method: fdr
                                  Value Df   Chisq Pr(>Chisq)
cntrl-T2D :         CRP : BL  0.10055  1  0.9692  0.4548408
cntrl-T2D :         SAA : BL  0.02445  1  0.0573  0.8408554
cntrl-T2D :         SAP : BL  0.03505  1  0.1178  0.8192347
cntrl-T2D :       IL-10 : BL  0.22686  1  4.5353  0.1147158
cntrl-T2D :       MCP-1 : BL  0.22233  1  4.3563  0.1147158
cntrl-T2D :        IL-6 : BL -0.03152  1  0.0911  0.8214965
cntrl-T2D :        IL-8 : BL  0.05120  1  0.2403  0.7596016
cntrl-T2D :      Leptin : BL  0.21108  1  4.0849  0.1211532
cntrl-T2D :       TNF-a : BL  0.08585  1  0.6757  0.5481172
cntrl-T2D :        VEGF : BL  0.06215  1  0.3542  0.7022486
cntrl-T2D : Adiponectin : BL  0.04489  1  0.1847  0.7785805
cntrl-T2D :    Resistin : BL -0.17308  1  2.7453  0.2100884
cntrl-T2D :       PAI-1 : BL  0.14393  1  1.9793  0.2910029
cntrl-T2D :         BNP : BL -0.22981  1  4.8151  0.1147158
cntrl-T2D :         CRP : EP -0.58020  1 25.2564  1.405e-05 ***
cntrl-T2D :         SAA : EP -0.49631  1 18.4808  0.0002403 ***
cntrl-T2D :         SAP : EP -0.26561  1  5.2930  0.1147158
cntrl-T2D :       IL-10 : EP -0.12707  1  1.1232  0.4262436
cntrl-T2D :       MCP-1 : EP -0.18977  1  2.5053  0.2269275
cntrl-T2D :        IL-6 : EP -0.24225  1  4.6944  0.1147158
cntrl-T2D :        IL-8 : EP -0.14622  1  1.7102  0.2970406
cntrl-T2D :      Leptin : EP -0.15006  1  1.8014  0.2957245
cntrl-T2D :       TNF-a : EP -0.15477  1  1.9161  0.2910029
cntrl-T2D :        VEGF : EP -0.19418  1  3.0162  0.1923490
cntrl-T2D : Adiponectin : EP  0.00717  1  0.0041  0.9488547
cntrl-T2D :    Resistin : EP -0.43514  1 15.1386  0.0009324 ***
cntrl-T2D :       PAI-1 : EP -0.20432  1  3.1304  0.1923490
cntrl-T2D :         BNP : EP -0.39436  1 12.4239  0.0029672 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The output from **testInteractions** above is a data frame with four columns: **Value**, **Df**, **Chisq**, and **Pr(>Chisq)**. The column **Value** contains the difference between the means of interest. The rest of the columns show the multivariate test information applied

to the contrasts: the degrees of freedom (Df), the chi square statistics, and the associated *p-value*. The row names give, first, the pairwise contrast being considered, and in this instance, all rows represent the difference between **cntrl** and **T2** patients; this is then followed by a colon ":" and then the analyte considered in the contrast, followed by another ":" and then the **T.pt.** value used for that comparison. For example, the first row gives the contrast between **cntrl** and **T2D** for **CRP** with respect to baseline readings (**BL**). After the baseline results the end point (**EP**), contrasts are given. From the above table, we see no significant analyte difference between control and T2D patients at baseline ($p\text{-value} < 0.05$), but there are four analytes **CRP**, **SAA**, **SAP**, **Resistin**, and **BNP** significantly different with respect to patient type according to the end point samples. These results go toward answering objectives 1 and 2; for objectives 3 and 4, a call to **testInteractions**, but where the roles of **T.pt.** and **Cond** are reversed in the above, call to **testInteractions** is all that's needed.

3.7.6 Visualization of Multivariate Regression Models

With a simple linear regression model, it is very useful to plot the regression line overlaid on scatterplot of the data, as it informs very quickly how well the regression model fits the data. For multivariate regression models, it is a lot more difficult to view the results graphically. To do this the dependent variable and an explanatory variable are generally viewed while holding the other variables constant. This produces as conditional plot, since it shows the relationship between the outcome and explanatory variables conditional on the other explanatory variables being held constant at particular values:

```
> visreg(mod.lmer, 'Cond', by = 'Cytokine', cond=list(T.pt='EP'),
         scales = scales, as.table=TRUE)
```

The result of the above command is given in Fig. 13, where the interaction between **Cond** and **Cytokine** is visualized, with the condition that **T.pt.** = "EP." The scales parameter is as given previously. By default, conditional plots in **visreg** are constructed by filling in other explanatory variables with the median (for numeric variables) or most common category (for factors), but this is easily overridden by specifying their values using the **cond** parameter.

