

Paul C. Guest *Editor*

# Multiplex Biomarker Techniques

Methods and Applications

# 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>

# **Multiplex Biomarker Techniques**

## **Methods and Applications**

Edited by

**Paul C. Guest**

*Laboratory of Neuroproteomics, University of Campinas (UNICAMP), Campinas, Brazil*



*Editor*

Paul C. Guest  
Laboratory of Neuroproteomics  
University of Campinas (UNICAMP)  
Campinas, Brazil

ISSN 1064-3745

Methods in Molecular Biology

ISBN 978-1-4939-6729-2

DOI 10.1007/978-1-4939-6730-8

ISSN 1940-6029 (electronic)

ISBN 978-1-4939-6730-8 (eBook)

Library of Congress Control Number: 2016958277

© Springer Science+Business Media LLC 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

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

Due to continuous technical developments and new insights into the high complexity of many diseases, there is an increasing need for multiplex biomarker readouts for improved clinical management and to support the development of new drugs by pharmaceutical companies. The initial rollout of these techniques has led to promising results by helping to read patients as deeply as possible and provide clinicians with information relevant for a personalized medicine approach. This book describes the basic technology platforms being applied in the fields of genomics, proteomics, transcriptomics, metabolomics, and imaging, which are currently the methods of choice in multiplex biomarker research. It also describes the chief medical areas in which the greatest progress has been made and highlights areas where further resources are required.

More than 1000 biomarker candidates for various diseases have been described in the scientific literature over the last 20 years. However, the rate of introduction of new biomarker tests into the clinical arena is much lower with less than 100 such tests actually receiving approval and appearing on the marketplace. This disconnect is most likely due to inconsistencies at the discovery end, such as technical variations within and between platforms, a lack of validation of biomarker candidates, as well as a lack of awareness within the research community of the criteria and regulatory matters for integrating biomarkers into the pipeline [1]. Another reason relates to the fact that many diseases are heterogeneous in nature and comprised of different subtypes. This can cause difficulties in studies attempting to identify biomarkers since different investigators may analyze cohorts comprised of unique or even mixed subtypes of a particular disease, and this can make comparisons both within and across studies invalid. Furthermore, the use of patient and control groups in clinical studies which have not been properly stratified according to biomarker profiling is one of the biggest causes of failure in the development of new drugs [2–9].

One way of addressing these issues is through the increasing use of multiplex biomarker tests which can provide a more complete picture of a disease. Multiplex biomarker assays can simultaneously measure multiple analytes in one run on a single instrument as opposed to methods that measure only one analyte at a time or multiple analytes at different times. The simultaneous measurement of different biomarkers in a multiplex format allows for lower sample and reagent requirements along with reduced processing times on a per assay basis (Table 1). In contrast, testing for single analytes can be laborious, time-consuming, and expensive in cases where multiple assays for different molecules are required.

So how does multiplexing improve classification of diseases?

Multiplexing allows for higher sample throughput with greater cross-comparability within and across experiments since each of the component assays are processed, read, and analyzed under identical conditions and at the same time. This obviates traditional problems of comparing the results of single assays within a given study, which may be subject to procedural inconsistencies in sampling, methodology, or data analysis. Most importantly, the use of multiple biomarkers allows for greater accuracy in the diagnosis of complex diseases by providing more complete information about the perturbed physiological pathways in a shorter time period. This includes in-depth attempts to decipher pathological changes

**Table 1**  
**Characteristics of single versus multiplex immunoassays**

|                  | <b>Advantages</b>  | <b>Disadvantages</b>   |
|------------------|--|--|
| Single assays    | <p>Greater sensitivity because there is no competition of different analytes for reagents</p> <p>Useful as a validation test after identification of biomarker candidates</p>  | <p>Requires prior knowledge to target specific analytes</p> <p>Requires greater amounts of sample per analyte</p> <p>Requires greater amounts of reagents per analyte</p> <p>Greater amount of time required for analysis of multiple analytes (in proportion to analyte number)</p> <p>Low cross-comparability of multiple assays as each one is run under different conditions and at a different time</p> |
| Multiplex assays | <p>No prior knowledge required as it can be used for screening</p> <p>Greater cross-comparability across analytes as all are run simultaneously under the same conditions</p> <p>More understanding of physiological pathways affected in disease due to higher number of simultaneous analyte measurements</p> <p>Lower amounts of sample required per analyte</p> <p>Lower amounts of reagents required per analyte</p> <p>Lower time required for analysis of multiple analytes</p> | <p>Requires more complex and stringent statistical analyses</p> <p>Often requires bioinformatic analyses to identify over-represented pathways</p> <p>Requires validation of analytes identified as significant during screens using an alternate technology</p>   |

at the level of the DNA sequence [10], epigenome [11], transcriptome [12], proteome [2], and metabolome [13]. Thus, we are now moving away from single biomarker tests to more comprehensive multiplex biomarker analyses in order to better classify and combat these disorders. This works in the same way that a complete fingerprint allows for more accurate identification of a suspect in a criminal investigation as opposed to a partial print which may not be resolvable across multiple suspects.

However, there are still challenges ahead. While some diseases are increasingly being treated according to biomarker profiling patterns, the “one-size-fits-all” approach is still the standard treatment for most diseases. Many diseases such as cancers [14–16], heart disease [17], diabetes and neurological disorders [18–20] present difficult problems when it comes to deciding on treatment options since multiple molecular pathways of complex network signaling cascades can be affected. In addition, as these disorders can affect all age groups and both sexes, even more variables can occur, leading to even greater variability. In order to deal with this issue, collaborative research networks should be established for multiplexing efforts to better integrate biomarker discovery in real time to targeted therapeutics.

In the United States, the Clinical Laboratory Improved Amendments (CLIA) act was passed by Congress in 1988 as a means of integrating quality testing for all laboratories and to ensure accuracy, reproducibility, and speed of patient testing results [21]. The Food and Drug Administration (FDA) is the responsible agency for applying these regulations for the purpose of categorizing biomarker assays based on technological complexity and ease of operation. Laboratory-developed tests have not necessarily received automatic approval and have traditionally been endorsed only at the FDA's discretion. This is because the clinical validation of multiplex biomarker tests will require the participation of multiple laboratories, and the resulting platforms are likely to need simplification stages and demonstration of increased robustness to merit extensive clinical applications. Multiplex tests may also require the use of an algorithm to derive a composite "score" representing the multiple values of each component assay for a classification or diagnosis. For example, scores of 100 and 0 would mean a 100% and 0% chance respectively that the disease is present. Of course, scores in the middle range would be less precise. Besides the multiplexing of analytes, another level of multiplexing can be achieved by running both patient and control samples in the same assay. For example, both cDNA arrays and two-dimensional gel electrophoresis (2D-DIGE) enable the analysis of hundreds of analytes simultaneously for up to three samples at the same time through the prelabeling of sample extracts with different spectrally resolvable fluorescent dyes.

The multiplex platforms for carrying out screening typically have medium to large footprints and require considerable expertise to operate. For transcriptomic or RNA-based profiling, these include quantitative PCR, cDNA microarray, and microRNA approaches. For proteomics, there are two-dimensional difference gel electrophoresis, multiplex immunoassay, label-free shot-gun mass spectrometry, selective reaction mass spectrometry, and labeled-based mass spectrometry platforms. For metabolomic screening, the main platforms in use are either mass spectrometry or proton nuclear magnetic resonance-based. For clinical applications and rollout of biomarker assays, it is becoming increasingly important that the platforms are small, user friendly, and fast so they can be used in a point-of-care setting. The latest developments along these lines include lab-on-a-chip and mobile phone applications. Detailed protocols describing both the discovery and point-of-care devices incorporating multiplexed assays are described in this book.

*Campinas, Brazil*

*Paul C. Guest*

## References

1. Boja ES, Jortani SA, Ritchie J, Hoofnagle AN, Težak Ž, Mansfield E et al (2011) The journey to regulation of protein-based multiplex quantitative assays. *Clin Chem* 57:560–567.
2. Lee JM, Kohn EC (2010) Proteomics as a guiding tool for more effective personalized therapy. *Ann Oncol* 21(Suppl 7):vii205–vii10. doi: 10.1093/annonc/mdq375. Review
3. Lee JM, Han JJ, Altwerger G, Kohn EC (2011) Proteomics and biomarkers in clinical trials for drug development. *J Proteomics* 74:2632–2641.
4. Kelloff GJ, Sigman CC (2012) Cancer biomarkers: selecting the right drug for the right patient. *Nat Rev Drug Discov* 11:201–214.
5. Begg CB, Zabor EC, Bernstein JL, Bernstein L, Press MF, Seshan VE (2013) A conceptual and methodological framework for investigating etiologic heterogeneity. *Stat Med* 32:5039–5052.
6. Henriksen K, O'Bryant SE, Hampel H, Trojanowski JQ, Montine TJ, Jeromin A et al (2014) Blood-based biomarker interest group. *Alzheimers Dement* 10:115–131.
7. Guest PC, Chan MK, Gottschalk MG, Bahn S (2014) The use of proteomic biomarkers for improved diagnosis and stratification of schizophrenia patients. *Biomark Med* 8:15–27.

8. Wang X, Adjei AA (2015) Lung cancer and metastasis: new opportunities and challenges. *Cancer Metastasis Rev* 34:169–171.
9. Hudler P (2015) Challenges of deciphering gastric cancer heterogeneity. *World J Gastroenterol* 21:10510–10527.
10. Hudson TJ (2013) Genome variation and personalized cancer medicine. *J Intern Med* 274:440–450.
11. Mummaneni P, Shord SS (2014) Epigenetics and oncology. *Pharmacotherapy* 34:495–505.
12. Hu YF, Kaplow J, He Y (2005) From traditional biomarkers to transcriptome analysis in drug development. *Curr Mol Med* 5:29–38.
13. Wishart DS (2008) Applications of metabolomics in drug discovery and development. *Drugs R D* 9:307–322.
14. Füzéry AK, Levin J, Chan MM, Chan DW (2013) Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. *Clin Proteomics* 10:13. doi: 10.1186/1559-0275-10-13.
15. Nolen BM, Lokshin AE (2013) Biomarker testing for ovarian cancer: clinical utility of multiplex assays. *Mol Diagn Ther* 17:139–146.
16. Ploussard G, de la Taille A (2010) Urine biomarkers in prostate cancer. *Nat Rev Urol* 7:101–109.
17. Vistnes M, Christensen G, Omland T (2010) Multiple cytokine biomarkers in heart failure. *Expert Rev Mol Diagn* 10:147–157.
18. Lee KS, Chung JH, Choi TK, Suh SY, Oh BH, Hong CH (2009) Peripheral cytokines and chemokines in Alzheimer's disease. *Dement Geriatr Cogn Disord* 28:281–287.
19. Kang JH, Vanderstichele H, Trojanowski JQ, Shaw LM (2012) Simultaneous analysis of cerebrospinal fluid biomarkers using microsphere-based xMAP multiplex technology for early detection of Alzheimer's disease. *Methods* 56:484–493.
20. Chan MK, Krebs MO, Cox D, Guest PC, Yolken RH, Rahmoune H et al (2015) Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. *Transl Psychiatry* 5:e601. doi: 10.1038/tp.2015.91.
21. <http://www.cms.gov/clia>.

---

# Contents

|                    |      |
|--------------------|------|
| Preface .....      | v    |
| Contributors ..... | xiii |

## PART I REVIEWS

|   |  |
|---|--|
| 1 Application of Multiplex Biomarker Approaches to Accelerate Drug Discovery and Development .....                            | 3<br><i>Hassan Rahmoune and Paul C. Guest</i>  |
| 2 The Application of Multiplex Biomarker Techniques for Improved Stratification and Treatment of Schizophrenia Patients ..... | 19<br><i>Johann Steiner, Paul C. Guest, Hassan Rahmoune, and Daniel Martins-de-Souza</i>         |
| 3 Multiplex Biomarker Approaches in Type 2 Diabetes Mellitus Research.....  | 37<br><i>Susan E. Ozanne, Hassan Rahmoune, and Paul C. Guest</i>                                 |
| 4 LC-MSE, Multiplex MS/MS, Ion Mobility, and Label-Free Quantitation in Clinical Proteomics .....                             | 57<br><i>Gustavo Henrique Martins Ferreira Souza, Paul C. Guest, and Daniel Martins-de-Souza</i> |
| 5 Phenotyping Multiple Subsets of Immune Cells In Situ in FFPE Tissue Sections: An Overview of Methodologies .....            | 75<br><i>James R. Mansfield</i>  |

## PART II STATISTICAL CONSIDERATIONS

|  |  |
|--|--|
| 6 Identification and Clinical Translation of Biomarker Signatures: Statistical Considerations..... | 103<br><i>Emanuel Schwarz</i>                  |
| 7 Opportunities and Challenges of Multiplex Assays: A Machine Learning Perspective .....           | 115<br><i>Junfang Chen and Emanuel Schwarz</i> |

## PART III PROTOCOLS

|  |  |
|--|--|
| 8 Multiplex Analyses Using Real-Time Quantitative PCR.....     | 125<br><i>Steve F.C. Hawkins and Paul C. Guest</i> |
| 9 Multiplex Analysis Using cDNA Transcriptomic Profiling ..... | 135<br><i>Steve F.C. Hawkins and Paul C. Guest</i> |
| 10 Multiplex Single Nucleotide Polymorphism Analyses.....      | 143<br><i>Steve F.C. Hawkins and Paul C. Guest</i> |

|    |  |     |
|----|--|-----|
| 11 | Pulsed SILAC as a Approach for miRNA Targets Identification<br>in Cell Culture . . . . .   | 149 |
|    | <i>Daniella E. Duque-Guimarães, Juliana de Almeida-Faria,<br/>Thomas Prates Ong, and Susan E. Ozanne</i>   |     |
| 12 | Blood Bio-Sampling Procedures for Multiplex Biomarkers Studies. . . . .  | 161 |
|    | <i>Paul C. Guest and Hassan Rahmoune</i>   |     |
| 13 | Multiplex Immunoassay Profiling . . . . .  | 169 |
|    | <i>Laurie Stephen</i>  |     |
| 14 | Multiplex Sequential Immunoprecipitation of Insulin Secretory<br>Granule Proteins from Radiolabeled Pancreatic Islets. . . . .   | 177 |
|    | <i>Paul C. Guest</i>   |     |
| 15 | Two Dimensional Gel Electrophoresis of Insulin Secretory Granule Proteins<br>from Biosynthetically-Labeled Pancreatic Islets . . . . .   | 187 |
|    | <i>Paul C. Guest</i>   |     |
| 16 | Depletion of Highly Abundant Proteins of the Human Blood Plasma:<br>Applications in Proteomics Studies of Psychiatric Disorders . . . . .  | 195 |
|    | <i>Sheila Garcia, Paulo A. Baldasso, Paul C. Guest,<br/>and Daniel Martins-de-Souza</i>  |     |
| 17 | Simultaneous Two-Dimensional Difference Gel Electrophoresis (2D-DIGE)<br>Analysis of Two Distinct Proteomes . . . . .  | 205 |
|    | <i>Adriano Aquino, Paul C. Guest, and Daniel Martins-de-Souza</i>  |     |
| 18 | Selective Reaction Monitoring for Quantitation of Cellular Proteins . . . . .  | 213 |
|    | <i>Vitor M. Faça</i>   |     |
| 19 | Characterization of a Protein Interactome by Co-Immunoprecipitation<br>and Shotgun Mass Spectrometry . . . . .   | 223 |
|    | <i>Giuseppina Maccarrone, Juan Jose Bonfiglio, Susana Silberstein,<br/>Christoph W. Turck, and Daniel Martins-de-Souza</i>   |     |
| 20 | Using <sup>15</sup> N-Metabolic Labeling for Quantitative Proteomic Analyses. . . . .  | 235 |
|    | <i>Giuseppina Maccarrone, Alon Chen, and Michaela D. Filion</i>  |     |
| 21 | Multiplex Measurement of Serum Folate Vitamers by UPLC-MS/MS. . . . .  | 245 |
|    | <i>Sarah Meadows</i>   |     |
| 22 | UPLC-MS/MS Determination of Deuterated 25-Hydroxyvitamin D<br>(d <sub>3</sub> -25OHD <sub>3</sub> ) and Other Vitamin D Metabolites for the Measurement<br>of 25OHD Half-Life. . . . . | 257 |
|    | <i>Shima Assar, Inez Schoenmakers, Albert Koulman, Ann Prentice,<br/>and Kerry S. Jones</i>  |     |
| 23 | iTRAQ-Based Shotgun Proteomics Approach for Relative<br>Protein Quantification . . . . .   | 267 |
|    | <i>Erika Velásquez Núñez, Gilberto Barbosa Domont, and Fábio César Sousa<br/>Nogueira</i>  |     |
| 24 | <sup>1</sup> H NMR Metabolomic Profiling of Human and Animal Blood<br>Serum Samples . . . . .  | 275 |
|    | <i>João G.M. Pontes, Antonio J.M. Brasil, Guilherme C.F. Cruz,<br/>Rafael N. de Souza, and Ljubica Tasic</i>   |     |

|    |  |     |
|----|--|-----|
| 25 | Lab-on-a-Chip Multiplex Assays . . . . .   | 283 |
|    | <i>Harald Peter, Julia Wienke, and Frank F. Bier</i>                                 |     |
| 26 | Multiplex Smartphone Diagnostics . . . . .   | 295 |
|    | <i>Juan L. Martinez-Hurtado, Ali K. Yetisen, and Seok-Hyun Yun</i>                   |     |
| 27 | Development of a User-Friendly App for Assisting Anticoagulation Treatment . . . . . | 303 |
|    | <i>Johannes Vegt</i>   |     |

#### PART IV FUTURE DIRECTIONS

|    |  |     |
|----|--|-----|
| 28 | Multiplex Biomarker Approaches to Enable Point-of-Care Testing and Personalized Medicine . . . . . | 311 |
|    | <i>Paul C. Guest</i>   |     |
|    | <i>Index</i> . . . . .   | 317 |

---

## Contributors

- JULIANA DE ALMEIDA-FARIA • *University of Cambridge Metabolic Research Laboratories and MRC Metabolic Diseases Unit, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; Faculty of Medical Sciences, Department of Pharmacology, State University of Campinas, Campinas, Brazil*
- ADRIANO AQUINO • *Laboratory of Neuroproteomics, Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas (UNICAMP), Campinas, SP, Brazil*
- SHIMA ASSAR • *MRC Elsie Widdowson Laboratory, Cambridge, UK*
- PAULO A. BALDASSO • *Laboratory of Neuroproteomics, Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas (UNICAMP), Campinas, SP, Brazil*
- FRANK F. BIER • *Fraunhofer Institute for Cell Therapy and Immunology, Branch Bioanalytics and Bioprocesses (IZI-BB), Potsdam, Germany*
- JUAN JOSE BONFIGLIO • *Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA), CONICET, Partner Institute of the Max Planck Society, Buenos Aires, Argentina*
- ANTONIO J.M. BRASIL • *Chemical Biology Laboratory, Institute of Chemistry, Organic Chemistry Department, State University of Campinas, Campinas, SP, Brazil*
- ALON CHEN • *Department Stress Neurobiology and Neurogenetics, Max Planck Institute of Psychiatry, Munich, Germany*
- JUNFANG CHEN • *Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany*
- GUILHERME C.F. CRUZ • *Chemical Biology Laboratory, Institute of Chemistry, Organic Chemistry Department, State University of Campinas, Campinas, SP, Brazil*
- GILBERTO BARBOSA DOMONT • *Laboratory of Protein Chemistry-Proteomics Unit, Chemistry Institute, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil*
- DANIELLA E. DUQUE-GUIMARÃES • *University of Cambridge Metabolic Research Laboratories and MRC Metabolic Diseases Unit, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; Department of Physiology and Biophysics, Institute of Biomedical Sciences, University of São Paulo, SP, Brazil*
- VITOR M. FAÇA • *Department of Biochemistry and Immunology, Ribeirão Preto Medical School, University of São Paulo, RibeirãoPreto, SP, Brazil; Center for Cell Based Therapy - Hemotherapy Center of Ribeirão Preto, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil*
- MICHAELA D. FILIOU • *Department Stress Neurobiology and Neurogenetics, Max Planck Institute of Psychiatry, Munich, Germany*
- SHEILA GARCIA • *Laboratory of Neuroproteomics, Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas (UNICAMP), Campinas, SP, Brazil*
- PAUL C. GUEST • *Laboratory of Neuroproteomics, Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas (UNICAMP), Campinas, SP, Brazil*

- STEVE F.C. HAWKINS • *Bioline Reagents Limited, Unit 16, The Edge Business Centre, London, UK*
- KERRY S. JONES • *MRC Elsie Widdowson Laboratory, Cambridge, UK*
- ALBERT KOULMAN • *MRC Elsie Widdowson Laboratory, Cambridge, UK*
- GIUSEPPINA MACCARRONE • *Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany*
- JAMES R. MANSFIELD • *Andor Technology, Concord, MA, USA*
- JUAN L. MARTINEZ-HURTADO • *Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK*
- DANIEL MARTINS-DE-SOUZA • *Department of Biochemistry and Tissue Biology, Neurobiology Center and Laboratory of Neuroproteomics, Institute of Biology, University of Campinas (UNICAMP), Campinas, SP, Brazil*
- SARAH MEADOWS • *MRC Human Nutrition Research, Cambridge, UK*
- FÁBIO CÉSAR SOUSA NOGUEIRA • *Laboratory of Protein Chemistry—Proteomics Unit, Chemistry Institute, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil*
- ERIKA VELÁSQUEZ NÚÑEZ • *Laboratory of Protein Chemistry—Proteomics Unit, Chemistry Institute, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil*
- THOMAS PRATES ONG • *University of Cambridge Metabolic Research Laboratories and MRC Metabolic Diseases Unit, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; Food Research Center (FoRC) and Faculty of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil*
- SUSAN E. OZANNE • *University of Cambridge Metabolic Research Laboratories and MRC Metabolic Diseases Unit, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK*
- HARALD PETER • *Fraunhofer Institute for Cell Therapy and Immunology, Branch Bioanalytics and Bioprocesses (IZI-BB), Potsdam, Germany*
- JOÃO G.M. PONTES • *Chemical Biology Laboratory, Institute of Chemistry, Organic Chemistry Department, State University of Campinas, Campinas, SP, Brazil*
- ANN PRENTICE • *MRC Elsie Widdowson Laboratory, Cambridge, UK*
- HASSAN RAHMOUNE • *Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK*
- INEZ SCHOENMAKERS • *MRC Elsie Widdowson Laboratory, Cambridge, UK*
- EMANUEL SCHWARZ • *Department of Psychiatry and Psychotherapy, Medical Faculty Mannheim, Central Institute of Mental Health, Heidelberg University, Mannheim, Germany*
- SUSANA SILBERSTEIN • *Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA), CONICET, Partner Institute of the Max Planck Society, Buenos Aires, Argentina*
- GUSTAVO HENRIQUE MARTINS FERREIRA SOUZA • *Mass Spectrometry Applications and Development Laboratory, Waters Corporation, São Paulo, SP, Brazil*
- RAFAEL N. DE SOUZA • *Chemical Biology Laboratory, Institute of Chemistry, Organic Chemistry Department, State University of Campinas, Campinas, SP, Brazil*
- JOHANN STEINER • *Department of Psychiatry, University of Magdeburg, Magdeburg, Germany*
- LAURIE STEPHEN • *Ampersand Biosciences, Saranac Lake, NY, USA*
- LJUBICA TASIC • *Chemical Biology Laboratory, Institute of Chemistry, Organic Chemistry Department, State University of Campinas, Campinas, SP, Brazil*

CHRISTOPH W. TURCK • *Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany*

JOHANNES VEGT • *Appamedix UG i.gr, Innovations-Centrum CHIC, Berlin, Germany*

JULIA WIENKE • *Fraunhofer Institute for Cell Therapy and Immunology, Branch Bioanalytics and Bioprocesses (IZI-BB), Potsdam, Germany*

ALI K. YETISEN • *Harvard Medical School and Wellman Center for Photomedicine, Massachusetts General Hospital, Cambridge, MA, USA; Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA*

SEOK H. YUN • *Harvard Medical School and Wellman Center for Photomedicine, Massachusetts General Hospital, Cambridge, MA, USA; Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA*

# **Part I**

## **Reviews**

# Chapter 1

## Application of Multiplex Biomarker Approaches to Accelerate Drug Discovery and Development

Hassan Rahmoune and Paul C. Guest

### Abstract

Multiplex biomarker tests are becoming an essential part of the drug development process. This chapter explores the role of biomarker-based tests as effective tools in improving preclinical research and clinical development, and the challenges that this presents. The potential of incorporating biomarkers in the clinical pipeline to improve decision making, accelerate drug development, improve translation, and reduce development costs is discussed. This chapter also discusses the latest biomarker technologies in use to make this possible and details the next steps that must be undertaken to keep driving this process forwards.

**Key words** Pharmaceutical company, Drug discovery, Biomarker, Genomics, Transcriptomics, Proteomics, Metabolomics

---

### 1 Introduction

Pharmaceutical companies are under pressure to improve their returns on existing and novel drug discovery efforts. This is an almost impossible task, considering that the average drug costs approximately one billion US dollars to develop and takes 10–15 years from initial discovery to the marketing phase [1]. This problem is compounded by the fact that around 70% of drugs do not recover their research and development costs and approximately 90% fail to yield an adequate return on investment. In addition, fewer than one in ten new drugs entering clinical trials make it to the market and some of those that do make it experience withdrawal and/or litigation [2–5]. Therefore, in order for the pharmaceutical companies to survive, minimizing these risks has become one of the most important objectives in drug discovery projects in recent years. For example, there has been considerable effort aimed at establishing standard operating procedures to plot a course through these problems and to help meet the intimidating regulatory demands. But the regulatory agencies have not just been standing by idling watching. In order to assist pharmaceutical

companies in this process, they have encouraged the incorporation of biomarker-based tests into the drug discovery pipeline and the Food and Drug Administration (FDA) has initiated efforts to modernize and standardize all involved procedures to facilitate delivery of more effective and safer drugs [6]. The FDA has estimated even a 10% improvement in the ability to predict failure of drug before it enters the clinical trial phases could save as much as one hundred million US dollars in development costs per drug [7].

---

## 2 A Brief History of Failed Drugs

The need for biomarker tests to guide drug development is perhaps best seen by recent major failures in this process. Over the last two decades more than 30 drugs have been withdrawn, mainly as a result of hepatotoxic or cardiotoxic effects [8]. In 1997, the FDA recommended that the antihistamine drug Terfendine be withdrawn from the market due to an association with heart arrhythmia, which could increase the risk of heart attacks and death [9]. In the year 2000, the antidiabetic and anti-inflammatory drug Troglitazone was withdrawn due to reports of liver toxicity [10]. In 2001, the 3-hydroxy-3-methylglutaryl-coenzyme A reductase inhibitor Cerivastatin, developed to treat high cholesterol levels, was withdrawn due to an increased risk of rhabdomyolysis, a severe condition that causes muscle pain and weakness and can sometimes result in renal failure and death [11, 12]. In 2003, the antidepressant drug Nefazodone was withdrawn due to liver toxicity [13]. One of the most infamous cases was the withdrawal of the anti-inflammatory drug Vioxx by Merck, due to reports about its increased risk of heart attack and stroke [14]. Merck agreed to pay 4.85 billion US dollars in damages three years after the withdrawal and also had to pay a further 285 million US dollars four years later in the face of charges that it “duped customers” into buying the drug [15]. Another infamous case was the serious adverse effects seen in phase I clinical studies of the monoclonal antibody TGN1412, produced by TeGenero [16]. TGN1412 was originally tested as a potential treatment for B cell chronic lymphocytic leukemia and rheumatoid arthritis and had shown no toxic effects in preclinical studies. The compound was withdrawn in 2006 after six healthy male volunteers who took part in the phase I trial experienced the beginnings of a “cytokine storm” within 90 min after receiving it. The phrase cytokine storm describes a proinflammatory effect, resulting in fever, pain, and organ failure. All of the volunteers required weeks of hospitalization. This and the other cases stated above indicate that these disasters may not have occurred if procedures been adopted using safety biomarkers to guide dosing and/or predict toxicities at an early stage in the drug discovery process.

---

### 3 Biomarker Impact in the Drug Discovery Process

Estimates are that the total number of potential biomarkers is higher than one million, which is clearly an overwhelming number. However, many researchers and pharmaceutical companies have been investing in multiplex “omics” technologies to assist in sorting through this mass of analytes and help in understanding diseases at a deeper level than ever before. These platforms consist of genomics, transcriptomics, proteomics, and metabolomics, along with others. All of these approaches involve identification of molecular fingerprints from clinical samples and convert this into information about physiological status. With the help of these multiplex biomarker approaches, we are just now beginning to able to better categorize diseases at the molecular level, rather than on symptoms alone. By finding molecular biomarkers of a disease, early detection and diagnosis could be improved by simply testing for the presence of the disease fingerprint. Biomarkers would also assist pharmaceutical companies who could now look for drugs which help to normalize disease-like signatures. These could be used in early preclinical stages of drug development such as by looking at the effect of test compounds in disease models. They could also assist in looking for markers of toxicity prior to entry of the drug into clinical trials if representative models could be developed. In the later stages, biomarker tests could be used to help stratify patient groups in order to identify those who are most likely to benefit from treatment. This is critical as too many trials may have failed simply due to the fact that the wrong patients were included in the study. This alone could save millions in costs since the phase II and phase III stages of clinical trials are normally the most expensive steps in the drug discovery process.

---

### 4 Multiplex Biomarker Techniques

Biomarkers are physical characteristics that can be measured in biosamples and used as an indication of physiological states such as good health, disease, or toxicity or to predict or monitor drug response [17]. For practical purposes, it is becoming increasing important that biomarkers can be measured with high accuracy and reproducibility, within a short timeframe and at an affordable cost. A biomarker should also reflect the underlying nature of the disorder or condition under investigation, at least to some extent. Many types of biomarker tests have emerged such as DNA sequencing for genomic studies and DNA microarrays and quantitative polymerase chain reaction (qPCR) for transcriptomic analyses. The sections below focus on methods that deliver proteomic and metabolomic profiles or “fingerprints” in blood samples. Since

changes in physiological states are dynamic in nature, these are likely to introduce alterations in numerous proteins and metabolites that converge on similar pathways. Most researchers now consider proteomics and metabolomic methods to be among the most informative regarding physiological status, considering that proteins and metabolites actually carry out, respond to, or are reflective of most processes of the body. Furthermore, most of the drugs in use today are designed to turn on or turn off proteins such as receptors or enzymes, or induce metabolic changes which can be seen as turnover of various proteins and small molecules.

#### **4.1 Multiplex Immunoassay Analysis of Proteins and Metabolites**

Blood serum and plasma contain several vital bioactive and regulatory molecules, including hormones, growth factors, and cytokines. The problem is that most of these molecules are present only at exceeding low concentrations and, therefore, measurement of these requires highly sensitive detection methods. One way of achieving this is through the use of multiplex immunoassay approaches [18]. These assays can target both proteins and metabolites. They are constructed and carried out as follows: (1) micro-sphere are loaded with different ratios of red and infrared dyes to give unique fluorescent signatures; (2) specific capture antibodies are attached to the surface of micro-spheres with specific signatures; (3) The antibody–sphere conjugates are mixed together to form the multiplex; (4) the sample is added and the target molecules bind to their respective antibody–sphere conjugates; (5) fluorescently labeled detection antibodies are added in a mixture and each of these binds to their target molecules in a sandwich format; and (6) the samples are streamed though a reader and the micro-spheres analyzed by two lasers for identification and quantification of the analyte present. In this final step, the lasers identify which analytes are present by measuring the unique signature of each dye-loaded micro-sphere and determine the amount of analyte bound by measuring the fluorescence associated with the fluorescent tags on the secondary antibodies (this is proportional to the analyte concentration).

#### **4.2 Two-Dimensional Gel Electrophoresis of Proteins**

Two-dimensional gel electrophoresis (2DE) works as follows: (1) protein mixtures in bio-samples are first applied to a strip gel allowing their separation according to their isoelectric points (this is pH at which no net charge occurs on the protein) using isoelectric focusing; (2) next the proteins are separated according to their apparent molecular weights by application of the strip to the top of a sodium dodecyl sulfate–polyacrylamide gel and second electrophoresis step; and (3) the protein spots in the gels can be visualized with any number of stains (e.g., Coomassie Blue or Sypro Ruby) and then quantitated using an imaging software. The 2DE technique allows the study of many tissue types but there are some problems with analysis of blood serum or plasma samples. This

occurs mainly due to the fact that blood contains a massive concentration range of proteins spanning at least 14 orders of magnitude [19]. This means that very abundant proteins such as albumin and the immunoglobulin chains would appear as large blobs on the gels and eclipse less abundant proteins such as the cytokines. However, a key advantage of 2DE is that it can generate information on intact proteins including any effects on posttranslational modifications, such as phosphorylation or glycosylation changes. This is not as simple using other proteomic methods such as shotgun mass spectrometry (below).

### 4.3 Mass Spectrometry

Just about the time that the human genome project was ending, a revolution in shotgun mass spectrometry began as this was developed as a sensitive and medium throughput approach for proteomic biomarker identification [20]. The term “shotgun” derives from the fact that the protein in bio-samples are cleaved with proteolytic enzymes to generate smaller peptides (analogous to shotgun pellets), which are the actual analytes. This is performed as most intact proteins are too large and complex in their structure to be ionized or analyzed directly in a mass spectrometer. After proteolysis, the resulting peptides are separated according to physiochemical properties, such as hydrophobicity, using liquid chromatography so that they enter the mass spectrometer more or less one at a time. As the peptides enter the mass spectrometer, they are ionized by a process such as electro-spray, which is basically application of an electric charge to evaporate the fluids, leaving the peptides in a charged plasma state. After this, the peptide ions are accelerated by magnets in the mass spectrometer towards a detector at a rate that is inversely proportional to their mass over charge ratios ( $m/z$ ). Quantitation can be carried out since the amount of each peptide hitting the detector per second is directly proportional to the quantity of the peptide and, by derivation, to that of the corresponding parent protein. At the same time, the sequence of the peptide can be determined by streaming in a gas such as nitrogen, which breaks the peptides into smaller pieces. The mass of each piece can then be used to derive the amino acid sequences that make up the peptides and these are used to search protein databases for identification of the parent proteins. The main advantage of this method is the ability to detect more difficult classes of proteins which are not detectable by 2DE approaches. The disadvantages include the loss of intact protein information since the proteins are enzymatically digested prior to analysis. This method is also used for metabolomic analysis although there is no need for enzymatic cleavage of the molecules as metabolites are normally of a manageable size and structure. In this case, the sample can be infused directly into the mass spectrometer, the quantity calculated as described above and the identity determined by comparisons to known standards in metabolomic databases.

#### 4.4 $^1\text{H}$ (Proton)-Nuclear Magnetic Resonance (NMR) Spectroscopy

As with mass spectrometry, proton NMR spectroscopy can be used for metabolomic and small molecule analyses, although in this method no separation or pre-fractionation of the molecules is required. Major advantages of this approach include the points that the sample preparation step is direct and simple and that it is highly reproducible at the analytical level. One drawback is that it is less sensitive compared to the mass spectrometry-based metabolomic techniques [21, 22]. The  $^1\text{H}$ -NMR method can yield information about the structural properties of metabolites and is therefore highly used for identification purposes. This works since the method can track the behavior of protons as the nuclei of each proton on the molecule of interest lines up in a strong magnetic field. The procedure begins with the application of radio frequency pulses to the sample. This induces the nuclei to change their rotation away from their equilibrium position in line with the axis of the magnetic field and the frequency of rotation is directly related to its physiochemical environment within the parent metabolite. Therefore, by using different combinations of radio pulses, one can determine how each atom interacts with the other atoms, yielding the structure and, consequently, the metabolite identity. Proton NMR spectroscopy can be used to monitor relative changes in the levels of small molecules and metabolites such as amino acids, vitamins, neurotransmitters, and neurotransmitter precursors, making it useful for biomarker studies of multiple disorders.

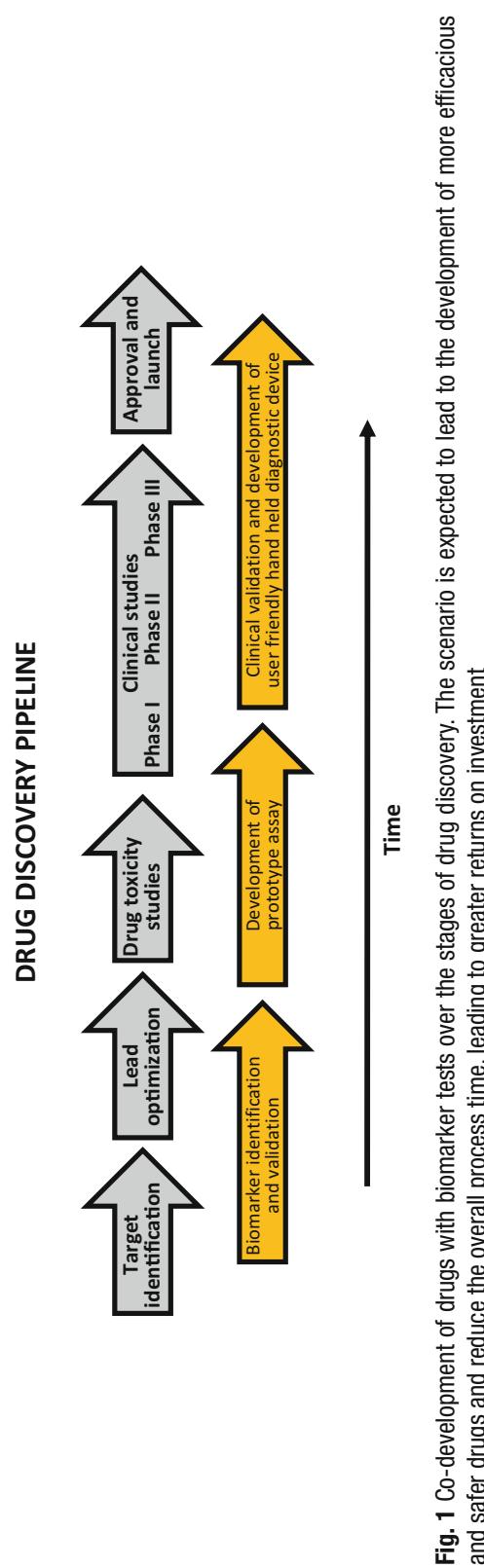
---

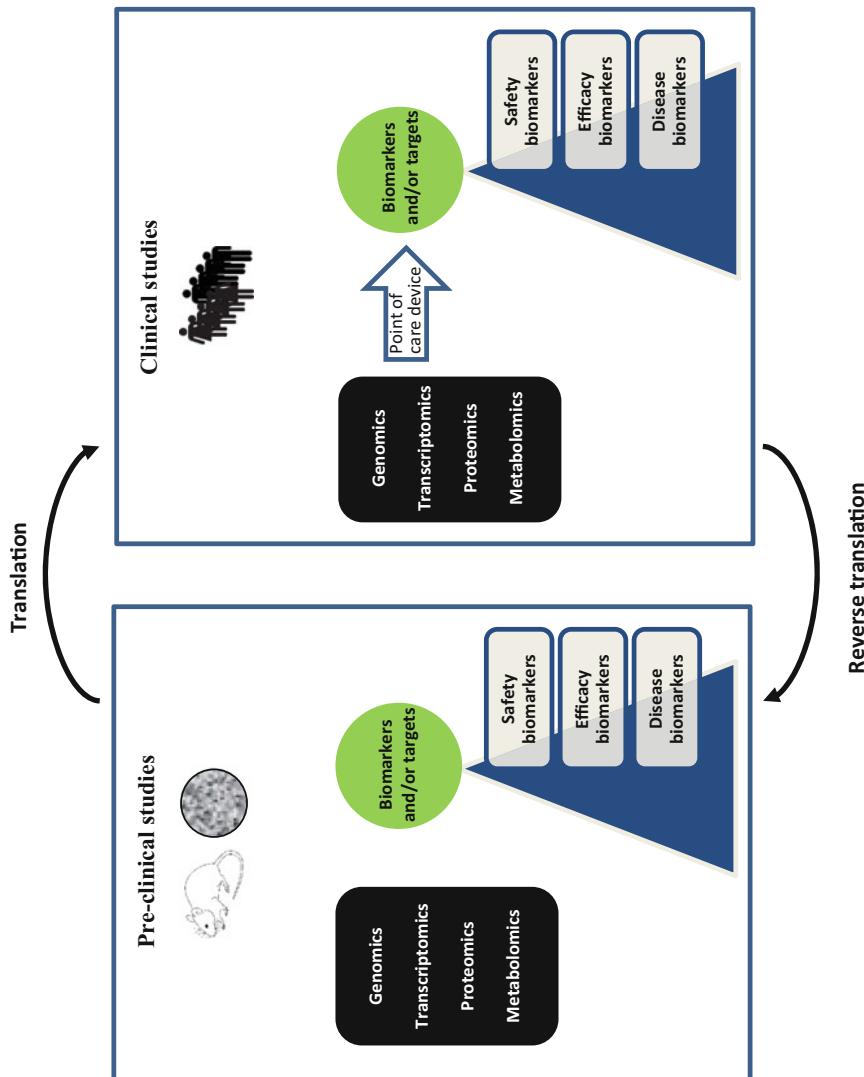
## 5 Use of Multiplex Biomarker Profiling in the Drug Discovery Process

Biomarker profiling can be used at multiple stages of drug discovery process as shown in Fig. 1 and described below in further detail. In the discovery phase, multiplex biomarker profiling could positively impact on target identification, target validation, lead compound prioritization, and efficacy screening of suitable pre-clinical models. In addition, applications in the development phase include the production of surrogate biomarkers for drug efficacy and for the validation of preclinical models of human diseases. Perhaps most importantly, any biomarker tests that arise from these earlier phase can be translated into user friendly point-of-care devices that can be used to identify disease signatures and for monitoring drug efficacy or toxicity in the clinical trial phases. In this way, mechanistic or targeted biomarkers can be used in pre-clinical or clinical development to validate the suitability of pre-clinical models and establish and facilitate translational medicine by providing pharmacological and biological biomarkers to predict clinical outcome (Fig. 2).

### 5.1 Target Validation

Most existing drug targets are components of a limited number of molecular networks that have been validated at the biological and





**Fig. 2** Translation and reverse translation of safety, efficacy, and disease biomarkers to characterize preclinical models and enable clinical development based on personalized medicine strategies

physiological levels [23]. However, many of these networks have not been fully elucidated in terms of the interacting transcript, protein and small molecule components and their relationship to other biological pathways. Therefore, there is a need for “mining” cell signaling and whole body networks further in order to identify novel tractable drug targets. Identification of molecules that operate as switch factors in the disease process is usually the first stage of target validation. This can be tested by manipulating the expression of the target molecules using gain or loss of function methods in an attempt to induce or reverse the disease phenotype [24]. Increasing the function of the molecule of interest could be achieved using agonist-type small molecules or by over-expression technologies. Alternatively, function could be knocked down using antagonist-like small molecules, ribozymes, small interfering RNAs, or genetic approaches. In each case, a multiplex molecular signature could be obtained for monitoring the resulting phenotype. Such approaches would give confidence that small molecules under drug development would have a similar effect and this would help to drive the project forward.

For successful validation and prioritization of novel drug targets, it is important to establish the molecular context or interaction pathways associated with potential drug targets. This involves understanding the disease at the functional level and confirming that the therapeutic concept works in preclinical models as well as in clinical proof-of-principle experiments. Genomic, transcriptomic, proteomic, and metabolomic profiling studies can provide this information by identifying components of cellular networks that could be targeted for possible therapeutic intervention. A single-cell transcriptomic profiling approach was used to validate the involvement of brown adipocyte tissue to protect against obesity and metabolic disease [25]. This study confirmed the presence of mRNAs encoding brown adipose tissue proteins such as uncoupling protein 1 and adrenergic receptor-beta 3 at both the mRNA and protein levels, and identified mRNAs encoding novel proteins such as orphan g-protein coupled receptors and other receptors regulated by neurotransmitters, cytokines and hormones. One study demonstrated that multiple methods are essential, including identification of cancer cell membrane proteins by mass spectrometry and phenotypic antibody screening, for the identification and validation of antibody tractable targets, which can significantly accelerate the therapeutic discovery process in cancer research [26]. In another investigation, a stable isotope-mass spectrometry metabolomic profiling approach was used to interrogate the mechanism of antibiotic action of d-cycloserine, a second line antibiotic used in the treatment of multidrug resistant *Mycobacterium tuberculosis* infections [27]. The authors used labeled <sup>13</sup>C α-carbon-<sup>2</sup>H-l-alanine for simultaneous tracking of alanine racemase and d-alanine:d-alanine ligase in *Mycobacterium tuberculosis* and found that the latter was more strongly inhibited than the former by d-cycloserine.

## 5.2 Lead Optimization

Many compounds fail in the later stages of drug development because of an unanticipated toxicity or poor efficacy [28]. This calls for a greater understanding of drug properties at an earlier stage in the development pipeline to help overcome these problems. One approach would be through the incorporation of appropriate multiplex biomarker tests into this stage of the pipeline. Such tests can be used to generate expression signatures from cells or tissues treated with new drugs for target identification and validation, and for delineating mechanism of action. Biomarker signatures can also be used in the identification and optimization of lead compounds by looking for correlations of specific molecular patterns with efficacy or specific toxicities. For example, monitoring the effects of developmental compounds on molecular patterns in the appropriate models might provide an early prediction of efficacy or toxicity [29]. Compounds which induce the same signature of protein expression changes are presumed to share the same mode of action and toxicity effects. Recently, Tang et al. reported on the development of a miniaturized Luminex assay consisting of a panel of multiplexed assays for measuring cytokines [30]. This assay facilitates high-throughput screening of compounds in cell models using cytokines as physiologically relevant molecular readouts. In addition, this multiplexed cytokine test can be used for profiling of bio-fluids such as blood serum and plasma for translational research. Cell models can provide useful screening platforms for drug profiling, using reporter systems for activation of receptor signaling or enzymatic cascades. This approach has been denoted as cyomics. By screening cell models with drug libraries, functional responses such as calcium flux, phosphorylation signaling cascades, mitochondrial membrane potential changes, receptor expression and/or internalization, ligand binding, apoptosis, oxidative stress, proliferation, and cell cycle status can be measured [31]. The ultimate aim is to use changes in molecular biomarker patterns to understand how drugs exert their effects at the molecular level.

## 5.3 Drug Toxicology Studies

Successful drugs should be potent, specific for their targets and bio-available with good pharmacokinetic profiles and low toxicity. In the ideal scenario, compounds lacking one or more of these traits should be identified during the early stages of the drug discovery pipeline so that only the most promising compounds are taken through to the clinical trial stages. However, toxicities usually become apparent only during the preclinical or clinical development stages when compound testing occurs in animal or cellular models or in humans. In some cases, toxicities may not even be detected until widespread distribution of the drug to the general public [32]. The reasons for this can be complex and on some occasions attributed to metabolism of the parent compound to toxic metabolites or to poor clearance. Multiplex omics methods

can accelerate development of the best lead compounds by facilitating screening methods for toxicity based signatures.

As a prime example, the EU Framework 6 Project: Predictive Toxicology (PredTox), studied the effects of 16 test compounds using both conventional toxicological parameters and multiplex biomarker approaches technologies [33]. They found three main classes of toxicity which were: (1) liver hypertrophy, (2) bile duct necrosis/cholestasis and (3) kidney proximal tubular damage. The results demonstrated that that the multiplex approaches can help drug companies to make better informed decisions during early phase toxicological studies. Toxicogenomics is the term applied to investigation of drug responses at gene expression level [34]. The liver is the main tissue targeted in this approach. DNA microarray profiling studies are now being carried out with known classes of toxicity inducing compounds with the objective that these can be referenced against novel compounds [35, 36]. In addition, the potential mechanisms of hepatotoxicity of doxorubicin-loaded microspheres in chemoembolization have been investigated by DNA microarray analyses combined with histological examinations [37]. This showed that doxorubicin caused lesions to the liver and disturbed liver metabolism-related enzymes. Another DNA microarray study investigated liver toxicity of rotodrine, a compound that has been used in preterm labor [38]. They found a specific increase in the levels of serum amyloid A, which was not induced by other hepatotoxic drugs like acetaminophen, valproic acid, or metformin. The increase in serum amyloid A was also more sensitive as a biomarker compared to the commonly measured liver enzymes aspartate aminotransferase and alanine aminotransferase. Accessible bio-fluids such as spittle, serum, plasma, and urine hold the most immediate promise for preclinical assessment in terms of better biomarkers. Andersson and colleagues analyzed plasma samples from 134 patients using proteomic and metabolomic approaches, with the aim of finding predictive biomarkers to explain the liver toxicity induced by ximelagatran, a compound developed for the prevention and treatment of thromboembolic conditions [39]. They found changes in 3-hydroxybutyrate, pyruvic acid, colony stimulating factor 1 receptor, l-glutamine, protein S, and alanine, as well as changes in other molecules. This approach helped to generate a new hypothesis for an unknown mechanism of toxicity.

In addition, cellular models could be useful in preclinical toxicity screening. Meneses-Lorente and coworkers used a two-dimensional differential in-gel electrophoresis and mass spectrometry profiling approach to identify a proteomic signature associated with hepatocellular steatosis in rats after dosing with a compound in pre-clinical development [40]. Within 6 h of dosing, livers showed hepatocellular vacuolation, which increased in extent and severity over time. Although alterations in alanine aminotransferase and

aspartate aminotransferase were not detected until day three, proteomic profiling changes were observed at the earliest time point and many were associated with liver steatosis. The proteins which showed increased levels were pyruvate dehydrogenase, phenylalanine hydroxylase and 2-oxoisovalerate dehydrogenase, which are all involved in acetyl-CoA production. One of the decreased proteins was sulfite oxidase, which is involved in triglyceride accumulation.

#### 5.4 Clinical Studies

Clinical applications of multiplex biomarker approaches include early detection of the disease using molecular signatures in biofluids as a complement to other measures carried out which specifically target the affected biological pathway in patients. Once a particular signature for compound efficacy has been established, this can be applied in a high throughput format such as screening with multiplexed immunoassays to help in the identification of compounds with optimum profiles. Also, prognostic biomarkers can be used to help predict drug efficacy in patients and to potentially identify which individuals are likely to benefit from treatment with a specific drug [41, 42]. Such approaches can help pave the way for development of more individualized therapies, using personalized medicine approaches [43]. The ultimate application of proteomics in drug discovery would be to identify biomarkers in a readily accessible body fluid, such as serum or plasma, and which can be correlated with the initiation of efficacy or severity of toxicity. Such biomarker signatures could also be used as surrogate markers or secondary endpoints to help predict the responses of individuals to treatment and allow adjustments of the therapy to achieve highest possible efficacy without reaching a level which elicits toxic side effects. Likewise, this approach could be used to facilitate identification of patient classes who will respond favorably to the drug in clinical trials as part of the personalized medicine strategy.

Kopetz et al. investigated the efficacy of fluorouracil (FU), leucovorin, irinotecan, and bevacizumab in a phase II trial in patients who were previously untreated for metastatic colorectal cancer, and measured changes in plasma cytokines and angiogenic factors as potential markers of treatment response using a multiplex immunoassay platform [44]. They found that elevated levels of interleukin 8 at baseline were associated with a shorter progression-free survival period and changes in basic fibroblast growth factor, hepatocyte growth factor, placental growth factor, stromal-derived factor-1, and macrophage chemoattractant protein-3 were associated with angiogenesis and myeloid recruitment in the progressive disease phase. Another clinical study investigated the effects of daily coffee consumption as potential risk factors for type II diabetes using gas chromatography mass spectrometry and multiplex immunoassay analysis approaches [45]. More recently,

analysis of the biomarker results from the AVADO phase III trial of first-line bevacizumab plus docetaxel for HER2-negative metastatic breast cancer showed that plasma levels of vascular endothelial growth factor (VEGF)-A and VEGF receptor-2 are potential predictive markers for bevacizumab efficacy [46].

---

## 6 Conclusions and Future Perspectives

This chapter describe the emerging use of multiplex biomarker profiling techniques as enabling platforms for use in all aspects of the drug discovery process. This is critical as current diagnostic procedures and strategies for developing novel medicines are in need of a paradigm change [47]. The regulatory health authorities now consider the incorporation of biomarkers into clinical platforms to be of high importance for the future of drug discovery and have introduced schemes to modernize methods, tools and techniques to achieve this goal.

Multiplex biomarker tests have now been available for more than two decades. In general most of the platforms have medium to large footprints and require expert technicians in order to operate them. Another drawback is that each of these methods has a typical turnaround time of approximately one day from the sample preparation stages to the final results presentation for a given sample. Within the last 5 years, multiplex biomarker tests have been miniaturized using microfluidic approaches to yield devices which are approximately the size of a small pamphlet or a credit card [48]. Most importantly, these devices are user friendly since no expertise is required for operation and the results can be returned in less than 15 min from a single drop of blood. There have also been recent developments which allow connection of these biomarker cards with smartphone technology using cleverly designed apps. For example, multiplex immunoassay-based tests have been developed on a handheld smartphone-based colorimetric reader using an optomechanical interface and this has been tested successfully in the case of detecting mumps, measles and herpes simplex I and II viruses [49]. It is not hard to imagine that similar tests for other diseases will be available in the not so distant future. Such devices would also meet the requirements of clinical studies and slot nicely into the pipeline in phase I–III clinical studies, considering their robustness, rapidity, and user friendly operation. This should help to inject renewed vigor into the pharmaceutical industry and most importantly help to improve the lives of the patients by enabling personalized medicine approaches.

## References

1. Hughes JP, Rees S, Kalindjian SB, Philpott KL (2011) Principles of early drug discovery. *Br J Pharmacol* 162:1239–1249
2. Giezen TJ, Mantel-Teeuwisse AK, Straus SM, Schellekens H, Leufkens HG, Egberts AC (2008) Safety-related regulatory actions for biologicals approved in the United States and the European Union. *JAMA* 300:1887–1896
3. McNaughton R, Huet G, Shakir S (2013) An investigation into drug products withdrawn from the EU market between 2002 and 2011 for safety reasons and the evidence used to support the decision-making. *BMJ Open* 4:e004221
4. Onakpoya IJ, Heneghan CJ, Aronson JK (2016) Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. *BMC Med* 14:10
5. Rawson NS (2016) Drug safety: withdrawn medications are only part of the picture. *BMC Med* 14:28
6. Ovens J (2006) Funding for accelerating drug development initiative critical. *Nat Rev Drug Discov* 5:271
7. <http://www.fda.gov/oc/initiatives/critical-path/whitepaper.html>
8. Need AC, Motulsky AG, Goldstein DB (2005) Priorities and standards in pharmacogenetic research. *Nat Genet* 37:671–681
9. [http://www.fda.gov/ohrms/dockets/ac/98/briefingbook/1998-3454B1\\_03\\_WL50.pdf](http://www.fda.gov/ohrms/dockets/ac/98/briefingbook/1998-3454B1_03_WL50.pdf)
10. Cohen JS (2006) Risks of troglitazone apparent before approval in USA. *Diabetologia* 49:1454–5145
11. Onakpoya IJ, Heneghan CJ, Aronson JK (2016) Worldwide withdrawal of medicinal products because of adverse drug reactions: a systematic review and analysis. *Crit Rev Toxicol* 3:1–13 [Epub ahead of print]
12. Marciante KD, Durda JP, Heckbert SR, Lumley T, Rice K, McKnight B et al (2011) Cerivastatin, genetic variants, and the risk of rhabdomyolysis. *Pharmacogenet Genomics* 21:280–288
13. Choi S (2003) Nefazodone (Serzone) withdrawn because of hepatotoxicity. *CMAJ* 169:1187
14. Dogné JM, Hanson J, Supuran C, Pratico D (2006) Coxibs and cardiovascular side-effects: from light to shadow. *Curr Pharm Des* 12:971–975
15. <http://www.bloomberg.com/news/articles/2013-07-18/merrick-pays-23-million-to-end-vioxx-drug-purchase-suits>
16. Sheridan C (2006) TeGenero fiasco prompts regulatory rethink. *Nat Biotechnol* 24:475–476
17. Atkinson AJ, Colburn WA, DeGruttola VG, DeMets DL, Downing GJ, Hoth DF, Biomarkers Definitions Working Group et al (2001) Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69:89–95
18. Fulton RJ, McDade RL, Smith PL, Kienker LJ, Kettman JR Jr (1997) Advanced multiplexed analysis with the FlowMetrix system. *Clin Chem* 43:1749–1756
19. Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 1:845–867
20. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR et al (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 17:676–682
21. Griffin JL (2003) Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol* 7:648–654
22. Beckonert O, Keun HC, Ebbels TM, Bundy J, Holmes E, Lindon JC et al (2007) Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc* 2:2692–2703
23. Imming P, Sinning C, Meyer A (2006) Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* 5:821–834
24. Sioud M (2007) Main approaches to target discovery and validation. *Methods Mol Biol* 360:1–12
25. Spaethling JM, Sanchez-Alavez M, Lee J, Xia FC, Dueck H, Wang W (2016) Single-cell transcriptomics and functional target validation of brown adipocytes show their complex roles in metabolic homeostasis. *FASEB J* 30:81–92
26. Rust S, Guillard S, Sachsenmeier K, Hay C, Davidson M, Karlsson A et al (2013) Combining phenotypic and proteomic approaches to identify membrane targets in a ‘triple negative’ breast cancer cell type. *Mol Cancer* 12:11
27. Prosser GA, de Carvalho LP (2013) Metabolomics reveal d-alanine:d-alanine ligase as the target of d-cycloserine in *Mycobacterium tuberculosis*. *ACS Med Chem Lett* 4:1233–1237
28. Amacher DE (2010) The discovery and development of proteomic safety biomarkers for the detection of drug-induced liver toxicity. *Toxicol Appl Pharmacol* 245:134–142

29. Meneses-Lorente G, Watt A, Salim K, Gaskell SJ, Muniappa N, Lawrence J et al (2006) Identification of early proteomic markers for hepatic steatosis. *Chem Res Toxicol* 19:986–998
30. Tang H, Panemangalore R, Yarde M, Zhang L, Cvijic ME (2016) 384-well multiplexed luminex cytokine assays for lead optimization. *J Biomol Screen.* pii: 1087057116644164. (Epub ahead of print)
31. Valet G (2006) Cytomics as a new potential for drug discovery. *Drug Discov Today* 11:785–791
32. Gale EA (2001) Lessons from the glitazones: a story of drug development. *Lancet* 357:1870–1875
33. Suter L, Schroeder S, Meyer K, Gautier JC, Amberg A, Wendt M et al (2010) EU framework 6 project: predictive toxicology (PredTox)--overview and outcome. *Toxicol Appl Pharmacol* 252:73–84
34. Afshari CA, Hamadeh HK, Bushel PR (2011) The evolution of bioinformatics in toxicology: advancing toxicogenomics. *Toxicol Sci* 120(S1):S225–S237
35. Bulera SJ, Eddy SM, Ferguson E, Jatkoe TA, Reindel JF, Bleavins MR et al (2001) RNA expression in the early characterization of hepatotoxins in Wistar rats by high-density DNA microarrays. *Hepatology* 33:1239–1258
36. Waring JF, Ciurlionis R, Jolly RA, Heindel M, Ulrich RG (2001) Microarray analysis of hepatotoxins in vitro reveals a correlation between geneexpression profiles and mechanisms of toxicity. *Toxicol Lett* 120:359–368
37. Verret V, Namur J, Ghegedian SH, Wassef M, Moine L, Bonneau M et al (2013) Toxicity of doxorubicin on pig liver after chemoembolization with doxorubicin-loaded microspheres: a pilot DNA-microarrays and histology study. *Cardiovasc Intervent Radiol* 36:204–312
38. Tsuchiya H, Sato J, Tsuda H, Fujiwara Y, Yamada T, Fujimura A et al (2013) Serum amyloid A upsurge precedes standard biomarkers of hepatotoxicity in ritodrine-injected mice. *Toxicology* 305:79–88
39. Andersson U, Lindberg J, Wang S, Balasubramanian R, Marcusson-Ståhl M, Hannula M et al (2009) A systems biology approach to understanding elevated serum alanine transaminase levels in a clinical trial with ximelagatran. *Biomarkers* 14:572–586
40. Meneses-Lorente G, Guest PC, Lawrence J, Muniappa N, Knowles MR, Skynner HA et al (2004) A proteomic investigation of drug-induced steatosis in rat liver. *Chem Res Toxicol* 17:605–612
41. Killestein J, Polman CH (2011) Determinants of interferon  $\beta$  efficacy in patients with multiple sclerosis. *Nat Rev Neurol* 7:221–228
42. Gerger A, Labonte M, Lenz HJ (2011) Molecular predictors of response to antiangiogenesis therapies. *Cancer J* 17:134–141
43. Flood DG, Marek GJ, Williams M (2011) Developing predictive CSF biomarkers-A challenge critical to success in Alzheimer's disease and neuropsychiatric translational medicine. *Biochem Pharmacol* 81:1422–1434
44. Kopetz S, Hoff PM, Morris JS, Wolff RA, Eng C, Glover KY et al (2010) Phase II trial of infusional fluorouracil, irinotecan, and bevacizumab for metastatic colorectal cancer: efficacy and circulating angiogenic biomarkers associated with therapeutic resistance. *J Clin Oncol* 28:453–459
45. Kempf K, Herder C, Erlund I, Kolb H, Martin S, Carstensen M et al (2010) Effects of coffee consumption on subclinical inflammation and other risk factors for type 2 diabetes: a clinical trial. *Am J Clin Nutr* 91:950–907
46. Miles DW, de Haas SL, Dirix LY, Romieu G, Chan A, Pivot X et al (2013) Biomarker results from the AVADO phase 3 trial of first-line bevacizumab plus docetaxel for HER2-negative metastatic breast cancer. *Br J Cancer* 108:1052–1060
47. Poste G (2011) Bring on the biomarkers. *Nature* 469:156–157
48. Schumacher S, Nestler J, Otto T, Wegener M, Ehrentreich-Förster E, Michel D et al (2012) Highly-integrated lab-on-chip system for point-of-care multiparameter analysis. *Lab Chip* 12:464–473
49. Berg B, Cortazar B, Tseng D, Ozkan H, Feng S, Wei Q et al (2015) Cellphone-based handheld micro-plate reader for point-of-care testing of enzyme-linked immunosorbent assays. *ACS Nano* 9:7857–7866

# **Chapter 2**

## **The Application of Multiplex Biomarker Techniques for Improved Stratification and Treatment of Schizophrenia Patients**

**Johann Steiner, Paul C. Guest, Hassan Rahmoune, and Daniel Martins-de-Souza**

### **Abstract**

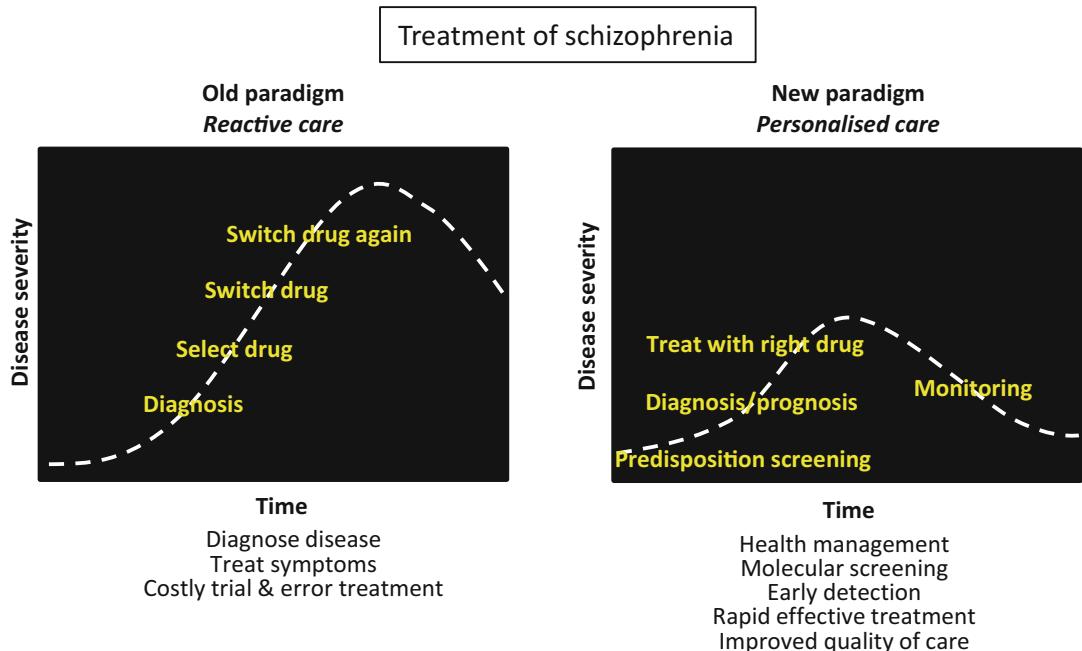
In the case of major psychiatric disorders such as schizophrenia, shortcomings in the conversion of scientific discoveries into newer and safer treatment options has led to a loss of confidence and precipitated a crisis for large pharmaceutical companies. This chapter describes how incorporation of multiplex biomarker approaches into the clinical pipeline can lead to better patient characterization, delivery of novel treatment approaches and help to renew efforts in this important area. The development of specific biomarker test panels for disease prediction should facilitate early intervention strategies, which may help to slow disease development or progression. Furthermore, the development of such tests using lab-on-a-chip and smartphone platforms will help to shift diagnosis and treatment of this major disorder into a point-of-care setting for improved patient outcomes.

**Key words** Schizophrenia, Blood-based biomarkers, Proteomics, Multiplex immunoassay, Mass spectrometry, Point-of-care, Lab-on-a-chip, Smartphone apps

---

### **1 Introduction**

Schizophrenia is a debilitating, mental health disorder which can strike individuals in their late teens or early adulthood and seriously impair medical health, quality of life, social well-being and productivity [1]. Clinical presentation usually occurs with symptoms such as hallucinations, delusions, anhedonia, social retreat, disorganized thinking and cognition impairment. At present, diagnosis is still based on expression of symptoms and is dependent on communications between the affected individual and the attending clinician or psychiatrist. This is usually achieved in an interview-like format using the Diagnostic and Statistical Manual of Mental Disorders (DSM) [2] or the International Classification of Diseases (ICD-10) [3] criteria as guidelines. However, these texts can only detail



**Fig. 1** Comparison of the old and new treatment paradigms involving schizophrenia patients, distinguished by the use of biomarkers for improved stratification

the symptoms of schizophrenia without pointing to the underlying molecular physiological pathways that may be affected. Furthermore, classification of a person as having schizophrenia can be confounded by the fact that individuals with other psychiatric disorders can share many of the same symptoms. For this reason, there are now concerted efforts to identify specific multiplex biomarker fingerprints that can potentially predict the onset of schizophrenia, improve diagnostic accuracy, monitor disease progression, and guide treatment options. The availability of such tests for use in blood serum or plasma would be ideal as this would facilitate use in clinical settings. This is because blood-based biomarkers would have high accessibility in clinical practice due to the low invasiveness of the sampling procedure and the low associated costs.

The application of biomarker-based diagnostic tests that can accurately classify patients according to the type of disorder or even disease subtype will help to reduce duration of untreated mental illness and improve individual responses by placing the right patients on the right treatments as early as possible. This is because there is a direct correlation between longer periods without treatment and poor outcomes [4]. It is thought that this will change the overall paradigm from reactive psychiatric care to a more optimized personalized treatment approach in the field of psychiatry as well as in other areas of medicine (Fig. 1). Also, implementation of earlier effective treatment should help to reduce patient referral to

secondary services such as hospitals, community groups, and crisis teams. Any reduction in the use of these expensive services will help to reduce the overall financial burden of psychiatric disorders, which surpassed 60 billion dollars per year in the 1990s in the USA alone [5]. More importantly, an early successful intervention will help to curb symptom severity. This is because schizophrenia may lead to decades of life disability [6], more than double that of cardiovascular disorders [7].

The discovery of validated biomarker tests that reflect the correlation between the patient clinical and molecular readouts, would also enhance future mental healthcare significantly if the resulting tests can be incorporated into standard operating systems and clinical decision making, as well as being deployed as fast, cost-effective, user friendly, point-of-care devices. The strictest classification of newly developed biomarker tests requires that results must be replicated in different laboratories and in different sites. In the case of psychiatric disorders like schizophrenia, this will be difficult to achieve. The major reason for this is that these conditions are poorly understood at the molecular level and there is high heterogeneity in the way that they are manifested in the affected persons [8]. In this chapter, we discuss the challenges and requirements of developing and rolling out molecular biomarker tests for schizophrenia. In addition, the chapter focuses on the use of biomarkers for improved classification and management of patients with schizophrenia for improved point-of-care treatment and as a means of rekindling drug discovery efforts within the pharmaceutical industry in the area of psychiatric disorders.

---

## 2 Current Diagnosis of Schizophrenia

Most psychiatrists and clinicians agree that schizophrenia is a general term for a mixture of mental conditions that present with similar symptoms, in the same way that most people with acute infectious disorders present with an elevated body temperature [8]. It should be emphasized that the enormous variety of psychopathological alterations summarized under the term “schizophrenia” do not represent a disease entity, but rather a hypothetical construct that was created many decades ago by leading authorities in the field and is now defined by international classifications committees, which have inclusion criteria that are changed from issue to issue. However, this crossover can lead to misdiagnosis in psychiatric practice. As an example, one investigation found that more than 30% of patients who actually had bipolar disorder were initially diagnosed as having schizophrenia [9]. Another study challenged the basis of the current classifications systems by pointing out that there are no current methods to validate the basic concepts which are independent of the same concepts [10]. In any event, psychiatrists do not always use these classification systems

for making a diagnosis. In many cases, diagnosis may be made based on experience and personal views in a more heuristic manner. Again, this is not ideal as it can result in errors based on misconceptions, biases or selective memories.

Aside from these issues, the DSM and ICD-10 classification systems work based on the framework that mental disorders such as schizophrenia are distinct diseases with common etiologies which can be defined by criteria based on signs and symptoms. In reality, it is often not the case that specific symptoms are linked to defined diseases. For example, individuals with traumatic disorders, infectious diseases, metabolic conditions or even those under the influence of certain substances can present with symptoms that occur in schizophrenia [11, 12]. In addition, it is not uncommon for a diagnosis to change over time. A long-term study found significant changes in diagnosis from major depressive disorder to bipolar disorder and schizophrenia [13] and another found that only half of the patients stayed on their initial diagnosis [14].

## **2.1 The Importance of Early Diagnosis**

The concordance rate for a diagnosis of schizophrenia in identical twins ranges from 10 to 70% [15–17]. Although this provides evidence that there can be a genetic predisposition for schizophrenia, it also indicates that an individual will not necessarily develop schizophrenia even when a potential genetic effect is present. In fact environmental and other nongenetic factors are also important. Factors which could precipitate schizophrenia include pregnancy or delivery complications, such as infections, hypoxia or malnutrition [18, 19], as well as nonbiological factors, including social stressors such as experiencing a natural disaster, loss of a family member, or the chronic experiences of an unbearable environment such as an intolerable work situation, a dysfunctional family life, or an abusive relationship [20]. On a positive note, the presence of an environmental component also suggests that disease prevention or minimization might be possible if the responsible factors can be identified and avoided.

It is not difficult to imagine that certain environmental factors such as poor nutrition, social stress, and physical trauma can affect a person's physiological state. Several research groups have now shown that metabolic abnormalities such as insulin resistance occur in 20–50% of schizophrenia patients at their first clinical presentation [21–23]. Furthermore, multiple research groups have found alterations in circulating inflammatory and immune response markers in first onset schizophrenia patients [24, 25]. Two studies have now shown that such changes can occur months to years before full clinical manifestation of schizophrenia symptoms, suggesting that perturbations in these molecular pathways may play a role in the disease etiology [26, 27]. This also gives hope for identifying individuals at risk of developing the disease at the earliest

possible phase, as described above. This is important as numerous reports have now described importance of early intervention therapeutics for individuals at high risk of developing schizophrenia [28–30]. Any delay in diagnosis can have detrimental effects on the lives of the patients, such as the patient experiencing a full blown psychosis leading to other problems including substance abuse, alienation from family and friends, increased accidents, and the potential of self-harm [31, 32]. There is the problem of misdiagnosis which can lead to inappropriate treatments, which can either be ineffective or even harmful to the patient. In addition, misdiagnosis followed by inappropriate treatment can have a number of socioeconomic consequences, such as inflated medical costs, work absence, and harmful effects on family and relationships [33].

## **2.2 The Development of Biomarker Assays for Diagnosis of Schizophrenia**

The European health authorities have lent support to the development and implementation of biomarkers through agencies such as the Innovative Medicines Initiative [34, 35]. This began as a partnership between the European Commission and the major pharmaceutical companies with the overall objective of promoting more efficient discovery and development of better medicines. A key objective is the discovery of translational biomarkers which, in this case, means incorporating them into drug discovery pipelines for use in human clinical studies. The European Commission contributed one billion Euros to this project and this has been matched in kind by contributions from the participating companies.

Diagnostic biomarker tests in the USA are regulated by the Clinical Laboratory Improved Amendments (CLIA) agency [36]. These imposed regulatory standards govern any tests that are performed in a clinical setting on human samples for the purpose of diagnosis, disease prevention, treatment or assessment of health. Commercially available tests marketed under CLIA are categorized by the FDA depending on the potential risks for health. The development of diagnostic biomarker tests for all diseases including psychiatric disorders requires repeated demonstrations of precise performance characteristics including scores such as sensitivity and especially specificity, given the symptomatic and molecular overlap among all psychiatric disorders. This is an absolute requirement since biomarker measurements can be affected by many factors including biological, ethnicity, gender, environmental, sample collection, and analytical variables. For example, development of multiplexed immunoassays requires the testing and validation of each component immunoassay as well as the combination of assays used in each multiplex to maximize repeatability, precision and accuracy. This includes selection and immobilization of capture ligands on microbeads, calibration steps, testing for reagent–antibody compatibility, and ensuring each individual assay has sufficient dynamic range and the required limits of detection [37].

Another criterion of biomarker tests that is often overlooked is that they must be in a format that is high throughput, accurate and user friendly to allow use by clinicians, hospital staff and scientists. Along these lines, we suggest that the discovery and implementation platforms should be different to maximize development of tests with the highest performance. Thus, although they are powerful discovery tools, mass spectrometry and two-dimensional gel electrophoresis techniques may be too cumbersome in their larger formats and may require a high level of expertise to be considered as realistic options for clinical use. In contrast, an automated platform based on multiplexed immunoassay is a more likely candidate as a clinically friendly platform as it has already shown some promise in this area. However, even this would be too slow in its discovery format.

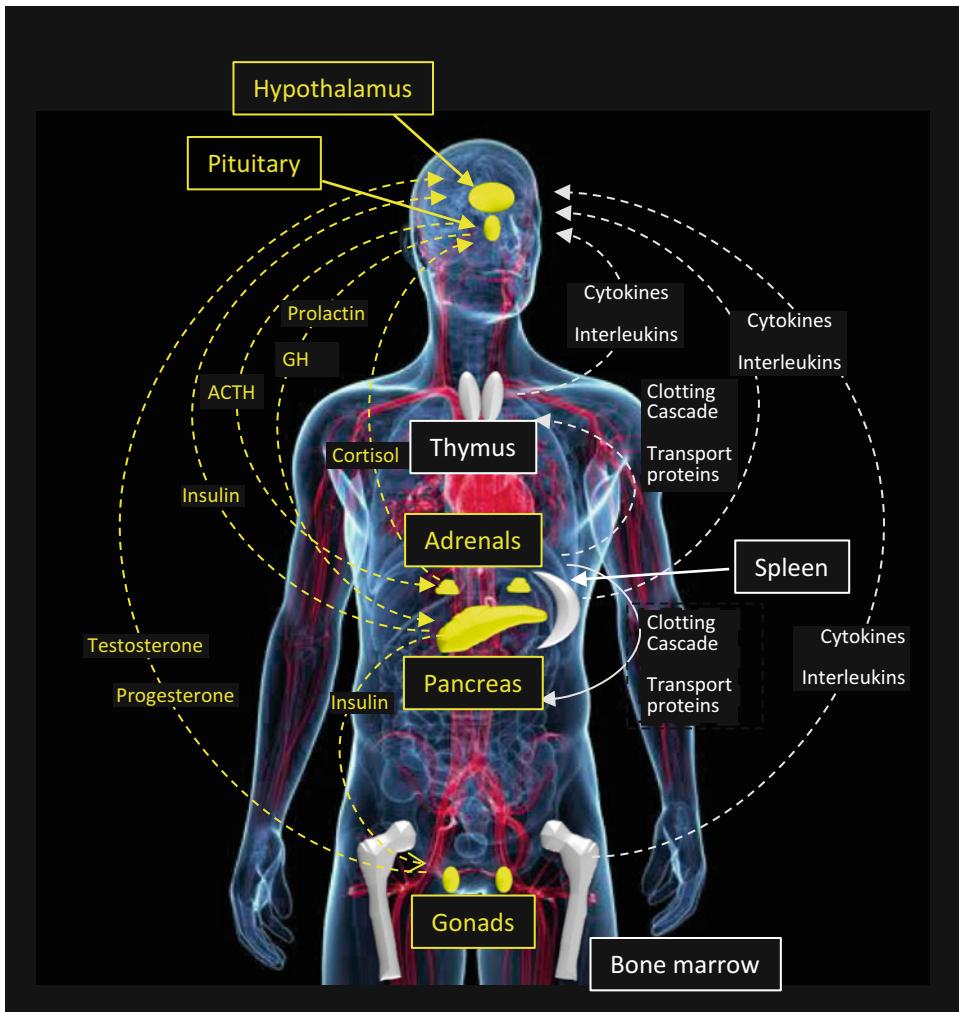
---

### 3 Biomarkers Identified for Schizophrenia

Although genomic studies are able to identify genes conferring susceptibility to a particular disease, the functional abnormalities of most disorders are reflected ultimately in the proteome and metabolome. This is because proteins and metabolites represent the molecular phenotype of a disease in parallel with the clinical manifestation. Recent years have seen the increasing use of proteomics as a tool for the discovery of biomarkers for diagnosis, monitoring disease progression, treatment response and for the identification of novel therapeutic targets. It is also important to remember that analysis of central nervous system (CNS) disorders is difficult as the brain is not readily accessible for invasive diagnostic purposes. Thus, sources such as serum and plasma have been undergoing increasing scrutiny as they have a higher utility in the clinic.

#### 3.1 *Inflammation Biomarkers in Schizophrenia*

A multiplex immunoassay profiling study which used cytokine arrays identified increased levels of interleukin (IL)-1 $\beta$  in cerebrospinal fluid from first onset schizophrenia patients, suggesting that the inflammation response may be perturbed in the brains of some patients [38]. This is consistent with other studies which demonstrated that brain development can be disturbed by changes in the balance of pro-inflammatory and anti-inflammatory cytokines [39, 40]. In addition, altered inflammation has been linked to changes in the glutamate system, the main excitatory neurotransmitter in the brain. Transcriptomic and proteomic profiling studies of post mortem brains from schizophrenia patients have identified increased levels of inflammation-related gene products in oligodendrocytes and endothelial cells in comparison to non-psychiatric control subjects [41, 42]. However, it is possible that this is a confounding factor of prolonged drug treatment or an unhealthy lifestyle, as often occurs in the chronic or latter stages of individuals suffering from this disorder (Fig. 2) [43].



**Fig. 2** Peripheral and central signaling molecules affected in schizophrenia with a focus on inflammation (white) and hormonal/metabolic (yellow) pathways. The dashed arrows indicate connections via the bloodstream. ACTH = adrenocorticotropic hormone; GH = growth hormone. Note that components of the interleukins, cytokines, transport proteins, and clotting cascade are not listed individually for presentation reasons. See text for more detail

The finding of circulating changes in molecules such as inflammatory factors in psychiatric disorders like schizophrenia is what makes biomarker testing for psychiatric disorders feasible [44, 45]. In addition, these factors may be informative as either trait or state biomarkers. A meta analysis of circulating inflammation-related changes in schizophrenia patients showed that cytokines including IL-12, soluble IL-2 receptor, interferon- $\gamma$ , and tumor necrosis factor- $\alpha$  may be useful as trait biomarkers, giving a stable indication that the disease is present [44]. In contrast, cytokines such as IL-1 $\beta$ , IL-6, and transforming growth factor- $\beta$  may represent state

biomarkers, which means that these may be used as readouts for acute changes in the disease. In addition, there have been many reports on the discovery of blood-based biomarker signatures comprised of a large number of inflammation-related proteins including some components of the clotting cascade and transport proteins in first onset schizophrenia patients [46, 47].

Inflammation in the periphery can affect brain function through effects on the hypothalamic–pituitary–adrenal (HPA) axis (see below) [48, 49]. Activation of inflammatory pathways stimulates release of corticotrophin releasing factor from the hypothalamic region of the brain and this initiates a cycle causing adrenocorticotrophic hormone (ACTH) to be released from the pituitary, which in turn drives cortisol release from the adrenal cortex [50]. Along with several other effects in the body, cortisol also exerts a negative feedback control on the HPA axis by binding to specific receptors in the brain and pituitary [51]. The link to psychiatric disorders comes from the fact that the HPA cycle also alters neurotransmitter systems throughout the brain, which are involved in regulation of mood and behavior. Given this link, it is not surprising that some investigators have tested the use of anti-inflammatory drugs such as aspirin or cyclooxygenase-2 inhibitors in combination with traditional antipsychotics as a novel treatment approach to relieve some symptoms of schizophrenia, with some success [52–55]. However, these findings require validation in further studies involving larger cohorts.

### **3.2 Neuroendocrine-Related Biomarkers**

Several studies have now shown effects on a number of hormonal systems related to metabolic homeostasis in schizophrenia. A number of studies over the past decade have shown that impaired fasting glucose tolerance, high insulin levels and insulin resistance occurs in both first onset [21, 22] and chronic schizophrenia patients [55, 57], as can occur in type 2 diabetes patients. One study showed the presence of hepatic insulin resistance in schizophrenia patients using a hyperinsulinemic clamp [58]. In terms of biomarkers, two studies found that first onset schizophrenia patients had increased levels of circulating insulinrelated peptides and high levels of chromogranin A, pancreatic polypeptide, prolactin, progesterone and cortisol, along with lower levels of growth hormone, in comparison to controls [23, 59]. This indicated altered secretion from several neuroendocrine glands including pancreatic  $\beta$  cells, pancreatic PP cells, the anterior pituitary, the sex organs and adrenal glands (Fig. 2). This could have important implications since chronically high insulin levels can have disruptive effects on brain function such as inducing increased brain inflammation, aberrant phosphorylation of filamentous structural proteins and increased deposition of amyloid plaques [60–62]. High insulin levels have also been found to lead to altered function of neurotransmitter pathways [63] and perturb synaptic plasticity

in brain regions such as the hippocampus [64]. The increased cortisol secretion is indicative of an activation of the HPA axis, as described above, which has been identified as a risk factor for schizophrenia in adolescents [65]. Another study showed gender-specific changes in the sex hormones estradiol and testosterone in schizophrenia patients, suggesting effects on the gonadal tissues [66]. More recent studies found decreased serum levels of thyroxine, triiodothyronine, and thyroid-stimulating hormone in schizophrenia patients [67], which may be tied in with the metabolism-related hormone changes described above. Another factor to consider is that many hormones are influenced by circadian rhythms and it is likely that some of those described above are co-regulated as part of an oscillatory feedforward–feedback mechanism between pancreatic islet cells, the pituitary and other components of the diffuse neuroendocrine system. For example, high insulin secretion has been associated with increased prolactin levels [68] and disrupted pulsatile release of growth hormone [69].

The repeated finding that hyperinsulinemia occurs in some first onset schizophrenia patients suggests that drugs which alleviate insulin resistance may offer a novel treatment approach. Furthermore, chronically treated patients can also exhibit high insulin levels since antipsychotic drugs can induce metabolic side effects such as insulin resistance and weight gain. Interestingly, the weight gain appears to be linked to antipsychotic therapeutic efficacy. One investigation showed that changes in body weight, blood glucose, and leptin levels were associated with improvement in both positive and negative symptoms of schizophrenia [70]. However these effects may not be an absolute requirement for improvement as studies that used the insulinsensitizing agents metformin and rosiglitazone to treat the antipsychotic-induced insulin resistance did so without disrupting the psycho-therapeutic benefits [71]. Therefore, the relationship between metabolism and psychiatric symptoms requires further scrutiny. It is possible that insulin-sensitizing agents may have a direct effect on alleviating some symptoms, such as the cognitive deficits. In support of this possibility, one study found that patients with mild Alzheimer's disease who were given pioglitazone showed improvements in cognition along with increased regional cerebral blood flow [72].