From the plot in Fig. 13 we can see that model, **mod.lmer**, appears to provide a reasonable fit to data because the scatter of points (residuals) around each conditional mean (red horizontal line) appears reasonably balanced. However, at times a more formal analysis of goodness of fit is required.

3.7.7 The Goodness of Fit

For mixed-effects models [24], there are two basic assumptions:

1. The within group errors are independent and identically normally distrubted with mean of zero and variance σ^2, and are independent of the random effects

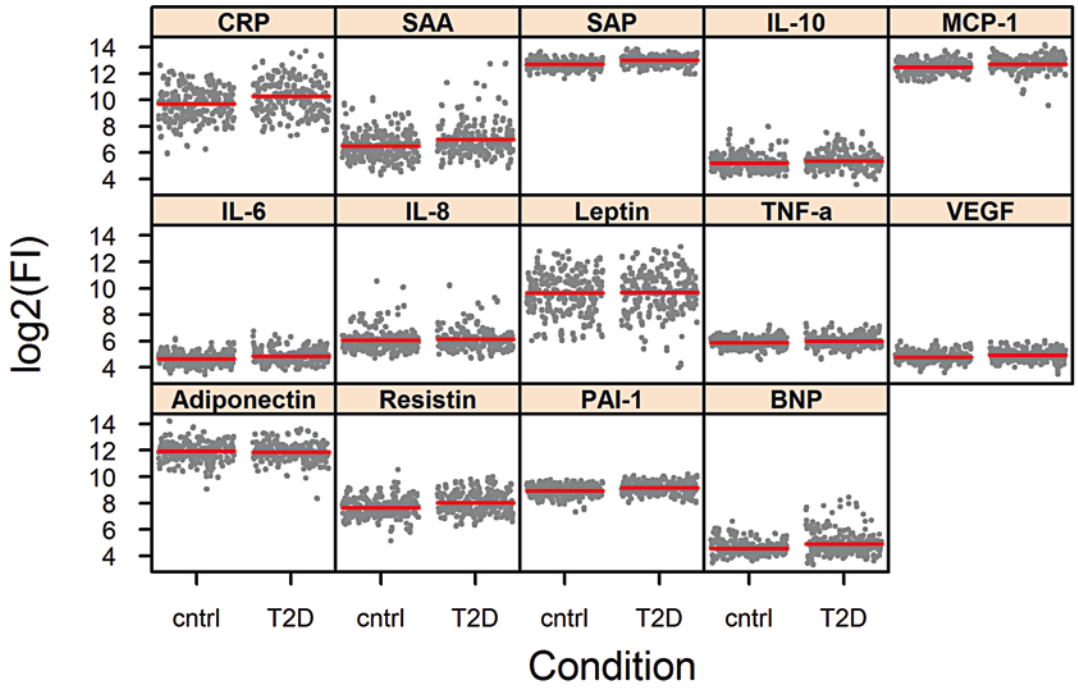


Fig. 13 Visreg visualization of a mixed-effect model with respect to the interaction between **Cytokine** and **Cond**, while **T.pt.** is held constant for end point readings, that is, **T.pt.** = "EP"'

2. The random effects are normally distributed with mean of zero and covariance matrix Ψ

While there are different functions in R for testing assumptions about distribution, most analysts agree that visual tests using diagnostic plots are best. For assumption 1 (the within group errors) below, we use R's plot function in the form **plot(model, formula)**; where **model** is the output R object from **lmer** and **formula** describing the components to be used in the plot. For example, **formula** = **y ~ x|g**. where **y** and **x** define the *y*-axis and *x*-axis components and **g** is an optional grouping term:

```
> plot(mod.lmer, Cytokine~resid(.), abline=0, col='red',as.table=T)
```

The result of the above command is given in Fig. 14a. The *x*-axis is defined by **resid(.)**, to represent the residuals from the **lmer** object **mod.lmer**. The argument **abline** = 0 is used to get a red vertical line at zero added to the plot. Note that the residuals for the **Cytokine** components are essentially centered on zero, and it appears that variability (IQR) for most analyte residuals is reasonably equivalent, but for **Leptin** and **CRP**, whose variances does appear larger.

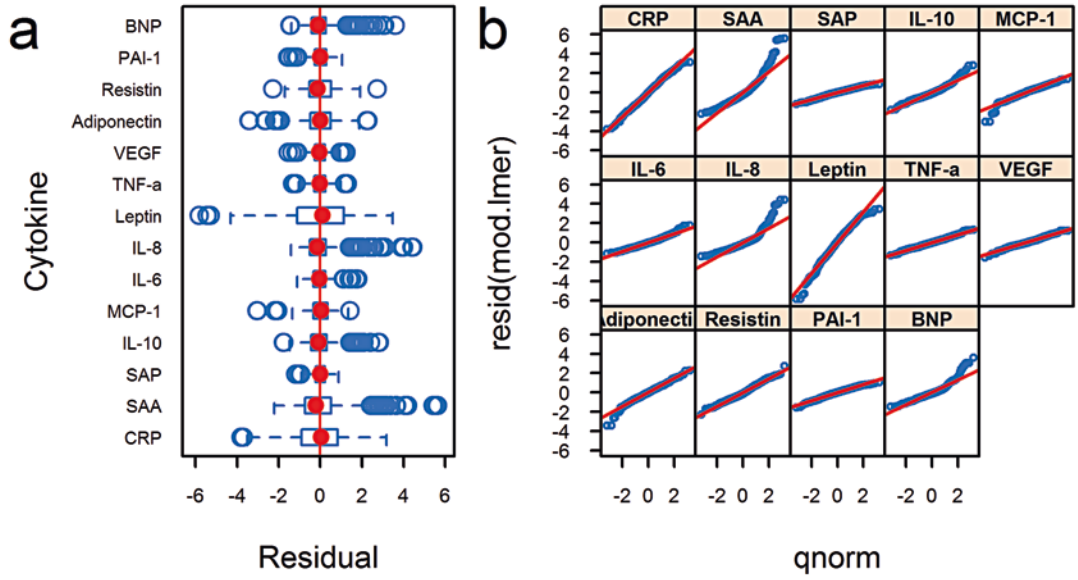


Fig. 14 Mod3 diagnostics tests for within group errors. (**a**) Boxplots for the residuals grouped by Cytokine; *Vertical red line* set on resid(.) = zero. (**b**) Normal QQ plot per analyte. *Red line* of fit passes through the first and third quartiles

The postulate of normality in assumption 1 can be assessed using R's qqmath function to obtain normal probability plots of the residual for each of the analytes:

```
> striptext=list(cex=0.8, font=2)
> qqmath(~resid(mod.lmer)|Cytokine, data=bp, type=c('r', 'p'),  lwd=1.5,
       par.settings=settings, par.strip.text=striptext, scales = scales,
       as.table=T)
```

The resulting plot from the above R session is given in Fig. 14b, where the red line, diagonal line, is used to indicate the expected outcome. The closer the observed (blue) points are to the red line, the better the fit. As seen in Fig. 14b, the residuals for most of the analytes appear normal; but for the some, like **SAA**, **BNP**, and **IL-8**, the fit could be better.

Assumption 2 above says the random effects should be centered on zero, be randomly distributed, and have equal covariances. These can be visually inspected, at least for the intercept model random terms, using another call to **bplxMER**:

```
> par(mfrow=c(6,2), mar=c(2,4,2,1))
> bplxMER(bp, mod.lmer, alpha = 0.1, doOutlier=F, cex=0.5)
```

The first line in the previous code uses R's **par** function to set up the display to show six rows by two columns of plots via the **mfrow** parameter, and it is used to set the margin around each plot according to the **mar** parameter. After the call to **par**, **bplxMER** is called passing it **bp**, **mod.lmer**, setting **alpha** to