Drugs which target other hormone systems have also been tested as a novel means of treating schizophrenia symptoms. Dehydroepiandrosterone (DHEA), an adrenal steroid-like compound, has been tested as a potential add-on therapeutic with antipsychotics and this led to improvements in depression and anxiety symptoms in some schizophrenia patients [73]. Furthermore, treatment with the selective estrogen receptor modulator raloxifene resulted in reduced negative symptoms in postmenopausal females with schizophrenia compared to controls [74].

### ***3.3 Biomarkers for Prediction of Treatment Response***

Biomarker tests that can be used for better classification of schizophrenia patients opens up the possibility of better treatment options. For example, biomarkers that can be used to predict response of schizophrenia patients to treatment would be an important step forward for the well-being of the patients and it will assist the prescribing physicians, as well as pharmaceutical companies conducting clinical trials. Genetic studies have shown that polymorphisms in the histamine 2 receptor (*HRH2*) gene can be used to predict response to clozapine treatment in 76 % of schizophrenia cases [75]. Other studies have shown that variants in genes for dopamine receptors, serotonin receptors and enzymes involved in drug metabolism or neurotransmitter turnover can have influence of patient response to treatment including the propensity to develop certain side effects [76]. Another way of predicting response is through the use of physiometric measurements such as waist circumference, adiposity, body mass index (BMI), which have already been used to predict the development of side effects such as metabolic syndrome or other insulin resistance with good sensitivity and specificity [77, 78]. As for blood-based proteomic biomarkers, one study showed that schizophrenia patients with higher levels of serum prolactin have a better outcome following 5 years of antipsychotic treatment [79]. Two multiplex immunoassay serum profiling studies found that the levels of insulin were predictive of improvement in negative symptoms [80] and those of specific apolipoproteins, growth factors, hormones and interleukins could be used to predict weight gain [81] in first-onset schizophrenia patients after 6 weeks of antipsychotic treatment (Table 1). Another study showed that the levels of fatty acid binding protein could be used to predict response to olanzapine [82]. It should be kept in mind that these three investigations involved study of first or recent onset patients and biomarker profiles may be different for more chronic patients. Further studies aimed at retesting these prototype biomarker panels may lead to development of validated molecular tests that can be used to identify those patients who are more likely to respond to particular antipsychotic medications as well as those who are likely to benefit from add-on compound that target either the inflammatory or metabolic symptoms. This could also lead to the opportunity for clinicians to take actions such as patient assessment, counseling, or even readjusting treatments in accordance with measured biomarker readouts.

---

## **4 Point-of-Care Methods for Use in Schizophrenia**

For decades psychiatrists have acted on the assumption that psychiatric disorders such as schizophrenia are caused by defects in brain. However, developments over recent years have resulted in a new concept involving the whole body in the precipitation and

**Table 1**

**Significant associations between molecular levels at baseline and changes in either (A) psychopathology scores (positive and negative syndrome scale—PANSS) or (B) body mass indices (BMI) after 6 weeks of antipsychotic treatment.** *R*=Spearman correlation coefficient. *NS*=not significant. *ANCOVA*=analysis of covariance [80, 81]

| <b>A</b>          |         |   |                   |         |       |
|-------------------|---------|---|-------------------|---------|-------|
| Positive symptoms |         |   | Negative symptoms |         |       |
| Protein           | P-value | R | Protein           | P-value | R     |
| Insulin           | NS      | – |                   | 0.005   | –0.37 |

| <b>B</b>                     |        |       |
|------------------------------|--------|-------|
| Protein                      | ANCOVA | R     |
| Apolipoprotein CIII          | 0.019  | –0.33 |
| Apolipoprotein H             | 0.005  | –0.33 |
| Epidermal growth factor      | 0.025  | –0.28 |
| Follicle-stimulating hormone | 0.043  | –0.28 |
| Interleukin 18               | 0.015  | 0.24  |
| Interleukin 25               | 0.024  | –0.26 |
| Interleukin 6 receptor       | 0.031  | –0.30 |
| Matrix metalloproteinase 1   | 0.011  | –0.24 |
| Placental growth factor      | 0.016  | –0.24 |
| Thyroid-stimulating hormone  | 0.026  | –0.23 |

progression of these conditions. This is because the brain is holistically integrated in most fundamental biological functions of the body and therefore the functioning of this organ can be monitored by examining changes in the molecular composition of the blood. This is the basis for the use of blood serum or plasma in the study of psychiatric diseases [46, 47, 83]. This is useful since blood can be taken from living patients at different stages of the disease or throughout a treatment course. In the foreseeable future, it is likely that increased biomarker testing by clinicians will lead to more extensive “bio-” signatures in individuals that reflect the physiological status occurring in health or disease. Blood serum and plasma samples contain many molecules such as hormones, growth factors and cytokines which can only be detected using methods

that are highly sensitive. One of the best methods to achieve this is the sandwich format of immunoassay [84, 85] and this is the basis for the multiplex immunoassay platform described above.

#### ***4.1 Credit Card-Sized Devices and Mobile Phone Apps for User Friendly Rapid Testing***

Multiplex immunoassay biomarker tests have now been available for more than a decade on medium-sized laboratory equipment and with typical turnaround times of around 1 week from the sample preparation stages to the final results analysis. More recently, multiplex methods have been developed using microfluidic approaches to yield a devices which are approximately the size of a credit card [86]. This offers the possibility of inexpensive analysis using either electrochemical or optical read-outs. Most importantly, this approach is user friendly as no expertise is required for operation. The protocol involves application of a blood drop to the card followed by insertion of the card into a book-sized reader/ analyzer and then a diagnosis score can be read out in less than 15 min. The major benefit of this approach is the rapid turnover time and this will help to minimize waiting times for lab test results, which can often take several days or even weeks using standard methods. Furthermore, these devices can connect to a computer for transmission of data to a smartphone device. Large consumer market companies like Apple and Google are now showing interest in the diagnostic market and are exploiting the potential of linking diagnostic test results with an app driven by smart software. This would allow testing using real-time, multiplexed sensors, linked with artificial intelligence through mobile communication systems. This is of particular relevance to mental disorders such as schizophrenia, since it is generally a long term disease that requires constant monitoring and treatment. A recent review of trials involving medical care interventions facilitated by smartphones showed that patient outcomes were improved more than 60% of the time [87]. Recently, multiplex immunoassay based tests have been developed on a handheld smartphone-based colorimetric reader using a 3D-printed optomechanical interface [88]. To date, this approach has been tested successfully in a clinical microbiology laboratory using mumps, measles and herpes simplex I and II virus immunoglobulin tests. It is not hard to imagine that similar tests for other diseases such as psychiatric disorders will be available in the not so distant future.

---

## **5 Conclusions**

This chapter describes recent advances using biomarker tests which can be used for improved diagnosis and classification of individuals with schizophrenia. The ultimate goal is to provide more informed treatment options for improved patient outcomes. The use of the multiplex biomarker approach provides a way of unraveling the

convoluted array of molecular pathways affected in this disorder and potentially facilitate identification of disease subtypes which require different treatment approaches. For example, many patients show distinct patterns of blood-based molecules which suggest the presence of perturbed inflammation- or metabolism-related pathways as described in this chapter. Thus, improved classification of such patients based on biomarker profiling would enable selection of better treatment options such as the potential of including add-on therapeutics targeting these pathways. Finally, the use of multiplex tests on handheld devices capable of distinguishing schizophrenia patients who are most likely to respond to specific psychiatric medications would be an important breakthrough for point-of-care applications. This could help to improve the lives of individuals suffering from this debilitating disorder and have beneficial effects on society as well as significant cost savings for the healthcare services in general.

---

## Acknowledgments

D.M.S. and the Laboratory of Neuroproteomics, UNICAMP are funded by FAPESP (São Paulo Research Foundation) grant number 13/08711-3.

## References

1. van Os J, Kapur S (2009) Schizophrenia. *Lancet* 374:635–645
2. American Psychiatric Association (2013) Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). American Psychiatric Publishing, Arlington, VA. ISBN 10: 0890425558
3. World Health Organization (1 Jan. 1992) ICD-10: The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. World Health Organisation. ISBN-10: 9241544228
4. Penttilä M, Jääskeläinen E, Hirvonen N, Isohanni M, Miettunen J (2014) Duration of untreated psychosis as predictor of long-term outcome in schizophrenia: systematic review and meta-analysis. *Br J Psychiatry* 205:88–94
5. Jablensky A (2000) Epidemiology of schizophrenia: the global burden of disease and disability. *Eur Arch Psychiatry Clin Neurosci* 250:274–285
6. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE et al (2013) Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* 382:1575–1586
7. Liu Y, Dalal K, Stollenwerk B (2013) The association between health system development and the burden of cardiovascular disease: an analysis of WHO country profiles. *PLoS One* 8:e61718
8. Tsuang MT (1975) Heterogeneity of schizophrenia. *Biol Psychiatry* 10:465–474
9. Gonzalez-Pinto A, Gutierrez M, Mosquera F, Ballesteros J, Lopez P, Ezcurra J et al. (1998) First episode in bipolar disorder: misdiagnosis and psychotic symptoms. *J Affect Disord* 50:41–44
10. Follette WC, Houts AC (1996) Models of scientific progress and the role of theory in taxonomy development: a case study of the DSM. *J Consult Clin Psychol* 64: 1120–1132
11. Yolken RH, Dickerson FB, Fuller TE (2009) Toxoplasma and schizophrenia. *Parasite Immunol* 31:706–715
12. Lovatt A, Mason O, Brett C, Peters E (2010) Psychotic-like experiences, appraisals, and trauma. *J Nerv Ment Dis* 198:813–819

13. Clayton PJ, Guze SB, Cloninger CR, Martin RL (1992) Unipolar depression: diagnostic inconsistency and its implications. *J Affect Disord* 26:111–116
14. Bromet EJ, Kotov R, Fochtman LJ, Carlson GA, Tanenberg-Karant M, Ruggero C et al. (2011) Diagnostic shifts during the decade following first admission for psychosis. *Am J Psychiatry* 168:1186–1194
15. Torrey EF (1992) Are we overestimating the genetic contribution to schizophrenia? *Schizophr Bull* 18:159–170
16. McGue M (1992) When assessing twin concordance, use the probandwise not the pairwise rate. *Schizophr Bull* 18:171–176
17. Tsuang M (2000) Schizophrenia: genes and environment. *Biol Psychiatry* 47:210–220
18. Dauncey MJ, Bicknell RJ (1999) Nutrition and neurodevelopment: mechanisms of developmental dysfunction and disease in later life. *Nutr Res Rev* 12:231–253
19. Schlotz W, Phillips DI (2009) Fetal origins of mental health: evidence and mechanisms. *Brain Behav Immun* 23:905–916
20. Koenig JI, Kirkpatrick B, Lee P (2002) Glucocorticoid hormones and early brain development in schizophrenia. *Neuropsychopharmacology* 27:309–318
21. Ryan MC, Collins P, Thakore JH (2003) Impaired fasting glucose tolerance in first-episode, drug-naïve patients with schizophrenia. *Am J Psychiatry* 160:284–289
22. Spelman LM, Walsh PI, Sharifi N, Collins P, Thakore JH (2007) Impaired glucose tolerance in first-episode drug-naïve patients with schizophrenia. *Diabet Med* 24:481–445
23. Guest PC, Wang L, Harris LW, Burling K, Levin Y, Ernst A et al. (2010) Increased levels of circulating insulin-related peptides in first-onset, antipsychotic naïve schizophrenia patients. *Mol Psychiatry* 15:118–119
24. Szulc A, Galińska B, Konarzewska B, Gudel-Trochimowicz I, Popławska R (2001) Immunological marker activity in first episode schizophrenic patients. *Pol Merkur Lekarski* 10:450–452
25. Van Venrooij JA, Fluitman SB, Lijmer JG, Kavelaars A, Heijnen CJ, Westenberg HG et al. (2012) Impaired neuroendocrine and immune response to acute stress in medication-naïve patients with a first episode of psychosis. *Schizophr Bull* 38:272–279
26. Perkins DO, Jeffries CD, Addington J, Bearden CE, Cadenhead KS, Cannon TD et al. (2014) Towards a psychosis risk blood diagnostic for persons experiencing high-risk symptoms: preliminary results from the NAPLS project. *Schizophr Bull* 41: 419–428
27. Chan MK, Krebs MO, Cox D, Guest PC, Yolken RH, Rahmoune H et al. (2015) Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. *Transl Psychiatry* 5, e601. doi:[10.1038/tp.2015.91](https://doi.org/10.1038/tp.2015.91)
28. Agius M, Shah S, Ramkisson R, Murphy S, Zaman R (2007) Three year outcomes of an early intervention for psychosis service as compared with treatment as usual for first psychotic episodes in a standard community mental health team. Preliminary results. *Psychiatr Danub* 19:10–19
29. Salokangas RK, McGlashan TH (2008) Early detection and intervention of psychosis. A review. *Nord J Psychiatry* 62:92–105
30. Yap HL (2010) Early psychosis intervention. *Singapore Med J* 51:689–693
31. Thomas P (2004) The many forms of bipolar disorder: a modern look at an old illness. *J Affect Disord* 79(Suppl 1):S3–S8
32. Hirschfeld RM (2001) Bipolar spectrum disorder: improving its recognition and diagnosis. *J Clin Psychiatry* 62(Suppl 14):5–9
33. Post RM (2005) The impact of bipolar depression. *J Clin Psychiatry* 66(Suppl 5):5–10
34. Kamel N, Compton C, Middelveld R, Higenbottam T, Dahlén SE (2008) The Innovative Medicines Initiative (IMI): a new opportunity for scientific collaboration between academia and industry at the European level. *Eur Respir J* 31:924–926
35. Hunter AJ (2008) The Innovative Medicines Initiative: a pre-competitive initiative to enhance the biomedical science base of Europe to expedite the development of new medicines for patients. *Drug Discov Today* 13:371–373
36. <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/IVDRegulatoryAssistance/ucm124105.htm>
37. Ellington AA, Kullo IJ, Bailey KR, Klee GG (2010) Antibody-based protein multiplex platforms: technical and operational challenges. *Clin Chem* 56:186–193
38. Söderlund J, Schröder J, Nordin C, Samuelsson M, Walther-Jallow L, Karlsson H et al. (2009) Activation of brain interleukin-1beta in schizophrenia. *Mol Psychiatry* 14:1069–1071
39. Merrill JE (1992) Tumor necrosis factor alpha, interleukin 1 and related cytokines in brain

- development: normal and pathological. *Dev Neurosci* 14:1–10
40. Meyer U, Feldon J, Yee BK (2009) A review of the fetal brain cytokine imbalance hypothesis of schizophrenia. *Schizophr Bull* 35:959–972
41. Saetre P, Emilsson L, Axelsson E, Kreuger J, Lindholm E, Jazin E (2007) Inflammation-related genes up-regulated in schizophrenia brains. *BMC Psychiatry* 7:46
42. Schmitt A, Leonardi-Essmann F, Durrenberger PF, Parlapani E, Schneider-Axmann T, Spanagel R et al. (2011) Regulation of immunemodulatory genes in left superior temporal cortex of schizophrenia patients: a genome-wide microarray study. *World J Biol Psychiatry* 12:201–215
43. Montejo AL (2010) The need for routine physical health care in schizophrenia. *Eur Psychiatry* 25(Suppl 2):S3–S5
44. Pedrini M, Massuda R, Fries GR, de Bittencourt-Pasquali MA, Schnorr CE, Moreira JC, et al. (2012) Similarities in serum oxidative stress markers and inflammatory cytokines in patients with overt schizophrenia at early and late stages of chronicity. *J Psychiatr Res* 46:819–824
45. Miller BJ, Buckley P, Seabolt W, Mellor A, Kirkpatrick B (2011) Meta-analysis of cytokine alterations in schizophrenia: clinical status and antipsychotic effects. *Biol Psychiatry* 70: 663–671
46. Schwarz E, Izmailov R, Spain M, Barnes A, Mapes JP, Guest PC et al. (2010) Validation of a blood-based laboratory test to aid in the confirmation of a diagnosis of schizophrenia. *Biomark Insights* 5:39–47
47. Schwarz E, Guest PC, Rahmoune H, Harris LW, Wang L, Leweke FM et al. (2012) Identification of a biological signature for schizophrenia in serum. *Mol Psychiatry* 17: 494–502
48. Späth-Schwalbe E, Born J, Schrezenmeier H, Bornstein SR, Stromeyer P, Drechsler S et al. (1994) Interleukin-6 stimulates the hypothalamus-pituitary-adrenocortical axis in man. *J Clin Endocrinol Metab* 79:1212–1214
49. Straub RH, Buttigereit F, Cutolo M (2011) Alterations of the hypothalamic-pituitary-adrenal axis in systemic immune diseases – a role for misguided energy regulation. *Clin Exp Rheumatol* 29:S23–S31
50. Bremner JD (2006) Traumatic stress: effects on the brain. *Dialogues Clin Neurosci* 8: 445–461
51. Spijker AT, van Rossum EF (2012) Glucocorticoid sensitivity in mood disorders. *Neuroendocrinology* 95:179–186
52. Müller N, Riedel M, Schwarz MJ (2004) Psychotropic effects of COX-2 inhibitors – a possible new approach for the treatment of psychiatric disorders. *Pharmacopsychiatry* 37: 266–269
53. Akhondzadeh S, Tabatabaei M, Amini H, Ahmadi Abhari SA, Abbasi SH, Behnam B (2007) Celecoxib as adjunctive therapy in schizophrenia: a doubleblind, randomized and placebo-controlled trial. *Schizophr Res* 90: 179–185
54. Müller N, Krause D, Dehning S, Musil R, Schennach-Wolff R, Obermeier M et al. (2010) Celecoxib treatment in an early stage of schizophrenia: results of a randomized, double-blind, placebocontrolled trial of celecoxib augmentation of amisulpride treatment. *Schizophr Res* 121:118–124
55. Laan W, Grobbee DE, Selten JP, Heijnen CJ, Kahn RS, Burger H (2010) Adjuvant aspirin therapy reduces symptoms of schizophrenia spectrum disorders: results from a randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry* 71:520–527
56. Arranz B, Rosel P, Ramirez N, Duenas R, Fernandez P, Sanchez JM et al (2004) Insulin resistance and increased leptin concentrations in noncompliant schizophrenia patients but not in antipsychotic-naïve first episode schizophrenia patients. *J Clin Psychiatry* 65: 1335–1342
57. Cohn TA, Remington G, Zipursky RB, Azad A, Connolly P, Wolever TM (2006) Insulin resistance and adiponectin levels in drug-free patients with schizophrenia: a preliminary report. *Can J Psychiatry* 51:382–386
58. van Nimwegen LJ, Storosum JG, Blumer RM, Allick G, Venema HW, de Haan L et al (2008) Hepatic insulin resistance in antipsychotic naïve schizophrenic patients: stable isotope studies of glucose metabolism. *J Clin Endocrinol Metab* 93:572–577
59. Guest PC, Schwarz E, Krishnamurthy D, Harris LW, Leweke FM, Rothermundt M (2011) Altered levels of circulating insulin and other neuroendocrine hormones associated with the onset of schizophrenia. *Psychoneuroendocrinology* 2011(36):1092–1096
60. Taguchi A, Wartschow LM, White MF (2007) Brain IRS2 signaling coordinates life span and nutrient homeostasis. *Science* 317:369–372
61. Convit A (2005) Links between cognitive impairment in insulin resistance: an explanatory model. *Neurobiol Aging* 26(Suppl 1):31–35
62. Craft S (2007) Insulin resistance and Alzheimer's disease pathogenesis: potential

- mechanisms and implications for treatment. *Curr Alzheimer Res* 4:147–152
63. Bello NT, Hajnal A (2006) Alterations in blood glucose levels under hyperinsulinemia affect accumbens dopamine. *Physiol Behav* 2006(88):138–145
  64. O'Malley D, Shanley LJ, Harvey J (2003) Insulin inhibits rat hippocampal neurones via activation of ATP-sensitive K<sup>+</sup> and large conductance Ca<sup>2+</sup>-activated K<sup>+</sup> channels. *Neuropharmacology* 44:855–863
  65. Corcoran CM, Smith C, McLaughlin D, Auther A, Malaspina D, Cornblatt B (2012) HPA axis function and symptoms in adolescents at clinical high risk for schizophrenia. *Schizophr Res* 135:170–174
  66. Gorobets LN, Matrosova MI (2010) Specialties of prolactin secretion and peripheral reproductive sex hormones in patients with first episode of schizophrenia. *Zh Nevrol Psichiatr Im S S Korsakova* 110:17–22
  67. Akiibinu MO, Ogundahunsi OA, Ogunyemi EO (2012) Inter-relationship of plasma markers of oxidative stress and thyroid hormones in schizophrenics. *BMC Res Notes* 5:169
  68. Ben-Jonathan N, Hugo ER, Brandebourg TD, LaPensee CR (2006) Focus on prolactin as a metabolic hormone. *Trends Endocrinol Metab* 17:110–116
  69. Tannenbaum GS, Martin JB, Colle E (1976) Ultradian growth hormone rhythm in the rat: effects of feeding, hyperglycemia, and insulin-induced hypoglycemia. *Endocrinology* 99:720–727
  70. Meltzer HY, Perry E, Jayathilake K (2003) Clozapine-induced weight gain predicts improvement in psychopathology. *Schizophr Res* 59:19–27
  71. Bahtiyar G, Weiss K, Sacerdote AS (2007) Novel endocrine disrupter effects of classic and atypical antipsychotic agents and divalproex: induction of adrenal hyperandrogenism, reversible with metformin or rosiglitazone. *Endocr Pract* 13:601–608
  72. Sato T, Hanyu H, Hirao K, Kanetaka H, Sakurai H, Iwamoto T (2011) Efficacy of PPARgamma agonist pioglitazone in mild Alzheimer disease. *Neurobiol Aging* 32: 1626–1633
  73. Nachshoni T, Ebert T, Abramovitch Y, Assael Amir M, Kotler M, Maayan R et al (2005) Improvement of extrapyramidal symptoms following dehydroepiandrosterone (DHEA) administration in antipsychotic treated schizophrenia patients: a randomized, double-blind placebo controlled trial. *Schizophr Res* 79: 251–256
  74. Usall J, Huerta-Ramos E, Iniesta R, Cobo J, Araya S, Roca M et al (2011) Raloxifene as an adjunctive treatment for postmenopausal women with schizophrenia: a double-blind, randomized, placebo-controlled trial. *J Clin Psychiatry* 72:1552–1557
  75. Arranz MJ, Munro J, Birkett J, Bolonna A, Mancama D, Sodhi M et al (2000) Pharmacogenetic prediction of clozapine response. *Lancet* 355:1615–1616
  76. Zhang JP, Malhotra AK (2011) Pharmacogenetics and antipsychotics: therapeutic efficacy and side effects prediction. *Expert Opin Drug Metab Toxicol* 7:9–37
  77. Gebhardt S, Haberhausen M, Heinzel-Gutenbrunner M, Gebhardt N, Remschmidt H, Krieg JC et al (2009) Antipsychotic-induced body weight gain: predictors and a systematic categorization of the long-term weight course. *J Psychiatr Res* 43:620–626. doi:[10.1016/j.jpsychires.2008.11.001](https://doi.org/10.1016/j.jpsychires.2008.11.001), Epub 2008 Dec 24
  78. Lau SL, Muir C, Assur Y, Beach R, Tran B, Bartrop R et al (2016) Predicting weight gain in patients treated with clozapine: the role of sex, body mass index, and smoking. *J Clin Psychopharmacol* 36:120–124
  79. Shrivastava A, Johnston M, Bureau Y, Shah N (2012) Baseline serum prolactin in drug naive, first-episode schizophrenia and outcome at five years: is it a predictive factor? *Innov Clin Neurosci* 2012(9):17–21
  80. Schwarz E, Guest PC, Steiner J, Bogerts B, Bahn S (2012) Identification of blood based molecular signatures for prediction of response and relapse in schizophrenia patients. *Transl Psychiatry* 2, e82
  81. Schwarz E, Steiner J, Guest PC, Bogerts B, Bahn S (2015) Investigation of molecular serum profiles associated with predisposition to antipsychotic-induced weight gain. *World J Biol Psychiatry* 16:22–30
  82. Tomasik J, Schwarz E, Lago SG, Rothermundt M, Leweke FM, van Beveren NJ et al (2016) Pretreatment levels of the fatty acid handling proteins H-FABP and CD36 predict response to olanzapine in recent-onset schizophrenia patients. *Brain Behav Immun* 52:178–186
  83. Domenici E, Willé DR, Tozzi F, Prokopenko I, Miller S, McKeown A et al (2010) Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections. *PLoS One* 5(2):e9166
  84. Salmon SE, Mackey G, Fudenberg HH (1969) “Sandwich” solid phase radioimmunoassay

- for the quantitative determination of human immunoglobulins. *J Immunol* 103:129–137
85. Salmon SE, Smith BA (1970) Sandwich solid phase radioimmunoassays for the characterization of human immunoglobulins synthesized in vitro. *J Immunol* 104: 665–672
86. Schumacher S, Nestler J, Otto T, Wegener M, Ehrentreich-Förster E, Michel D et al (2012) Highly-integrated lab-on-chip system for point-of-care multiparameter analysis. *Lab Chip* 12: 464–473
87. Krishna S, Boren SA, Balas EA (2009) Healthcare via cell phones: a systematic review. *Telemed J E Health* 15:231–240
88. Berg B, Cortazar B, Tseng D, Ozkan H, Feng S, Wei Q et al (2015) Cellphone-based handheld micro-plate reader for point-of-care testing of enzyme-linked immunosorbent assays. *ACS Nano* 9:7857–7866

# Chapter 3

## Multiplex Biomarker Approaches in Type 2 Diabetes Mellitus Research

Susan E. Ozanne, Hassan Rahmoune, and Paul C. Guest

### Abstract

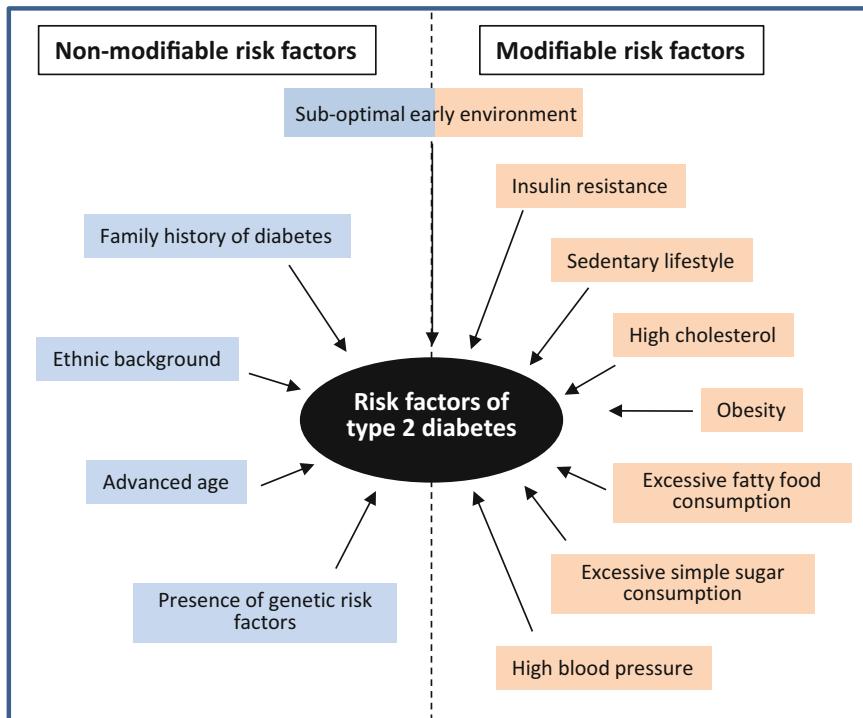
Type 2 diabetes mellitus is a multifactorial condition resulting in high fasting blood glucose levels. Although its diagnosis is straightforward, there is not one set of biomarkers or drug targets that can be used for classification or personalized treatment of individuals who suffer from this condition. Instead, the application of multiplex methods incorporating a systems biology approach is essential in order to increase our understanding of this disease. This chapter reviews the state of the art in biomarker studies of human type 2 diabetes from a proteomic and metabolomic perspective. Our main focus was on biomarkers for disease prediction as these could lead to early intervention strategies for the best possible patient outcomes.

**Key words** Type 2 diabetes mellitus, Biomarkers, Drug targets, Genomics, Proteomics, Metabolomics

---

### 1 Introduction

Type 2 diabetes mellitus is a complex metabolic disorder resulting from a combination of insulin resistance and insulin deficiency. This type of diabetes accounts for 90 % of the cases worldwide and can be distinguished from type 1 diabetes by the fact that it is not autoimmune in origin. After a marked increase over the last three decades, the prevalence of this disorder has now reached epidemic proportions, with almost 400 million adult cases throughout the world [1]. One reason that has been proposed to account for this increase is a possible epidemiologic transition away from communicable diseases as being the major cause of early death. However, it is also likely that this is due to a transition towards increasingly unhealthy dietary habits and lower levels of physical activity [2, 3]. Therefore, there is an urgent need for the development of early intervention strategies to address the public health, economic, scientific and ethical calls to the individual, as well as the societal and healthcare burdens associated with this disease.



**Fig. 1** Schematic diagram showing the impact of various risk factors for development of type 2 diabetes. The factors on the *right* and *left* are listed as non-modifiable and modifiable, respectively but it should be noted that these are generalizations

To date, type 2 diabetes has been difficult to predict and there is no cure. However, abnormalities in glucose metabolism including states of moderate hyperglycaemia are known to occur well before disease onset [4, 5]. Therefore, availability of a reliable biomarker test for identification prior to development of hyperglycaemia would enable high risk individuals to adopt certain nutritional and physical activity lifestyle changes to delay or even halt progression of the disease [6–8]. In turn, this would help to minimize its impact and improve the long term health outcome of the individual.

Known risk factors for type 2 diabetes include advanced age, high body mass index (BMI), poor diet, low physical activity, unfavorable intrauterine environment and certain adverse genetic associations [1, 7, 9–12] (Fig. 1). Molecular studies have also pointed at insulin resistance and knock-on effects in liver, skeletal muscle and adipose tissues as potential predictive markers of type 2 diabetes development [4, 5, 13–15]. Genome Wide Association Studies (GWAS) have identified several genetic risk allele [14]. Many of these are thought to be involved in regulation of pancreatic  $\beta$ -cell function, although the mechanisms by which they do so are poorly defined.

To improve patient outcomes, prediction of diabetes should be achieved at the earliest possible stage of the disease. Here we review the various multiplex proteomic and metabolomic studies which have been performed in an effort to identify predictive biomarkers for this disorder. These two approaches were chosen since proteomic and metabolomic markers can be detected and routinely monitored in the blood, making these more amenable for clinical investigations. Nevertheless, the results of other biomarker-related approaches will be presented to demonstrate consistency of the biological findings.

---

## 2 Epigenetic Biomarkers for Prediction of Type 2 Diabetes

Most genetic studies of type 2 diabetes have not attempted to account for information on interaction with potential environmental factors such as nutrition, body mass, activity levels or lifestyle. However, it is likely that the genetic variants associated identified through GWAS act in combination with physiological perturbations caused by environmental factors such as poor nutrition or high body mass indices to increase the risk of diabetes. The interaction between the environment has been termed epigenetics and explains why one genotype can give rise to multiple phenotypes. The main epigenetic modifications that mediate this interaction between the genotype and the environment are DNA methylation (on cytosine residues), histone acetylation/methylation and alterations in levels of small noncoding RNAs such as miRNAs.

Several studies have now found significant differences in site-specific DNA methylation patterns between tissues from type 2 diabetics and glucose tolerant individuals as well as during aging in humans [16–21]. Although in many cases the analysis has been of a clinically accessible tissue such as peripheral blood mononuclear cells (PBMCs) some have been carried out in more metabolically relevant tissues such as pancreatic islets, adipose tissue liver and skeletal muscle [22–26]. There is much debate as to how reflective DNA methylation changes in PBMCs are of the same loci in tissues involved in glucose homeostasis such as liver, muscle, adipose tissue and pancreatic islets. However regardless of whether the changes are indicative of the underlying biology they could still represent useful biomarkers [6, 27]. One study using PBMCs from more than 800 individuals who were not diabetic identified an association between differential methylation in a cholesterol transport gene (*ABCG1*) and fasting insulin levels [28]. Another investigation found methylation differences in the nuclear factor-kB pathway gene (*MALTI*) and the G-protein receptor 6 gene (*GPR61*) in PBMCs from twins who were discordant for type 2 diabetes [29] and another identified DNA hypo-methylation of specific loci in young individuals who later developed diabetes [30]. In addition,

a study identified methylation differences in genes such as *PPARG* and *IRSI* in adipose tissue from individuals with type 2 diabetes compared to nondiabetic controls [25]. Both of these genes encode proteins that are involved in insulin signaling.

Other studies have identified small noncoding RNAs (e.g., microRNAs) as another possible source of biomarkers for type 2 diabetes [31–33]. Guay and colleagues reviewed the potential of using circulating microRNA profiles as a means of monitoring specific aspects of health and they identified patterns which appeared to be predictive of long-term complications in patients with type 1 and type 2 diabetes [34].

---

### 3 Proteomic and Metabolomic Technologies in Type 2 Diabetes Research

Biomarkers are physical characteristics or molecules that can be monitored and used as an indication of physiological parameters such as health, disease or drug response. For practical purposes, biomarkers should be easily accessible due to the potential need of routine sampling. This section will give a brief description of proteomic and metabolomic profiling in blood samples (for more in depth descriptions, see other chapters in this volume). It should be kept in mind that changes in physiological state can be dynamic and thus methods that are capable of coping with this are essential. Most researchers now study proteins and metabolites, considering that these molecules actually carry out or respond to most processes in cells, tissues or organisms. In addition, most drugs in use today target proteins such as enzymes or receptors and induce metabolic change, evidenced as the synthesis or degradation of both proteins and metabolites. With this in mind, most diabetes researchers have employed methods such as multi-protein arrays, two dimensional gel electrophoresis (2DGE), mass spectrometry or nuclear magnetic resonance (NMR) profiling platforms.

#### 3.1 Multiplex Immunoassay

Blood serum and plasma contain many important bioactive or regulatory molecules, such as hormones, growth factors and cytokines which are present only at low concentrations. Therefore, detection of these by high sensitivity platforms is essential. This can be achieved by multiplex immunoassay as described by Fulton and coworkers in 1997 [35]. The assay consists of multiple specific antibodies each covalently linked to spectrally distinct microbeads. When the sample is added the target molecules will bind to the appropriate beads and fluorescently labeled detection antibodies are then added which bind to the target molecules in a sandwich-like format. Finally, the beads are passed through a reader in a narrow stream for analysis by two lasers for identification and quantification of the targeted analyte.

### **3.2 Two-Dimensional Gel Electrophoresis (2DE)**

In this method, protein samples are applied to a strip gel with an immobilized-pH gradient and the resident proteins are separated according to their isoelectric points (pI) by isoelectric focusing. Next the strips are applied onto a sodium dodecyl sulfate (SDS)-polyacrylamide gel and electrophoresed to separate the proteins according to their molecular weights (MW). The resulting protein spot constellations can be visualized with post-electrophoretic stains such as Coomassie Blue or Sypro Ruby and quantitated using appropriate imaging software. Alternatively, multiple protein samples can be pre-labeled with spectrally distinct fluorescent dyes and then electrophoresed on the same gels to allow direct comparison of the different proteomes [36]. This latter technique known as 2D-DIGE (two-dimensional difference gel electrophoresis) helps by eliminating the need for gel to gel comparisons. The resolution power of 2DE and 2D-DIGE are remarkable for proteome investigations but limitations include the poor detection of very acidic and very basic proteins as well as low abundant proteins [37]. Also, there is a need to use other techniques to identify the separated proteins using a technique such as mass spectrometry.

### **3.3 Mass Spectrometry**

In proteomics, mass spectrometry was initially used to identify proteins previously separated by 2DE, employing peptide mass fingerprints [38]. Given 2DE limitations and the fact that mass spectrometers equipped with ESI sources could be connected online to liquid chromatography systems, the concept of shotgun proteomics or shotgun mass spectrometry has emerged [39]. This has revolutionized the field towards the end of the Human Genome Project, due to its sensitivity and high throughput power for proteomic biomarker identification [39]. The term “shotgun” comes from the point that the target proteins can be cleaved with enzymes to generate smaller peptides, which are the actual analytes in this approach (see other chapters in this book for more detailed information). Cleavage is performed as most intact proteins are too large and complex at the structural level to enable direct analysis. After cleavage, the peptides are separated according to their physiochemical properties by liquid chromatography and they enter the mass spectrometer by electrospray ionization. This ionization stage allows the peptides to be accelerated by magnets in the mass spectrometer towards a detector at a speed that is inversely proportional to their mass/charge ratio. Quantitation occurs at the detector essentially via the number of impacts of each given peptide ion. At the same time, the peptide sequence can be determined by bombardment with a streaming gas to allow fragmentation into smaller pieces. Determining the masses of these fragments can then be used to derive the amino acid sequences that make up the peptides, which could then be used to search a protein database to obtain the identification of the parent protein. This method can also be used for metabolomic analysis, with some modifications.

In this case, there is no need for enzymatic cleavage of the molecules as these are already of a manageable size. Thus, the sample can be infused directly into the mass spectrometer, the quantity measured as above, and the identity of the metabolite can be determined through comparison with known standards.

### **3.4 *<sup>1</sup>H-Nuclear Magnetic Resonance (NMR) Spectroscopy***

<sup>1</sup>H-NMR spectroscopy is used for metabolite and small molecule analyses. One major advantage of this method is the simplicity and reproducibility of the sample preparation stage. However, it is less sensitive than mass spectrometry approaches for metabolite detection [40, 41]. The technique gives structural information about the target molecules and so it is ideal for identification purposes. This works by monitoring the behavior of protons on molecules in a strong magnetic field. The nuclei of the protons line up with the magnetic field in a similar manner as a compass needle aligns with the Earth's magnetic field. The NMR procedure is initiated through application of radio pulses to the sample and this stimulates the nuclei to rotate around the axis of the magnetic field with a frequency related to the physiochemical environment of the surrounding atoms within the molecule. NMR can be used to monitor relative changes in the levels of small molecules such as amino acids, sugars and lipids.

---

## **4 Proteomic Biomarkers**

### **4.1 *Biomarkers for Prediction of Type 2 Diabetes***

There is no question about whether beta cell dysfunction leading to impaired insulin production and insulin resistance are established risk factors for type 2 diabetes. However, since insulin is known to be involved in regulation of metabolism and/or growth of virtually all cells in the body, there are likely to be many more associated proteins which can be detected in serum or plasma that are biomarkers of type 2 diabetes. An early study used surface enhanced laser desorption/ionization time-of-flight (SELD-TOF) mass spectrometry for identification of potential serum biomarkers associated with type 2 diabetes [42]. This led to detection of four proteins which showed significant quantitative differences between the diabetic and control groups and these were identified by peptide mass fingerprinting as albumin, apolipoprotein C3, transferrin, and transthyretin (Table 1). Another investigation found decreased levels of apolipoprotein A1 and increased levels of apolipoprotein E, leptin and C-reactive protein in patients with type 2 diabetes compared to controls, using a combination of 2DGE and matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry [43]. Furthermore they validated these changes using immunoassays. Another 2DGE study analyzed levels of proteins in plasma samples from patients with poorly controlled diabetes and nondiabetic controls [44]. This resulted in identification of increased levels of fibrinogen and hap-

**Table 1**  
**Potential serum/plasma proteomic biomarkers for risk of type 2 diabetes and associated complications**

| Biomarker                                  | Function            | Citation |
|--|---------------------|----------|
| Type 2 diabetes                            |                     |          |
| $\alpha$ 1-antitrypsin                     | Inflammation        | [41]     |
| Albumin                                    | Molecular transport | [39]     |
| Albumin-cysteinylated                      | Molecular transport | [42]     |
| Albumin-glycated                           | Molecular transport | [42]     |
| Apolipoprotein A1                          | Molecular transport | [40, 41] |
| Apolipoprotein A1-glycated                 | Molecular transport | [42]     |
| Apolipoprotein A1-oxidized                 | Molecular transport | [42]     |
| Apolipoprotein C3                          | Molecular transport | [39]     |
| Apolipoprotein E                           | Molecular transport | [40]     |
| C-reactive peptide                         | Inflammation        | [40]     |
| Hemoglobin-glycated                        | Molecular transport | [42]     |
| Hemoglobin-nitrosylated                    | Molecular transport | [42]     |
| Leptin                                     | Hormone signaling   | [40]     |
| Transferrin                                | Molecular transport | [39]     |
| Transthyretin                              | Molecular transport | [39, 41] |
| Vitamin D binding protein                  | Molecular transport | [41]     |
| Diabetes-associated cardiovascular disease |                     |          |
| Interleukin 6                              | Inflammation        | [51]     |
| Plasminogen activator inhibitor-1          | Coagulation         | [51]     |
| Tumor necrosis factor- $\alpha$            | Inflammation        | [51]     |
| von Willebrand factor                      | Coagulation         | [51]     |
| Diabetes-associated nephropathy            |                     |          |
| Albumin                                    | Molecular transport | [15, 53] |
| C-reactive protein                         | Inflammation        | [60]     |
| E-selectin                                 | Inflammation        | [60]     |
| Intercellular cell adhesion molecule 1     | Inflammation        | [58, 59] |
| Interleukin 6                              | Inflammation        | [58, 59] |
| Interleukin 18                             | Inflammation        | [54–58]  |

(continued)

**Table 1**  
**(continued)**

| Biomarker                                      | Function            | Citation |
|--|---------------------|----------|
| Tissue plasminogen activator                   | Coagulation         | [60]     |
| Vascular cell adhesion molecule 1              | Inflammation        | [58, 59] |
| von Willebrand factor                          | Coagulation         | [58–60]  |
| Diabetes-associated retinopathy                |                     |          |
| $\alpha$ 1-antitrypsin                         | Inflammation        | [62]     |
| $\alpha$ 2-macroglobulin                       | Inflammation        | [62]     |
| Afamin   | Molecular transport | [62]     |
| Antithrombin-III                               | Coagulation         | [62]     |
| Apolipoprotein A1                              | Molecular transport | [62]     |
| Apolipoprotein A4                              | Molecular transport | [64]     |
| Apolipoprotein E                               | Molecular transport | [62]     |
| CD40 ligand                                    | Inflammation        | [63]     |
| Clusterin                                      | Inflammation        | [64]     |
| Complement C1/C3                               | Inflammation        | [64)     |
| Complement C7                                  | Inflammation        | [64]     |
| Fibrinogen                                     | Coagulation         | [62]     |
| Gelsolin                                       | Coagulation         | [62]     |
| Haptoglobin                                    | Molecular transport | [62]     |
| Inter- $\alpha$ -trypsin inhibitor H2          | Inflammation        | [64]     |
| Interleukins 6, 7, 9, 13, 15, 17, S2R $\alpha$ | Inflammation        | [63]     |
| Kininogen-1                                    | Coagulation         | [62]     |
| Leucine-rich alpha-2-glycoprotein              | Inflammation        | [62]     |
| Monocyte chemoattractant protein-1             | Inflammation        | [63]     |
| Protein arginine N-methyltransferase 5         | Growth signaling    | [62]     |
| Serum amyloid P                                | Molecular transport | [62]     |
| Tumor necrosis factor- $\beta$                 | Inflammation        | [63]     |
| Vitamin D-binding protein                      | Molecular transport | [62]     |
| Vitronectin                                    | Inflammation        | [62]     |

toglobin along with decreased levels of  $\alpha$ 1-antitrypsin, apolipoprotein A1, transthyretin, and vitamin D binding protein in plasma from the diabetic individuals. A more recent study described the development of a new on-chip liquid chromatography-nanoelectrospray ionization mass spectrometry blood-based glycaemia monitoring assay which afforded simultaneous measurement of glucose and glycated forms of hemoglobin (HbA1c), albumin and apolipoprotein A1 [45]. This study incorporated assays for cysteinylated albumin, S-nitrosylated hemoglobin, and methionine-oxidized apolipoprotein A1 for assessing oxidative stress and cardiovascular risk [45]. The results showed that this assay was capable of distinguishing type 2 diabetes patients from healthy controls although further testing using additional cohorts in multiple clinical settings is required to confirm the findings.

#### **4.2 Biomarkers for Assessing the Risk of Diabetes-Associated Cardiovascular Disease**

Diabetes is the fifth leading cause of death worldwide being associated with more than 5 % of all deaths. This is because diabetes is a known risk factor for development of several other disorders, including cardiovascular disease, stroke and peripheral vascular disease [46]. Approximately two-thirds of the deaths attributed to type 2 diabetes result from comorbid cardiovascular events such as coronary artery disease [47]. Thus, the life expectancy for persons with type 2 diabetes is around 5–10 years lower than for non-diabetic individuals [48]. However, the underlying mechanisms linking diabetes with cardiovascular conditions have not been completely elucidated. Considering this, there is a critical need for further studies of molecular risk factors and the development of biomarker assays that can be used to predict, characterize and monitor this disorder along with its associated comorbidities and outcomes. Of particular note, several studies over the past few years have reported that treatment of type 2 diabetes patients using glucose lowering agents alone does not lead to a reduction in severity of a heart disease-related comorbidity [49–52]. This finding suggests that therapeutic measures in addition to glycemic control are needed. Furthermore, the Food and Drug Administration has now issued a new guidance to drug companies concerning restrictions on the development of novel antidiabetic pharmaceuticals [53]. In essence, this states that new drugs in development must be evaluated for any cardiovascular disease risk. Therefore, there is also a clinical need for the development of new biomarker tests which can be used to predict the propensity towards this risk as well as the physiological effects of novel antidiabetic compounds.

Recent proteomic studies have suggested that there is an interrelation between the metabolic disruptions in type 2 diabetes and inflammation, which leads to a disease state through an increased coagulation response and consequential detrimental effects on endothelial cells and the vasculature (Table 1) [15]. The increased expression of inflammatory cytokines and adhesion molecules can

amplify the inflammatory responses, leading to an aggravation of diabetic vascular complications. In turn, this may initiate a process of atherosclerosis and the formation of arterial thrombi. This appears to result in the changes in circulating pro-coagulant and inflammatory biomarkers including von Willebrand factor, interleukin-6, tumor necrosis factor- $\alpha$ , and plasminogen activator inhibitor-1. In addition several of the serum and plasma proteins detected as potential biomarkers for type 2 diabetes have also been implicated in separate studies of coronary heart disease. This includes the proteins albumin and C-reactive protein, as well as many others associated with coagulation or inflammation [54]. Such molecules may serve as potential biomarkers for predicting the risk of cardiovascular and renal perturbations in diabetic patients and for the monitoring of patient responses to therapeutics targeting these widespread complications.

#### **4.3 Biomarkers for Prediction of Diabetic Nephropathy**

Diabetic nephropathy is the leading cause of chronic renal failure in Western countries. It has been associated with early death and can also negatively affect patient quality of life, as well as being a significant burden on the healthcare systems [55]. One of the first clinical signs of such microvascular damage in diabetes is an increase in the levels of albumin [56, 57]. This has led to screening of albumin levels in patients with diabetes to identify those at risk of microvascular complications such as nephropathy [53]. Recent longitudinal studies have also been carried out and these have led to the identification of proteomic and metabolomic biomarkers which can predict the onset or progression of nephropathy in patients with type 2 diabetes. The most robust markers for prediction of onset of neuropathy identified to date are plasma asymmetric dimethylarginine, serum interleukin 18 and urinary ceruloplasmin, immunoglobulin G and transferrin (Tables 1 and 2) [58–62]. For predicting neuropathy progression, plasma levels of intercellular cell adhesion molecule 1, interleukin 6, plasma asymmetric dimethylarginine, vascular cell adhesion molecule 1, and von Willebrand factor have been consistently identified as the most robust biomarkers [62, 63]. For prediction of both the onset and progression of neuropathy, the most accurate biomarkers were plasma C-reactive protein, E-selectin, tissue-type plasminogen activator, triglycerides, and von Willebrand factor [64]. This indicated that the lists of biomarkers for predicting onset and progression were distinct. Another study carried out multiplex immunoassay analyses of the levels of 27 cytokines in urine from patients with type 2 diabetes and micro- or normo-albuminuria [65]. They found that the levels of eotaxin, granulocyte colony-stimulating factor, interferon- $\gamma$ -inducible protein 10, interleukin 8, monocyte chemoattractant protein-1, RANTES and tumor necrosis factor- $\alpha$  were increased significantly in micro-albuminuric patients compared to levels in normo-albuminuric diabetic patients or controls. On the other

**Table 2**  
**Potential plasma/serum metabolomic biomarkers for risk of type 2 diabetes and associated complications**

| Biomarker                                  | Function                  | Citation     |
|--|---------------------------|--------------|
| Type 2 diabetes                            |                           |              |
| $\alpha$ -hydroxybutyric acid              | Lipid-related             | [70, 71]     |
| Acetone                                    | Lipid-related             | [70]         |
| Acetoacetate                               | Lipid-related             | [70]         |
| Acyl-alkyl-phosphatidylcholine             | Lipid-related             | [66]         |
| $\beta$ -hydroxybutyrate                   | Lipid-related             | [70]         |
| Diacylphosphatidylcholine C32:1            | Lipid-related             | [66]         |
| Glucose                                    | Sugar                     | [65]         |
| Glycine                                    | Amino acid                | [66, 69]     |
| Isoleucine                                 | Branched chain amino acid | [65–67]      |
| Leucine                                    | Branched chain amino acid | [65–67]      |
| Linoleoyl-glycerophosphocholine            | Lipid-related             | [71]         |
| lysophosphatidylcholine C18:2              | Lipid-related             | [65, 66, 69] |
| Phenylalanine                              | Aromatic amino acid       | [65, 66]     |
| Sphingomyelin C16:1                        | Lipid-related             | [66]         |
| Tyrosine                                   | Aromatic amino acid       | [65, 66]     |
| Valine                                     | Branched chain amino acid | [65–67]      |
| Diabetes-associated cardiovascular disease |                           |              |
| No distinct markers                        | NA                        | [74]         |
| Diabetes-associated nephropathy            |                           |              |
| $\alpha$ -hydroxybutyric acid              | Lipid-related             | [76, 77]     |
| Asymmetric dimethylarginine                | Oxidation/reduction       | [54–59]      |
| Isoleucine                                 | Branched chain amino acid | [76, 77]     |
| Leucine                                    | Branched chain amino acid | [76, 77]     |
| Phenylalanine                              | Aromatic amino acid       | [76, 77]     |
| Triglycerides                              | Lipid-related             | [60]         |
| Tyrosine                                   | Aromatic amino acid       | [76, 77]     |
| Valine                                     | Branched chain amino acid | [76, 77]     |
| Diabetes-associated retinopathy            |                           |              |
| Ascorbic acid                              | Oxidation/reduction       | [78]         |
| Galactitol                                 | Sugar                     | [78]         |

hand, the levels of granulocyte-macrophage colony-stimulating factor, macrophage inflammatory protein-1 $\alpha$  and macrophage inflammatory protein-1 $\beta$  levels were elevated in micro-albuminuric patients in comparison to controls. Furthermore, the levels of interferon- $\gamma$ -inducible protein 10 and monocyte chemoattractant protein-1 were significantly correlated with the urinary albumin excretion and estimated glomerular filtration rates, and the levels of eotaxin, granulocyte-macrophage colony-stimulating factor, interferon- $\gamma$ -inducible protein 10, monocyte chemoattractant protein-1, and RANTES were correlated with glycated hemoglobin (HbA1c). These findings suggest that monitoring of cytokine levels in urine may be helpful for early diagnosis and treatment of patients with diabetic nephropathy.

#### **4.4 Biomarkers for Determining Risk of Diabetic Retinopathy**

A proteomics study using a two-dimensional difference gel electrophoresis (2D-DIGE) approach found 28 plasma protein differences between diabetic patients with and without retinopathy and these proteins were identified by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry [66]. The proteins which were altered in the patients with retinopathy were mainly involved in the inflammation response and the coagulation cascade (Table 1). This included identification of multiple proteins that had been reported previously as diabetic retinopathy markers, including  $\alpha$ 1-antitrypsin,  $\alpha$ 2-macroglobulin, antithrombin-III, apolipoprotein A1, apolipoprotein E, complement C1/C3, fibrinogen, haptoglobin, kininogen-1, serum amyloid P-component, and vitronectin. This demonstrated the consistency of these findings with previous studies. In addition, this study also identified potential novel diabetic retinopathy biomarkers such as afamin, gelsolin, leucine-rich alpha-2-glycoprotein, protein arginine N-methyltransferase 5 and vitamin D-binding protein. These studies are still in an early phase and awaiting validation testing using additional cohorts but could in the future be useful clinically. However, the finding that some of these had already been identified by previous investigations lends some support to the possibility that these proteins and their associated pathways may be involved in the onset and progression of diabetes-associated retinopathy.

One study assessed the levels of plasma cytokines in 59 diabetic patients and 19 nondiabetic controls using a multiplex immunoassay approach [67]. They found that the levels of monocyte chemoattractant protein-1, tumor necrosis factor- $\beta$ , soluble CD40 ligand, and the interleukins 6, 7, 9, 13, 15, 17 and soluble 2R $\alpha$  were increased significantly in diabetic patients compared to the controls. The authors also found that the plasma levels of tumor necrosis factor- $\alpha$  plasma were significantly higher in patients with diabetic retinopathy compared to patients without retinopathy. In contrast, the levels of Flt-3 ligand and interleukins 1Ra, 3, 5, and

12 (p40) were present at lower levels in the diabetes patients compared to the control group. Jin and colleagues performed a comprehensive proteomic analysis using a selective reaction monitoring mass spectrometry approach to discover biomarkers for diabetic retinopathy [68]. The authors targeted 96 vitreous humor proteins which were also expressed in plasma and found that a 4-plex panel comprised of apolipoprotein A4, complement C7, clusterin, and inter-alpha-trypsin inhibitor H2 could be used to distinguish diabetes patients with early stages of retinopathy from controls. A drawback of many of these studies is that they include samples that are collected from people who already have diabetes. In these situations it is not clear if the changes identified are simply a consequence of the disease. To identify biomarkers of future disease risk longitudinal prospective cohort studies are required so that samples are available from an individual both prior to and after development of type 2 diabetes.

---

## 5 Metabolomic Biomarkers

### **5.1 Metabolite Biomarkers for Prediction of Type 2 Diabetes**

Metabolomic studies have identified blood-based biomarkers which may have predictive value in type 2 diabetes (Table 2). For example, increased levels of branched-chain amino acids (leucine, isoleucine, and valine) have been identified by several different research groups and in some cases these changes were found to occur more than 10 years before clinical diagnosis of the disease [69–71]. It is possible that changes in these amino acids are involved in the pathogenesis of type 2 diabetes although similar changes have also been associated with insulin resistance which can occur in a variety of disorders [72]. Metabolomic studies have also identified elevated levels of circulating aromatic amino acids such as phenylalanine and tyrosine several years ahead of actual diagnosis [69, 70]. Only one amino acid, glycine, was found to be reduced before the onset of type 2 diabetes and this may be a result of increased gluconeogenesis [70, 73].

A large number of metabolomic profiling studies have also found increased levels of  $\alpha$ -hydroxybutyric acid up to 10 years before clinical presentation of the disease [74, 75]. This is thought to occur due to perturbed redox pathways resulting in increased oxidative stress and lipid oxidation [75, 76]. Several studies have also found serum or plasma changes in  $\beta$ -hydroxybutyrate, acetone and acetoacetate prior to overt manifestation of type 2 diabetes, and these changes are also likely to be associated with increased lipid oxidation [74]. Alterations in other circulating lipids have also been identified in metabolomic profiling studies, including elevated levels diacylphosphatidylcholine C32:1 [70] and decreased levels of sphingomyelin C16:1 [70], acyl-alkyl-phosphatidylcholine [70], lysophosphatidylcholine C18:2 [69, 70, 73], and linoleoyl-

glycerophosphocholine [71]. The latter molecule may be linked to diabetes as its normal function is to increase glucose-stimulated insulin secretion from pancreatic  $\beta$ -cells [75].

One of the most common means to assess predictive power of potential biomarker signatures is the application of receiver operating characteristic-area under the curve (ROC-AUC) statistics. Using this approach, one study found that a metabolomic signature could predict the development of diabetes with a similar ROC-AUC score as found using the Diabetes Risk Score, which is based on lifestyle, diet and anthropometric factors [77]. However, by combining a biomarker signature composed of fasting glucose, HbA1c, and metabolites with the Diabetes Risk Score, the predictive power could be increased by approximately 6% to a value of 0.912 (1.0=perfection) [69]. Another study found a ROC-AUC of 0.82 using a metabolomic signature for prediction of type 2 diabetes compared to a slightly lower score of 0.79 using only fasting glucose levels [74]. Although these improvements in predictive power are only marginal, they support the need for identification of further biomarker-based signatures associated with disease risk and highlight the potential of combining molecular, dietary, lifestyle, and anthropomorphic data into predictive algorithms.

## **5.2 Metabolite Biomarkers of Diabetes-Related Cardiovascular Disease**

One study attempted to identify metabolomic signatures associated with cardiovascular risk factors such as high blood pressure, liver complications and coronary heart disease in patients with type 2 diabetes [78]. Although the researchers were successful and found distinct metabolomic signatures between patients with these different comorbidities, no conclusions could be made given the absence of a control group with diabetes and no comorbidities. However, a number of metabolomic studies of risk factors for cardiovascular disorders have been carried out and these have identified several markers which also related to diabetes risk, such as branched chain and aromatic amino acids [79, 80]. It is clear for the scarcity of research in this particular aspect of type 2 diabetes that considerable further work is required. This is especially important in this particular case considering the devastating effects of heart disease throughout the world at the individual, societal, and economic levels.

## **5.3 Metabolite Biomarkers for Prediction of Diabetes-Related Renal Disease**

Another study screened urine metabolites of diabetic patients with and without chronic renal disease and this led to identification of 13 molecules that were present at significantly lower levels in those with the kidney disorder [81]. Most of these molecules were involved in mitochondrial function consistent with the possibility that impaired mitochondrial metabolism was an underlying factor in this comorbidity. In addition, a prospective metabolomic study identified increased levels of uremic solutes and acylcarnitines prior to end-stage kidney disease in type 2 diabetes patients, compared

to matched controls with no renal dysfunction [82]. The same study also found decreased levels of branched chain and aromatic amino acids and the derivative  $\alpha$ -hydroxybutyric acid, suggestive of increased mitochondrial amino acid  $\beta$ -oxidation. These findings were consistent with another metabolomic study in which plasma from diabetes patients with and without diabetic nephropathy was compared (Table 2) [83].

#### **5.4 Metabolite Biomarkers for Prediction of Diabetes-Related Retinopathy**

A metabolomic profiling study of vitreous humor carried out on diabetic retinopathy patients found reduced levels of galactitol and ascorbic acid (Table 2) [84]. However, the findings may have been limited by the fact that the control group used in this study was made up of patients with macular degeneration, rather than diabetic patients without retinopathy. Nevertheless, further studies using definitive disease and control cohorts should be carried out since the production of a predictive biomarker test and application to individual risk assessment has the potential of initiating early interventions which could improve patient outcomes.

---

## **6 Conclusions and Future Directions**

Novel research in the fields of proteomics and metabolomics has opened up new opportunities of identifying biomarkers for risk and progression of type 2 diabetes mellitus. Through profiling of blood-based biomarkers, new information about these markers and how they interact with age or lifestyle-related risk factors have now been identified. Further insights into these associations will lead to improvements in existing predictive biomarkers of type 2 diabetes and associated comorbidities and should therefore help to prevent or delay onset in individuals at high risk. Most of this research has revealed a tight link with disturbances in inflammatory response and redox potential in the disease itself as well as in the common comorbidities. However, considerable further work is required combining the use of the omic platforms in multicentre and longitudinal studies to provide specific biomarker panels that can distinguish the risk potential for each of these possibilities. Furthermore, translation of these biomarkers onto rapid and user-friendly platforms for point-of-care testing is essential to optimize potential clinical applications. This may include the lab-on-a-chip and mobile phone applications as described in the last few chapters of this book.

Future challenges include targeted research that will bring us closer to a better understanding of the link between molecular biomarkers and the link with physiometric or lifestyle-related factors at the level of the individual patient. This is important as biomarker testing has already established that algorithms that combine such different factors can lead to tests with greater performance, com-

pared to those using a single factor alone. Therefore, future research on biomarkers and type 2 diabetes has the potential to produce an abundance of specific biomarkers that might aid in the individually targeted prevention of the disease and the associated comorbidities. This is critical considering that this condition is so devastating to individual health as well as to society and the health-care services throughout the world.

## Acknowledgments

SEO is funded by the UK Medical Research Council (MC\_UU\_12012/4).

## References

1. <http://www.diabetesatlas.org/>
2. Steyn NP, Mann J, Bennett PH, Temple N, Zimmet P, Tuomilehto J et al (2004) Diet, nutrition and the prevention of type 2 diabetes. *Public Health Nutr* 7:147–165
3. Mattei J, Malik V, Wedick NM, Hu FB, Spiegelman D, Willett WC et al (2015) Reducing the global burden of type 2 diabetes by improving the quality of staple foods: the global nutrition and epidemiologic transition initiative. *Global Health* 11:23
4. Perseghin G, Ghosh S, Gerow K, Shulman GI (1997) Metabolic defects in lean non Diabetic offspring of NIDDM parents: a cross-sectional study. *Diabetes* 46:1001–1009
5. Henriksen JE, Levin K, Thye-Rønn P, Alford F, Hother-Nielsen O, Holst JJ et al. (2000) Glucose-mediated glucose disposal in insulin-resistant normoglycemic relatives of type 2 diabetic patients. *Diabetes* 49:1209–1218
6. Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P (2001) Prevention of type 2 diabetes mellitus by changes in life style among subjects with impaired glucose tolerance. *N Engl J Med* 344:1343–1350
7. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA et al (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 346:393–403
8. Diabetes Prevention Program Research Group, Knowler WC, Fowler SE, Hamman RF, Christophe CA, Hoffman HJ et al (2009) 10-year follow-up of diabetes incidence and weight loss in the diabetes prevention program outcomes study. *Lancet* 374:1677–1686
9. Hales CN, Barker DJ (1992) Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Diabetologia* 35:595–601
10. Nielsen JH, Haase TN, Jakobs C, Nalla A, Søstrup B, Nalla AA et al (2014) Impact of fetal and neonatal environment on beta cell function and development of diabetes. *Acta Obstet Gynecol Scand* 93:1109–1122
11. Vaag A, Brons C, Gillberg L, Hansen NS, Hjort L AGP et al (2014) Genetic, nongenetic and epigenetic risk determinants in developmental programming of type 2 diabetes. *Acta Obstet Gynecol Scand* 93:1099–1108
12. Coop A, Torsoni AS, Velloso L (2016) Mechanisms in endocrinology: metabolic and inflammatory pathways on the pathogenesis of type 2 diabetes. *Eur J Endocrinol* 174:R175–R187
13. DeFronzo RA (1992) Pathogenesis of type 2 (non-insulin dependent) diabetes mellitus: a balanced overview. *Diabetologia* 35:389–397
14. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert MF et al (2014) Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* 63:2158–2171
15. Domingueti CP, Dusse LM, Carvalho MD, de Sousa LP, Gomes KB, Fernandes AP (2016) Diabetes mellitus: the linkage between oxidative stress, inflammation, hypercoagulability and vascular complications. *J Diabetes Complications* 30:738–745

16. Ling C, Poulsen P, Simonsson S, Ronn T, Holmkvist J, Almgren P et al (2007) Genetic and epigenetic factors are associated with expression of respiratory chain component NDUFB6 in human skeletal muscle. *J Clin Invest* 117:3427–3435
17. Ronn T, Poulsen P, Hansson O, Holmkvist J, Almgren P, Nilsson P et al (2008) Age influences DNA methylation and gene expression of COX7A1 in human skeletal muscle. *Diabetologia* 51:1159–1168
18. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A et al (2012) Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci USA* 109:10522–10527
19. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K et al (2012) Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol* 13:R97.10
20. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S et al (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 49:359–367
21. Dayeh T, Volkov P, Salo S, Hall E, Nilsson E, Olsson AH et al (2014) Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet* 10:e1004160
22. Small KS, Hedman AK, Grundberg E, Nica AC, Thorleifsson G, Kong A et al (2007) Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet* 43:561–564
23. Johansson A, Enroth S, Gyllensten U (2013) Continuous aging of the human DNA methylome throughout the human lifespan. *PLoS One* 8:e67378
24. Steegenga WT, Boekschoten MV, Lute C, Hooiveld GJ, de Groot PJ, Morris TJ et al (2014) Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr)* 36:9648
25. Horvath S (2013) DNA methylation age of human tissues and cell types. *Genome Biol* 14:R115
26. Voight BF et al, MAGIC investigators, GIANT Consortium (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42: 579–589
27. Ozanne SE (2015) Epigenetics and metabolism in 2014: Metabolic programming--knowns, unknowns and possibilities. *Nat Rev Endocrinol* 11:67–68
28. Hidalgo B, Irvin MR, Sha J, Zhi D, Aslibekyan S, Absher D et al (2014) Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the genetics of lipid lowering drugs and diet network study. *Diabetes* 63:801–807
29. Yuan W, Xia Y, Bell CG, Yet I, Ferreira T, Ward KJ et al (2014) An integrated epigenomic analysis for type 2 diabetes susceptibility loci in monozygotic twins. *Nat Commun* 5:5719
30. Toporoff G, Aran D, Kark JD, Rosenberg M, Dubnikov T, Nissan B et al (2012) Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. *Hum Mol Genet* 21:371–383
31. Ferland-McCollough D, Ozanne SE, Siddle K, Willis AE, Bushell M (2010) The involvement of microRNAs in Type 2 diabetes. *Biochem Soc Trans* 38:1565–1567
32. Esguerra JL, Mollet IG, Salunkhe VA, Wendt A, Eliasson L (2014) Regulation of pancreatic beta cell stimulus-secretion coupling by microRNAs. *Genes (Basel)* 5:1018–1031
33. Esguerra JL, Eliasson L (2014) Functional implications of long non-coding RNAs in the pancreatic islets of Langerhans. *Front Genet* 5:209
34. Guay C, Regazzi R (2013) Circulating microRNAs as novel biomarkers for diabetes mellitus. *Nat Rev Endocrinol* 9:513–521
35. Fulton RJ, McDade RL, Smith PL, Kienker LJ, Kettman JR Jr (1997) Advanced multiplexed analysis with the FlowMetrix system. *Clin Chem* 43:1749–1756
36. Knowles MR, Cervino S, Skynner HA, Hunt SP, de Felipe C, Salim K et al (2003) Multiplex proteomic analysis by two-dimensional differential in-gel electrophoresis. *Proteomics* 3: 1162–1171
37. Oliveira BM, Coorssen JR, Martins-de-Souza D (2014) 2DE: the phoenix of proteomics. *J Proteomics* 104:140–150
38. Larsen MR, Roepstorff P (2000) Mass spectrometric identification of proteins and characterization of their post-translational modifications in proteome analysis. *Fresenius J Anal Chem* 366:677–690
39. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR et al (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 17:676–682
40. Griffin JL (2003) Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol* 7:648–654
41. Beckonert O, Keun HC, Ebbels TM, Bundy J, Holmes E, Lindon JC et al (2007) Metabolic profiling, metabolomic and metabonomic

- procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc* 2:2692–2703
42. Sundsten T, Eberhardson M, Göransson M, Bergsten P (2006) The use of proteomics in identifying differentially expressed serum proteins in humans with type 2 diabetes. *Proteome Sci* 4:22
  43. Riaz S, Alam SS, Akhtar MW (2010) Proteomic identification of human serum biomarkers in diabetes mellitus type 2. *J Pharm Biomed Anal* 51:1103–1107
  44. Bhonsle HS, Korwar AM, Chougale AD, Kote SS, Dhande NL, Shelgikar KM et al (2013) Proteomic study reveals downregulation of apolipoprotein A1 in plasma of poorly controlled diabetes: a pilot study. *Mol Med Rep* 7:495–498
  45. Mao P, Wang D (2014) Top-down proteomics of a drop of blood for diabetes monitoring. *J Proteome Res* 13:1560–1569
  46. Erion DM, Park HJ, Lee HY (2016) The role of lipids in the pathogenesis and treatment of type 2 diabetes and associated co-morbidities. *BMB Rep* 49(3):139–148
  47. Coccheri S (2007) Approaches to prevention of cardiovascular complications and events in diabetes mellitus. *Drugs* 67:997–1026
  48. Tuomilehto J, Lindstrom J, Qiao Q (2005) Strategies for the prevention of type 2 diabetes and cardiovascular disease. *Eur Heart J Suppl* 7:D18–D22
  49. Brunner EJ, Shipley MJ, Witte DR, Fuller JH, Marmot MG (2006) Relation between blood glucose and coronary mortality over 33 years in the Whitehall Study. *Diabetes Care* 29: 26–31
  50. Nissen SE, Wolski K (2007) Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 356:2457–2471
  51. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, Byington RP, Goff DC Jr, Bigger JT et al. (2008) Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med* 358:2545–2559
  52. Duckworth W, Abraira C, Moritz T, Reda D, Emanuele N, Reaven PD et al (2009) Glucose control and vascular complications in veterans with type 2 diabetes. *N Engl J Med* 360: 129–139
  53. FDA Guidance for Industry: Diabetes Mellitus — Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER); 2008
  54. Donahue MP, Rose K, Hochstrasser D, Vonderscher J, Grass P, Chibout SD et al (2006) Discovery of proteins related to coronary artery disease using industrial-scale proteomics analysis of pooled plasma. *Am Heart J* 152:478–485
  55. Gordois A, Scuffham P, Shearer A, Oglesby A (2004) The health care costs of diabetic nephropathy in the United States and the United Kingdom. *J Diabetes Complications* 18:18–26
  56. Mogensen CE, Neldam S, Tikkanen I, Oren S, Viskoper R, Watts RW et al (2000) Randomised controlled trial of dual blockade of renin-angiotensin system in patients with hypertension, microalbuminuria, and non-insulin dependent diabetes: the candesartan and lisinoprilmicroalbuminuria (CALM) study. *Br Med J* 321:1440–1444
  57. Ninomiya T, Perkovic V, de Galan BE, Zoungas S, Pillai A, Jardine M et al (2009) Albuminuria and kidney function independently predict cardiovascular and renal outcomes in diabetes. *J Am Soc Nephrol* 20:1813–1821
  58. van der Velde M, Halbesma N, de Charro FT, Bakker SJ, de Zeeuw D, de Jong PE et al (2009) Screening for albuminuria identifies individuals at increased renal risk. *J Am Soc Nephrol* 20:852–862
  59. Kazumi T, Hozumi T, Ishida Y, Ikeda Y, Kishi K, Hayakawa M et al (1999) Increased urinary transferrin excretion predicts microalbuminuria in patients with type 2 diabetes. *Diabetes Care* 22:1176–1180
  60. Narita T, Hosoba M, Kakei M, Ito S (2006) Increased urinary excretions of immunoglobulin G, ceruloplasmin, and transferrin predict development of microalbuminuria in patients with type 2 diabetes. *Diabetes Care* 29: 142–144
  61. Araki S, Haneda M, Koya D, Sugimoto T, Isshiki K, Chin-Kanasaki M et al (2007) Predictive impact of elevated serum level of IL-18 for early renal dysfunction in type 2 diabetes: an observational follow-up study. *Diabetologia* 50:867–873
  62. Hanai K, Babazono T, Nyumura I, Toya K, Tanaka N, Tanaka M et al (2009) Asymmetric dimethylarginine is closely associated with the development and progression of nephropathy in patients with type 2 diabetes. *Nephrol Dial Transplant* 24:1884–1888
  63. Persson F, Rossing P, Hovind P, Stehouwer CD, Schalkwijk CG, Tarnow L et al (2008) Endothelial dysfunction and inflammation predict development of diabetic nephropathy in

- the Irbesartan in patients with type 2 diabetes and microalbuminuria (IRMA 2) study. *Scand J Clin Lab Invest* 68:731–738
64. Stehouwer CD, Gall MA, Twisk JW, Knudsen E, Emeis JJ, Parving HH (2002) Increased urinary albumin excretion, endothelial dysfunction, and chronic low-grade inflammation in type 2 diabetes: progressive, interrelated, and independently associated with risk of death. *Diabetes* 51:1157–1165
65. Liu J, Zhao Z, Willcox MD, Xu B, Shi B (2010) Multiplex bead analysis of urinary cytokines of type 2 diabetic patients with normo- and microalbuminuria. *J Immunoassay Immunochem* 31:279–289
66. Lu CH, Lin ST, Chou HC, Lee YR, Chan HL (2012) Proteomic analysis of retinopathy-related plasma biomarkers in diabetic patients. *Arch Biochem Biophys* 529:146–156
67. Hang H, Yuan S, Yang Q, Yuan D, Liu Q (2014) Multiplex bead array assay of plasma cytokines in type 2 diabetes mellitus with diabetic retinopathy. *Mol Vis* 20:1137–1145
68. Jin J, Min H, Kim SJ, Oh S, Kim K, Yu HG et al (2016) Development of diagnostic biomarkers for detecting diabetic retinopathy at early stages using quantitative proteomics. *J Diabetes Res* 2016:6571976. doi:[10.1155/2016/6571976](https://doi.org/10.1155/2016/6571976), Epub 2015 Nov 9
69. Walford GA, Porneala BC, Dauriz M, Vassy JL, Cheng S, Rhee EP et al. (2014) Metabolite traits and genetic risk provide complementary information for the prediction of future type 2 diabetes. *Diabetes Care* 37:2508–2514
70. Floegel A, Stefan N, Yu Z, Mühlenbruch K, Drogan D, Joost HG et al (2013) Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* 62:639–648
71. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E et al (2011) Metabolite profiles and the risk of developing diabetes. *Nat Med* 17:448–453
72. Batch BC, Hyland K, Svetkey LP (2014) Branch chain amino acids: biomarkers of health and disease. *Curr Opin Clin Nutr Metab Care* 17:86–89
73. Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y et al (2012) Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* 8:615
74. Padberg I, Peter E, González-Maldonado S, Witt H, Mueller M, Weis T et al (2014) A new metabolomic signature in type-2 diabetes mel-
- litus and its pathophysiology. *PLoS One* 9:e85082
75. Ferrannini E, Natali A, Camasta S, Nannipieri M, Mari A, Adam KP et al (2013) Early metabolic markers of the development of dysglycemia and type 2 diabetes and their physiological significance. *Diabetes* 62:1730–1737
76. Gall WE, Beebe K, Lawton KA, Adam KP, Mitchell MW, Nakhe PJ et al (2010) Alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One* 5:e10883
77. Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Möhlig M et al (2007) An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 30:510–515
78. Wu T, Xie G, Ni Y, Liu T, Yang M, Wei H et al (2015) Serum metabolite signatures of type 2 diabetes mellitus complications. *J Proteome Res* 14:447–456
79. Shah SH, Sun JL, Stevens RD, Bain JR, Muehlbauer MJ, Pieper KS et al (2012) Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. *Am Heart J* 163:844–850
80. Würtz P, Havulinna AS, Soininen P, Tynkkynen T, Prieto-Merino D, Tillin T et al (2015) Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 131:774–785
81. Sharma K, Karl B, Mathew AV, Gangoiti JA, Wassel CL, Saito R et al (2013) Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease. *J Am Soc Nephrol* 24:1901–1912
82. Niewczas MA, Sirich TL, Mathew AV, Skupien J, Mohney RP, Warram JH et al (2014) Uremic solutes and risk of end-stage renal disease in type 2 diabetes: metabolomic study. *Kidney Int* 85:1214–1224
83. Huang M, Liang Q, Li P, Xia J, Wang Y, Hu P et al (2013) Biomarkers for early diagnosis of type 2 diabetic nephropathy: a study based on an integrated biomarker system. *Mol Biosyst* 9:2134–2141
84. Barba I, García-Ramírez M, Hernández C, Alonso MA, Masmiquel L, García-Dorado D et al (2010) Metabolic fingerprints of proliferative diabetic retinopathy: an  $^1\text{H}$ -NMR-based metabonomic approach using vitreous humor. *Invest Ophthalmol Vis Sci* 51:4416–4421

# Chapter 4

## LC-MS<sup>E</sup>, Multiplex MS/MS, Ion Mobility, and Label-Free Quantitation in Clinical Proteomics

Gustavo Henrique Martins Ferreira Souza, Paul C. Guest,  
and Daniel Martins-de-Souza

### Abstract

Proteomic tools can only be implemented in clinical settings if high-throughput, automated, sensitive, and accurate methods are developed. This has driven researchers to the edge of mass spectrometry (MS)-based proteomics capacity. Here we provide an overview of recent achievements in mass spectrometric technologies and instruments. This includes development of high and ultra definition-MS<sup>E</sup> (HDMS<sup>E</sup> and UDMS<sup>E</sup>) through implementation of ion mobility (IM) MS towards sensitive and accurate label-free proteomics using ultra performance liquid chromatography (UPLC). Label free UPLC-HDMS<sup>E</sup> is less expensive than labeled-based quantitative proteomics and has no limits regarding the number of samples that can be analyzed and compared, which is an important requirement for supporting clinical applications.

**Key words** Proteomics, Label-free, MS<sup>E</sup>, DIA, HDMS<sup>E</sup>, UDMS<sup>E</sup>, Data independent acquisition

---

### 1 Introduction

Proteomics has emerged as a promising field for biomarker identification in the post-genomic era. By the end of the 1990s, the combination of liquid chromatography and electrospray ionization mass spectrometry (ESI-MS) became the main tool for proteomic characterization. Technological advances in this area and the number of potential proteomic biomarkers have grown substantially in recent years, revealing a broader understanding of diseases. Considering the heterogeneity among proteomic datasets originating from these studies, scientists must now focus their efforts on validating this information, and establish the relevance of these proteins through applications in clinical practice [1, 2]. In order to achieve this, proteomic-based studies must first demonstrate improvements in design and standardized protocols to produce uniform and reliable results. Researchers should pay attention to pivotal details when defining patient cohorts, such as clinical stage, inclusion/exclusion criteria, sample quality, quality of controls,

sample preservation/storage methods, protocol complexity, reproducibility, and adequacy of statistical analyses. This is a daunting task considering as most clinical studies are complex and can sometimes require large numbers of samples/participants as well as extended data processing periods.

Although proteomics is in an early stage of exponential growth, the field currently offers extensive capabilities for elucidating functional protein interactions and for discovering novel biomarkers for therapeutics. In addition, proteomics can be used to determine the functional status of a disease and subsequently help to select treatment options. This approach has been termed pharmacoproteomics [3]. Despite sophisticated proteomic study designs using bottom-up shotgun proteomics, subsequent research has challenged the reliability or reproducibility of the results. These problems impact all of the “omics” fields and not just one technology [4]. This chapter describes some of the latest developments in the field of label-free MS approaches which may help to improve the discovery and validation of multiplex biomarkers for use in laboratory, preclinical and clinical studies.

---

## 2 Systems Biology

The use of quantitative high-throughput MS-based proteomics and comprehensive systems biology approaches has generated complex datasets which have the possibility to assist researchers in discovering the meaning of disease-related molecular changes, especially when combined with information generated by other omic approaches. In this context, clinical proteomics has the potential to enable early disease discovery through the development of multiplex assays that can aid clinical decisions. However, the term “biomarker” can be applied only to those that have been validated and approved following the rules of regulatory agencies. The discovery, validation and translation of biomarkers depend on pairing of the correct proteomic approach with a reliable study design. The throughput achieved in protein MS enables the discovery of new biomarkers in virtually all types of biological samples. Recent MS-based approaches have allowed the identification of new biomarkers not previously described with high sensitivity and confidence. Furthermore, the development of tools for *in silico* analysis of datasets allows for the interconnection of the identified proteins to increase our understanding their interactions in a systems biology manner. The completion of this cycle could lead to the identification of a biomarker with real clinical significance and one that discriminates between disease and healthy controls with higher confidence. Major diseases are not caused by the action of single genes, but rather by alterations in the functioning of a complex web of networks and pathways. Thus, measurement at the

level of proteome and metabolome presents new opportunities for both data integration and the potential translation of findings to medical practice. Although these strategies are useful in identifying disease-associated pathways or biomarkers, bench scientists and clinicians recognize the need to translate these methods and knowledge into clinical practice. However, the interdisciplinary nature of these studies and the lack of robust tools have impeded progress. Nevertheless, recent examples show that reliability of the data can be achieved and efforts are underway to improve translation of multiplex biomarkers into clinical practice.

---

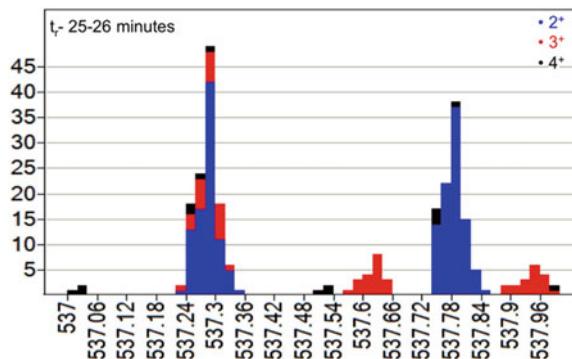
### 3 Recent Advances in MS-Based Proteomic Analysis Applied to Biomarker Discovery

Proteomics has evolved to focus on functionality of the huge data-sets acquired through a variety of analytical technologies. The quality of procedures, selectivity and specificity has recently received increased attention. The potential of protein analysis using complex samples such as chromosome-based, cell and biological fluid-based proteomics is only just now coming to fruition. Many researchers claim to have identified proteins but the lack of analytical quality and statistics is high. For example, the cell is a complex structure and working with cellular proteins remains challenging. Proteins are more difficult to understand than DNA or RNA. Proteins sometimes act alone, in one-to-one interactions or in large groups. If we want to understand cell behavior, we first need to understand proteins, taking them into consideration as an entire system. Consequently, the protein analysis problem is complex and the vast combination of the common and uncommon amino acids increases this complexity. This includes the fact that proteins can undergo a large range of enzymic and non-enzymic interactions, not all tissues express the proteins at the same time, and the same protein can be expressed at different levels in different tissues. In addition, not all proteins are expressed at the same concentration in the cell and the concentration of a given protein can change over time. A mass spectrometer is ideally suited to acquire sufficient data to cover the maximum dynamic range as needed. It has been noted that an increased dynamic range in the analytical instrument increases the chances of success due to access to proteins of lower abundance, especially in the label-free quantitation approaches. MS as a non-target approach relies on methods such as data dependent acquisition (DDA), data-independent acquisition (DIA), multiplexed  $\text{MS}^{\text{E}}$  /  $\text{HDMS}^{\text{E}}$  acquisitions, or selected reaction monitoring (SRM). These approaches assist in increasing the analytical capability of MS-base approaches and therefore allow more thorough analyses of samples in biomarker research.

## 4 Shotgun Proteomics: Complex Samples Data Acquisition and Molecular Ion Density

For complex samples, data-dependent acquisitions (DDA) are largely incapable of accommodating the chimeric and stochastic nature of the data and peptide mass sufficiency distribution (Fig. 1). Peptides with an overlapping mass-to-charge ratio ( $m/z$ ) and similar retention times make assigning more than one peptide per MS/MS spectrum far more difficult to interpret, even with application of complex algorithms [5]. However, it has been found that DDA and DIA product ion spectra have high quality similarity when one peptide is isolated at the selection window.