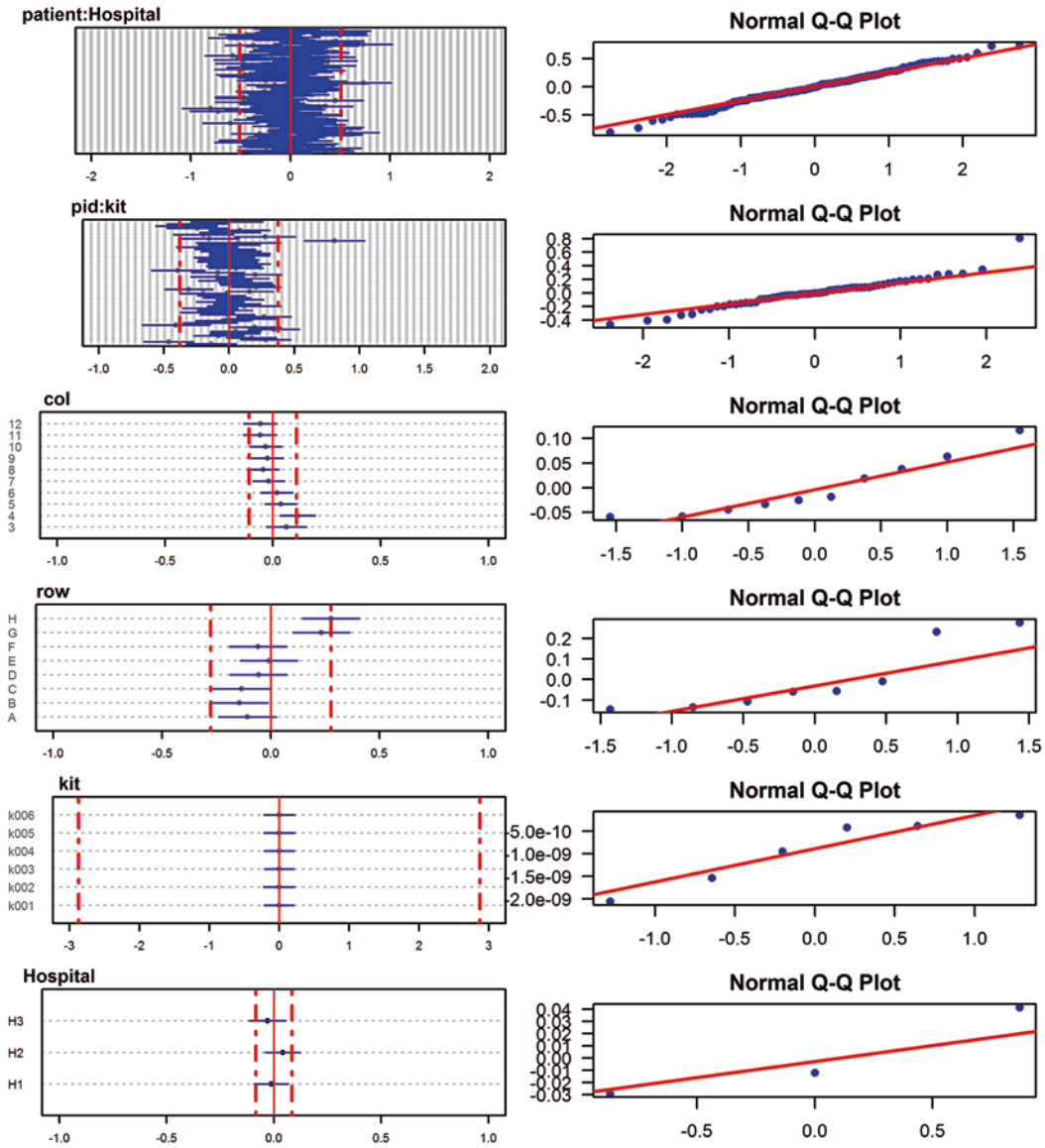


Fig. 15 Analysis of random effects assumptions: *left column* gives the random effects, for each effect. Where each effect should be centered on zero and the width of confidence intervals should be equal for each level of the effect. *Right column* gives the associated normal plots for each effect shown in the *left column*. The *red line* (*right column*) is used to indicate normality

0.1, and turning of the **doOutlier** analysis and sets **cex** to 0.5. The result of the above call is given in Fig. 15, where it is seen in the left column that most of the random effects are centered on zero. The **col.** and **row** random effects deviate slightly but essentially remain, within the 90% confidence limits. The QQ plot tests for normality appear reasonable. Therefore, overall it appears that the model, **mod.lmer**, is a reasonable fit to data.

4 Notes

A factor is just a categorical independent variable. When two factors are crossed, then every category of one factor co-occurs with every category (level) in the other factor. That is, there is at least one observation for all combinations for the two factors. When a factor is nested in another, then the first one coincides with one level of the other. For fixed effect modeling, nested factors can't be used to form interactions as not all combinations of the two factors have observations. Similarly, for partially crossed factors, that is, only fully crossed factors can be used to form interactions. However, with respect to random effects, the lme4 package is different from most other software for fitting mixed models in that it handles fully crossed and partially crossed random effects gracefully.

Acknowledgments

I would like to thank ResMed Science Center, ResMed Ltd., for allowing the use of the data generated for one of ResMed's clinical trial studies for this nonclinical manuscript. This work was undertaken at the Australian Proteome Analysis Facility (APAF) at Macquarie University, the infrastructure provided by the Australian Government through the Collaborative Research Infrastructure Strategy (CRIS) program.

References

1. Suh C-H, Kim H-A (2008) Cytokines and their receptors as biomarkers of systemic lupus erythematosus. Expert Rev Mol Diagn 8(2):189–198
2. Vignali DA (2000) Multiplexed particle-based flow cytometric assays. J Immunol Methods 243(1):243–255
3. Hanley B (2007) Variance in multiplex suspension array assays: carryover of microspheres between sample wells. J Negat Results Biomed 6(1):1
4. Breen EJ, Tan W, Khan A (2016) The statistical value of raw fluorescence signal in Luminex xMAP based multiplex immunoassays. Sci Rep 6:26996
5. Sanz H, Aponte J, Harezlak J, Dong Y, Murawska M, Valim C (2015) drLumi: multiplex immunoassays data analysis. Available from: https://CRAN.R-project.org/package=drLumi
6. Team RC (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.r-project.org/
7. Breen EJ, Polaskova V, Khan A (2015) Bead-based multiplex immuno-assays for cytokines, chemokines, growth factors and other analytes: median fluorescence intensities versus their derived absolute concentration values for statistical analysis. Cytokine 71(2):188–198
8. Ayroles JF, Gibson G (2006) Analysis of variance of microarray data. Methods Enzymol 411:214–233
9. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE et al (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet 11(10):733–739
10. Liu Q, Markatou M (2016) Evaluation of methods in removing batch effects on RNA-seq data. Infect Dis Transl Med 2(1):3–9
11. Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R (2006) Statistical practice in high-throughput screening data analysis. Nat Biotechnol 24(2):167–175

12. Altman N, Krzywinski M (2015) Points of significance: sources of variation. Nat Methods 12(1):5–6
13. Walker A (2015) openxlsx: read, write and edit XLSX files. Available from: https://CRAN.R-project.org/package=openxlsx
14. De Rosario-Martinez H (2015) phia: post-hoc interaction analysis. R package. Available from: http://CRAN.R-project.org/package=phia
15. Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:48
16. Venables WN, Ripley BD (2002) Modern applied statistics with S. Fourth, editor. Springer, New York
17. {others} GRWaBBaGGaGGaAKaTLaDMaA MaJRa (2015) gdata: various R programming tools for data manipulation. Available from: https://CRAN.R-project.org/package=gdata
18. Sarkar D (2008) Lattice: multivariate data visualization with R. Springer Science & Business Media, New York
19. Wickham H (2007) Reshaping data with the reshape package. J Stat Softw 21(12):1–20
20. Venables WN, Ripley BD, Venables WN (2002) Modern applied statistics with S. 4th ed. Springer, New York. xi, 495p
21. Fox J, Weisberg S (2011) An R companion to applied regression, 2nd edn. Sage, Thousand Oaks, CA
22. Krzywinski M, Altman N (2014) Points of significance: visualizing samples with box plots. Nat Methods 11(2):119–120
23. Cook RD, Weisberg S (1982) Residuals and influence in regression. Chapman and Hall, New York
24. Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. Springer, New York. xvi, 528p
25. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate – a practical and powerful approach to multiple testing. J Roy Stat Soc B Met 57(1):289–300
26. Rosenberg-Hasson Y, Hansmann L, Liedtke M, Herschmann I, Maecker HT (2014) Effects of serum and plasma matrices on multiplex immunoassays. Immunol Res 58(2–3):224–233
27. Fichorova RN, Richardson-Harman N, Alfano M, Belec L, Carbonneil C, Chen S et al (2008) Biological and technical variables affecting immunoassay recovery of cytokines from human serum and simulated vaginal fluid: a multicenter study. Anal Chem 80(12):4741–4751
28. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC (2016) nlme: linear and nonlinear mixed effects models. Available from: http://CRAN.R-project.org/package=nlme
29. Gelman A (2005) Analysis of variance—why it is more important than ever. Ann Stat 33(1):1–53
30. Liu P, Hwang JT (2007) Quick calculation for sample size while controlling false discovery rate with application to microarray analysis. Bioinformatics 23(6):739–746

Appendix A
Standard Operating Procedures for Plasma Collection in Clinical Research

The Early Detection Research Network (EDRN) Standard Operating Procedure (SOP) for Collection of EDTA Plasma

Abstract

The variables surrounding collection and processing of blood specimens affect blood chemistry and the proteome in a way that introduces systematic changes that may be mistakenly attributed to a particular physiopathological condition. A limiting factor of current clinical proteomic studies has been the lack of accepted pre-analytical and analytical guidelines. Recent worldwide efforts have been made to standardize blood collection, processing, and storage conditions for case and control samples as part of the HUPO Plasma Proteome Project initiative. Given the complexity of the blood proteome, collection and handling present a broad range of specific pre-analytical technical challenges, from venipuncture to the use of blood derivatives, protease inhibitors, and processing specifications and storage conditions. As the areas of clinical validation of different disease states from blood-derived sources (i.e., disease biomarkers) move toward validation stages, the importance of controlled and standardized protocols is imperative. The establishment of standard operating procedures (SOPs) for the collection and processing of plasma and sera allows for systematic analysis of samples without the potential of bias and variance. This protocol is the SOP for plasma/serum use in clinical proteome research, based on the Early Detection Research Network (EDRN) and Specimen Collection and Handling Committee (SCHC) of the HUPO Plasma Proteome Project [1].

General Requirements

- Gloves must be worn at all times when handling specimens. This includes during removal of the rubber stopper from the blood tubes, centrifugation, pipetting, disposal of contaminated tubes, and cleanup of any spills. Tubes, needles, and pipettes must be properly disposed of in biohazard containers, in accordance with institutional requirements.
- Universal precautions and OSHA (Occupational Safety and Health Administration) and institutional requirements (http://www.osha.gov/SLTC/biologicalagents/index.html) should be followed, including gloves, eye protection, or working in a biosafety cabinet for blood processing.