A significant source of error in proteomic experiments arises from the algorithmic interpretation of product ion spectra derived from the chimeric and composite MS spectra. In the case of complex mixture experiments, most search engines do not acknowledge the fact that a typical DDA product ion spectrum is most likely to arise from co-fragmented peptides. Approximately two-thirds of all precursor ion detections in a complex protein digest mixture are at least 2.5 orders of magnitude lower in intensity than the most abundant ions. Consequently, the incidence of overlapping isotopic clusters of similar  $m/z$  and intensity is significant. The specificity of DDA acquisitions is challenged under such conditions, especially when the search engine peptide score is based primarily on the intensity of the matched product ions relative to that of the unmatched ions. The ion density together with dynamic range and molecular weight distributions of proteins in a complex mixture, allow many analytes to be packaged in a small analytical



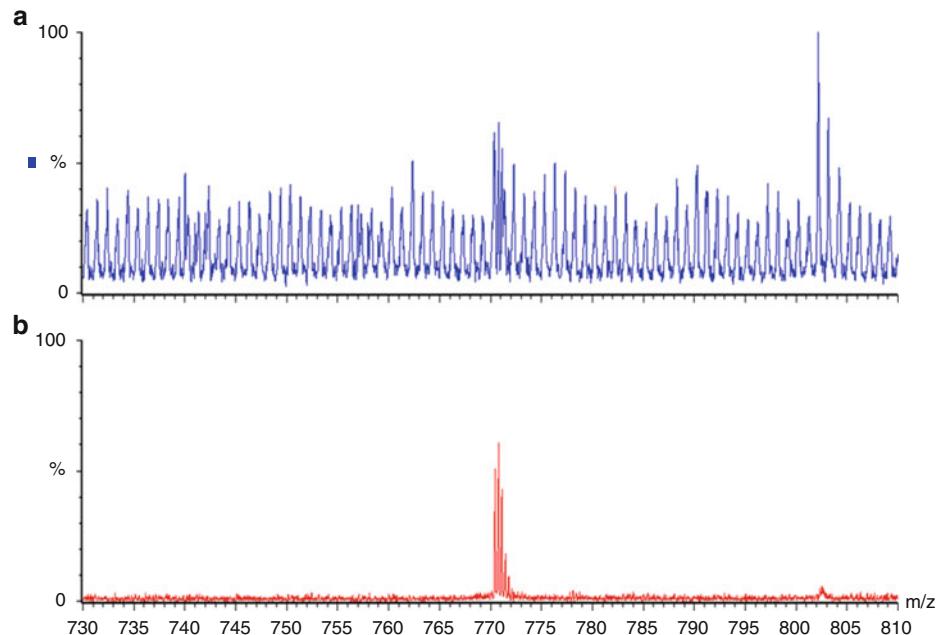
**Fig. 1** Mass sufficiency distribution—1 Da wide. The bar plot illustrates the ions of  $m/z$  537–538 assuming a mass resolving power of 25,000 FWHM at mass 500. There are 50 bins, each bin is 20 millidaltons wide and the colors indicate charge-state. Over 45 independent ions on  $m/z$  537.28  $\pm$  20 mDa within that 1 min of gradient elution are represented. For a given charge-state of an accurate mass measurement, an orthogonal approach such as ion mobility separation could help to increase selectivity

space, which includes both chromatography and mass analysis. Since these are unevenly distributed, areas of high analyte density can also suffer from issues such as the dynamic range of the sample being greater than the dynamic range of MS. Besides this, the data density can be followed by constricted  $m/z$  spaces where two or more peptides are fragmented at once (this is called the chimeracy effect). According to Michalski et al., up to 50% of total peptides are chimeric and this problem is not solved by speed or sensitivity. The DDA method plays a key role making analyses irreproducible due to the stochastic nature of DDA MS/MS [6]. In contrast, DIA methods are theoretically limited only by instrument peak capacity [7]. Acquiring DDA data more quickly or with higher sensitivity insignificantly reduces these sources of analytical mistakes. An increase in speed and sensitivity without a concurrent increase in specificity will generally produce compromised information by generating supplementary low abundance mixed mass spectra. Thus, modern acquisition techniques such as a multiplex high-resolution format MSE and high-definition HDMS<sup>E</sup> (drift time-conformation and charge differences) are recommended for proteomics and complex samples [8].

---

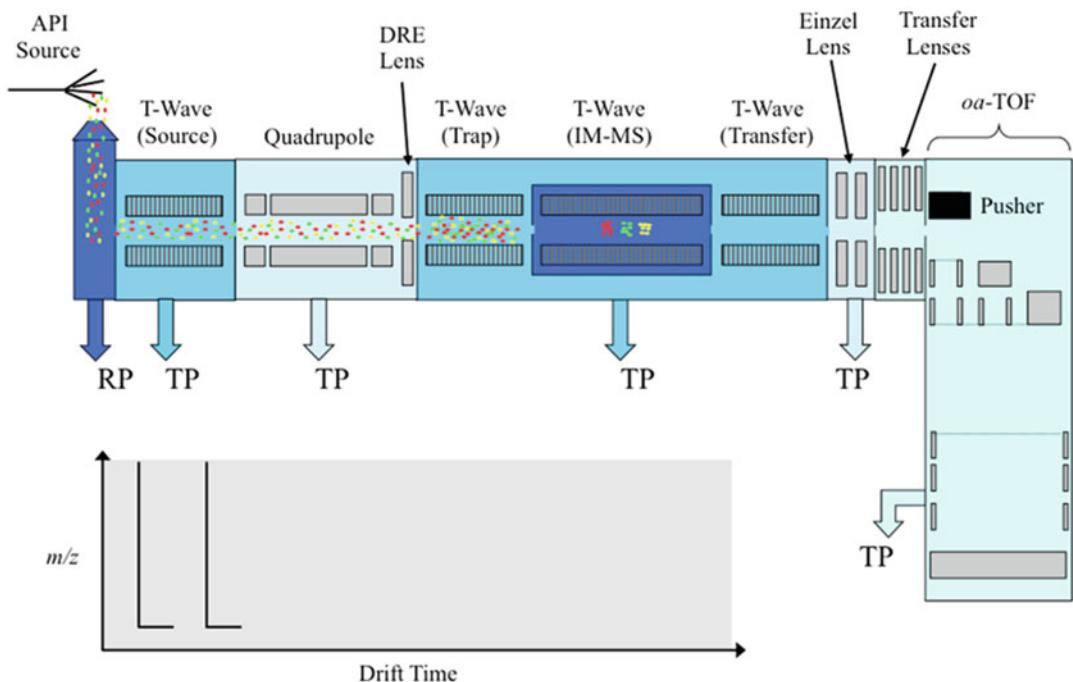
## 5 Ion Mobility: Conformation-Based Separations in Gas-Phase

Known as “ion chromatography” and “plasma chromatography,” ion-mobility spectrometry began as far back as 1930, when Tyndall and Powell published studies showing that different conformations could be discriminated in the gas phase due to the inference of DC and specific differential apertures capable of accelerating ions into a low-pressure cell [9, 10]. In 1932, Bradbury and Neilson described the first ion-gate that resulted in development of the ion trap prior to IM-MS separation [11, 12]. In 1967, the first commercial mass spectrometer was produced which was capable of producing an outstanding 1 Torr (1.33 mbar) on a drift cell device with a maximum pressure and this could be coupled to a time of flight (TOF) analyzer. In 1991, Von-Helden and collaborators described an ion-mobility mass spectrometer with drift-cell capable of handling up to 4 Torr and carried out the first experimental measurements of cross section values from different conformations and molecular species [13]. Next, Valentine and collaborators described development of the first ion-mobility MS with a dynamic accumulating ion trap upstream to increase duty-cycle. In 2001, an important article came out [14] describing separation of peptides from different charge states that dramatically expanded the applications for clinical and medical fields by making it applicable for both proteomic and metabolomic analyses. The first commercially available instrument was released in 2006 which had additional benefits such as improved accuracy and specificity due to the discrimination of s/n ratios (Fig. 2).



**Fig. 2** Improvements of signal-to-noise ratio (s/n). The MS spectra demonstrate the separation at gas phase from a peptide. (a) Standard MS spectrum. (b) Drift-time stripped Ion-Mobility based MS spectrum. Both datasets were acquired with a 100 attomoles of  $\beta$ -lactoglobulin nanoESI(+)

As mentioned above, the introduction of Synapt High Definition mass spectrometer in 2006 provided a new way to visualize mass spectrometric data and calculating molecular cross sections by combining the techniques of ultra performance liquid chromatography, ion mobility spectrometry with MS (Fig. 3) [15]. By definition, ion-mobility spectrometry is capable of separating gas-phase ionic species as they drift through a gas under positive pressure and under the influence of an electric field. The rate of the drift is proportional and dependent on the mobility through the gas. Mobility is driven based on mass, charge, conformation, gas pressure and polarizability, electrical field such as wave velocity of the drift cell (m/s) and height (V). Thus with the correct parameters, it is possible to discriminate different charges and isobaric species due to the specificity of different cross sections. For protein and peptide measurements, measuring the mobility of molecular ions can yield structural information, since small and compact ions drift quicker than large extended ones. Some classical limitations of ion-mobility separation in standard drift-tubes is low duty cycle, ions lost at the gate, short separation times such as 10–20 ms, diffusive losses that cause the ion cloud to be drawn out during separation and losses on the walls and the exit aperture before reaching the detector. Also, one of the main requirements was the need for a high voltage potential gradient across a long drift tube for efficient separation. Current ion-mobility devices are more effective as these use radiofrequencies (RF) applied over



**Fig. 3** Schematic of an instrument with ion mobility separation. The scheme demonstrates the drift time pusher extraction pulses and molecular ions being discriminated (red, green and yellow) by size, shape and charge. Arrival times of ions are recorded by synchronization of the *oa*-TOF acquisitions with gated release of ions from the Trap T-wave to the IM-MS. After the gate pulse, the subsequent 200 orthogonal acceleration pushes of the TOF analyzer are recorded giving an overall mobility recording time of:  $200 \times \text{tpp}$  where tpp is the pusher frequency. Following the next gated release of ions a further 200 mass spectra are acquired and added to the corresponding spectra from the previous acquisition. This process is repeated for a predefined period and subsequently 200 spectra are saved and the next summation period begins. RP Rotary Pump, TP Turbo-molecular pump, *oa*-TOF: orthogonal acceleration time-of-flight

the driving field extension at the electrodes, giving radial confinement and greater sensitivity by allowing low and high impact ion currents to transit across the electric field more efficiently.

## 6 Dynamic Range, Resolution, and Collision Energy

Advances in ion-mobility separation, conferring significantly increased peak capacity and peptide identification has shown that resolution is dependent only on  $m/z$  but also on collision cross sections and multiplex DIA acquisitions such as high and ultra definition- $\text{MS}^{\text{E}}$  (HDMS $^{\text{E}}$  and UDMS $^{\text{E}}$ ) [16–20]. More specifically the UDMS $^{\text{E}}$  approach allows all of the benefits of HD with a more effective collision energy control at the transfer cell. In HDMS $^{\text{E}}$ , the ions are separated based on mobility in the IM-MS cell via a traveling voltage wave and the precursors are subjected to fragmentation in the transfer cell normally through application of

ramped collision energies (eV) [21]. On the other hand in UDMS<sup>E</sup>, drift time-specific collision energies are used to fragment precursor ions prior to TOF-analysis. Ion mobility roughly correlates with *m/z* for all charge states. Thus, it is possible to assign quasi *m/z*-specific collision energies to the different drift time bins, increasing the intensity and MS/MS performance for peptide sequencing correlation.

Ion mobility approaches such as traveling wave ion mobility MS (TWIM-MS) are usually located inside the mass spectrometer [22]. The IM-MS located after the quadrupole and before a TOF analyzer increases selectivity since discrimination between charged molecules can be performed even for those that elute with the same retention times as well those with the same *m/z*, given different cross sections relative to conformation. For example, peptides from complex samples with the same constitutive amino acids but different amino acid positions will be orthogonally separated and discriminated into two or more different signals through drift time (dt). Until recently, separated ions were typically detected using a device called a multichannel plate (MCP). Briefly, this type of detector corresponds to two V-shaped aligned plates behaving as a fast-electron multiplier, such that a single ion event can result in  $1 \times 10^7$  electrons being produced over 4–5 ns. This event is followed by a time-to-digital converter (TDC) that is a positive chemical ionization (PCI)-based acquisition device. A high-impact ion flux caused by multi-charged ions such as those from peptides, strikes the front MCP, causing an electron shower within the plate that is then multiplied across a pair of plates. The voltage strike created on the anode is recorded by the TDC as an ion arrival event and this information is combined with TOF data, buffered and passed to the host embedded PC acquisition system (EPCAS), for processing and data-event recording [23]. This ion flux momentum over a single MS spectra-detection amplitude range corresponds to at least 3 logs of dynamic range ( $\log_{10\max}=3$ ). New detectors such as the ETP-MagneTOF and the hybrid analog-to-digital converter (ADC) combined with ion mobility dramatically enhances the duty-cycle and peak capacity, contributing to the accuracy of measuring ion arrival times in a mass spectrometer with a wider dynamic range ( $\log_{10\max}=4–6$ ). With the unbiased DIA acquisition with alternating low and high energy, the detection dynamic range over each sample can also be calculated. All of those features have now been combined into one instrument.

---

## 7 Quantitative Proteomics and Peak Capacity

Absolute quantitation can assist in understanding proteomic data [24]. For an LC-MS system, 1D and 2D nano chromatography and can be employed [25]. In 1D chromatography, a column with

an inner diameter of 75 µm can usually be loaded with up to 500 ng of total protein and this can be increased in accordance with the number of fractions. This helps to avoid saturation of the column and concomitantly increases identification of peptides due to enhanced peak-to-peak chromatographic resolution. The number of peptides identified in a protein should be proportional to the intact protein's molecular weight and the amount of that particular protein in the column, as described previously [26]. In order to address peak capacity and avoid charging effects, stacked-ion ring traveling-wave (T-wave) [22] devices have been used recently to increase the ion optics capability, thus enabling the amount of ions that the mass spectrometer can handle. In the example of "step wave" (SW) devices, the conjoined ion optics transmission T-wave helps to increase ion sensitivity, focusing and robustness. The SW function is based on stacked ring ion guides, which are designed to maximize ion transmission from the source to the mass analyzer. It also allows for active removal of neutral contaminants, enhancing the overall signal to noise ratio. The patented design enables efficient capture of the diffuse ion cloud entering the first stage, which is then focused into the upper ion guide for transfer to the mass analyzer. It actively extracts the ion beam into the upper stage and passes the high gas flow, excess solvent and neutral species to the exhaust. This protects the critical upper ion guide and the subsequent aperture from the direct line of sight of contaminants, ensuring that the methods remain robust for longer, even with complex matrices as in the case of omics samples.

In order to facilitate biomarker identification, MS discovery and validation platforms must be developed in parallel. The ion optics must be equal or superior to those in SRM analysis and this can be achieved with t-wave devices (such as the StepWave™ technology mentioned above). This has many applications besides ion mobility and ion optics and can be used as collision cells in trap collision-induced dissociation (CID) or electron transfer dissociation (ETD), as well as in transfer cells in CID that can be turned on individually or in groups [27]. As previously mentioned, hybrid ADC detectors dramatically enhance peak capacity and detection with a dynamic range of up to 4–6 logs. This target can be achieved in a straightforward manner for analysis of complex samples using high-resolution ion mobility instruments with a continuous ion current, low energy and high energy DIA multiplex MS/MS.

---

## 8 Data Quality Control in Label-Free Proteomics

Figures of merit (FOM) are based on how consistent a proteomics experiment can be graded in terms of consistency and robustness by the end of the study. At the peptide level, the following points may be considered:

- Percentage of missed cleavages.
  - Percentage of peptides that have fragmented in the source.
  - Mass accuracy (normality of distribution across 0 ppm).
  - Mass error distribution across  $m/z$  range.
  - Percentage of detected peptides at a maximum of 5 ppm.
- The following can also be considered at the protein level
- Average peptide per protein.
  - The calculated false positive/discovery rate (FPR/FDR).
  - Total number of proteins identified.
  - Dynamic range of the analysis.
  - Protein chosen to normalize data.
  - Coefficient of variation of all proteins and the chosen protein (spectra count or ion-accounting).

These considerations can assist in understanding the quality and integrity of MS data and allow choices to be made for further validation [6]. The FOM can also be enhanced using the increased power of identification algorithms, such as those based on ion accounting [28], and the inclusion of higher numbers of samples. Due to the complexity of the samples and low component concentrations, the data can sometimes be incomplete or can include interferences, and the assignment of data to peptides or clusters is often uncertain. Thus, reproducibility of the data is important when searching for a potential biomarker. The identification-based algorithms must rely not only on the mass accuracy of the precursors and fragment ions but also on other previously described physicochemical parameters. After a biomarker is chosen, a new method is developed and the process advances to the target analysis stage. Here, the focus is on measuring a target peptide and a related protein biomarker across several replicates and conditions. Recent instrument developments and LC-MS-based proteomics techniques have considerably improved the speed of analysis, depth of protein coverage and amount of information that can be obtained from complex biological sample mixtures. Despite these developments, variations in identification and quantification remain a concern, and alternative and complementary methods such as SRM are required. The value and use of a data-independent fragment ion repository has therefore been explored. In line with this, the required sensitivity and selectivity for the purpose of protein identification and quantification have been defined. Conceptually, the following requirements can be considered as guidelines for validation, hypothesis-driven studies or SRM method development [29]:

- A minimum of three peptides and three product ions should be identified for each protein (fragment selection is based on the highest replication rate and the smallest signature product ion variation across experiments and samples).
- Similar precursor and product ions should be selected from the second and third proteins that are consistently present with the protein of interest (together, these proteins outline a fragment ion signature).
- Unknown samples can be subsequently mapped against the fragmentation database signature to validate the presence of the target protein.

Various statistical and computational tools and methods are currently being tested using the fragment ion repository to facilitate the processes outlined above and to increase mathematical accounting information in the relational database [30]. Currently, the database can only be populated with qualitative results obtained through DIA experiments, specifically LC-MSE and/or LC-HDMS<sup>E</sup> approaches. DDA experiments generally do not yield precursor and product ion intensity measurements across the complete chromatographic peak or MS and MS/MS intensity recording for the same amount of time using the same gain. As spectral libraries and fragment ion repositories are more widely used in proteomics, the process will improve [31]. In addition to the identification of fragments, peptides, and proteins, data-independent fragment ion repositories have great potential in regard to the quantification of protein abundances, stoichiometry, and the reliable quantification of posttranslational modifications [32].

---

## 9 Applications of Label-Free Shotgun MS in Biomarker Studies

### 9.1 Analysis of Biofluids

In accordance with the need to identify biomarkers in clinically accessible biosamples, label-free shotgun MS profiling analyses have been applied routinely in the analysis of media which are obtainable from subjects with minimal invasiveness, such as urine, cerebrospinal fluid, sweat, serum, and plasma. This has been particularly useful in the study of cancer, considering the great need for such biomarkers to support clinical studies and the discovery of new drugs. One label-free profiling study investigated urine from prostate cancer patients in an attempt to identify urinary biomarkers [33]. Twenty potential biomarkers were identified in the prostate cancer patients with fold differences of more than two standard deviations observed for 17 proteins using an intensity-based absolute quantification approach. In a similar manner, Beretov et al. performed a comparative proteomic analysis using ion count relative quantification label free liquid chromatography

MS analysis and identified 59 urinary proteins that were present at significantly different levels in breast cancer compared to control females [34]. Some of these markers were associated with pre- or early-invasive phases and others were linked with metastatic stages of the cancer. Therefore, further work in this area could lead to identification of serum biomarkers associated with staging of breast cancer. In another study, Fan et al. carried out a high-performance liquid chromatography label-free MS profiling analysis of serum from colorectal cancer patients and healthy controls and found 69 proteins that were linked to the cancer [35]. Two of these proteins (macrophage mannose receptor 1 and S100A9) were validated as being increased by western blot analysis and could be used to distinguish colorectal cancer patients from controls with high accuracy through immunoassay analyses.

Studies of other diseases have also been performed such as comparative label-free liquid chromatography-tandem MS analysis of cerebrospinal fluid from patients with sporadic amyotrophic lateral sclerosis (ALS) [36]. This study identified several proteins which were significantly altered in ALS patients compared to patients with other neurological disorders and controls. These proteins were linked with diverse biological pathways such as inflammation, neuronal activity, and extracellular matrix regulation. The authors also used the identified profile to create a support vector machine classifier which could distinguish the ALS from the non-ALS individuals with 83% sensitivity and 100% specificity in an independent test set, and they validated four of the classifier proteins in an independent test set. Another report described a label-free MS profiling approach to provide information about the normal composition of human sweat, as a means of investigating this chemical barrier against pathogens on the surface of skin [37]. These researchers identified 95 proteins and 20 of these were novel. The most abundant proteins in this medium were dermcidin, apolipoprotein D, clusterin, prolactin-inducible protein, and serum albumin, which comprised more than 90% of secreted sweat proteins and these were validated using other multiplex biomarker profiling approaches. Another label-free MS investigation of serum showed that alpha-1-antitrypsin was the most discriminative protein found in patients with tuberculosis compared to controls [38]. In addition, Zhang et al. applied a similar shotgun profiling analysis of plasma which led to identification of three glycated albumin peptides which could be used for early diagnosis of type 2 diabetes mellitus [39].

Although psychiatric and neurological disorders have traditionally been considered to be diseases of the central nervous system, serum and plasma can be used to study such conditions. This is possible since the bloodstream contains a large number of molecules which are involved in linking peripheral and central functions such as mood, behavior, and memory. For example, one

group analyzed serum from depressed patients and control subjects using a label-free liquid chromatography-tandem MS profiling method and identified ten proteins that were present at significantly different levels in the depressed patients [40]. The differences in three of these proteins were validated by immunoassay and these were ceruloplasmin, inter-alpha-trypsin inhibitor heavy chain H4 and complement component 1qC, which were all increased in the depressed patients. Another investigation analyzed blood plasma samples of patients with mild and severe forms of cognitive impairment using a label-free shotgun proteomics approach [41]. This led to selection of a multivariate model which demonstrated an accuracy of 79% in predicting progressive cognitive impairment. Within the model, sex-specific protein biomarkers were also identified which included alpha-2-macroglobulin as correlating with progression to more severe forms of impairment in females only. Another study found sex-specific serum biomarker profiles in patients with Asperger syndrome using a label-free shotgun proteomics approach [42]. Twelve proteins were altered in Asperger syndrome females and one protein was altered in Asperger syndrome males. These results indicate that the search for biomarkers or novel drug targets in autism-related disorders may require stratification into male and female subgroups, which could lead to the production of new targeted approaches to treatment.

## **9.2 Analysis of Tissues**

Label-free liquid chromatography MS profiling has also been applied in the study of tissues in several distinct medical areas. As with biofluids, one of the largest areas of use in the analysis of tissue samples has been in the field of cancer. One investigation used a label-free MS profiling approach to study renal carcinoma tissue in comparison to normal tissue from the same patients [43]. The authors found changes in proteins indicative of mitochondrial dysfunction and cell death, along with perturbations in metabolism and acetylation. In a similar approach, Dai and coworkers analyzed proteins from matched pairs of human gastric cancer and adjacent tissues [44]. They found 146 proteins which were altered by more than twofold in the cancerous tissue. The changes in four nucleic acid binding proteins (heterogeneous nuclear ribonucleoprotein (hnRNP)A2B1, hnRNPD, hnRNPL, and Y-box binding protein 1) were validated by other multiplex biomarker techniques such as quantitative polymerase chain reaction and immunoassay. Other researchers have used the label-free method to study normal biological processes. In an investigation of the effects of aging, Theron and coworkers carried out a shotgun proteomic analysis of muscle tissue from mature and older females and found that 35 proteins were linked to aging in muscle and most of these showed decreased levels with age [45]. This included proteins involved in energy metabolism and myofilament and cytoskeleton functions.

One area of widespread use of the label-free shotgun proteomic profiling approach has been in the analysis of tissues associated with the central nervous system. This has mainly involved the study of psychiatric and neurological disorders. Martins-de Souza and coworkers presented the first characterization of the human occipital lobe (primary visual cortex) and cerebellum proteomes using a label-free shotgun MS profiling approach [46]. They identified proteins which have been associated previously with conditions such as neurological disorders, progressive motor neuropathy, Parkinson's disease, and schizophrenia. Therefore, these proteins may serve as biomarkers in the study of neurological processes and identify potential novel drug targets for therapeutic treatment approaches. Two label-free shotgun MS studies of postmortem pituitary tissues resulted in identification of candidate biomarkers for schizophrenia [47], bipolar disorder and major depressive disorder [48]. Schizophrenia patients had altered levels of pro-adrenocorticotrophic hormone, arginine vasopressin precursor, agouti-related protein, growth hormone, prolactin, and secretagogin. By comparison, bipolar disorder patients had significantly increased levels of pro-opiomelanocortin (POMC) and galanin, and major depressive disorder patients had significantly decreased levels of the prohormone-converting enzyme carboxypeptidase E and decreased activity of prolyl-oligopeptidase convertase. Given that the pituitary directly releases a variety of hormones and other bioactive molecules into the circulation, many of the proteins identified in the above study could serve as focal points in the search for peripheral biomarkers in clinical or drug treatment studies of psychiatric patients. Also in the field of psychiatric studies, Broek and colleagues carried out a label-free shotgun proteomic analysis of postmortem brain samples and found opposite directional changes in proteins which have a role in synaptic connectivity and myelination in the prefrontal cortex and cerebellum [49]. This is consistent with existing theories of the occurrence of altered structural and/or functional connectivity in the prefrontal cortex and cerebellum in autism patients. As with all biomarker-related studies, the findings of the above studies can only be considered as preliminary and require validation using larger independent cohorts of patient and control samples in follow up investigations.

---

## 10 Conclusions and Perspectives

In conclusion, repositories are a valuable addition to the requirements of system biology, not only allowing quantitative analysis of low-abundant proteins but also delivering reliable quantitative data when proteins are analyzed across multiple samples in multiple laboratories. When requiring more than one peptide for identification, the label-free approach gives superior information particu-

larly when coverage is taken into account. Both the label and label-free approaches can provide relative quantification datasets although the latter has advantages in terms of sample requirements, sample preparation and instrumental time. Due to significant sample complexity and wide concentration ranges, confidence in qualitative protein identifications and accuracy of the quantitative assessment of such proteins is imperative. A multiplexed acquisition approach such as label-free MS<sup>E</sup> can facilitate more protein identifications by offering greater selectivity and specificity while also providing the desired quantitative reproducibility for further validation and SRM/MRM acquisitions. Spectra databanks can increase the data integrity for quantitative proteomics discovery and reduce cost, sample size, instrument time, and sample preparation. This will increase confidence in identification and quantification, the average number per protein identified and reproducibility. Also will be most important for clinical applications and the use of these approaches in several areas of medicine has increased markedly over the last decade.

---

## Acknowledgments

Prof. Daniel Martins-de-Souza is supported by the São Paulo Research Foundation (FAPESP) grants 13/08711-3 and 14/10068-4 and by the Brazilian National Council for Scientific and Technological Development (CNPq) grant 460289/2014-4.

## References

1. Lee JM, Kohn EC (2010) Proteomics as a guiding tool for more effective personalized therapy. *Ann Oncol* 21:205–210
2. Tchabo NE, Liel MS, Kohn EC (2005) Applying proteomics in clinical trials: assessing the potential and practical limitations in ovarian cancer. *Am J Pharmacogenomics* 5:141–148
3. Wulfkuhle JD, Edmiston KH, Liotta LA, Petricoin EF 3rd (2006) Technology insight: pharmacoproteomics for cancer—promises of patient-tailored medicine using protein microarrays. *Nat Clin Pract Oncol* 3:256–268
4. Patel VJ, Thalassinos K, Slade SE, Connolly JB, Crombie A, Murrell JC et al (2009) A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J Proteome Res* 8:3752–3759
5. Michalski A, Cox J, Mann M (2011) More than 100,000 detectable peptides species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J Proteome Res* 10:1785–1793
6. Geromanos SJ, Hughes C, Golick D, Ciavarini S, Gorenstein MV, Richardson K et al (2011) Simulating and validating proteomics data and search results. *Proteomics* 11:1189–1211
7. Geromanos SJ, Vissers JP, Silva JC, Dorschel CA, Li GZ, Gorenstein MV et al (2009) The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependent LC-MS/MS. *Proteomics* 9:1683–1695
8. Chakraborty AB, Berger SJ, Gebler JC (2007) Use of an integrated MS-multiplexed MS/MS data acquisition strategy for high-coverage peptide mapping studies. *Rapid Commun Mass Spectrom* 21:730–744
9. Tyndall AM, Starr LH, Powell CF (1928) The mobility of ions in air. Part IV. Investigations by two new methods. *Proc R Soc Lond A* 121:172–184
10. Tyndall AM, Grindley GC, Sheppard PA (1928) The mobility of ions in air. Part V. The transformation of positive Ions at short ages. *Proc R Soc Lond A* 121:185–194

11. Bradbury NE (1932) Photoelectric currents in gases between parallel plates as a function of the potential difference. *Phys Rev* 40:980
12. Bradbury NE, Nielsen RA (1936) Absolute values of the electron mobility in hydrogen. *Phys Rev* 49:388–393
13. von Helden G, Hsu M-T, Kemper PR, Bowers MT (1991) Structures of carbon cluster ions from 3 to 60 atoms: linears to rings to fullerenes. *J Chem Phys* 95:3835–3837
14. Valentine SJ, Kulchania M, Srebalus Barnes CA, Clemmer DE (2001) Multidimensional separations of complex peptide mixtures: a combined high-performance liquid chromatography/ion mobility/time-of-flight mass spectrometry approach. *Int J Mass Spectrom* 212:97–109
15. Ruotolo BT, Giles K, Campuzano I, Sandercock AM, Bateman RH, Robinson CV (2005) Evidence for macromolecular protein rings in the absence of bulk water. *Science* 310:1658–1661
16. Giles K, Williams JP, Campuzano I (2011) Enhancements in travelling wave ion mobility resolution. *Rapid Commun Mass Spectrom* 25:1559–1566
17. Ruotolo BT, Benesch JL, Sandercock AM, Hyung SJ, Robinson CV (2008) Ion mobility-mass spectrometry analysis of large protein complexes. *Nat Protoc* 3:1139–1152
18. Lalli PM, Corilo YE, de Sa GF, Daroda RJ, de Souza V, Souza GHMF et al (2011) Intrinsic mobility of gaseous cationic and anionic aggregates of ionic liquids. *ChemPhysChem* 12:1444–1447
19. Lalli PM, Corilo YE, Fasciotti M, Riccio MF, de Sa GF, Daroda RJ et al (2013) Baseline resolution of isomers by traveling wave ion mobility mass spectrometry: investigating the effects of polarizable drift gases and ionic charge distribution. *J Mass Spectrom* 48:989–997
20. Distler U, Kuharev J, Navarro P, Levin Y, Schild H, Tenzer S (2014) Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat Methods* 11:167–170
21. Distler U, Kuharev J, Navarro P, Tenzer S (2016) Label-free quantification in ion mobility-enhanced data-independent acquisition proteomics. *Nat Protoc* 11:795–812
22. Giles K, Pringle SD, Worthington KR, Little D, Wildgoose JL, Bateman RH (2004) Applications of a travelling wave-based ratio-frequency-only stacked ring ion guide. *Rapid Commun Mass Spectrom* 18:2401–2414
23. Morris HR, Paxton T, Dell A, Langhorne J, Berg M, Bordoli RS et al (1996) High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom* 10:889–896
24. Yu YQ, Gilar M, Lee PJ, Bouvier ES, Gebler JC (2003) Enzyme-friendly, mass spectrometry-compatible surfactant for ion-solution enzymatic digestion of proteins. *Anal Chem* 75:6023–6028
25. Gilar M, Olivova P, Daly AE, Gebler JC (2005) Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J Sep Sci* 28:1694–1703
26. Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ (2006) Absolute quantification of proteins by LCMS<sup>E</sup>: a virtue of parallel MS acquisition. *Mol Cell Proteomics* 5:144–156
27. Williams JP, Brown JM, Campuzano I, Sadler PJ (2010) Identifying drug metallation sites on peptides using electron transfer dissociation (ETD), collision induced dissociation (CID) and ion mobility-mass spectrometry (IM-MS). *Chem Commun* 46:5458–5460
28. Li GZ, Vissers JP, Silva JC, Golick D, Gorenstein MV, Geromanos SJ (2009) Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* 9:1696–1719
29. Turtoi A, Mazzucchelli GD, De Pauw E (2010) Isotope coded protein label quantification of serum proteins—comparison with the label-free LC-MS and validation using the MRM approach. *Talanta* 80:1487–1495
30. Lucas JE, Thompson JW, Dubois LG, McCarthy J, Tillmann H, Thompson A et al (2012) Metaprotein expression modeling for label-free quantitative proteomics. *BMC Bioinformatics* 13:74
31. Benjamin AM, Thompson JW, Soderblom EJ, Geromanos SJ, Henao R, Kraus VB et al (2013) A flexible statistical model for alignment of label-free proteomics data—incorporating ion mobility and product ion information. *BMC Bioinformatics* 14:364
32. Thalassinos K, Vissers JP, Tenzer S, Levin Y, Thompson JW, Daniel D et al (2012) Design and application of a data-independent precursor and product ion repository. *J Am Soc Mass Spectrom* 23:1808–1820
33. Adeola HA, Soares NC, Paccez JD, Kaestner L, Blackburn JM, Zerbini LF (2015) Discovery of novel candidate urinary protein biomarkers for prostate cancer in a multiethnic cohort of South African patients via label-free mass spectrometry. *Proteomics Clin Appl* 9:597–609

34. Beretov J, Wasinger VC, Millar EK, Schwartz P, Graham PH, Li Y (2015) Proteomic analysis of urine to identify breast cancer biomarker candidates using a label-free LC-MS/MS approach. *PLoS One* 10:e0141876
35. Fan NJ, Chen HM, Song W, Zhang ZY, Zhang MD, Feng LY, Gao CF (2016) Macrophage mannose receptor 1 and S100A9 were identified as serum diagnostic biomarkers for colorectal cancer through a label-free quantitative proteomic analysis. *Cancer Biomark* 16:235–243
36. Collins MA, An J, Hood BL, Conrads TP, Bowser RP (2015) Label-free LC-MS/MS proteomic analysis of cerebrospinal fluid identifies protein/pathway alterations and candidate biomarkers for amyotrophic lateral sclerosis. *J Proteome Res* 14:4486–4501
37. Csősz É, Emri G, Kalló G, Tsaprailis G, Tőzsér J (2015) Highly abundant defense proteins in human sweat as revealed by targeted proteomics and label-free quantification mass spectrometry. *J Eur Acad Dermatol Venereol* 29:2024–2031
38. Song SH, Han M, Choi YS, Dan KS, Yang MG, Song J (2014) Proteomic profiling of serum from patients with tuberculosis. *Ann Lab Med* 34:345–353
39. Zhang M, Xu W, Deng Y (2013) A new strategy for early diagnosis of type 2 diabetes by standard-free, label-free LC-MS/MS quantification of glycated peptides. *Diabetes* 62:3936–3942
40. Lee J, Joo EJ, Lim HJ, Park JM, Lee KY, Park A et al (2015) Proteomic analysis of serum from patients with major depressive disorder to compare their depressive and remission statuses. *Psychiatry Investig* 1:249–259
41. Yang H, Lyutvinskiy Y, Herukka SK, Soininen H, Rutishauser D, Zubarev RA (2014) Prognostic polypeptide blood plasma biomarkers of Alzheimer's disease progression. *J Alzheimers Dis* 40:659–666
42. Steeb H, Ramsey JM, Guest PC, Stocki P, Cooper JD, Rahmoune H et al (2014) Serum proteomic analysis identifies sex-specific differences in lipid metabolism and inflammation profiles in adults diagnosed with Asperger syndrome. *Mol Autism* 5:4
43. Zhao Z, Wu F, Ding S, Sun L, Liu Z, Ding K, Lu J (2015) Label-free quantitative proteomic analysis reveals potential biomarkers and pathways in renal cell carcinoma. *Tumour Biol* 36:939–951
44. Dai P, Wang Q, Wang W, Jing R, Wang W, Wang F et al (2016) Unraveling molecular differences of gastric cancer by label-free quantitative proteomics analysis. *Int J Mol Sci* 17(1):E69. doi:10.3390/ijms17010069, pii
45. Théron L, Gueugneau M, Coudy C, Viala D, Bijlsma A, Butler-Browne G et al (2014) Label-free quantitative protein profiling of vastuslateralis muscle during human aging. *Mol Cell Proteomics* 13:283–294
46. Martins-de-Souza D, Guest PC, Guest FL, Bauder C, Rahmoune H, Pietsch S et al (2012) Characterization of the human primary visual cortex and cerebellum proteomes using shotgun mass spectrometry-data-independent analyses. *Proteomics* 12:500–504
47. Krishnamurthy D, Harris LW, Levin Y, Koutroukides TA, Rahmoune H, Pietsch S et al (2013) Metabolic, hormonal and stress-related molecular changes in post-mortem pituitary glands from schizophrenia subjects. *World J Biol Psychiatry* 14:478–489
48. Stelzhammer V, Alsaif M, Chan MK, Rahmoune H, Steeb H, Guest PC et al (2014) Distinct proteomic profiles in post-mortem pituitary glands from bipolar disorder and major depressive disorder patients. *J Psychiatr Res* 60:40–48
49. Broek JA, Guest PC, Rahmoune H, Bahn S (2014) Proteomic analysis of post mortem brain tissue from autism patients: evidence for opposite changes in prefrontal cortex and cerebellum in synaptic connectivity-related proteins. *Mol Autism* 5:41

# **Chapter 5**

## **Phenotyping Multiple Subsets of Immune Cells *In Situ* in FFPE Tissue Sections: An Overview of Methodologies**

**James R. Mansfield**

### **Abstract**

The recent clinical success of new cancer immunotherapy agents and methods is driving the need to understand the role of immune cells in solid tissues, especially tumors. Immune cell phenotyping via flow cytometry, while a cornerstone of immunology, is not spatially resolved and cannot analyze immune cell subsets *in situ* in clinical biopsy sections or to determine their interrelationships. To address this problem, a number of methodologies have been developed in attempts to phenotype immune and other cells in images acquired from tissue sections and to assess their organization in the tumor and its microenvironment. This chapter review the staining and multiplex image analysis methods that have been developed for phenotyping immune and other cells in formalin-fixed, paraffin-embedded (FFPE) tissue sections.

**Key words** Immunotherapy, Cancer, Tumor, Immune cells, Multiplex imaging

---

### **1 Introduction**

The growth in cancer immunotherapy and immunology as treatment methodology and paradigm for cancer research, respectively, is unprecedented. From its humble beginnings over a century ago through the work of William B. Coley to an increasing rate of US Food and Drug Association (FDA) approvals for cancer immunotherapies across several cancer types, cancer immunotherapy has taken its place as one of the five pillars of cancer therapy, along with surgery, radiotherapy, chemotherapy, and targeted therapies. It was chosen as the Breakthrough of the Year by Science Magazine in 2013 and continues to play a central role at major cancer conferences such as the American Association of Cancer Researchers (AACR) or the American Society for Clinical Oncology (ASCO), with many of the advances being treatment of otherwise untreatable solid tumors. Many excellent reviews on the subject have been written [1–4] and immune-oncology and the tumor microenvironment are now included as a part of the hallmarks of cancer [1].

Given the new importance of the role of the immune system in solid-tumor oncology, and its continued importance in many other diseases, it is vital to be able to understand the specific role each different type of immune cells plays *in situ* in the tumor and its microenvironment, something that is still practically challenging. There have been decades of research on the different roles of each type of immune cell [1], and many thousands of immune cell types have been described. Flow cytometry has been a key technology used in immunology and has been employed for decades as one of the primary tools used to differentiate the various immune cells [2]. The technique relies on the multiplex fluorescence labeling of cells followed by passing them one at a time through a flow chamber where the fluorescent signal of each marker is measured [3]. The determination of the phenotype of each cell is made through a process of “gating,” or separating the various cell types based on the fluorescence intensity of each marker or marker combination. The markers used in flow cytometry are typically cell membrane-bound antigens and it is the differential expression of these cell surface markers that defines the phenotype of the cell [4].

However, flow cytometry cannot give any information about the distribution of the various cells in solid tissues. In order to be analyzed by flow cytometry, a solid tissue must first be disaggregated and then the cells passed through the flow cytometer. In the field of solid-tumor oncology, it is critical to be able to visualize the phenotypic distributions of the various immune and other cells and to investigate how these relate to the tumor and its microenvironment. This requires being able to obtain the same kind of phenotypic information that is already well understood from flow cytometry of individual cells but to do so on cells still *in situ*. It is also important to be able to obtain an image of the cells in the context of the tissue architecture so that the spatial relationships of those cells can be viewed and evaluated in some manner.

Traditional pathology methods for the visualization of cellular phenotypes *in situ* utilize a one-marker-at-a-time approach in thin formalin-fixed, paraffin-embedded (FFPE) sections, typically using 3,3'-diaminobenzidine (DAB) to stain for a single protein marker and hematoxylin as a counterstain to display cell nuclei and some tissue architecture. This is a useful approach as long as a single protein is enough to accomplish the objectives. However, for cancer immunology and immunotherapy research this is typically insufficient as the phenotypes in question are those from flow cytometric analyses and can require knowing the expression of several, if not dozens, of proteins simultaneously in a single cell. A good example of this is the flow cytometric definition of what constitutes a regulatory T cell (CD3+, CD4+, CD25 high, and FOXP3+), something beyond a one-marker-at-a-time approach. It is this requirement to be able to measure the level of multiple proteins in a specific cell, while at the same time seeing the cell *in situ* in the context of the other cells and tissues in a sample, which is driving the development of multiplexed imaging methods.

This chapter covers the range of technologies that have been developed to try to determine complex phenotypes of cells *in situ* in FFPE tissue sections. A wide range of technologies have been applied to this problem, which will be grouped into two major categories: (1) simultaneous acquisition and (2) sequential acquisition of imagery. In simultaneous acquisition, all of the markers are present in the sample at one time and the images or data are collected from the sample in one pass. Since there are a number of technologies which are simultaneous, this section will be subdivided into staining, imaging and image analysis subsections. For sequential acquisition, markers are present in small groups (typically 2–4), imagery is acquired and the sample cleared or wiped of the first group of markers to prepare for the next group. The final images or data require registering the resulting set into one larger dataset.

---

## 2 Simultaneous Acquisition Methods

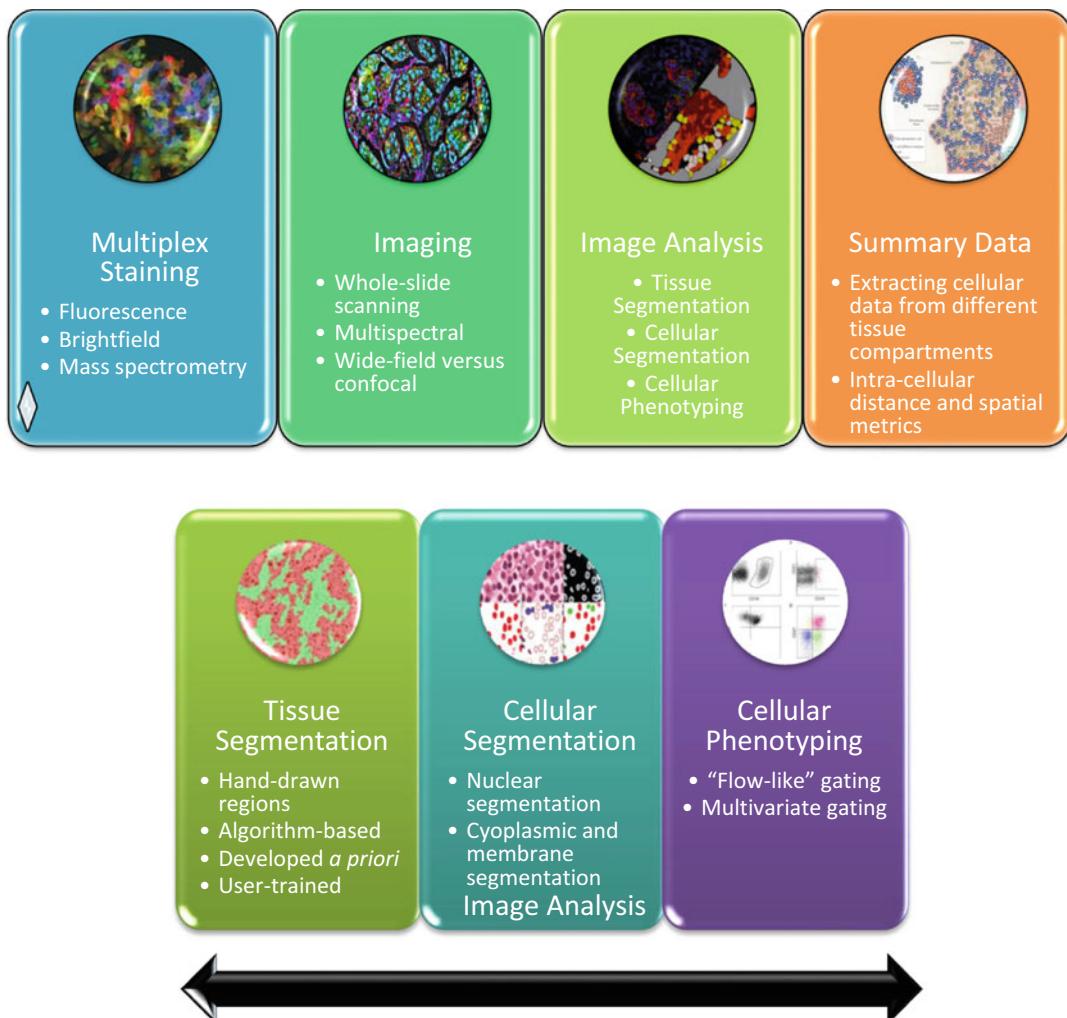
In simultaneous acquisition methods, all of the markers on which the per-cell analyses will be based are applied onto the tissue section prior to imaging and then an appropriate imaging method acquires data from all markers in a single step. The multimarker staining is typically done via some form of immunohistochemistry, using antibodies which bind to specific antigens in the tissue section, followed by the application of some form of marker (chromogenic, fluorescent or metal isotopes) which will then be imaged using an appropriate technology. Choosing the most appropriate staining method depends on the number of markers one wishes to image.

For a small number of markers (three or fewer), chromogenic methods are typically easiest to implement and best matched to the types of samples used most often in pathology, although fluorescence can also be used. For a medium number of markers [4–12], fluorescence is typically easiest to implement, although there are examples of chromogenic staining being used for up to 6 markers. Chromogenically and fluorescently labeled samples can each be imaged using commercially available optical imaging systems. For large numbers of markers (6 or more), there have been a number of publications describing the use of using metal isotopes to label the sample and then some form of mass spectrometry imaging to acquire the data. Mass spectrometry imaging can theoretically assess more than 100 markers simultaneously, although the practical limit for staining has typically been around 40.

In addition to the markers required for each antigen one wishes to label, a method of staining the nucleus is also needed since downstream image analysis methods all rely on using nuclei to find cells. For chromogenic methods, nuclear counterstaining is typically carried out in a blue color using hematoxylin, although there are other colors/reagents available (methyl green, nuclear fast red, etc.). For fluorescence methods, 4',6-diamidino-2-phenylindole

(DAPI) is typically used as the nuclear counterstain, although other colors/regents are also available (Hoechst, Draq5, etc.). For mass spectrometry imaging, one typically needs to use an antibody for a nuclear protein to delineate the nucleus. It should be noted that using an antibody to demarcate the nucleus would be a problem for chromogenic or fluorescent methods as only a limited number of markers can be applied. However, in cases where it is possible to use up to 40 markers, as is the case for mass spectrometry imaging, using one of these analytes for nuclear counterstaining is feasible.

Figure 1 shows a schematic of sequential methods. The top panel shows the overview of the staining, imaging, image analysis and summary data analyses that are typically performed for this



**Fig. 1** The top panel shows a schematic of the workflow for a simultaneously labeled sample, involving staining, imaging, image analysis, and summary data analysis. The image analysis portion is further broken out into the lower panel, showing tissue segmentation, cellular segmentation, and cellular phenotyping

kind of assay. The lower panel shows a further break down of the image analysis step into tissue segmentation, cellular segmentation and cellular phenotyping.

---

### 3 Staining and Imaging: Chromogenic

The simplest approach to visualize or quantitate specific cell types *in situ* is to use standard chromogenic methods. The most common method is to use a DAB chromogen as substrate for an immunoenzyme reaction. This deposits a brown, alcohol- and water-insoluble precipitate at the site of enzymatic activity. A primary antibody for the protein of interest is placed onto the slide and then a secondary antibody that is selective for the species or type of the primary antibody is applied. The secondary antibody is typically conjugate to an enzyme such as horseradish peroxidase (although others are used), such that the immunoreaction marks the distribution of the protein with the chromogen. By choosing antibodies for proteins that are specific for a particular cell type, one can delineate specific cell phenotypes, such as CD4 for helper T cells, CD8 for cytotoxic T cells and cytokeratin for epithelial cells. A blue hematoxylin counterstain is commonly used to highlight the nuclei in the sample and to enable visualization of the tissue architecture. The vast majority of immunohistochemistry in pathology is done using DAB as a marker and hematoxylin as a counterstain, considering the utility of this approach. It is easily assessed visually, either through the oculars of a microscope, or by using a color brightfield digital slide scanner and viewing the resulting image on a computer screen. For samples which have been digitized into an image, it is possible to perform analysis to quantitate the protein distribution, either through counting positive cells or determining the amount of chromogen present in each cell or subcellular structure.

Despite its ease of use and ubiquity, DAB has some well-known limitations. Although it is possible to measure the intensity of the DAB staining and quantitate protein expression on a per-cell basis, the dynamic range of expression levels that can be measured using DAB is less than that achieved with fluorescence [5]. DAB is also problematic for tissues that have natural brown pigments, such as skin or other epithelial tissues that contain melanin. In those cases, a red chromogenic substrate can be used. However, the main limitation of DAB is that it can only mark the distribution of a single protein, which can make differentiating between complex cell types difficult or impossible. To assess the distributions of more than one protein in a tissue, a commonly used method is to stain serial sections of a specimen using different antibodies and then interpolating between the sections to get an estimate of the inter-distribution of the proteins. This works well on a gross morphology level.

However, flow cytometry and other techniques have given researchers a wide range of cellular phenotypes that are important, particularly in immunology, and understanding these phenotypes requires using more than one marker per individual cell. By taking serial sections, one has to rely on a particular cell being split between the two sections, with one portion in one section and another portion in the sequential section, and then finding these cells and making the connection between the halves. This is a difficult task and only works for two markers at a time. Many immune cells are smaller than the thickness of a tissue section and are not split across sections. To assess these cells, the markers need to be deposited on the same microscope slide and in the same cell using a multiplex chromogenic approach.

There are many methods of performing multiplex chromogenic immunohistochemistry, dating back to the 1970s. It is beyond the scope of this chapter to describe all of these, but there are a number of excellent review articles [6–8]. Two of the more common and useful methods will be described here.

### **3.1 Double for Dummies**

This was so named by the original author because it is the simplest means of performing double staining and can be done using any pair of different species antibodies (e.g., one mouse and one rabbit antibody) [9]. The two primary antibodies are put on the sample together, either in a cocktail or individually. The markers are visualized using two separate processes for chromogen deposition, first with a secondary antibody for one species (e.g., anti-mouse) and one color, and then using a secondary antibody for the other species (e.g., anti-rabbit) and a different color chromogen. This has the advantage of not requiring any extra heat steps to strip previously used primary and secondary antibodies. However, it does require having primary antibodies of different species, which can be difficult for some marker combinations for which monoclonal rabbit antibodies may be the only ones obtainable.

### **3.2 Sequential Staining Using Antibody Stripping**

This is probably the most difficult of the methods, but it is the most generalizable and can be used for up to six different antibodies of the same species. A simple version is a universal type of sequential double alkaline phosphatase staining. This includes visualization in red and blue, and any cross-reaction between the two staining sequences is abolished by a heat-induced epitope retrieval step. This removes both primary and secondary antibodies but does not generally affect the quality of subsequent steps [10, 11]. This can be extended to use of a larger number of colors, with three immuno-markers being practical and 4–6 being maximal [12–14].

Of course there are many other methods for multiplex immunohistochemistry. For a more detailed discussion, the reader is encouraged to review the work of the late Chris van der Loos, who was widely acknowledged to be the leader in this field [15, 16].

---

## 4 Imaging Considerations in Brightfield

If all that is required is a photograph of a sample, then any color camera on a microscope can be used to take an image. Care should be taken for white balancing of the image and for flat-fielding (removing the observed intensity differences in the white light distribution across the sample), but these are optional as long as the image quality is good. However, if one wishes to perform some kind of analysis on the image, more care must be taken to provide quantitative data. Color balancing and image flat-fielding are standard on all whole-slide scanning systems used in digital pathology, so the use of those kinds of imaging systems is preferable to employing regular microscopes and cameras.

### 4.1 Optical Density (OD) Conversion

The immune-enzyme methods used in chromogenic staining deposit an amount of dye that is proportionate to the amount of antigen present in the sample [11]. According to the Beer–Lambert law, the amount of color measured in an image should translate to the amount of antigen present in each pixel of an image [17, 18]. To be able to apply the Beer–Lambert relationship, the light intensities must first be converted to absorbance (A) or optical density (OD) units, after which the absorbance of the sample will be proportional to the amount of antigen present [19]. This can only be a relative quantitation unless there is a reference standard or series of standards that can be used to calibrate the absorbance, which has proved elusive in FFPE sections. Despite this limitation, relative quantitation of antigen intensity within an image, sample or between samples can be useful.

### 4.2 Color Versus Multispectral Imaging

The quantitative imaging of a chromogenic sample stained singly with either DAB or Fast Red, for example, and then counter-stained with hematoxylin, requires only a color imaging system. In general, the quantitative separation of co-localized colors requires at least one more wavelength than the number of colors being imaged [20]. Therefore, separating brown and blue using a three-wavelength RGB system is possible and all standard digital pathology systems are capable of acquiring images which can be used for this. However, if there are more than two co-localized colors, more wavelengths are required which necessitates using a multispectral approach [21, 22]. Although quantitative color separation can be useful for the visualization of multiple markers in a single sample, the most important aspect of color separation is for the quantitation of color intensity. If two or more colors are co-localized, it is difficult to measure the absorbance of any of the colors unless they are separated. For a two-color system, this can be done with a red–green–blue (RGB) camera, but more than two co-localized colors requires multispectral imaging and unmixing

or some other quantitative color separation method [23]. There are a number of algorithms that attempt to do a blind separation of multiple colors in multispectral imaging [24] and these can be useful for data exploration. However, because these algorithms attempt to decompose or separate each individual multispectral image, they are prone to errors when unanticipated components are present in an image or when one or more of the colors is not present. For intensity quantitation, it is more reliable to use an unmixing, or separation, method that utilizes a basis set, or spectral library, that has been developed in advance to match the samples being quantitated. A predeveloped spectral library will ensure that all samples are quantitated identically and intensity measurements can then be compared between images. Although theoretically determined spectra can be used, this is not optimal. There are both subtle and not-so-subtle differences in the spectral properties of various commercial chromogens, depending on the sample and the specific staining method used, which necessitates the use of sample-specific spectral libraries, usually developed from singly stained control samples [24]. In addition, DAB can display different spectra depending on many factors, which make it less than optimal for spectral imaging and unmixing [20].

The second problem with multispectral imaging systems is that they are generally slow compared with regular whole-slide scanning systems. At the time of writing, there are no commercial multispectral whole-slide scanning systems on the market. However, there are a few systems which combine the ability to do multispectral imaging on a per-field-of-view basis with whole-slide scanning [13, 25]. These so-called “survey and drill” systems use a regular color imaging system to generate a low-magnification overview of the sample and then return to acquire higher resolution multispectral images from a subset of fields of view. This typically yields 10–50 fields of multispectral data, compared to the sometimes thousands of fields obtained with a traditional whole-slide scanner. However, a pre-commercial system has been developed which can scan an entire microscope multispectrally in a reasonable timeframe [26]. This will help with the expansion of these methods into more routine use in pathology research.

Despite all of the potential pitfalls, it is possible to obtain quantitative separations of many chromogens in a single sample. The biggest challenges are typically with optimization and validation. There are simple methods available for two markers and it can be fairly straightforward to develop a three-marker (four-color) assay. However, it can be difficult and time-consuming to develop staining assays for more than four colors. For more than three markers, a fluorescence approach is more common and generally considered to be easier.

---

## 5 Staining and Imaging: Fluorescence

For fluorescence imaging, there are a large number of options for labeling multiple fluorophores in FFPE tissue sections. Each of these has their pros and cons, depending on the level of multiplexing required and the abundance of the protein being labeled.

### 5.1 Direct Labeling

One of the simplest methods for multiplexed fluorescence staining is to use antibodies which have already been conjugated to fluorophores of different colors. Other than issues with spectral overlap between channels during imaging, there is no theoretical limit to the number of markers which can be used in this manner. However, there are some practical limitations to this. First of all, one must have antibodies which have been directly conjugated to the fluorophores. There are a few antibody-fluorophore conjugates which are available commercially, although the vast majority of antibodies one would wish to use are not available in pre-conjugated form. This means that the conjugation must be done in the lab, using one of the commercial conjugation kits or by developing the chemistry independently. In addition, since this method results in only one or a few fluorophores for each antibody bound to the sample, there may not be enough fluorescence to detect a signal, especially against an autofluorescent sample background.

### 5.2 Primary–Secondary Conjugation Methods

One means of avoiding a conjugation step to directly label an antibody with a fluorophore is to use primary–secondary antibody methodologies. For this, a primary antibody of a specific species is introduced into the sample and then a secondary antibody directed against that species and labeled with a fluorophore is used. The secondary antibody binds to the primary antibodies in the sample, labeling them with a fluorophore. This methodology can be extended to multiple fluorophores by using primary antibodies of different species, and then using secondary antibodies (with different colored fluorophores) specific for those same species. This relies on having a separate species of primary antibody for each marker to be labeled. It can be challenging to find appropriate rabbit, mouse, rat, goat, chicken, camel combinations for the exact markers one needs. In addition, this method can still suffer from sensitivity problems, as it results in only a single antibody–fluorophore-labeling event for each primary antibody present.

### 5.3 Quantum Dots

Quantum dots (QDs), small nanoscale semiconductor devices with useful fluorescence properties, have been used in an effort to address both the issue of needing multiple nonoverlapping fluorophores and the sensitivity problems of direct and primary–secondary methods. By controlling the size of the nanoparticle, the Stokes shift of QDs can be “tuned” to have them emit at narrow and

specific wavelengths using a single excitation wavelength, enabling the use of many of these in a single sample without significant cross talk. In addition, QDs are typically brighter than the organic small molecule fluorophores normally used in microscopy. QDs have been used in FFPE tissue sections to achieve multiplexing of up to six colors in a single sample [27, 28]. However, there are still some limitations to this technology. The QDs are large, often 20–30 nm in diameter (once coatings have been put around them to prevent their oxidization and make them soluble in water and biologically compatible), and this can prevent their penetration into the specimen, limiting their access to only the outermost portion of the tissue. QDs can be used both for direct labeling and for primary–secondary staining methods. Although individual QD molecules can have narrow fluorescence spectra (2–3 nm), the commercial production is based on separating QDs of similar size into lots. The end result is that a range of sizes is separated into a single lot, giving a distribution of individually narrow fluorescent peaks. This distribution can be 10–30 nm in width, which still limits the number of QDs which can be multiplexed in a single sample [29]. Despite their potential advantages, QDs do not seem to work for all markers and are not a single solution to multiplexed fluorescence imaging. However, they can be used in combination with other methods and labels, which can make them useful in some circumstances. Nonetheless, QDs have been used in a large number of studies describing the use of multiplexed QD immunohistochemistry in pathology [28, 30, 31].

#### **5.4 Immuno-enzyme Amplification**

In general, the problem of “one fluorophore per antibody” can limit the utility of direct and primary–secondary methods. Abundant proteins are not too much of a problem but for proteins which have a very low copy number in a cell, regular conjugation methods are insufficient. These sensitivity issues can largely be overcome using an enzyme-mediated amplification method involving HRP analogous in some ways to that used for chromogenic staining [32]. Tyramide signal amplification (TSA) involves the same workflow as the chromogenic amplification method: application of a primary antibody, conjugation with a species-specific secondary antibody which contains HRP and then binding of multiple fluorophores to the sample using an HRP-catalyzed reaction [33]. The primary use of TSA has been for signal amplification. However, as a part of the tyramide reaction, the fluorophores are covalently bound to the sample, unlike direct and regular primary–secondary methods, in which the fluorophore is only weakly bound through non-covalent mechanisms. As a result, tyramide-reaction-bound fluorophores will remain in place on the sample during a boiling/heating procedure that will strip the primary and secondary antibodies. This will leave the sample prepared for a second round of TSA primary–secondary-amplification process using a primary antibody of any species, including the same species as the other primaries being

used. The heat step, typically through boiling the sample in a buffer, is similar to those used for chromogenic multiplexing and similar to the heat-induced epitope retrieval step often performed on FFPE tissues. This species-independent multiplexing, combined with the signal amplification benefits of TSA, makes it useful for multiplexed immunofluorescence [34]. The primary drawback of TSA multiplexing is that it is a sequential method, requiring the repetition of the entire staining cycle for each color/marker being used. For a 5- or 6-color staining protocol, this can take 18–24 h of processing time. In addition, sometimes the heating step is performed using a microwave, which can hinder automation of the process.

### 5.5 Autostaining

Although not an immunohistochemistry method per se, autostainers are an important part of *in situ* multiplexing. To be able to utilize a multiplex immunohistochemical approach in pathology and medical research, it is important to be able to fully automate the staining of the sample since a clinical study can involve hundreds or thousands of samples, each with one or more tissue sections to be analyzed. Having to manually stain these sections, particularly with a sequential method like TSA, can be onerous at best and practically impossible in the worst-case scenario. Fortunately, many of these methods can be automated in commercial automated slide staining systems [35]. In addition, there are now commercially available solutions for multiplexed TSA staining for up to five markers, with future expansions planned.

### 5.6 Combining IHC and ISH

Phenotype visualization and quantitation via immunohistochemistry is valuable. However, its utility can be increased when combined with *in situ* hybridization for mRNA or DNA, giving a read out on the genotype, phenotype and the mechanism by which one is translated into the other. Although relatively new, there have already been a number of publications showing the utility of this kind of multiplexing [36–38].

---

## 6 Imaging Considerations in Fluorescence

Most conventional microscopes use a fluorescence imaging scheme that involves having one emission wavelength for each excitation wavelength. This is typically achieved by using a filter cube with an excitation filter, a dichroic mirror and an emission filter. The excitation filter selects which excitation wavelength range will be extracted from a broad-spectrum excitation light source. The excitation light is shone onto the dichroic mirror, which is chosen to reflect the excitation light and transmit the emission light. The dichroic mirror reflects the excitation light through the objective lens onto the sample. The fluorescence emission thus generated is collected by the objective lens and

transmitted to the dichroic mirror, which passes the light to the emission filter. The emission filter is chosen to absorb light in the range of the excitation wavelengths (to prevent those from reaching the camera) and to pass a typically narrow range of emission wavelengths on to the camera. The excitation, dichroic mirror, and emission filters are carefully chosen to match the absorbance and emission properties of the fluorophore of interest for that filter cube. This type of fluorescence imaging utilizes a monochrome camera to take an image of the sample with each filter cube. The resulting gray-scale images, often referred to as “channels,” are then typically pseudo-colored into any color of choice. In this way, a multilayer pseudo-color image is created with one layer per filter cube (and per fluorophore).

In conventional fluorescence microscopy, methods have been worked out for imaging four colors simultaneously with little problem in non-fixed samples such as tissues or cells, and there have been a high number of publications using those methods. Unlike cell samples, for FFPE tissue sections there is the problem of tissue autofluorescence, which can mask or completely obscure signals of interest and can interfere with quantitation of even those signals that can be seen above the autofluorescence. In addition, the desire to multiplex more than four colors at one time can make cross-talk between channels harder to manage. To address this, there are a number of cross-talk correction algorithms based on linear unmixing that can be used to ensure that the fluorophores are viewed or quantitated in the correct channel [39, 40].

### **6.1 Color Versus Multispectral Imaging**

In pathology, fluorescently labeled tissue sections are most often imaged using a commercial whole-slide scanning system, which can rapidly digitize the entire tissue section. The resulting whole-slide can be used for viewing or quantitation. If the fluorophores in the sample are bright enough to be easily seen above the intrinsic tissue autofluorescence then visualization is easy. If the fluorophores are weak relative to the autofluorescence, then slide visualization can be difficult or impossible with a regular whole-slide scanner. Fluorescence images from a regular microscope or a whole-slide scanner can also be used for intensity quantitation of a pixel or region of the sample. Again, if the fluorescence signals in the sample are strong relative to the intrinsic autofluorescence, then quantitation can be straight forward. However, even small amounts of tissue autofluorescence in a given channel can cause problems. A 5% contribution of autofluorescence is probably not very important. However, given the strong autofluorescence in formalin-fixed tissue, many samples can contain 20%, 50% or even 90% autofluorescence in a given channel, especially in the green-yellow emission range. These interfering signals can make signal level quantitation impossible on a regular microscope or whole-slide scanner.

For situations where autofluorescence is a problem, either for simple visualization or for intensity quantitation, a multispectral imaging approach is required [41, 42]. By acquiring more spectral information about the sample, it is possible to unmix or separate the autofluorescence from the fluorophores of interest, increasing the contrast and legibility of a slide and improving intensity quantitation. The same unmixing algorithms that can correct for auto-fluorescence can be used to correct cross-talk between channels. The use of multispectral imaging to remove autofluorescence has been applied over many years in a number of fields, including small animal imaging and in microscopy [41]. It has proved especially useful in pathology, with hundreds of publications using this approach over the last 10 years. This has been documented in a recent review article [43].

Although excellent for fluorescence imaging of FFPE tissue sections, there are two major drawbacks with multispectral imaging approaches. Firstly, obtaining correct results with linear unmixing requires having the correct spectral “library,” or spectral basis set. The preferred method for developing a spectral library is to use samples that have been labeled with only one fluorophore as a “spectral control.” In the case of autofluorescent samples, however, this can be difficult as there are typically no pixels in an image that contain the fluorophore of interest and are also free of auto-fluorescence. To overcome this, an unstained sample is used to obtain a spectrum of autofluorescence and then that is used to isolate purified signatures from the control sample spectra [42]. There is a debate over how many extra wavelengths need to be acquired in a multispectral approach for autofluorescence removal. Theoretically, a minimum of  $N + 1$  wavelengths for  $N$  fluorophores is required. However, in practice more wavelengths are required to sufficiently measure the spectral properties of each fluorophore and to quantitatively separate them from each other and from auto-fluorescence. However, it is not yet clear just how many are required for a quantitative separation in the case of FFPE autofluorescence and the kinds of fluorophores used in pathology.

## 6.2 Confocal and Wide-Field Imaging

Both confocal and widefield fluorescence systems have been widely used for multiplexed imaging in the life sciences. However, the confocal approach has been typically focused on cellular imaging, while widefield fluorescence has been used for pathology samples, although there are some notable and interesting exceptions [44, 45]. Confocal systems are designed to eliminate interference from out-of-focus light in a sample through the use of pinholes to reject light from regions above and below the focal plane. This means that the majority of light detected comes from a specified voxel in the sample, with localization in all three dimensions. Most commercial confocal systems have the ability to scan in all three dimensions and can therefore build up a three-dimensional map of the

sample [46]. For thick samples ( $>10\text{ }\mu\text{m}$ ) or when full three-dimensional information is required, this is clearly an advantage. However, when imaging with a  $20\times$  objective, which is the *de facto* standard in pathology, most if not all of a  $5\text{-}\mu\text{m}$  section is in focus in a non-confocal image even for high numerical aperture objectives, which means that there is little advantage to using a confocal system. However for  $40\times$  or  $60\times$  objectives, there can be a significant amount of out-of-focus light that will cause blur in an image. In addition, it can be advantageous to be able to do three-dimensional co-localization analyses. In those situations, confocal imaging can be useful. In particular, three-dimensional resolution of small spots can be assisted by confocal imaging. DNA or RNA *in situ* hybridization (FISH or CISH) spots can be small and at  $20\times$  it can be difficult to resolve individual foci, particularly when they are aligned vertically in the sample. This problem can be exacerbated when imaging  $5\text{-}\mu\text{m}$  sections because of the way sectioning can leave only portions of individual cells in the sample. Confocal approaches can enable the use of thicker sections containing cells which can be analyzed in their entirety, albeit at the cost of taking more time to acquire imagery.

---

## 7 Highly Multiplexed Imaging

Traditional immunohistochemistry and imaging, as described above, is an excellent means of combining pathology and tissue architecture with the visualization and quantitation of protein distributions. However, even the most ambitious of methods will be limited in the number of markers that can be assessed, with somewhere between 6 and 12 being a practical limit. However, there are good reasons for wanting to use more than 6–12 markers. Fluorescence flow cytometry can handle up to 19 or 20 markers and newer mass spectrometry-based flow cytometric methods have been demonstrated with 40 markers, up to a limit of potentially around 100 [47–50]. These higher multiplexing methods have been shown to be useful, particularly for cancer immunological applications. Mass spectroscopy imaging has been applied to FFPE tissue sections for a number of years and is approaching a more main-stream acceptance, as instrumentation becomes less complex and more affordable [51, 52]. There have been two good recent reviews of mass spectroscopy imaging [53, 54].

Another new technology involves using UV light to release previously bound antibodies and their bar-code fluorescent tags from spatially resolved regions of an FFPE section. The released markers are then collected and analyzed in a separate analysis instrument, thus giving a read out of up to 800 markers simultaneously from a small region of the sample. The analysis technology (nCounter<sup>TM</sup>) has been in use for a number of years, and has

numerous applications in pathology for protein, DNA and RNA quantitation in any combination [55–57]. The new version of that technology can analyze the same markers as before, but provides the information from smaller regions of an FFPE section. Although new, this technology has the potential to deliver spatially resolved phenotypic and genotypic information from regions as small as single cells.

---

## 8 Sequential Acquisition Methods

Another approach to obtaining complex phenotypic information from cells *in situ* is to acquire images from a sample containing a smaller number of markers (typically 2–4) and then remove these markers and re-image the same field of view with a new set of markers, repeating as many cycles as necessary. One such study showed the detection of 61 protein epitopes in a single FFPE section using pair-wise immunofluorescence [58]. This technology has been commercialized under the name MultiOmyx™ and sold as a service [59]. This kind of technology provides for a highly multiplexed assay, with up to 100 markers acquired simultaneously. There are other methods for achieving high multiplicity in single cells using cyclic immunofluorescence, in which two- or four-color staining alternates with inactivation of fluorophores to progressively build a multichannel image. The use of standard reagents and instrumentation may be suitable for high-throughput assays and screening applications [60, 61].

Given the sequential nature of these processes, it can be slow to acquire all of the imagery needed for a single field of view from a single sample. In addition, care needs to be taken with image registration to ensure that the data from each pixel in each of the sequential images is coming from the exact same spot on the sample. Despite these limitations, a number of publications have come out using this approach [62–65].

---

## 9 Image Analysis Methods

Multiplex staining and image acquisition methods are just two parts of the story. The more colors that are put into a single slide, the more complicated it becomes to be able to visually assess the slide, particularly if there is some co-localization of the markers. Even with 4–6 markers plus a counterstain, it can be difficult to make sense of the images by eye. However, once the very highly multiplexed methods come into more mainstream use, it will be virtually impossible to visualize these markers all at once using any kind of standard histopathology display method. Therefore, some kind of image analysis and redisplay is required. In addition, the

primary driver of adding so many markers to FFPE samples is to better understand the phenotype of each cell, something that is not obvious from simply viewing markers. Again, this highlights the importance of image analysis in multiplexed imaging.

Image analysis in pathology can be divided broadly into three general steps: (1) tissue segmentation and quantitation; (2) cellular segmentation and quantitation; and (3) image/slide summary and meta-analysis. The goal in tissue segmentation is to use some kind of tissue morphology assessment algorithm to subdivide the image into morphologically relevant regions that can then be analyzed individually for area and as a part of the per-cell summary information. Cellular segmentation covers a range of analyses applied to individual cells in an image. It is often a multistep process, with a variety of algorithms employed, typically starting with the finding of each cell's nucleus based on the nuclear counterstain. Cellular cytoplasm and membrane regions are then found based on the nuclear location and shape. Once each cellular region has been defined, the intensity of each marker can be extracted. This subcellular per-marker intensity information can then be translated into a multimarker phenotype for each cell. Once each cell's phenotype is known, summary data can be compiled for each field of view and for each microscope slide, counting the number of each cell type in each morphologic region of the sample. For example, this approach can be used to determine how many CD4+/CD25+/FOXP3+ cells are in a tumor, tumor margin, and/or stroma on a slide.

Tissue, or morphologic, segmentation is often the most difficult step. The morphology of clinical FFPE sections is complex at best as this varies from sample to sample and tumor type to tumor type, and can be equivocal even for experienced pathologists [66, 67]. One approach is to have an expert manually outline the morphologic regions of interest in each image. This can work if there are only a few images that need analyzing but is evidently impractical for larger data sets with thousands of images. Developing an automated algorithm that will assess an image and segment it into morphologic regions is difficult, but there are successful examples.

There are two general approaches. One is parametric: an algorithm is developed using human-chosen image criteria (nuclear size, shape, distance metrics, colors, etc.) to segment unknown images [68, 69]. Another method is to employ user-drawn training regions as examples of each morphological class and then to use a computer program to determine an effective, typically black-box algorithm for distinguishing example regions. This is sometimes termed “machine-learning” or “deep learning” and has been applied in a range of image analysis problems, including pathology [70, 71]. A number of groups and commercial software packages have incorporated one or both of these schemes, which can be applied to a range of digital pathology image and sample types [72–77].

Cellular segmentation and analysis is typically the next step. In this, the location and shape of each cell is determined using a cell segmentation algorithm, typically utilizing the nuclear counterstain to denote the nuclear location and then finding the rest of the cellular subcompartments using a variety of other methods. Cellular segmentation is not new and has been applied to a large variety of cell types and situations [78]. However, cells *in situ* in FFPE tissue sections present a more difficult challenge than live or fixed cells on a glass slide or in a well. A 5  $\mu\text{m}$  section of FFPE tissue cuts through cells, leaving partial cells in the section and the possibility of spatial overlap between cells. This makes nuclear segmentation difficult, requiring special algorithms for FFPE tissue compared to other samples [79, 80]. In addition, because of the three-dimensional aspects of a section and the partial cells found there, performing cytoplasmic and membrane segmentation on FFPE sections is also more difficult. Many nuclear segmentation algorithms make the assumption that “one nucleus equals one cell,” which can make segmentation of multinucleated cells difficult. Cytoplasmic segmentation tends to be done by a watershed expansion starting from each cell’s nucleus, which requires an active membrane signal to determine correct cellular boundaries. Membrane segmentation typically requires having a good membrane stain from which to work, a strategy that works well when there is a clear membrane pattern, but which can be problematic for cells with no obvious membrane staining. In spite of these complications, there are many commercial and user-developed software packages and algorithms available which do an effective job of performing cellular segmentation on brightfield and fluorescence images from FFPE sections.

Once the cellular and subcellular boundaries and compartments have been decided, then it is a relatively simple matter to extract the fluorescence or absorbance data from each specific cell and subcellular (e.g., nuclear, cytoplasmic and membrane) region. This gives a table of data in which each row represents the data for a specific cell, which is analogous to a table of data obtained from flow cytometry. For example, a sample with 5-plex staining and a nuclear counterstain, and subdivision of each cell into nuclear, cytoplasmic and membrane compartments, this would result in 20 variables (nuclear intensities for all five fluors, cytoplasmic intensity for all five fluors, the membrane intensity for all five fluors and the total intensity for all five fluors). In addition, it is often possible to extract other summary data about the fluors in each subcompartment (minimum, maximum, standard deviation, and sometimes texture-based readings), and each cell typically comes with a range of shape/size variables (area, long axis, short axis, circularity, etc.), as well as texture variables about the staining. This results in a very long list of variables for each cell in the image.

This table of per-cell data forms the basis of cellular phenotyping. The simplest approach for analyzing this kind of table is to use a flow-cytometry sequential gating approach. Most cell

segmentation data can be exported and read into flow-cytometry commercial software for analysis [81, 82]. However, direct application of the sequential gating methods used so effectively in cytometry to cellular data derived from cells *in situ* in FFPE sections may not be satisfactory [83]. In addition to the three-dimensional problems of overlapping cells, there is a large cellular contact problem. In a flow cytometer, the cells are analyzed one at a time and the levels of each marker are measured for that cell without interference from other cells. However in tissue sections, cells can touch each other and, in the case of smaller immune cells, this can involve many cells. The nature of membrane segmentation of these cells *in situ* means that they share the membrane with neighboring cells, which means that if a CD4+/CD8– cell shares 25 % of its membrane region with a touching CD4-/CD8+ cell, the nice, neat dual cluster scatter plot seen for CD4/CD8 differentiation will not exist for these cells in tissue. The two neat clusters will spread into one another, depending on the sharing of membranes and spatial overlap.

One approach to dealing with this tissue cytometry problem was developed by Ron Germain and coworkers at the National Institute of Health and has been applied to a number of immunology applications. These include the role of immune homeostasis by co-localized effector and regulatory T cells [84] and assessment of dendritic cell subpopulations [45]. Another approach is to have a user-guided training of an algorithm based on examples the user selects for each multimarker phenotype of interest. The computer then generates an algorithm which creates the best multivariate gating scheme to separate the cells into phenotypes based on the long list of per-cell variables in the data table. One implementation of this machine learning for cellular phenotyping involved using a multinomial logistic regression algorithm to find the multivariate gating scheme in a 6-plex assay to divide cells into categories of tumor cells, cytotoxic T cells, helper T cells, regulatory T cells, and macrophages, producing cell phenotype maps retaining spatial arrangements. This kind of multivariate gating scheme requires users to select a minimum number of examples of the cells exemplifying the phenotypes of interest to train the algorithm. This is likely to be the most promising approach to this kind of highly multiplexed imaging and analysis problem.

Once each cell has been located and phenotyped, with its marker intensities extracted on a subcellular basis, there are a range of options for analysis. A simple visual picture of the distributions of the various cell types in the tumor and tumor microenvironment can be useful. However, visual assessment of the inter-distributions of multiple cell types is difficult. For this reason a number of summary data analyses have been developed to begin to address this data analysis problem. These kinds of summary analyses can be done on a per-field (per-image) basis or across all of the images

from a section, or even across a cohort of patients in a study. A simple metric is to simply count the number of each phenotypical cell in each morphological region. These summary metrics can be augmented by a range of spatial mapping algorithms. Typically these can be used to explore whether or not the distances between the cells of a specific phenotype are related to a clinical/experimental parameter such as survival, dose or age. These can be simple spatial distance metrics such as a Euclidean or Mahalanobis distances or more complicated metrics involving the relationships between multiple cellular phenotypes and tissue morphological regions. There are also hypothesis-free methods which can be used to determine what spatial patterns of cells correlate with clinical or experimental parameters such as survival [85, 86]. This all contributes to being able to map the spatial heterogeneity in the tumor microenvironment, something which is understood to be of increasing importance [87].

---

## 10 Literature Highlights

Two groups at the National Institute of Health have developed multiplexed imaging and per-cell analysis technologies utilizing confocal microscopes [88]. Sometimes termed histo-cytometry, this method has been used to stain and image up to ten fluorescent markers, and then open source software is used to perform per-cell analysis [45]. These have been primarily used in neuroscience and immune-oncology applications, from investigating how helper B cells can promote cytotoxic T cell survival and proliferation independently of antigen presentation through CD27/CD70 interactions [89], to how immune homeostasis is enforced by co-localized effector and regulatory T cells [84] to investigations of the spatio-temporal basis of immunity in lymphoid tissue [90].

Multispectral imaging has been used for over a decade as the basis for multiplexed imaging in pathology, with hundreds of papers published. Full coverage of the uses of this technology for the multiplexed imaging of clinical biopsy samples is outside of the scope of this chapter but there has been a recent literature review which addresses this [43].

One recent example of multiplexed imaging and per-cell analysis is the investigation of tumor infiltrating plasma cells and their association with tertiary lymphoid structures, cytolytic T cell responses and prognosis in ovarian cancer [91]. Another is a study into tumor-derived lipocalin-2 and whether or not it promotes breast cancer metastases. To this end, co-localized PyMT- and Ki-67-double positive tumor cells were detected and quantified [92]. In immune-oncology, the T cell landscape in primary melanoma was investigated and found to predict the survival of patients with metastatic disease after their treatment with dendritic cell

vaccines [93]. In another immune-oncology paper, the authors discuss how CD169 can identify an activated CD8+ T cell subset in lymph nodes which can predict a favorable prognosis in colorectal cancer [94]. This same set of technologies has been applied to the studying the effects of transforming growth factor- $\beta$  to limit secretion of lumican by activated stellate cells within primary pancreatic adenocarcinoma tumors [95].

Recent work from the group of Bernard Fox has shown a methodology that can be consistently applied to investigate FFPE tissues by multiplexing up to six markers. This helped to enumerate the complicated phenotypes of immune cell subsets and allowed spatial distribution analyses. They validated the use of multiplex immunohistochemistry for detection of CD3+CD4+ and CD3+/CD8+ T cell subsets in murine spleen and tumors [96] and applied the same methodology to analyzing tumor-infiltrating lymphocytes in melanoma [97]. The institute led by James Allison at MD Anderson has been among the leaders in immune-oncology research for many years. In a recent publication, they analyzed immune signatures in longitudinal tumor samples for insight into biomarkers of response and mechanisms of resistance to immune checkpoint blockade. This work demonstrated that adaptive immune signatures in tumor biopsies can be predictive of response to checkpoint blockade treatment. This study is also of high interest because it combined multiplexed protein analysis, per-cell quantitation and phenotyping with highly multiplexed non-imaging analyses of homogenized tissues [98]. These combinations of traditional pathology imaging and analysis with newer multiplexed quantitative pathology methods along with multiplexed analyses of homogenized tissues are likely to become increasingly useful as tumor-immune interactions are investigated further.

---

## 11 Conclusions

The field of multiplexed imaging and per-cell phenotyping of cells *in situ* in FFPE sections has undergone a rapid growth thanks to the interest in immune-tumor interactions, fuelled by the success of new cancer immunotherapies. There are a range of technologies that have been developed, including simple two- or three-plex chromogenic staining with a visual assessment all the way to highly multiplexed methods which can analyze hundreds of markers at a time from individual cells. All of these techniques require some kind of marker labeling strategy, an imaging strategy and an analysis strategy, each with its own pros and cons. Up to six-plex per-cell phenotyping using immunofluorescence labeling, multi-spectral imaging, and morphologic and per-cell analysis software has become attainable, using a few different commercial staining methods and image analysis packages. Beyond that, there are a

few highly multiplexed methods that are just becoming available. Although these can provide potentially up to 800-plex analyses, the acquisition time is slow and the scope of an experiment using these approaches may be limited. One area that has not seen much development so far is in the combination of genotyping and phenotyping of cells—combining ISH for DNA and RNA, and immunohistochemistry for protein staining along with imaging and cellular analyses. These hold the potential to expand our understanding of tumor-immune interactions even further. However, further developments are necessary due to the spot-counting nature of many ISH analysis problems and the three-dimensional problems of cells in the context of the tumor microenvironment. Addressing these will require the development of three-dimensional imaging methods for thicker (10–30 µm) samples.