Reproduced and adapted with permission from [1].

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5, © Springer Science+Business Media LLC 2017

- All equipment (storage, shipping, and centrifuge) must be labeled as biohazard.
- It is important to take steps to prevent hemolysis in these samples. A vacutainer is recommended. If a needle is used, a 21 gauge needle is recommended.

EDTA Plasma Collection

Supplies

- EDTA Blood Collection Tubes (e.g., BD vacutainers catalog # 366450).
- Centrifuge with swinging bucket rotor.
- 15 mL polypropylene conical tubes (e.g., Corning 430052, Fisher catalog #05-538-53D).
- Sterile cryovials with writing surface (e.g., Simport T311-2 or Fisher #05-669-57).
- 2, 5, and 10 mL pipettes (e.g., Fisher catalog #13-678-11C, 13-678-11D, 13-678-11E).
- Disposable transfer pipettes (e.g., Fisher catalog #13-711-20).
- Automatic pipette aid.
- Small ice bucket.

Plasma Separation Procedure

1. After collection, gently mix the blood by inverting the tube 8–10 times. Store vacutainer tubes upright at 4 °C until centrifugation. Blood samples should be centrifuged within 4 h of blood collection.
2. Centrifuge blood samples in a horizontal rotor (swing-out head) for 10–20 min at 1100–1300 × *g* at room temperature (20 °C).
   Warning: Excessive centrifuge speed (over 2000 × *g*) may cause tube breakage and exposure to blood and possible injury. If needed, RCF for a centrifuge can be calculated. For an online calculator tool, please refer to:
   http://www.changbioscience.com/cell/rcf.html.
3. After centrifugation, plasma layer will be at the top of the tube. Mononuclear cells and platelets will be in a whitish layer, called the "buffy coat," just under the plasma and above the red blood cells (additional processing of these cell fractions is optional).
4. Carefully collect the plasma layer with an appropriate transfer pipette without disturbing the buffy coat layer. If more than one tube is collected, pool the plasma samples from both tubes into a 15 mL conical tube and mix. Pipette the plasma into appropriate sized aliquots in labeled cryovials. Aliquot volume is recommended to be 100 or 250 μL; however, some sites may

determine that 1 mL aliquot sizes are needed. Close the caps tightly and place on ice. This process should be completed within 1 h of centrifugation.

5. Check that all aliquot vial caps are secure and that all vials are labeled.
6. Place all aliquots upright in a specimen box or rack in an −80 °C or colder freezer. All specimens should remain at −80 °C or colder prior to shipping. The samples should not be thawed prior to shipping. (Plasma will be shipped on dry ice. Refer to SOP for "Shipping" instructions.)

Data Points

1. Date and time of blood collection.
2. Number and volume of aliquots prepared.
3. Date and time into −80 °C.
4. Date and time of shipping.
5. Any freeze-thaw that occurs with a sample for any reason.
6. Any variations or deviations from the SOP, problems, or issues.

Notes

- Sterile, disposable droppers, Pipetman, pipette aid, and Eppendorf repeater are examples of ways to aliquot. It depends on the aliquot size and volume and the plasma volume.
- Plasma should not undergo freeze-thaw cycles, so choose the aliquot volume carefully.
- Freezers need to have a backup generator or other emergency system options: Create emergency management plan, such as moving to a new freezer or adding dry ice in the event of a freezer failure.

Reference

1. Tuck MK, Chan DW, Chia D et al (2009) Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. J Proteome Res 8(1):113–117

Additional References

Hulmes JD, Bethea D, Ho K et al (2004) An investigation of plasma collection, stabilization, and storage procedures for proteomic analysis of clinical samples. In: Liotta LA, Petricoin E (eds) Clinical proteomics, vol 1. Humana Press, Totowa, NJ, pp 17–32

Mischak H, Apweiler R, Banks RE et al (2007) Clinical proteomics: a need to define the field and to begin to set adequate standards. Proteomics Clin Appl 1(1):148–156

Rai AJ, Gelfand CA, Haywood BC et al (2005) HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics 5(13):3262–3277

Tammen H (2008) Specimen collection and handling: standardization of blood sample collection. Methods Mol Biol 428:35–42

Vitzthum F, Siest G, Bunk DM et al (2007) Metrological sharp shooting for plasma proteins and peptides: the need for reference materials for accurate measurements in clinical proteomics and in vitro diagnostics to generate reliable results. Proteomics Clin Appl 1(9):1016–1035

Appendix B
Standard Operating Procedures for Serum Collection in Clinical Research