## References

- Medzhitov R, Shevach EM, Trinchieri G, Mellor AL, Munn DH, Gordon S et al (2011) Highlights of 10 years of immunology in *nature reviews immunology*. *Nat Rev Immunol* 11:693–702
- Mandy FF, Bergeron M, Minkus T (1997) Evolution of leukocyte immunophenotyping as influenced by the HIV/AIDS pandemic: a short history of the development of gating strategies for CD4<sup>+</sup> T-cell enumeration. *Cytometry* 30:157–165
- Loken MR (1990) Immunofluorescence Techniques. In: Flow cytometry and sorting. 2nd edn. Wiley-Blackwell; ISBN-13: 978-0471562351, pp 341–353
- Zola H, Swart B, Banham A, Barry S, Beare A, Bensussan A et al (2007) CD molecules 2006—human cell differentiation molecules. *J Immunol Methods* 318:1–5
- Rimm DL (2006) What brown cannot do for you. *Nat Biotechnol* 24:914–916
- Taylor CR, Levenson RM (2006) Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment II. *Histopathology* 49:411–424
- Potts SJ, Johnson TD, Voelker FA, Lange H, Young GD (2011) Multiplexed measurement of proteins in tissue in a clinical environment. *Appl Immunohistochem Mol Morphol* 19:494–498
- Dixon AR, Bathany C, Tsuei M, White J, Barald KF, Takayama S (2015) Recent developments in multiplexing techniques for immunohistochemistry. *Expert Rev Mol Diagn* 15:1171–1186
- van der Loos CM, Naruko T, Becker AE (1996) The use of enhanced polymer one-step staining reagents for immunoenzyme double-labelling. *Histochem J* 28:709–714
- van der Loos CM, Teeling P (2008) A generally applicable sequential alkaline phosphatase immunohistochemical double staining. *J Histotechnol* 31:119–127. doi:[10.1179/his.2008.31.3.119](https://doi.org/10.1179/his.2008.31.3.119)
- Osman TA, Øijordsbakken G, Costea DE, Johannessen AC (2013) Successful triple immunoenzymatic method employing primary antibodies from same species and same immunoglobulin subclass. *Eur J Histochem* 57:e22
- Blenman KR, Lee PP (2014) Quantitative and spatial image analysis of tumor and draining lymph nodes using immunohistochemistry and high-resolution multispectral imaging to predict metastasis. *Methods Mol Biol* 1102:601–621
- Setiadi AF, Ray NC, Kohrt HE, Kapelner A, Carcamo-Cavazos V, Levic EB et al (2010) Quantitative, architectural analysis of immune cell subsets in tumor-draining lymph nodes from breast cancer patients and healthy lymph nodes. *PLoS One* 5:e12420
- Oguejiofor K, Hall J, Slater C, Betts G, Hall G, Slevin N (2015) Stromal infiltration of CD8 T cells is associated with improved clinical outcome in HPV-positive oropharyngeal squamous carcinoma. *Br J Cancer* 113:886–893
- van der Loos CM (2010) Chromogens in multiple immunohistochemical staining used for visual assessment and spectral imaging: the colorful future. *J Histotechnol* 33:31–40. doi:[10.1179/his.2010.33.1.31](https://doi.org/10.1179/his.2010.33.1.31)
- van der Loos CM (2008) Multiple immunoenzyme staining: methods and visualizations for the observation with spectral imaging. *J Histochem Cytochem* 56:313–328
- Farkas DL, Ballou B, Du C, Fisher GW, Lau C, Levenson RM (1997, September). Optical

- image acquisition, analysis and processing for biomedical applications. In: Image analysis and processing. Session 11: biomedical applications image analysis and processing. Volume 1311 of the series lecture notes in computer science, pp 663–671
18. Mansfield JR, Sowa MG, Scarth GB, Somorjai RL, Mantsch HH (1997) Fuzzy C-means clustering and principal component analysis of time series from near-infrared imaging of forearm ischemia. *Comput Med Imaging Graph* 21:299–308
  19. Mansfield JR, Hoyt C, Levenson RM (2008) Visualization of microscopy-based spectral imaging data from multi-label tissue sections. *Curr Protoc Mol Biol Chapter 14:Unit 14.19.* doi:[10.1002/0471142727.mb1419s84](https://doi.org/10.1002/0471142727.mb1419s84)
  20. Neher R, Neher E (2004) Optimizing imaging parameters for the separation of multiple labels in a fluorescence image. *J Microsc* 213:46–62
  21. Rothmann C, Bar-Am I, Malik Z (1998) Spectral imaging for quantitative histology and cytogenetics. *Histol Histopathol* 13:921–926
  22. Macville MV, Van der Laak JA, Speel EJ, Katzir N, Garini Y, Soenksen D (2001) Spectral imaging of multi-color chromogenic dyes in pathological specimens. *Anal Cell Pathol* 22:133–142
  23. Wemmert C, Kruger JM, Forestier G, Sternberger L, Feuerhake F, Gançarski P (2013). Stain unmixing in brightfield multiplexed immunohistochemistry. In: Image processing (ICIP), 2013 20th IEEE International Conference. IEEE pp. 1125–1129. Doi: <http://dx.doi.org/10.1016/j.compmedimag.2015.04.001>
  24. Choi S, Cichocki A, Park H-M, Lee SY (2005) Blind source separation and independent component analysis: a review. *Neural Inform Process - Lett Rev* 6(1):1–57
  25. Ghaznavi F, Evans A, Madabhushi A, Feldman M (2013) Digital imaging in pathology: whole-slide imaging and beyond. *Annu Rev Pathol* 8:331–359
  26. Lindner M, Shotan Z, Yuval G (2016) Rapid microscopy measurement of very large spectral images. *Opt Express* 24:9511–9527
  27. Vu TQ, Lam WY, Hatch EW, Lidke DS (2015) Quantum dots for quantitative imaging: from single molecules to tissue. *Cell Tissue Res* 360:71–86
  28. Zhou L, Yan J, Tong L, Han X, Wu X, Guo P (2016) Quantum dot-based immunohistochemistry for pathological applications. *Cancer Transl Med* 2:21–28. doi:[10.4103/2395-3977.177562](https://doi.org/10.4103/2395-3977.177562)
  29. Resch-Genger U, Grabolle M, Cavaliere-Jaricot S, Nitschke R, Nann T (2008) Quantum dots versus organic dyes as fluorescent labels. *Nat Methods* 5:763–775
  30. Byers RJ, Hitchman ER (2011) Quantum dots brighten biological imaging. *Prog Histochem Cytochem* 45:201–237
  31. Tholouli E, MacDermott S, Hoyland J, Yin JL, Byers R (2012) Quantitative multiplex quantum dot in-situ hybridisation based gene expression profiling in tissue microarrays identifies prognostic genes in acute myeloid leukaemia. *Biochem Biophys Res Commun* 425:333–339
  32. Chao J, DeBiasio R, Zhu Z, Giuliano KA, Schmidt BF (1996) Immunofluorescence signal amplification by the enzyme-catalyzed deposition of a fluorescent reporter substrate (CARD). *Cytometry* 23:48–53
  33. Warford A, Akbar H, Riberio D (2014) Antigen retrieval, blocking, detection and visualisation systems in immunohistochemistry: a review and practical evaluation of tyramide and rolling circle amplification systems. *Methods* 70:28–33
  34. Wang G, Achim CL, Hamilton RL, Wiley CA, Soontornniyomkij V (1999) Tyramide signal amplification method in multiple-label immunofluorescence confocal microscopy. *Methods* 18:459–464
  35. Yarilin D, Xu K, Turkekul M, Fan N, Romin Y, Fijisawa S (2015) Machine-based method for multiplex in situ molecular characterization of tissues by immunofluorescence detection. *Sci Rep* 5:9534. doi:[10.1038/srep09534](https://doi.org/10.1038/srep09534)
  36. Stempel AJ, Morgans CW, Stout JT, Appukuttan B (2014) Simultaneous visualization and cell-specific confirmation of RNA and protein in the mouse retina. *Mol Vis* 20:1366–1373
  37. Kunju LP, Carskadon S, Siddiqui J, Tomlins SA, Chinnaian AM, Palanisamy N (2014) Novel RNA hybridization method for the in situ detection of ETV1, ETV4, and ETV5 gene fusions in prostatecancer. *Appl Immunohistochem Mol Morphol* 22:e32–e40
  38. Li P, Janczewski WA, Yackle K, Kam K, Pagliardini S, Krasnow MA et al (2016) The peptidergic control circuit for sighing. *Nature* 530:293–297
  39. Zimmermann T, Rietdorf J, Pepperkok R (2003) Spectral imaging and its applications in live cell microscopy. *FEBS Lett* 546:87–92
  40. Zimmermann T (2005) Quantum dot-based immunohistochemistry for pathological applications. *Adv Biochem Engin Biotechnol* 95:245–265
  41. Dickinson ME, Bearman G, Tille S, Lansford R, Fraser SE (2001) Multi-spectral imaging

- and linear unmixing add a whole new dimension to laser scanning fluorescence microscopy. *Biotechniques* 31:1272–1279
42. Mansfield JR, Gossage KW, Hoyt CC, Levenson RM (2005) Autofluorescence removal, multiplexing, and automated analysis methods for in-vivo fluorescence imaging. *J Biomed Opt* 10:41207
  43. Mansfield JR (2014) Multispectral imaging: a review of its technical aspects and applications in anatomic pathology. *Vet Pathol* 51:185–210
  44. Gerner MY, Torabi-Parizi P, Germain RN (2015) Strategically localized dendritic cells promote rapid T cell responses to lymph-borne particulate antigens. *Immunity* 42:172–185
  45. Gerner MY, Kastenmuller W, Ifrim I, Kabat J, Germain RN (2012) Histo-cytometry: a method for highly multiplex quantitative tissue imaging analysis applied to dendritic cell subset microanatomy in lymph nodes. *Immunity* 37:364–376
  46. Sheppard CJR, Gan X, Gu M, Roy M, Pawley JB (2006) In: Pawley JB (ed) *Handbook of biological confocal microscopy*, 3rd edn. Springer, New York, pp 442–451. ISBN ISBN-13: 978-0387259215
  47. Giesen C, Wang HA, Schapiro D, Zivanovic N, Jacobs A, Hattendorf B et al (2014) Highly multiplexed imaging of tumor tissues with sub-cellular resolution by mass cytometry. *Nat Methods* 11:417–422
  48. Tanner SD, Baranov VI, Ornatsky OI, Bandura DR, George TC (2013) An introduction to mass cytometry: fundamentals and applications. *Cancer Immunol Immunother* 62:955–965
  49. Schwamborn K, Caprioli RM (2010) Molecular imaging by mass spectrometry—looking beyond classical histology. *Nat Rev Cancer* 10:639–646
  50. Spitzer MH, Nolan GP (2016) Mass cytometry: single cells, many features. *Cell* 165:780–791
  51. Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, Borowsky AD (2014) Multiplexed ion beam imaging of human breast tumors. *Nat Med* 20:436–442
  52. Schüffler PJ, Schapiro D, Giesen C, Wang HA, Bodenmiller B, Buhmann JM (2015) Automatic single cell segmentation on highly multiplexed tissue images. *Cytometry A* 87:936–942
  53. Norris JL, Caprioli RM (2013) Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research. *Chem Rev* 113:2309–2342
  54. Bodenmiller B (2016) Multiplexed epitope-based tissue imaging for discovery and health-care applications. *Cell Syst* 2:225–238
  55. Ullal AV, Peterson V, Agasti SS, Tuang S, Juric D, Castro CM et al (2014) Cancer cell profiling by barcoding allows multiplexed protein analysis in fine-needle aspirates. *Sci Transl Med* 6(219):219ra9
  56. Reis PP, Waldron L, Goswami RS, Xu W, Xuan Y, Perez-Ordonez B et al (2011) mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *BMC Biotechnol* 11:46. doi:[10.1186/1472-6750-11-46](https://doi.org/10.1186/1472-6750-11-46)
  57. Kolbert CP, Feddersen RM, Rakhshan F, Grill DE, Simon G, Middha S et al (2013) Multi-platform analysis of microRNA expression measurements in RNA from fresh frozen and FFPE tissues. *PLoS One* 8:e52517
  58. Gerdes MJ, Sevinsky CJ, Sood A, Adak S, Bello MO, Bordwell A et al (2013) Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc Natl Acad Sci U S A* 110:11982–11987
  59. Hollman-Hewgley D, Lazare M, Bordwell A, Zebadua E, Tripathi P, Ross AS et al (2014) A single slide multiplex assay for the evaluation of classical Hodgkin lymphoma. *Am J Surg Pathol* 38:1193–1202
  60. Lin JR, Fallahi-Sichani M, Sorger PK (2015) Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nat Commun* 6:8390. doi:[10.1038/ncomms9390](https://doi.org/10.1038/ncomms9390)
  61. Clarke GM, Zubovits JT, Shaikh KA, Wang D, Dinn SR, Corwin AD et al (2014) A novel, automated technology for multiplex biomarker imaging and application to breast cancer. *Histopathology* 64:242–255
  62. Sood A, Miller AM, Brogi E, Sui Y, Armenia J, McDonough E, Pang Z et al (2016) Multiplexed immunofluorescence delineates proteomic cancer cell states associated with metabolism. *JCI Insight* 1(6):e87030
  63. Adams A, Bordwell A, Bouman D, Fisher D, Hollman D, Lazare M et al (2013) MultiOmx™: a novel chemistry and visualization tools to aid resolution of discrepant Hodgkin lymphoma cases—a single slide multiplex assay for the evaluation of classical Hodgkin lymphoma. *Blood* 122:2994
  64. Gerdes MJ, Sood A, Sevinsky C, Pris AD, Zavodszky MI, Ginty F (2014) Emerging understanding of multiscale tumor heterogeneity. *Front Oncol* 4:366. doi:[10.3389/fonc.2014.00366](https://doi.org/10.3389/fonc.2014.00366)
  65. Uhlik MT, Liu J, Falcon BL, Iyer S, Stewart J, Celikkaya H et al (2016) Stromal-based signatures for the classification of gastric cancer. *Cancer Res* 76:2573–2586
  66. Rizzardi AE, Johnson AT, Vogel RI, Pambuccian SE, Henriksen J, Skubitz AP et al

- (2012) Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. *Diagn Pathol* 7:1596–1597
67. Mengel M, von Wasielewski R, Wiese B, Rüdiger T, Müller-Hermelink HK, Kreipe H (2002) Inter-laboratory and inter-observer reproducibility of immunohistochemical assessment of the Ki-67 labelling index in a large multi-centre trial. *J Pathol* 198:292–299
  68. Beck AH, Sangi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ et al (2011) Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med* 3:108ra113
  69. Baatz M, Zimmermann J, Blackmore CG (2009) Automated analysis and detailed quantification of biomedical images using definisens cognition network technology. *Comb Chem High Throughput Screen* 12:908–916
  70. Fernandez DC, Bhargava R, Hewitt SM, Levin IW (2005) Infrared spectroscopic imaging for histopathologic recognition. *Nat Biotechnol* 23:469–474
  71. Levenson R (2008) Putting the “more” back in morphology: spectral imaging and image analysis in the service of pathology. *Arch Pathol Lab Med* 132:748–757
  72. Feldman MD (2008) Beyond morphology: whole slide imaging, computer-aided detection, and other techniques. *Arch Pathol Lab Med* 132:758–763
  73. Veta M, Pluim JP, van Diest PJ, Viergever MA (2014) Breast cancer histopathology image analysis: a review. *IEEE Trans Biomed Eng* 61:1400–1411
  74. Pantanowitz L, Valenstein PN, Evans AJ, Kaplan KJ, Pfeifer JD, Wilbur DC et al (2011) Review of the current state of whole slide imaging in pathology. *J Pathol Inform* 2:36. doi:[10.4103/2153-3539.83746](https://doi.org/10.4103/2153-3539.83746)
  75. Isse K, Lesniak A, Grama K, Roysam B, Minervini MI, Demetris AJ (2012) Digital transplantation pathology: combining whole slide imaging, multiplex staining and automated image analysis. *Am J Transplant* 12:27–37
  76. Bueno G, Fernández-Carrobles MM, Deniz O, García-Rojo M (2016) New trends of emerging technologies in digital pathology. *Pathobiology* 83:61–69
  77. Akbar S, Jordan LB, Purdie CA, Thompson AM, McKenna SJ (2015) Comparing computer-generated and pathologist-generated tumour segmentations for immunohistochemical scoring of breast tissue microarrays. *Br J Cancer* 113:1075–1080
  78. Meijering E (2012) Cell segmentation: 50 years down the road. *IEEE signal processing magazine*, 29(5), pp. 140–145. September 2012
  79. Irshad H, Veillard A, Roux L, Racoceanu D (2014) Methods for nuclei detection, segmentation, and classification in digital histopathology: a review-current status and future potential. *IEEE Rev Biomed Eng* 7:97–114
  80. Xing F, Yang L (2016) Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. *IEEE Rev Biomed Eng* Jan 6. [Epub ahead of print]
  81. Lugli E, Roederer M, Cossarizza A (2010) Data analysis in flow cytometry: the future just started. *Cytometry A* 77:705–713
  82. Aghaeepour N, Finak G, FlowCAP Consortium, DREAM Consortium, Hoos H, Mosmann TR et al (2013) Critical assessment of automated flow cytometry data analysis techniques. *Nat Methods* 10:228–238
  83. Hedley DW (1989) Flow cytometry using paraffin-embedded tissue: five years on. *Cytometry* 10:229–241
  84. Liu Z, Gerner MY, Van Panhuys N, Levine AG, Rudensky AY, Germain RN (2015) Immune homeostasis enforced by co-localized effector and regulatory T cells. *Nature* 528:225–230
  85. Rose CJ, Naidoo K, Clay V, Linton K, Radford JA, Byers RJ (2013) A statistical framework for analyzing hypothesized interactions between cells imaged using multispectral microscopy and multiple immunohistochemical markers. *J Pathol Inform* 4(Suppl):S4. doi:[10.4103/2153-3539.109856](https://doi.org/10.4103/2153-3539.109856)
  86. Nelson LS, Mansfield JR, Lloyd R, Oguejiofor K, Salih Z, Menasce LP, Linton KM et al (2015) Automated prognostic pattern detection shows favourable diffuse pattern of FOXP3(+) tregs in follicular lymphoma. *Br J Cancer* 113:1197–1205
  87. Heindl A, Nawaz S, Yuan Y (2015) Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology. *Lab Invest* 95:377–384
  88. Germain RN (2005) Imaging dynamic interactions in immune responses. *Semin Immunol* 17:385–386
  89. Deola S, Panelli MC, Maric D, Selleri S, Dmitrieva NI, Voss CY et al (2008) Helper B cells promote cytotoxic T cell survival and proliferation independently of antigen presentation through CD27/CD70 interactions. *J Immunol* 180:1362–1372
  90. Qi H, Kastenmüller W, Germain RN (2014) Spatiotemporal basis of innate and adaptive immunity in secondary lymphoid tissue. *Annu Rev Cell Dev Biol* 30:141–167

91. Kroeger DR, Milne K, Nelson BH (2016) Tumor infiltrating plasma cells are associated with tertiary lymphoid structures, cytolytic T cell responses, and superior prognosis in ovarian cancer. *Clin Cancer Res* 22:3005–3015
92. Ören B, Urosevic J, Mertens C, Mora J, Guiu M, Gomis RR et al (2016) Tumour stroma-derived lipocalin-2 promotes breast cancer metastasis. *J Pathol* 239:274–285
93. Vasaturo A, Halilovic A, Bol KF, Verweij DI, Blokx WA, Punt CJ et al (2016) T cell landscape in a primary melanoma predicts the survival of patients with metastatic disease after their treatment with dendritic cell vaccines. *Cancer Res* 76:3496–3506
94. Zhang J, Xu J, Zhang RX, Zhang Y, Ou QJ, Li JQ et al (2016). CD169 identifies an activated CD8+ T cell subset in regional lymph nodes that predicts favorable prognosis in colorectal cancer patients. *OncoImmunology* (in press)
95. Kang YA, Roife D, Lee Y, Lv H, Suzuki R, Ling J et al (2016) Transforming growth factor- $\beta$  limits secretion of lumican by activated stellate cells within primary pancreatic adenocarcinoma tumors. *Clin Cancer Res*, Apr 28. pii: clincanres.2780.2015. [Epub ahead of print]
96. Feng Z, Jensen SM, Messenheimer DJ, Farhad M, Neuberger M, Bifulco CB et al (2016) Multispectral imaging of T and B cells in murine spleen and tumor. *J Immunol* 196:3943–3950
97. Feng Z, Puri S, Moudgil T, Wood W, Hoyt CC, Wang C et al (2015) Multispectral imaging of formalin-fixed tissue predicts ability to generate tumor-infiltrating lymphocytes from melanoma. *J Immunother Cancer* 3:47. doi:10.1186/s40425-015-0091-z
98. Chen PL, Roh W, Reuben A, Cooper ZA, Spencer CN, Prieto PA et al (2016) Analysis of immune signatures in longitudinal tumor samples yields insight into biomarkers of response and mechanisms of resistance to immune checkpoint blockade. *Cancer Discov* 6:827–837, Jun 14. pii: CD-15-1545. [Epub ahead of print]

# **Part II**

## **Statistical Considerations**

# Chapter 6

## Identification and Clinical Translation of Biomarker Signatures: Statistical Considerations

Emanuel Schwarz

### Abstract

Powerful machine learning tools exist to extract biological patterns for diagnosis or prediction from high-dimensional datasets. Simultaneous advances in high-throughput profiling technologies have led to a rapid acceleration of biomarker discovery investigations across all areas of medicine. However, the translation of biomarker signatures into clinically useful tools has thus far been difficult. In this chapter, several important considerations are discussed that influence such translation in the context of classifier design. These include aspects of variable selection that go beyond classification accuracy, as well as effects of variability on assay stability and sample size. The consideration of such factors may lead to an adaptation of biomarker discovery approaches, aimed at an optimal balance of performance and clinical translatability.

**Key words** Biomarker discovery, Machine learning, Variable selection, Classification, Molecular diagnostics

---

### 1 Introduction

Technological advancements in high-throughput technology have tremendously accelerated the search for biological patterns that have clinical utility for diagnosis and prediction. Among these are multiplexed assays that facilitate simultaneous measurement of analytes in small sample volumes, with high-throughput and low variability often comparable to the single-plex gold standard methodology. The developments in these approaches have been paralleled by a tremendous increase in application of multivariate methods to identify biological signatures within the generated high-dimensional datasets, a process that has been accelerated by the availability of complex algorithms in standard software packages. These facilitate the extraction of complex biological patterns from high-dimensional data that can already be transferred efficiently into dedicated multiplexed measurement systems. To aid in this process, computational pipelines have been developed that support translation from study design and initial biomarker

screening to clinically applicable multiplexed tests [1]. However, such transferability is platform dependent and results from high-throughput profiling within a research setting may not be easily transferred into clinically usable assay systems [2]. Despite these technological advances, the translation of biomarker candidates into clinical tests has been slow. For example in the cancer field, many biomarker candidates that showed promise during initial stages of the development process did not turn out to be clinically useful, and this was frequently not realized until the later stages of the process [3]. While there are numerous aspects that affect the discovery and clinical translation of biomarker signatures, including clinical, methodological, and regulatory challenges [4–8], this chapter will focus on statistical considerations regarding the optimal identification of biomarker sets.

The computational identification of biological signatures typically falls within the realm of supervised machine learning, where a given part of labeled data is used for “training,” to select an “optimal” combination of measurements for a predefined classification or prediction task. The algorithm is subsequently tested in a part of the data not used during training and the accuracy of the prediction in this test data is used as an estimation of how the classifier will perform in future, independent datasets. In practice, training and testing is usually performed by splitting datasets into a training and a test set, by cross-validation or similar techniques. In cross-validation, the data is split randomly into a given number of junks and each of these junks is used as a test set until all subjects have been classified once. Estimates of classifier accuracy are then determined across the entire cohort used during cross-validation. The identification of biomarker signatures using machine learning methodology is influenced by a significant number of choices, such as algorithm selection, data transformation, or parameter specification. As these aspects have been extensively reviewed elsewhere [9–11], this chapter focuses on areas that are less frequently considered during identification of machine learning classifiers but which may be of significant importance for translating signatures efficiently into clinically usable tests.

---

## 2 Considerations for Identifying Optimal Predictive Signatures

The application of machine learning methods to identify biological signatures with potential clinical relevance is common in current research and reaches across all areas of medicine [12]. There are a large variety of different algorithms that are used for the identification of such biomarker signatures. Although benchmarking studies have compared the performance of these algorithms using diverse datasets [13], the choice for or against a given algorithm is not trivial as it frequently depends on the data under investigation

[14]. For high-dimensional data, assessment of the dependency between data dimensionality and accuracy has shown that some algorithms give higher performance than others [15]. Nevertheless, for nearly all algorithms, there is an advantage in applying so-called variable selection, the identification of an optimal subset of all measured variables [16]. The selection of a smaller number of variables is frequently beneficial for performance and generalizability of algorithms. This is due to the fact that algorithms are affected less by sources of variability compared to those incorporating a large number of predictors. Therefore, subject-specific fluctuations in the measured analyte levels or random effects from variables that do not show substantial association with the outcome are less likely to impact on classification or prediction. On the other hand, for accurate classification of complex illnesses, or other multifactorial clinical outcomes, such as treatment response, it is beneficial to include larger numbers of predictors that each contributes a small piece of independent information. As a consequence, to find an optimal balance between accuracy and generalizability, biomarker panels identified in the literature frequently feature between 5 and 80 different predictors.

In these approaches, the defining criterion for selecting a given set of predictors is classification accuracy. However, in high-dimensional data it is frequently observed that there is not a single optimal solution to the variable selection problem as different combinations of variables may lead to comparable performance estimates. For this reason, there is an opportunity to utilize other metrics during the variable selection process that capture important properties of a given predictor set beyond the classification accuracy. One of these metrics is the stability of the variable selection under sampling variability [17], an issue that is particularly prominent for the usual scenario of variable numbers largely exceeding the number of investigated subjects [18]. The general idea of stability assessment is the repeated sampling of training sets from a given dataset and to explore how consistently a given variable gets selected as an important predictor. Numerous methods exist to quantify variable selection stability and these are reviewed in [17] and [19]. Importantly, different variable selection methods differ regarding the stability and accuracy of the selected variables [20]. Among the reasons for this difference is their differential treatment of correlated variables with some methods explicitly trying to avoid redundancy. This may in turn lead to a loss of stability as, for example, many genes may encode functionally related proteins [20].

Another potentially relevant property of predictor sets is cost-effectiveness. In particular for multiplexing, addition of a given analyte may be associated with a disproportionately higher cost compared to other predictors. Therefore, it may be desirable to include an estimation of cost-effectiveness that quantifies the balance between the cost of utilizing a given predictor and its impact

on classification performance in the variable selection process [21]. Such considerations are also relevant for multimodal classifiers that aim to extract predictive patterns across multiple technological platforms, including proteomics, RNA sequencing, genome-wide association data, or neuroimaging. For efficient clinical translation of classifiers, an estimation of cost-effectiveness during variable selection is almost mandatory. In addition to including such metrics during variable selection, other approaches have been developed that may show utility for such applications. Examples include the integration of machine learning with constraint programming, where optimal variable combinations are found within the limits of prespecified constraints [22, 23]. Another interesting approach is the utilization of more complex, multimodal data during training of a classifier, but not during testing. “Learning with privileged information” facilitates an adaptation of the parameters learned on such less complex classifiers through the additional data available during training [24–26]. This would allow the utilization of a wider spectrum of data during classifier design, while profiting from improved accuracy, higher cost-effectiveness, and simpler logistics during classifier application. Another related approach is that of multi-task learning, which is based on the assumption that a single classifier cannot learn a given task well from a complex dataset [27, 28]. Instead, multiple classifiers are learned on different tasks that may or may not be related [29]. During this process, information is exchanged between the classifiers. This leads to an inductive adaptation of the original classifier based on the additionally learned classifiers and has previously been applied for multimodal machine learning in the imaging genetics field [30]. An interesting adaptation of this method called robust multi-task learning explores the actual dependency between tasks and may outperform standard multi-task learning [31, 32].

Another interesting consideration is cross-platform transferability. For clinical translation of biomarker tests, signatures that have initially been identified by comprehensive high-throughput platforms, such as DNA microarrays, need to be transferred to more targeted platforms with lower variability like quantitative real-time PCR. Another example where platform migration has been performed successfully is the development of a dedicated 51-plex multiplexed immunoassay panel for aiding the diagnostic process of schizophrenia [33]. In this context, metrics such as the log-ratio discrepancy have been devised for prediction of how successfully a given model can be migrated [34]. The log-ratio discrepancy quantifies the absolute difference in log-ratio as measured between platforms, averaged over all possible pairs of samples and has a minimum value of zero when there is perfect agreement between platforms. However, more research is needed to identify algorithms that can accurately predict platform migration issues using results from a given omic platform, prior to availability of data on the same subjects from the migration target.

---

### 3 Measurement Error, Sample Size, and Generalizability

The main reasons for performing high-throughput experiments such as those with multiplexed assays include the increase in efficiency, the reduction in required sample volume, and, potentially, a reduction of measurement variability. One aspect that is frequently overlooked is that the joint measurement of multiple analytes gives rise to correlated measurement errors, but the effect on subsequent inference of outcome associations is typically not considered. Pollack et al performed an interesting investigation of such effects through modeling of correlated measurement error, correlations between predictors, and measurement error as a function of predictor level under varying levels of overall errors in measurement [35]. This study showed that while bias may be substantial under different settings of correlation and measurement error, outcome associations were mostly biased toward the null. In addition, bias was nearly absent when the true underlying association had an odds ratio of 1. Therefore, correlated measurement error is not likely to lead to false-positive, but rather to false-negative findings [35]. Another interesting conclusion was that measurements that are subject to a lower limit of detection were less affected by bias when values below these limits were replaced by statistically valid substitution values. The transformation of data values regarding effects of measurement reliability and accuracy is an important aspect that should be considered during classifier design.

The variability of measurements that includes both the biological inter-subject variability as well as measurement-related variability also has a crucial impact on the sample size required for particular classification problems. In practice, machine learning approaches are most frequently performed post-hoc on an already existing database. Here, sample size is typically determined by practical considerations (i.e., the patient recruitment rate within a fixed time-window) or more formal considerations, such as sample-size calculation for univariate analysis. However, such considerations rarely include estimations of required sample size for accurate prediction based on multivariate methods. This is statistically an arguably more complex task compared to univariate sample size estimation. Statistical learning theory provides bounds on the generalization performance of a given classifier that depends on the sample size and these bounds can in turn be used to estimate sample sizes for a given target performance [36].

However, such bounds on classification performance are known to be conservative and frequently higher generalizability is observed than predicted by the model. As a consequence, the required sample number would likely be lower to achieve a given performance than predicted by the model. For example, the model predicts an approximation toward the optimal performance proportional to the inverse of the sample number. In practice, such approximation can, however, be observed to be exponential with the sample number, at least in some cases [37].

The required sample number clearly depends on factors such as the complexity of the classification task, the underlying variability in the investigated cohort, and platform measurement specific effects [38]. Therefore, strategies that simulate the required sample size based on pilot data or that perform an adaptive sample size determination may be of higher practical utility compared to theoretical bounds. For example, Shao and coworkers determined the minimum required sample size for microarray prediction of outcomes with various complexities using an adaptive procedure. A signal-to-noise-ratio-based metric that is used to quantify the observed difference between classes was used as a criterion to stop or continue sample collection to obtain an optimally sized training sample [39]. Similarly, Dobbin et al showed that the identification of good classifiers from high-dimensional microarray data depends on standardized fold change, class prevalence, and number of genes or features on the arrays [40]. From this they proposed a method for ex post facto assessment of whether the size of a training set used to develop a classifier was adequate. Hwang et al used power analysis to estimate the minimum sample size required to build a discriminant analysis algorithm on transcriptomics data with statistical reliability [41]. De Valpine applied Monte Carlo simulation to model required sample sizes for classification and concluded that many existing studies will identify suboptimal classification performance, leading to poor validation results, due to limitations in sample size [42]. More recently, Beleites et al used learning curves to estimate the impact of sample size on generalization performance. It was found that the typical sample sizes of up to 25 subjects per class may lead to acceptable classification performance but learning curves can easily be masked by the large uncertainty due to the small sample numbers. In fact, they estimated that 75-100 samples will be required to achieve good classification performance and such samples can subsequently be used for more refined sample size estimation [43]. The simulation of suitable sample sizes should not only consider platform-specific methodological aspects or the nature of the classification task but also the expected heterogeneity of the effect across the target population. For example, a frequent observation is that biomarker signatures vary substantially across clinical sites, even if strict standard operating procedures are followed for sample collection and data acquisition. In particular for complex algorithms that integrate a large number of predictors that show a heterogeneous response to site-to-site variability, normalization of such effects may be difficult and larger training samples will only be useful to the extent that they reflect the heterogeneity in the target population. Variable selection strategies that reduce the complexity of algorithms can help to reduce this problem. In particular, they allow identification of variable subsets that are reproducibly altered across clinical sites, a measure of variable stability in the context of population heterogeneity.

---

## 4 Longitudinal Stability of Classifiers

This section explores a variability-related problem that impacts on classifier performance. High-throughput technologies for identification of biomarker signatures have a non-hypothesis-driven nature and are typically applied during an early stage of clinical tool development. However, high-throughput profiling platforms are for measurement variability, cost, and logistics reasons less suitable for clinical translation of biomarker signatures. These issues may be overcome through migration of the biomarker panel to dedicated measurement platforms, such as multiplex systems targeting the analytes used within the respective classifiers. This is also important from a regulatory perspective, since clinical translation of a given biomarker signature requires fixed algorithms that are specified on one occasion with little possibility for recalibration. This requires high measurement accuracy and substantial stability over extended periods of time. While variability cannot be avoided for many reasons, such as changes in reagent lots, little is known about its effects on the performance of multivariate classifiers. For example, an important question is whether or not the algorithm performance is more sensitive to future measurement noise when they incorporate a higher number of variables to achieve the same given performance. Such error accumulation would have an important impact already on classifier design and implementation of dedicated measurement platforms, since performance optimization would not only relate to current performance on a given dataset but also to future performance under methodological noise.

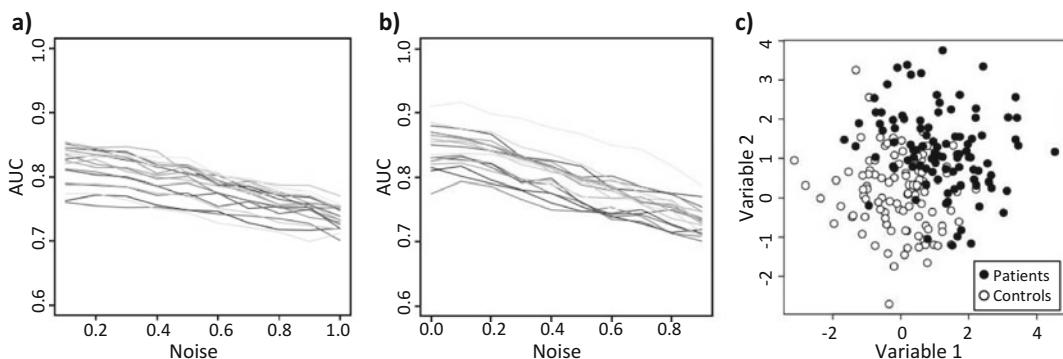
To explore this, we performed a simulation for training classifiers of different complexity and evaluated the impact of noise during testing. The simulation was performed as follows:

1. A training dataset was created that contained 100 “patients” and 100 “controls.”
2. For each group a number ( $n$ ) of uncorrelated variables (with  $n=5$  through  $n=100$ ) were drawn randomly from a normal distribution with a standard deviation of 1.
3. The mean of the distribution was 0 for controls and dynamically adjusted for patients.
4. The adjustment of the patient mean was performed such that the classification accuracy in the training data was 80%, averaged over many repetitions of the procedure.
5. Classification was performed using linear discriminant analysis and accuracy was assessed using cross-validation (the dynamic adjustment of the patient mean, which corresponds in this case to the effect size, is necessary since classifiers with higher number of variables can achieve the same classification accuracy at lower effect sizes across its predictors).

6. A test dataset was simulated using the same parameters as identified for the training data.
7. To each observation in the test data noise from a normal distribution with a mean of 0 and variable standard deviation was added.
8. The classifier trained on the training data was then applied to the test data and classification performance assessed using the area under the receiver operating characteristic curve (AUC).

Figure 1a shows that the decrease in classification performance with higher levels of added noise does not depend on the number of predictors in the classifier. In this example, noise was additive, meaning that the introduced amount of noise did not depend on the magnitude or concentration of the variable it was added to. In practice, variability effects are frequently multiplicative, affecting variables with higher magnitude more compared to those at lower magnitudes. Another characteristic that is often found in real-life datasets is that distributions of variables are heavy-tailed and primarily observations within the tail differentiate patients from control subjects. For this reason, the same simulation described above was performed with the following modifications.

9. During creation of the training data, the absolute value of the normal random variables was used, sampling for both groups from a distribution with a mean of 0, although in this case the standard deviation was dynamically adjusted in patients, leading to increasing portions of the patient group that differed from the controls.



**Fig. 1** Simulation of longitudinal measurement variability effects performance of classifiers with differing numbers of predictors. The figure shows the dependency of performance (AUC) on noise that was additive (panel **a**; noise is taken from a normal distribution with mean 0 and standard deviation is plotted along the x axis) and multiplicative (**b**; noise is taken from a normal distribution with mean 1 and standard deviation is plotted along the x axis). (**c**) Example of a decision boundary for two-dimensional normal distributions associated with average cross-validation AUC of 0.80. In panels (**a**) and (**b**), darker gray colors refer to decision rules with higher numbers of predictors

10. Noise was sampled from a normal distribution with a mean of 1 and varying standard deviation and the absolute values of such noise were multiplied with the simulated test data.

Figure 1b shows that the decrease in classification performance with increasing amounts of noise did not depend on the numbers of predictors within the classifier, as described under the original conditions. These results suggest that noise encountered after classifier design will not affect complex classifiers more than those with fewer predictors. In practice, the situation may be different due to several factors. First, more complex rules frequently integrate a large number of predictors that are substantially more difficult to measure and, therefore, affected by disproportionately higher amounts of longitudinal measurement variability. The expected measurement variability and variance of a given analyte in the population may already be a useful metric during variable selection for designing the classifier. Secondly, the example presented here is based on simulation of simple, linear decision boundaries. A two-dimensional example of this is shown in Fig. 1c. Here, the effect size in each variable was adapted such that linear discriminant analysis yielded an average AUC of 0.80. With non-linear decision boundaries, effects of noise may be stronger, in particular when more advanced algorithms, such as random Forests or neural networks, need to be used. For this reason, it may be advantageous to perform classification with simpler algorithms, such as linear support vector machines, even if variables are selected using more complex methods, such as random Forest variable importance measures. Such hybrid machine learning approaches aim to combine advanced variable selection with improved generalizability and may perform better than either of the methods individually [44].

---

## 5 Conclusions

The application of machine learning algorithms for identification of biological signatures with clinical utility is widespread and spans across most high-throughput profiling technologies. Freely available software packages facilitate intuitive use of such methods making them accessible to a broad community of scientists without the requirement for in-depth knowledge of the algorithms' inner workings. However, the increasing popularity of these approaches and the potential clinical relevance of biomarker signatures also contribute to the problem that machine learning is applied on studies that were not designed for such analyses, which are frequently too small and lead to irreproducible findings. Here some considerations were discussed that may have utility for classifier

development throughout the different phases of biomarker discovery and clinical translation. These comprise approaches to estimate sample sizes appropriate for multivariate classification and potential bounds on test performance. They also include properties of predictors that go beyond their impact on classification accuracy such as stability, cost-effectiveness, or the possibility for platform migration. While these properties are frequently overlooked in favor of optimizing classification accuracy, they play a crucial role when translating a biomarker signature into an assay system for clinical use. Finally, a simulation was performed to estimate the effect of longitudinal measurement variation on a defined classifier. These results suggest that there is no accumulation of noise that would affect algorithms with higher numbers of predictors more than those with fewer. This finding is generally positive for biomarker discovery in complex illnesses, where large numbers of individually weak predictors need to be combined to achieve clinically relevant predictive performance. In practice, it will still be desirable to reduce the complexity of decision rules as far as possible and the considerations discussed here may aid in driving this selection process toward an optimal balance of performance and clinical translatability.

---

## Acknowledgments

This study was supported by the DFG Emmy-Noether-Program SCHW 1768/1-1.

## References

1. Cohen Freue GV, Meredith A, Smith D, Bergman A, Sasaki M, Lam KK et al (2013) Computational biomarker pipeline from discovery to clinical implementation: plasma proteomic biomarkers for cardiac transplantation. *PLoS Comput Biol* 9:e1002963
2. Zhang Z, Chan DW (2010) The road from discovery to clinical diagnostics: lessons learned from the first FDA-cleared in vitro diagnostic multivariate index assay of proteomic biomarkers. *Cancer Epidemiol Biomarkers Prev* 19:2995–2999
3. Alymani NA, Smith MD, Williams DJ, Petty RD (2010) Predictive biomarkers for personalised anti-cancer drug use: discovery to clinical implementation. *Eur J Cancer* 46:869–879
4. Deyati A, Younesi E, Hofmann-Apitius M, Novac N (2013) Challenges and opportunities for oncology biomarker discovery. *Drug Discov Today* 18:614–624
5. Jin G, Zhou X, Wang H, Wong STC (2010) The challenges in blood proteomic biomarker discovery. In: Pham T (ed) *Comput Biol*. Springer, New York, pp 273–299
6. 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
7. Füzéry AK, Levin J, Chan MM, Chan DW (2013) Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. *Clin Proteomics* 10:13
8. Goodsaid F, Mattes WB (2013) The path from biomarker discovery to regulatory qualification. 1 edn., Academic Press. Accessed 16 July 2013. ISBN: 0123914965
9. Kotsiantis SB (2007) Supervised machine learning: a review of classification techniques. *Informatica* 31:249–268. doi:[10.1111/j.1369-124X.2007.01559\\_160.x](https://doi.org/10.1111/j.1369-124X.2007.01559_160.x)

10. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. *Elements* 1:337–387. doi:[10.1007/b94608](https://doi.org/10.1007/b94608)
11. Saeys Y, Inza I, Larrañaga P (2007) A review of feature selection techniques in bioinformatics. *Bioinformatics* 23:2507–2517
12. Kononenko I (2015) Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med* 23:89–109. doi:[10.1016/S0933-3657\(01\)00077-X](https://doi.org/10.1016/S0933-3657(01)00077-X)
13. Caruana R, Niculescu-Mizil A (2006) An empirical comparison of supervised learning algorithms. Proceedings of 23rd international conference machine learning. pp 161–168. doi:[10.1145/1143844.1143865](https://doi.org/10.1145/1143844.1143865)
14. Guo Y, Gruber A, McBurney RN, Balasubramanian R (2010) Sample size and statistical power considerations in high-dimensionality data settings: a comparative study of classification algorithms. *BMC Bioinformatics* 11:447
15. Caruana R, Karampatziakis N, Yessenalina A (2008) An empirical evaluation of supervised learning in high dimensions. Proceedings of 25th international conference machine learning. pp 96–103. doi: [10.1145/1390156.1390169](https://doi.org/10.1145/1390156.1390169)
16. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3:1157–1182
17. He Z, Yu W (2010) Stable feature selection for biomarker discovery. *Comput Biol Chem* 34:215–225
18. Loscalzo S, Yu L, Ding C (2009) Consensus group stable feature selection. Proceedings of 15th ACM SIGKDD international conference on knowledge discovery and data mining. pp 567–576
19. Awada W, Dittman D, Wald R, Napolitano A, Khoshgoftaar TM(2012) A review of the stability of feature selection techniques for bioinformatics data. In: Proceedings of 2012 IEEE 13th international conference information reuse and integration IRI 2012. pp 356–363
20. Haury AC, Gestraud P, Vert JP (2011) The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures. *PLoS One* 6(12), e28210
21. Braun DC, Reynolds JD (2012) Cost-effective variable selection in habitat surveys. *Methods Ecol Evol* 3:388–396
22. Guns T, Nijssen S, De Raedt L (2011) Itemset mining: a constraint programming perspective. *Artif Intell* 175:1951–1983
23. Talbi EG (2013) Combining metaheuristics with mathematical programming, constraint programming and machine learning. *4OR Q J Oper Res* 11:101–150
24. Lapin M, Hein M, Schiele B (2014) Learning using privileged information: SV M+ and weighted SVM. *Neural Netw* 53:95–108
25. Pechony D, Vapnik V (2010) On the theory of learning with privileged information. *Nips* pp 1894–1902
26. Vapnik V, Vashist A (2009) A new learning paradigm: learning using privileged information. *Neural Netw* 22:544–557
27. Chapelle O, Shivaswamy P, Vadrevu S, Weinberger K, Zhang Y, Tseng B (2011) Boosted multi-task learning. *Mach Learn* 85:149–173
28. Evgeniou T, Pontil M (2004) Regularized multi-task learning. Proceedings of 10th ACM SIGKDD pp 109–117
29. Romera-Paredes B, Argyriou A, Pontil M, Berthouze N (2012) Exploiting unrelated tasks in multi-task learning. Proceedings of 15th international conference of artificial intelligence statistics, vol 22, pp 951–959
30. Wang H, Nie F, Huang H, Risacher SL, Saykin AJ, Shen L et al (2012) Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. *Bioinformatics* 28:i127–i136
31. Gong P, Ye J, Zhang C (2012) Robust multi-task feature learning. *KDD* 2012:895–903
32. Ishibuchi H, Nojima Y (2007) Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *Int J Approx Reason* 44:4–31
33. Schwarz E, Izmailov R, Spain M, Barnes A, Mapes JP, Guest PC et al (2010) Validation of a blood-based laboratory test to aid in the confirmation of a diagnosis of schizophrenia. *Biomark Insights* 5:39–47
34. Gyorffy B, Molnar B, Lage H, Szallasi Z, Eklund AC (2009) Evaluation of microarray preprocessing algorithms based on concordance with RT-PCR in clinical samples. *PLoS One* 4(5):e5645
35. Pollack AZ, Perkins NJ, Mumford SL, Ye A, Schisterman EF (2013) Correlated biomarker measurement error: an important threat to inference in environmental epidemiology. *Am J Epidemiol* 177:84–92
36. Shawe-Taylor J, Anthony M, Biggs NL (1993) Bounding sample size with the Vapnik-Chervonenkis dimension. *Discret Appl Math* 42:65–73
37. Cohn D, Tesauro G (1991) How tight are the Vapnik-Chervonenkis bounds? *Neural Comput* 4:249–269
38. Dobbin K, Simon R (2005) Sample size determination in microarray experiments for class comparison and prognostic classification. *Biostatistics* 6:27–38

39. Shao L, Fan X, Cheng N, Wu L, Cheng Y (2013) Determination of minimum training sample size for microarray-based cancer outcome prediction—an empirical assessment. *PLoS One* 8:e68579
40. Dobbin KK, Zhao Y, Simon RM (2008) How large a training set is needed to develop a classifier for microarray data? *Clin Cancer Res* 14:108–114
41. Hwang D, Schmitt WA, Stephanopoulos G, Stephanopoulos G (2002) Determination of minimum sample size and discriminatory expression patterns in microarray data. *Bioinformatics* 18:1184–1193
42. De Valpine P, Bitter HM, Brown MPS, Heller J (2009) A simulation-approximation approach to sample size planning for high-dimensional classification studies. *Biostatistics* 10:424–435
43. Beleites C, Neugebauer U, Bocklitz T et al (2013) Sample size planning for classification models. *Anal Chim Acta* 760:25–33
44. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W et al (2009) A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* 10:213

# **Chapter 7**

## **Opportunities and Challenges of Multiplex Assays: A Machine Learning Perspective**

**Junfang Chen and Emanuel Schwarz**

### **Abstract**

Multiplex assays that allow the simultaneous measurement of multiple analytes in small sample quantities have developed into a widely used technology. Their implementation spans across multiple assay systems and can provide readouts of similar quality as the respective single-plex measures, albeit at far higher throughput. Multiplex assay systems are therefore an important element for biomarker discovery and development strategies but analysis of the derived data can face substantial challenges that may limit the possibility of identifying meaningful biological markers. This chapter gives an overview of opportunities and challenges of multiplexed biomarker analysis, in particular from the perspective of machine learning aimed at identification of predictive biological signatures.

**Key words** Biomarker discovery, Machine learning, Confounding, Bias, Multiplex

---

### **1 Introduction: Opportunities of Multiplexed Assay Systems**

Due to substantial advancements of high-throughput omics technologies, the acquisition of high-dimensional biological datasets has become routine practice. Genetic association, expression, or methylation testing can be performed at the level of the genome- and proteomics often measures in excess of 1000 variables in a given sample [1–3]. The availability of such high-dimensional datasets affords tremendous possibilities of performing a non-hypothesis-driven search for biological patterns that predict a given clinical outcome. This strategy particularly applies to complex clinical outcomes in which individual predictors may have low effect sizes and a combination of numerous predictors is needed to achieve clinically useful accuracy. However, high-throughput measurement of molecular concentrations using proteomic or transcriptomic techniques can be affected by substantial measurement variability and therefore such data has its greatest use during the initial stages of the biomarker development process. As the development of accurate

classifiers is dependent on the longitudinal stability of measured concentrations, a central aim is the reduction of measurement noise. This is an important requirement for validation of biological patterns as otherwise a subtle but predictive biological signature may be drowned out by experimental variability. In this scenario, multiplex assays have substantial utility as they can facilitate the transition from a global biomarker screening tool to a dedicated measurement platform with low variability.

Some previous studies have found the analytical performance of multiplex assays to be comparable to that of single-plex assays [4, 5]. However, such comparability of analytical performance may depend on the measured analytes and the experimental procedures [6]. In addition to reducing measurement noise, multiplexed measurement of biological analytes may be of help to improve throughput. For example, some liquid chromatography-based mass spectrometry methods may not be able to generate datasets of a size sufficient for the identification of accurate biological signatures. Meaningful application of machine learning tools to high-dimensional biological datasets requires large sample numbers, in particular when effects of individual predictors are small. For example, Ein-Dor et al. have shown that thousands of samples are needed for robust prediction of outcome in cancer [7]. Similarly, Kim and coworkers showed that independently generated gene signatures predictive of breast cancer using 600 samples per experiment show an overlap of only 16.5% [8]. While such sample sizes may already exceed the technological possibilities of some omics assay systems, the required sample size is far higher in practice. This is due to the fact that for development of algorithms that generalize well to new samples, the data used for algorithm training needs to reflect the properties of the target population. In practice, this means that the spectrum of physiological inter-individual variability needs to be reflected in both the data and the differences between factors such as ethnicities or clinical sites, an effect that can be substantial even if standard operating procedures are in place [9]. In addition, diagnostic classifiers are rarely useful when purely focusing on case-control differentiation and therefore, training data should incorporate measurements from relevant differential diagnoses. Under these circumstances, multiplex immunoassays offer the possibility of translating promising omics results into an assay system with which high-quality, large-scale data can be acquired for more extensive assessment of predictive algorithm performance. Such assessment can help to overcome small sample effects in training data that are known to frequently lead to false-positive findings [10].

---

## 2 Recent Examples of Multiplex Assays for Disease Diagnosis

An example of multiplexed assay-based biomarker development is the development of a test for schizophrenia [11]. Here, a multiplex panel of 51 analytes was created that combined promising features from proteomics and immunoassay measurements into a single measurement system. This system was then used to acquire protein concentration data on over 800 subjects for algorithm training. Similarly, Surinova et al used a phased mass-spectrometry approach to identify a biomarker signature for colorectal cancer. Candidates derived from an initial profiling of about 300 secreted and cell surface candidate glycoproteins were translated into an 88-plex targeted SRM assay to measure the concentrations of these in large populations [12]. This was ultimately transferred into a five-protein signature with high accuracy.

One of the primary strengths of multiplexed assay systems is the possibility to customize the set of measurement targets to suit the investigated biological system ideally. While the simultaneously measured analytes may cover the most predictive set as described above, it may also try to capture important elements of a given biological pathway or known molecular targets of a given illness. For example, mass spectrometry-based multiplex selected reaction monitoring (SRM) has been used to identify dysregulation of glycolysis-associated enzymes in schizophrenia [13]. Hembrough and colleagues reported a significant improvement in tumor tissue analysis due to the use of multiplexed SRM that enables the accurate measurement of the expression levels of a panel of oncological protein targets [14]. Similarly, Xie et al designed a novel multiplex assay that quantifies autoantibodies against a large collection of clinically vital tumor-associated antigens, combined with a classical cancer biomarker [15]. More recently, Arjomandi et al proposed a novel algorithm to improve the performance of an ovarian cancer detection test using a multiplex approach [16]. In this study, measurements of autoantibodies to p53 in sera of patients against selected confirmatory epitopes were acquired using a multiplexed-based immunoassay. Such a focus on a selected set of target analytes facilitates biologically meaningful stratification of biological measures for multivariate or functional downstream analysis. One example of this is biologically stratified association analyses (also known as “pathway-wide association study,” PWAS), which can aid in differentiating illness relevant variants from background [17, 18]. A related analysis focusing on genetic co-expression networks (“integrative network-based association study,” INAS) has been used to identify illness-associated genes from tissue-to-tissue co-expression networks [17, 19]. Similarly, meaningful biological stratification of measured analytes may help machine learning approaches to

identify biological signatures and remove noise from measurements that are not likely to be illness related. This particularly applies for interactions between measured analytes, which are difficult to extract from noisy, high-dimensional datasets. The integration of interactions in machine learning models such as support vector machines or random forests can aid to improve performance [20–22].

These studies show that the application of multiplexed immunoassays for biomarker development goes beyond large-scale validation of candidate markers identified through omics techniques. In fact, Stahl-Zeng and coworkers have shown that multiplexed selected reaction monitoring of N-glycosites can be used to detect plasma proteins at concentrations in the ng/mL or sub-ng/mL range and quantify their concentrations accurately over five orders of magnitude [23]. Methodological innovations, such as data acquisition paradigms to simultaneously quantify and confirm the identity of the targeted peptides can further expand the multiplex capacity of SRM approaches [24].

---

### 3 Challenges: Biological and Technical Confounders

Multiplexed measurement of biological analytes gives rise to several challenges some of which may affect particularly multivariate downstream analysis. One of the most prominent factors is cross-reactivity and analytes that interact with other antibodies or irrelevant interfering factors that may have to be excluded from analysis [25]. For example, serum rheumatoid factor and other heterophilic antibodies can cause significant interference with antibody-based multiplex immunoassays through binding to detection reagents [26, 27]. Similar cross-reactivity effects are known for non-immunoassay-based measurement platforms, such as SRM. In the case of the latter, ion suppression can impact on the formation of the analyte ions because of sample-matrix interactions [28, 29].

Another potential challenge affecting multiplex assays is that assay performance may vary between different laboratories or platforms and may therefore impede multi-site biomarker development efforts. Zhang et al. investigated the impact of variability between three different laboratories and multiplex systems on the clinical interpretation of pneumococcal serology assay measurements in 57 antibody-deficient patients [30]. Despite substantial variability of quantitative antibody levels, a decision concordance of up to 80% was determined across laboratories. This suggests that the integration of multiplex data using multivariate algorithms may overcome the problems of variable individual predictors. However, this study also highlights the need for novel paradigms to control multiplex data quality across reference laboratories to help reduce experimental variation [30, 31].

While technical variation is typically lower than biological differences, some sources of such variability do not only hinder identification of predictive signatures but they can lead to discovery of artefactual illness fingerprints [32, 33]. Frequently, multivariate and machine learning analyses are performed using multiplex data that were not acquired for such analyses and therefore study design may not be suitable to exclude some confounding factors. This particularly relates to batch effects that are virtually unavoidable for high-throughput measurements [34] and which can be substantially amplified through aggregation of their effects across measured analytes. Such effects include changes in personnel, different laboratories, different run dates, and inconsistent experimental design between assays or laboratories. If such effects are associated with the outcome of interest they may lead to substantial bias and application of statistical remedies may not be sufficient to remove their effects on the data [34].

Clarke and colleagues systematically investigated technical confounding factors in multiplex immunoassay data [33]. Mixed-effects modeling was used to eliminate the effects of both the technical and biological sources of variation followed by the analytical evaluation of normalized multiplex data. Similarly, Browne et al. applied a statistical batch correction method to reduce variance between measurement plates for multiplex cytokine assays in serum and saliva [35]. While such methods reduce the impact of batch effects on downstream analysis, residual variation in strongly confounded data may still impact particularly on sensitive machine learning algorithms. Soneson et al. have used different machine learning algorithms to explore the influence of batch effect severity on classification performance [36]. When high levels of the confounding factor existed, the performance estimates obtained from cross-validation were found to be highly biased. Parker and coworkers developed a novel method of batch effect removal for multivariate prediction that they called frozen surrogate variable analysis [37]. This approach uses a training dataset to correct batch effects at the individual subject level. As batch effects may remain undetected especially in smaller cohorts, multiplex assays and careful study design may help to identify less biased biological signatures that show better generalizability to new samples. Also, some multiplex systems have been optimized for improved longitudinal measurement stability, thereby reducing problems caused by batch-to-batch assay variation [38].

Finally, an interesting challenge of multiplexed assays is that a given factor of variability might simultaneously affect the different measured analytes and therefore lead to correlated measurement error. Pollack et al. explored the impact of such effects on downstream inference on outcome associations and found that outcome associations were mostly biased toward the null [39]. This effect is therefore more likely to lead to false-negative rather than false-positive outcome associations.

## 4 Conclusions

The combination of experimental and computational methods is indispensable for the identification of novel biomarkers and the exploration of biological mechanism of complex illnesses. Due to their high throughput, typically low technical variation, and the possibility to efficiently acquire data on biological analytes that are difficult to measure otherwise, multiplex assays can be an important element of biomarker development strategies [40]. At the same time, their application faces several challenges that may particularly impact on sensitive downstream machine learning analysis. These may reduce the ability to identify biological illness signatures or lead to artefactual findings when sources of variation are associated with the clinical outcome of interest. While statistical methods exist to remove such effects in multiplex data, residual confounding factors may remain and require the generation of independent data for validation of biomarker signatures. Finally, sharing of data, protocols, and experimental designs as well as strategies to ensure consistent data quality across laboratories would aid in validation efforts of promising biomarker signatures.

## Acknowledgments

This study was supported by the DFG Emmy-Noether-Program SCHW 1768/1-1.

## References

- Gutstein HB, Morris JS, Annangudi SP, Sweedler JV (2008) Microproteomics: analysis of protein diversity in small samples. *Mass Spectrom Rev* 27:316–330
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM et al (2011) High density DNA methylation array with single CpG site resolution. *Genomics* 98:288–295
- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530–536
- Elshal MF, McCoy JP (2006) Multiplex bead array assays: performance evaluation and comparison of sensitivity to ELISA. *Methods* 38:317–323
- Dupont NC, Wang K, Wadhwa PD, Culhane JF, Nelson EL (2005) Validation and comparison of luminex multiplex cytokine analysis kits with ELISA: determinations of a panel of nine cytokines in clinical sample culture supernatants. *J Reprod Immunol* 66:175–191
- Tighe PJ, Ryder RR, Todd I, Fairclough LC (2015) ELISA in the multiplex era: potentials and pitfalls. *Proteomics Clin Appl* 9:406–422
- Ein-Dor L, Zuk O, Domany E (2006) Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci U S A* 103:5923–5928
- Kim SY (2009) Effects of sample size on robustness and prediction accuracy of a prognostic gene signature. *BMC Bioinformatics* 10:147
- Ellington AA, Kullo IJ, Bailey KR, Klee GG (2009) Measurement and quality control issues in multiplex protein assays: a case study. *Clin Chem* 55:1092–1099
- Mollenhauer B, Parnetti L, Rektorova I, Kramberger MG, Pikkarainen M, Schulz-Schaeffer WJ et al (2015) Biological confounders for the values of cerebrospinal fluid proteins in Parkinson's disease and related disorders. *J Neurochem*. doi: [10.1111/jnc.13390](https://doi.org/10.1111/jnc.13390). [Epub ahead of print]

11. Schwarz E, Izmailov R, Spain M, Barnes A, Mapes JP, Guest PC et al (2010) Validation of a blood-based laboratory test to aid in the confirmation of a diagnosis of schizophrenia. *Biomark Insights* 12:39–47
12. Surinova S, Choi M, Tao S, Schüffler PJ, Chang CY, Clough T et al (2015) Prediction of colorectal cancer diagnosis based on circulating plasma proteins. *EMBO Mol Med* 7:1166–1178
13. Martins-de-Souza D, Alsaif M, Ernst A, Harris LW, Aerts N, Lenaerts I et al (2012) The application of selective reaction monitoring confirms dysregulation of glycolysis in a preclinical model of schizophrenia. *BMC Res Notes* 5:146
14. Hembrough T, Thyparambil S, Liao WL, Darfler MM, Abdo J, Bengali KM et al (2013) Application of selected reaction monitoring for multiplex quantification of clinically validated biomarkers in formalin-fixed, paraffin-embedded tumor tissue. *J Mol Diagn* 15:454–465
15. Xie C, Kim HJ, Haw JG, Kalbasi A, Gardner BK, Li G et al (2011) A novel multiplex assay combining autoantibodies plus PSA has potential implications for classification of prostate cancer from non-malignant cases. *J Transl Med* 9:43
16. Arjomandi A, Delanoy ML, Walker RP, Binder SR (2015) A novel algorithm to improve specificity in ovarian cancer detection. *Clin Ovarian Other Gynecol Cancer* DOI: <http://dx.doi.org/10.1016/j.cogc.2015.05.002>
17. Califano A, Butte AJ, Friend S, Ideker T, Schadt E (2012) Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat Genet* 44:841–847
18. Wang K, Li M, Hakonarson H (2010) Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 11:843–854
19. Dobrin R, Zhu J, Molony C, Argman C, Parrish ML, Carlson S et al (2009) Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome Biol* 10:R55
20. Zhao Z, Liu H (2007) Searching for interacting features. Published in: IJCAI'07 Proceedings of the 20th international joint conference on artificial intelligence. pp 1156–1161
21. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
22. Ho TK (1995) Random decision forests. In: Document analysis and recognition, proceedings of the third international conference on IEEE, vol 1, pp 278–282
23. Stahl-Zeng J, Lange V, Ossola R, Eckhardt K, Krek W, Aebersold R et al (2007) High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics* 6:1809–1817
24. Kiyonami R, Schoen A, Prakash A, Zabrouskov V, Peterman S, Picotti P et al (2011) Increased selectivity, analytical precision, and throughput in targeted proteomics. *Mol Cell Proteomics* 10:M110–M002931. doi:[10.1074/mcp.M110.002931](https://doi.org/10.1074/mcp.M110.002931). Epub Jul 27
25. Skogstrand K, Thorsen P, Nørgaard-Pedersen B, Schendel DE, Sørensen LC, Hougaard DM (2005) Simultaneous measurement of 25 inflammatory markers and neurotrophins in neonatal dried blood spots by immunoassay with xMAP technology. *Clin Chem* 51:1854–1866
26. Todd DJ, Knowlton N, Amato M, Frank MB, Schur PH, Izmailova ES et al (2011) Erroneous augmentation of multiplex assay measurements in patients with rheumatoid arthritis due to heterophilic binding by serum rheumatoid factor. *Arthritis Rheum* 63:894–903
27. Churchman SM, Geiler J, Parmar R, Horner EA, Church LD, Emery P et al (2012) Multiplexing immunoassays for cytokine detection in the serum of patients with rheumatoid arthritis: lack of sensitivity and interference by rheumatoid factor. *Clin Exp Rheumatol* 30:534–542
28. Annesley TM (2003) Ion suppression in mass spectrometry. *Clin Chem* 49:1041–1044
29. Grebe SK, Singh RJ (2011) LC-MS/MS in the clinical laboratory—where to from here. *Clin Biochem Rev* 32:5–31
30. Zhang X, Simmerman K, Yen-Lieberman B, Daly TM (2013) Impact of analytical variability on clinical interpretation of multiplex pneumococcal serology assays. *Clin Vaccine Immunol* 20:957–961
31. Hill HR, Pickering JW (2013) Reference laboratory agreement on multianalyte pneumococcal antibody results: an absolute must! *Clin Vaccine Immunol* 20:955–956
32. Blankley RT, Fisher C, Westwood M, North R, Baker PN, Walker MJ et al (2013) A label-free selected reaction monitoring workflow identifies a subset of pregnancy specific glycoproteins as potential predictive markers of early-onset pre-eclampsia. *Mol Cell Proteomics* 12:3148–3159
33. Clarke DC, Morris MK, Lauffenburger DA (2010) Normalization and statistical analysis of multiplexed bead-based immunoassay data using mixed-effects modeling. *Mol Cell Proteomics* 12:245–262
34. 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* 11:733–739
- 35. Browne RW, Kantarci A, LaMonte MJ, Andrews CA, Hovey KM, Falkner KL et al (2013) Performance of multiplex cytokine assays in serum and saliva among community-dwelling postmenopausal women. *PLoS One* 8, e59498
  - 36. Soneson C, Gerster S, Delorenzi M (2014) Batch effect confounding leads to strong bias in performance estimates obtained by cross-validation. *PLoS One* 9, e100335
  - 37. Parker HS, Corrada Bravo H, Leek JT (2014) Removing batch effects for prediction problems with frozen surrogate variable analysis. *Peer J* 2, e561
  - 38. Cham GK, Kurtis J, Lusingu J, Theander TG, Jensen AT, Turner L (2008) A semi-automated multiplex high-throughput assay for measuring IgG antibodies against *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) domains in small volumes of plasma. *Malar J* 7:108
  - 39. Pollack AZ, Perkins NJ, Mumford SL, Ye A, Schisterman EF (2013) Correlated biomarker measurement error: an important threat to inference in environmental epidemiology. *Am J Epidemiol* 177:84–92
  - 40. Mani A, Ravindran R, Manepalli S, Vang D, Luciw PA, Hogarth M et al (2015) Data mining strategies to improve multiplex microbead immunoassay tolerance in a mouse model of infectious diseases. *PLoS One* 10:e0116262

# **Part III**

## **Protocols**

# Chapter 8

## Multiplex Analyses Using Real-Time Quantitative PCR

Steve F.C. Hawkins and Paul C. Guest

### Abstract

Quantitative polymerase chain reaction (qPCR) is a routinely used method for the detection and quantitation of gene expression in real time. Multiplex qPCR requires the use of probe-based assays, in which each probe is labeled with a unique fluorescent dye, resulting in different observed colors for each assay. The signal from each dye is used to quantitate the amount of each target separately in the same tube or well. The availability to multiplex therefore allows the measurement of the expression levels of several targets or genes of interest quickly. Here, we describe a method using the SensiFAST and SensiFAST One-Step probe kits which allows simultaneous real-time quantitation of up to 5 amplicons.

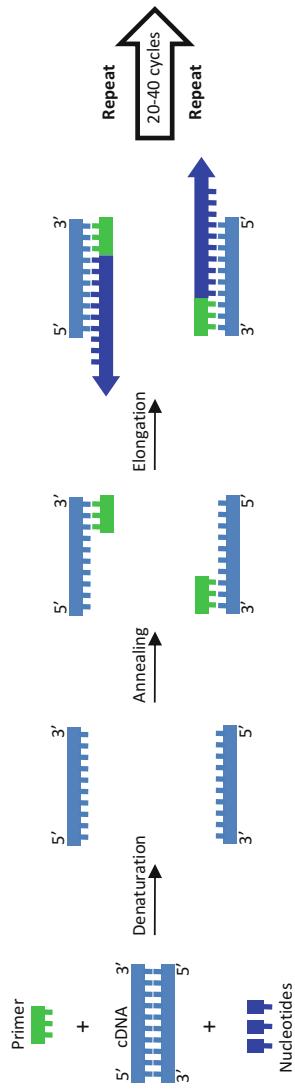
**Key words** qPCR, Fluorescent dyes, Taq polymerase, Quantitation, mRNA, cDNA, Amplicon, Multiplex analysis

---

### 1 Introduction

Polymerase chain reaction (PCR) method was a revolutionary innovation by Kary Mullis in the 1980s [1, 2]. Since this time, it has seen widespread use in biomedical research since it can detect and quantify small amounts of specific nucleic acid sequences. For example, small levels of messenger RNA (mRNA) can be quantified through the combination of reverse transcription (RT) to yield complementary DNA (cDNA) and PCR amplification to produce exponentially higher levels of these cDNA strands [3] (Fig. 1). In addition to increased levels of the amplified products (amplicons), the reliability and reproducibility of measurements between different laboratories are essential, especially if the method is to be performed in a clinical setting. This is critical for patient outcomes as well as for reducing healthcare costs since approximately one third of medical care budgets result from measurements and tests associated with diagnosis [4].

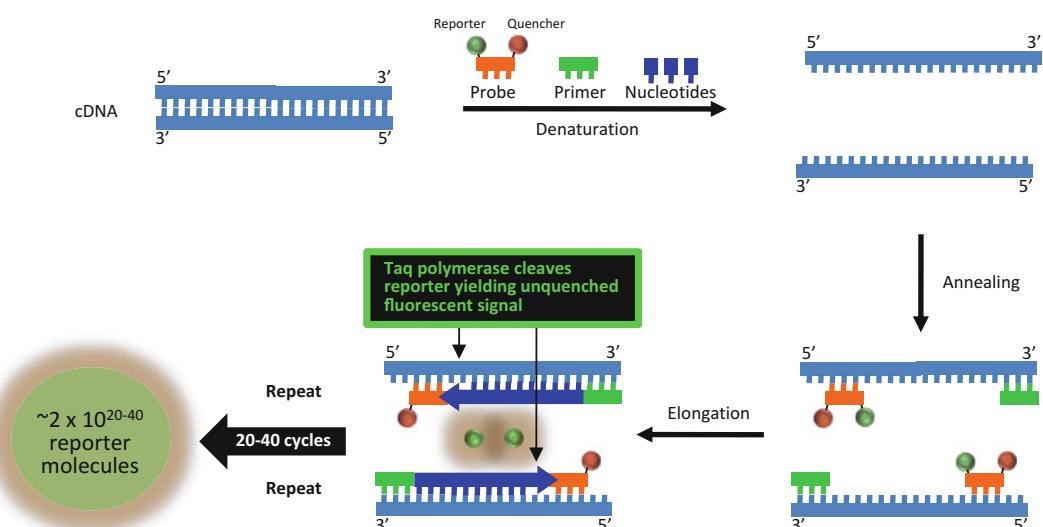
Quantitative PCR (qPCR) is a later development of the method that allows users to monitor the progress of a PCR reaction in real time [5]. In brief, the method uses a DNA-based sequence-specific



**Fig. 1** Schematic diagram of PCR

probe with a fluorescent reporter molecule at one end and a molecule that quenches this fluorescence at the other. The proximity of the reporter to the quench molecule prevents the detection of fluorescence and cleavage of the probe by the 5' to 3' exonuclease activity of Taq polymerase results in unquenched emission of fluorescence. Thus, the increase in the cDNA amplicon targeted by the reporter probe during each PCR cycle leads to a proportional increase in fluorescence due to cleavage of the probe and release of the reporter (Fig. 2). The available fluorescent reporter molecules include dyes that bind to double-stranded DNA such as SYBR® Green (Thermo Fisher Scientific; Waltham, MA, USA) or sequence specific probes like Molecular Beacons (Newark, NJ, USA), Scorpions (DxS Ltd), or TaqMan® Probes (Roche Molecular Diagnostics; Basel, Switzerland). As with standard PCR, qPCR is normally performed using a thermal cycler, which can rapidly heat and cool samples to allow the melting, annealing, and extension phases of replication. However in the case of qPCR, the thermocycler should also have the ability to illuminate each sample with specific wavelengths of light for the detection of the fluorescence emitted following excitation of the probe.

PCR normally consists of a series of temperature changes that are repeated approximately 30 times. Each cycle consists of two or three steps. In the three step cycling approach, the first step is carried out at approximately 95 °C, which allows separation of the double-stranded nucleic acid chains (denaturation). The second phase is performed at around 55 °C to allow binding of the primers to the DNA/cDNA template (annealing). Finally, the third step is carried out at 72 °C to facilitate polymerization using DNA polymerase (elongation). In the two step cycling method, the



**Fig. 2** Schematic diagram of real-time qPCR procedure

annealing and elongation steps are combined at the temperature annealing temperature. In qPCR, it should be noted that 40 cycles are performed and that the temperatures and associated times used in each cycle depend on a variety of factors, such as the polymerase used, the concentration of deoxyribonucleotides (dNTPs), and the optimum binding temperature of the primers.

In general, two basic methods are used in qPCR and these are based on either relative quantification and absolute quantification [6]. Relative quantification is based on comparisons with standard DNA/cDNAs within the sample for measurement of ratiometric differences. The absolute quantitation approach can yield the precise number of resulting amplicons by comparison with DNA standards using a calibration curve. This requires that PCR of the DNA/cDNA in the sample and the standard have the same amplification efficiency. In addition to widespread use in research studies, qPCR has already been applied in many studies for the discovery of biomarkers for applications in clinical studies such as evaluating the status of certain cancers or for monitoring disease progression or treatment response [7–9]. Here, we describe the use of the SensiFAST™ Probe (Bioline; London, UK) that uses a unique buffer chemistry to enable fast and reproducible multiplex qPCR determinations. This property makes this an ideal approach for routine clinical use.

---

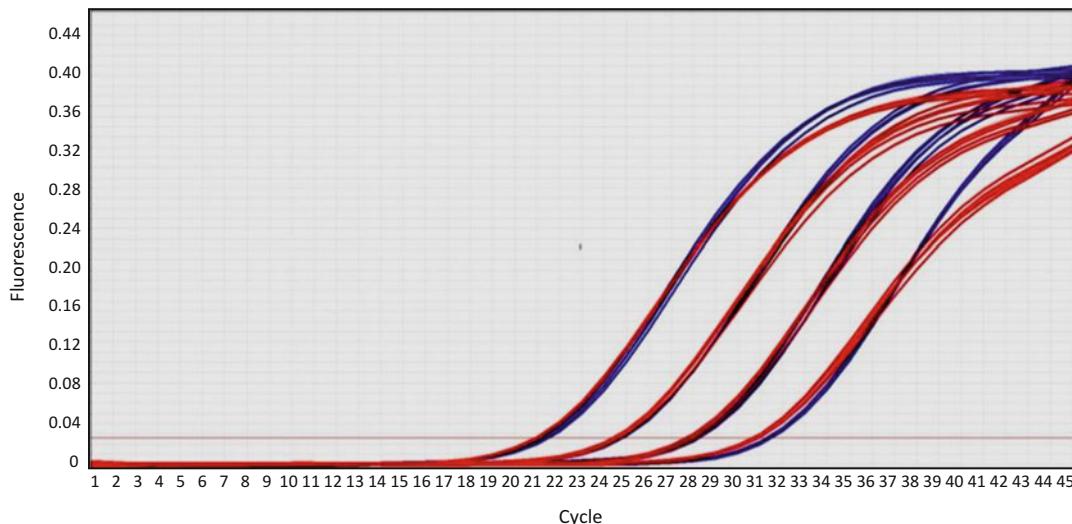
## 2 Materials (See Note 1)

1. 400 nM oligonucleotide primers (*see Notes 2 and 3*).
2. 100 nm probes (*see Notes 4 and 5*).
3. Templates: approximately 1 mg genomic DNA or 100 ng cDNA or  $1 \times 10^{-6}$ –1.0 µg total RNA or 0.01 pg mRNA (*see Note 6*).
4. 1× SensiFAST Probe Mix, containing hot-start DNA polymerase, dNTPs, stabilizers, and enhancers.
5. Reverse transcriptase (*see Note 7*).
6. RNase inhibitor (*see Note 7*).
7. qPCR thermocycler (*see Note 8*).

---

## 3 Methods (See Note 9)

1. Isolate DNA or RNA as required using standard methods.
2. Select amplicons of interest (*see Note 10*).
3. For DNA and cDNA templates, prepare a PCR master mix based on a standard 20 µL final reaction volume containing the primers, probes, template, and probe mix (*see Notes 6 and 11*).



**Fig. 3** Five replicates were run using a conventional TaqMan primer/probe set under fast cycling conditions (3 min 95 °C followed by 45 cycles at 95 °C for 10 s and 60 °C for 10 s). Singleplex reactions (blue line) and quadruplex reaction (red line) for the  $\gamma$ -actin and JOE dye are indistinguishable. However, slightly lower fluorescence intensity is often seen for the multiplex reactions, as reagents are consumed more quickly

4. For total RNA and mRNA templates, prepare a PCR master mix based on a standard 20  $\mu$ L final reaction volume containing the 1:100 reverse transcriptase, 1:50 RiboSafe RNase Inhibitor, primers, probes, template, and probe mix (*see Notes 6 and 11*).
5. Suggested thermal cycling conditions for DNA and cDNA: 1 cycle at 95 °C for 2–5 min for polymerase activation, then 40 cycles at 95 °C for 10 s for denaturation and 60 °C for 20–50 s for annealing/extension (*see Notes 12 and 13*) (Fig. 3).
6. Suggested thermal cycling conditions for RNA: 1 cycle at 45 °C for 10 min for reverse transcription, 1 cycle at 95 °C for 2 min for polymerase activation, then 40 cycles at 95 °C for 5 s for denaturation and 60 °C for 20 s for annealing/extension (*see Notes 13 and 14*).
7. Data analysis (*see Note 15*).

#### 4 Notes

1. These guidelines refer to the design and setup of TaqMan probe-based PCR. Please refer to the relevant literature when using other probe types. The specific amplification, yield, and overall efficiency of any qPCR can be critically affected by the sequence and concentration of the probes and primers, and amplicon length.

2. Use primer-design software, such as Primer3 (<http://frodo.wi.mit.edu/primer3/>) or visual OMP™ (<http://dnasoftware.com/>). Primers should have a melting temperature (Tm) of approximately 60 °C; the Tm of the probe should be approximately 10 °C higher than that of the primers. Tm can be determined using software; however, a good approximation is  $(2 \times \text{number of As and Ts}) + (4 \times \text{number of Gs and Cs})$ . Importantly, a 40–60% GC content is recommended for all primers and to avoid long stretches of any one base. There is a range of about 6–8 °C over which the PCR will work well. The closer you are to the top of this range, the more specificity you will have. For fast reaction kits such as the SensiFAST we also recommend adding a further 5 °C as they have a higher salt concentration.
3. A final primer concentration of 400 nM is suitable for most probe-based reactions; however to determine the optimal concentration we recommend titrating in the range 200–1000 nM. The forward and reverse primers concentration should be equimolar. We recommend aliquoting the primers to avoid repeated freeze/thaw of the primary source, as this will effect PCR efficiency and sensitivity. Aliquots should be used for up to six freeze/thaw cycles.
4. A probe concentration of 100 nM is recommended for multiplexing since higher concentrations can result in cross-channel fluorescence. Probe sequence should be designed as above.
5. For each probe, consider the spectral properties of the dyes in terms of intensity of fluorescence and spectral overlap with other dyes within the reaction. It is important to determine the dyes for which your qPCR instrument has been calibrated or is capable of detecting once calibrated. The manufacturer can provide instrument excitation and detectable emission wavelengths (Table 1). Some of the older instruments require the use of a passive reference such as ROX or fluorescein, to normalize expression levels between wells of a 96 or 384-well plate. If normalization is required with these instruments, this will reduce the selection of fluorescent dyes that can be used.
6. It is important that the DNA template is suitable in terms of purity and concentration. The template must be devoid of any contaminating PCR inhibitors (e.g., EDTA). The recommended amount of template for PCR is dependent upon the type of DNA used. For genomic DNA, use up to 1 mg extracted DNA using a kit such as the Bioline ISOLATE II Genomic DNA Kit or a phenol/chloroform-based method such as Bioline TriSure (ensure that samples are washed thoroughly as even small amounts of phenol are inhibitory to PCR). For cDNA, highly pure RNA is recommended and to perform a two-step RT-PCR. It is also important to use a pre-optimized

**Table 1**  
**Common fluorophores and quenchers used for qPCR probes**

| Fluorophore | Absorption (nm) | Emission (nm) | Suggested compatible quencher |
|-------------|-----------------|---------------|-------------------------------|
| FAM         | 495             | 517           | TAMRA, BHQ-1, Dabcyl          |
| JOE         | 520             | 548           | TAMRA, BHQ-1, Dabcyl          |
| VIC         | 528             | 546           | TAMRA, BHQ-1, Dabcyl          |
| HEX         | 537             | 553           | TAMRA, BHQ-1, Dabcyl          |
| NED         | 546             | 575           | TAMRA, BHQ-1, Dabcyl          |
| TAMRA       | 550             | 576           | BHQ-2                         |
| Cy3         | 550             | 570           | BHQ-2                         |
| ROX         | 581             | 607           | BHQ-2                         |
| Cy5         | 650             | 667           | BHQ-2/BHQ-3                   |

mix such as the SensiFAST cDNA Synthesis Kit for reverse transcription. The optimal amount of cDNA to use in a single PCR is dependent upon the copy number of the target gene. We suggest using 100 ng cDNA per reaction; however, it may be necessary to vary this amount. For RNA, it is important that the template is intact and devoid of DNA or contaminating inhibitors of both reverse transcription and PCR. The recommended amount of template for one-step real-time RT-PCR is dependent upon the type of RNA used. For total RNA, we recommend using  $1 \times 10^{-6}$  pg to 1  $\mu$ g and for mRNA 0.01 pg per 20  $\mu$ L reaction.

7. For use with RNA templates. It is important to use an RNAase inhibitor such as the RiboSafe RNase inhibitor (Bioline) although others can be used.
8. Many instruments can be used here such as the 7500 FAST, 7900HT FAST, ViiA7™, and StepOne™ from Applied Biosystems (Waltham, Massachusetts, USA), the Mx4000™ from Stratagene/Agilent (Santa Clara, California, USA), the iCycler™ and MyiQ5™ from Bio-Rad (Hercules, California, USA), the LightCycler® from Roche (Basel, Switzerland), the RotorGene™ from Qiagen (Hilden, Germany), and the MIC from Bio Molecular Systems (Upper Coomera, Queensland, Australia). Although other instruments can be used but it is important to check with the manufacturer for compatibility, as described above.
9. qPCR is extremely sensitive and so to help prevent any carry-over DNA contamination, separate areas for reaction setup, PCR amplification and any post-PCR gel analysis should be maintained. It is essential that any tubes containing amplified

PCR product are not opened in the PCR setup area. As with all types of PCR, follow the three-room rule. One of the biggest causes of contamination and background is from using the same pipettes for extraction, PCR setup, and post-run analysis. Even if aerosol resistant tips are used all the time, this is not a good idea. Instead, you should have a dedicated set of pipettes for each stage. In addition to pipettes, you should have a different location, either hoods with UV lamps (or preferably a completely different room) for extractions, PCR setup, and any post PCR analysis. In addition, it is important to detect the presence of contaminating DNA that may affect the reliability of the data by including a no-template control reaction, replacing the template with PCR-grade water.

10. For multiplex analyses, the length of the amplicons should be similar and between 50 and 150 bp for optimal PCR efficiency. This is important since each amplicon competes for the same reagents in the probe mix (dNTPs and polymerase).
11. Always mix the reagents well before use. This may sound obvious but this is a very sensitive system and the reagents contain dyes, dNTPs, and enzymes that may have settled while sitting in the freezer or refrigerator.
12. The conditions are suitable for the SensiFAST Probe Kit targeting amplicons up to 200 bp. For the polymerase activation step, 2 min are required for cDNA and 5 min for genomic DNA. For all other steps, the temperatures may vary depending on the primer sequences and up to 50 cycles may be required in multiplex experiments.
13. When testing a mix, template, or primers, it is important to amplify from a tenfold template dilution series. Loss of detection at low template concentrations is the only direct measurement of sensitivity and can also indicate the presence of inhibitors. If inhibition is observed, either the DNA needs to be used at lower concentrations or it requires re-purification. Ideally, samples should cross the threshold ( $C_t$ ) between cycles 20–30. Therefore, individual reactions should be optimized prior to multiplexing, with efficiencies as close to 100% as possible.
14. The conditions are suitable for the SensiFAST Probe One-Step Kit, targetting amplicons of up to 200 bp. However, they can be varied to suit different machine-specific protocols. The reverse transcription reaction time can be extended up to 20 min and/or the temperature can be increased up to 48 °C. For the annealing/extension stage, temperatures may vary depending on primer sequences and up to 50 s may be necessary for multiplexing with more than two probes.

15. Optimal analysis settings, such as baseline and threshold values, for each primer/probe set are a prerequisite for accurate quantification data. It is therefore important to analyze the data for each channel separately as the qPCR instrument default settings may not provide accurate results. It is recommended to keep the multiplex reactions after amplification so that if there is any doubt in the results, the PCR products can be checked on an agarose gel.

## References

1. Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 155:335–350
2. Mullis KB (1990) The unusual origin of the polymerase chain reaction. *Sci Am* 262:56–61
3. Higuchi R, Fockler C, Dollinger G, Watson R (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology* 11:1026–1030
4. <http://www.managedcaremag.com/archives/2009/5/managing-cost-diagnosis>
5. Valasek MA, Repa JJ (2005) The power of real-time PCR. *Adv Physiol Educ* 29:151–159
6. Wong ML, Medrano JF (2005) Real-time PCR for mRNA quantitation. *Biotechniques* 39:75–85
7. Deng Q, Yang H, Lin Y, Qiu Y, Gu X, He P et al (2014) Prognostic value of ERCC1 mRNA expression in non-small cell lung cancer, breast cancer, and gastric cancer in patients from Southern China. *Int J Clin Exp Pathol* 7:8312–8321
8. Böttcher R, Henderson DJ, Dulla K, van Strijp D, Waanders LF, Tevez G et al (2015) Human phosphodiesterase 4D7 (PDE4D7) expression is increased in TMPRSS2-ERG-positive primary prostate cancer and independently adds to a reduced risk of post-surgical disease progression. *Br J Cancer* 113:1502–1511
9. Bahnassy A, Mohanad M, Ismail MF, Shaarawy S, El-Bastawisy A, Zekri AR (2015) Molecular biomarkers for prediction of response to treatment and survival in triple negative breast cancer patients from Egypt. *Exp Mol Pathol* 99:303–311

# Chapter 9

## Multiplex Analysis Using cDNA Transcriptomic Profiling

Steve F.C. Hawkins and Paul C. Guest

### Abstract

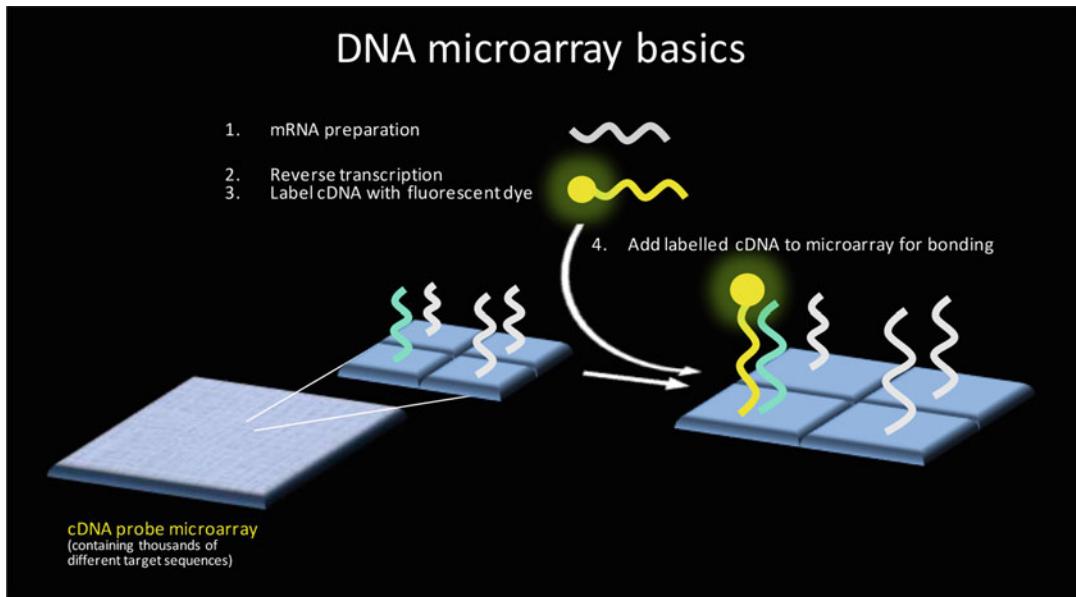
DNA microarrays contain microscopic DNA spots attached to a solid surface. Each spot contains picomolar levels of a specific DNA probe sequence and hybridization to the corresponding gene products can be detected and quantitated through the use of fluorescently labeled target DNA. In this format, DNA microarrays can be used to measure the expression level of thousands of genes in a single experiment. Here, we present a method to detect the mRNA transcriptional changes in neuronal precursor cells following differentiation using high density cDNA microarrays.

**Key words** Fluorescent dyes, Taq polymerase, Quantitation, mRNA, cDNA, Multiplex analysis

---

### 1 Introduction

The DNA microarray technique was developed originally in a rudimentary format in 1975 [1] but emerged more or less in its present form beginning in the 1990s [2, 3]. This approach gave scientists the capability of simultaneously profiling the expressed genes, or messenger RNAs (mRNAs), within a cell, tissue, or organism for the first time. As with most molecular biology detection methods, the DNA microarray approach works by exploiting the quality of nucleic acid sequences to hybridize with other nucleic acids and form complementary sequences. Microarrays consist of a pre-designed set of synthetic nucleic acid probes immobilized and arrayed in a grid-like pattern on a solid surface [4]. Microarrays were first made by immobilizing probes onto filter paper although this progressed to attaching the probes to solid surfaces, such as glass or silicon chips [5–7]. Currently, most approaches use a robot to print nucleic acid probes onto a chemical matrix or they employ a photoactivated chemistry and masking technique to synthesize probes directly on the surface of choice in predesignated locations. In a typical protocol, reverse transcription (RT) of extracted mRNA is carried out in the presence of the four nucleotides, one of which is labeled with a fluorescent dye, to generate labeled complementary

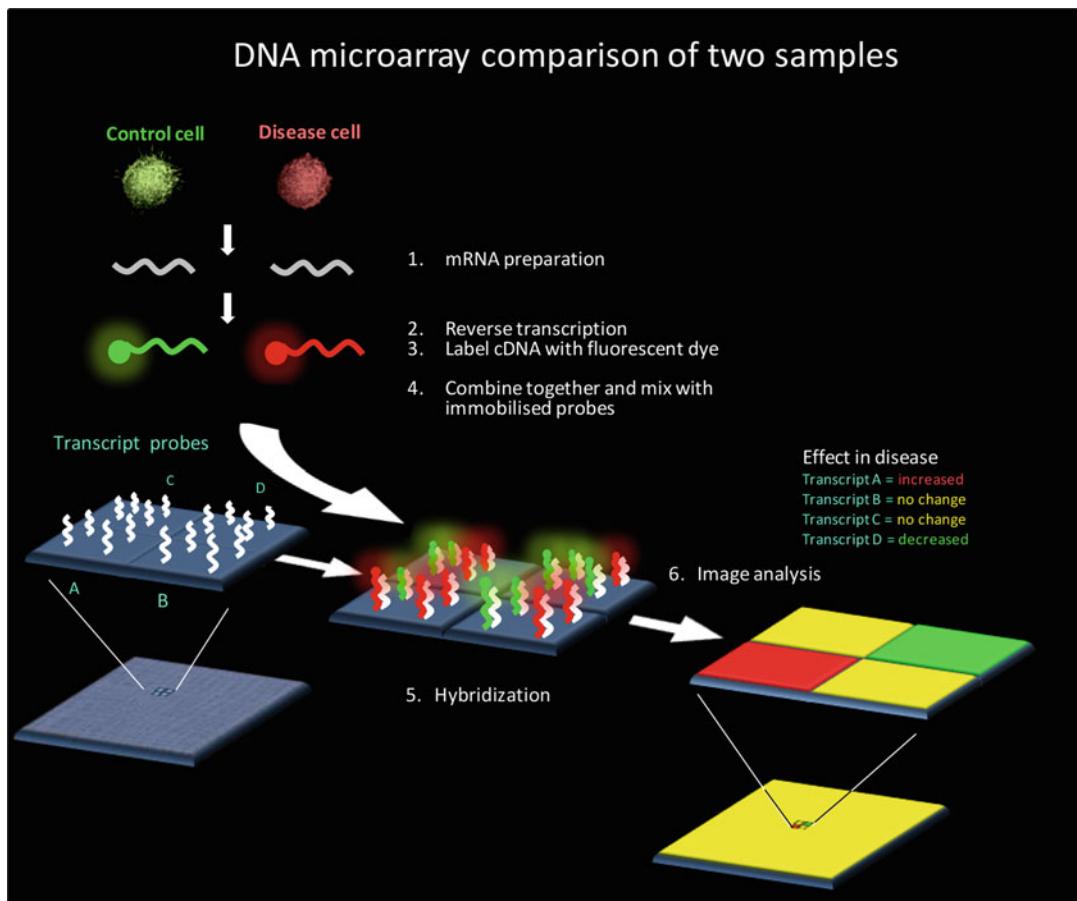


**Fig. 1** Schematic diagram of the cDNA microarray technique

DNA (cDNA) that can then be hybridized with the probe array (Fig. 1).

In comparative microarray studies, one sample usually serves as the control (e.g., a sample from a healthy person) and the other as an experimental sample (e.g., a sample from an individual with a specific disease). This begins with isolation of mRNA from the two samples followed by RT using a nucleotide mixture labeled with dyes that fluoresce at different wavelengths. This is done so that the two samples can be hybridized and eventually distinguished on the same microarray [8]. The hybridization step involves addition of the labeled cDNA mixture onto the microarray where this will bind to the attached cDNA probes, followed by a series of washes to remove the nonspecific cDNAs. Following this, the fluorescent labels on captured cDNA strands are excited by a laser and these release light at a specific wavelength and intensity. The strength of the released light depends upon the amount of target sample binding to the probes present on the corresponding spot in the microarray. This correlates to the amount of each specific mRNA in the sample. Therefore, calculating the ratio of the two dyes in a comparative DNA array experiment yields the relative levels of the same mRNA in the associated samples (Fig. 2).

Microarray technology has high multiplexing capacity with the capability of simultaneously profiling thousands of mRNA transcripts. It has already been used in numerous studies covering such applications as single nucleotide polymorphism identification, studies of cell fate and disease mechanisms [5, 9, 10]. Here, we

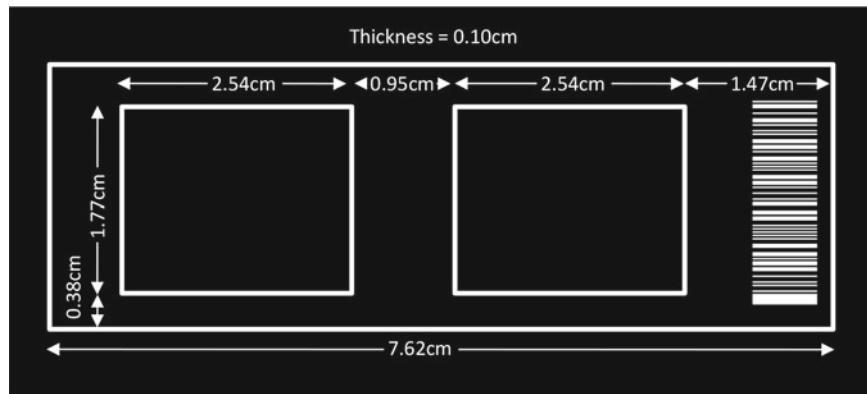


**Fig. 2** Schematic diagram of a comparative microarray study. In the example shown, the transcript levels in a control and diseased cell are compared. If a particular mRNA is present at a higher level due to the disease, the spot on that part of the microarray will appear to have a higher amount of red coloration due to binding of a greater amount of red-labeled cDNA. In contrast, higher levels of an mRNA in the control will appear to be green on the corresponding position on the microarray and those with approximately equal levels in the disease and control will appear as yellow (an equal mixture of red and green)

present the transcriptional changes in adult mouse subventricular zone progenitor cells induced by differentiation [11] using the Agilent high density cDNA microarrays.

## 2 Materials (see Note 1)

1. DNAase- and RNase-free water (*see Note 2*).
2. 20 µg total RNA samples (*see Note 3*).
3. 6 µM oligo<sub>20</sub> dT primers.
4. 20 nM cyanine 3-dCTP (*see Note 4*).
5. 20 nM cyanine 5-dCTP (*see Note 4*).



**Fig. 3** Schematic of Agilent high density microarray slide

6. First strand synthesis master mix: 1× first strand reaction buffer, 10 mM dithiothreitol, 400 U Moloney murine leukemia virus reverse transcriptase, 40 U RiboSafe RNase inhibitor, 40 nM ATP, 40 nM TTP, 40 nM GTP and 20 nM CTP (*see Note 5*).
7. cDNA wash buffer PE (Qiagen; Manchester, UK).
8. Elution buffer: 10 mM Tris–HCl, pH 8.5.
9. QIAquick spin column (Qiagen).
10. Deposition hybridization buffer (*see Note 6*).
11. cDNA microarray on  $7.62 \times 2.54 \times 0.10$  cm glass slides and appropriate slide cover slips (Agilent Technologies; Santa Clara, CA, USA) (Fig. 3) (*see Note 7*).
12. Microarray hybridization chamber (*see Note 8*).
13. Wash solution 1: 0.5× saline sodium citrate (SSC) and 0.01% sodium dodecyl sulphate (SDS).
14. Wash solution 2: 0.06× SSC.

### 3 Methods

1. Combine RNA and oligo dT primer in a sterile nuclease-free tube in 25 mL total volume and incubate at 70 °C for 10 min and then place on ice for 5 min (*see Note 9*).
2. Add cyanine 3-dCTP to the RNA sample from undifferentiated cells and add cyanine 5-dCTP to the RNA sample from differentiated cells (*see Note 10*).
3. Add the master mix to each sample and incubate at 42 °C for 60 min.

4. Heat at 70 °C for 10 min and then place on ice for 5 min (*see Note 11*).
5. Centrifuge briefly to ensure all liquids are at the bottom of the tube and add 20 U RNase 1A, mix by pipetting gently up and down, and incubate at room temperature for 30 min (*see Note 12*).
6. Combine the cyanine 3- and cyanine 5-cDNA reactions for each microarray hybridization experiment into one tube to give a final volume of 100 µL (*see Note 13*).
7. Centrifuge at 10,000×*g* for 1 min using a QIAquick column (*see Note 14*).
8. Discard flow-through and wash the column with 400 µL of wash buffer, centrifuge for 60 s, and repeat (*see Note 15*).
9. Elute the cyanine-labeled cDNA by adding 30 µL of elution buffer, leaving this for 1 min, and centrifuging for 60 s, collecting the eluate into a nuclease-free collection tube.
10. Repeat wash and centrifugation, and combine the two eluates from each sample into a single collection tube.
11. Add 7.5 µL of each cyanine-labeled sample to a fresh tube containing 1× deposition hybridization buffer and nuclease-free water to give a final volume of 25 µL.
12. Mix well, incubate at 98 °C for 2 min, and leave the solution at room temperature until use (*see Note 16*).
13. Place the microarray slide so that the barcode side is facing down and add the hybridization mixture onto the center of each microarray section (Fig. 3) and place a glass cover slip over each one in use (*see Note 17*).
14. Place the slide into the hybridization chamber, which contains nuclease-free water, close the chamber, and incubate at 65 °C overnight or for approximately 16 h (*see Note 18*).
15. The next day, remove the slides from the chamber and remove cover slips by gently dipping in wash solution 1.
16. Place the slide in a rack in a container containing wash solution 1 on a magnetic stirrer and stir for 5 min.
17. Transfer to the same setup in a container containing wash solution 2 and stir for 2 min.
18. Transfer the slides to plastic containers in a swinging bucket rotor and centrifuge at 400×*g* for 2 min (*see Note 19*).
19. Scan the slides using an appropriate imager and analyze the spot volumes in the channels for cyanine 3 and cyanine 5 to quantitate ratios of expression levels of specific transcripts in the two samples under comparison (Table 1) (*see Note 20*).

**Table 1**

**Top differentially (Diff) regulated genes in neural progenitor cells (NPCs) following a 24 h differentiation protocol. The values are given as the levels in Diff/NPC cells**

| Symbol           | Name  | Ratio |
|------------------|---|-------|
| <i>ALDH1A3</i>   | Aldehyde dehydrogenase family 1, subfamily A3                 | 10.44 |
| <i>CHGB</i>      | Chromogranin B  | 9.43  |
| <i>PADI2</i>     | Peptidyl arginine deiminase, type II                          | 7.47  |
| <i>TRP53INPI</i> | Transformation related protein 53 inducible nuclear protein 1 | 7.39  |
| <i>POU3F1</i>    | POU domain, class 3, transcription factor 1                   | 6.97  |
| <i>GFAP</i>      | Glial fibrillary acidic protein                               | 6.83  |
| <i>NDRG2</i>     | N-myc downstream regulated gene 2                             | 6.54  |
| <i>DUSP4</i>     | Dual specificity phosphatase 4                                | 6.35  |
| <i>EDNRB</i>     | Endothelin receptor type B                                    | 6.28  |
| <i>AGT</i>       | Angiotensinogen   | 6.15  |
| <i>IER3</i>      | Immediate-early response 3                                    | 5.89  |
| <i>APLN</i>      | Apelin  | 5.81  |
| <i>SPRY1</i>     | Sprouty homologue 1 (Drosophila)                              | 4.96  |
| <i>SPP1</i>      | Secreted phosphoprotein 1                                     | 4.84  |
| <i>C1QR1</i>     | Complement component 1, subcomponent q1                       | 4.71  |

#### 4 Notes

1. The protocol below is specific for Agilent cDNA microarrays on  $7.62 \times 2.54 \times 0.10$  cm glass slides containing amplified Human UniGene 1, Human Drug Target, and Human Foundation Series clone sets [12, 13]. Each microarray contains approximately 22,500 cDNAs and each spot corresponds to a different clone. For other microarrays the protocol may be slightly different.
2. Care should be taken throughout due to the ease of losing DNA and RNA to unwanted nuclease activity.
3. We used RNA extracted from neuronal precursor cells under proliferation (culture in the presence of 20 ng/mL fibroblast growth factor 2 and 20 ng/mL epidermal growth factor) and differentiating (culture in the absence of growth factors) conditions. The RNA can be extracted using standard protocols.
4. Minimize exposure to light as both the cyanine dyes are photolabile.

5. This step is for use with total RNA templates and conversion to cDNA. It is important to use an RNase inhibitor such as the RiboSafe RNase inhibitor (Bioline) although others can be used.
6. The buffer contains lithium chloride and lithium lauryl sulphate, so the usual laboratory care should be taken.
7. Each slide contains two microarrays so two 2.4×3.0 cm cover slips are required.
8. Many chambers can be used but these must be compatible with slide/cover slip hybridization and watertight.
9. This step denatures the template and primers.
10. It is recommended to carry out reciprocal labeling reactions to account for any preferential labeling of either of the cyanine dyes. In this case one comparison would be Cy3-undifferentiated versus Cy5-differentiated as shown and another could be Cy5-undifferentiated versus Cy3-differentiated.
11. This step inactivates the enzyme.
12. This step degrades the RNA.
13. This allows hybridization of the two different samples to the same microprobes on the cDNA array.
14. In this step, the cyanine-labeled DNA is bound to the column. Here, we use the Qiagen quick spin columns but other columns from other manufacturers would also work well.
15. We use the Qiagen buffer PE but other wash buffers would also work here (these buffers usually contain salts, ethanol, and other reagents).
16. This is to denature the DNA.
17. Be careful not to touch the surface of the microarray as it is easily damaged. The presence of two sections on each slide allows the reciprocal labeling strategy to be easily applied (as described above).
18. The chamber used must be watertight when sealed and the addition of water to the chamber helps to prevent evaporation.
19. This step is for drying the slides and must be performed quickly to avoid leaving a residue on the slides which could interfere with scanning.
20. There are many scanners available on the market for this purpose, although these should be capable of exciting and detecting the hybridized cyanine 3- (550 nm and 570 nm, respectively) and cyanine 5- (650 nm and 670 nm, respectively) labeled cDNAs.

## References

1. Grunstein M, Hogness DS (1975) Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc Natl Acad Sci U S A* 72:3961–3965
2. Craig AG, Nizetic D, Hoheisel JD, Zehetner G, Lehrach H (1990) Ordering of cosmid clones covering the herpes simplex virus type I (HSV-I) genome: a test case for fingerprinting by hybridization. *Nucleic Acids Res* 18:2653–2660
3. Lennon GG, Lehrach H (1991) Hybridization analyses of arrayed cDNA libraries. *Trends Genet* 7:314–317
4. Leung YF, Cavalieri D (2003) Fundamentals of cDNA microarray data analysis. *Trends Genet* 19:649–659
5. Bumgarner R (2013) Overview of DNA microarrays: types, applications, and their future. *Curr Protoc Mol Biol Chapter 22:Unit 22.1.* doi:[10.1002/0471142727.mb2201s101](https://doi.org/10.1002/0471142727.mb2201s101)
6. Goldmann T, Gonzalez JS (2000) DNA-printing: utilization of a standard inkjet printer for the transfer of nucleic acids to solid supports. *J Biochem Biophys Methods* 42: 105–110
7. Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, Pitas A et al (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res* 12:1749–1755
8. Shalon D, Smith SJ, Brown PO (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* 6:639–645
9. Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghazal G et al. (1999) Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nat Genet* 22:164–167
10. de Souto MC, Costa IG, de Araujo DS, Ludermir TB, Schliep A (2008) Clustering cancer gene expression data: a comparative study. *BMC Bioinformatics* 9:497
11. Bonnert TP, Bilsland JG, Guest PC, Heavens R, McLaren D, Dale C (2006) Molecular characterization of adult mouse subventricular zone progenitor cells during the onset of differentiation. *Eur J Neurosci* 24:661–675
12. <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL10>
13. <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL875>

# Chapter 10

## Multiplex Single Nucleotide Polymorphism Analyses

Steve F.C. Hawkins and Paul C. Guest

### Abstract

Quantitative polymerase chain reaction (qPCR) is a routinely used method for detection and quantitation of gene expression in real time. This is achieved through the incorporation and measurement of fluorescent reporter probes in the amplified cDNA strands, since the fluorescent signals increase as the reaction progresses. The availability of multiple probes that fluoresce at different wavelengths allows for multiplexing as this gives rise to amplicons with unique fluorescent signatures. Here, we describe a method using the SensiFAST and SensiFAST One-Step probe kits which allows simultaneous real-time quantitation of up to 5 amplicons.

**Key words** Single nucleotide polymorphism, Allele, PCR, Fluorescent dyes, Taq polymerase, Quantitation, Multiplex analysis

---

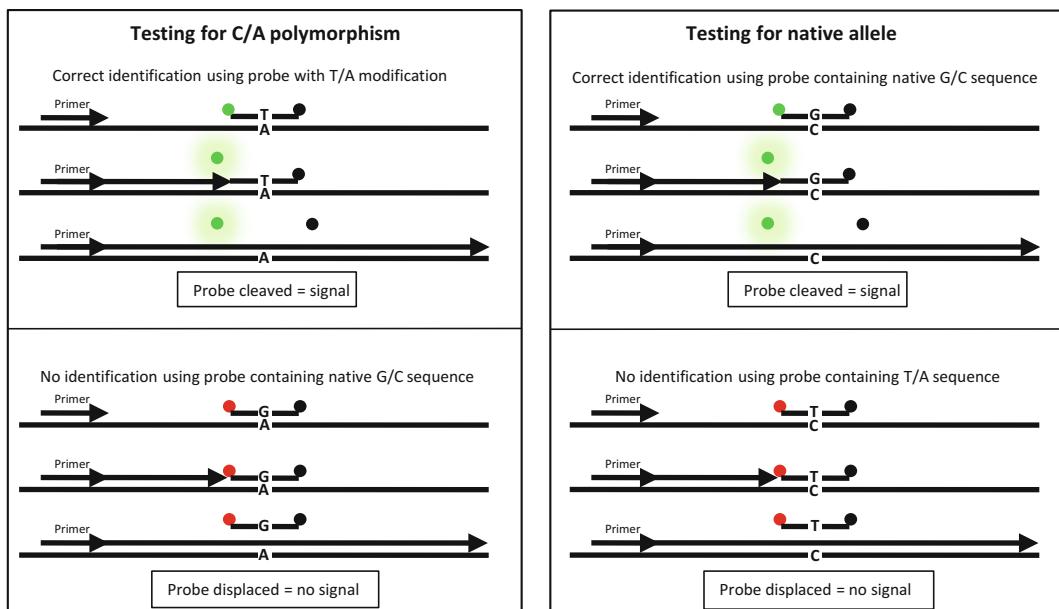
### 1 Introduction

A single nucleotide polymorphism (SNP) is a variation in a single nucleotide at a specific position in the genome [1]. As an example, the base cytosine (C) may be present at a specific position in the human genome although this is replaced by an adenine (A) base in a small percentage of individuals. Many SNPs are benign although some may underlie differences across individuals in their susceptibility to specific diseases [2], such as Alzheimer disease [3], sickle-cell anaemia [4], certain metabolic disorders [5], various types of cancer [6], and heart disease [7]. For example, a C/T substitution at amino acid position 130 of the *APOE* gene is associated with a higher risk for Alzheimer's disease [8].

Several platforms are now in use for detection of SNPs, the most efficient way to link a SNP with phenotypes is the so-called genome-wide association (GWA) study, in which hundreds of thousands or even millions of polymorphisms are scanned per sample using DNA microarrays. For determination of new SNPs, next generation sequencing (NGS) is used; however for high-throughput screening of individual SNPs (often linked to diseases) qPCR is the

method of choice. The TaqMan® SNP genotyping approach uses the 5' nuclease activity of Taq polymerase to produce a fluorescent signal during the polymerase chain reaction (PCR) stage of the analysis [9]. The assay normally uses two probes that are identical except for the sequence at the site of the allele in question. One probe is complementary to the native sequence whereas the other targets the mutant allele. Each probe also contains a distinct 5' fluorescent reporter dye and a 3' quencher dye. As with standard qPCR (see Chapter 11), fluorescence of the reporter dye is suppressed in intact probes due to the proximity of the quencher. In the annealing stage of PCR, the probes hybridize to the site in question. During the extension stage, the reporter and quencher are released by the 5' nuclease activity of Taq polymerase, resulting in fluorescence of the reporter dye. However, cleavage of the probe occurs only in the case of perfectly matched probes as only these will be recognized by the polymerase (Fig. 1). At the end of the PCR experiment, the fluorescent signal for the two reporter dyes is measured. The ratio of the signals will be indicative for the genotype of the sample. For increased accuracy, the DNA sample should be analyzed with two probes. One of these should contain the native sequence and the other should contain the mutant allele.

Here, we describe the use of the SensiFAST Genotyping Kit for high throughput of SNPs (see Note 1). This assay is used to detect a C/T transition substitution on gene *SLCO2B1* (a solute carrier organic anion transporter family member 2B1). The sequence around the substitution is given below:



**Fig. 1** Schematic diagram showing how the levels of a mutant (**C/A**) and native (**T/A**) allele can be determined by PCR analysis of the same DNA sample using two probes specific for each of these sequences

A G G A T G C C A G G G T A G T T A A C C C G G [C / T]  
GAGGCTGAAGTCTAAATAACTGGAA

---

## 2 Materials (See Note 1)

1. 0.9 µM forward primer (ThermoFisher Scientific; Loughborough, Leicestershire, UK; assay: M\_25345745\_10; catalogue number: 4351384) (*see Note 2*).
2. 0.9 µM reverse primer (ThermoFisher Scientific; assay: catalogue number: 4351384).
3. 0.2 µM allele 1 probe (ThermoFisher Scientific; catalogue number: 4351384) (*see Note 3*).
4. 0.2 µM allele 2 probe (ThermoFisher Scientific; catalogue number: 4351384).
5. 20 ng cDNA template (*see Note 4*).
6. 1× SensiFAST Genotyping Mix, containing hot-start DNA polymerase, dNTPs, stabilizers, and enhancers.
7. qPCR thermocycler (*see Note 5*).

---

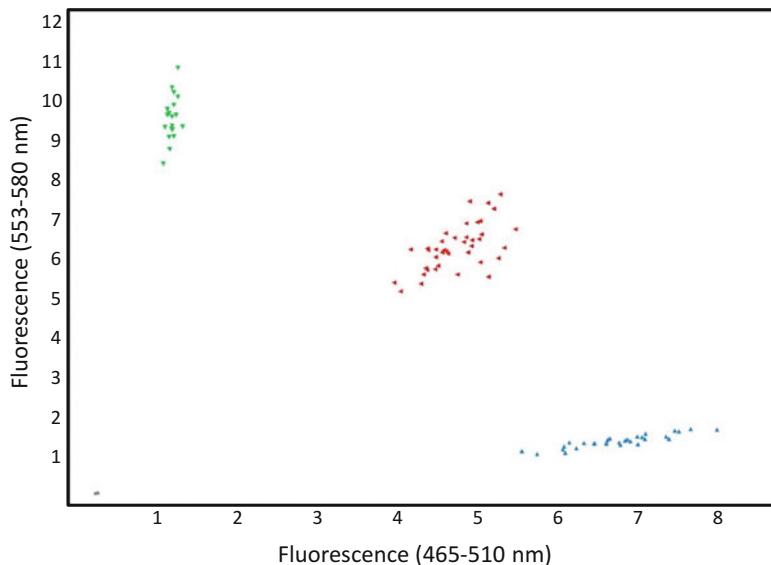
## 3 Methods (See Note 6)

1. Prepare a PCR master mix based on a standard 20 µL final reaction volume containing the primers, probes, template, and Genotyping mix (*see Notes 4 and 7*).
2. Suggested thermal cycling conditions for DNA: 1 cycle at 95 °C for 2–5 min for polymerase activation, then 40 cycles at 95 °C for 10 s for denaturation and 60 °C for 20–50 s for annealing/extension (*see Notes 8 and 9*).
3. Data analysis (Fig. 2) (*see Note 10*).

---

## 4 Notes

1. The SensiFAST Genotyping Kit is compatible with many dual-label probe assays. The overall yield and efficiency relates to the sequence and concentration of all reagents including template, primers, and probes, as well as the targeted amplicon length. Please refer to the appropriate literature during the design and set-up stages. The following information relates to the design and setup of TaqMan probe-based qPCR.
2. The length, sequence, and concentration of the primers and probes are critical for specific amplification. Here, we have used primers and probes from the indicated ThermoFisher kit,



**Fig. 2** DNA samples from 96 individuals were genotyped using a custom TaqMan drug metabolism genotyping assay. The Sensifast Genotyping Kit was used on the Roche LightCycler® 480 (autocaller confidence level 95 %). Cycling conditions were: 95 °C for 3 min, then 35 cycles at 95 °C for 10 s and 60 °C for 30 s. The results show that the Sensifast Genotyping Kit was able to generate distinct clusters across the 96 individuals

targeting the specific mutation site within the *SLCO2B1* gene. For user-defined targets that are not commercially available, we recommend the use of primer-design software, such as Primer3 (<http://frodo.wi.mit.edu/primer3/>) or visual OMP™ (<http://dnasoftware.com/>). For the best results primers and probes should have melting temperatures of approximately 60 °C and the targeted amplicon length should be 80–200 base pairs. For the *SLCO2B1* gene, the amplicon length is approximately 150 base pairs (SNP ID: rs4226825, chromosome 7: 99658318 on NCBI Build 37, polymorphism: C/T transition substitution). Forward and reverse primers should be equimolar and titration in the range of 0.2–1 µM should be used to find the ideal concentration.

3. For user-defined targets that do not have commercially available reagents, we recommend the use of probe design software, as indicated above. A final concentration of 0.2 µM for each probe is sufficient for most reactions although these should be at least twofold lower than the primer concentrations. As for the primers, the concentration of both probes should be equimolar and determined by titration as above. Each probe should contain dyes that are similar in intensity of fluorescence but with a distinct spectral separation. Here, we have used probes from the indicated ThermoFisher kit, target-

ing the specific mutation site within the SLCO2B1 gene. The probes are labeled with VIC® (528–546 nm) and 6-carboxyfluorescein (FAM; 495–517 nm), which are spectrally distinct and can be quenched with tetramethylrhodamine (TAMRA).

4. The SensiFAST Genotyping Kit can be used with sample lysates or purified genomic DNA. It is important that the DNA template does not contain any inhibitors, such as EDTA.
5. Many instruments can be used here but it is important to check with the manufacturer for compatibility, as described above. In this study, we used the Roche LightCycler® 480.
6. Due to the sensitivity of qPCR, clean room conditions should be used to avoid any contamination of the samples.
7. As reagents can settle or partition during storage, always mix well when setting up reactions. However, do not use sonication methods for this due to the likely outcome of shearing the DNA template.
8. Positive and controls of known genotype should be included to ensure distinct genotype calling by the instrument. A control that does not contain any template should also be used to detect any contamination.
9. Cycling conditions can be varied to suit machine-specific protocols. Initially, running 30 cycles with a 30 s extension is recommended and then adding cycles in increments of five, as required. Longer extension times may be required for amplicons larger than 200 bp. For low concentrations of template (<1 ng), up to 45 cycles may be necessary. However, it is not recommended to exceed a total of 45 cycles for optimal calling. Here are the conditions that we used for this experiment: 95 °C for 3 min, then 35 cycles at 95 °C for 10 s and 60 °C for 30 s.
10. Most qPCR machines will create scatter plots and use algorithms to assign genotypes based on reporter probe signals at the end of amplification steps. Failure of this autocalling is usually due to problems associated with outliers that skew the clusters. Removing the outliers and reanalyzing the data should allow the program to adjust the scaling so that distinct clusters can be seen.

## References

1. Ligtenberg MJ, Gennissen AM, Vos HL, Hilkens J (1991) A single nucleotide polymorphism in an exon dictates allele dependent differential splicing of episitin mRNA. *Nucleic Acids Res* 19:297–301
2. Bacolod MD, Schemmann GS, Giardina SF, Paty P, Notterman DA, Barany F (2009) Emerging paradigms in cancer genetics: some important findings from high-density single nucleotide polymorphism array studies. *Cancer Res* 69:723–727

3. Li L, Yin Z, Liu J, Li G, Wang Y, Yan J, Zhou H (2013) CYP46A1 T/C polymorphism associated with the APOE ε4 allele increases the risk of Alzheimer's disease. *J Neurol* 260:1701–1708
4. Vicari P, Adegoke SA, Mazzotti DR, Cançado RD, Nogutti MA, Figueiredo MS (2014) Interleukin-1 $\beta$  and interleukin-6 gene polymorphisms are associated with manifestations of sickle cell anemia. *Blood Cells Mol Dis* 54:244–249
5. Ghalandari H, Hosseini-Esfahani F, Mirmiran P (2015) The association of polymorphisms in leptin/leptin receptor genes and ghrelin/ghrelin receptor genes with overweight/obesity and the related metabolic disturbances: a review. *Int J Endocrinol Metab* 13:e19073
6. Chen Y, Zhong H, Gao JG, Tang JE, Wang R (2015) A systematic review and meta-analysis of three gene variants association with risk of prostate cancer: an update. *Urol J* 12: 2138–2147
7. Zhang MM, Xie X, Ma YT, Zheng YY, Yang YN, Li XM, Fu ZY, Liu F, Chen BD (2015) Association of COX-2 -765G>C genetic polymorphism with coronary artery disease: a meta-analysis. *Int J Clin Exp Med* 8:7412–7418
8. Namboori PK, Vineeth KV, Rohith V, Hassan I, Sekhar L, Sekhar A, Nidheesh M (2011) The ApoE gene of Alzheimer's disease (AD). *Funct Integr Genomics* 11:519–522
9. Glaab WE, Skopek TR (1999) A novel assay for allelic discrimination that combines the fluorogenic 5' nuclease polymerase chain reaction (TaqMan) and mismatch amplification mutation assay. *Mutat Res* 430:1–12

# Chapter 11

## Pulsed SILAC as a Approach for miRNA Targets Identification in Cell Culture

Daniella E. Duque-Guimarães, Juliana de Almeida-Faria,  
Thomas Prates Ong, and Susan E. Ozanne

### Abstract

Pulsed stable isotope labeling by amino acids in cell culture (pSILAC) comprises a variation of the classical SILAC proteomic methodology that enables the identification of short-term proteomic responses such as those elicited by micro RNAs (miRNAs). Here, we describe a detailed pSILAC protocol for global identification and quantification of protein translation alterations induced by a miRNA using 3T3-L1 pre-adipocytes as a model system.

**Key words** Pulsed stable isotope labeling by amino acids, Cell culture, MicroRNAs, Targets, Proteomics

---

### 1 Introduction

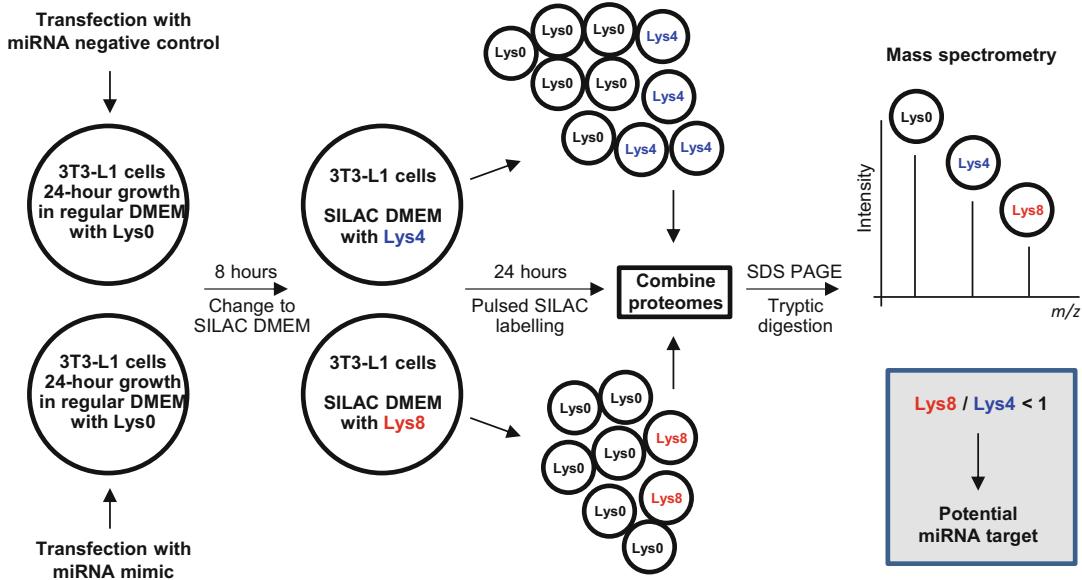
MicroRNAs (miRNAs) have been shown to play a key role in the control of gene expression, modulating the expression of many proteins in a wide range of biological processes [1]. These small noncoding RNA molecules can individually regulate multiple target genes by binding to target sites found within the 3' untranslated region of the targeted mRNA, resulting in posttranscriptional gene silencing of gene networks. The posttranscriptional effect results from degradation of the targeted mRNA or repression of its translation [2, 3]. Most of the targets contain perfect complementary sites in their 3'-untranslated regions to the seed sequence of the miRNA. This conserved seed sequence of typically six to eight nucleotides is often used as the main feature for miRNA target site prediction [4]. Several computational algorithms such as TargetScan [5] and Diana-MicroT [6] have been developed to predict targets based on a combination of different parameters including these sequences. However, additional factors can impede these bioinformatic approaches, resulting in false positives and thus

limiting their usefulness [7]. Experimental methods are therefore required for robust target predication. Approaches used to date include in vitro luciferase reporter assays for individual miRNAs and targets or high-throughput experimental methods such as next-generation sequencing following immunoprecipitation [8]. However, these have their limitations and the methodology is still challenging. There is no clear consensus that a functional miRNA target will be always identified.

Stable Isotope Labeling by Amino acids in Cell culture (SILAC) is a proteomic approach that has been used to identify miRNA targets and investigate the consequences on protein levels [9, 10]. SILAC is a quantitative method based on whole proteome metabolic labeling with stable isotope-labeled amino acids in cell culture [11]. This method is based on the introduction of a mass difference between two proteomes, which creates two versions of every peptide (e.g.,  $^{12}\text{C}$  versus  $^{13}\text{C}$ ) that can be distinguished in mass spectrometry (MS)-based analysis [11, 12]. The classical version of SILAC is based on growing cells in media with natural “light” or “heavy” amino acids for several days or for approximately five cell doublings. This allows virtually complete incorporation of the heavy amino acids into cellular proteomes. SILAC-based approaches have also been used to identify miRNA targets in many cell systems including HeLa cells [9], as well as in SW480 colon cancer [13] and MiaPaCa-2 pancreatic cancer [14] cells.

In recent years, an increasing number of studies have adopted a variation of the classical SILAC method termed pulsed SILAC (pSILAC) to investigate miRNA proteomic effects [10, 15, 16]. The pSILAC method is based on growing two populations of cells (e.g., miRNA-treated and control cells) that are pulse-labeled with two different heavy SILAC amino acids [15]. Both populations of cells are grown initially in standard media containing normal light amino acids [16]. After miRNA transfection, the culture medium from control and miRNA-transfected cells is replaced with SILAC medium containing either medium-heavy or heavy amino acids, respectively. After the pulse labeling step, all newly synthesized proteins will incorporate medium-heavy or heavy amino acids, while proteins that were synthesized before the labeling step will contain only light amino acids and are not considered for protein quantification [15, 16].

In this chapter, we describe a detailed pSILAC protocol for global identification and quantification of protein translational alterations induced by miRNA using 3T3-L1 pre-adipocytes as a model-system (Fig. 1). Cells are grown in two cell culture dishes for 24 h and then transfected with either the miRNA mimic or a transfection control. After 8 h, the normal growth media from both cell populations is replaced with media containing either heavy lysine (Lys8) or medium-heavy lysine (Lys4) for the miRNA mimic transfected and miRNA control cells, respectively. After



**Fig. 1** pSILAC experimental design. Cells are cultivated in regular DMEM media containing light lysine (Lys 0) in two cell culture dishes for 24 h and then transfected with the miRNA mimic of interest or the negative control. After 8 h, normal growth medium from both cell populations is replaced with SILAC media containing either heavy lysine [ $^{13}\text{C}_6\text{ }^{15}\text{N}_2$ -l-lysine (Lys8); red font] and medium-heavy lysine [ $^2\text{H}_4$ -l-lysine (Lys4); blue font] for the miRNA mimic cells and miRNA negative control cells, respectively. After 24 h, protein extracts from both cell populations are combined and subjected to SDS-PAGE separation, digested with trypsin and submitted to high performance liquid chromatography tandem mass spectrometry (HPLC/MS/MS). The ratio of heavy and medium-heavy peptides reflects differences in the translation of the corresponding protein induced by the miRNA mimic and the light peptides are ignored. A ratio of heavy (Lys8)/medium-heavy (lys4) peptides smaller than 1 indicates a potential miRNA target

24 h, protein extracts from both cell populations are combined and subjected to sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) to separate the proteins. The protein bands are then excised from the gels, enzymatically digested with trypsin and submitted to high performance liquid chromatography tandem mass spectrometry (HPLC/MS/MS) analysis. The ratio of heavy and medium-heavy peptides reflects differences in the translation of the corresponding protein induced by the miRNA, while the light peptides are ignored.

## 2 Materials (See Note 1)

### 2.1 pSILAC Reagents

1. L-arginine-HCL (Arg0) in phosphate buffered saline (PBS) (*see Note 2*).
2. 4, 4, 5, 5-D4 L-Lysine-2HCl (Lys4) in PBS (*see Note 2*).
3.  $^{13}\text{C}_6\text{ }^{15}\text{N}_2$  l-Lysine-2HCl (Lys8) in PBS (*see Note 2*).

4. Dulbecco's Modified Eagle Medium (DMEM), containing 10% heat-inactivated calf serum, antibiotic-free (*see Note 3*).
5. SILAC DMEM media, lacking lysine and arginine.
6. Dialyzed fetal bovine serum (FBS) (*see Note 4*).
7. Medium-heavy Lys4 SILAC DMEM: 0.08 mg/mL L-arginine, 0.15 mg/mL Lys4, and 10% FBS in 500 mL SILAC DMEM (0.22 µm filter-sterilized).
8. Heavy Lys8 SILAC DMEM: 0.08 mg/mL L-arginine, 0.15 mg/mL Lys8, and 10% FBS in 500 mL SILAC DMEM (0.22 µm filter-sterilized).

## 2.2 Cell Culture and miRNA Mimic Transfection

1. Undifferentiated 3T3-L1 pre-adipocytes (*see Note 5*).
2. Sterile 10 cm diameter cell culture dishes, 15 mL-capacity Falcon tubes, and 1.5 mL-capacity microcentrifuge tubes.
3. Tissue culture trypsin solution for dissociation of cells.
4. Opti-MEM® medium and Lipofectamine®RNAiMAX reagent (Invitrogen).
5. 10 µM mirVana™ miRNA126 mimic and 10 µM mirVana™ miRNA negative control #1 in nuclease-free water (ThermoFisher Scientific) stored at or below 20 °C (*see Note 6*).
6. BLOCK-iT™ Alexa Fluor® Red Fluorescent Oligo (Invitrogen) (*see Note 7*).
7. Epifluorescence inverted microscope.
8. RIPA buffer: 25 mM Tris-HCl (pH 7.6), 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS.
9. 100× Halt™ protease and phosphatase single-use inhibitor cocktail (ThermoFisher Scientific).
10. Bicinchoninic acid (BCA) reagent, 0.125–1.0 mg/mL bovine serum albumin protein standards.
11. TRI® reagent (Sigma-Aldrich).
12. Direct-zol™ RNA MiniPrep (Zymo Research).
13. TaqMan® microRNA reverse transcription kit (ThermoFischer Scientific).

## 2.3 SDS-PAGE

1. 4× Novex Nu PAGE LDS sample buffer (Thermo Fisher Scientific).
2. Proteomics grade SDS-PAGE gel solutions or precast gels (*see Note 8*).
3. Stain solution: 1 mg/mL CoomassieBlue R250, 40% methanol, 10% acetic acid (filter sterilized).
4. Destain solution: 20% methanol, 10% acetic acid (filter sterilized).

## 2.4 In-Gel Trypsin Digestion

1. 100 mM ammonium bicarbonate.
2. 100 mM ammonium bicarbonate, 50% acetonitrile.
3. 20 mM ammonium bicarbonate.
4. Reducing solution: 100 mM ammonium bicarbonate, 10 mM dithiothreitol (*see Note 9*).
5. Alkylation solution: 100 mM ammonium bicarbonate, 20 mM iodoacetamide (*see Note 9*).
6. Trypsin digestion solution: 12.5 µg/mL trypsin (mass spectrometry grade), 20 mM ammonium bicarbonate (*see Note 10*).
7. 1% formic acid.

## 2.5 HPLC, MS Analysis, and Data Processing

1. MS solvent: 0.1% trifluoroacetic acid, 3% acetonitrile.
2. Solvent A: 0.1% formic acid.
3. Solvent B: 100% acetonitrile.
4. NanoAcuity uPLC with a 100 Å, 5 µm, 180 µm × 20 mm C18 Symmetry trap and 130 Å, 1.7 µm, 75 µm × 250 mm C18 BEH analytical columns (Waters Corporation).
5. Accurate mass/high resolution MS instrument such as the LTQ-OrbiTrap XL (ThermoFisher Scientific).
6. Maxquant software (version 1.5.3.30 or higher; [www.maxquant.org](http://www.maxquant.org)).

## 2.6 Analysis of pSILAC Data

1. Several software options: Microsoft Office Excel (<https://www.office.com/>), R (<https://www.r-project.org>) in conjunction with Bioconductor (<http://www.bioconductor.org>), Spotfire (<http://spotfire.tibco.com/>), Matlab (<http://uk.mathworks.com/products/matlab/>), and Graphpad Prism (<http://www.graphpad.com/scientific-software/prism/>).

## 3 Methods

### 3.1 Cell Culture and miRNA Mimic Transfection

1. Seed  $0.4 \times 10^6$  3T3-L1 cells in two dishes in 10 mL antibiotic-free DMEM and culture for 24 h at 37 °C under 5% CO<sub>2</sub> (*see Note 11*).
2. Add 50 µL of miRNA mimic stock solution to 500 µL Opti-MEM medium in a sterile 1.5 mL tube.
3. Add 43 µL of Lipofectamine RNAiMAX to 500 µL Opti-MEM medium in a sterile 1.5 mL tube.
4. Add diluted miRNA mimic solution to diluted Lipofectamine RNAiMAX solution, gently mix and incubate for 5 min at room temperature.
5. Add the miRNA mimic-lipid complex to 60–80% confluent cells, gently mix, and incubate for 8 h at 37 °C under 5% CO<sub>2</sub>.

6. Add 50  $\mu$ L of miRNA control stock solution to 500  $\mu$ L Opti-MEM medium in a sterile 1.5 mL tube.
7. Add 43  $\mu$ L of Lipofectamine RNAiMAX to 500  $\mu$ L Opti-MEM medium in a sterile 1.5 mL tube.
8. Add diluted miRNA control solution to diluted Lipofectamine RNAiMAX solution, gently mix, and incubate for 5 min at room temperature.
9. Add the miRNA control-lipid complex to 60–80% confluent cells, gently mix, and incubate for 8 h at 37 °C under 5% CO<sub>2</sub>.
10. Remove the regular medium from the dish containing miRNA mimic-transfected cells, replace with 10 mL pre-warmed (37 °C) heavy Lys8 SILAC medium, and incubate for 24 h at 37 °C under 5% CO<sub>2</sub>.
11. Remove the regular medium from the dish containing miRNA control-transfected cells, replace with 10 mL pre-warmed (37 °C) medium-heavy Lys4 SILAC medium, and incubate for 24 h at 37 °C under 5% CO<sub>2</sub> (*see Notes 12 and 13*).

### **3.2 Protein and RNA Extraction**

1. After 24 h, add trypsin solution to remove cells from the plates and wash by centrifugation in ice-cold PBS at 750  $\times g$ , for 5 min at 4 °C, in a 15 mL Falcon tube.
2. Suspend the pellets gently in PBS and repeat the centrifugation wash step.
3. Resuspend cells in 10 mL ice-cold PBS, transfer 5 mL homogeneous aliquots to two different Falcon tubes designated for protein or RNA extraction, and pellet the cells by centrifugation at 750  $\times g$  for 5 min at 4 °C.
4. For protein extraction, add 250  $\mu$ L cold RIPA buffer containing protease and phosphatase inhibitors to one aliquot of cells, swirl on ice for 5 min, centrifuge at 13,000  $\times g$  to remove the cell debris, and transfer the supernatants to fresh tubes (*see Note 14*).
5. Mix 50  $\mu$ g proteins extracted from cells transfected with miRNA mimic with 50  $\mu$ g of proteins extracted from cells transfected with control.
6. For total RNA extraction, add 500  $\mu$ L TRI reagent to the cell aliquot and extract mRNAs and small RNAs using the Direct-zol RNA kit according to the manufacturer's instructions.
7. Prepare cDNA using the reverse transcription kit and carry out polymerase chain reactions following the manufacturer's protocol (*see Note 15*).

### **3.3 SDS-PAGE and Treatment of Gel Pieces**

1. Add the necessary volume of 4 $\times$  sample buffer to the miRNA mimic/control mixture.
2. Electrophorese the samples.

3. Stain the gel for approximate 2 h, incubate in destain solution overnight on a rocking table, and take a photograph of the protein bands.
4. Slice the gel lanes into the desired number of horizontal pieces and then chop these further to produce 1 mm<sup>3</sup> (approximate) pieces using a new scalpel blade for each slice.
5. The gel pieces from each slice can now be used immediately for the next step or stored at –80 °C.
6. Wash the gel pieces from each slice with 1 mL water for 15 min, centrifuge at 750 × g for 5 min, and discard the supernatant.
7. Add 300 µL of acetonitrile to the gel pieces, wash for 15 min, centrifuge as above, and discard the supernatant.
8. Wash the pieces with 300 µL of 100 mM ammonium bicarbonate for 15 min, centrifuge as above, and discard the supernatant.
9. Wash the pieces with 300 µL of 100 mM ammonium bicarbonate, 50% acetonitrile for 15 min, centrifuge as above, and discard the supernatant (*see Note 16*).
10. If the gel pieces are still blue, repeat steps 8 and 9.
11. Wash the pieces as above with 100 µL of acetonitrile for 5 min (which will cause them to shrink and turn white), centrifuge as above, and discard the supernatant.
12. Dry the pieces in a sterile hood for 10–15 min.
13. Add 50 µL of reducing solution and incubate for 1 h at 56 °C, centrifuge as above, and discard the supernatant.
14. Add 50 µL alkylation solution, incubate for 30 min at room temperature, centrifuge as above, and discard the supernatant.
15. Wash the pieces twice with 300 µL of 100 mM ammonium bicarbonate for 15 min, centrifuge as above, and discard the supernatant.
16. Wash the pieces with 300 µL of 100 mM ammonium bicarbonate, 50% acetonitrile for 15 min, centrifuge as above, and discard the supernatant.
17. Add 100 µL acetonitrile, incubate for 5 min, centrifuge as above, and discard the supernatant.
18. Dry the gel pieces in a sterile hood for 10–15 min.

### **3.4 In-Gel Trypsin Digestion**

1. Add 30 µL digestion solution to the gel pieces from each slice and let stand for 30 min at room temperature (*see Note 17*).
2. Add a small volume of 20 mM ammonium bicarbonate (without trypsin) so that this solution just covers the gel pieces and incubate at 30 °C overnight (at least 16 h) (*see Note 18*).

3. Add an equal volume of acetonitrile and incubate at 30 °C for 30 min.
4. Centrifuge as above and transfer the supernatant containing peptide digests to a fresh tube.
5. Add 50 µL 1% formic acid to the gel pieces and incubate at room temperature for 20 min.
6. Centrifuge as above and add the supernatant containing peptide digests to the same tube as above.
7. Repeat steps 3–6.
8. Add 50 µL of acetonitrile to the gel pieces and incubate for 10 min (they should shrink and turn white again).
9. Centrifuge as above and add the supernatant to the same tube as above giving a final volume of 200–400 µL containing the peptide digest corresponding to each gel slice.
10. Dry samples in a vacuum centrifuge at 60 °C (*see Note 19*).
11. Suspend the peptide pellets in 25–50 µL of 1% formic acid, agitate gently for 30 min at room temperature, and store the samples at –80 °C if required.

### **3.5 HPLC, MS Analysis, and Data Processing**

1. Thaw the samples and centrifuge at 13,000 ×*g* for 10 min.
2. Take aliquots from the supernatant gently and transfer to an MS tube.
3. Dry the tryptic peptides almost to completion in a centrifugal vacuum concentrator and suspend in 10 µL MS solvent.
4. Analyze peptides using solvents A and B over the desired linear gradient using liquid chromatography tandem mass spectrometry.
5. Raw MS files can be processed in parallel using MaxQuant and data can be searched against the International Protein Index mouse database using MASCOT Daemon and/or Andromeda softwares (*see Note 20*).

---

## **4 Notes**

1. Make up all solutions and dilutions with ultra-pure water unless another reagent is required. It is important to work clean and sterile and all solutions must be made fresh before use.
2. These are the stock solutions and can be stored at –20 °C. Different companies (ThermoFisher, Cambridge Isotope Laboratories, Life Technologies, among others) provide SILAC kits, usually containing only one heavy amino acid (e.g., Lys6). Although they are suited for classical SILAC experiments, they normally lack a second heavy

amino acid that is needed for pSILAC experiments. Thus, in our laboratory we acquire the Lys4 (medium-heavy) and Lys8 (heavy) reagents for the pSILAC experiments separately. This introduces a 4 Da mass shift in the heavy Lys-labeled peptide as compared to the medium-heavy Lys-labeled peptide, which allows identification and quantification of peptides in MS analysis.

3. Because antibiotics can interfere with the miRNA mimic transfection based on the liposome reagent, these should not be added to the regular DMEM media in which cells are grown initially.
4. Dialyzed fetal bovine serum should be used in pSILAC experiments to avoid incorporation of unlabeled Lys0 that is normally present in this medium. Importantly, some cell lines may display altered growth patterns in SILAC media because dialyzed serum lacks growth factors. Therefore, when optimizing pSILAC experiments, it is recommended to verify if the cells under study are affected by the pSILAC media conditions.
5. The cell line should be chosen according to their suitability for the miRNA(s) of interest. To overexpress a specific miRNA, it is preferable to choose a cell line that has low endogenous expression of the same miRNA. In contrast, to inhibit a miRNA of interest, choose a cell line with high expression of that miRNA.
6. It is important to optimize the correct dose and incubation time for the mimic and/or inhibitor chosen.
7. BLOCK-iT™ Alexa Fluor® Red Fluorescent Oligo is a double stranded RNA oligonucleotide that can provide a good indication of transfection efficiency (as indicated by a strong fluorescence) [15].
8. Precast gels are the best option but it is possible to use home-made gels if desired. In this case, prepare “clean” stocks of all gel solutions and store as recommended.
9. Approximately 200 µL will be needed per gel slice. Prepare this solution fresh on the day of use.
10. Prepare 40 µL of this solution for each gel slice fresh on the day of use.
11. One dish is for miRNA mimic transfection and the other is for miRNA negative control transfection. Transfection should be performed when cells reach 60–80% confluence after a 24 h growth period. Thus, the starting amount of cells should be defined for each cell line in use. For undifferentiated 3T3-L1 pre-adipocytes, we use  $0.4 \times 10^6$  cells in a 10 mL dish and this yields approximately 70% confluence at the time of transfection.

12. In our laboratory we run at least four independent biological replicates.
13. For each biological replicate, a transfection efficiency control with BLOCK-iTAlexa Fluor Red Fluorescent Oligo can be run in parallel. Efficiency can be visually monitored using an epifluorescence microscope 8 h after transfection and 24 h after pulse labeling with SILAC media.
14. For each biological sample, it is important to save at least 60 µg of protein for validation studies (e.g., by Western blot analysis). Based on the final pSILAC analysis, it is possible to select the proteins showing the most significant changes in abundance and validate these using commercially available antibodies.
15. Performing reverse transcription polymerase chain reaction analysis of the miRNA of the interest is a good way to check success of the transfection. It is also important to save RNA for further mRNA expression studies of dysregulated proteins since targets are usually regulated at both the mRNA and protein levels by miRNAs. In this case, use High-Capacity cDNA reverse transcription kit.
16. The gel pieces should shrink during this stage.
17. In this step, the gel pieces should be restored to their original sizes.
18. Check after 30 min to determine if there is enough solution covering the gel pieces and, if not, add more. Make a note of how much buffer was added in total to each set of gel pieces.
19. This could take 2–3 h.
20. Based on the final list of dysregulated proteins, it is possible to select potential direct targets. Check if the 3' un-translated region of the gene matches the miRNA seed sequence for the miRNA of interest. While Western blot analyses can be performed for general target validation, luciferase assays can be used to validate direct targets.

---

## Acknowledgments

DDG was the recipient of a FAPESP Post-Doctoral Fellowship (BEPE-PD 2014/20380-5) and is funded by the BHF (PG/14/20/30769. JdAF was the recipient of a FAPESP Doctoral Fellowship (BEPE-DR 2014/17012-4), TPO was a recipient of a Visiting Scientist CAPES Science Without Borders Fellowship (BEX 11766-13-1). SEO is funded by the UK Medical Research Council (MC\_UU\_12012/4).

## References

1. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281–297
2. Flynt AS, Lai EC (2008) Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat Rev Genet* 9:831–842
3. Pasquinelli AE (2012) MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet* 13:271–282
4. Wang X (2014) Composition of seed sequence is a major determinant of microRNA targeting patterns. *Bioinformatics* 30:1377–1383
5. <http://www.targetscan.org>
6. <http://www.microrna.gr/webServer>
7. Doran J, Strauss WM (2007) Bio-informatic trends for the determination of miRNA-target interactions in mammals. *DNA Cell Biol* 26:353–360
8. Thomson DW, Bracken CP, Goodall GJ (2011) Experimental strategies for microRNA target identification. *Nucleic Acids Res* 39:6845–6853
9. Vinther J, Hedegaard MM, Gardner PP, Andersen JS, Arctander P (2006) Identification of miRNA targets with stable isotope labeling by amino acids in cell culture. *Nucleic Acids Res* 34:e107
10. Kaller M, Oeljeklaus S, Warscheid B, Hermeking H (2014) Identification of microRNA targets by pulsed SILAC. *Methods Mol Biol* 1188:327–349
11. Ong SE, Mann M (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc* 1:2650–2660
12. Ong SE, Mann M (2007) Stable isotope labeling by amino acids in cell culture for quantitative proteomics. *Methods Mol Biol* 359:37–52
13. Bauer KM, Hummon AB (2012) Effects of the miR-143/-145 microRNA cluster on the colon cancer proteome and transcriptome. *J Proteome Res* 11:4744–4754
14. Huang TC, Renuse S, Pinto S, Kumar P, Yang Y, Chaerkady R et al (2015) Identification of miR-145 targets through an integrated omics analysis. *Mol Biosyst* 11:197–207
15. Selbach M, Schwahnässer B, Thierfelder N, Fang Z, Khanin R, Rajewsky N (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature* 455:58–63
16. Schwahnässer B, Gossen M, Dittmar G, Selbach M (2009) Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* 9:205–2099

# **Chapter 12**

## **Blood Bio-Sampling Procedures for Multiplex Biomarkers Studies**

**Paul C. Guest and Hassan Rahmoune**

### **Abstract**

A major challenge in single or panel of biomarker discovery and validation is the inherent biological complexity underlying disease heterogeneity and inconsistent responses to treatment. Moreover, the lack of standardization in the sampling, processing, and storage of biological fluids such as plasma and serum disrupts the discovery and validation of blood-based biomarker tests in preclinical and clinical settings. This chapter presents a reproducible sample collection and handling procedure that aims to enhance analyte stability and ensure compatibility with the corresponding multiplex biomarker profiling platforms. The importance of defining bio-sample acquisition and processing, study design, and profiling platform guidelines for blood-based biomarker measurements is paramount for the success of personalized health-care strategy and development of companion diagnostics.

**Key words** Bio-sampling, Bio-processing, Serum, Plasma, Biomarkers

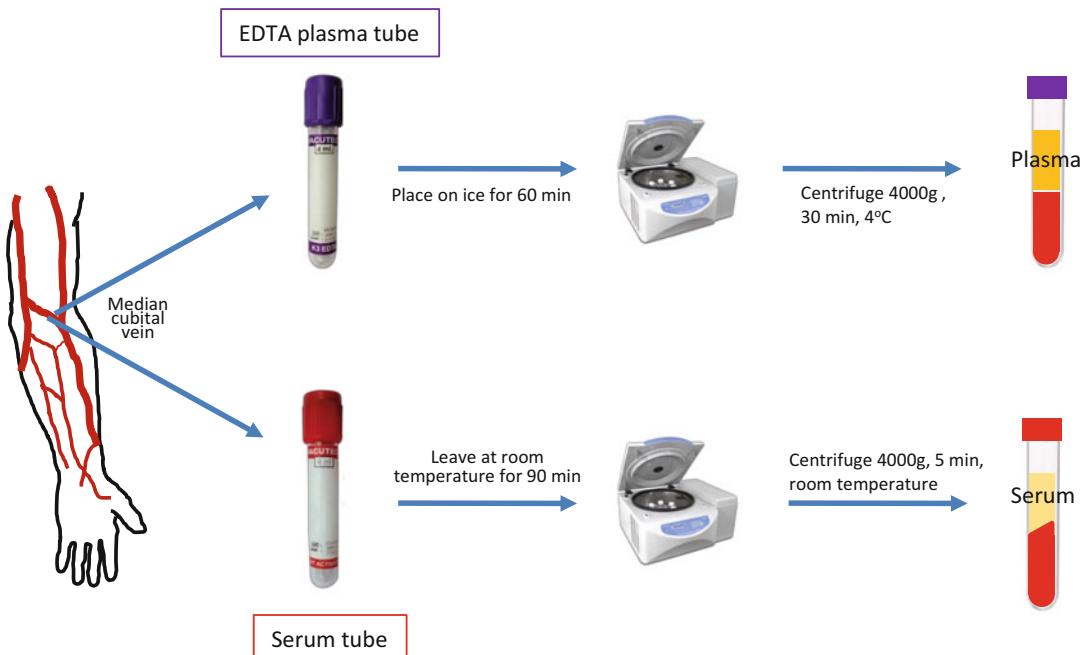
---

### **1 Introduction**

Multiplex proteomic and metabolomic techniques have increased in their use in recent years in the ongoing quest of finding novel biomarkers for life-threatening diseases. In clinical development, biomarkers discovery, validation, and translating the findings into a companion diagnostic requires reproducible sample collection and handling procedures. Whenever possible, it is important that bio-fluids such as serum or plasma are chosen for these studies as blood is readily accessible from most study populations by standard venous puncture techniques. Therefore, these are likely to result in discovery and development of biomarker tests with greater clinical utility and even support point-of-care use [1, 2] and offer a rapid means of analysis in emergency room situations [3]. However, it is vital that procedures for blood draw, plasma and serum preparation and storage are standardized to allow successful comparison of results across different laboratories in validation and product development studies.

Plasma is derived from blood by the addition of an anticoagulant such as EDTA, heparin, or sodium citrate to inhibit the clotting process, followed by removal of blood cells and cell debris by centrifugation (Fig. 1). This procedure is carried out rapidly and samples can be aliquoted in low-binding tubes and stored at  $-80^{\circ}\text{C}$  for several months or even years. Production of serum is different due to the fact that no anticoagulant is added and the resulting coagulated material, which is comprised mostly of cells, cell debris, and clotting factors, is removed by the centrifugation step [4]. This method requires leaving the samples for approximately 90 min to allow clotting or adding a clot activator to reduce the clotting time, prior to aliquotting and storage as above. Here, we present standard protocols for venous puncture along with plasma and serum preparation and storage techniques, which can be used in multiplex immunoassay [5], two-dimensional gel electrophoresis [6], tandem mass spectrometry [7], selected reaction monitoring mass spectrometry [8], proton nucleic acid resonance spectroscopy [9], and many other multiplex biomarker profiling approaches.

It should be stressed that selecting the adequate bio-sampling and processing methods can influence the results of biomarker studies. Therefore, compatibility between the preparation and analytical phases should be established prior to carrying out studies [10–12]. To ensure uniformity and consistency of bio-fluid processing, standard operating procedures (SOPs) are generally set up and applied uniformly. Specimens such as serum and plasma should



**Fig. 1** Schematic diagram showing blood collection for plasma and serum preparation

be prepared and stored according to the SOP and amendments added and recorded to trace back any discrepancies that might influence study outcomes.

## 2 Materials

### **2.1 Blood Draw**

1. Sterile blood draw needles (14–20 gauge).
2. Holder/adapter for use with the collection system (*see Note 1*).
3. Tourniquet.
4. Alcohol wipes (70% isopropyl alcohol).
5. Gloves (Latex, rubber or vinyl).

### **2.2 Serum and/or Plasma Preparation (See Note 2)**

1. Evacuated collection tubes (*see Note 3*).
2. Sterile serological pipettes of appropriate volumes.
3. Benchtop centrifuge (*see Note 4*).
4. Storage tubes or cryovials (*see Note 5*).

### **2.3 Serum Preparation Only**

1. Tubes with no additive or clotting agent (*see Note 6*).
2. Serum Vacutainer tubes with clot activator +/-gel separator (*see Note 7*).

### **2.4 Plasma Preparation Only (See Note 8)**

1. Tubes with chelating agent +/-gel separator (*see Note 9*).

## 3 Methods

### **3.1 Blood Draw (See Note 10)**

1. Prior to blood draw, record each donor's demographic details, physiological status, and other metadata in a Clinical laboratory worksheet approximately as described in Table 1 (*see Note 11*).
2. Perform venipuncture selecting the most appropriate arm vein of the participant (*see Note 18*).
3. Clean the participant's arm with alcohol in a circular fashion, beginning at the site and working outward and allow to air dry.
4. Insert the needle at an angle that is 20–30° of the vein, avoid trauma and excessive probing (*see Note 19*).
5. Draw 8–10 mL of whole blood for each 4–5 mL of serum or plasma required

### **3.2 Serum Preparation**

1. Collect up to 8 mL of whole blood in serum tubes.
2. Immediately after collection, invert the tube 8–10 times.
3. Allow tubes to clot in the vertical position for 90 min at room temperature (*see Note 20*).

**Table 1****Typical metadata and dietary/lifestyle that may be collected for a biomarker study (see Note 12)**

|  |   |
|--|---|
| Samples                                | 1. Type of sample collected (e.g., serum and/or plasma) ( <i>see Note 13</i> )<br>2. Date and time of sample withdrawal<br>3. Time until freezing<br>4. Date of collection ( <i>see Note 14</i> )<br>5. Additives to sample ( <i>see Note 15</i> )  |
| Demographics                           | 1. Gender<br>2. Ethnicity ( <i>see Note 16</i> )<br>3. Height, weight and body mass index (BMI)<br>4. Hip and waist measurement<br>5. Fasting status<br>6. Smoking status (number cigarettes/day, duration)<br>7. Alcohol consumption<br>8. Hormonal status (e.g., menstruation, menopause, hormonal treatment)<br>9. Pregnancy or breastfeeding<br>10. Disease onset, duration and current medication (including dosage)<br>11. Co-morbidities (e.g., presence of respiratory/cardiovascular disease, diabetes, osteoarthritis, chronic inflammatory diseases, cancer, mental diseases) including disease duration and regular medications taken |
| Physiological/<br>biochemical analysis | 1. Blood pressure diastolic/ systolic (mm Hg)<br>2. Blood clinical laboratory readouts such as blood count, urea, creatinine, glucose, lipids<br>3. Total protein, albumin<br>4. Sodium, potassium<br>5. Liver enzymes<br>6. Glucose tolerance test<br>7. C-reactive protein  |
| Dietary and lifestyle                  | 1. Type of food and beverages consumed prior to sample collection<br>2. If possible, participants should fast overnight before specimen withdrawal ( <i>see Note 17</i> )<br>3. If applicable, record medications type and dose<br>4. Participants should refrain from heavy exercise, alcohol, tobacco and nicotine use for 12 h prior to specimen collection<br>5. Use and levels (when possible) of illicit drugs should be recorded   |

4. Centrifuge at  $1100 \times g$  for 15 min at 4 °C.
5. Transfer 0.5 mL aliquots of the top serum layer to pre-labeled 1.5 mL-capacity Eppendorf Lobind tubes on ice (*see Note 5*).
6. Record and discard samples that are hemolyzed (red or pink tingeing) or those that show lipemia (floating milky white substance formed by the accumulation of lipoprotein particles).
7. Place aliquots immediately on dry ice and transfer to a –80 °C freezer until analysis (*see Note 21*).

### **3.3 Plasma Preparation**

1. Collect up to 10 mL of whole blood in plasma tubes.
2. Immediately after collection, invert the tube 8–10 times.
3. Place tubes on wet ice for 30 min (*see Note 22*).

4. Centrifuge at  $1100 \times g$  for 15 min at 4 °C.
5. Transfer 0.5 mL aliquots of top plasma layer to pre-labeled 1.5 mL-capacity Eppendorf LoBind tubes on ice (*see Notes 5 and 23*).
6. Record and discard samples that are hemolyzed or show lipemia as above.
7. Place aliquots immediately on dry ice and transfer to a –80 °C freezer until analysis as described above for serum.

---

## 4 Notes

1. Evacuated systems are available for use with a syringe, single draw, or butterfly system.
2. Blood proteome-based biomarker profiles can be influenced by the choice of serum versus plasma. This is due to differences in both content and stability of the resident molecules [4].
3. Tube type with and without stabilizer should be determined at the start of the study and kept in constant use throughout. This is also important for comparison of results across studies. For example, Hab et al reported marked differences in analyte measurements using immunoassays of EDTA-, heparin- and citrate-plasma, and serum [13].
4. Centrifugation at room versus refrigerated temperature can affect the stability of blood biomarkers (e.g., 4 °C is better suited for platelet preparation).
5. For volumes less than 1 mL, 1.5 mL-capacity Eppendorf Lo Bind tubes can be used. For volumes greater than 1 mL, cryovials can be used. It is important to pre-label all tubes.
6. We normally use 10 mL Vacutainer Plus tubes with a clear cap. These can be purchased with or without a gel separator. We suggest not to use tubes with a gel separator as this can interfere with some assays (e.g., for Pharmacokinetic and pharmacodynamic investigations).
7. We normally use the 10 mL BD Vacutainer Plus plastic serum tubes with either a red or mottled red/gray cap.
8. The choice of chelating agent may be important as this could affect performance of the chosen biomarker profiling platform (e.g., the use of EDTA tubes leads to more reproducible results compared to citrate or heparin tubes with the Luminex multiplex immunoassay system). We suggest contacting the manufacturer of the intended platform for advice in the matter prior to initiation of the study. It should also be noted that protease inhibitors are not generally needed when storing plasma in EDTA tubes as this reagent can inhibit most proteases through chelation of metal ion co-factors, required for protease activity.

9. We normally use 10 mL Vacutainer K<sub>2</sub>EDTA tubes from BD Bioscience or equivalent. Tubes are also available containing sodium citrate or sodium heparin. The choice is dependent on compatibility of the chelating reagent with the biomarker profiling platform, as described above.
10. It is important to remember that safety always comes first. All bio-samples and materials should be handled as if capable of transmitting infection and disposed of with proper precaution in accordance with state and local regulations. It is important to avoid contact of bio-samples with skin and mucous membranes. Clinical sample disposals are usually performed in accordance with the local guidance and rules for the safe use and disposal at containment level 2 or level 3 as necessary.
11. Consistency is important. Ensure that all measurements are carried out using the same systems and operators if possible. In addition, attempt to collect all data from all participants to avoid missing information. This is important to allow corrections for potential confounding factors during the data analysis phase. Up to 46% of laboratory errors are known to be associated with pre-analytic processing [14] and can even generate false readings of blood biomarker levels [15].
12. The median cubital vein is used most commonly due to ease of location and size (Fig. 1).
13. This stage should be performed by a trained phlebotomist.
14. Clotting time can be reduced to approximately 30 min if tubes containing a clotting factor are used. However, these tubes should be used consistently throughout the study.
15. There can be problems using temperatures ranging from -20 °C to -30 °C, such as increased degradation or cryoprecipitation of the molecular content, and this can affect biomarker profiling results [16–19]. Problems can also occur with multiple freeze-thaw cycles so these should be minimized. Freezers should be monitored by an automated security alarm and back-up systems, such as spare freezers, should be in place for emergencies. Upon first thaw, protease inhibitors can be added as needed and if these do not interfere with the intended biomarker profiling platform. Using additives such protease and phosphatase inhibitors may help to stabilize proteome profiles [19].
16. Times can vary but the main point is to keep this consistent.
17. When acquiring the plasma, take care not to contact the mono-nuclear cells and platelets, which occur in a white-colored layer just under the plasma layer.
18. This is a general guideline and the information required may vary from study to study. The important point is consistency.

19. Collect samples from test and control cases randomly to avoid biased statistical outcome.
20. Seasonal effects can sometimes alter the results of biomarker testing.
21. Additives can also affect profiling results.
22. Participant distribution should be equally represented to facilitate successful validation phases or repeat studies at alternate sites.
23. If participants have not fasted it is important to record this along with the time/nature of the last meal and potentially take measurements of glucose and insulin levels.

## References

1. Guest PC, Guest FL, Martins-de Souza D (2015) Making sense of blood-based proteomics and metabolomics in psychiatric research. *Int J Neuropsychopharmacol* pii:pyv138. doi:[10.1093/ijnp/pyv138](https://doi.org/10.1093/ijnp/pyv138)
2. Guest FL, Guest PC, Martins-de-Souza D (2016) The emergence of point-of-care blood-based biomarker testing for psychiatric disorders: enabling personalized medicine. *Biomark Med* 10:431–443
3. Singhal N, Saha A (2014) Bedside biomarkers in pediatric cardio renal injuries in emergency. *Int J Crit Illn Inj Sci* 4:238–246
4. Alsaif M, Guest PC, Schwarz E, Reif A, Kittel-Schneider S, Spain M et al (2012) Analysis of serum and plasma identifies differences in molecular coverage, measurement variability, and candidate biomarker selection. *Proteomics Clin Appl* 6:297–303
5. Fulton RJ, McDade RL, Smith PL, Kienker LJ, Kettman JR Jr (1997) Advanced multiplexed analysis with the flowmetrix system. *Clin Chem* 43:1749–1756
6. Unlü M, Morgan ME, Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18:2071–2077
7. Paulo JA, Kadiyala V, Banks PA, Steen H, Conwell DL (2012) Mass spectrometry-based proteomics for translational research: a technical overview. *Yale J Biol Med* 85:59–73
8. Ji QC, Rodila R, Gage EM, El-Shourbagy TA (2003) A strategy of plasma protein quantitation by selective reaction monitoring of an intact protein. *Anal Chem* 75:7008–7014
9. Griffiths WJ, Koal T, Wang Y, Kohl M, Enot DP, Deigner HP (2010) Targeted metabolomics for biomarker discovery. *Angew Chem Int Ed Engl* 49:5426–5445
10. Beltran A, Suarez M, Rodríguez MA, Vinaixa M, Samino S, Arola L et al (2012) Assessment of compatibility between extraction methods for NMR- and LC/MS-based metabolomics. *Anal Chem* 84:5838–5844
11. López E, Madero L, López-Pascual J, Latterich M (2012) Clinical proteomics and OMICS clues useful in translational medicine research. *Proc Natl Acad Sci U S A* 10:35. doi:[10.1186/1477-5956-10-35](https://doi.org/10.1186/1477-5956-10-35)
12. Hebels DG, Georgiadis P, Keun HC, Athersuch TJ, Vineis P, Vermeulen R et al (2013) Performance in omics analyses of blood samples in long-term storage: opportunities for the exploitation of existing biobanks in environmental health research. *Environ Health Perspect* 121:480–487
13. Haab BB, Geierstanger BH, Michailidis G, Vitzthum F et al (2005) Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: systematic variation between sample types and calibration of mass spectrometry data. *Proteomics* 5:3278–3291
14. Bevan-McBride K (1999) Laboratory sampling: does the process affect the outcome? *J Intraven Nurs* 22:137–142
15. Bowen RA, Hortin GL, Csako G, Otañez OH, Remaley AT (2010) Impact of blood collection devices on clinical chemistry assays. *Clin Biochem* 43:4–25
16. Zander J, Bruegel M, Kleinheimpel A BS, Petros S, Kortz L et al (2014) Effect of biobanking conditions on short-term stability of biomarkers in human serum and plasma. *Clin Chem Lab Med* 52:629–639

17. Kang HJ, Jeon SY, Park JS, Yun JY, Kil HN, Hong WK et al (2013) Identification of clinical biomarkers for pre-analytical quality control of blood samples. *Biopreserv Biobank* 11:94–100
18. Hsieh SY, Chen RK, Pan YH, Lee HL (2006) Systematical evaluation of the effects of sample collection procedures on low-molecular-weight serum/plasma proteome profiling. *Proteomics* 6:3189–3198
19. Rai AJ, Gelfand CA, Haywood BC, Warunek DJ, Yi J, Schuchard MD et al (2005) HUPO plasma proteome project specimen collection and handling: towards the standardization of parameters for plasma proteome samples. *Proteomics* 5:3262–3277

# Chapter 13

## Multiplex Immunoassay Profiling

Laurie Stephen

### Abstract

Multiplex immunoassays allow for the rapid profiling of biomarker proteins in biological fluids, using less sample and labor than single immunoassays. This chapter details the methods to develop and manufacture multiplex assays for the Luminex® platform. Although assay development is not included here, the same methods can be used to covalently couple antibodies to the Luminex beads and to label antibodies for the screening of sandwich pairs, if needed. The assay optimization, detection of cross-reactivity, and minimizing antibody interactions and matrix interferences will be addressed.

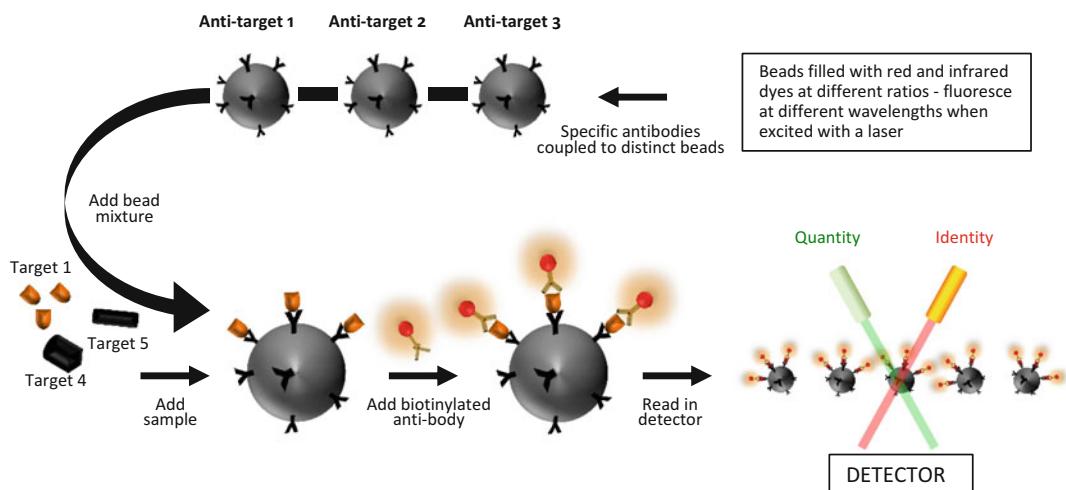
**Key words** Disease, Multiplex assay, Antibody, Biomarker, Luminex® assay

---

### 1 Introduction

A large majority of immunoassays rely on antibodies to capture and detect the analyte of interest in a biological matrix. Traditional immunoassays detect the presence of a single analyte and most rely on enzyme-driven detections. New technologies have been developed that allow multiple analytes to be measured simultaneously on a single sample, all in a single reaction vessel. The method described here is for the development of a multiplex assay for the Luminex® system, although the same principal will apply to other technologies (Fig. 1). For information on multiplexing technologies, there are several recent publications that review the latest developments [1–4].

As with any new technology, there are unique advantages and disadvantages that the user encounters. For example, the ability to simultaneously measure multiple analytes in a single sample maximizes the amount of information that can be obtained from single sample, reduces laboratory analysis time and sample volume requirement, and provides cost savings. However, multiplexed assays also present unique challenges for the user that would not be encountered if single assays were used for each individual analyte. Examples of these include different detection ranges, the potential



**Fig. 1** Overview of multiplex immunoassay technique. Samples are added to dye-coded microbead-antibody conjugates that capture specific targets. Following incubation with a second antibody containing a biotinylated label to form a “sandwich” configuration, the mixtures are streamed through the Luminex instrument that uses lasers for the identification of the antibody-microbead conjugates and quantitation of the bound molecules. The example shows a triplex assay capable of binding targets 1, 2, and 3. Since the sample only contains target 1, this is the only target bound and quantified

for cross-reactivity, increased matrix interference, and potential for false positives due to antibody interactions. These challenges, if not carefully addressed, can generate misleading results. Several papers have addressed the issue of increased interference in multiplex immunoassays [5–7] and others have described the additional challenges involved in the validation stage [8–10]. This chapter will describe a method to develop a multiplex immunoassay and strategies for minimizing interference and interactions within the assay.

## 2 Materials

### 2.1 Bead Conjugation

1. Magnetic separator.
2. Sonicating bath.
3. Copolymer tubes and labels.
4. 1–4 mL of magnetic  $12.5 \times 10^6$  beads/mL Luminex microspheres/beads.
5. 125 µg/mL capture antibodies (*see Note 1*).
6. Sulfo-NHS, N-hydroxysulfosuccinimide.
7. N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide (EDC).
8. Activation buffer: 100 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 6.0).
9. Coupling buffer: 0.05 M 2-morpholino-ethane-sulfonic acid mono-hydrate (MES) (pH 5.0).

10. Blocking/storage buffer: 10 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 7.4), 150 mM NaCl, 0.02% Tween 20, 0.1% bovine serum albumin (BSA), and 0.05% NaN<sub>3</sub> or Proclin (PBS-TBN).

## **2.2 Biotinylation of Detection Antibody**

1. Sulfo-NHS-LC biotin (ThermoFisher Scientific; Waltham, MA, USA).
2. Dialysis unit.
3. Phosphate-buffered saline (PBS) (pH 7.4).

## **2.3 Multiplex Development**

1. Assay buffer: PBS, 1% BSA.
2. Wash buffer: PBS, 0.02% Tween-20.
3. 100 µg/mL Streptavidin, R-phycoerythrin (SAPE).
4. 96-well plate.
5. Plate washer or magnetic separator for 96-well plate.
6. Recombinant protein standards for each assay in the multiplex.
7. Representative serum and /or plasma samples from rheumatoid arthritis (RA) patients or with known rheumatoid factor and controls.
8. Animals sera (e.g., horse, goat, mouse, rabbit, donkey, fetal calf).
9. Heterophilic blocking reagents (such as IIR from Bioreclamation; Tru-Block from Meridian Life Sciences or HBR from Scantibodies).

## **3 Methods**

### **3.1 Bead Conjugation**

1. Allow 1–4 mL of stock microspheres, to settle by placing vials upright on a flat magnetic separator for 2 min (*see Note 2*).
2. Taking care not to disturb settled beads remove and discard 0.8 mL of buffer for every 1 mL stock beads.
3. Transfer and pool the remaining volume into a single copolymer tube.
4. Place tube in the magnetic separator and allow separation to occur for 30–60 s.
5. With tube in separator remove and discard supernatant.
6. Add 0.5 mL of activation buffer, resuspend with vortexing and sonication, and place in the magnetic separator for 30–60 s.
7. With tube in separator remove and discard supernatant, and resuspend in 0.4 mL of activation buffer with vortexing and sonication.
8. Add activation buffer to Sulfo-NHS for a final concentration of 50 mg/mL and add 50 µL of this to the tube and vortex.