The Early Detection Research Network (EDRN) Standard Operating Procedure (SOP) for Collection of Serum

Abstract

The variables surrounding collection and processing of blood specimens affect blood chemistry and the proteome in a way that introduces systematic changes that may be mistakenly attributed to a particular physiopathological condition. A limiting factor of current clinical proteomic studies has been the lack of accepted pre-analytical and analytical guidelines. Recent worldwide efforts have been made to standardize blood collection, processing, and storage conditions for case and control samples as part of the HUPO Plasma Proteome Project initiative. Given the complexity of the blood proteome, collection and handling present a broad range of specific pre-analytical technical challenges, from venipuncture to the use of blood derivatives, protease inhibitors, and processing specifications and storage conditions. As the areas of clinical validation of different disease states from blood-derived sources (i.e., disease biomarkers) move toward validation stages, the importance of controlled and standardized protocols are imperative. The establishment of standard operating procedures (SOPs) for the collection and processing of plasma and sera allows for systematic analysis of samples without the potential of bias and variance. This protocol is the SOP for plasma/serum use in clinical proteome research, based on the Early Detection Research Network (EDRN) and Specimen Collection and Handling Committee (SCHC) of the HUPO Plasma Proteome Project [1].

General Requirements

- Gloves must be worn at all times when handling specimens. This includes during removal of the rubber stopper from the blood tubes, centrifugation, pipetting, disposal of contaminated tubes, and cleanup of any spills. Tubes, needles, and pipettes must be properly disposed of in biohazard containers, in accordance with institutional requirements.
- Universal precautions and OSHA (Occupational Safety and Health Administration) and institutional requirements (http://www.osha.gov/SLTC/biologicalagents/index.html) should be followed, including gloves, eye protection, or working in a biosafety cabinet for blood processing.

Reproduced and adapted with permission from [1].

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5, © Springer Science+Business Media LLC 2017

- All equipment (storage, shipping, and centrifuge) must be labeled as biohazard.
- It is important to take steps to prevent hemolysis in these samples. A vacutainer is recommended. If a needle is used, a 21 gauge needle is recommended.

Serum Collection

Supplies

- Red Top Vacutainer (NOT SST (serum separator tubes); these contain polymeric gels with several constituents to adjust viscosity, density, and other physical properties) (e.g., BD vacutainers catalog #366430).
- Centrifuge with swinging bucket rotor.
- 15 mL polypropylene conical tubes (e.g., Corning 430052, Fisher catalog #05-538-53D).
- Sterile cryovials with writing surface (e.g., Simport T311-2 or Fisher #05-669-57).
- 2, 5, and 10 mL pipettes (e.g., Fisher catalog #13-678-11C, 13-678-11D, 13-678-11E)..
- Disposable transfer pipettes (e.g., Fisher catalog #13-711-20).
- Automatic pipette aid.
- Small ice bucket.

Serum Separation Procedure

1. Filled red top blood collection tubes ("vacutainers") should sit upright after the blood is drawn at room temperature (20 °C) for a minimum of 30 to a maximum of 60 min to allow the clot to form.

   Note: Use red top (serum) tubes (silicon-coated)—no additives and not SST (serum separator tubes). These tubes, without additives, allow the red blood cells to form a clot. The clot also includes white blood cells, platelets, etc. After centrifuging, the clot is at the bottom of the tube, and the serum is on top of the clot. The red top tubes do not have to be full to be used.

2. Centrifuge the blood sample at the end of the clotting time (30–60 min) in a horizontal rotor (swing-out bucket) for 20 min at 1100–1300 × *g* at room temperature. If blood is not centrifuged immediately after the clotting time (30–60 min at room temperature), tubes should be refrigerated (4 °C) for no longer than 4 h.

   Warning: Excessive centrifuge speed (over 2000 × *g*) may cause tube breakage and exposure to blood and possible injury. If

needed, RCF for a centrifuge can be calculated. For an online calculator tool, please refer to: http://www.changbioscience.com/cell/rcf.html.

3. Use pipette to transfer the serum (recommendation: do not pour!). If more than one tube is drawn, pull the serum from both tubes into a 15 mL conical tube and mix. Pipette serum into the labeled cryovials, filling the vials in sequential order. Aliquot volume is recommended to be 100 or 250 μL. Close the caps on the vials tightly. This process should be completed within 1 h of centrifugation.
4. *Note*: Be very careful not to pick up red blood cells when aliquoting. This can be done by keeping the pipette above the red blood cell layer and leaving a small amount of serum in the tube. Check that all aliquot vial caps are secure and that all vials are labeled.
5. Place all aliquots upright in a specimen box or rack in an −80 °C or colder freezer. All specimens should remain at −80 °C or colder prior to shipping. The samples should not be thawed prior to shipping.