9. Add activation buffer to EDC for a final concentration of 10 mg/mL and add 50 µL of this to the tube and vortex.
10. Incubate 20 min in the dark while rotating at room temperature and prepare the protein during the incubation.
11. Place tube in a magnetic separator for 30–60 s, remove the supernatant, and add 0.5 mL of coupling buffer.
12. Repeat for a total of two washes and resuspend in 0.45 mL of coupling buffer.
13. Add 0.2 mL of capture antibody to the activated microspheres with immediate vortexing (*see Note 2*).
14. Incubate for 2 h in the dark while rotating at room temperature.
15. Place tube in a magnetic separator for 30–60 s, remove supernatant, and add 1.0 mL of blocking/storage buffer.
16. Resuspend with vortexing and sonication and incubate for 30 min in the dark with rotation at room temperature.
17. Place tube in magnetic separator for 30–60 s, remove supernatant, and wash twice with 0.25 mL blocking/storage buffer.
18. Resuspend in 0.25 mL blocking/storage buffer, count the beads with a hemocytometer, adjust to  $50 \times 10^6$  beads/mL, and store at 2–8 °C.

### 3.2 Biotinylation

1. Immediately before use, prepare a 10 mM solution of the biotin reagent.
2. Add 10 mM biotin reagent solution to the antibody solution at a 20:1 biotin:antibody molar ratio (*see Note 3*).
3. Incubate reaction on ice for two hours or at room temperature for 30 min.
4. Remove excess biotin by dialysis in PBS using a minimum of three buffer exchanges.
5. Add BSA to a final concentration of 1% and preservative for long-term stability.

### 3.3 General Protocol

1. Create a capture bead mini-pool by adding 5 µL of each bead solution to a final volume of 1.4 mL in assay buffer.
2. Make an 8 standard (S8) mini-pool by adding 0.2 µg of each recombinant protein to a final volume of 0.2 mL in assay buffer and do seven 10-fold serial dilutions to create a standard curve.
3. Make a mini-pool mix of detection antibodies by adding 5 µg of each biotinylated antibody to a final volume of 5 mL in assay buffer.
4. Produce 1:5, 1:10, 1:100, and 1:1000 serial dilutions of serum and plasma samples in assay buffer (*see Note 4*).
5. Add 30 µL of standard or sample to a well of the 96-well plate.

6. Add 10 µL of blocking solution and then add 10 µL of capture beads.
7. Incubate the plate for 1 h on a plate shaker at room temperature.
8. Wash three times with 100 µL wash buffer, add 40 µL of the detection mini-pool to each well, and incubate the plate for 1 h on a plate shaker at room temperature.
9. Add 20 µL SAPE to each well plate and mix for 30 min on a plate shaker at room temperature (*see Note 5*).
10. Wash three times with 100 µL wash buffer and add 100 µL assay buffer.
11. Incubate the plate for 2–5 min on a plate shaker at room temperature and then analyze on the Luminex 100 Analyzer.

### **3.4 Curve Optimization (See Note 6)**

1. Examine the standard curve median fluorescence intensities (MFI) and choose the best four points for each that cover the range of your sample signals.
2. Make a new standard mini-pool based on the above (Table 1) (*see Note 7*).

**Table 1**  
**Screen of samples and animal sera**

| Name          | Analyte 1 | Analyte 2 | Analyte 3 | Analyte 4 |
|---------------|-----------|-----------|-----------|-----------|
| S8            | 6581      | 18605     | 24370     | 8669      |
| S7            | 2965      | 8044      | 19134     | 3009      |
| S6            | 1242      | 3301      | 9323      | 1128      |
| S5            | 528       | 1276      | 2855      | 430       |
| S4            | 250       | 554       | 805       | 185       |
| S3            | 123       | 257       | 265       | 104       |
| S2            | 86        | 143       | 155       | 78        |
| S1            | 76        | 109       | 121       | 67        |
| Serum 1 1:5   | 2894      | 181       | 281       | 224       |
| Serum 1 1:10  | 1127      | 80        | 197       | 121       |
| Rabbit serum  | 2689      | 67        | 154       | 145       |
| Goat serum    | 116       | 59        | 116       | 195       |
| Hamster serum | 89        | 79        | 138       | 265       |
| Mouse serum   | 82        | 71        | 112       | 49        |
| Rat serum     | 330       | 71        | 289       | 23        |
| Donkey serum  | 30        | 65        | 80        | 35        |

3. By examining where the samples fall on the curve, several mini-pools of detections can be made to determine the lowest concentration of each antibody needed (*see Note 8*)

### **3.5 Blocker Optimization (See Note 9)**

1. Make various blocker formulations with and without 1% donkey serum and include assay buffer as one formulation.
2. Run samples (including RA samples) at two dilutions with each blocker to determine the optimal blocker formulation.
3. Calculate the sample linearity and choose the blocker that gives linearity values 80–120% and the highest sample signals.

### **3.6 Assessing Cross-Reactivity**

1. Make a multiplex standard mini-pool along with single mini-pools of each standard at the same concentrations as those in the multiplex.
2. Make a detection mini-pool with single mini-pools of each detection antibody at the same concentrations as in the multiplex.
3. In order to assess standard cross-reactivity, run the assay with multiplex beads and compare the results with each single assay standard curve (*see Note 10*).
4. In order to assess detection cross-reactivity, run the assay with multiplex beads and standards, along with several positive serum or plasma samples and test each single detection pool.
5. Calculate any signals and compare to the multiplex curve to calculate cross-reactivity (*see Note 11*).

### **3.7 Packaging and Use**

1. Batch bead and detection mixes: package individually in assay buffer using the volumes listed in the general protocol above.
2. Batch blocker and S8: package in standard diluent.
3. Standard diluent, assay buffer, SAPE and blocker: package individually.
4. Store beads, SAPE and assay buffer at 4 °C, and all other components at –80 °C (*see Note 12*).
5. Make two to three levels of assay quality controls (QC) by spiking recombinant protein into serum samples and package these individually and store at –80 °C.
6. Proceed with assay validation and the generation of assay acceptance criteria by testing a minimum of 20 of each QC in duplicate over a minimum of 3 days.

---

## **4 Notes**

1. Prior to coupling or biotinylation of the antibodies, they should be free of any amines and other proteins. If using a Tris-based buffer, dialyze the antibody into 1× PBS using a minimum of

three buffer exchanges. If the antibody preparation contains stabilizer proteins such as BSA or gelatin, purify with Protein A or Protein G columns, followed by dialysis.

2. 1 mL of microspheres is sufficient for ~40 plates of assays.
3. The molecular weight of immunoglobulin G (IgG) is 150,000 kDa, so the amount of biotin required for a 1 mL of a 1 mg/mL antibody solution could be calculated as follows:  

$$1 \text{ mL IgG} \times 1 \text{ mg/mL} \times 20 \text{ mmol biotin} / 1 \text{ mmol IgG} \times 1 \text{ mmol IgG} / 150,000 \text{ mg IgG} \times 1000 \text{ ul/mL} = 0.133 \text{ mmol biotin} = 13 \mu\text{L of the } 10 \text{ mM biotin solution.}$$
4. The assay will be multiplex based on the dilutions requirement of the samples. In general, if the expected levels are in the pg/mL range, they will be at the same dilution; however, it is unlikely that an analyte in the pg/mL range will have the same dilution requirement as an analyte in the µg/mL range.
5. The SAPE concentration will vary with the number of analytes in the multiplex.
6. For optimization of the curve, repeat the assay with the indicated changes.
7. In the example in Table 1, the sample dilution was determined to be 1:5. In order to match the matrix and minimize curve background, the standard curve should be made up in 20% donkey serum.
8. The detection concentration for Analyte 1 can likely be lowered, whereas that of Analyte 2 should be increased.
9. Repeat the assay with optimized detection and standard concentrations using the indicated changes.
10. If cross-reactivity is greater than 10%, a different recombinant standard may be needed.
11. If cross-reactivity is greater than 10%, the detection in question can be titrated down and the assay repeated. If there is no improvement, the assay may need to be removed from the multiplex.
12. The detection buffer can also be stored at 4 °C if space is limiting.

## References

1. Tighe P, Negm O, Todd I, Fairclough L (2013) Utility, reliability and reproducibility of immunoassay multiplex kits. *Methods* 61:23–29
2. Ellington AA, Kullo IJ, Bailey RK, Klee GG (2010) Antibody-based protein multiplex platforms: technical and operational challenges. *Clin Chem* 56:186–193
3. Spindel S, Sapsford KE (2014) Evaluation of optical detection platforms for multiplexed detection of proteins and the need for point-of-care biosensors for clinical use sensors. *Sensors (Basel)* 14:22313–22341
4. Marquette CA, Corgier BP, Blum LJ (2012) Recent advances in multiplex immunoassays. *Bioanalysis* 4:927–936

5. Todd DJ, Knowlton N, Amato M, Frank MB, Schur PH, Izmailova ES et al (2011) Erroneous augmentation of multiplex assay measurements in patients with rheumatoid arthritis due to heterophilic binding by serum rheumatoid factor. *Arthritis Rheum* 63:894–903
6. Fraser S, Soderstrom C (2014) Due diligence in the characterization of matrix effects in a total IL-13 Singulex™ method. *Bioanalysis* 6:1123–1129
7. Krika LJ (1999) Human anti-animal antibody interferences in immunological assays. *Clin Chem* 45:942–956
8. Jani D, Allinson J, Berisha F, Cowan KJ, Devanarayan V, Gleason C et al (2016) Recommendations for use and fit-for-purpose validation of biomarker multiplex ligand binding assays in drug development. *AAPS J* 18:1–14
9. Bastarache JA, Koyama Y, Wickersham NE, Ware LB (2014) Validation of a multiplex electrochemiluminescent immunoassay platform in human and mouse samples. *J Immunol Methods* 408:13–23
10. Tighe PJ, Ryder RR, Todd I, Fairclough LC (2015) ELISA in the multiplex era: potentials and pitfalls. *Proteomics Clin Appl* 9:1862–8354

# Chapter 14

## Multiplex Sequential Immunoprecipitation of Insulin Secretory Granule Proteins from Radiolabeled Pancreatic Islets

Paul C. Guest

### Abstract

Pulse radiolabeling of cells with radioactive amino acids is a common method for tracking the biosynthesis of proteins. Specific proteins can then be immunoprecipitated and analyzed by electrophoresis and imaging techniques. This chapter presents a protocol for the biosynthetic labeling of pancreatic islets with  $^{35}\text{S}$ -methionine, followed by multiplex sequential immunoprecipitation of insulin and three other secretory granule accessory proteins. This provided a means of distinguishing those pancreatic islet proteins with different biosynthetic rates in response to the media glucose concentrations.

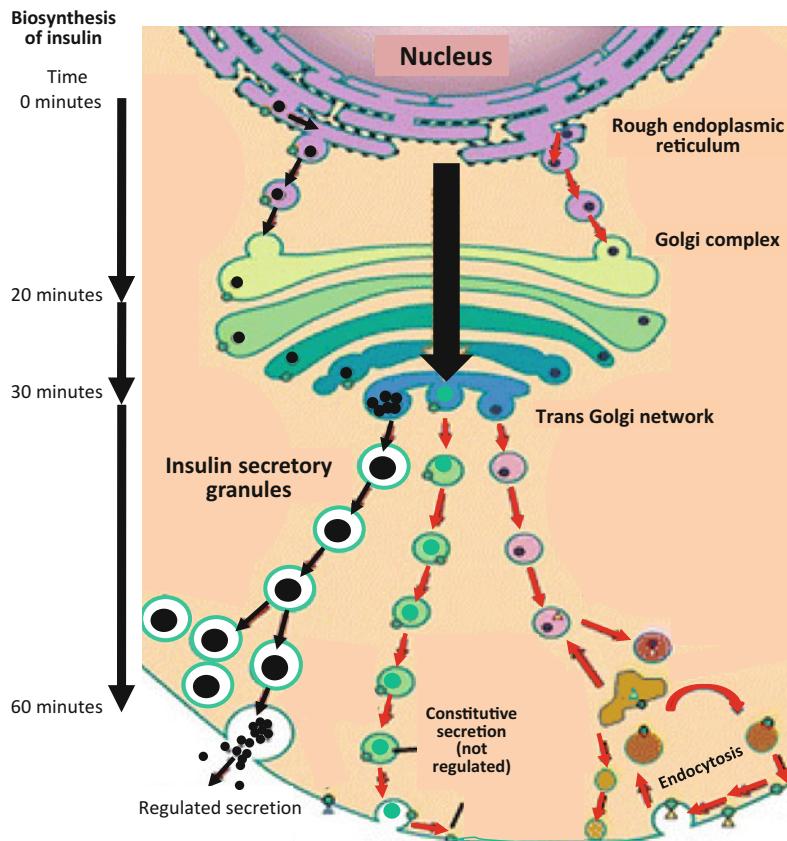
**Key words** Pulse radiolabeling, Immunoprecipitation, Electrophoresis, Pancreatic islets, Insulin, Secretory granule proteins

---

### 1 Introduction

Pulse radiolabeling of cells followed by immunoprecipitation is a means of looking at the biosynthesis of targeted proteins [1, 2]. In this method, a radiolabeled amino acid is usually added to the medium so this can be incorporated into nascent proteins as they are being synthesized. Then the newly synthesized proteins can be immunoprecipitated by direct or indirect means and analyzed at any time, thereby providing a means of tracking their fate inside the cell. Direct immunoprecipitations are carried out by covalently linking an antibody of interest to a solid matrix such as cyanogen bromide (CNBr)-activated Sepharose [3, 4]. In this approach, antibodies are coupled directly to the resin through primary amines. Indirect immunoprecipitations are carried out by non-covalently binding the antibody to an affinity-based matrix such as Protein A Sepharose. Then in both direct and indirect protocols, cell lysates can be incubated with the resulting immunoabsorbents for binding, elution, and subsequent analysis of the target proteins.

This chapter describes a 20 min pulse labeling of pancreatic islets with  $^{35}\text{S}$ -methionine and the sequential immunoprecipitation of the hormone insulin and the secretory granule accessory proteins chromogranin A, prohormone convertase (PC)1 and PC2, as described in previous studies [5–7]. Enzymological analyses have shown that production of mature insulin requires cleavage of proinsulin by PC1 and PC2 on the carboxy-terminal side of Arg<sup>31</sup>-Arg<sup>32</sup> and Lys<sup>64</sup>-Arg<sup>65</sup>, respectively [8]. PC1 has a pH optimum of 5.5 and requires mM calcium ion concentration, which coincides with the environment inside insulin secretory granules [2]. This fits with finding that final conversion of proinsulin to insulin does not begin to occur until approximately 30 min after initial synthesis on the rough endoplasmic reticulum (Fig. 1). Thus, the 20 min labeling period employed here helps to ensure that insulin is still present mostly in its precursor form, making



**Fig. 1** Diagram of a pancreatic islet cell showing the biosynthesis of insulin and transport through the regulated secretory pathway

quantitative studies more direct. Here, the preparation of immunoadsorbents, pulse radiolabeling of islets, immunoprecipitation, and gel-based analyses is presented.

---

## 2 Materials

1. Purified monoclonal antibody for insulin (*see Note 1*).
2. Polyconal antisera for chromogranin A, PC1 and PC2 (*see Note 2*).
3. CNBr-Activated Sepharose 4 (GE Healthcare; Little Chalfont, Bucks, UK).
4. Activation solution: 1 mM HCl.
5. Coupling buffer: 100 mM NaHCO<sub>3</sub> (pH 8.3) and 500 mM NaCl.
6. Quenching buffer: 100 mM Tris-HCl (pH 8.0).
7. Wash buffer 1: 100 mM Tris-HCl (pH 8.0) and 500 mM NaCl.
8. Wash buffer 2: 100 mM NaOAc (pH 4.0) and 500 mM NaCl.
9. Storage buffer: 10 mM NaH<sub>2</sub>PO<sub>4</sub>, 2 mM K<sub>2</sub>HPO<sub>4</sub>, 137 mM NaCl, 2.7 mM KCl.
10. Modified Kreb's bicarbonate buffer: 25 mM NaHCO<sub>3</sub> (pH 7.4), 115 mM NaCl, 5.9 mM KCl, 1.2 mM MgCl<sub>2</sub>, 1.2 mM NaH<sub>2</sub>PO<sub>4</sub>, 1.2 mM Na<sub>2</sub>SO<sub>4</sub>, 2.5 mM CaCl<sub>2</sub>, and 0.1% bovine serum albumin.
11. Islet lysis buffer: 25 mM Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub> (pH 9), 3% BSA, 1% Tween-20, 1 mM phenylmethanesulfonyl fluoride, 0.1 mM E-64, 1 mM EDTA, and 0.1% NaN<sub>3</sub> (*see Note 3*).
12. Post-immunoprecipitation wash buffer: 50 mM Tris-HCl (pH 7.5), containing 150 mM NaCl, 1% Triton X-100, 1% deoxycholate, 0.1% SDS, and 5 mM-EDTA.
13. Insulin elution reagent: 25% acetic acid (*see Note 4*).
14. Protein A Sepharose (*see Note 5*).
15. Protein A Sepharose rehydration buffer: 20 mM NaH<sub>2</sub>PO<sub>4</sub> (pH 8.0) and 150 mM NaCl.
16. Reagent for elution of other islet proteins: 20 mM HCl.
17. <sup>35</sup>S-methionine.
18. Alkaline-urea electrophoresis gel: polymerized from 7.5% acrylamide and 0.20 NN'-methylenebisacrylamide, containing 12.5 mM Tris/80 mM glycine (pH 8.6) and 8 M urea.
19. Alkaline-urea gel tank buffer: 12.5 mM-Tris/80 mM glycine (pH 8.6).
20. Alkaline urea gel loading buffer: 2.5 mMTris-HCl (pH 8.6), 8 M urea, and 0.001% Bromophenol Blue.

21. SDS polyacrylamide electrophoresis: gels polymerized from 15% acrylamide and 0.080 NN'-methylenebisacrylamide in a Tris-glycine buffer system using the discontinuous buffer system of Laemmli [9].
22. Fluorography solution: 20% 2,5-diphenyloxazole in acetic acid.
23. MSE Sonifier (Crawley, UK) with a microprobe (*see Note 6*).

---

### 3 Methods

#### 3.1 Preparation of Anti-Insulin Immunoadsorbent

1. Dialyze the antibody into two changes of coupling buffer over 6 h at 4 °C (*see Note 7*).
2. Measure the absorbance at 280 nm of the final antibody solution and calculate the concentration (*see Note 8*).
3. Add 20 mL ice-cold activation solution to 1 g of dried resin and gently mix for 2 h at 4 °C (*see Note 9*).
4. Centrifuge the resin at  $1000 \times g$  for 5 min and remove the supernatant.
5. Add the dialyzed antibody to the resin at a concentration of 2 mg antibody/mL swollen resin and mix overnight at 4 °C.
6. Centrifuge the resin at  $1000 \times g$  for 5 min and remove the supernatant (*see Note 10*).
7. Add 20 mL of coupling buffer to the resin and gently mix for 30 min at room temperature.
8. Centrifuge at  $1000 \times g$  for 5 min and remove the supernatant.
9. Add 20 mL of quenching buffer and gently mix for 2 h at room temperature.
10. Centrifuge at  $1000 \times g$  for 5 min and remove the supernatant.
11. Suspend the resin in 20 mL wash buffer 1, centrifuge at  $1000 \times g$  for 5 min and remove the supernatant.
12. Suspend the resin in 20 mL wash buffer 1, centrifuge at  $1000 \times g$  for 5 min and remove the supernatant.
13. Repeat steps 11 and 12 twice and suspend the resin in 20 mL of storage buffer.
14. Centrifuge at  $1000 \times g$  for 5 min, remove the supernatant and suspend the resulting immunoadsorbent in 8 mL of storage buffer, and store at 4 °C (*see Note 11*).

#### 3.2 Preparation of Immunoadsorbent for Accessory Secretory Granule Proteins

1. Add 20 mL of rehydration buffer to 4 g of Protein A Sepharose powder and leave gently rocking for 30 min.
2. Centrifuge at  $1000 \times g$  for 5 min in a swinging bucket rotor, remove the supernatant, and suspend in 20 mL of the same buffer.

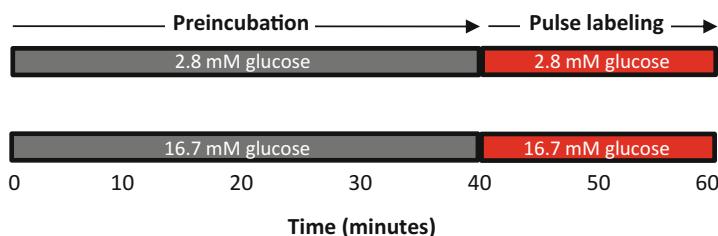
3. Repeat the centrifugation and washing steps twice and store in 16 mL of rehydration buffer at 4 °C (*see Note 11*).
4. Add 200 µL of suspended Protein A Sepharose slurry to microcentrifuge tubes containing 15 mL of each polyclonal antiserum and incubate with gentle rocking overnight at 4 °C (*see Note 12*).
5. Centrifuge at 1000 ×  $\text{g}$  for 5 min, remove supernatant, and wash three times by centrifugation and resuspension in 500 mL of the rehydration buffer.
6. Centrifuge at 1000 ×  $\text{g}$  for 5 min and remove the supernatant, leaving a packed gel of approximately 50 µL for immediate use.

### **3.3 Biosynthetic Radiolabeling of Pancreatic Islets**

1. Preincubate 100 isolated islets per experimental condition for 40 min in 100 µL of Kreb's bicarbonate buffer containing 2.6 or 16.7 mM glucose at 37 °C in microcentrifuge tubes under 95% O<sub>2</sub>/5% CO<sub>2</sub> (*see Note 13*).
2. Recover the islets by centrifugation at 100 ×  $\text{g}$  for 10 s and resuspend in 100 µL of the same pre-warmed medium containing 150 µCi of <sup>35</sup>S-methionine and incubate for 20 min at 37 °C under 95% O<sub>2</sub>/5% CO<sub>2</sub> (Fig. 2).
3. Recover the islets by centrifugation at 100 ×  $\text{g}$  for 10 s, carefully remove the radioactive supernatant, and gently resuspend the islets in 500 µL of the same ice-cold buffer containing 2 mM methionine (*see Note 14*).
4. Repeat this process twice and place the tubes containing the islet pellets on dry ice.
5. Add 200 µL lysis buffer and sonicate for 15 s at approximately 25 W (*see Note 15*).
6. Centrifuge the lysates at 13,000 ×  $\text{g}$  for 5 min and retain the supernatants for immunoprecipitation.

### **3.4 Immunoprecipitation of Insulin and Other Pancreatic Islet Proteins**

1. Incubate islet lysates for 1 h at room temperature in microcentrifuge tubes with 50 µL of a100 mg/mL suspension of Cowan-strain Staphylococcus aureus cells (*see Note 16*).
2. Centrifuge the samples at 13,000 ×  $\text{g}$  for 5 min and retain the supernatants.

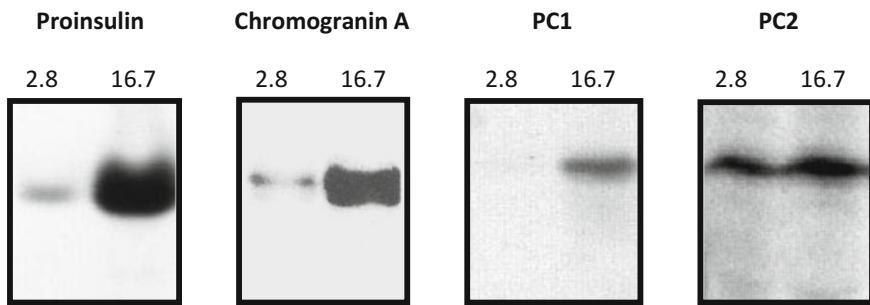


**Fig. 2** Schematic diagram showing the <sup>35</sup>S-methionine pulse labeling protocol in either low (2.8 mM) or high (16.7 mM) glucose concentrations

3. Carrying out immunoprecipitation of insulin-related molecules by adding the lysates to the 50 µL packed gel of respective immunoadsorbent and incubate overnight at 4 °C.
4. Centrifuge the anti-insulin immunoadsorbent in a swinging bucket rotor at  $500 \times g$  for 1 min and retain the supernatant for the next immunoprecipitation.
5. Wash the anti-insulin immunoadsorbent by repeated centrifugation and resuspension in  $4 \times 1$  mL of lysis buffer,  $2 \times 1$  mL of immunoadsorbent wash buffer, and then with  $2 \times 1$  mL of distilled water.
6. Elute with  $2 \times 1$  mL of 25% acetic acid, freeze dry and reconstitute in 50 mL of alkaline-urea gel loading buffer.
7. Incubate the supernatant obtained after immunoprecipitation of the insulin-related molecules overnight at 4 °C with 50 mL packed gel of anti-chromgranin A immunoadsorbent.
8. Wash the anti-chromgranin A immunoadsorbent as above and retain the supernatant for the next immunoprecipitation.
9. Elute the chromogranin A-related peptides with  $2 \times 100$  µL of 20 mM HCl.
10. Combine the two eluates, freeze dry and reconstitute in 50 µL of 125 mM Tris-HCl pH 6.8, containing 2% SDS, 0.25 M sucrose, 5 mM-EDTA, 65 mM-dithiothreitol, and 0.005% Bromophenol Blue.
11. Repeat Subheading 3.4, step 8 through Subheading 3.4, step 10 for immunoprecipitation of PC1 and PC2.

### **3.5 Electrophoresis and Fluorography of 3%S-Labeled Immunoprecipitates**

1. Prerun alkaline-urea gels in tank buffer for 600 V-hours, replace the upper tank buffer, and load the insulin immunoprecipitates in alkaline-urea gel loading buffer and subject to electrophoresis for 1000 V-hours (see Note 17).
2. Disassemble the gel plates and shake the gels for  $2 \times 5$  min in acetic acid, 2 h in fluorography solution, and then leave for 30 min under cold running water (see Note 18).
3. Vacuum dry and expose the gel to Cronex 4 X-ray film (Dupont; Stevenage, Herts, UK) for 6 to 72 h (see Note 19).
4. For the other islet proteins, heat the samples at 95 °C for 5 min and electrophorese to the point where the dye front just reaches the bottom of the gel.
5. Perform fluorography as above for 3–14 days (Fig. 3) (see Note 20).



**Fig. 3** Immunoprecipitation of proinsulin, chromogranin A, PC1, and PC2 from  $^{35}\text{S}$ -methionine-labeled islets. The concentrations of glucose (mM) used for the labeling are shown on the top of each track on the images (see Note 21)

#### 4 Notes

1. We used the clone 3B7 that recognizes epitopes in the rat proinsulin structure [8].
2. For chromogranin A, antiserum was raised in guinea pigs to a beta-galactosidase fusion protein incorporating amino acids 60-234 of the rat chromogranin A sequence as described by Hutton et al [5]. For PC1 and PC2, antisera were raised in rabbits against glutathione S-transferase fusion proteins incorporating amino acids 111-137 and 162-388 of rat PC1 and PC2 respectively. These fusion proteins were produced using the bacterial expression vector as described by Bennett et al [6].
3. Large amounts of BSA are added to prevent loss of protein on tube walls during immunoprecipitation. Insulin is known to be a “sticky” molecule.
4. Insulin elution requires highly acetic conditions due to the high affinity of the monoclonal antibody and the poor solubility of the insulin molecule.
5. To be used in indirect immunoprecipitation of the pancreatic islet proteins chromogranin A, PC1, and PC2.
6. Any sonication probe device can be used but the probe should fit inside a 1.5 mL-capacity microcentrifuge tube and therefore be not more than 2 mm in diameter at the tip.
7. It is important to remove all traces of Tris buffer from the antibody because this contains primary amines and will therefore react with the activated resin.
8. The ideal concentration is 2 mg antibody/mL swollen resin.
9. 1 g of dried resin will swell to give a volume of approximately 4 mL.
10. This should be saved in case the coupling did not work. This can be measured by reading the optical density at 280 nm in a

spectrophotometer and by looking for the loss of the monoclonal antibody from the solution. A good coupling efficiency would be greater than 80%.

11. Store for up to 1 month if not using preservatives.
12. In step, the immunoglobulin fraction in the serum is bound to the resin.
13. We have used isolated rat islets in this study although rat pancreatic beta cell lines can also be used. If this is the case,  $5 \times 10^5$  cells would be approximately equivalent to 100 islets as each islet contains around 5000 cells.
14. Use appropriate precautions when handling and disposing of the radioactive materials. The addition of ice-cold medium containing nonradioactive methionine stops the uptake of  $^{35}\text{S}$ -methionine and halts metabolic activity of the islet cells.
15. Adjust power setting accordingly using other sonication probe devices.
16. This is to preclear the supernatants by removing any immunoglobulin like molecules that could interfere with immunoprecipitation experiments.
17. This is equivalent to approximately 1 and 3/4 dye front lengths. We stop the electrophoresis after the dye has reached the bottom of the gel and then add new dye to a blank well at the end of the gel and restart the electrophoresis for the remaining three fourths gel length run.
18. The gel turns white during this final step.
19. Other films can be used but check the manufacturer's specifications. In addition, obtaining the best exposure may require a few attempts and adjusting the times accordingly.
20. Longer exposure periods may be necessary as these islet proteins are much lower in abundance compared to insulin. However, the exposure period should not exceed the half life of  $^{35}\text{S}$ , which is 87 days.
21. The biosynthesis of proinsulin, chromogranin A, and PC1 was stimulated 10-30 fold at the higher glucose concentration, consistent with previous studies [4, 7, 10].

## References

1. Shanmugam G, Vecchio G, Attardi D, Green M (1972) Immunological studies on viral polypeptide synthesis in cells replicating murine sarcoma-leukemia virus. *J Virol* 10:447–455
2. Hutton JC (1994) Insulin secretory granule biogenesis and the proinsulin-processing endopeptidases. *Diabetologia* 37(Suppl 2):S48–S56
3. Houwen B, Goudeau A, Dankert J (1975) Isolation of hepatitis B surface antigen (HBsAg) by affinity chromatography on antibody-coated immunoabsorbents. *J Immunol Methods* 8:185–194
4. Guest PC, Rhodes CJ, Hutton JC (1989) Regulation of the biosynthesis of insulin secre-

- tory granule proteins: co-ordinate translational control is exerted on some, but not all, granule matrix constituents. *Biochem J* 257:432–437
5. Hutton JC, Peshavaria M, Johnston CF, Ravazzola M, Orci L (1988) Immunolocalization of betagranin: a chromogranin A-related protein of the pancreatic B-cell. *Endocrinology* 122:1014–1020
  6. Bennett DL, Bailyes EM, Nielsen E, Guest PC, Rutherford NG, Arden SD et al (1992) Identification of the type 2 proinsulin processing endopeptidase as PC2, a member of the eukaryote subtilisin family. *J Biol Chem* 267:15229–15236
  7. Guest PC, Abdel-Halim SM, Gross DJ, Clark A, Poitout V, Amaria R et al (2002) Proinsulin processing in the diabetic Goto-Kakizaki rat. *J Endocrinol* 175:637–647
  8. Davidson HW, Rhodes CJ, Hutton JC (1988) Intraorganellar calcium and pH control proinsulin cleavage in the pancreatic beta cell via two distinct site-specific endopeptidases. *Nature* 333:93–96
  9. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685
  10. Guest PC, Arden SD, Bennett DL, Clark A, Rutherford NG, Hutton JC (1992) The post-translational processing and intracellular sorting of PC2 in the islets of Langerhans. *J Biol Chem* 267:22401–22406

# Chapter 15

## Two Dimensional Gel Electrophoresis of Insulin Secretory Granule Proteins from Biosynthetically-Labeled Pancreatic Islets

Paul C. Guest

### Abstract

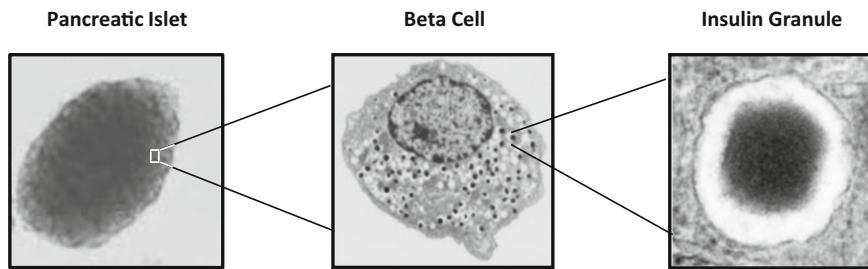
Pulse-chase radiolabeling of cells with radioactive amino acids is a common method for tracking the biosynthesis of proteins. Radiolabeled newly synthesized proteins can be analyzed by a number of techniques such as two dimensional gel electrophoresis (2DE). This chapter presents a protocol for the biosynthetic labeling of pancreatic islets with  $^{35}\text{S}$ -methionine in the presence of basal and stimulatory concentrations of glucose, followed by subcellular fractionation to produce a secretory granule fraction and analysis of the granule protein contents by 2DE. This provides a means of determining whether or not the biosynthetic rates of the entire granule constituents are coordinately regulated.

**Key words** Pulse-chase radiolabeling, Pancreatic islets, Subcellular fractionation, Insulin secretory granules, 2D electrophoresis

---

### 1 Introduction

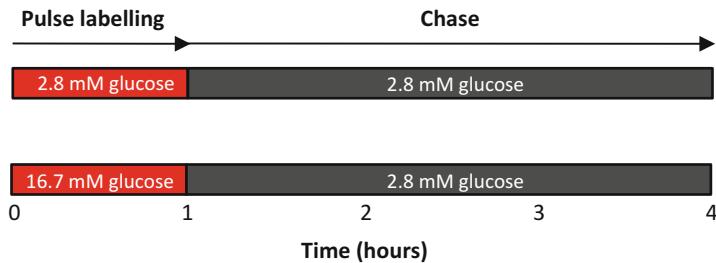
The secretory granules of pancreatic B cells perform a specialized function in the packaging, storage, and secretion of the hormone insulin (Fig. 1) [1]. In addition to insulin and the connecting (C-) peptide, these subcellular organelles contain more than 150 polypeptides [2]. This includes the proteases involved in proinsulin-to-insulin conversion, proinsulin conversion intermediates, other polypeptide precursor proteins, minor cosecreted peptides, membrane proteins involved in cell trafficking, and ion translocating proteins involved in regulation of the intragranular environment. Most of these proteins are likely to be synthesized, transported, and packaged into nascent granules in a coordinated manner to ensure correct functioning of the granule. Insulin biosynthesis is regulated by many circulating nutrients and other factors although glucose is the most important physiologically [3]. However, it is



**Fig. 1** Images of (a) pancreatic islet as seen through a light microscope, (b) pancreatic beta cell visualized by electron microscopy, and (c) insulin secretory granule obtained by high power electron microscopy. Each islet contains around 5000 cells of which 70–80 % are comprised of the insulin-producing pancreatic beta cells. The granules contain a dense core of insulin hexamers which comprise around 80 % of the protein mass and this is surrounded by less dense material containing approximately 150 different granule accessory proteins [1]

not known whether just some or all of the granule constituents are affected in a similar manner.

This chapter addresses this question by two dimensional gel electrophoresis (2DE) [4] of secretory granule subcellular fractions prepared from  $^{35}\text{S}$ -methionine-labeled rat islets, as described by Guest and coworkers [5]. The protocol employed is a 1 h pulse labeling of pancreatic islets with  $^{35}\text{S}$ -methionine in the presence of either low or high glucose concentrations, followed by a chase period of 3 h in nonradioactive medium containing a low glucose concentration (Fig. 2). The production of mature insulin requires cleavage of proinsulin by the endoproteases prohormone convertase 1 (PC1) and prohormone convertase 2 (PC2) on the carboxy-terminal side of Arg<sup>31</sup>-Arg32 and Lys<sup>64</sup>-Arg<sup>65</sup>, respectively, followed by the removal of the exposed basic residues by the exopeptidase carboxypeptidase H [6, 7]. This conversion is optimal in the low pH and high  $\text{Ca}^{2+}$  environment in the late trans Golgi network and secretory granule compartments [1]. This fits with finding that final conversion of proinsulin to insulin does not begin to occur until approximately 30 min after initial synthesis on the rough endoplasmic reticulum and transport to these compartments. Thus, the 1 h pulse labeling and 3 h chase employed here ensures that insulin and most of the other secretory proteins have had sufficient time to reach the granule compartment. In addition, the chase under low glucose conditions minimizes secretion of these newly synthesized proteins, thereby ensuring that they are retained within the granules. Here, the pulse chase radiolabeling of islets, subcellular fractionation, and 2DE analyses are presented.



**Fig. 2** Schematic diagram showing the  $^{35}\text{S}$ -methionine pulse chase labeling protocol

## 2 Materials

1. 400 rat pancreatic islets per condition (*see Note 1*).
2. Post-nuclear fraction from 0.5 g of rat insulinoma tissue (*see Note 2*).
3. High glucose modified Kreb's bicarbonate buffer: 25 mM  $\text{NaHCO}_3$  (pH 7.4), 115 mM NaCl, 5.9 mM KCl, 1.2 mM  $\text{MgCl}_2$ , 1.2 mM  $\text{NaH}_2\text{PO}_4$ , 1.2 mM  $\text{Na}_2\text{SO}_4$ , 2.5 mM  $\text{CaCl}_2$ , 16.7 mM glucose, and 0.1% bovine serum albumin.
4. Low glucose modified Kreb's bicarbonate buffer: 25 mM  $\text{NaHCO}_3$  (pH 7.4), 115 mM NaCl, 5.9 mM KCl, 1.2 mM  $\text{MgCl}_2$ , 1.2 mM  $\text{NaH}_2\text{PO}_4$ , 1.2 mM  $\text{Na}_2\text{SO}_4$ , 2.5 mM  $\text{CaCl}_2$ , 2.8 mM glucose, and 0.1% bovine serum albumin.
5. Low glucose chase incubation buffer: Dulbecco's modified Eagle's medium, containing 10% newborn calf serum and 2.8 mM-glucose.
6.  $^{35}\text{S}$ -methionine (*see Note 3*).
7. Islet homogenization medium: 10 mM KMes (pH 6.5) containing 0.3 M sucrose, 1 mM  $\text{MgSO}_4$ , and 1 mM EGTA.
8. 4.4, 8.8, and 17.7% Nycodenz (Nyegaard Diagnostica; Oslo, Norway) in homogenization medium.
9. Subcellular fractionation wash buffer: 10 mM KMes (pH 6.5) containing 0.25 M sucrose.
10. Isoelectric focussing buffer: 9.5 M urea, 5% 2-mercaptoethanol, 0.40% pH 3–10 range Pharmalytes (Pharmacia Fine Chemicals; Uppsala, Sweden), 1.6% pH 5–7 range Ampholines (Pharmacia LKB Biotechnology, Bromma, Sweden), and 0.001% Bromophenol Blue.
11. Isoelectric focussing tube gel buffer: 8 M urea, 20% pH 3–10 range Pharmalyte and 1% each of pH 3.5–5.0, 5–7, and 7–9 range Ampholines.
12. Isoelectric focussing lower tank buffer: 0.085% phosphoric acid.

13. Isoelectric focussing upper tank buffer: 0.2 N NaOH.
14. Second dimension equilibration buffer: 50 mM Tris–Cl, pH 6.8, 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS, 0.01% Bromophenol Blue.
15. Second dimension tank buffer: 25 mM Tris/192 mM glycine (pH 8.3) and 0.1% SDS.
16. Fluorography solution: 20% 2,5-diphenyloxazole in acetic acid.
17. Nunc Cryotubes (Gibco; Paisley, Scotland, UK).
18. 1 mL capacity glass tube homogenizer (*see Note 4*).
19. 1.2 cm × 5.0 cm polypropylene centrifuge tubes (Beckman Instruments; Palo Alto, CA, USA) (*see Note 5*).
20. Swinging bucket rotor (*see Note 6*).
21. MSE Sonifier (Crawley, UK) with a microprobe (*see Note 7*).
22. Vertical isoelectric focussing tube gel system using 15 cm × 0.15 cm inner diameter glass tubes (*see Note 8*).
23. Second dimension gels cast in 15 cm × 15 cm × 0.15 cm glass plates: linear 5–20% acrylamide gradient, containing 0.065% N,N'-methylenebisacrylamide, 0.375 M Tris–HCl pH 8.8, and 0.2% SDS (*see Note 9*).
24. Second dimension electrophoresis tank for running gels cast in 15 cm × 15 cm × 0.15 cm glass plates (*see Note 10*).

### 3 Methods

#### **3.1 Biosynthetic Radiolabeling of Pancreatic Islets**

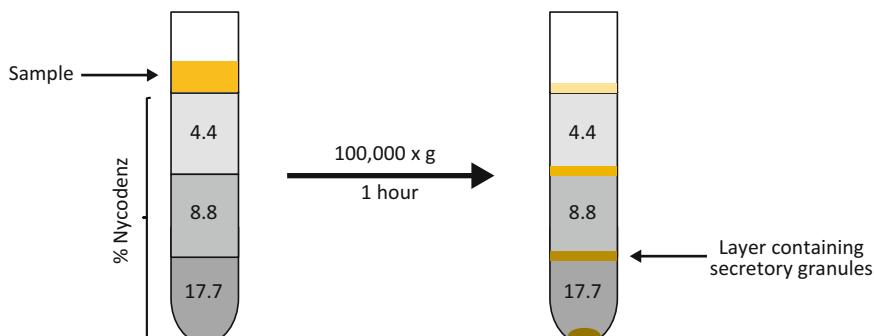
1. Preincubate 400 isolated islets for 60 min in 500 µL of modified Kreb's bicarbonate buffer containing 2.6 or 16.7 mM glucose at 37 °C in Cryotubes under 95% O<sub>2</sub>/5% CO<sub>2</sub>.
2. Recover the islets by centrifugation at 100×*g* for 10 s and suspend in 200 µL of the same pre-warmed medium containing 200 µCi of <sup>35</sup>S-methionine and incubate for 1 h at 37 °C in Cryotubes under 95% O<sub>2</sub>/5% CO<sub>2</sub> (Fig. 2) (*see Note 11*).
3. Recover the islets by centrifugation at 100×*g* for 10 s, carefully remove the radioactive supernatant, and gently suspend the islets in 500 µL of chase buffer and incubate for 3 h under 95% O<sub>2</sub>/5% CO<sub>2</sub> (*see Note 12*).
4. Terminate the incubations by the addition of 1 mL of ice-cold Kreb's low glucose incubation medium followed by centrifugation for 10 s at 3300×*g* in a swinging bucket rotor and discard the media.
5. Wash the islet pellets by two further cycles of resuspension and centrifugation as above and subject immediately to subcellular fraction (*see Note 13*).

### 3.2 Subcellular Fractionation

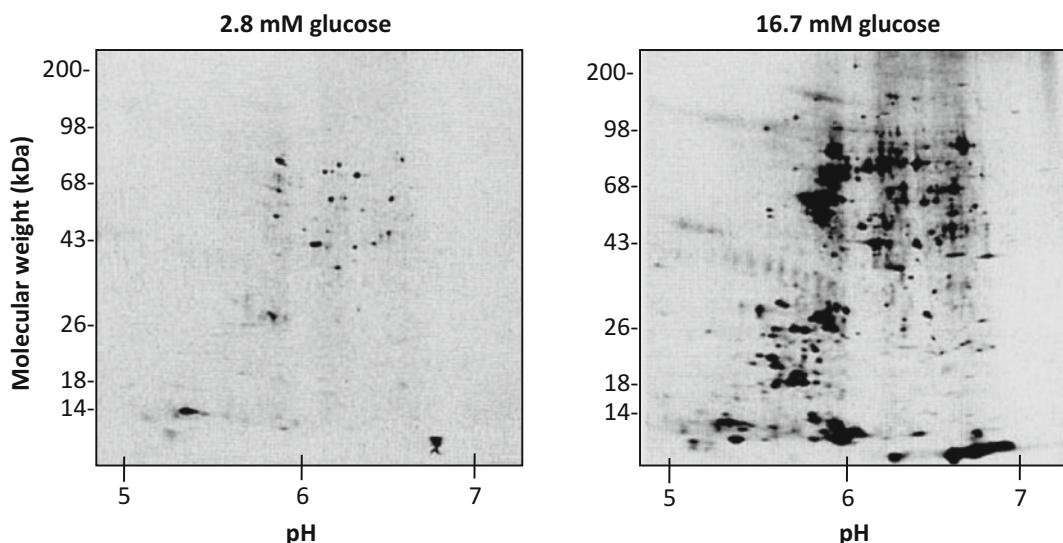
1. Combine radiolabeled slets with the insulinoma cells and homogenize in the glass tube homogenizer using ten strokes of a Teflon pestle driven at 600 rpm at 4 °C (*see Note 14*).
2. Centrifuge the homogenates for 5 min at  $1700 \times g$  at 4 °C to remove unbroken cells and nuclei.
3. Transfer the supernatants to the  $1.2 \times 5$  cm centrifuge tubes and add 1.3 mL portions of each of the 4.4, 8.8, and 17.7% NycoDenz solutions (in that order), loading from the bottom (Fig. 3) (*see Note 15*).
4. Centrifuge the gradients for 1 h at  $100,000 \times g$  in the swinging bucket rotor.
5. Collect the material at the 8.8/17.7% NycoDenz interface and suspend this in subcellular fractionation wash buffer (*see Note 16*).
6. Centrifuge for 20 min at  $50,000 \times g$  in the swinging bucket rotor.
7. Resuspend the particulate material containing enriched secretory granules in the subcellular fractionation wash buffer, centrifuge again, and store the final pellets at –80 °C prior to 2DE analysis.

### 3.3 Two Dimensional Gel Electrophoresis of $^{35}\text{S}$ -Labeled Secretory Granule Proteins

1. Homogenize the secretory granule pellets by sonication in 100  $\mu\text{L}$  of isoelectric focussing buffer.
2. Centrifuge at  $13,000 \times g$  to remove the particulate material.
3. Subject the samples to 2DE analysis by isoelectric focusing in the first dimension and SDS polyacrylamide gel electrophoresis in the second dimension (*see Note 17*).
4. For isoelectric focussing, prefocus the gels in the assembled tube gel apparatus for 1 h at 200 V.
5. Add up to 50  $\mu\text{L}$  of sample to the top of each tube using a Hamilton syringe and carry out isoelectric focussing at 800 V for approximately 6 h until the Bromophenol Blue has reached the bottom of the tube (*see Note 18*).



**Fig. 3** Schematic diagram showing subcellular fractionation protocol



**Fig. 4** Islets were labeled with  $^{35}\text{S}$ - methionine for 1 h in either 2.8 mM glucose or 16.7 mM-glucose and then incubated for a further 3 h in nonradioactive medium containing 2.8 mM glucose. Subcellular fractions were prepared and those enriched in secretory granules were subjected to 2DE followed by fluorography (28 days exposure). The image shows that the biosynthesis of the most of the proteins was stimulated 10–30 fold at the higher glucose concentration. Previous pulse chase labeling and immunoprecipitation studies have shown that this is known to occur for the secretory granule proteins insulin [8], chromogranin A [8], secretory granule membrane protein 110 [9], and PC1 [10]

6. Using a syringe of the appropriate diameter, apply pressure to the top of the glass tubes to extrude the gels into the second dimension equilibration buffer (*see Note 19*).
7. Incubate the gels for 5 min in the equilibration buffer and then load on to the top of the second dimension gel using a 0.15 cm thick spacer to nudge the gels into position (*see Note 20*).
8. Carry out electrophoresis at 60 V for 1 h followed by 120 V until the dye front reaches the bottom of the gel.
9. Disassemble the gel plates and immerse the gels for 2 × 5 min in acetic acid, 2 h in fluorography solution, and then leave for 30 min under cold running water (*see Note 21*).
10. Vacuum dry and expose the gel to Cronex 4 X-ray film (Dupont; Stevenage, Herts, UK) for 2–8 days (Fig. 4) (*see Note 22*).

#### 4 Notes

1. Rat islets were obtained from 10 to 12-week-old New England Deaconess Hospital rats by a collagenase digestion technique, as described by Guest et al [8]. However, other protocols can be used, providing that these can yield the required number of large intact islets.

2. This was prepared as described by Hutton et al [2] for combination with radiolabeled rat islets prior to homogenization and density gradient centrifugation to facilitate efficient recovery. Other rat beta cell lines can be used as a substitute.
3. Use appropriate precautions when handling and disposing of radioactive materials.
4. We used a 1 mL capacity glass tube homogenizer and Teflon pestle from Jencons Scientific (Leighton Buzzard, Beds, UK) although similar products available from other suppliers would work just as well.
5. Other tube and swinging bucket rotor combinations can be used but ensure these have appropriate specifications.
6. We used a Beckman SW 50.1 rotor. Other swinging bucket rotors can be used but ensure that these are an adequate match for the centrifuge tubes.
7. Any sonication probe device can be used but the probe should fit inside a 1.5 mL-capacity microcentrifuge tube and therefore be not more than 2 mm in diameter at the tip.
8. Although we used a homemade device for this, tube gel systems are available from many suppliers. In addition, it is likely that the instruments designed for use of immobilized pH gradient strips could also be employed. Keep in mind to designate the chosen device for use with radioactive materials.
9. The gels used for this study were cast in the lab. However, precast 5–20% gradient gels are also available from many suppliers.
10. Again, many systems could be used for this although the size should be compatible with the isoelectric focussing stage in terms of tube gel or strip length.
11. Use recommended precautions and dispose of radioactive material appropriately.
12. The addition of the chase medium containing nonradioactive amino acids stops the uptake of  $^{35}\text{S}$ -methionine and halts metabolic activity of the islet cells.
13. Do not freeze the islets as this will disrupt cellular and intracellular membranes, which will disrupt the subcellular fractionation step.
14. The clearance of the glass tube and pestles allows for disruption of the cell membranes but leaves most intracellular organelles, such as the secretory granules, intact.
15. The under-layering approach helps to form sharp boundaries between layers, which may be more difficult to achieve loading using the over-layering method.

16. All fractions and the pellet should be collected for further analyses but we are only presenting the results using the layer most enriched in secretor granules [5].
17. We used the method as described by Anderson et al [11].
18. This is a continuous isoelectric focussing system as it does not use an immobilized pH gradient. Care should be taken not to run samples into the lower tank buffer. As a useful guide, Bromophenol Blue will turn green at pH 4 and yellow at pH 3 and the run should be timed to terminate when a green-yellow band reaches the bottom of the tube gel. However, it is best to carry out a time course study to determine optimum running time when you are running a new kind of sample.
19. A 200 µL pipette tip works well for this.
20. Be careful not to damage the tube gel and ensure that there are no air bubbles between the tube gel and the second dimension gel.
21. The gel turns white during this final step.
22. Other films can be used. Please check the manufacturer's specifications. Obtaining the best exposure may require multiple attempts and adjusting the times accordingly.

## References

1. Hutton JC (1994) Insulin secretory granule biogenesis and the proinsulin-processing endopeptidases. *Diabetologia* 37(Suppl2):S48–S56
2. Hutton JC, Penn EJ, Peshavaria M (1982) Isolation and characterisation of insulin secretory granules from a rat islet cell tumour. *Diabetologia* 23:365–373
3. Hedeskov CJ (1980) Mechanism of glucose-induced insulin secretion. *Physiol Rev* 60:42–509
4. O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250:4007–4021
5. Guest PC, Bailyes EM, Rutherford NG, Hutton JC (1991) Insulin secretory granule biogenesis. Co-ordinate regulation of the biosynthesis of the majority of constituent proteins. *Biochem J* 274:73–78
6. Davidson HW, Rhodes CJ, Hutton JC (1988) Intraorganellar calcium and pH control proinsulin cleavage in the pancreatic beta cell via two distinct site-specific endopeptidases. *Nature* 333:93–96
7. Bennett DL, Bailyes EM, Nielsen E, Guest PC, Rutherford NG, Arden SD et al (1992) Identification of the type 2 proinsulin processing endopeptidase as PC2, a member of the eukaryote subtilisin family. *J Biol Chem* 267:15229–15336
8. Guest PC, Rhodes CJ, Hutton JC (1989) Regulation of the biosynthesis of insulin secretory granule proteins: co-ordinate translational control is exerted on some, but not all, granule matrix constituents. *Biochem J* 257:432–437
9. Hutton JC, Bailyes EM, Rhodes CJ, Rutherford NG, Arden SD, Guest PC (1990) Biosynthesis and storage of insulin. *Biochem Soc Trans* 18:122–124
10. Guest PC, Abdel-Halim SM, Gross DJ, Clark A, Poitout V, Amaria R et al (2002) Proinsulin processing in the diabetic Goto-Kakizaki rat. *J Endocrinol* 175:637–647
11. Anderson NG, Anderson NL, Tollaksen SL (1979) Operation of the isodalt system, publication ANL-BIM-79-2, division of biological and medical research. Argonne National Laboratory, Argonne, IL, USA

# Chapter 16

## Depletion of Highly Abundant Proteins of the Human Blood Plasma: Applications in Proteomics Studies of Psychiatric Disorders

**Sheila Garcia, Paulo A. Baldasso, Paul C. Guest,  
and Daniel Martins-de-Souza**

### Abstract

Psychiatric disorders are complex diseases involving exogenous and endogenous factors. Biomarkers for diagnosis or prediction of successful treatment are not existent. In addition, the molecular basis of these diseases is still poorly understood. Blood plasma represents the most complex proteome as it contains subproteomes from several body tissues. However, the high abundance of some little proteins can obscure the analysis of hundreds of low abundance proteins, which are potential biomarkers. Therefore, removal of these high abundance proteins is pivotal in any proteomic study of plasma. Here, we present a method of depleting these proteins using immunoaffinity liquid chromatography.

**Key words** Plasma, Protein depletion, Immunoaffinity chromatography, Neuropsychiatric disorders, Plasma biomarker, Neuroproteomics

---

### 1 Introduction

Psychiatric disorders are heterogeneous diseases involving genetic and environmental components. Due to this complexity, it is difficult to understand the molecular basis of these diseases and, choose an adequate treatment [1]. The misdiagnosis rate is high among patients with these disorders, being as high as 40% for schizophrenia and 69% for bipolar disorder [2, 3]. This occurs due to the poor understanding of the biochemical pathways involved in these diseases [4, 5]. In addition, the number of non-responders to treatment is high. In schizophrenia, for example, the proportion of patients who fail to respond adequately is 25–40% and the treatment dropout rate is around 33% [6, 7].

Another complicating factor in the majority of psychiatric disorders is that the choice of drug used for treatment is random and based on patient clinical history and/or the choice of the physician [8]. In addition, treatment can sometimes lead to strong

side effects. Therefore, it is essential to find biomarkers that improve treatment outcomes [9] using blood proteome profiling methods such as mass spectrometry [10].

Research has shown that some proteins present in blood of individuals with schizophrenia are altered in abundance compared to healthy individuals. Some examples are the cytokines and interleukins associated with the inflammation response and circulating hormones such as insulin, cortisol, and follicle-stimulating hormone (FSH) that can affect brain function [11, 12]. This suggests that although psychiatric disorders appear in the brain, the effects can be observed in other parts of the body including the peripheral blood system [5, 13, 14].

Body fluids have been used in molecular biology studies of psychiatric disorders for more than a decade. Serum and plasma are easier and more accessible to work with than solid tissues such as the brain and their use allows the follow-up of patients during the disease course or following treatment [15]. However, most of the blood proteins are present only at low levels, suggesting the need for more sensitive technologies [15]. Approximately 99% of blood proteins are represented by 30 highly abundant proteins, with concentrations spanning at least 12 orders of magnitude. Human serum albumin (HSA) and immunoglobulin G (IgG) are among the highly abundant proteins and together comprise 70% of the total blood protein mass [14, 16]. This is a problem as the current dynamic range of most mass spectrometry instruments is lower than five orders of magnitude. Furthermore, many of the highly abundant proteins obscure those of lower abundance. As a consequence, the direct analysis of crude plasma or serum is not possible until these high abundance proteins are depleted [17]. Thus, methods of fractionation, such as immunoaffinity chromatography, are often used to deplete the levels of these abundant proteins prior to analysis [16, 18].

Here, we present a protocol using the commercially available Multiple Affinity Removal System (MARS®; Agilent Technologies; Santa Clara, CA, USA) that can be used in conjunction with most high-performance liquid chromatography (HPLC) systems. The column matrix contains polyclonal antibodies that target the 14 most abundant proteins in plasma or serum. This method can be applied prior to most proteomic investigations of serum or plasma to enrich the low abundance proteins.

---

## 2 Materials

1. Human plasma or serum samples.
2. Buffer solution A: 0.2 M sodium phosphate pH 7.4, 0.5 M sodium chloride, 0.02% sodium azide, store for up to 30 days after preparation.

3. Buffer solution B: 2 M Urea and 0.5 M glycine (pH 2.25), store for up to 30 days after preparation.
4. 50 mM ammonium bicarbonate.
5. HPLC system (we use a Waters 2487 Dual  $\lambda$  Absorbance Detector) and a manual injection system using a Hamilton syringe.
6. Agilent MARS Human 14 column (4.6 mm inner diameter, length 100 mm).
7. Concentration centrifuge tubes with 3000 Da molecular weight cutoff and 6 mL capacity.
8. Microcentrifuge and large centrifuge rotor for up to 15 mL capacity centrifuge tubes.
9. Eppendorf LoBind tubes 0.5 mL and other 1.5 mL capacity microcentrifuge tubes as needed.
10. Protein assay quantitation equipment and reagents.
11. SDS-PAGE gel equipment.
12. Gel reagents: acrylamide (30 %), sodium dodecyl sulfate (SDS), dithiothreitol (DTT), Tris-Base, chlorine hydroxide (HCl), bromophenol blue, glycine, commasie blue dye, methanol, ammonium persulfate, tetramethylenediamine (TEMED).
13. Sample buffer: 10 % sodium dodecyl sulfate (SDS), 10 mM dithiothreitol (DTT), 20 % glycerol, 0.2 M Tris-HCl (pH 6.8), and 0.05 % bromophenol blue.
14. Stock solution of running buffer (10 $\times$ ): 250 mM Tris-HCl, 2 M glycine, and 10 % of SDS.

---

### 3 Methods (See Note 1)

1. Dilute 30  $\mu$ L of plasma or serum four times with buffer solution A in a 1.5 mL tube (*see Note 2*).
2. Centrifuge at 21,000  $\times g$  for 15 min, then recover the supernatant without disrupting the pellet (*see Note 3*).
3. Transfer the supernatant to a new 0.5 mL LoBind tube and proceed with **step 9** or keep the sample on ice if performing the experiment on the same day (*see Note 4*).
4. With the pumps connected in a bottle containing water, purge HPLC lines at a flow rate of 1 mL/min for 5 min each (*see Note 5*).
5. With the pumps connected in buffers solution A and B, purge HPLC lines with buffer solution A at a flow rate of 1 mL/min for 5 min each.
6. Connect the MARS human 14 column in the HPLC system and make sure to drip test when installing each extremity of the column to avoid air bubbles in the system (*see Note 6*).

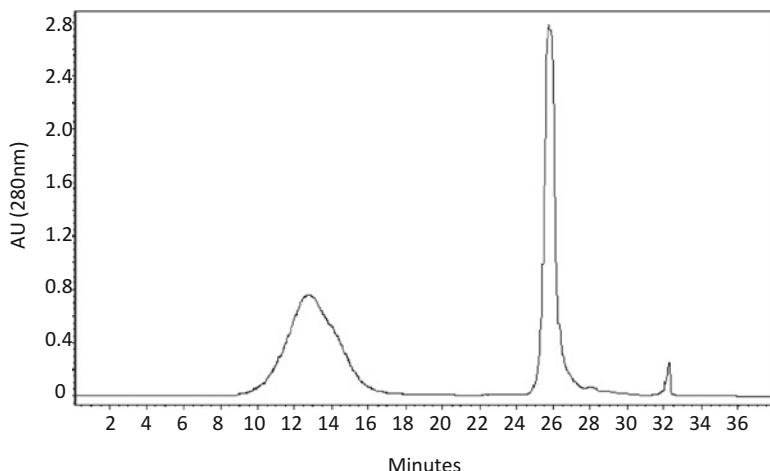
**Table 1**  
**LC wash and equilibration conditions**

| Time  | % Buffer B | Flow rate (mL/min) | Max pressure | Curve |
|-------|------------|--------------------|--------------|-------|
| 0.00  | 100        | 1.00               | 60           | 6     |
| 15.00 | 100        | 1.00               | 60           | 6     |
| 16.00 | 0          | 1.00               | 60           | 6     |
| 31.00 | 0          | 1.00               | 60           | 6     |
| 32.00 | 0          | 0.2                | 60           | 6     |

**Table 2**  
**LC method**

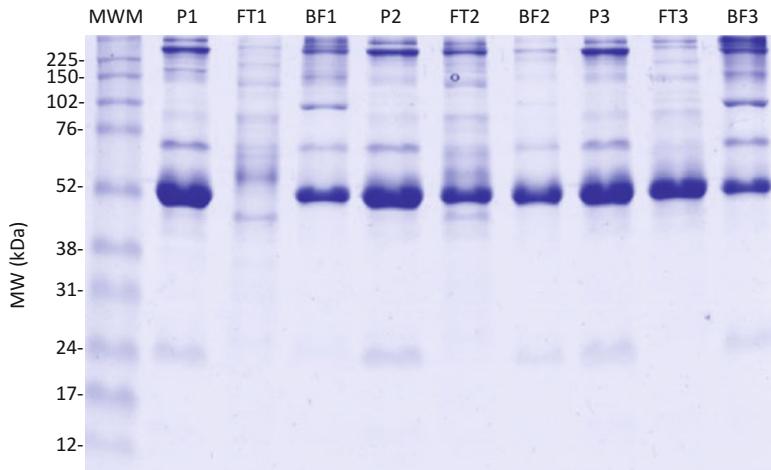
| Phase                          | Time  | % Buffer B | Flow rate (mL/min) | Max pressure | Curve |
|--------------------------------|-------|------------|--------------------|--------------|-------|
| Low abundant fraction elution  | 0.00  | 0          | 0.125              | 60           | 6     |
|                                | 18.00 | 0          | 0.125              | 60           | 6     |
| Washing                        | 18.01 | 0          | 1.0                | 60           | 6     |
|                                | 20.00 | 0          | 1.0                | 60           | 6     |
| High abundant fraction elution | 20.01 | 100        | 1.0                | 60           | 6     |
|                                | 27.00 | 100        | 1.0                | 60           | 6     |
| Column regeneration            | 27.01 | 0          | 1.0                | 60           | 6     |
|                                | 38.00 | 0          | 1.0                | 60           | 6     |

7. Set up the LC method and run a wash method to equilibrate the column (Table 1) (*see Note 7*).
8. Inject ~120 µL of prepared sample into HPLC sample loop (*see Note 8*) and set up the LC method below as shown in Table 2.
9. Set up to initiate the depletion run method (*see Note 9*).
10. Collect the fractions containing low abundant proteins that should elute within the first 2.25 mL/18 min of the method and then collect the fractions containing high abundant proteins that should elute between 7.25 and 11.25 mL/23–27 min (the high abundance proteins will bind to the antibodies on the column matrix) (Fig. 1) (*see Note 10*).
11. Collect low and high abundance protein fractions for buffer exchange into sodium bicarbonate and storage at –80 °C (*see Note 11*).
12. Wash the concentration centrifuge tubes using a wash bottle containing MilliQ water.



**Fig. 1** Representative chromatogram of depletion runs on the MARS Hu14 column. The first peak contains the low-abundance proteins (flowthrough fraction), the second peak contains the high abundance proteins (bound fraction), and the last peak is due to the buffer solution

13. Add 3 mL of 50 mM of ammonium bicarbonate buffer solution (natural pH 7.5) and centrifuge at  $6600 \times g$  for 40 min or until remaining volume is 300–600  $\mu$ L and discard the remaining volume.
14. Add the low abundance protein fractions (~1.5 mL) to these tubes. (*see Note 12*).
15. Centrifuge at  $6600 \times g$  for approximately 30 min or until the remaining volume is 300–600  $\mu$ L.
16. Complete the volume of tubes to 6 mL and pipette up and down to wash proteins off the filter membrane.
17. Centrifuge at  $6600 \times g$  for 90 min or until remaining volume is 300–600  $\mu$ L.
18. Pipette up and down several times to resuspend proteins and transfer to new 1.5 mL tubes.
19. Measure protein concentrations in duplicates using a protein assay.
20. Digest the proteins in each sample using a protease such as trypsin and/or store at –80 °C.
21. After the last run, wash the column with buffer A in isocratic mode at 1 mL/min over 20 min.
22. Annotate the MARS column usage in a logbook to check column performance in accordance with the manufacturer's guarantee of 200 runs (*see Note 13*).
23. Flush the HPLC system with water at 1 mL/min over 30 min.



**Fig. 2** SDS-PAGE of human plasma protein fractions from the MARS Hu14 column. An equal amount (15 µg) of crude plasma (P), flow-through (FT), and bound fraction (BF) were separated on 12 % SDS-polyacrylamide gel. The proteins were stained with Comassie blue R-250 dye. *MWM*: molecular weight markers

**Table 3**  
Standard concentrations of acrylamide used for resolving proteins of specific molecular weights

| Acrylamide (%) | M.W. Range (kDa) |
|----------------|------------------|
| 7              | 50–500           |
| 10             | 20–300           |
| 12             | 10–200           |
| 15             | 3–100            |

24. To use the MARS Hu14 column more than 200 runs, it is important to check column efficiency in each set of sample runs.
25. Prepare or purchase a 12 % sodium dodecylsulfate (SDS) polyacrylamide gel, load 10–20 µg protein from each fraction, carry out electrophoresis, and stain proteins with Coomassie Blue R-250 (Fig. 2; Table 3).
26. If preparing the gel, make 5 mL of stacking gel and 10 mL of running gel according to Tables 4 and 5, respectively.
27. Pipette the running gel solution between glass plates and wait approximately 30 min for polymerization and then repeat this procedure for the stacking gel, which is layered on the top (*see Note 14*).

**Table 4**  
**Composition of the stacking gel**

| Component                                | Volume (mL) |
|--|-------------|
| 0.5 M Tris-HCl, pH 6.8                   | 1.25        |
| 10% (w/v) SDS                            | 0.05        |
| Acrylamide/bis-acrylamide (30%/0.8% w/v) | 0.67        |
| H <sub>2</sub> O                         | 2.975       |
| Just before pouring the gels, add:       |             |
| 10% (w/v) ammonium persulfate (APS)      | 0.05        |
| TEMED                                    | 0.005       |

**Table 5**  
**Composition of the running gel**

| Acrylamide %                             | 8       | 10      | 12      | 15      |
|--|---------|---------|---------|---------|
| 1.5 M Tris-HCl, pH 8.8                   | 2.6 mL  | 2.6 mL  | 2.6 mL  | 2.6 mL  |
| 10% (w/v) SDS                            | 0.1 mL  | 0.1 mL  | 0.1 mL  | 0.1 mL  |
| Acrylamide/Bis-acrylamide (30%/0.8% w/v) | 2.6 mL  | 3.4 mL  | 4 mL    | 5 mL    |
| H <sub>2</sub> O                         | 4.6 mL  | 3.8 mL  | 3.2 mL  | 2.2 mL  |
| Just before pouring the gels, add:       |         |         |         |         |
| 10% (w/v) ammonium persulfate (APS)      | 0.1 mL  | 0.1 mL  | 0.1 mL  | 0.1 mL  |
| TEMED                                    | 0.01 mL | 0.01 mL | 0.01 mL | 0.01 mL |

28. Insert the well comb without introducing air bubbles and leave in place until the stacking gel has set completely (*see Note 15*).
29. Mix the sample with sample buffer.
30. Heat sample for 5–10 min at 95 °C to denature the proteins.
31. Load a protein molecular weight marker into the first well and the prepared samples into the others, and load empty channels with sample buffer using the same volume (*see Note 16*).
32. Load the gel platform with running buffer.
33. Set a voltage or current to run the electrophoresis.
34. Stop the running gel when the final band of protein marker has reached the bottom of glass plates.
35. Remove the gel from the plates and stain proteins using Coomassie Blue R-250 dye for at least 2 h.

36. Discard the dye and destain the gel with 10–20% methanol or bleach solution for approximately 2 h, changing this solution approximately three times (*see Note 17*).
37. Soak the gel briefly in water prior to imaging (*see Note 18*).

---

#### 4 Notes

1. All steps should be carried out at 4 °C except the HPLC method that should be performed at room temperature.
2. In applying 40 µL of crude plasma, we observed the presence of HSA in the flow through fraction of one sample. Thus, we suggest using lower amounts (such as 30 µL) to avoid this.
3. We use a step of centrifugation but recommend using spin cartridges to avoid introducing particulate matter on the column.
4. Turn on the HPLC system (pumps, controller, recorder, and collector). Before starting a run, place the LoBind tubes in the collector and set up a method according to manufacturer's protocol.
5. The buffer solutions do not need degassing. If the HPLC is purged with ethanol or isopropanol, it is necessary to wash the HPLC system with water to avoid salt precipitation within the system.
6. We use the MARS 14 column in a medium temperature room (17 °C) but the manufacturer suggests room temperature use (21 °C).
7. This allows the complete elution of the residual matter within the column.
8. We use a loop of 120–500 µL depending on sample size as the loop volume should be at least twice the volume of sample. Before starting each run, wash the sample loop three times the loop volume.
9. The Waters 2487 Dual  $\lambda$  Absorbance Detector HPLC system requires a manual injection with a Hamilton syringe and the sample must be injected within the holder with the sample valve at load position. After injection, leave the syringe in the holder and switch the sample valve to the inject position when starting the run. This step requires care to prevent air bubbles within sample loop. To set up the LC method and start the run we used the software Empower Pro 2 (Waters Corporation® v. 6.0). Before removal of the syringe at the end of a run, turn the sample valve back to the load position. Also, it is important to wash the syringe at least five times with buffer A between each run to avoid cross contamination between samples.
10. Use 1.5 mL LoBind tubes for fraction collection to minimize sample loss. We use 38 LoBind tubes made of nontoxic polypropylene, nuclease-free and pyrogen free (Axygen®).

11. Store the high abundance protein fraction at  $-80^{\circ}\text{C}$  after buffer exchange for potential future analysis. This desalting step can be performed using 6 mL centrifugal concentrator tubes (e.g., Vivastop® 6 Centrifugal Concentrator).
12. The buffer exchange should be done at  $4^{\circ}\text{C}$  to minimize protein degradation.
13. The manufacturer guarantees 200 runs, but we have experience of more than 2000 runs with the same column without loss of efficiency. For this, we test the efficiency of the column by SDS page electrophoresis of fractions used on each set of runs.
14. Use the manufacturer's protocol as appropriate.
15. Before removing the comb, make sure that the gel has polymerized.
16. All wells must be loaded to ensure that the proteins run into the gel with minimal spreading.
17. Observe if the destaining solution is blue and replace it.
18. The gel profile suggests two possible problems. First, there may be a low efficiency of immunodepletion, as seen by the presence of high abundance protein bands in the flow-through fraction of P2 and P3. Second, the profiles of the different samples may be different due to biological variation, as seen by differences in the flow-through fractions. We suggest running each sample twice to help determine which of these possibilities is most likely. Overall, the results show that the MARS Hu14 column is efficient (calculations not shown).

## Acknowledgments

DMS was funded by FAPESP (São Paulo Research Foundation, grants 2013/08711-3, 2013/25702-8 and 2014/10068-4).

## References

1. Nascimento JM, Martins-de-Souza D (2015) The proteome of schizophrenia. *NPJ Schizophr* 1:14003
2. Hirschfeld RMA, Lewis L, Vornik LA (2003) Perceptions and impact of bipolar disorder: How far have we really come? Results of the National Depressive and Manic-Depressive Association 2000 Survey of individuals with bipolar disorder. *J Clin Psychiatry* 64: 161–174
3. Soares-Weiser K, Maayan N, Bergman H, Davenport C, Kirkham AJ, Grabowski S et al (2015) First rank symptoms for schizophrenia (cochrane diagnostic test accuracy review). *Schizophr Bull* 41:792–794
4. Cosgrove VE, Suppes T (2013) Informing DSM-5: biological boundaries between bipolar I disorder, schizoaffective disorder, and schizophrenia. *BMC Med* 11:127
5. Guest PC, Guest FL, Martins-de-Souza D (2015) Making sense of blood-based proteomics and metabolomics in psychiatric research. *Int J Neuropsychopharmacol*. doi:[10.1093/ijnp/pv138](https://doi.org/10.1093/ijnp/pv138)
6. Chan MK, Gottschalk MG, Haenisch F, Tomasik J, Ruland T, Rahmoune H et al

- (2014) Applications of blood-based protein biomarker strategies in the study of psychiatric disorders. *Prog Neurobiol* 122:45–72
7. Regier D (2012) Merging categorical and dimensional diagnoses of mental disorders. *Epidemiol Psychiatr Sci* 21:267–269
  8. Hyman SE (2007) Can neuroscience be integrated into the DSM-V? *Nat Rev Neurosci* 8:725–732
  9. Schwarz E, Guest PC, Rahmoune H, Martins-de-Souza D, Niebuhr DW, Weber NS et al (2012) Identification of a blood-based biological signature in subjects with psychiatric disorders prior to clinical manifestation. *World J Biol Psychiatry* 13:627–632
  10. Martins-De-Souza D (2012) Proteomics tackling schizophrenia as a pathway disorder. *Schizophr Bull* 38:1107–1108
  11. Potvin S, Stip E, Sepehry AA, Gendron A, Bah R, Kouassi E (2008) Inflammatory cytokine alterations in schizophrenia: a systematic quantitative review. *Biol Psychiatry* 63:801–808
  12. Schwarz E, Guest PC, Steiner J, Bogerts B, Bahn S (2012) Identification of blood-based molecular signatures for prediction of response and relapse in schizophrenia patients. *Transl Psychiatry* 2, e82
  13. Miller BJ, Buckley P, Seabolt W, Mellor A, Kirkpatrick B (2011) Meta-analysis of cytokine alterations in schizophrenia: clinical status and antipsychotic effects. *Biol Psychiatry* 70: 663–671
  14. Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 1:845–867
  15. Jaros JA, Martins-de-Souza D, Rahmoune H, Schwarz E, Leweke FM, Guest PC et al (2012) Differential phosphorylation of serum proteins reflecting inflammatory changes in schizophrenia patients. *Eur Arch Psychiatry Clin Neurosci* 262:453–455
  16. Jaros JA, Guest PC, Bahn S, Martins-de-Souza D (2013) Affinity Depletion of Plasma and Serum for Mass Spectrometry-Based Proteome Analysis. *Biomarkers: In Medicine, Drug Discovery, and Environmental Health* 1002, (Humana Press, 2013).
  17. Williamson JC, Edwards AV, Verano-Braga T, Schwämmle V, Kjeldsen F, Jensen ON et al (2016) High-performance hybrid Orbitrap mass spectrometers for quantitative proteome analysis: observations and implications. *Proteomics*. doi: [10.1002/pmic.201400545](https://doi.org/10.1002/pmic.201400545)
  18. Urbas L, Brne P, Gabor B, Barut M, Strlic M, Petric TC et al (2009) Depletion of high-abundance proteins from human plasma using a combination of an affinity and pseudo-affinity column. *J Chromatogr A* 1216:2689–2694

# Chapter 17

## Simultaneous Two-Dimensional Difference Gel Electrophoresis (2D-DIGE) Analysis of Two Distinct Proteomes

Adriano Aquino, Paul C. Guest, and Daniel Martins-de-Souza

### Abstract

This chapter describes the basics, applications, and limitations of two-dimensional gel electrophoresis (2DE) and two-dimensional difference gel electrophoresis (2D-DIGE) for multiplex analysis of distinct proteomes. We also propose a basic protocol for 2D-DIGE, technique that allows the analysis of paired protein extracts, which are labeled with fluorescent Cy3 and Cy5 dyes and electrophoresed with a Cy2-labeled standard extract on the same 2DE gels. Scanning the gels at wavelengths specific for each dye allows direct overlay the two different proteomes and the differences in abundance of specific protein spots can be determined.

**Key words** Proteomics, 2D-DIGE, 2DE, 2D difference gel electrophoresis, CyDyes

---

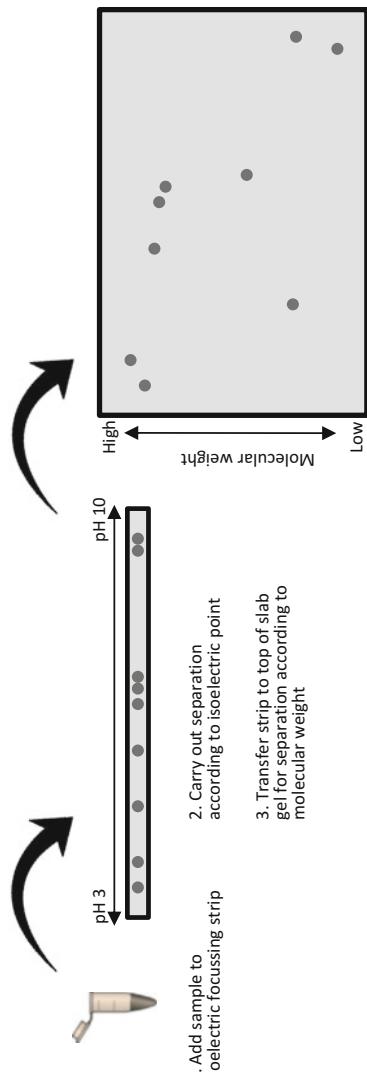
### 1 Introduction

It is likely that Marc Wilkins did not expect the term “proteome” that he coined during his Ph.D. in 1994 would give raise to whole science that we now know as “proteomics” [1]. In reality, proteomics had already been in action for almost 20 years before this after Patrick O’Farrell introduced a technique of protein separation in 1975 called two-dimensional gel electrophoresis (2DE) [2]. During the genome era, 2DE was the most used method for comparative global proteome analyses, combined with mass spectrometry (MS) for protein identification. Since this time, the technique has undergone many optimization steps, including the use of carrier ampholytes, immobilized pH gradients (IPG), and development of IPG acrylamide strips [3, 4]. Perhaps the most significant advancement has been the introduction of sample labeling with fluorophores, which makes 2DE more sensitive, precise, and replicable.

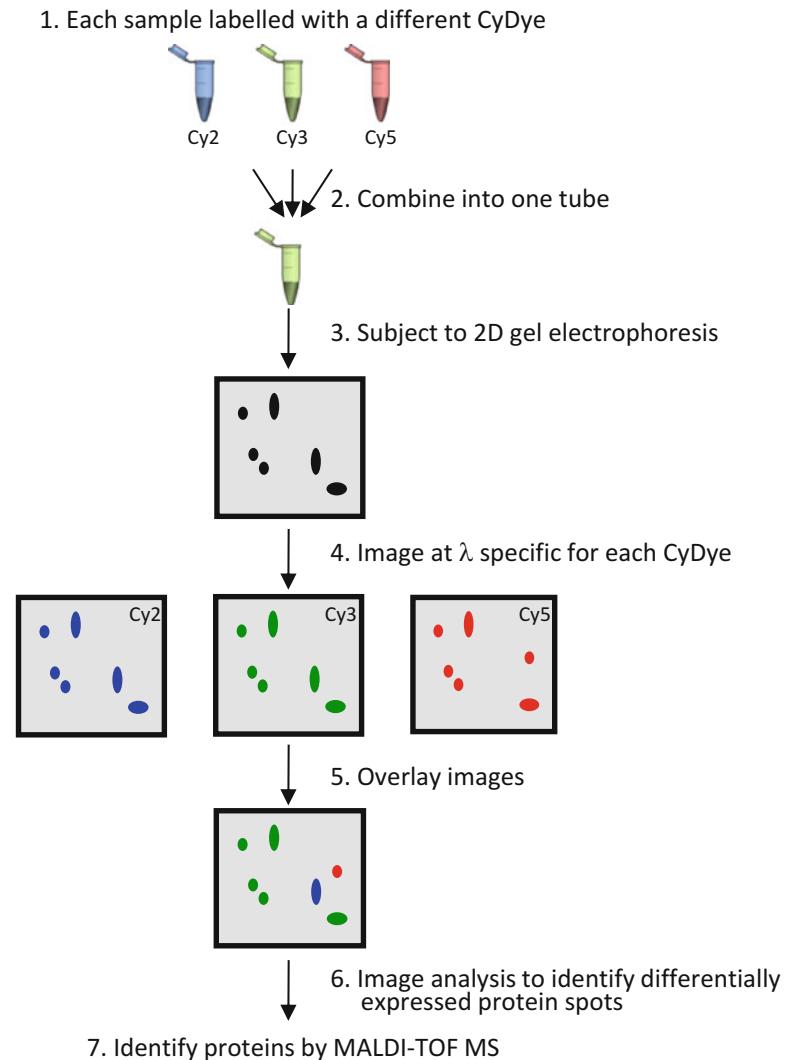
2DE involves separation of proteins in two dimensions according to their isoelectric point (isoelectric focusing—IEF) and apparent molecular size generally by polyacrylamide gel electrophoresis (PAGE) (Fig. 1). This enables the simultaneous separation and display of hundreds of proteins on a single gel. IEF was originally performed in a capillary-gel format, with a pH gradient generated by ampholytes in a gel-matrix. These first systems had low reproducibility due to handling difficulties. As a result, IPGs were developed which are comprised of ampholytes covalently bonded to the gels. The second dimension of the method separates proteins based on their apparent molecular weight in a polyacrylamide gel matrix, in the presence of sodium dodecyl sulfate detergent (SDS). In SDS-PAGE or 1D-PAGE, large amounts of the SDS are necessary to mask the charges of the proteins via formation of anionic complexes. The electrophoretic mobility of the proteins in the polyacrylamide gel is then dependent on the molecular weight of the protein and the separation occurs when a current is applied. The reproducibility of SDS-PAGE and IEF has improved by the development of precast gels. Protein visualization or detection after 2DE separation is commonly carried out using organic dyes, silver staining, radioisotope tagging, or by fluorescent and chemiluminescent labeling agents, as discussed below [5].

After running the gels are digitalized and the resulting images can be analyzed using a variety of software. This allows comparison of protein spot densities from test and control samples across different gels. After this, spots of interest can be excised from the gel and the proteins digested for identification by MS [6, 7]. 2DE-MS remained as the main proteomic technique for many years due to advances in protein mass spectrometry, the availability of complete genome sequences, and the development of computational programs for correlation and analysis of data [7, 8].

Until the late 1990s, identification of statistically significant differences across two or more proteomes required the running and analysis of many 2DE gels. This was difficult due to technical variations in sample preparation and gel running conditions [7]. In 1997, Unlu and colleagues presented an effective method of reducing gel-to-gel variation in a technique that they called difference gel electrophoresis [9]. This method was later developed by Amersham Biosciences (now GE Healthcare) and is now called differential in-gel electrophoresis (DIGE). In a typical experiment, up to three protein extracts can be compared on a single gel by covalently labeling these prior to electrophoresis with size- and charge-matched spectrally resolvable fluorescent dyes (Cy2, Cy3, and Cy5) [10]. Therefore, fluorescent imaging of the gel at the wavelengths specific for each CyDye generates separate images for each proteome and these can be overlaid directly for display of any differentially expressed proteins and these proteins identified by MS as described above (Fig. 2). Here we present a typical 2D-DIGE protocol.



**Fig. 1** Basic experimental flow in 2D gel electrophoresis



**Fig. 2** Experimental flow of 2D-DIGE procedure

## 2 Materials

1. Protein samples of interest.
2. Extraction buffer: 30 mM Tris pH 8.0, 8 M urea and complete EDTA-free protease inhibitors (Roche Diagnostics, Mannheim, Germany).
3. CyDyes (GE Healthcare; Little Chalfont, Bucks, UK): RPK0272—25 nmol of Cy2™ DIGE Fluor minimal dye; RPK0273—25 nmol of Cy3™ DIGE Fluor minimal dye; RPK0275—25 nmol of Cy5™ DIGE Fluor minimal dye.
4. Sonicator with micro-probe that can fit inside a 1.5 mL-capacity Eppendorf tube.
5. 24 cm IPG strips (pH 4–7) (*see Note 1*).