Data Points

1. Is the serum **hemolyzed? If yes, sample cannot be used.**
2. Date and time of blood collection.
3. Number and volume of aliquots prepared.
4. Date and time into −80 °C.
5. Date and time of shipping.
6. Any freeze-thaw that occurs with a sample for any reason.
7. Any variations or deviations from the SOP, problems, or issues.

Notes

- Eppendorf repeater are examples of ways to aliquot. It depends on the aliquot size and volume and the sera volume.
- Serum should not undergo freeze-thaw cycles, so choose the aliquot volume carefully.
- Freezers need to have a backup generator or other emergency system options: Create emergency management plan, such as moving to a new freezer or adding dry ice in the event of a freezer failure.

Reference

1. Tuck MK, Chan DW, Chia D et al (2009) Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. J Proteome Res 8(1):113–117

Additional References

Hulmes JD, Bethea D, Ho K et al (2004) An investigation of plasma collection, stabilization, and storage procedures for proteomic analysis of clinical samples. In: Liotta LA, Petricoin E (eds) Clinical proteomics, vol 1. Humana Press, Totowa, NJ, pp 17–32

Mischak H, Apweiler R, Banks RE et al (2007) Clinical proteomics: a need to define the field and to begin to set adequate standards. Proteomics Clin Appl 1(1):148–156

Rai AJ, Gelfand CA, Haywood BC et al (2005) HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. Proteomics 5(13):3262–3277

Tammen H (2008) Specimen collection and handling: standardization of blood sample collection. Methods Mol Biol 428:35–42

Vitzthum F, Siest G, Bunk DM et al (2007) Metrological sharp shooting for plasma proteins and peptides: the need for reference materials for accurate measurements in clinical proteomics and in vitro diagnostics to generate reliable results. Proteomics Clin Appl 1(9):1016–1035

Appendix C
Reference Ranges for Blood Tests Are Sorted by Mass and Molarity

Hormones predominate at the left part of the scale, shown at ng/L or pmol/L, being in very low concentration. There appears to be the greatest cluster of substances at μg/L or nmol/L range, becoming less so toward mg/L or μmol/L. However, there is another cluster containing many metabolic substances like cholesterol and glucose at the limit g/L or mmol/L range.

To translate a substance from the molar to the mass concentration scale above:

- Numerically: *molar concentration* × *molar mass* = *mass concentration*.
- Measure directly in distance on the scales.

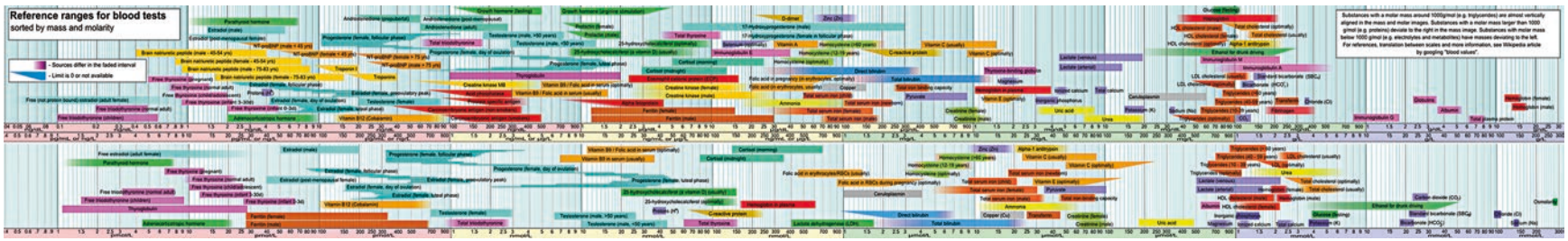
Author: Mikael Häggström

Date: April 2010

Website: http://en.wikipedia.org/wiki/Reference_ranges_for_blood_tests

Reference range list from Uppsala University Hospital ("Laborationslista"). Artnr 40284 Sj74a. Issued on April 22, 2008.

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5, © Springer Science+Business Media LLC 2017


Reference ranges for blood tests
sorted by mass and molarity
- Sources differ in the faded interval
- Limit is 0 or not available
Substances with a molar mass around 1000g/mol (e.g. triglycerides) are almost vertically aligned in the mass and molar images. Substances with a molar mass larger than 1000 g/mol (e.g. proteins) deviate to the right in the mass image. Substances with molar mass below 1000 g/mol (e.g. electrolytes and metabolites) have masses deviating to the left.
For references, translation between scales and more information, see Wikipedia article by googling "blood values".

INDEX

David W. Greening and Richard J. Simpson (eds.), *Serum/Plasma Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1619, DOI 10.1007/978-1-4939-7057-5, © Springer Science+Business Media LLC 2017

E

F

G

H

I

L

M

N

P

Q

R

S

T

U

V