6. Strip rehydration buffer: 30 mM Tris pH 8.0, 7 M urea, 2 M thiourea, 4% CHAPS, 2% DTT, 2% pH 3–10 IPG buffer.
7. SDS equilibration buffer: 50 mM Tris-Cl, pH 8.8, 6 M urea, 30% (v/v) glycerol, 2% (w/v) SDS, 0.01% bromophenol blue.
8. Strip equilibration buffer 1: SDS equilibration buffer with 2% iodoacetamide.
9. Strip equilibration buffer 2: SDS equilibration buffer with 5% iodoacetamide, 0.01% and bromophenol blue (*see Note 2*).
10. 50 mM Tris-HCl, pH 6.8.
11. IPGbox (GE Healthcare) or similar.
12. First dimension isoelectric focusing (IEF), Ettan™ IPGphor™ 3 Isoelectric Focusing System (GE Healthcare) or similar.
13. 12.5% acrylamide, 0.33 N,N'-methylenebisacrylamide, 0.375 M Tris-HCl pH 8.8, 0.1% SDS, 0.1% (w/v) ammonium persulfate, 0.025–0.09% (v/v) tetramethylethylenediamine (TEMED) (*see Note 3*).
14. 0.5% (w/v) agarose in Laemmli SDS-PAGE electrode buffer.
15. Second dimension, Ettan DALT6 Electrophoresis System (GE Healthcare) or similar.
16. Typhoon FLA 9500 Imager (GE Healthcare).
17. DeCyder™ 2D Software (GE Healthcare).

---

### 3 Methods

1. Sonicate tissue samples in 7:1 (w:v) extraction buffer and centrifuge at approximately  $13,000 \times g$  for 20 min to pellet insoluble debris.
2. Transfer supernatants to fresh 1.5 mL-capacity microcentrifuge tubes.
3. Estimate the protein concentration of the samples using a standard protein measurement kit (*see Note 4*).
4. Dilute each CyDye with 25  $\mu$ L of fresh anhydrous dimethyl formamide immediately prior to the labeling reaction to obtain a stock solution of 1 nmol (*see Note 5*).
5. Add 0.4  $\mu$ L of 1 nmol Cy2 to 50  $\mu$ g of sample 1.
6. Add 0.4  $\mu$ L of 1 nmol Cy3 to 50  $\mu$ g of sample 2.
7. Add 0.4  $\mu$ L of 1 nmol Cy5 to 50  $\mu$ g of a mixture 1:1 (25  $\mu$ g each) of samples 1 and 2 (*see Note 6*).
8. Incubate samples on ice for 30 min in the dark.
9. Quench samples by adding 1  $\mu$ L of 10 mM lysine and leave the samples on ice for 10 min in the dark.
10. Complete the volume of each the samples to 150  $\mu$ L with rehydration buffer.

11. Incubate the samples on ice for 15 min in the dark.
12. Mix all three samples together (*see Note 7*).
13. Add 450 µL of the sample mixture to IPG strips and allow to hydrate at 20 °C for 12 h in an IPGbox.
14. Run the IPG strip on the IPGphor at 200 V for 1 h, 500 V for 1 h, 1000 V for 1 h, and 8000 V for 8 h at 20 °C using a maximum current setting of 50 mA/strip (*see Note 8*).
15. Strips can be frozen at –80 °C or **step 14** can be carried out immediately.
16. Equilibrate the IPG strip in two steps: first, soak the strip in 100 mL equilibration buffer 1 for 10 min.
17. Incubate the strips for 10 min in equilibration buffer 2 (*see Note 9*).
18. Prepare resolving acrylamide gel solution as needed (*see Subheading 2, item 12*) (*see Note 10*).
19. Add APS and TEMED last to the solution and pour this between assembled low fluorescence glass plates to within approximately 2 cm from the top.
20. Carefully layer butanol on top to help achieve a flat surface on the gel upon polymerization.
21. Rinse the butanol off the gel surface, apply the equilibrated IPG strips so it is sitting immediately on top of the gel, and seal with 0.5 % agarose in SDS running buffer on top (*see Note 11*).
22. Carry out electrophoresis at 60 V for 1 h followed by 30 µA/gel until the dye front reaches the end of the gel.
23. Scan the gels directly between the glass plates (*see Notes 12* and *13*) using filters specific for the excitation and emission wavelengths of Cy2 (480 and 530 nm, respectively), Cy3 (540 and 590 nm, respectively), and Cy5 (620 and 680 nm, respectively).
24. Export the images as 16-bit tagged image file format (TIFF) files for analysis.
25. Analyze the images automatically using the DeCyder Batch Processor and Differential In-Gel Analysis (DIA) software tools.
26. Compare the protein spot volumes on the Cy5-labeled pooled standard image with matching spots on the Cy2- or Cy3-labeled images.
27. Match the images from each gel with the biological variation analysis (BVA) software using the Cy5-labeled pooled standard image for normalization of each protein spot.
28. Use the software to identify protein spots with differences in abundance. Within gels, this is achieved by direct overlay of spots and across gels by land marking, warping, and matching spots using the Biological Variation Analysis function of the DeCyder software (*see Note 14*).

---

## 4 Notes

1. The IPG strip pH range should be chosen to maximize resolution of the proteins of interest. The pH 4–7 strips will provide better resolution of acidic protein spots.
2. The bromophenol blue is added to provide a visible dye front during electrophoresis.
3. The acrylamide and *N,N'*-methylenebisacrylamide concentrations should be chosen based on the desired molecular weight range separation (low percentages of acrylamide allow resolution of high molecular weight proteins and high percentages resolve proteins in the lower molecular weight range). Ammonium persulfate and TEMED should be added just before use since these initiate the gel polymerization process.
4. The presence of urea in the extraction buffer might interfere with some assays, which could lead to false readings.
5. It is best to use DMF as fresh as possible after opening. We noticed decreased labeling efficiency even if the DMF was only 1 month old.
6. The use of an internal standard helps to minimize false positives and false negatives since it can serve as a control for each protein spot on all gels in the analysis. The standard is usually made by combining equal volumes of each extract.
7. It is possible that the CyDyes may show preferential labeling of some proteins although this can be accounted for by reversing the dye:extract combinations.
8. A step voltage gradient was used as we and others have noticed that this helps to avoid horizontal streaking of protein spots on the final 2D gel image.
9. This stage serves the same purpose as a stacking gel in 1D gel electrophoresis. The spots only begin to resolve once they reach the higher pH in the resolving gel.
10. Lower percentage gels favor resolution of high molecular weight proteins and higher percentages can resolve proteins of lower molecular weights.
11. Ensure that there are no air bubbles trapped between the strips and the gel tops to avoid distortion in the protein spot patterns.
12. The second-dimension gels should be poured between low fluorescent Pyrex glass plates to minimize background fluorescence during scanning. Furthermore, the Ettan Dalt II system allowed simultaneous running of multiple plates. This is important as it means that all second dimension gels can be run under approximately the same conditions, which allows for better matching of the gel images in subsequent stages.

13. A key advantage of the 2D DIGE technique is that gels can be imaged after electrophoresis without disassembly of the low-fluorescence glass plates. This ensures that the gels are not deformed or damaged during imaging and also minimizes the possibility of contamination. Furthermore, gels can be scanned for different lengths of time to maximize detection of high- and low-abundance protein spots.
14. Protein identification can be achieved using any kind of protein mass spectrometry. We suggest preparing a gel containing approximately 200 µg of the standard pool followed by colloidal Coomassie Blue staining for excision of spots. Spots must be digested in gel prior to mass spectrometry analysis for protein identification [11].

## Acknowledgments

Dr. Adriano Aquino and Prof. Daniel Martins-de-Souza are funded by FAPESP (São Paulo Research Foundation), grants 15/09159-8 and 13/08711-3.

## References

1. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF et al (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 13:19–50
2. O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250:4007–4021
3. Görg A, Postel W, Weser J, Gunther S, Strahler JR, Hanash SM et al (1987) Elimination of point streaking on silver stained two-dimensional gels by addition of iodoacetamide to the equilibration buffer. *Electrophoresis* 8:122–124
4. Görg A, Postel W, Günther S (1988) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 9:531–546
5. López JL (2007) Two-dimensional electrophoresis in proteome expression analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* 849:190–202
6. Thomas A, Lenglet S, Chaurand P, Deglon J, Mangin P, Mach F et al (2011) Mass spectrometry for the evaluation of cardiovascular diseases based on proteomics and lipidomics. *Thromb Haemost* 106:20–33
7. Penque D (2009) Two-dimensional gel electrophoresis and mass spectrometry for biomarker discovery. *Proteomics Clin Appl* 3:155–172
8. Oliveira BM, Coorssen JR, Martins-de-Souza D (2014) 2DE: the phoenix of proteomics. *J Proteomics* 104:140–150
9. Unlü M, Morgan ME, Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18:2071–2077
10. Knowles MR, Cervino S, Skynner HA, Hunt SP, de Felipe C, Salim K et al (2003) Multiplex proteomic analysis by two-dimensional differential in-gel electrophoresis. *Proteomics* 3:1162–1171
11. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (1996) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1:2856–2860

# Chapter 18

## Selective Reaction Monitoring for Quantitation of Cellular Proteins

Vitor M. Faça

### Abstract

Proteins and proteomes are dynamic and complex. The accurate identification and measurement of their properties such as abundance, location, and turnover are challenging tasks. Even though high-throughput proteomics has significantly evolved, the technique still lacks fully quantitative and reproducible qualities. A mass spectrometry-based targeted proteomic strategy called selective reaction monitoring (SRM) has emerged in recent years as an important multiplex platform to precisely quantify sets of proteins in multiple samples. This has several advantages in terms of sensitivity, reproducibility, and sample consumption compared to classical methods including those based on antibodies. Here, we present a detailed protocol for quantitation of panels of proteins from cell line extracts using the SRM targeted proteomics approach.

**Key words** Mass spectrometry, Protein quantitation, Selective reaction monitoring, Targeted proteomics

---

### 1 Introduction

Proteomics has evolved progressively and currently allows identification and quantification of large sets of proteins with good precision and across several orders of magnitude. This process reached its highpoint recently with the publication of the draft of human proteome [1, 2]. However, the proteome is dynamic as it changes constantly and it is significantly different in both composition and expression levels of individual components across cells, tissues, organs, and individuals. Because of this, the capabilities of high-throughput proteomic methods based on liquid chromatography tandem mass spectrometry (LC-MS/MS) to analyze multiple samples in a timely and reproducible manner are still considered to be limited [3]. In order to address this, targeted proteomics approaches have now focused on precise quantitation of specific sets of proteins in multiple samples or time points. In this way, the monitoring of dynamic proteome changes has emerged as a complementary strategy to complete and validate proteomic and transcriptomic datasets.

The mass spectrometry-based strategy termed selected reaction monitoring (SRM), also known as multiple reaction monitoring (MRM), has been applied frequently for accurate quantitation of small molecules such as drugs or metabolites, and it is now being applied progressively more to the study of peptides and proteins [4]. This increasing interest in options for direct quantitation of proteins or peptides is also driven by limitations and lack of reproducibility of the antibody-based methods [5]. However, it should be stated that all methods have their limitations. The SRM method is mainly based on selecting and quantifying a specific set of peptides (proteotypic peptides) derived from a target list of proteins of interest using LC-MS/MS analysis [6]. SRM takes advantage of the ion filtering capabilities of tandem quadrupole-based equipment, allowing selection of precursors and respective fragment ions produced by CID (collision-induced dissociation) to detect low abundance species among complex mixtures and across four to five orders of magnitude. Due to the rapid tandem quadrupole duty cycle that occurs over milliseconds and the online chromatographic separation of peptides, numerous precursor–fragment ion pairs known as SRM transitions can be monitored simultaneously. This facilitates quantitative experiments involving several different samples in a timely manner.

Since SRM methods efficiently quantitate peptides, the accurate monitoring of proteomic changes in complex samples requires the inclusion of reproducible and efficient enzymatic digestion of proteins in the method. Also, the peptides selected as indicators of protein abundance must meet several characteristics, including the possibility of chemical synthesis to generate these as standards for absolute quantitation. Here, a detailed protocol is presented for quantitative SRM analysis of cellular proteins using, as an example, identification of the proteomic changes in mammary epithelial cells following treatment with transforming growth factor-beta (TGF- $\beta$ ).

---

## 2 Materials

### 2.1 General

1. Use high-performance liquid chromatography (HPLC) or mass spectrometry grade reagents if possible and prepare fresh solutions using MilliQ water on the day of use, maintain at room temperature, unless indicated otherwise.

### 2.2 Cell Culture Components

1. Breast cell lineage MCF-10A (acquired from ATCC-CRL-10317<sup>TM</sup>) (*see Note 1*).
2. Cell culture media: Mammary Epithelial Cell Growth Medium (MEGM<sup>TM</sup>; Lonza; São Paulo, Brazil), supplemented with 100 ng/mL cholera toxin and MEGM kit CC-3150.

2. Tissue culture dishes for adherent cell culture (60.1 cm<sup>2</sup> area) (*see Note 1*).
3. Cell wash solution: phosphate buffered saline (PBS) (pH 7.4).
1. Bradford protein quantification kit (Bio-Rad; Hercules, CA, USA) (*see Note 2*).
2. Denaturation buffer: 8 M urea, 0.15 M Tris-HCl (pH 8.5).
3. Sigma Protease Inhibitor Cocktail for use with mammalian cell and tissue extracts or similar.
4. Trypsin solution: sequencing grade modified 100 µg/mL trypsin (Promega; Madison, WI, USA) in 0.1 M ammonium bicarbonate solution (*see Note 3*).
5. Reducing solution: 10 mg/mL dithiotreitol in 0.15 M Tris-HCl (pH 8.5) (*see Note 4*).
6. Alkylation solution: 20 mg/mL iodoacetamide in 0.15 M Tris-HCl (pH 8.5) (*see Note 4*).
7. Solid phase purification: OASIS HLB solid-phase-extraction columns (Waters Corporation; Milford, MA, USA).
8. Equilibration solution: 95 % water, 5 % acetonitrile, 0.1 % formic acid.
9. Elution solution: 50% water, 50% acetonitrile, 0.1% formic acid.

#### **2.4 Targeted Mass Spectrometry Components**

1. Equipment: tandem quadrupole Xevo-TQs mass spectrometer coupled to a Class I Ultra Performance Chromatographic System (UPLC) (Waters Corporation).
2. Chromatographic column: ACQUITY UPLC HSS C<sub>18</sub>column (1.8 µm particle size, 1 mm inner diameter × 150 mm length) (Waters Corporation).
3. Reconstitution buffer: 3% acetonitrile, 0.1% formic acid.
4. Solvent A for reverse phase chromatography: 95% water, 5% acetonitrile, 0.1% formic acid.
5. Solvent B for reverse phase chromatography: 99.9% acetonitrile, 0.1% formic acid.

### **3 Methods**

#### **3.1 Sample Preparation for SRM Proteomic Analysis of Cell Lines**

1. Cultivate cancer cell lineages in 60 cm<sup>2</sup> plates in MEGM supplemented with 10% FBS according to standard protocols for cell culture and until confluence.
2. Remove old media and cultivate cells in fresh media containing specific treatments and/or for the period of interest in time-course experiments (*see Note 5*).

3. If proteins of interest are also present in the cell secretion, save conditioned media and remove debris first by centrifugation at  $12,000 \times g$  and then by filtration through a 0.22 mm filter (*see Note 6*).
4. Collect cells after appropriate time or treatment by washing three times with ice-cold PBS, drain any residual solution, add 1 mL of denaturation buffer containing freshly added protease inhibitor cocktail, and remove the cells from the plate surface using a cell scraper (*see Note 7*).
5. Carry out cell lysis and total protein extraction performing three cycles of sonication and then immerse the samples in an ice bath (*see Note 8*).
6. Quantitate total proteins in cell extracts by protein assay and use 50 µg aliquots for SRM analysis (*see Note 9*).
7. Reduce the protein cysteine residues by adding 5 µL of reduction solution and maintaining the reaction at 37 °C for 1 h.
8. Perform protein alkylation by adding 10 µL of alkylating solution and incubate the samples at room temperature for an additional 1 h.
9. Dilute samples to approximately 0.6 M by the addition of 0.15 mM Tris-HCl (pH 8.5) (*see Note 10*).
10. Perform trypsin digestion by adding 10 µL trypsin solution to each sample to a final enzyme:protein ratio of 1:50 (w/w) and incubate for 2 h at 37 °C, then add an additional 5 µL of trypsin solution to each digest and incubate for an additional 16 h or overnight at 37 °C.
11. Desalt samples using solid-phase extraction in OASIS columns as follows: (1) condition the column with 1 mL acetonitrile; (2) equilibrate the column with 1.6 mL equilibration solution; (3) apply the sample to the column; (4) wash the column with 1.6 mL of equilibration solution; (5) elute the peptides with 1.2 mL of elution solution; and (6) dry the eluted peptides in a speedvac (*see Note 11*).

### **3.2 SRM Method Development and Data Analysis**

1. Select protein targets for SRM analysis and identify the proteotypic peptides (Table 1) (*see Note 12*).
2. Synthesize or obtain the selected peptides commercially if quantitative analysis is of interest (*see Note 13*).
3. Perform a collision energy and chromatographic separation standardization for the selected set of proteotypic peptides (Table 1) (*see Note 14*).
4. Suspend the samples for SRM analysis in 50 µL of reconstitution buffer, vortex thoroughly, centrifuge at  $12,000 \times g$  for 15 min, and transfer the supernatants to mass spectrometer compatible injection vials.

**Table 1**

**Proteotypic peptides and analytical parameters utilized in the study with MCF-10A cells. The protein set was selected based on high-throughput proteomic experiments (unpublished data). Peptides were synthesized and analyzed individually to maximize efficiency of fragmentation. The selection of 2 SRM transitions per proteotypic peptide improved method reliability**

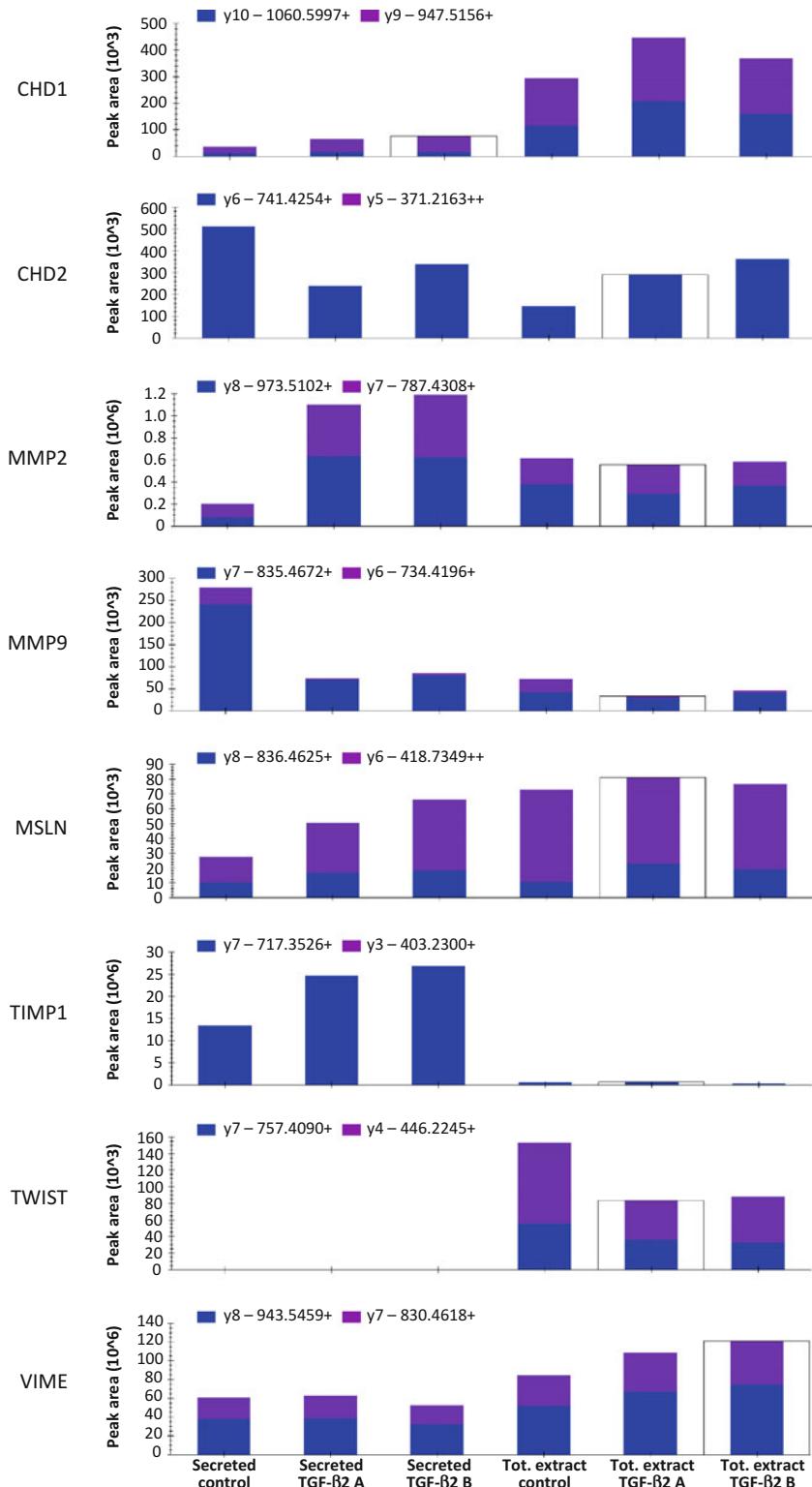
| SwissProt ID | Protein | Sequence        | m/z      | Charge | Col. energy | Trans.1     | Trans. 2     |
|--------------|---------|-----------------|----------|--------|-------------|-------------|--------------|
| P12830       | CDH1    | NTGVISVVTGGLDR  | 716.39 2 | 22     |             | 716.4>947.6 | 716.4>1060.7 |
| P19022       | CDH2    | GPFPPQELVR      | 521.79 2 | 20     |             | 521.8>371.4 | 521.8>741.6  |
| P08253       | MMP2    | AFQVWSDVTPLR    | 709.88 2 | 20     |             | 709.9>787.5 | 709.9>973.6  |
| p14780       | MMP9    | AFALWSAVTPLTFTR | 840.96 2 | 21     |             | 840.9>734   | 840.9>835.6  |
| Q13421       | MSLN    | LLGPHVEGLK      | 531.82 2 | 18     |             | 531.8>418.6 | 531.6>836.6  |
| P01033       | TIMP1   | GFQALGDAADIR    | 617.32 2 | 14     |             | 617.3>404.3 | 617.3>717.6  |
| Q15672       | TWIST   | IIPTLPSDK       | 492.29 2 | 26     |             | 429.3>446.2 | 429.3>757.6  |
| P08670       | VIME    | ILLAELEQLK      | 585.36 2 | 18     |             | 585.4>830.6 | 585.4>943.6  |

5. Run appropriate calibration curves with standard peptides (*see Note 15*).
6. Inject 10 µL of samples in triplicate (or more) to facilitate good statistical analysis (Fig. 1) (*see Note 16*).
7. Analyze the data using the MS vendor software or using the Skyline software [7] (Fig. 1) (*see Note 17*).

---

#### 4 Notes

1. Cell culture in MEGM media is standard for MCF-10A cells. Adapt to appropriate cell culture media following ATCC specifications ([www.atcc.org](http://www.atcc.org)) for other cell lines as necessary. In general, cells are cultivated at a density of  $2\text{--}5 \times 10^6$  cells per plate, producing approximately 200–500 µg of protein in total cell extracts.
2. Other protein assay reagents can be used but the user must ensure that it is compatible with the reagents present in the protein extracts of interest.
3. Once prepared, pass the solution through a 0.22 µm syringe filter and use the filtrate.
4. Always prepare fresh reducing and alkylation solutions to guarantee efficient reaction with cysteine residues. Do not store these solutions.
5. For the purpose of demonstrating the protocol, we performed experiments with the breast epithelial cell line MCF-10A treated with 10 ng/mL of TGF-β2 for 72 h.



6. The analysis of proteins present in conditioned media requires special steps for sample processing. Please see Faça et al. for a detailed protocol [8].
7. The step of cell recovery from culture plate is important for reproducibility. Since cells may vary significantly in terms of plate adherence, refine the protocol if necessary.
8. The sonication step can be performed in an ultrasonic bath or with an ultrasonic probe. Note that sonication can heat samples and promote protein carbamylolation by urea present in denaturation buffer. Therefore, avoid heating by using short (e.g., 30 s) cycles of sonication followed by cooling the samples in an ice bath.
9. Quantification of total protein is important to allow efficient reduction, alkylation, and enzymatic protein digestion reactions.
10. Dilution of the solution guarantees good trypsin activity, which is close to 100% in urea solutions at concentration less than 1 M according to the manufacturer.
11. Samples can be stored dried at -80 °C until ready for SRM analysis.
12. In order to develop a SRM method for target proteins, use the following workflow: (1) perform a virtual tryptic protein digestion using for example the PeptideMass tool available in UNIPROT ([www.uniprot.org](http://www.uniprot.org)); (2) select tryptic peptides containing 10–20 amino acids since these are easier to obtain by solid-phase peptide synthesis; (3) if possible, select peptides containing proline residues, which generate intense fragment peaks; (4) avoid peptides containing methionine or N-terminal glutamine, as these accumulate in source modifications during MS ionization; and (5) check previous detection and predicted suitability for selected peptides using the SRMatlas databank ([www.srmatlas.org](http://www.srmatlas.org)) [9].
13. Although the use of synthetic peptides representing the target proteins is not obligatory, they are useful for method development and refinement. Peptides can be synthesized rapidly in

←  
**Fig. 1** SRM analysis of MCF-10A cell line treated with TGF- $\beta$ 2. The protein set was selected based on high-throughput proteomic experiments (unpublished data). Cells were treated with TGF- $\beta$ 2 for 72 h. Both total cell extract (50  $\mu$ g of protein) and conditioned media (2 mL) were analyzed for the same set of proteins. The data were collected for two biological replicates (TGF- $\beta$ 2 A and TGF- $\beta$ 2 B) and samples were injected in triplicate. Secreted proteins such as TIMP1 were detected only in the media, while the nuclear transcription factor TWIST was detected only in total cell extracts. The full experiment was performed in approximately 6 h using a 20 min gradient for peptide separation. The effects on protein levels induced by TGF- $\beta$  in MCF-10A cells were consistent with the results of previous experiments and were validated by Western blotting for some proteins

house using standard solid-phase Fmoc chemistry. Preselection of short peptides (<20 amino acids) facilitates synthesis and generally produces good yields. If absolute quantitation is of interest, purify the peptides by reverse phase chromatography and use an accurate method to quantify these, such as amino acid analysis. Also, the use of isotopically labeled peptides (e.g., Lys<sup>13</sup>C<sub>6</sub>) allows more precise quantification since this approach incorporates internal standards.

14. The study of ideal collision energy for SRM studies is essential for good sensibility. Also, it is important to perform these studies in the local instrument since several tuning parameters can affect the fragmentation efficiency. The optimization of a chromatographic gradient with the specific set of peptides is also important to guarantee quick runs and good reproducibility. For these steps, consider using the software Skyline [7] which is designed to facilitate method development in addition to performing efficient data analysis across multiple platforms.
15. Chromatography in micro or nano-scale is susceptible to micro-particles originating from insoluble material or solid-phase extraction column leaking. This simple centrifugation step improves the column lifetime significantly.
16. The injection of 10 µL corresponds to 10 µg of protein digest. This amount is below the loading capacity of the UPLC HSS C<sub>18</sub> Column [1 mm (inner diameter) × 150 mm length] to obtain best chromatographic separations.
17. The use the multi-platform Skyline software to integrate, calibrate, and quantify samples analyzed by SRM streamlines the development and application of the method.

---

## Acknowledgments

This research was supported by FAPESP (Young Scientist Grant – Proc. No. 2011/0947-1), Center for Cell Based Therapy-CTC-CEPID (Proc. FAPESP 2013/08135-2), and CISBi-NAP. V.M.F. received fellowships from CNPq, Proc. No. 308561/2014-7.

## References

1. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R et al (2014) A draft map of the human proteome. *Nature* 509:575–581
2. Wilhelm M, Schlegl J, Hahne H, Moghaddas Gholami A, Lieberenz M, Savitski M et al (2014) Mass-spectrometry-based draft of the human proteome. *Nature* 509:582–587
3. Cox J, Mann M (2011) Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* 80:273–299
4. Picotti P, Aebersold R (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* 9:555–566

5. Baker M (2015) Reproducibility crisis: blame it on the antibodies. *Nature* 521:274–276
6. Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D et al (2007) Computational prediction of proteotypic peptides for quantitative proteomics. *Nat Biotechnol* 25:125–131
7. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B et al (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26:966–968
8. Faça VM, Palma CS, Albuquerque D, Canchaya GN, Grassi ML, Epifânio VL et al (2014) The secretome analysis by high-throughput proteomics and multiple reaction monitoring (MRM). *Methods Mol Biol* 1156:323–335
9. Picotti P, Lam H, Campbell D, Deutscher EW, Mirzaei H, Ranish J et al (2008) A database of mass spectrometric assays for the yeast proteome. *Nat Methods* 5:913–914

# Chapter 19

## Characterization of a Protein Interactome by Co-Immunoprecipitation and Shotgun Mass Spectrometry

Giuseppina Maccarrone, Juan Jose Bonfiglio, Susana Silberstein,  
Christoph W. Turck, and Daniel Martins-de-Souza

### Abstract

Identifying the partners of a given protein (the interactome) may provide leads about the protein's function and the molecular mechanisms in which it is involved. One of the alternative strategies used to characterize protein interactomes consists of co-immunoprecipitation (co-IP) followed by shotgun mass spectrometry. This enables the isolation and identification of a protein target in its native state and its interactome from cells or tissue lysates under physiological conditions. In this chapter, we describe a co-IP protocol for interactome studies that uses an antibody against a protein of interest bound to protein A/G plus agarose beads to isolate a protein complex. The interacting proteins may be further fractionated by SDS-PAGE, followed by in-gel tryptic digestion and nano liquid chromatography high-resolution tandem mass spectrometry (nLC ESI-MS/MS) for identification purposes. The computational tools, strategy for protein identification, and use of interactome databases also will be described.

**Key words** Mass spectrometry, Co-immunoprecipitation, Interactome, Proteome, Computational tools

---

### 1 Introduction

The interactome may be defined as the complete set of molecular interactions of a molecule in a given cell or other biological environment [1]. The interactome of a protein represents the proteins and molecules that interact with it to promote, regulate, and inhibit its activity or expression. Alternatively, members of the interactome may be substrates or molecular components regulated by the protein of interest.

Identifying the interactome of a given protein may provide leads about the protein's function and the molecular mechanisms in which it is involved. The protein interaction map of *Drosophila melanogaster* encompasses 556 protein complexes and can assign functional

processes to almost 600 protein-coding genes, many of which previously lacked annotation. The data gained from the *Drosophila melanogaster* interactome may also provide information about human cell interactomes [2] and enrich genome databases by generating annotations to non-annotated protein-coding genes. An extensive proteomic survey that used affinity tag purification of *E. coli* strains identified 5,993 protein interactions and allowed proteins to be assigned functions such as protein synthesis, amino acid metabolism, biofilm formation, and motility [3]. More specific to a protein, novel partner proteins in brain tissue were unraveled to collapsin response mediator protein-2 (CRMP2/DPYSL2) (Reference: Martins-de-Souza D, Cassoli JS, Nascimento JM, Hensley K, Guest PC, Pinzon-Velasco AM, Turck CW. The protein interactome of collapsin response mediator protein-2 (CRMP2/DPYSL2) reveals novel partner proteins in brain tissue. *Proteomics Clin Appl.* 2015 Oct;9(9–10):817–31). The study of protein interactomes in diseases may reveal dysfunctional pathways, further insights into regulation, and the possible role that protein partners play in the disease [4].

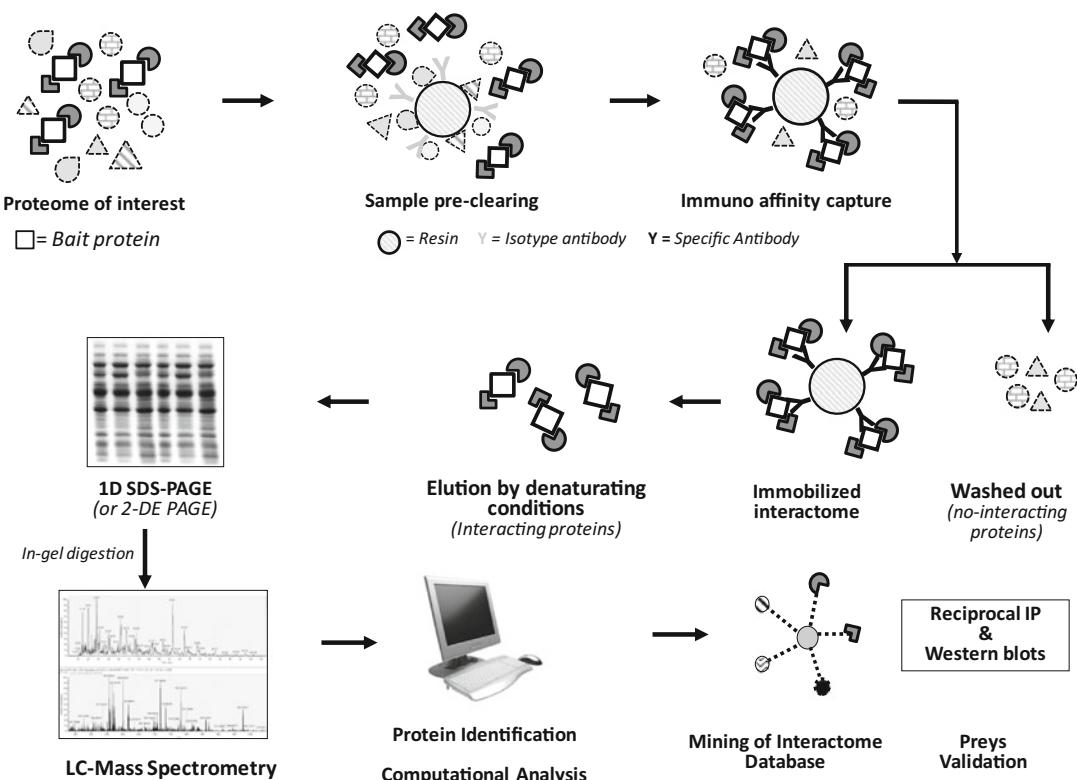
Many interactome studies have been performed with yeast two-hybrid screening (Y2H) and Tandem Affinity Purification (TAP) [5], although co-immunoprecipitation (co-IP) is also a reliable method for studying protein interactomes with its own advantages [6]. Briefly, the co-IP of intact protein complexes employs an immobilized antibody that targets a specific protein. The sample of interest is passed through or incubated with the antibody under non-denaturating conditions. Next, the target protein (bait) binds to the antibody along with its partners (prey), while noninteracting proteins pass through or do not bind to the matrix. The captured protein complex may be analyzed further by proteomic analysis (Fig. 1). Co-IP and subsequent mass spectrometry (MS) have been applied to proteins of interest in diseases. Recently, the B-Raf interactome was characterized in mouse hippocampal cells [7, 8]. This protein is involved in a cellular signaling cascade, known to play a role in cancer [9]. The interactomes of the G-protein beta subunit [10], neurocalcin delta [11], dopamine transporter [12], and dynamin-1 [13], which are involved in neurological disorders, have also been characterized by co-IP and MS.

Here, we describe a protocol (shown in Fig. 1) that can be used to investigate the interactome of a protein of interest.

## 2 Materials

### 2.1 Immunoprecipitation Sample Preparation

1. Ice-cold phosphate buffered saline, PBS: 137 mM NaCl, 10 mM phosphate, 2.7 mM KCl, pH 7.4.
2. Modified RIPA buffer: 50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1 mM EDTA, 1% NP-40, 1/25 protease inhibitor cocktail (Roche Molecular Biomedicals, Mannheim, Germany), 1 mM phenylmethylsulfonyl fluoride, and 1 mM activated sodium orthovanadate.



**Fig. 1** Steps involved in the characterization of a bait protein interactome. Firstly, the cell lysate is precleared with A/G beads and nonimmune rabbit IgG. The precleared lysate is incubated with the A/G beads and the antibody against the bait protein to fish protein preys. The precleared lysate is again incubated with the nonimmune rabbit IgG to fish nonspecific binders (Mock-IP fraction, not shown). The immobilized interactome is washed and eluted from the beads, and proteins are separated before identification using LC-ESI-MS/MS. Data analysis includes the identification of the protein interactome, data filtering considering mock-IP results, mining of the interactome database available in the public domain and validation

## 2.2 Immunoprecipitation Assay

1. Immunoprecipitating antibody: anti B-Raf antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA).
2. Nonimmune antibody produced in the same species of the immunoprecipitating antibody: normal rabbit IgG.
3. Protein A/G plus agarose.

## 2.3 Proteomics Sample Preparation

1. Electrophoresis equipment.
2. Laemmli sample buffer 2x: 0.125 M Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue.
3. Coomassie Brilliant Blue R-250.
4. 25 mM ammonium bicarbonate/50% acetonitrile.
5. 25 mM ammonium bicarbonate.
6. 10 mM dithiothreitol.

7. 50 mM iodoacetamide.
8. 50% acetonitrile/2% formic acid.
9. Trypsin solution: 5 ng/ $\mu$ L trypsin in 25 mM ammonium bicarbonate.

#### **2.4 nLC ESI-MS/MS**

1. Nano LC system.
2. Trap column RP-C18.
3. In-house packed nano-column (75  $\mu$ m i.d.  $\times$  15 cm).
4. Column packing material RP-C18, 3  $\mu$ m.
5. Picofrit self-packed column (75  $\mu$ m i.d., 10  $\mu$ m tip).
6. Distilled H<sub>2</sub>O, 0.1% HCOOH.
7. Eluent A: H<sub>2</sub>O/ACN (95/5), 0.1% HCOOH.
8. Eluent B: ACN, 0.1% HCOOH.
9. Centrifuge filter: 0.22  $\mu$ m.
10. LTQ Orbitrap XL mass spectrometer (Thermo Fisher, Germany).
11. Nano electrospray ionization (ESI) source (Thermo Fisher).

#### **2.5 Mass Spectrometry Software**

1. LC-ESI-MS/MS data acquisition software: Xcalibur v. 2.07 (Thermo Fisher).
2. Nano-LC controlling software: LC-Eksigent v.2.09 (Eksigent, Dublin, CA, USA).
3. Raw MS and MS/MS spectra processing software: Bioworks v. 3.3 (Thermo Fisher).
4. Protein database search engine: Mascot Demon 2.2.2 (Matrix Science, UK).

#### **2.6 Western Blot Analysis**

1. Immobilon PVDF membrane.
2. TBS-T buffer: 137 nM NaCl, 20 mM Tris-HCl, 0.05% Tween-20, pH 8.0.
3. 5% nonfat dry milk in TBS-T.
4. ECL Plus reagent kit.

### **3 Methods (See Note 1)**

#### **3.1 Immunoprecipitation Sample Preparation**

1. Grow cells to 70% confluence in a 100 mm cell culture plate (see Note 2).
2. Aspirate culture medium and wash cells with 6 mL of ice-cold PBS.
3. Harvest cells on ice by gently scraping them off with a rubber policeman in 1 mL of ice-cold complete modified RIPA buffer (see Note 3).

4. Lyse cells for 2 h at 4 °C with gentle agitation followed by centrifugation at  $3500 \times g$  for 5 min at 4 °C.
5. Collect the supernatant in a 1.5 mL microcentrifuge tube.
6. Take an aliquot of this supernatant for future studies (e.g., Western blot validation).

### **3.2 Co-immunoprecipitation Assay**

1. Preclear the resulting lysate by incubation for 1 h at 4 °C with 1 µg of nonimmune rabbit polyclonal antibody (nonimmune antibody produced in the same species of the anti-B-Raf antibody used).
2. After 1 h, add 30 µL of protein A/G plus agarose for a 2 h incubation at 4 °C with gentle agitation (*see Note 4*).
3. Separate nonspecific proteins bound to the antibody and protein A/G plus agarose beads by centrifugation at  $3500 \times g$  for 5 min at 4 °C, and discard the pellet containing nonspecific proteins, antibody, and protein A/G plus agarose.
4. Collect the supernatant and separate into two equal parts for the specific and control IPs.
5. The specific co-IP procedure is performed by incubating the supernatant with 1 µg of anti B-Raf antibody overnight at 4 °C, with gentle agitation (*see Note 5*).
6. The control IP is carried out by incubating an aliquot of the supernatant with 1 µg of nonimmune rabbit polyclonal antibody overnight at 4 °C with gentle agitation (label the immunoprecipitated fraction resulting from the control IP as “mock IP” (*see Note 6*)).
7. Fish the B-Raf antibody, the B-Raf protein, and B-Raf interactors by adding 35 µL of protein A/G plus agarose, incubate for 2 h at 4 °C with gentle agitation, and then pellet by centrifugation at  $3500 \times g$  for 5 min at 4 °C (save the supernatant for future studies).
8. Perform the same procedure for the mock IP.
9. To reduce nonspecific binding, wash the immunoprecipitate three times with 500 µL of complete modified RIPA buffer and once with 1 mL of ice-cold PBS (*see Note 7*).
10. Apply the same wash procedure for the mock IP pellet.

### **3.3 Proteomics Sample Preparation**

1. Suspend the pellet/immunoprecipitate in 30 µL 2× Laemmli sample buffer (*see Note 8*).
2. Before electrophoresis, heat the sample for 5 min at 95 °C, centrifuge and separate from the agarose beads.
3. Resolve the proteins in the total sample by 10% SDS-PAGE and then stain in-gel with Coomassie Brilliant Blue R-250 for 1 h at room temperature with gentle agitation (*see Note 9*).

4. Slice the gel lane is sliced into ~20 fractions, cut into small pieces and wash twice with 25 mM ammonium bicarbonate/50% acetonitrile.
5. Reduce sulfhydryl bonds in the proteins within the gel by incubating with 10 mM dithiothreitol for 30 min at 56 °C.
6. Alkylate the reduced sulfhydryl groups by incubating with 50 mM iodoacetamide for 30 min at room temperature, wash twice, dry and rehydrate in 20 µL trypsin solution for 15 min on ice.
7. Remove the surplus trypsin solution, cover the gel pieces with 25 mM ammonium bicarbonate, and allow digestion to proceed for 4–6 h at 37 °C (see Note 10).
8. Extract the resulting peptides twice with 50% acetonitrile/2% formic acid and then lyophilize to dryness.

### 3.4 nLC ESI-MS/MS

1. Dissolve the dried peptides in 10 µL 0.1% HCOOH and pass through the filter before analysis by nLC ESI-MS/MS (see Note 11).
2. Separate the peptides according to their physicochemical properties using a nano HPLC system equipped with a trap column and a picofrit nano-column coupled online to an LTQ-Orbitrap mass spectrometer via a nano ESI source (see Note 12).
3. Load 5 µL of each sample onto the RP-C18 trap column and wash for 10 min with eluent A at a flow-rate of 3 µL/min.
4. Separate the peptides on the RP-C18 nano-column by applying a linear gradient of eluent B from 2 to 10% in 5 min and 10–40% in 98 min, at a flow rate of 200 nL/min (see Note 13).
5. Operate the mass spectrometer in positive ion mode, using a set method that includes a data-dependent acquisition of a full scan (FS) and MS/MS scans of each peptide.
6. The FS scan is recorded at the Orbitrap analyzer in the mass range from  $m/z$  380 to 1600 at resolution 60,000 (FMHW,  $m/z$  400) and the FS data are acquired in profile mode.
7. The fragment spectra (MS/MS) are acquired at the ion trap (LTQ) in centroid mode (see Note 14).
8. The fragments are generated using collision-induced dissociation mode, with no fragmentation at the source considered.
9. The five most intensive ions per FS are selected for the fragmentation such that each mass ion with intensity greater than 500 counts is fragmented once and inserted into a dynamic exclusion list for 60 s (see Note 15).
10. Other parameters set for the fragmentation are as follows: (1) 30 s repeat duration time; (2) 2 µm isolation width ( $m/z$ ); (3) 30 ms activation time; (4) 35 V normalized collision energy; and (5) Q=0.250 activation.

### **3.5 Protein Identification**

1. Convert the MS raw data to \*.XML format using BioWorks tool or similar before searching against a decoy UniProt mouse protein database (*see Note 16*).
2. Set the search parameters as follows: (1) trypsin as enzyme; (2) one missed cleavage allowed; and (3) cysteine carbamidomethylation (fixed) and methionine oxidation (variable) chosen as modifications.
3. Precursor and fragment ion mass accuracies are set to 15 ppm and 0.8 Da, respectively.
4. Peptides are accepted as accurate if the confidence interval (CI) is 95%, the Mascot score is greater than 30, and if at least two unique and nonredundant peptides can be matched to the protein.

### **3.6 Computational Analysis (See Note 17)**

1. In each co-IP MS dataset, check for the presence of the bait protein in terms of peptide signal intensities and sequence coverage (*see Note 18*).
2. Identify and discard “mock IP” proteins and other proteins introduced during sample handling (e.g., keratin, IgGs) identified in multiple co-IPs and MS experiments.
3. Generate a single protein list by determining common proteins identified in multiple, independent co-IP MS experiments, with candidates identified in more than one co-IP MS replicate selected as the most reliable (*see Note 19*).
4. Sort proteins on the basis of the sequence coverage as those identified by only a small number of peptides may be weak or dynamic interactors.
5. Use interactome databases to look for the bait protein and known interactors (*see Note 20*).
6. Perform a detailed sequence analysis of the interactors and sort those that share subdomains as this may be an additional indicator of the validity of the identification.
7. Make a list of the novel interactors.
8. For novel protein-protein interactions, use reverse co-IP following the above steps to verify the specificity of the interaction with the bait protein.

### **3.7 Western Blot (See Note 21)**

1. Separate a fraction of the resuspended pellet containing the immunoprecipitated proteins by SDS-PAGE and electrotransfer the resulting protein bands to an Immobilon PVDF membrane at 100 V for 1 h at 4 °C.
2. Block the membranes for 1 h at room temperature with 5% nonfat dry milk in TBS-T.
3. Incubate with the selected primary antibody against the candidate protein overnight at 4 °C and then for 1 h at room temperature with the corresponding secondary antibody.
4. Detect the immune complexes using ECL Plus.

### 3.8 Reciprocal IP Followed by Western Blot

1. Perform reciprocal IPs followed by Western blot to further validate a potential interacting molecule that has been isolated and identified with the co-IP methods described above (*see Note 22*).

---

## 4 Notes

1. This protocol can be used to identify associated molecules of any protein of interest as long as an antibody against the protein is available. Likewise, any kind of biological material can be used, such as cells, a tissue, or extracellular matrix. As an example, we describe a workflow for the interactome analysis of B-Raf performed in HT22 cell lysates. The steps include immunoprecipitation of B-Raf and its protein partners under native conditions using co-IP followed by shotgun mass spectrometry for the identification of the interactome. B-Raf interacting partners were validated with non-MS-based approaches such as Western blot and reciprocal IP followed by Western blot.
2. If the bait protein is expressed at low levels, it may be desirable to use three or more 100 mm plates and harvest them in smaller volumes of modified RIPA buffer (0.5 or 0.75 mL per plate).
3. Buffer composition (salinity, detergent, and concentration) is critical for a successful co-IP experiment. Optimal conditions are empirical and should be determined for each investigated protein/antibody complex. The buffer composition suggested in this protocol is a good starting point. Normally, reducing salt concentration allows for the immunoprecipitation of more proteins. However, the chance of identifying false positives (non-true preys) is also increased. If there is a special interest in identifying phosphorylated proteins, the modified RIPA buffer should contain higher concentrations of activated sodium orthovanadate (1/25 of stock solution).
4. Proteins A and G exhibit variation in binding strength to different immunoglobulins (Ig). This variation exists between different species and between different antibody subclasses from the same species. The decision to use proteins A, G, or A/G depends on the species in which the antibody was produced [14] (*see Table 1*).
5. The concentration of immunoprecipitation antibody and the incubation time may be specific for each protein target. Optimal conditions need to be set up experimentally. However, the amounts and times suggested here can be considered good starting points.

**Table 1**

**Protein A and protein G exhibit variation in binding strength to different immunoglobulins (Ig) [14, 15]. This exists both between different species and between different antibodies subclasses from the same species. The decision of using protein A or G or A/G in an IP experiment depends on the specie in which the antibody was produced. W = weak binding, M = medium binding, S = strong binding, NB = no binding**

| Species | Antibody Class     | Protein A | Protein G | Protein A/G |
|---------|--------------------|-----------|-----------|-------------|
| Human   | Total IgG          | S         | S         | S           |
|         | IgG1, IgG2         | S         | S         | S           |
|         | IgG3               | W         | S         | S           |
|         | IgG4               | S         | S         | S           |
|         | IgM                | W         | NB        | W           |
|         | IgD                | NB        | NB        | NB          |
|         | IgA                | W         | NB        | W           |
|         | IgA1, IgA2         | W         | NB        | W           |
|         | IgE                | M         | NB        | M           |
|         | Fab                | W         | W         | W           |
| Mouse   | ScFv               | W         | NB        | W           |
|         | Total IgG          | S         | S         | S           |
|         | IgM                | NB        | NB        | NB          |
|         | IgG1               | W         | M         | M           |
|         | IgG2a, IgG2b, IgG3 | S         | S         | S           |
| Rat     | Total IgG          | W         | M         | M           |
|         | IgG1               | W         | M         | M           |
|         | IgG2a              | NB        | S         | S           |
|         | IgG2b              | NB        | W         | W           |
|         | IgG2c              | S         | S         | S           |
| Cow     | Total IgG          | W         | S         | S           |
|         | IgG1               | W         | S         | S           |
|         | IgG2               | S         | S         | S           |
| Goat    | Total IgG          | W         | S         | S           |
|         | IgG1               | W         | S         | S           |
|         | IgG2               | S         | S         | S           |
| Sheep   | Total IgG          | W         | S         | S           |
|         | IgG1               | W         | S         | S           |

(continued)

**Table 1**  
**(continued)**

| Species    | Antibody Class  | Protein A | Protein G | Protein A/G |
|------------|-----------------|-----------|-----------|-------------|
|            | IgG2            | S         | S         | S           |
| Horse      | Total IgG       | W         | S         | S           |
|            | IgG(ab), IgG(c) | W         | NB        | W           |
|            | IgG(T)          | NB        | S         | S           |
| Rabbit     | Total IgG       | S         | S         | S           |
| Guinea pig | Total IgG       | S         | W         | S           |
| Hamster    | Total IgG       | M         | M         | M           |
| Pig        | Total IgG       | S         | W         | S           |
| Donkey     | Total IgG       | M         | S         | S           |
| Cat        | Total IgG       | S         | W         | S           |
| Dog        | Total IgG       | S         | W         | S           |
| Monkey     | Total IgG       | S         | S         | S           |
| Chicken    | Total IgY       | NB        | NB        | NB          |

6. The mock IP sample can be used both as a negative control in Western blot for the validation of binding proteins (preys) and in the identification of the nonspecific binding proteins by LC-ESI-MS/MS.
7. To reduce the number of nonspecific proteins, washes can be performed with gentle agitation at 4 °C for 5 min. Furthermore, additional washes can be implemented to determine the specificity and strength of the identified partners.
8. The pellet can be stored at –20 °C for a maximum of 7 days.
9. In some cases, loading smaller amounts of sample may result in a better separation in the gel, especially if the concentration of immunoprecipitating antibody is high or the immunoprecipitated protein interacts with high numbers of other proteins, such as molecular chaperones.
10. Overnight trypsin digestion does not yield more peptides but results in more autolysis trypsin peptides.
11. Before filtering the samples, wet the filter membrane device by adding 0.1% HCOOH, then centrifuge and discard it. This additional step increases sample recovery.

12. The picofrit columns are not coated. In our system, the ionization occurs via the liquid junction and is highly stable across the LC-MS/MS analysis.
13. The use of HPLC-grade bottled water reduces the likelihood of clogging the nano-column by particulate contamination. Particulate contamination can be carbon particles released from the in-house meg-ohm water system.
14. The centroid spectra have a small file size because they contain less information describing the m/z signal. It is advisable to reduce the file size of an LC-MS/MS run by running the FS in profile mode and the MS/MS spectra in centroid mode. The MS/MS information content in centroid spectra is sufficient to identify the fragment sequence.
15. Mono-charged ions are rejected for MS/MS.
16. We use an in-house version of the Mascot search engine for this.
17. A crucial part of the interactome analysis is to filter the nonspecific binding molecules from true physiological interactors. A minimum of two independent co-IP MS analyses should be run to generate a list of interactors (preys) of a target protein. Peptides are considered if they are identified with a false discovery rate (FDR) value between 0.5 and 1 %.
18. This is a critical requirement for the subsequent validation studies.
19. A large bait interactome (>100 interactors) may indicate the identification of proteins that cross-interact, either directly or indirectly [6]. The BioVenn tool (<http://www.cmbi.ru.nl/cdd/biovenn/>) can be used to compare data from a maximum of three experiments.
20. The interactome databases IntAct [15], PINA, [16], BioGRID [17]. Report the methods used for the identification of the interacting proteins. This information can also confirm the validity of the identification. Interactors identified by more than one method are more reliable than those identified by a single method and any previously identified interactor found in the dataset serves as an internal validation and an assessment of the data quality.
21. Selected candidates are validated by Western blot of resolved co-immunoprecipitated proteins.
22. In this case, the reciprocal IP should be performed in cell lysates and the proteins should be resolved by SDS-PAGE and analyzed by Western blot using antibodies against the bait protein (B-Raf in our example). Furthermore, the reciprocal IP stage can help to circumvent the problem of immunodetecting proteins that might be obscured due to their migration close to the IgG heavy or light chains during SDS-PAGE (e.g., vimentin).

## Acknowledgments

D.M.S. is funded by FAPESP (São Paulo Research Foundation, grants 2013/08711-3, 2013/25702-8, and 2014/10068-4).

## References

1. Sanchez C, Lachaize C, Janody F, Bellon B, Röder L, Euzenat J et al (1999) Grasping at molecular interactions and genetic networks in *Drosophila melanogaster* using FlyNets, an Internet database. *Nucleic Acids Res* 27:89–94
2. Guruharsha KG, Rual JF, Zhai B, Mintseris J, Vaidya P, Vaidya N et al (2011) A protein complex network of *Drosophila melanogaster*. *Cell* 147:690–703
3. Hu P, Janga SC, Babu M, Díaz-Mejía JJ, Butland G, Yang W et al (2009) Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol* 7, e96. doi:[10.1371/journal.pbio.1000096](https://doi.org/10.1371/journal.pbio.1000096)
4. Coulombe B (2011) Mapping the disease protein interactome: toward a molecular medicine GPS to accelerate drug and biomarker discovery. *J Proteome Res* 10:120–125
5. Williamson MP, Sutcliffe MJ (2010) Protein-protein interactions. *Biochem Soc Trans* 38:875–878
6. Markham K, Bai Y, Schmitt-Ulms G (2007) Co-immunoprecipitations revisited: an update on experimental concepts and their implementation for sensitive interactome investigations of endogenous proteins. *Anal Bioanal Chem* 389:461–473
7. Bonfiglio JJ, Maccarrone G, Rewerts C, Holsboer F, Arzt E, Turck CW et al (2011) Characterization of the B-Raf interactome in mouse hippocampal neuronal cells. *J Proteomics* 74:186–198
8. Bonfiglio JJ, Inda C, Senin S, Maccarrone G, Reijojo D, Giacomini D et al (2013) B-Raf and CRHRI internalization mediate biphasic ERK1/2 activation by CRH in hippocampal HT22 Cells. *Mol Endocrinol* 27:491–510
9. Garnett MJ, Marais R (2004) Guilty as charged: B-Raf is a human oncogene. *Cancer Cell* 6:313–319
10. Zhang JH, Simonds WF (2000) Copurification of brain G-protein beta5 with RGS6 and RGS7. *J Neurosci* 20:RC59
11. Ivings L, Pennington SR, Jenkins R, Weiss JL, Burgoyne RD (2002) Identification of Ca<sup>2+</sup>-dependent binding partners for the neuronal calcium sensor protein neurocalcin delta: interaction with actin, clathrin and tubulin. *Biochem J* 363:599–608
12. Hadlock GC, Nelson CC, Baucum AJ 2nd, Hanson GR, Fleckenstein AE (2011) Ex vivo identification of protein-protein interactions involving the dopamine transporter. *J Neurosci Methods* 196:303–307
13. Gorini G, Ponomareva O, Shores KS, Person MD, Harris RA, Mayfield RD (2010) Dynamin-1 co-associates with native mouse brain BKCa channels: proteomics analysis of synaptic protein complexes. *FEBS Lett* 584:845–851
14. Richman DD, Cleveland PH, Oxman MN, Johnson KM (1982) The binding of staphylococcal protein A by the sera of different animal species. *J Immunol* 128:2300–2305
15. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C et al (2011) The IntAct molecular interaction database in 2012. *Nucleic Acids Res* 40:D841–D846
16. Cowley MJ, Pinece M, Kassahn KS, Waddell N, Pearson JV, Grimmond SM et al (2012) PINA v2.0: mining interactome modules. *Nucleic Acids Res* 40:D862–D865
17. Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS et al (2010) The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res* 39: D698–D704

# Chapter 20

## Using $^{15}\text{N}$ -Metabolic Labeling for Quantitative Proteomic Analyses

Giuseppina Maccarrone, Alon Chen, and Michaela D. Filiou

### Abstract

Quantitative proteomics has benefited from the application of stable isotope labeling-based approaches. Using stable isotopically labeled material as an internal standard in proteomic comparisons allows an unbiased and accurate quantification of protein expression level changes. Here, we describe the use of *in vivo*  $^{15}\text{N}$  metabolic labeling to generate labeled protein standards from mice. We then present a protocol including sample preparation, mass spectrometry, and data analysis workflows using these standards to compare unlabeled proteomes. We focus on mouse brain tissue and plasma samples, although this conceptual framework can be applied to most organisms.

**Key words**  $^{15}\text{N}$  metabolic labeling, Quantitative proteomics, Mass spectrometry, Peptide quantification

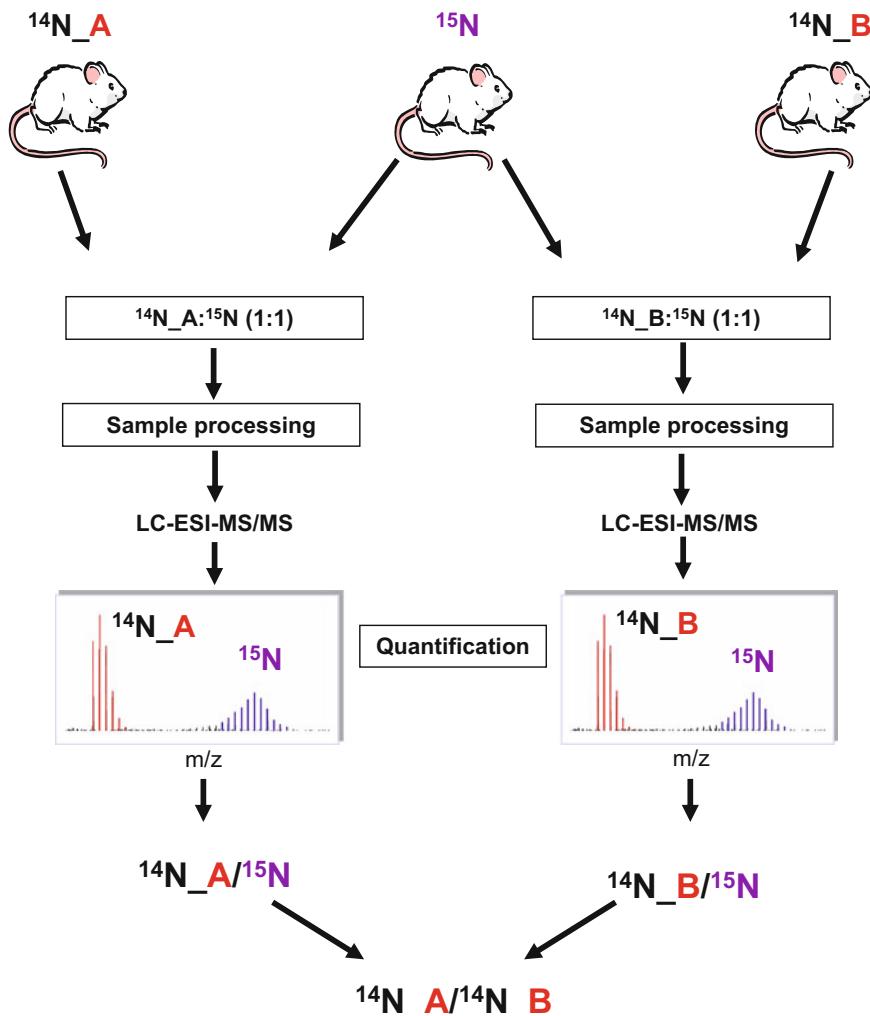
---

### 1 Introduction

Quantitative proteomics enables the comparison of proteomes to identify altered protein expression levels at a given time point. State-of-the-art quantitative proteomic methodologies include the use of stable nonradioactive isotopes to generate labeled internal standards for subsequent comparisons of unlabeled proteomes. The advantage of this approach is the high quantification accuracy due to the fact that handling biases during sample preparation are eliminated [1, 2]. Here, we describe a quantitative proteomic workflow using the  $^{15}\text{N}$  nitrogen isotope, which contains one more neutron compared to the most abundant  $^{14}\text{N}$  form found in nature.

To generate  $^{15}\text{N}$ -labeled material, mice are fed with a  $^{15}\text{N}$ -labeled, bacteria-based diet [3]. Then,  $^{15}\text{N}$ -labeled biosamples such as brain tissue or plasma are used as internal standards for quantitative proteomic comparisons of two unlabeled groups of interest (hereafter named **A** and **B**). Briefly, an unlabeled ( $^{14}\text{N}$ ) version of proteome **A** is mixed at a 1:1 ratio based on tissue weight or protein content with the  $^{15}\text{N}$ -labeled internal standard ( $\text{A}/^{15}\text{N}$ ).

The  $\text{A}/^{15}\text{N}$  protein mix is then resolved according to molecular weight by gel electrophoresis and the protein bands are excised and enzymatically digested to produce peptides, which are analyzed by mass spectrometry. The same procedure is applied to the second group  $\text{B}$  ( $\text{B}/^{15}\text{N}$ ). In both cases, analysis of the  $^{14}\text{N}/^{15}\text{N}$  peptide mixtures by liquid chromatography-tandem mass spectrometry (LC-MS/MS) results in the unlabeled ( $^{14}\text{N}$ ) and the  $^{15}\text{N}$ -labeled versions of the same peptide being separated in the mass spectrometer due to their different mass to charge ratios ( $m/z$ ), leading to generation of two distinct signals. In silico quantification of the  $^{14}\text{N}$  and  $^{15}\text{N}$  peptide signals reveals their relative abundance in the mixture. Next, the data from the  $^{14}\text{N}/^{15}\text{N}$   $\text{A}$  and  $\text{B}$  comparisons are combined in silico to provide an indirect comparison between the two proteomes. The methodological concept followed in this study is shown in Fig. 1.



**Fig. 1** Principle of peptide/protein quantification using  $^{15}\text{N}$ -labeled internal standards

Protein quantification by <sup>15</sup>N metabolic labeling is characterized by high repeatability that depends on the complexity of the biological material analyzed [4]. Importantly, by using a <sup>15</sup>N-labeled internal standard, any existing <sup>15</sup>N isotope effect on the labeled proteome will not influence the quantification and data interpretation steps. We have previously used *in vivo* <sup>15</sup>N metabolic labeling to identify protein expression changes in mouse models of neuropsychiatric phenotypes [5–11].

## 2 Materials

### 2.1 <sup>15</sup>N Metabolic Labeling: Sample Acquisition

1. <sup>15</sup>N-labeled, bacteria-based diet U-15 N-SILAM-Mouse (Silantes GmbH; Munich, Germany) (*see Note 1*).
2. Standard mouse diet (Altromin; Lage, Germany).
3. 0.9% NaCl solution for perfusion.

### 2.2 Sample Preparation for Proteomic Analysis

1. Protein content estimation assay reagents (*see Note 2*).
2. Reagents and equipment for carrying out electrophoresis of biosamples (*see Note 3*).
3. 4× sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer: 125 mM Tris-HCl (pH 6.8), 9.2% SDS (w/v), 6% dithiothreitol (w/v), 40% glycerol (v/v), and 0.01% Bromophenol blue (w/v).
4. Costar SpinX tubes (Corning Inc; Glendale, AZ, USA).
5. Multiple Affinity Removal Spin cartridge mouse 3 (MARS3; Agilent Technologies; Santa Clara, CA, USA) (*see Note 4*).
6. Buffer A and Buffer B (Agilent Technologies).
7. Coomassie Brilliant Blue R-250.
8. Ammonium bicarbonate (Merck Millipore; Darmstadt, Germany).
9. Dithiothreitol (BioRad).
10. Iodoacetamide (BioRad).
11. Formic acid (Merck Millipore) (*see Note 5*).
12. Acetonitrile (Merck Millipore) (*see Note 6*).
13. Porcine trypsin (Promega; Madison, WI, USA).

### 2.3 Mass Spectrometry: Proteomics Data Analysis

1. LC buffer: 95% acetonitrile/0.1% formic acid.
2. In-house packed fused silica 3 μm RP-C18 (0.075 mm × 15–20 cm) chromatographic columns (Maisch; Monheim, Germany).
3. LC-MS/MS: nano HPLC-2D system (Eksigent; Dublin, CA, USA) coupled online to an LTQ-Orbitrap mass spectrometer via a nano electrospray ion source (ThermoFisher Scientific; Bremen, Germany).

4. Databases: BioWorks (v3.3.1) and SEQUEST (v28) (Thermo Fischer Scientific; San Jose, CA, USA).

---

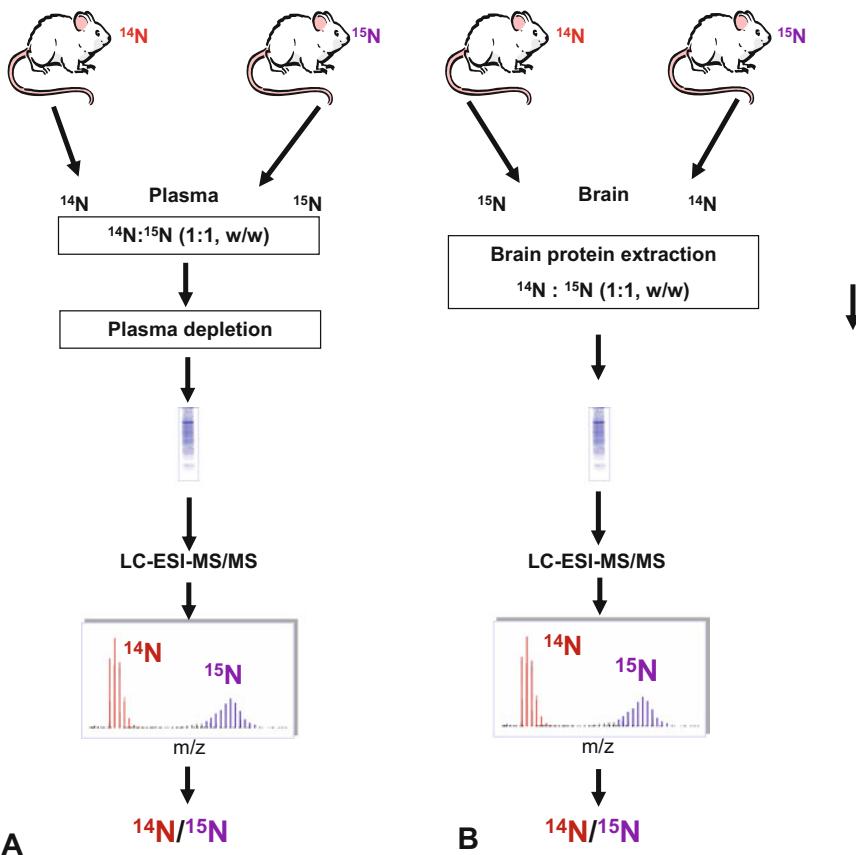
### 3 Methods

#### 3.1 $^{15}\text{N}$ -Metabolic Labeling: Sample Collection

1. Mice are mated and once pregnancy is detected, remove the male mice from the cages and subject the pregnant females to an ad libitum diet containing a 1:1 mixture of standard food and the  $^{15}\text{N}$ -labeled bacteria-based diet for 4 days.
2. Following this 4 day period, feed the pregnant mice exclusively with the  $^{15}\text{N}$ -labeled bacteria-based diet throughout pregnancy until weaning and then feed the offspring with the same diet for a total of 56 days post-partum [3] (*see Note 7*).
3. Anesthetize the  $^{15}\text{N}$ -labeled 2-month-old mice and the brain following perfusion with the NaCl solution.
4. Dissect brain areas of interest according to the mouse brain atlas [12].
5. Acquire plasma by centrifuging blood samples at  $13,000 \times g$  for 10 min at  $4^\circ\text{C}$ .
6. Collect the plasma supernatants, snap-freeze these in liquid nitrogen, and store at  $-80^\circ\text{C}$ .

#### 3.2 Sample Preparation for Proteomic Analysis

1. Plasma samples (Fig. 2a): For depletion of highly abundant plasma proteins, combine 400  $\mu\text{g}$  of  $^{14}\text{N}$  and  $^{15}\text{N}$ -labeled plasma protein samples at a 1:1 (w/w) ratio (*see Note 8*).
2. Dilute with 200  $\mu\text{L}$  of buffer A and filter using the Costar SpinX tubes according to the manufacturer's instructions.
3. Apply the flow through to the MARS3 cartridge according to the manufacturer's instructions.
4. Collect the cartridge flow through volume corresponding to the depleted proteome and estimate the protein concentration by Bradford assay (*see Note 9*).
5. Wash the cartridge to elute the bound proteins using buffer B (*see Note 10*).
6. Brain samples (Fig. 2b): Enrich brain region subproteomes if required using appropriate homogenization buffers and standard procedures.
7. For the enriched brain and depleted plasma proteomes, mix 50–150  $\mu\text{g}$  of the  $^{14}\text{N}/^{15}\text{N}$  protein extracts in the 4 $\times$  SDS-PAGE sample buffer to achieve a final 1x concentration, heat at  $95^\circ\text{C}$  for 5 min, and leave to cool at room temperature.
8. Load the  $^{14}\text{N}/^{15}\text{N}$  protein extracts onto mini 12.5% SDS-PAGE gels and resolve the proteins by gel electrophoresis (*see Note 11*).



**Fig. 2**  $^{14}\text{N}/^{15}\text{N}$  relative protein quantification workflow for mouse plasma and brain. **(a)** Plasma samples are mixed 1:1 based on protein content and the  $^{14}\text{N}/^{15}\text{N}$  mixture is depleted of three highly abundant plasma proteins. **(b)** Brain samples are either mixed 1:1 based on tissue weight and then the subproteome of interest is enriched or the subproteome of interest is extracted first and then the protein extracts are mixed 1:1 based on protein content. Protein mixtures are separated based on molecular weight by gel electrophoresis and analyzed by LC-MS/MS

9. Stain the protein bands in the gel with Coomassie.
10. Slice each gel lane into approximately 20 fractions and cut these further to produce smaller pieces.
11. Wash the gel pieces twice with 25 mM ammonium bicarbonate in 50% acetonitrile.
12. Perform reduction with 10 mM dithiothreitol in 25 mM ammonium bicarbonate for 30 min at 56 °C (*see Note 12*).
13. Perform carboxymethylation with 50 mM iodoacetamide in 25 mM ammonium bicarbonate for 30 min at room temperature (*see Note 13*).
14. Wash the gel pieces twice 25 mM ammonium bicarbonate in 50% acetonitrile and dry them in a vacuum centrifuge.

15. Rehydrate the gel pieces in 5 ng/ $\mu$ L trypsin in 25 mM ammonium bicarbonate using a volume that is just enough to cover the pieces.
16. Leave for 15 min on ice, remove the surplus trypsin solution, and cover the gel pieces with 25 mM ammonium bicarbonate.
17. Incubate the samples for 4–6 h at 37 °C (*see Note 14*).
18. Extract the resulting peptides with 50% acetonitrile/2% formic acid twice (*see Note 15*).
19. Lyophilize the peptide extracts and store them at –80 °C for mass spectrometry analysis (*see Note 16*).

### **3.3 Mass Spectrometry (LC-ESI-MS/MS): Proteomics Data Analysis**

1. Dissolve the peptides in 10  $\mu$ L 0.1% formic acid.
2. Load 5  $\mu$ L per fraction onto an in-house packed column and analyze using a 2.5 h gradient that includes washing with 0.1% formic acid for 20 min followed by elution with 95% acetonitrile/0.1% formic acid from 2 to 45% at a flow rate of 200 nl/min.
3. Operate the LTQ-Orbitrap in positive ion mode, using a data-dependent automatic scan switch between the MS and MS/MS acquisitions.
4. Record full scans in the mass analyzer with a m/z range of 380–1600 in profile mode.
5. Use a TOP5 method during which the MS/MS analysis of the five most intense peptide ions for each scan is recorded in the LTQ mass analyzer in centroid mode (*see Note 17*).
6. Search each raw file generated by mass spectrometry twice: once against a  $^{14}\text{N}$  and once against a  $^{15}\text{N}$  decoy mouse database using BioWorks (v3.3.1) and SEQUEST (v28) software (Thermo Fischer Scientific, San Jose, CA). Use the following parameters: peptide mass tolerance = 20 ppm; fragment ion mass tolerance = 1 Da; enzyme = trypsin; missed cleavage sites = 2; only tryptic peptides allowed; static modification = cysteine carboxyamidomethylation; variable modification = methionine oxidation.
7. Export  $^{14}\text{N}$  and  $^{15}\text{N}$  identified peptides as DTA files and use DTSelect [13] to filter the peptide identifications and assemble the peptides into proteins using the following filtering parameters: minimum delta CN = 0.08; cross correlation score (Xcorr) = 1.90 ( $z=1+$ ), 2.70 ( $z=2+$ ), 3.50 ( $z \geq 3+$ ); purging duplicate spectra on the basis of Xcorr ( $-t=2$ ); minimum charge state = 1; maximum charge state = 6; minimum of 1 identified peptide per protein (*see Note 18*).
8. Use ProRata software [14] for ion chromatogram extraction and relative  $^{14}\text{N}/^{15}\text{N}$  peptide and protein quantification (*see Note 19*).

9. Combine the two indirect <sup>14</sup>N/<sup>15</sup>N comparisons (**A**/<sup>15</sup>N and **B**/<sup>15</sup>N) to generate protein quantification differences between groups **A** and **B**.

---

## 4 Notes

1. The <sup>15</sup>N-labeled bacteria-based diet should be stored long term at 4 °C.
2. We use the Bradford assay although many protein assay kits can be used. The user should ensure compatibility of the reagents in the kit with those in the sample extracts.
3. We use reagents and equipment from Bio-Rad (Hercules, CA, USA) although many others can be used.
4. This is for depletion of the three abundant plasma proteins (albumin, IgG, and transferrin).
5. Other suppliers can be used although all reagents should be HPLC grade.
6. Other suppliers can be used although all reagents should be HPLC grade.
7. This feeding protocol typically results in greater than 90% <sup>15</sup>N incorporation rates into mouse brain and plasma. <sup>15</sup>N incorporation rates in a sample can be assessed by using the QuantiSpec [15] or Atomizer [16] software.
8. This can be done based on original tissue weight in the case of brain samples or on protein content for both brain and plasma samples. Protein concentrations can be determined using the Bradford assay. When <sup>14</sup>N and <sup>15</sup>N-labeled samples are mixed according to protein content, homogenization buffers compatible with the Bradford assay should be used. In case of analyzing a brain subproteome (e.g., synaptosomes), it is also advised to enrich the same subproteome for the <sup>15</sup>N-labeled brain to avoid dilution effects when using a whole <sup>15</sup>N-labeled brain region extract.
9. This step is recommended to expand the dynamic range of the plasma proteome analysis.
10. This step regenerates the column for future use.
11. In case a higher resolution is required for proteins of a certain molecular weight, gels with different percentages of acrylamide can be used.
12. This step reduces the disulfide bonds in proteins.
13. This step alkylates the reduced sulfide groups.
14. Digestion with trypsin overnight does not increase the number of generated peptides but results instead in a higher number of autolysis (self-digestion of trypsin) products.

15. To reduce the mass spectrometry run time if required, peptide extracts from individual gel slices can be combined and analyzed together in a single run.
16. Storage of lyophilized extracts in  $-80^{\circ}\text{C}$  is the most reliable long-term storage option to preserve sample quality.
17. All other parameters should be set as described previously [17, 18].
18. Manual validation of differential expression is necessary to avoid false positives.
19. In order to report protein groups instead of individual proteins, the Trans-Proteomic Pipeline (TPP) should be used for peptide filtering and grouping [19].

## Acknowledgments

This work was funded by the Max Planck Society and a BMBF *Quant Pro* grant. The authors thank Prof. Daniel Martins de Souza for insightful comments and discussions.

## References

1. Filiou MD, Martins-de-Souza D, Guest PC, Bahn S, Turck CW (2012) To label or not to label: applications of quantitative proteomics in neuroscience research. *Proteomics* 12:736–747
2. Filiou MD (2013) The potential of  $^{15}\text{N}$  metabolic labeling for schizophrenia research. *Arch Clin Psychiatr* 40:51–52
3. Frank E, Kessler MS, Filiou MD, Zhang Y, Maccarrone G, Reckow S et al (2009) Stable isotope metabolic labeling with a novel  $^{15}\text{N}$ -enriched bacteria diet for improved proteomic analyses of mouse models for psychopathologies. *PLoS ONE* 4, e7821
4. Filiou MD, Soukupova M, Rewerts C, Webhofer C, Turck CW, Maccarrone G (2015) Variability assessment of  $^{15}\text{N}$  metabolic labeling-based proteomics workflow in mouse plasma and brain. *Mol Biosyst* 11:1536–1542
5. Filiou MD, Teplytska L, Otte DM, Zimmer A, Turck CW (2009) Myelination and oxidative stress alterations in the cerebellum of the G72/G30 transgenic schizophrenia mouse model. *J Psychiatr Res* 46:1359–1365
6. Filiou MD, Zhang Y, Teplytska L, Reckow S, Gormanns P, Maccarrone G et al (2011) Proteomics and metabolomics analysis of a trait anxiety mouse model reveals divergent mitochondrial pathways. *Biol Psychiatry* 70:1074–1082
7. Zhang YY, Filiou MD, Reckow S, Gormanns P, Maccarrone G, Kessler MS et al (2011) Proteomic and metabolomic profiling of a trait anxiety mouse model implicate affected pathways. *Mol Cell Proteomics* 10:M111.008110.
8. Lopes S, Teplytska L, Vaz-Silva J, Dioli C, Trindade R, Moraes M et al. (2016) Tau deletion prevents stress-induced dendritic atrophy in prefrontal cortex: Role of synaptic mitochondria. *Cereb Cortex* Apr 12. pii: bhw057 (Epub ahead of print)
9. Filiou MD, Turck CW, Martins-de-Souza D (2011) Quantitative proteomics for investigating psychiatric disorders. *Proteomics Clin Appl* 5:38–49
10. Kao CY, He Z, Henes K, Asara JM, Webhofer C, Filiou MD et al (2016) Fluoxetine treatment rescues energy metabolism pathway alterations in a posttraumatic stress disorder mouse model. *Mol Neuropsychiatry* 2:46–59
11. Nussbaumer M, Asara JM, Teplytska L, Murphy MP, Logan A, Turck CW et al (2016) Selective mitochondrial targeting exerts anxiolytic effects in vivo. *Neuropsychopharmacology* 41:1751–1758
12. Paxinos G, Franklin K (2001) The mouse brain in stereotaxic coordinates, 2nd edn. Academic, San Diego, CA. ISBN 13: 978-0123887214

13. Tabb DL, McDonald WH, Yates JR 3rd (2002) DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* 1:21–26
14. Pan C, Kora G, McDonald WH, Tabb DL, VerBerkmoes NC, Hurst GB et al (2006) ProRata: a quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation. *Anal Chem* 78:7121–7131
15. Haegler K, Mueller NS, Maccarrone G, Hunyadi-Gulyas E, Webhofer C, Filiou MD et al (2009) QuantiSpec – quantitative mass spectrometry data analysis of <sup>15</sup>N-metabolically labeled proteins. *J Proteomics* 71:601–608
16. MacCoss MJ, Wu CC, Matthews DE, Yates JR 3rd (2005) Measurement of the isotope enrichment of stable isotope-labeled proteins using high-resolution mass spectra of peptides. *Anal Chem* 77:7646–7653
17. Filiou MD, Bisle B, Reckow S, Teplytska L, Maccarrone G, Turck CW (2010) Profiling of mouse synaptosome proteome and phosphoproteome by IEF. *Electrophoresis* 31:1294–1301
18. Filiou MD, Turck CW (2012) Psychiatric disorder biomarker discovery using quantitative proteomics. *Methods Mol Biol* 829:531–539
19. Keller A, Eng J, Zhang N, Li XJ, Aebersold R (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol* 1:2005.0017, Epub 2005 Aug 2

# Chapter 21

## Multiplex Measurement of Serum Folate Vitamers by UPLC-MS/MS

Sarah Meadows

### Abstract

The implementation of a liquid chromatography–tandem mass spectrometry (LC-MS/MS) method to measure six folate vitamers in serum samples allows a more individual and accurate measurement than the commonly used immunoassays. In the described method, serum samples undergo solid phase extraction followed by liquid chromatography coupled with electrospray ionization tandem mass spectrometry with a run time of 3.5 min. Recovery is 95% for the most important folate metabolite, 5-methyltetrahydrofolate (MTHF), and greater than 78% for other minor folate forms. The limit of detection ranges from 0.2 to 0.4 nmol/L with a intra-batch imprecision of less than 7% for all analytes and calibration ranges of 1–100 nmol/L for MTHF and 0.5–20 nmol/L for the minor folate forms, with greater than 0.99 R<sup>2</sup> linearity.

**Key words** Folate, 5-Methyltetrahydrofolate, Folic acid, Solid phase extraction, LC-MS/MS

---

### 1 Introduction

Folate is the general term for a water-soluble B vitamin naturally found in foods such as leafy vegetables, legumes, egg yolks, liver and some citrus fruits. Folic acid itself does not occur naturally but can be obtained from vitamin supplements or by eating fortified foods [1, 2]. A number of critical cellular pathways depend on folate, including DNA, RNA, and protein methylation as well as DNA synthesis and maintenance [2, 3]. Because of this, there are many health consequences of folate deficiency among all age groups. These include megaloblastic anemia, depression, cognitive impairment, low birth weight, risk of placental abruption, neural tube defects, and other birth-related abnormalities, including orofacial clefts and heart defects [4, 5].

The most widely publicized issue surrounding folate is that of low levels during pregnancy causing birth defects associated with the nervous system. Folate is an essential micronutrient during fetal development due to its importance in transmethylation reactions and synthesis of DNA in growing cells [3]. A significant

portion of the 300,000 neural tube defects (NTDs) that occur yearly worldwide continue to be a great public health burden although these are preventable by the periconceptual consumption of folic acid [2]. NTDs occur early in pregnancy (neural tube closure is completed by embryonic day 21–28) [6, 7] and arise when the tube fails to close properly leading to anencephaly, encephalocele and spina bifida. The demonstration that periconceptional supplementation with folic acid dramatically reduces the incidence of NTDs has generated considerable clinical and public health interest, which has led to fortification with folic acid of the food supply in the USA and some other countries [8]. The recent WHO guidelines for optimal serum and red blood cell folate levels in women of reproductive age for prevention of neural tube defects, states that in 2012 an estimated 270,358 deaths globally were attributable to congenital abnormalities during the first 28 days of life.

In addition to neural tube closure, B vitamins including folate are required for brain metabolic pathways and are fundamental in all aspects of brain development and maintenance of brain health throughout life. Observational and animal evidence appears to support a role for maternal folate status in later cognitive performance of the resulting children and there are also studies linking low maternal folate status with a higher incidence of behavioral and emotional problems, inattention, and hyperactivity in the offspring [7]. Recent studies have also shown links between low plasma folate and poor cognitive performances in children and adolescents, as well as a positive association between higher dietary folate intake and academic achievement [7].

Mental dysfunction in the elderly ranging from cognitive impairment to dementia is also a matter of concern. Brain changes occur long before the diagnosis of dementia is made and, given the current increases in life expectancy, the numbers of individuals suffering from this condition is expected to double by 2025. Therefore it is important to find early biomarkers which would enable timely interventions to delay the onset or slow the progression of the disease. There is emerging evidence suggesting that suboptimal status of folate and metabolically related B vitamins may be linked with cognitive dysfunction and dementia. If cognitive dysfunction can be slowed or prevented by improving the B vitamin status in healthy older people it could offer a cost effective preventative public health strategy in aging populations [7].

Evidence showing that supplementation with folic acid protects against NTDs has led to worldwide governmental recommendations advising all women planning a pregnancy to consume 400 µg of folic acid per day from the time of preconception until the end of the first trimester of pregnancy [2, 7, 9–11]. Even with this knowledge public health campaigns remain largely unsuccessful [9] since up to 50% of all pregnancies are unplanned in countries such as the USA [2]. Mandatory fortification programs have been implemented in many countries to improve folate status and reduce

the current high costs associated with prevention programs such as education campaigns and other interventions that require behavioral change [2]. The Scientific Advisory Committee on Nutrition has called for mandatory fortification in the UK to replace voluntary fortification in a bid to increase the country's folate status [10].

Despite the unequivocal success of folic acid in reducing NTD rates, several studies have questioned the harmful role of unmetabolized folic acid in blood [3, 11]. Concerns have been raised that due to fortification, the subsequent increase in folic acid intakes across the population may have harmful effects on health, such as the masking of pernicious anemia, colorectal cancer promotion in people with preexisting lesions or adverse cognitive effects in the elderly with low vitamin B<sub>12</sub> status [4, 9, 11, 12]. Measurement of unmetabolized folic acid has been suggested as a way of monitoring folic acid intake [2, 5]. At a time when there are still questions around the safety of high levels of folic acid in the blood, the ability to differentiate between this and endogenous folates is valuable.

Folate has traditionally been measured using a microbiological assay but since the 1970s commercial protein binding assays on automated clinical analysers have been widely employed due to the ease of use of these platforms and the increased throughput they offer. The microbiological assays are considered more accurate as they can recover folate vitamers equally [13]. In contrast the protein binding assays generally underestimate folate concentrations due to the different affinities of the folate vitamers for the antibody used [1, 6]. Serum folate is considered to be an indicator of recent folate intake whereas red blood cell folate concentrations indicate long term status [6, 10]. Red cell folate is normally calculated using whole blood folate, serum folate, and hemocrit. However, repeated low concentrations of serum folate occurring within an individual over the course of a month are also indicative of folate depletion [6, 14]. It has proven to be more technically challenging to measure whole blood folate by liquid chromatography tandem mass spectrometry (LCMS/MS) than it is to measure serum folate. This is because red blood cells require hemolysis to release the polyglutamate folates, which then need to be deconjugated to monoglutamates without any folate loss prior to analysis [6, 8]. This has led to whole blood tests only being carried out in specialist laboratories which mostly use microbiological assays.

In contrast to both the microbiological assay and the radioisotope protein-binding assay, chromatography techniques are able to differentiate between individual folate species [6] and are now often coupled to mass spectrometers with high sensitivity, specificity, and selectivity compared to other detection methods [6, 8, 15]. The advantage of measuring the different folate species is likely to become greater in the future as more information on genetic polymorphisms that affect nutritional status and folate distributions become available [1]. Initial steps to standardize folate methods

began with the development of higher order reference methods that use isotope dilution with LC-MS/MS analysis. In addition, recent advances in sample clean up procedures have led to increasing use of LC/MS or LC-MS/MS procedures [5].

The method described here is for a routine LC-MS/MS analysis for determination and quantitation of six folate vitamers in serum: 5-methyltetrahydrofolate (MTHF), folic acid, tetrahydrofolate (THF), 5-formyltetrahydrofolate (FTHF), 5, 10 methenyltetrahydrofolate ( $\text{CH} + \text{THF}$ ), and pyrazino-s-triazine derivative of  $4\alpha$ -hydroxy-5-methyl tetrahydrofolate (MeFox).

---

## 2 Materials

### 2.1 Specialized Equipment

1. Waters HSS T3  $1.8 \mu 2.1 \times 100$  mm with Waters Acuity HSS T3 Vanguard precolumn (*see Note 1*).
2. AB Sciex 5500 Qtrap mass spectrometer.
3. 96-well plate positive pressure processor.
4. 96-well plate vacuum manifold and pump.
5. 50  $\mu\text{g}$  phenyl 96-well SPE Plates.
6. PVDF 96-well filter plates.

### 2.2 Reagents

1. Stock 10% ammonium formate, pH 3.2 (adjust using ammonium hydroxide). Store at room temperature for up to 6 months (*see Note 2*).
2. Conditioning buffer: 1% ammonium formate pH 3.2. Make up fresh for each assay.
3. Sample diluent: 1% ammonium formate pH 3.2, 0.5% ascorbic acid. Make up fresh for each assay.
4. Wash buffer: 0.05% ammonium formate pH 3.4, 0.1% ascorbic acid. Make up fresh for each assay.
5. Mobile phase solvent: water–methanol–acetonitrile (4.95/4.0/1.0). Degas using sonication. Keep for up to 2 weeks at room temperature.
6. Working mobile phase: 5 mM ammonium formate, 0.5% acetic acid in mobile phase solvent. Make up fresh for each assay.
7. Elution buffer: 0.5% acetic acid, 0.5% ascorbic acid in mobile phase solvent. Make up fresh for each assay.
8. Working standard diluents: phosphate buffered saline (PBS) pH 7.2, 4% albumin, 0.5% ascorbic acid. Make up fresh each assay.
9. Strong needle wash: 100% methanol.
10. Weak needle wash: 65% acetonitrile.
11. Seal wash: 10% methanol.

**2.3 Standards and Internal Standards**  
**(See Notes 2–6 and Table 1)**

1. 0.1 M potassium phosphate pH 7.2. Store at room temperature for up to 1 month.
2. 20–40 µg/mL folic acid/<sup>13</sup>C5 folic acid (**Stock I**). Remove 1 mL for spectrophotometric analysis and store the remaining solution in 5 mL aliquots in cryovials at –80 °C for up to 5 years. Calculate the exact concentration from the absorbance reading on the spectrophotometer and make 20 µM solutions (**Stock II**). Store in 300 µL aliquots in cryovials at –80 °C for up to 1 year (see Notes 3–5).
3. 200–400 µg/mL MeFox / <sup>13</sup>C5 MeFox (**Stock I**). Remove 1 mL for spectrophotometric analysis and store the remaining solution at 4 °C. Calculate the exact concentration from the absorbance reading on the spectrophotometer and make 100 µg/mL (**Stock II**) and 20 µM (**Stock III**) solutions (see Table 2) Store in 1.2 mL aliquots in cryovials at –80 °C for up to 5 years.
4. 200–400 µg/mL MTHF/<sup>13</sup>C5 MTHF, FTHF/<sup>13</sup>C5 FTHF, THF/<sup>13</sup>C5 THF, and 5,10-methenylTHF/<sup>13</sup>C5 5,10-methenylTHF (**Stock I**). Remove 1 mL for spectrophotometric analysis. To the remaining solution add 1% ascorbic acid and store at 4 °C. Calculate the exact concentration from the absorbance reading on the spectrophotometer (see Notes 6 and 7) and make 100 µg/mL (**Stock II**) and 20 µM (**Stock III**) solutions (see Table 2). Store in 1.2 mL aliquots in cryovials at –80 °C for up to 5 years.
5. Mixed standard: 2 µM MTHF and 0.4 µM of other folates in sample diluents. Make up fresh for each assay, using a new set of stock aliquots each time.

**Table 1**  
**Diluents required for standard preparation**

| Analyte/labeled analyte | Stock I diluent                              | Stock II diluent            | Stock III diluent               |
|-------------------------|--|-----------------------------|---------------------------------|
| MTHF                    | 20 mM phosphate buffer pH7.2 + 0.1% cysteine | 1 % aqueous ascorbic acid   | 0.5 % aqueous ascorbic acid     |
| Folic Acid              | 20 mM phosphate buffer pH7.2                 | Deionized water             | N/A                             |
| FTHF                    | 20 mM phosphate buffer pH7.2 + 0.1% cysteine | 1 % aqueous ascorbic acid   | 0.5 % aqueous ascorbic acid     |
| THF                     | 20 mM phosphate buffer pH7.2 + 0.1% cysteine | 1 % aqueous ascorbic acid   | 0.5 % aqueous ascorbic acid     |
| CH + THF                | 1 M HCl                                      | 1 M HCl + 1 % ascorbic acid | 0.5 M HCl + 0.5 % ascorbic acid |
| MeFox                   | 0.1 M NaOH                                   | Deionized water             | 0.1 % aqueous ascorbic acid     |

**Table 2**  
**Composition of stock II and III standard solutions**

| Analyte                   | Volume of Stock II ( $\mu\text{L}$ ) |
|---------------------------|--------------------------------------|
| MTHF                      | 919                                  |
| FTHF                      | 947                                  |
| THF                       | 891                                  |
| CH + THF                  | 911                                  |
| MeFox                     | 947                                  |
| $^{13}\text{C}5$ MTHF     | 929                                  |
| $^{13}\text{C}5$ FTHF     | 957                                  |
| $^{13}\text{C}5$ THF      | 901                                  |
| $^{13}\text{C}5$ CH + THF | 923                                  |
| $^{13}\text{C}5$ MeFox    | 957                                  |

**Table 3**  
**Volume of stock II required to make 20 $\mu\text{M}$  stock III solutions**

| Standard        | Conc of MTHF/other folates (nmol/L) | Standard to dilute | Volume standard ( $\mu\text{L}$ ) | Volume of PBS + 0.4% albumin ( $\mu\text{L}$ ) |
|-----------------|-------------------------------------|--------------------|-----------------------------------|--|
| S1 Top Standard | 100/20                              | Mixed standard     | 50                                | 950  |
| S2              | 75/15                               | Mixed standard     | 30                                | 770  |
| S3              | 50/10                               | Mixed standard     | 25                                | 975  |
| S4              | 25/5                                | Mixed standard     | 12.5                              | 987.5  |
| S5              | 10/2                                | Mixed standard     | 10                                | 1990   |
| S6              | 5/1                                 | S1                 | 50                                | 950  |
| S7              | 1/0.2                               | S1                 | 10                                | 990  |

6. Working standards. Make up fresh each assay as shown in Table 3.
  7. Working internal standard: 60 nM  $^{13}\text{C}5$  MTHF and 20 nM all other  $^{13}\text{C}5$  folates in 0.1% ascorbic acid. Make up fresh for each assay, using a new set of stock aliquots each time (Table 4).
- 2.4 Quality Control (QC) Material**
1. In house pooled anonymized serum/plasma: (a) 10–15 nM and 1–2 nM MTHF and folic acid, respectively; and (b) 20–30 nM and 8–12 nM MTHF and folic acid respectively. Dispense into 350  $\mu\text{L}$  aliquots in amber glass UPLC vials

**Table 4**  
**Parameters for calculating folate concentration from absorbance readings**

|                             | MTHF/ <sup>13</sup> C5<br>MTHF | Folic acid/ <sup>13</sup> C5<br>folic acid | THF/ <sup>13</sup> C5<br>THF | FTHF/ <sup>13</sup> C5<br>FTHF | CH + THF/ <sup>13</sup> C5<br>CH + THF | MeFox/ <sup>13</sup> C5<br>MeFox |
|-----------------------------|--------------------------------|--|------------------------------|--------------------------------|--|----------------------------------|
| MW                          | 459.4/464.4                    | 441.4/446.4                                | 445.4/450.4                  | 473.4/478.4                    | 455.5/460.5                            | 473.4/478.4                      |
| $\lambda_{\text{max}}$ (nm) | 290                            | 282 and 346                                | 298                          | 285                            | 288 and 348                            | 280                              |
| Absorption coefficient      | 31700                          | 27600@282<br>7200@346                      | 25000                        | 37200                          | 13500@288<br>26500@348                 | 19365                            |

(excess can be stored in glass scintillation vials and aliquoted as required) and store at –80 °C for up to 5 years.

2. QC spike solution: 4 µM MTHF and 0.4 µM all other folates, 0.1 % ascorbic acid.
3. Spiked serum/plasma from anonymized donor samples: Create a pool with low endogenous levels of folates and use the spike mix to create approximate spiked concentrations of 50 nM MTHF and 5 nM all other folates. Store in 350 µL aliquots in amber glass UPLC vials (excess pooled plasma can be stored in glass scintillation vials and aliquoted as required) at –80 °C for up to 5 years.
4. Standard reference material 1950: purchased from the National Institute of Standards and Technology (NIST) in 1 mL vials with known values of MTHF and folic acid of 27 nM and 3.5 nM respectively. Store at –80 °C (*see Note 8*).

### 3 Methods

#### 3.1 Sample Extraction

1. Make up a worksheet to include the reagent blank, diluent (S0), working standards (S1–S7), QCs, and samples (*see Notes 9 and 10*).
2. Use separate 12 × 75 mm borosilicate glass tubes for reagent blank, standards, QCs, and samples.
3. Add 150 µL standard, QC or sample into the tube (*see Note 11*).
4. Add 50 µL working internal standard into each tube except the reagent blank.
5. Add 200 µL water into the reagent blank tube.
6. Add 100 µL methanol into each tube.
7. Cover the tubes with Parafilm and mix for 5 min on a multi-tube vortexer.
8. Centrifuge at 1650 × g for 5 min.

9. Decant the supernatant from each tube into a second set of labeled borosilicate glass tubes.
10. Pipette 350 µL sample diluent into each tube.
11. Cover the tubes with Parafilm and mix for 30 s on a multi-tube vortexer.
12. Let stand for 20 min at room temperature (*see Note 12*).
13. Condition SPE columns on the positive pressure processor with 500 µL acetonitrile, 500 µL methanol, and 2×650 µL conditioning buffer. Push each addition through the column (*see Note 13*).
14. Load 550 µL of sample into each well of the SPE plate. Stand for 2 min
15. Push the samples through using minimal pressure (*see Note 14*).
16. Wash the samples with 2×650 µL of wash buffer. Push through using minimal pressure.
17. Elute the samples into a 96-well filter plate using 2×150 µL elution buffer. Turn the pressure down to zero and up slowly until the liquid just starts to drip through (*see Note 14*). The elution buffer needs to drip through slowly for optimum elution.
18. Filter the samples under vacuum into a 2 mL 96-well plate.
19. Seal the plate.
20. Mix on the plate shaker for 10 min.
21. Load onto the UPLC and inject 15 µL into the LC-MS instrument (*see Fig. 1* for an example chromatogram).

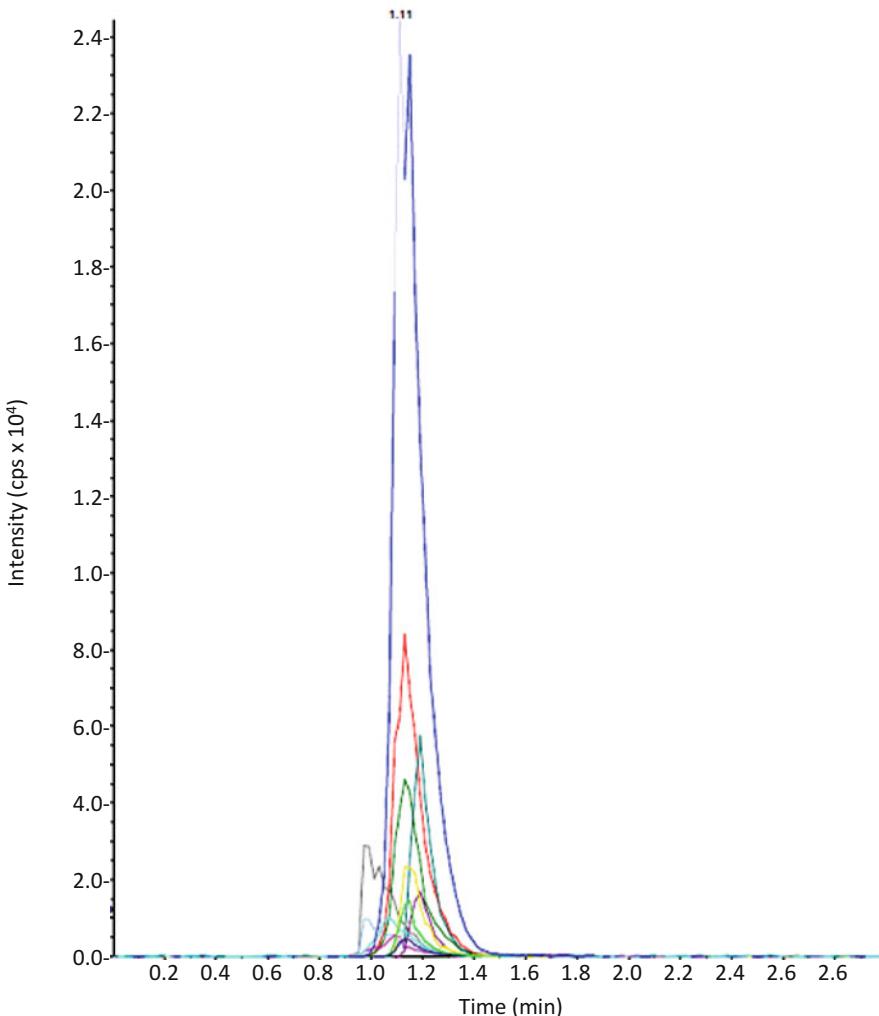
### **3.2 UPLC Setup for Waters Acquity H Class**

Set running conditions as follows:

1. Autosampler temp: 7 °C.
2. Column oven: 30 °C.
3. Mobile phase: isocratic.
4. Flow: 250 µL/min.
5. Run time: 3.5 min (*see Note 15*).

### **3.3 Mass Spectrometer Setup for AB Sciex 4000**

1. Analyte parameters (Table 5) (*see Note 16*).
2. Polarity: ESI in positive mode.
3. Curtain gas: 10.
4. Collision gas: high.
5. Ion spray voltage: 5500.
6. Temperature: 650.
7. Ion source gas 1: 55.
8. Ion source gas 2: 45.
9. Interface heater: on.



**Fig. 1** Example chromatogram

---

#### 4 Notes

1. Comparable C8 LC columns manufactured by Phenomenex and Waters with the same particle size, diameter, and length yield slightly poorer chromatography for THF. Only consider using these if the minor folate forms are of no interest.
2. Make each analyte and its corresponding internal standard up on the same day. Only make one solution at a time and clean the bench area thoroughly with water between each analyte and its labeled internal standard to prevent cross-contamination. Use designated volumetric flasks for all the analytes and internal standards to prevent contamination. These should be rinsed

**Table 5**  
**Analyte parameters**

| Analyte                     | Transition          | Declustering potential | Collision energy | Collision cell exit potential | Entrance potential |
|-----------------------------|---------------------|------------------------|------------------|-------------------------------|--------------------|
| MTHF                        | 460.2 > 131.2/180.1 | 56                     | 29/51            | 18/12                         | 10                 |
| Folic Acid                  | 442.3 > 295.1/176.0 | 61                     | 21/57            | 18/8                          | 10                 |
| FTHF                        | 474.4 > 299.2/165.8 | 61                     | 45/65            | 16/28                         | 10                 |
| THF                         | 446.2 > 299.2/120.1 | 61                     | 27/55            | 22/6                          | 10                 |
| CH + THF                    | 456.2 > 412.3/482.0 | 96                     | 41/69            | 10/14                         | 10                 |
| MeFox                       | 474.2 > 284.2/132.1 | 61                     | 53/73            | 15/22                         | 10                 |
| <sup>13</sup> C5 MTHF       | 465.2 > 313.1       | 81                     | 25               | 8                             | 10                 |
| <sup>13</sup> C5 Folic Acid | 447.2 > 295.4       | 56                     | 19               | 16                            | 10                 |
| <sup>13</sup> C5 FTHF       | 479.2 > 299.2       | 76                     | 47               | 16                            | 10                 |
| <sup>13</sup> C5 THF        | 451.1 > 299.1       | 51                     | 29               | 17                            | 10                 |
| <sup>13</sup> C5 CH + THF   | 461.3 > 416.3       | 106                    | 41               | 12                            | 10                 |
| <sup>13</sup> C5 MeFox      | 479.2 > 284.4       | 76                     | 53               | 22                            | 10                 |

with deionized water after use, left to drain and stored until needed again. DO NOT USE DETERGENT.

3. The absorbance of the stock I standards is measured at  $\frac{1}{2}$  and  $\frac{1}{5}$  dilution for folic acid and  $\frac{1}{20}$  and  $\frac{1}{50}$  dilution for all the other folates using the same diluent as in the stock solution spectrophotometrically at their respective  $\lambda_{\text{max}}$  and concentration is calculated using the following formula (Table 4):

$$\text{Conc}(\mu\text{g / mL}) = (\text{abs} \times \text{Dilution factor} \times 1000 \times \text{MW}) \times \text{molar absorption coefficient}$$

For MTHF, THF, FTHF, CH + THF and MeFox and their corresponding internal standards, once the concentration has been determined, make as much of Stock I into Stock II as possible and discard the rest.

4. Folate solids that come as a calcium salt may need glacial acetic acid to aid dissolution in the buffer. Add 5  $\mu\text{L}$  at a time and mix well after each addition until the salt is dissolved.
5. Folic acid is already oxidized and not very soluble at low pH so ascorbic acid is not required when making the standard solutions.
6. 5,10 methenylTHF stock I dissolves better if the flask is heated. We find using a beaker of hot water sufficient.

7. 5-Methyl THF and its labeled analog should be measured at 290 nm and 245 nm and the ratio calculated (290/245) to ensure there is no oxidation to dihydrofolate. The ratio should not exceed 3.3
8. We run SRM 1950 along with in house QCs as there is no commercial QC material available with LC-MS/MS assigned values.
9. We run QCs after the standards and at the end of the plate with the exception of the SRM which is only run once on the plate, at the beginning as an accuracy check. No drift was observed across the plate so duplicate analysis is unnecessary.
10. All standards, QCs, and samples are run singly. With the QC protocol described above this leaves enough space for 80 samples on a full 96-well plate.
11. Minimize the time samples are out at room temperature defrosting as there is a loss of folates in the sample at prolonged times.
12. Do not leave the diluted samples standing on the bench for longer than 30 min as there is a loss of folates in the sample at prolonged room temperature.
13. Do not keep the air / nitrogen running any longer than necessary when pushing liquids through the columns to prevent the columns from drying out.
14. The samples should slowly drip through the column, I look for about one drip per second. If the samples are eluted too quickly the folates will not bind to the columns reducing the sensitivity of the assay.
15. Washing the column with 65% acetonitrile for 1.5 min after each injection prolongs the column life and minimizes back pressure increases during the run.
16. A second transition is only routinely measured for MTHF as the sensitivity is too low for the other analytes. If a peak of >5 nmol/L for folic acid, THF, FTHF, and CH<sub>2</sub>THF or >10 nmol/L for MeFox is observed, then the sample is immediately reinjected along with the standards and a set of QCs using a method with the full set of transitions for all analytes. In this way a quantifier peak is obtained with which to double check the initial result.

---

## Acknowledgments

This work was funded by the Medical Research Council  
MRC\_MC\_U105960384

## References

1. Shane B (2011) Folate status assessment history: implications for measurement of biomarkers in NHANES. *Am J Clin Nutr* 94:337S–342S
2. Crider KS, Bailey LB, Berry RJ (2011) Folic acid food fortification—its history, effect, concerns, and future directions. *Nutrients* 3:370–384
3. Obeid R, Kasoha M, Kirsch SH, Munz W, Herrmann W (2010) Concentrations of unmetabolized folic acid and primary folate forms in pregnant women at delivery and in umbilical cord blood. *Am J Clin Nutr* 92:1416–1422
4. Smith AD (2007) Folic acid fortification: the good, the bad, and the puzzle of vitamin B-12. *Am J Clin Nutr* 85:3–5
5. de Benoist B (2008) Conclusions of a WHO Technical Consultation on folate and vitamin B12 deficiencies. *Food Nutr Bull* 29(2 Suppl):S238–S244
6. WHO (2015) Guideline: optimal serum and red blood cell folate concentrations in women of reproductive age for prevention of neural tube defects. WHO, Geneva. ISBN 9789241549042
7. McGarel C, Pentieva K, Strain JJ, McNulty H (2015) Emerging roles for folate and related B-vitamins in brain health across the lifecycle. *Proc Nutr Soc* 74:46–55
8. Bailey LB (2009) Folate in health and disease, 2nd edn. CRC, Boca Raton, FL. ISBN 10: 1420071246
9. Hopkins SM, Gibney MJ, Nugent AP, McNulty H, Molloy AM, Scott JM et al (2015) Impact of voluntary fortification and supplement use on dietary intakes and biomarker status of folate and vitamin B-12 in Irish adults. *Am J Clin Nutr* 101:1163–1172
10. Clarke R, Bennett D (2014) Folate and prevention of neural tube defects. *BMJ* 349:g4810
11. <http://www.efsa.europa.eu/en/events/event/corporate090121>
12. Bailey RL, Mills JL, Yetley EA, Gahche JJ, Pfeiffer CM, Dwyer JT et al. (2012) Serum unmetabolized folic acid in a nationally representative sample of adults  $\geq 60$  years in the United States, 2001–2002. *Food Nutr Res* 56. doi: 10.3402/fnr.v56i0.5616. Epub 2012 Apr 2
13. Pfeiffer CM, Hughes JP, Lacher DA, Bailey RL, Berry RJ, Zhang M et al (2012) Estimation of trends in serum and RBC folate in the U.S. population from pre- to postfortification using assay-adjusted data from the NHANES 1988–2010. *J Nutr* 142:886–893
14. McDowell MA, Lacher DA, Pfeiffer CM, Mulinare J, Picciano MF, Rader J et al (2008) Blood folate levels: the latest NHANES results. *NCHS Data Brief* 6:1–8
15. Wang X, Zhang T, Zhao X, Guan Z, Wang Z, Zhu Z et al (2014) Quantification of folate metabolites in serum using ultraperformance liquid chromatography tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 962:9–13

# Chapter 22

## UPLC-MS/MS Determination of Deuterated 25-Hydroxyvitamin D ( $d_3$ -25OHD $_3$ ) and Other Vitamin D Metabolites for the Measurement of 25OHD Half-Life

Shima Assar, Inez Schoenmakers, Albert Koulman, Ann Prentice, and Kerry S. Jones

### Abstract

Plasma 25-hydroxyvitamin D (25OHD) half-life (25OHD $t_{1/2}$ ) is a dynamic marker of vitamin D metabolism that can be used to assess vitamin D expenditure and help inform vitamin D requirements. Our group recently established an approach to determine the 25OHD $t_{1/2}$  as an alternative biomarker of 25OHD expenditure in humans. The approach uses a small oral dose of stable isotope labeled 25OHD $_3$  [ $3\text{-}^2\text{H}$ -25OHD $_3$ ] (6,19,19-d $_3$ ) ( $d_3$ -25OHD $_3$ ) (tracer), which is distinguishable from endogenous 25OHD by liquid chromatography tandem-mass spectrometry (LC-MS/MS). We report here the method, which relies on protein precipitation, purification with solid phase extraction, derivatization with 4-phenyl-1,2,4-triazoline-3,5-dione (PTAD), and determination of the compounds by isotope-dilution UPLC-MS/MS. The method proved to be rapid and sensitive (LOQ 0.2 nmol/L) for the quantification of this tracer as well as the other vitamin D metabolites: 25OHD $_3$ , 25OHD $_2$ , and 24,25(OH) $_2$ D $_3$  in human plasma.

**Key words** Vitamin D, Stable isotope of 25OHD $_3$ , 25OHD $_3$ , 25OHD $_2$ , 24,25(OH) $_2$ D $_3$ , Biomarker, PTAD derivatization, Liquid chromatography mass spectrometry

---

### 1 Introduction

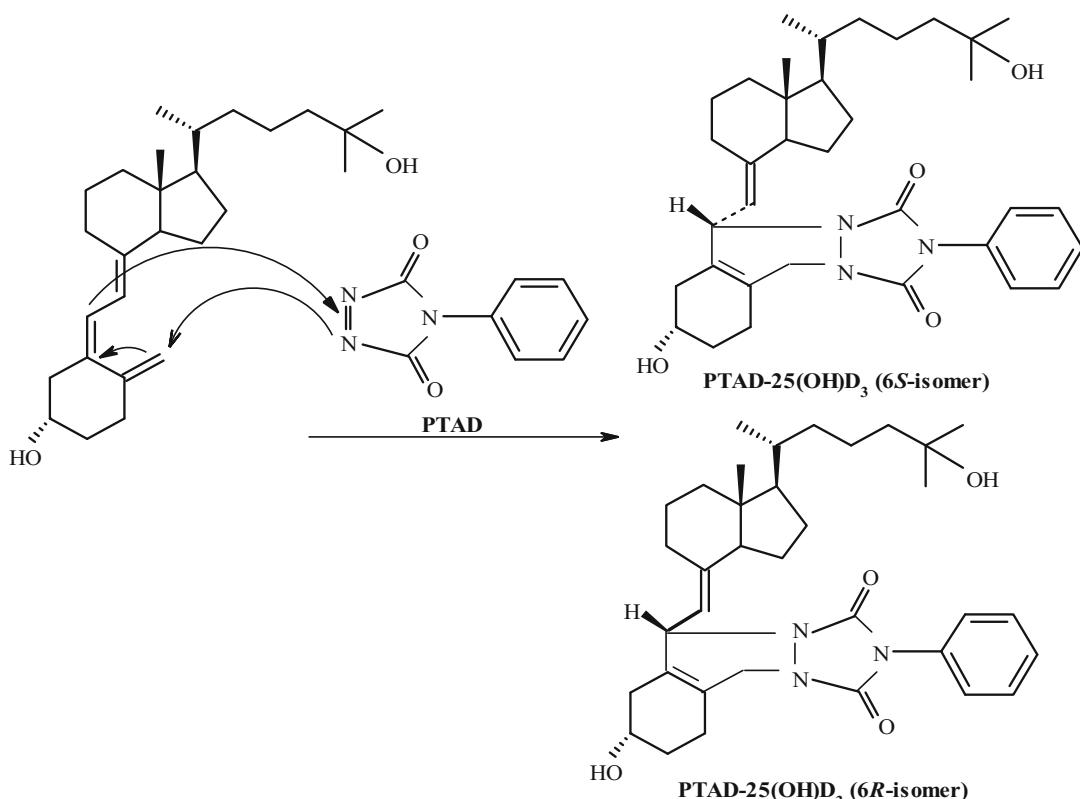
25-hydroxyvitamin D (25OHD) is the most widely assessed biomarker of vitamin D status in biological samples. This metabolite has relatively high serum concentration that is not under tight homeostatic control and a relatively long half-life (2–3 weeks) [1–4]. 25OHD concentration reflects vitamin D supply from cutaneous synthesis, diet, and usage through metabolism, excretion, and sequestration into body tissues [5, 6]. To investigate the link between the vitamin D supply and tissue requirements with vitamin D metabolism, our group has proposed the plasma half-life of 25OHD as a dynamic biomarker of vitamin D expenditure [5]. The plasma half-life of 25OHD has been measured previously

using radiolabeled and unlabeled compounds and has been reviewed by Jones et al. [7]. The use of radiolabeled compounds is no longer considered acceptable for human use, particularly in children, pregnant women, and other vulnerable groups. The use of unlabeled compounds is not recommended because it is not possible to distinguish the administered dose from endogenous vitamin D metabolites and large doses are required that may perturb normal vitamin D metabolism [7]. The use of stable isotope labeled compounds together with mass spectrometry is the preferred choice for research of vitamin absorption and metabolism [8]. Recently, stable isotope labeled vitamin D compounds have become available for research. We designed and performed studies to measure 25OHD half-life ( $25\text{OHD}_{t_{1/2}}$ ) using an oral dose of stable isotope labeled 25OHD ( $^2\text{H}-25\text{OHD}_3$  (6,19,19-d<sub>3</sub>), (d<sub>3</sub>-25OHD<sub>3</sub>), as a tracer in humans [7, 9, 10].

A number of experimental factors may affect the usefulness of tracer methods to determine half-life [7]. Two points are of relevance here. Firstly, quantification of the tracer compound should be highly specific and analytical methods should distinguish tracer from endogenous compounds. Secondly, in order not to perturb normal metabolism, the tracer dose should not exceed more than 10% of endogenous levels and thus high sensitivity is required. This can be achieved through the coupling of ultra-performance (sub 2 μm particle size columns) liquid chromatography with electrospray-based tandem mass spectrometry (UPLC-MS/MS).

LC-MS/MS is frequently referred to as the “gold standard” method for the determination of the 25OHD concentration in biological samples. However, the lipophilic nature, low ionization efficiency of the vitamin D metabolites, low abundance in human plasma, and potential interferences from co-eluting isobaric compounds make development of a LC-MS/MS method challenging [11, 12]. To overcome these issues, derivatization with Cookson-type triazoline-diones (TADs) reagent has been reported for the analysis of vitamin D metabolites. These reagents introduce polar groups into the analytes and therefore increase the ionization efficiency, sensitivity, and specificity for the analysis of vitamin D compounds [13, 14]. 4-phenyl-1,2,4-triazoline-3,5-dione (PTAD) (Fig. 1) is a readily available and cost-effective derivatizing reagent applicable to the analysis of vitamin D.

Here, we describe a fast and sensitive LC-MS/MS method that simultaneously determines low abundance tracer ( $^2\text{H}-25\text{OHD}_3$ ) and 25(OH)D<sub>3</sub>, 25(OH)D<sub>2</sub> and 24,25(OH)<sub>2</sub>D<sub>3</sub> in human plasma and is applicable to tracer experiments of vitamin D metabolism. In summary, plasma samples were subjected to protein precipitation with acetonitrile and followed by purification with solid-phase extraction. The purified samples were derivatized with PTAD solution prepared in acetonitrile and derivatized samples were analyzed on an UPLC system interfaced to a triple quadrupole mass spectrometer.



**Fig. 1** Derivatization of  $25\text{OHD}_3$  with PTAD derivatization reagent. Adapted from [15]

## 2 Materials

### 2.1 Sample Pretreatment

- Human plasma samples (*see Note 1*).
- Acetonitrile (AcN) (*see Note 2*).
- Ethyl acetate.
- Methanol (MeOH).
- 30% methanol (*see Note 3*).
- 4-phenyl-1,2,4-triazoline-3,5-dione (PTAD) (Sigma-Aldrich, Gillingham, UK) solution: 0.5 mg in 1 mL AcN (*see Note 4*).
- Oasis HLB SPE cartridge (1 mL, 30 mg sorbent, Waters, Elstree, UK).
- SPE vacuum manifold.
- $\text{N}_2$  sample concentrator.

### 2.2 Standards

- $25\text{OHD}_3$ ,  $25\text{OHD}_2$ ,  $24,25(\text{OH})_2\text{D}_3$ ,  $d_3\text{-}25\text{OHD}_3$  (Sigma-Aldrich).
- $d_6\text{-}25\text{OHD}_3$ ,  $d_6\text{-}24,25(\text{OH})_2\text{D}_3$  (Sigma-Aldrich).

**2.3 UPLC-MS/MS**

1. 0.1% formic acid (FA) in solvent A and B (below).
2. 5 mM monomethylamine (MMA) in solvent A and B (below).
3. Solvent A: 0.1% FA, 5 mM MMA in H<sub>2</sub>O.
4. Solvent B: 0.1% FA and 5 mM MMA in mixture of MeOH:H<sub>2</sub>O:AcN (97:2:1, v:v:v).
5. Strong/weak and needle wash solvent: AcN:H<sub>2</sub>O (1:1, v/v).
6. 0.2 µm Supelco Nylon 66 membrane (Supelco, Bellefonte, PA, USA).
7. 200 µL utosampler glass micro insert.
8. 2 mL autosampler amber glass vials with cap and septa.
9. Acquity UPLC™ BEH C18 column (2.1 × 100 mm, 1.7 µm) (Waters Corporation; Milford, MA, USA).
10. Acquity UPLC module (Waters Corporation).
11. 5500 QTRAP quadrupole-linear ion trap mass spectrometer (AB Sciex; Concord, ON, Canada).

**3 Methods****3.1 Preparation of Standard and Internal Standard Solutions**

1. Prepare main stock solutions of each standard (i.e., 25OHD<sub>3</sub>, 25OHD<sub>2</sub>, 24,25(OH)<sub>2</sub>D<sub>3</sub>, and d<sub>3</sub>-25OHD<sub>3</sub>) and internal standard (i.e., d<sub>6</sub>-25OHD<sub>3</sub> and d<sub>6</sub>-24,25(OH)<sub>2</sub>D<sub>3</sub>) by dissolving the compounds in AcN and store at -20 °C until use.
2. Prepare working solutions of each standard and internal standard by making serial AcN dilutions of the main stock solutions to obtain the required concentration for the calibration curves (*see Note 5*).
3. Prepare six external standards containing the analytes of interest with different ranges of concentrations (*see Note 5*) and their corresponding internal standards (*see Note 6*) to generate the standard curves.
4. Dry the standard solutions under gentle stream of nitrogen gas using sample concentrator and apply for derivatization process.
5. Prepare quality control material (*see Note 7*).

**3.2 Pretreatment of Plasma**

1. Transfer 250 µL plasma to a 2 mL Eppendorf tube and spike with internal standard solution containing d<sub>6</sub>-25OHD<sub>3</sub> and d<sub>6</sub>-24,25(OH)<sub>2</sub>D<sub>3</sub> (*see Note 6*).
2. Mix for a few seconds using a mixer vortex and incubate the spiked plasmas at room temperature for 1 h on a multi-tube vortexer.

### 3.3 Protein Precipitation

1. Add 600  $\mu$ L AcN to spiked plasma and mix for approximately 1 min using a vortexer.
2. Centrifuge the mixture at  $10,000 \times g$  for 15 min to precipitate the protein.
3. Transfer the supernatant to a 2 mL Eppendorf tube and evaporate AcN under a gentle flow of nitrogen gas at room temperature to a volume of around 200  $\mu$ L.
4. Add 800  $\mu$ L of deionized H<sub>2</sub>O to remaining solution and mix using a vortexer.

### 3.4 Solid Phase Extraction

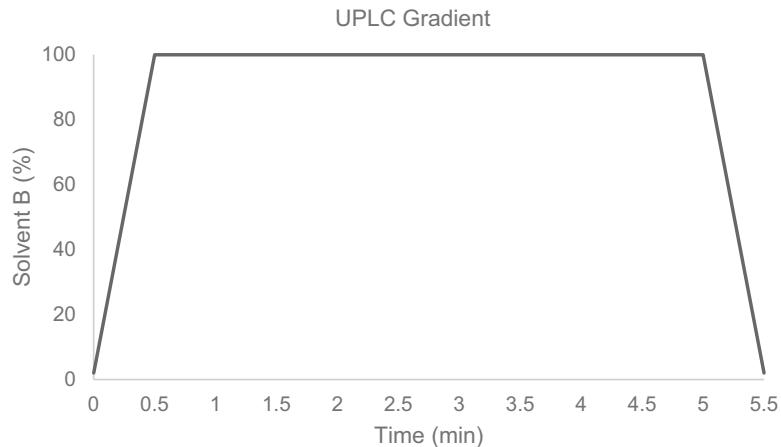
1. Precondition Oasis HLB SPE cartridges with 1 mL ethyl acetate and allow to pass through the cartridges and then add 1 mL MeOH.
2. Equilibrate the cartridges with 1 mL deionized H<sub>2</sub>O.
3. Load the prepared samples onto the cartridge and allow the solution to pass through the cartridges to the waste vessel.
4. Add 1 mL deionized H<sub>2</sub>O onto cartridges and allow to pass through to the waste vessel to remove the unwanted compounds.
5. Repeat step 4 with 1 mL aqueous 30% MeOH.
6. Replace waste vessels with 2 mL Eppendorf tubes.
7. Add 1 mL AcN followed by 500  $\mu$ L ethyl acetate to elute the analytes of interest.
8. Dry the eluate under gentle stream of nitrogen gas at room temperature in the preparation for the derivatization step.

### 3.5 Derivatization

1. Add 50  $\mu$ L 0.5 mg/mL PTAD solution to the eluate residue of plasmas, in-house quality control material and the six prepared standard solutions (*see Note 8*).
2. Mix gently using a multi-channel vortexer for 1 h at room temperature.

### 3.6 LC-MS/MS Analysis

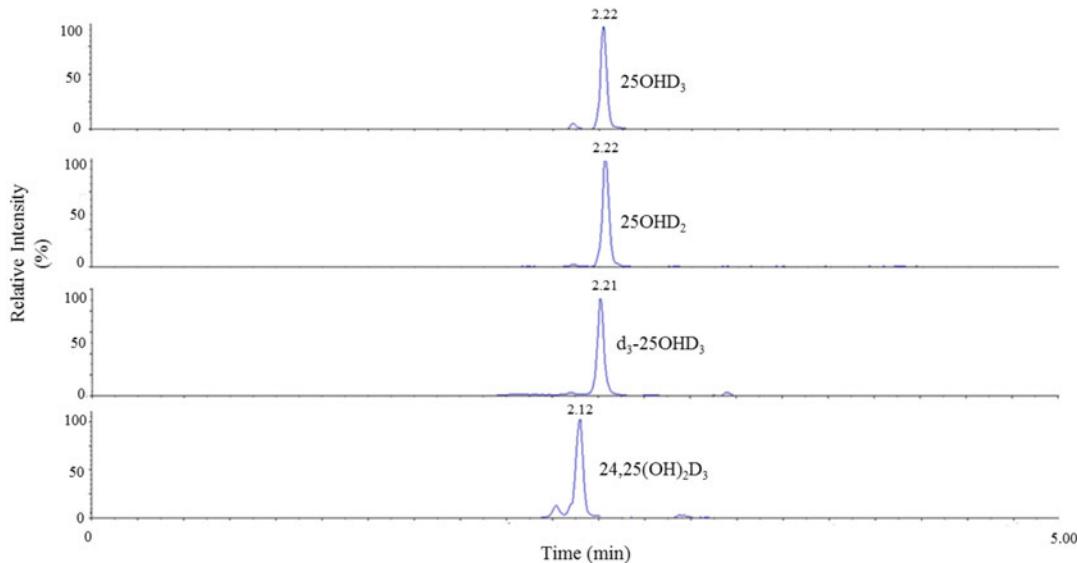
1. Transfer 50  $\mu$ L derivatized plasma and calibration standard solutions into 200  $\mu$ L inserts and load in a 2 mL amber color capped vial (*see Note 9*).
2. Load the samples on an LC plate and perform MS run (*see Note 10*).
3. LC parameters: LC gradient programme: solvent B; initial 2% for 0.5 min and increase to 100% at 0.5 min, remain at 100% for 4.5 min, and decrease to 2% for 0.5 min (Fig. 2); flow rate at 350  $\mu$ L.minute<sup>-1</sup>; run time of 5.5 min; target sample temperature 5 °C; column temperature 45 °C; injection mode:



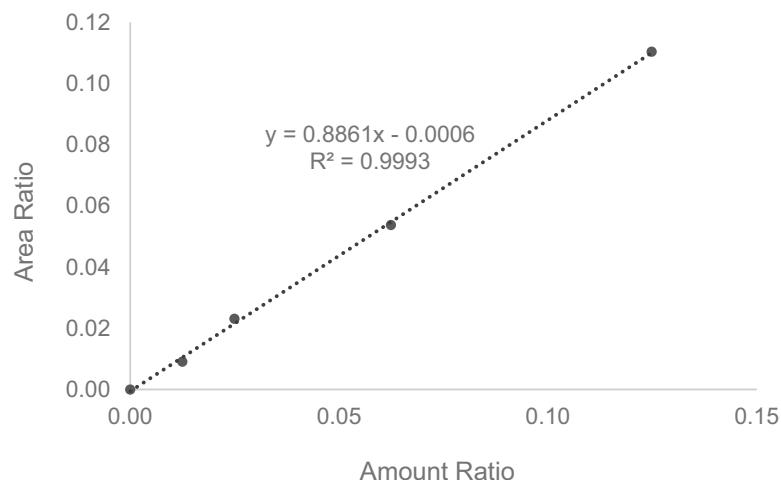
**Fig. 2** LC Gradient programme used to elute vitamin D metabolites over 5.5 min. The initial and end concentrations selected for solvent B were 2%. The Y-axis indicates the solvent B percentage and the X-axis is the elution time

partial loop; loop size: 20  $\mu$ L; injection volume: 10  $\mu$ L; equilibration time: 5 min.

4. MS operating parameters (*see Note 11*): electrospray ionization (ESI) positive polarity; ion spray voltage, 5500 V; source temperature, 550 °C; curtain gas 30 psi; ion source gas1, 70 psi and ion source gas2, 60 psi; selected mass transitions for analysis of PTAD-derivatized vitamin D metabolites were as 607 > 298 for 25OHD<sub>3</sub>, 613 > 298 for d<sub>6</sub>-25OHD<sub>3</sub>, 610 > 301 for d<sub>3</sub>-25OHD<sub>3</sub>, 623 > 298 for 24,25(OH)<sub>2</sub>D<sub>3</sub>, and 629 > 298 for d<sub>6</sub>-24,25(OH)<sub>2</sub>D<sub>3</sub> with dwell time 100 ms.
- ### 3.7 Data Analysis
1. Multiple ways exist to carry out the data analysis and the quantification method (*see Fig. 3* as an example for the chromatograms obtained from analysis of a plasma sample) (*see Note 12*).
  2. Derive the external standard curves by plotting the peak area ratio against their amount ratio (i.e., analyte: corresponding internal standard) (*see Fig. 4* as an example for the calibration curve) (*see Note 13*).
  3. Assess the precision and quality of the analysis by calculation of the coefficient variation percentage (CV %) between the results obtained from analysis of the set of standards and in-house quality control material (*see Note 10*).



**Fig. 3** LC-MS/MS chromatograms of 25OHD<sub>3</sub>, 25OHD<sub>2</sub>, d<sub>3</sub>-25OHD<sub>3</sub> (tracer), and 24,25(OH)<sub>2</sub>D<sub>3</sub> in pooled human plasma



**Fig. 4** Calibration curve for the determination of the tracer (d<sub>3</sub>-25OHD<sub>3</sub>) constructed by plotting the peak area ratio against amount ratio for d<sub>3</sub>-25OHD<sub>3</sub>:d<sub>6</sub>-25OHD<sub>3</sub>

#### 4 Notes

1. Methods for plasma sample collection were carried out as described elsewhere [3]. Details of the study design and half-life determination are reported by [3]. In summary, tracer (d<sub>3</sub>-25OHD<sub>3</sub>) was dissolved in vegetable oil and an aliquot equal to 40 nmol was given to healthy volunteers on a small

piece of bread. Fasting LiHep-blood samples of volunteers was collected and stored at  $-80^{\circ}\text{C}$  until use.

2. Use analytical HPLC grade for all reagents.
3. Prepare all aqueous solutions using deionized water to attain a sensitivity of  $18\text{ M}\Omega\text{-cm}$  at  $25^{\circ}\text{C}$ .
4. Prepare PTAD solution in AcN freshly on the day of sample LC-MS/MS analysis.
5. The concentration ranges were selected for the standard curves based on the expected concentration range in human plasma samples from our studies:  $25\text{OHD}_3$  ( $5\text{--}200\text{ nmol/L}$ ),  $d_3\text{-}25\text{OHD}_3$  ( $0.25\text{--}4\text{ nmol/L}$ ),  $24,25(\text{OH})_2\text{D}_3$  ( $1\text{--}20\text{ nmol/L}$ ).
6. The concentrations of internal standards, i.e.,  $d_6\text{-}25\text{OHD}_3$  and  $d_6\text{-}24,25(\text{OH})_2\text{D}_3$ , were calculated to be  $20\text{ nmol/L}$  in final  $50\text{ }\mu\text{L}$  PTAD solutions for both external standard solutions and plasma samples.
7. Pooled plasma, as in-house quality control material, was collected from healthy volunteers under a method development ethics approval and informed, written consent was obtained. In-house quality control sample was treated and prepared as mentioned above and analyzed with each batch of plasma samples.
8. Optimization of the PTAD solution concentration and mixing time have been reported elsewhere [16].
9.  $50\text{ }\mu\text{L}$  sample allows for two runs accounting for duplicate injections using partial loop injection mode and  $10\text{ }\mu\text{L}$  injection volume.
10. Run the set of external standards and in-house quality control material at the beginning and end of the batch to assess the quality of the run.
11. MS parameters were optimized to give the highest response for the determination of the compounds of interest.
12. The system control and peak integration were performed using Analyst version 1.6.2 (Applied Biosystems, Concord, ON, Canada).
13. Stable isotope labeled ( $d_6\text{-}25\text{OHD}_3$ ) was used as an internal standard for both the tracer ( $d_3\text{-}25\text{OHD}_3$ ) and  $25\text{OHD}_3$  and  $d_6\text{-}24,25(\text{OH})_2\text{D}_3$  for  $24,25(\text{OH})_2\text{D}_3$  quantification.

---

## Acknowledgments

We would like to thank the UK Medical Research Council for funding this study (programme number U105960371).

## References

1. Holick MF (2009) Vitamin D status: measurement, interpretation, and clinical application. *Ann Epidemiol* 19:73–78
2. van den Ouweland JM, Vogeser M, Bacher S (2013) Vitamin D and metabolites measurement by tandem mass spectrometry. *Rev Endocr Metab Disord* 14:159–184
3. Jones KS, Assar S, Harnpanich D, Bouillon R, Lambrechts D, Prentice A et al (2014) 25(OH)D<sub>2</sub> half-life is shorter than 25(OH)D<sub>3</sub> half-life and is influenced by DBP concentration and genotype. *J Clin Endocrinol Metab* 99:3373–3381
4. Aghajafari F, Field CJ, Rabi D, Kaplan BJ, Maggiore JA, O'Beirne M et al (2016) Plasma 3-Epi-25-hydroxycholecalciferol can alter the assessment of vitamin D status using the current reference ranges for pregnant women and their newborns. *J Nutr* 146:70–75
5. Prentice A, Goldberg GR, Schoenmakers I (2008) Vitamin D across the lifecycle: physiology and biomarkers. *Am J Clin Nutr* 88:500S–506S
6. Seamans KM, Cashman KD (2009) Existing and potentially novel functional markers of vitamin D status: a systematic review. *Am J Clin Nutr* 89:1997S–2008S
7. Jones KS, Schoenmakers I, Bluck LJ, Ding S, Prentice A (2012) Plasma appearance and disappearance of an oral dose of 25-hydroxyvitamin D<sub>2</sub> in healthy adults. *Br J Nutr* 107:1128–1137
8. Bluck LJ (2009) Recent progress in stable isotope methods for assessing vitamin metabolism. *Curr Opin Clin Nutr Metab Care* 12:495–500
9. Braithwaite V, Jones KS, Assar S, Schoenmakers I, Prentice A (2014) Predictors of intact and C-terminal fibroblast growth factor 23 in Gambian children. *Endocr Connect* 3:1–10
10. Jones KS, Assar S, Vanderschueren D, Bouillon R, Prentice A, Schoenmakers I (2015) Predictors of 25(OH)D half-life and plasma 25(OH)D concentration in The Gambia and the UK. *Osteoporos Int* 26:1137–1146
11. Volmer DA, Mendes LR, Stokes CS (2015) Analysis of vitamin D metabolic markers by mass spectrometry: current techniques, limitations of the “gold standard” method, and anticipated future directions. *Mass Spectrom Rev* 34:2–23
12. Qi Y, Geib T, Schorr P, Meier F, Volmer DA (2015) On the isobaric space of 25-hydroxyvitamin D in human serum: potential for interferences in liquid chromatography/tandem mass spectrometry, systematic errors and accuracy issues. *Rapid Commun Mass Spectrom* 29:1–9
13. Higashi T, Yamauchi A, Shimada K (2003) Application of 4-(4-nitrophenyl)-1,2,4-triazoline-3,5-dione to analysis of 25-hydroxyvitamin D<sub>3</sub> in human plasma by liquid chromatography/electron capture atmospheric pressure chemical ionization-mass spectrometry. *Anal Sci* 19:941–943
14. Aronov PA, Hall LM, Dettmer K, Stephensen CB, Hammock BD (2008) Metabolic profiling of major vitamin D metabolites using Diels-Alder derivatization and ultra-performance liquid chromatography-tandem mass spectrometry. *Anal Bioanal Chem* 391:1917–1930
15. Higashi T, Suzuki M, Hanai J, Inagaki S, Min JZ, Shimada K et al (2011) A specific LC/ESI-MS/MS method for determination of 25-hydroxyvitamin D<sub>3</sub> in neonatal dried blood spots containing a potential interfering metabolite, 3-epi-25-hydroxyvitamin D<sub>3</sub>. *J Sep Sci* 34:725–732
16. Ding S, Schoenmakers I, Jones K, Koulman A, Prentice A, Volmer DA (2010) Quantitative determination of vitamin D metabolites in plasma using UHPLC-MS/MS. *Anal Bioanal Chem* 398:779–789

# Chapter 23

## iTRAQ-Based Shotgun Proteomics Approach for Relative Protein Quantification

Erika Velásquez Núñez, Gilberto Barbosa Domont,  
and Fábio César Sousa Nogueira

### Abstract

Shotgun proteomics has a key role in quantitative estimation of proteins from biological systems under different conditions, which is crucial in the understanding of their functional roles. Isobaric tagging for relative and absolute quantitation (iTRAQ) mass spectrometry is based on pre-labeling of peptides with mass tags which allows the multiplex analysis of up to eight proteomes simultaneously. We describe here a detailed protocol for sample preparation and iTRAQ 4-plex labeling for relative quantification of multiple samples from human and plant tissues. We also present two strategies for peptide fractionation after the iTRAQ labeling protocol.

**Key words** iTRAQ, Animal tissues, Plant tissues, Stable isotope labeling, Quantitative proteomics, Mass spectrometry

---

### 1 Introduction

Shotgun proteomics is no longer merely descriptive, but has become quantitative [1]. The estimation of protein abundances in different biological conditions provides more information than in qualitative studies, in which only the presence or posttranslational modifications of proteins are determined. One of the turning points in quantitative proteomics was the implementation of isobaric methods for relative protein quantification. Different from other stable isotope labeling methods in which identical peptides with a mass variance from two or more samples can be distinguished in the mass spectrometry (MS) 1 spectrum, isobaric methods add isotope tags of the same mass to each peptide. In MS1 mode, each peptide labeled with different isobaric tags is undistinguishable but in MS2 mode reporter ions of distinct masses are released such that peptides from different samples can be differentiated and quantified, based on the intensity of these reporter ions [2]. Two of the most used isobaric labeling methods are isobaric

tags for relative and absolute quantitation (iTRAQ) [3] and tandem mass tags (TMT) [4, 5].

The widely employed iTRAQ approach is a reliable labeling technique employed in the analysis of up to four (iTRAQ 4-plex) [3] or eight samples (iTRAQ 8-plex) [6]. Compared to label-free mass spectrometry techniques, the multiplexing potential of isotopic labeling increases the statistical relevance and accuracy of results through the simultaneous analysis of different biological samples and through normalization to an internal standard. In the iTRAQ 4-plex approach, the reagent contains reporter (*N*-methylpiperazine), balance (carbonyl) and reactive groups (NHS ester). Each reagent has the same mass achieved by a combination of  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{18}\text{O}$  in the reporter ( $m/z$  114–117) and balance groups (28–31 Da). The labeled peptides have identical retention times in liquid chromatography and, since the tags are isobaric, the peptides appear as single peak with the same  $m/z$  in a MS1 spectrum. However, selection of the precursor ion for fragmentation produces a MS2 spectrum with reporter ion peaks at low mass region (114, 115, 116, and 117) and peptide backbone fragmentation peaks. Intensity of the reporter ion peaks directly reflects the abundance of the peptide in each sample (Fig. 1).

Here, we describe a detailed sample preparation and iTRAQ 4-plex labeling method for relative quantification of multiple samples from human and plant tissues. Additionally, we detail two strategies for peptide fractionation after iTRAQ labeling.

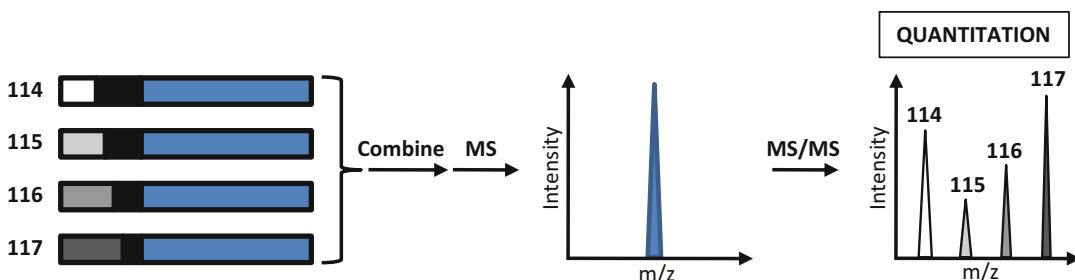
## 2 Materials (See Note 1)

### 2.1 Protein Extraction of Human Tissues

- Human tissues: RapiGest® (Waters Corporation).
- 1 M triethylammonium bicarbonate (TEAB; Sigma-Aldrich) (see Note 2).

### 2.2 Protein Extraction of Plant Tissues

- 50 mM pyridine (pH 5.0), 10 mM thiourea, 1% SDS, 5% polyvinylpolypyrrolidone (PVPP).
- 10% trichloroacetic acid (TCA) in acetone.
- 7 M urea/2 M thiourea solution containing 5% TEAB.



**Fig. 1** Principle of iTRAQ mass spectrometry technique

**2.3 Enzymatic Digestion**

1. Qubit® 2.0 fluorimetric assay kit (Invitrogen).
2. Reducing solution: 100 mM dithiothreitol (DTT) or 50 mM tris(2-carboxyethyl) phosphine (TCEP) (*see Notes 3 and 4*).
3. Alkylation solution: 400 mM iodoacetamide (*see Note 5*).
4. Sequencing grade modified trypsin (Promega).
5. Acetic acid.
6. 10% trifluoroacetic acid (TFA).

**2.4 iTRAQ Peptide Labeling**

1. 0.1% TFA.
2. 0.1% TFA/50% acetonitrile (ACN).
3. 0.1% TFA/70% ACN.
4. iTRAQ reagent 4-plex Kit (Applied Biosystems Sciex).
5. C18 macro-spin column (Harvard Apparatus) (*see Note 6*).
6. Strong cation exchange (SCX) macro-spin column (Harvard Apparatus) (*see Note 7*).
7. Buffer A: 5 mM KH<sub>2</sub>PO<sub>4</sub>/25% ACN (pH 3).
8. Buffer B: 1 M KCl stock solution.
9. Shimadzu LC-20AT high performance liquid chromatography (HPLC) instrument for hydrophilic interaction chromatography (HILIC) using a 3 μm × 5 cm × 2 mm TSKgel® amide-80 column (Sigma-Aldrich) (*see Note 8*).
10. Solvent A (HILIC-A): 90% ACN/0.1% TFA.
11. Solvent B (HILIC-B): 0.1% TFA.

**2.5 Labeled Peptide Analysis by Nano LC-MS/MS**

1. 2 cm length, 200 μm inner diameter EASY II trap-column (ThermoFisher Scientific).
2. 18 cm length, 100 μm inner diameter, 5 μm resin (ReproSil-Pur C18) pulled analytical capillary column (Dr. Maisch GmbH).
3. Phase A: 0.1% formic acid, 5% ACN.
4. Phase B 0.1% formic acid, 95% ACN.
5. LTQ Orbitrap Velos mass spectrometer (ThermoFisher Scientific).

**2.6 Data Analysis**

1. Data inspection: Xcalibur 2.1 software (ThermoFisher Scientific).
2. Database searches: Proteome Discoverer 2.1 software (ThermoFisher Scientific) with the SEQUEST algorithm.
3. Databases: Uniprot ([www.uniprot.org/](http://www.uniprot.org/)), NCBI ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)), and neXtProt ([www.nextprot.org/](http://www.nextprot.org/)).

---

### 3 Methods

#### 3.1 Protein Extraction in Human Tissues

1. Pulverize and macerate the tissues in liquid nitrogen [7].
2. Add 0.1% RapiGest in 50 mM TEAB (*see Note 9*).
3. Vortex the samples and centrifuge for 30 min at  $20,000 \times g$  at 4 °C.
4. Transfer the supernatant to another tube and take one aliquot for protein quantification.

#### 3.2 Protein Extraction in Human Tissues

1. Mix the sample at a 1:40 ratio with pyridine buffer [8].
2. Stir for 2 h at 4 °C and centrifuge at  $10,000 \times g$  for 40 min.
3. Precipitate proteins in ice-cold 10% TCA in acetone.
4. Wash the pellet with ice-cold acetone three times and dry under vacuum.
5. For protein quantification, dissolve the pellet in urea/thiourea solution.

#### 3.3 Protein Digestion with Trypsin

1. Quantitate proteins using the Qubit 2.0 fluorometric assay kit according to the manufacturer's instructions.
2. Disulfide bond reduction in proteins: incubate the samples with DTT or TCEP solution at a final concentration of 10 mM for 1 h at 30 °C.
3. Thiol group alkylation in proteins: incubate the samples with iodoacetamide solution at a final concentration of 40 mM for 30 min at room temperature in the dark.
4. Trypsin digestion: add trypsin at a trypsin:protein ratio of 1:50 and incubate for 12–18 h at 37 °C (*see Note 10*).
5. Stop the reaction by adding 10% TFA to give a final concentration of 0.1% (*see Note 11*).

#### 3.4 iTRAQ Peptide Labeling [9, 10]

1. Peptide cleaning: incubate C18 spin columns with 500 µL 100% ACN for 15 min and centrifuge at  $2000 \times g$  for 1 min.
2. Add the same amount of ACN and repeat the centrifugation step.
3. Equilibrate the columns with 150 µL 0.1% TFA and centrifuge at  $2000 \times g$ .
4. Repeat this step three times, add 75–150 µL sample, and centrifuge at  $2000 \times g$ .
5. Wash the columns using 0.1% TFA and centrifuge at  $2000 \times g$ .
6. Repeat the wash/centrifugation three times in total and elute the peptides in two successive steps into the same collection tube with 0.1% TFA/50% ACN and 0.1% TFA/70% ACN followed by centrifugation at  $2000 \times g$ .

7. Dry the peptides by vacuum centrifugation.
8. Peptide labeling: suspend peptides in 30  $\mu$ L of 20 mM TEAB (pH 8.5) (*see Note 12*) and quantify using the Qubit 2.0 fluorometric assay to normalize peptide amounts.
9. Briefly centrifuge the iTRAQ reagent solution at room temperature so that the content is collected in the bottom of the tube and add 70  $\mu$ L of ethanol to each vial.
10. Vortex the vials and centrifuge briefly as above.
11. Transfer the contents of each vial to a specific sample tube, vortex and centrifuge again (*see Note 13*).
12. Incubate the samples at room temperature for 1 h.
13. Stop the reaction by adding formic acid at a final concentration of 1%, vortex and centrifuge as above (*see Note 14*).
14. Combine the contents of all samples labeled with different iTRAQ tags into one tube, vortex and centrifuge.
15. Dry the contents in a vacuum centrifuge but stop before complete dryness is reached (*see Note 15*).

### ***3.5 iTRAQ-Labeled Peptide Fractionation***

1. Suspend the semidry pellets at approximately 1  $\mu$ g peptides/ $\mu$ L in 100  $\mu$ L of 5 mM KH<sub>2</sub>PO<sub>4</sub>/25% ACN solution and vortex.
2. Incubate the SCX spin column (Harvard Apparatus) with 500  $\mu$ L of the same solution for 15 min at room temperature.
3. Centrifuge at 2000  $\times g$  until all solution has passed through the column and repeat this step.
4. Add the sample to the spin column, centrifuge at 2000  $\times g$  and collect the column flow-through.
5. Carry out four sequential elution steps using 150  $\mu$ L of the 5 mM KH<sub>2</sub>PO<sub>4</sub>/25% ACN solution, containing first 75 mM KCl, second 150 mM KCl, third 250 mM KCl, and last 500 mM KCl, followed by centrifugation each time at 2000  $\times g$  and collecting the eluates in separate tubes.
6. Desalt the samples using the peptide cleaning step above.
7. Suspend the samples in 0.1% formic acid and quantify as above.
8. Suspend the samples at approximately 1  $\mu$ g/ $\mu$ L in 100  $\mu$ L of HILIC-A solution, vortex and centrifuge briefly to collect the sample at the bottom of the tube.
9. Load samples at a flow rate of 0.2 mL/min into the TSKgel Amide-80 column on the LC-20AT HPLC system.
10. Fractionate the peptides by applying 100% HILIC-A for 10 min, 12% HILIC-B for 2 min, 20% HILIC-B for 30 min, 30% HILIC-B for 30 min, 100% HILIC-B for 5 min, and return to 100% HILIC-A for 5 min.

### 3.6 Labeled Peptide Analysis by nanoLC-MS/MS

1. Load 1 µg labeled peptides onto the trap and capillary columns on the nano LC system coupled online to the LTQ Orbitrap Velos mass spectrometer.
2. For peptide elution, apply a gradient from 100 % phase A to 35 % phase B over 120 min at a flow rate of 200 nL/min.
3. After each run, wash the column with 90 % phase B and re-equilibrate with phase A.
4. Acquire spectra in positive mode applying a data-dependent automatic survey MS scan and MS/MS.
5. Set the resolution of the Orbitrap mass analyzer at 60,000 at m/z 400, automatic gain control target at  $1 \times 10^6$ , and maximum ion injection at 500 milliseconds (see Note 16).
6. Acquire MS/MS spectra with a resolution of 7500 at 400 m/z, a signal threshold of 30,000, normalized collision energy of 40, and dynamic exclusion enabled for 30 s with a repeat count of 1.
7. Place an Eppendorf tube covered with 5 % ammonia water solution under the nano ESI needle (see Note 17).

### 3.7 Data Analysis

1. Inspect the raw data using the Xcalibur software.
2. Perform database searches against target and decoy (reverse) databases from Uniprot, NCBI, and neXProt using the following search parameters: MS accuracy = 10 ppm; MS/MS accuracy = 0.1 Da; trypsin digestion with two missed cleavages allowed; fixed carbamidomethyl modification of cysteine; and variable modification of oxidized methionine.
3. For identification of iTRAQ labeled peptides, also include the iTRAQ 4-plex monoisotopic mass = 144.102 and variable modification for N-terminus, lysine and tyrosine.
4. Accept false discovery rates less than 1 % and peptide rank = 1.

---

## 4 Notes

1. Reagents should be of analytical grade, solvents HPLC or LC-MS grade and solutions should be prepared with ultrapure water (18 MΩ-cm at 25 °C). LC-MS solutions require LC-MS grade water.
2. It is advisable to use protease inhibitors in this step to prevent degradation of proteins caused by proteases in the sample. In addition, for phosphoproteomics, it is necessary to use phosphatase inhibitors to prevent dephosphorylation during preparation and handling of sample.
3. It is recommended to use a fresh DTT stock solution.

4. TCEP has the advantage of being a more powerful reducing agent than DTT, providing an irreversible reaction. In addition, it is more hydrophilic, active in alkaline and acidic conditions, and more resistant to air oxidation. Also, it does not reduce metals and is significantly more stable than DTT in the absence of a metal chelator.
5. It is necessarily to prepare the iodoacetamide solution immediately before use and keep it protected from light because it is unstable and light-sensitive. Also, the alkylation step must be performed in darkness.
6. The C18 Macro-Spin Column has a binding capacity of 30–300 µg of sample, accepting a sample volume of 70–150 µL. Review the manufacturer's specifications before use.
7. The SCX Macro-Spin Column has an ion capacity of 0.18–0.25 mmol (Cl<sup>-</sup>)/ml and a binding capacity of 30–300 µg of protein sample, accepting a sample volume of 70–150 µL. Review the manufacturer's specifications before use.
8. HILIC is recommended to remove excess iTRAQ reagent from iTRAQ-labeled peptides because the solutions involved in this procedure are compatible with mass spectrometric analysis. This eliminates an additional step of sample cleaning [11, 12].
9. We recommend addition of 200 µL of 0.1% RapiGest for 100 mg of tissue.
10. Samples in 8 M urea/ 2 M thiourea must be diluted to final concentrations lower than 1 M urea using 100 mM TEAB and avoid heating. Check the pH to ensure that it is close to 8.
11. For RapiGest samples, acidify the samples with TFA to a final concentration of 1% to stop the reaction and incubate for 40 min at room temperature. Centrifuge for 30 min at 20,000 ×*g* to remove insoluble material.
12. Before peptide labeling, ensure that the pH is close to 8.5.
13. It is advisable to use commercial peptides (e.g., Glu-1-fibrinopeptide B) at a known concentration at the time of labeling to serve as internal controls and facilitate data normalization.
14. It is advisable to analyze a peptide sample aliquot by MS before making the final mix of all iTRAQ labels to confirm the presence of labeled peptides with the appropriate m/z peaks for each reporter ion. If the labeling process was not successful, repeat the labeling procedure.
15. The peptide pellets are easier to resuspend if not completely dry.
16. Choose a method consisting of fragmentation of the ten most intense ions by high-energy collision dissociation.
17. The presence of 5% ammonia during analysis avoids the supercharge effect of the iTRAQ 4-plex tag [13].

## Acknowledgments

This work was funded by the CNPq (Grant number: 477325/2013-0) and FAPERJ (Grant number: E-26/202.801/2015).

## References

- Nogueira FC, Domont GB (2014) Survey of shotgun proteomics. *Methods Mol Biol* 1156:3–23
- Rauniyar N, Yates JR 3rd (2014) Isobaric labeling-based relative quantification in shotgun proteomics. *J Proteome Res* 13:5293–5309
- Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S et al (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3:1154–1169
- Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G et al (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 75:1895–1904
- Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF et al (2008) Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal Chem* 80:2921–2931
- Choe L, D'Ascenzo M, Relkin NR, Pappin D, Ross P, Williamson B et al (2007) 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics* 7:3651–3660
- Aquino PF, Lima DB, de Saldanha da Gama FJ, Melani RD, Nogueira FC, Chalub SR et al (2014) Exploring the proteomic landscape of a gastric cancer biopsy with the shotgun imaging analyzer. *J Proteome Res* 13:314–320
- Vasconcelos ÉAR, Nogueira FCS, Abreu EFM, Gonçalves EF, Souza PAS, Campos FAP (2005) Protein extraction from cowpea tissues for 2-D gel electrophoresis and MS analysis. *Chromatographia* 62:447–450
- Nogueira FC, Palmisano G, Schwämmle V, Campos FA, Larsen MR, Domont GB et al (2012) Performance of isobaric and isotopic labeling in quantitative plant proteomics. *J Proteome Res* 11:3046–3052
- Nogueira FC, Palmisano G, Schwämmle V, Soares EL, Soares AA, Roepstorff P et al (2013) Isotope labeling-based quantitative proteomics of developing seeds of castor oil seed (*Ricinus communis* L.). *J Proteome Res* 12:5012–5024
- Palmisano G, Lendal SE, Engholm-Keller K, Leth-Larsen R, Parker BL, Larsen MR (2010) Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat Protoc* 5:1974–1982
- Melo-Braga MN, Verano-Braga T, León IR, Antonacci D, Nogueira FC, Thelen JJ et al (2012) Modulation of protein phosphorylation, N-glycosylation and Lys-acetylation in grape (*Vitis vinifera*) mesocarp and exocarp owing to *Lobesia botrana* infection. *Mol Cell Proteomics* 11:945–956
- Thingholm TE, Palmisano G, Kjeldsen F, Larsen MR (2010) Undesirable charge-enhancement of isobaric tagged phosphopeptides leads to reduced identification efficiency. *J Proteome Res* 9:4045–4052

# Chapter 24

## **<sup>1</sup>H NMR Metabolomic Profiling of Human and Animal Blood Serum Samples**

**João G.M. Pontes, Antonio J.M. Brasil, Guilherme C.F. Cruz,  
Rafael N. de Souza, and Ljubica Tasic**

### **Abstract**

Nuclear magnetic resonance (NMR) spectroscopy techniques allow the acquisition of a large amount of data and when combined with multivariate statistical analysis, it is possible to process and interpret the obtained NMR data in accordance with the biological problem being investigated. In this chapter, the search for biologically relevant biomarkers is addressed using NMR spectroscopy-based metabolomics, due to their clinical relevance for either diagnosis or monitoring of diseases and disorders.

**Key words** Nuclear Magnetic Resonance, Metabolomics, Serum, Chemometrics, Biomarkers

---

### **1 Introduction**

The search for biomarkers that characterize metabolic conditions, disorders or diseases, has grown in the last years. Biotechnology tools such as genomics, proteomics, and metabolomics are the most used techniques to find biomarkers [1–3] and are complementary to each other. Metabolomics is a scientific approach that allows qualitative and quantitative identification of metabolites. Metabolites are products of cellular metabolic pathways. The first researchers to use the term “metabolome” to designate the set of all low-molecular-mass compounds synthesized by an organism were Oliver et al. in 1998 [4]. However, it was only in 2002 that Fiehn suggested the classification of analytical approaches and, thus, introduced the word “metabolomics” [4–6]. Metabolomics brought to science a new perspective for diseases diagnosis, since biofluids can be collected noninvasively and in small amounts, as is the case for urine and blood sampling. These biofluids are useful because they reflect the biochemical status of an organism and the potential homeostasis changes [7].

With non-destructive analysis, easier sample preparation, and selective experiments for biological samples, NMR has some advantages in analysis of biofluids compared to other techniques such as those based on mass spectrometry (MS). Although MS is more sensitive and requires only small sample amounts, the analysis is destructive [8]. NMR data interpretation sometimes can be simpler when compared to MS. However, a two-dimensional (2D) NMR analyses are needed for complex samples because they give more details about sample composition. Besides, there are more data indexed in databases, which provide more accurate and reliable analysis. However, 2D spectra acquisition requires a longer procedure, when compared to one-dimensional (1D) methods [8].

The NMR applications in metabolomics are based on the identification of small and macromolecules that caused group separation, which are present in one and absent in other group, or have different concentrations in the investigated groups. The chemical shift, coupling constants (splitting patterns) and different peak intensities in  $^1\text{H}$  NMR spectra reveal details about qualitative and quantitative relationships between intramolecular and intermolecular resonances. This information is acquired due to nuclei magnetic spin properties. When nuclei are submitted to a strong magnetic field their spin energy levels changes and this can be observed through radiofrequency waves ( $\lambda$  between  $10^2$  to  $10^4$  m). Pulsed NMR generates a Free Induction Decay (FID) as an observable signal of nonequilibrium nuclear spin magnetization. However, the signal output is in the time domain, which is not explainable. Therefore, the transformation to frequency domain through Fourier Transformer (FT) is needed [9, 10].

Spectra obtained by NMR offer the raw data to be processed in the next step in the search for biomarkers, known as chemometrics. This is a crucial method for obtaining useful information from the raw NMR data. Usually, statistical procedures include multivariate methods such as principal component analysis (PCA) [11] and partial least squares (PLS) analysis [12]. Even so, biofluid  $^1\text{H}$  NMR spectra may show variability, probably caused by differences in pH, metal-ion concentrations, and chemical exchange phenomena [13]. This requires baseline correction, alignment, binning, normalisation and scaling of the NMR data before applying chemometrics.

---

## 2 Materials

### 2.1 Collecting Blood Serum

1. Serum samples (see Notes 1 and 2).
2. Serum vacutainer tubes.
3. Sodium azide ( $\text{NaN}_3$ ).
4. Polypropylene tubes.
5. Biofreezer at  $-80^\circ\text{C}$  (see Notes 3 and 4).

## 2.2 Samples Preparation

1. Deuterium oxide ( $D_2O$ , 99.9 %).
2. NMR tubes (18 cm  $\times$  5 mm).
3. Phosphate buffered saline (PBS) (*see Note 5*)

## 2.3 One- and Two-Dimensional NMR Spectra Acquisition

1. Bruker® Avance III 600 MHz spectrometer (*see Note 6*).
2. Triple resonance Broadband Inverse (TBI) probe (*see Note 7*).
3. FID data interpreting software Bruker® TopSpin version 3.1.

## 2.4 Chemometric Analysis

1. Spreadsheet software, such as Microsoft® Office Excel and Origin®.
2. Chemometrics analysis software, such as Pirouette®, MATLAB®.

## 3 Methods

### 3.1 Sera Sample Collection

1. Collect the blood sample using a peripheral vein access in the morning (7–10 AM) after fast of 12 h (*see Note 8*).
2. Immediately put the sample recipient on ice and leave 1 h for coagulation.
3. Centrifuge, collect the supernatant fraction and add 0.05 mM sodium azide to avoid bacterial contamination [14].
4. Store the vacutainer at –80 °C (*see Note 9*).

### 3.2 $^1H$ NMR Spectra Acquisition and Processing

1. To a 250  $\mu$ L of animal/human serum sample add 250  $\mu$ L of deuterium oxide or PBS with 10 % of deuterium oxide.
2. Transfer the solution into an NMR tube.
3. For analysis of non-lyophilized blood serum samples, the WATERGATE pulse sequence should be used [15], which is a specific and appropriate pulsed gradient field technique for suppression of the water signal (*see Notes 10 and 11*).
4. Acquisition of  $^1H$  NMR spectra is performed in triplicate for monitoring reproducibility.
5. For  $^1H$  NMR acquisition use following parameters shown in Table 1 (*see Note 12*).
6. Adjust the baseline of all  $^1H$  NMR spectra acquired (*see Note 13*).
7. Reference the chemical shift using the lactate doublet (*see Note 14*).
8. For HSQC acquisition, use the parameters shown in Table 2.

### 3.3 Chemometrics Data Processing

1. After acquiring and preprocessing  $^1H$  NMR spectra, these are exported as American Standard Code for Information Interchange (ASCII) files, so intensity and chemical shift values are listed and the data organized using a spreadsheet software (*see Note 15*).

**Table 1**  
**Steps to 1D spectra acquisition**

| A  | Solvent ( $D_2O$ ).   |
|----|---|
| B. | Lock (commonly one can use automatic correcting)  |
| C  | Probe match/tune (manual or automatic correcting)   |
| D  | Shimming (automatic/manual command). Sometimes it is necessary to activate more than one time and adjust according to reference signal  |
| E  | Acquisition pars. This can be a pulse sequence that in this case was either WATERGATE or cpmg1d in $T_2$ analysis. The number of scans can change according to spectra precision. Generally, 2–4 scans are used to check the sample and 128 scans are used for data acquisition |
| F  | Prosol pars. Use the same number of scans utilized in acquisition pars  |
| G  | Receiver gain (automatic command)   |
| H  | Start acquisition   |

**Table 2**  
**Acquisition parameters 2D NMR–HSQC spectra acquisition**

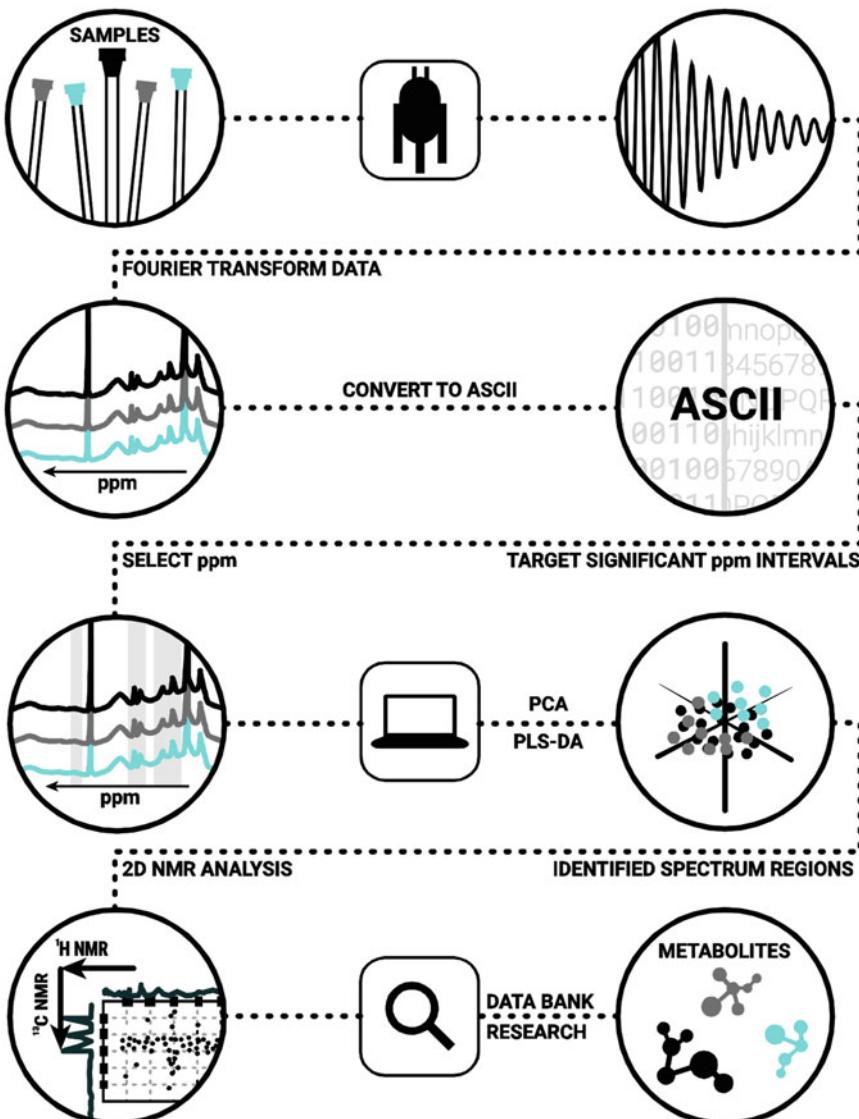
| Pulse sequence        | hsqcedetedgpson.3 |
|-----------------------|-------------------|
| Number of scans       | 64                |
| Number of dummy scans | 16                |
| Matching/tuning       | Manual            |

2. The chemometric analysis in data preprocessing is carried out according to the method of metabolomics analysis that is employed. For example: target analysis, metabolic profiling or metabolic fingerprinting.
3. Choose autoscale as preprocessing for Principal Component Analysis to data analysis at selected spectral range.
4. Exclude outliers and repeat the analysis.

### 3.4 Biomarkers Identification

1. The choice of a spectral region for chemometrics analysis in the search for biomarkers should be made in accordance with the biochemical and spectroscopical knowledge necessary to interpret the biological problem at question.
2. After the chemometrics analysis run, it is necessary to identify “*outlying*” samples, i.e., data that do not fit in the model. It is important to remove the outliers if any [16].
3. For biomarkers searching and comparison use any of the following databases: Human Metabolome Database (HMDB), Biological Magnetic Resonance Bank (BMRB) and Madison-Qingdao Metabolomics Consortium Database.

4. Finally, once the spectral range is determined, the peaks are assigned and the contour maps correlations used to compare experimental results with available data banks (Fig. 1) (see Note 16).



**Fig. 1** Summary of key steps in the metabolomics analysis. *Step 1* — Place sample into an NMR tube, insert this into the NMR instrument and record the  $^1\text{H}$  NMR spectra according to the instruction manual. *Step 2* — After Fourier Transformation of the data, each NMR spectrum must be referenced and exported as an ASCII file. Target the parts of the spectra that include differences between the groups of samples. *Step 3* — Run PCA and PLS-DA (or other chemometrics tool according to the model) over selected parts of the spectra and analyze groupings. Identify the NMR data responsible for the differences between the groups of samples. *Step 4* — If there are no rough deviations between the model and reality, assign and interpret the NMR data, using principally 2D NMR and search the appropriate NMR databases for identification of the metabolites

---

## 4 Notes

1. While handling animal and human biofluids it is necessary wear a lab coat, eyewear, and gloves for self-protection during all processes, up to the point of sample storage after NMR analysis.
2. It is necessary to obtain authorization of local ethics committee and a document with experiment details, such as start and final dates and procedures to be included in the research.
3. Sample freezing and thawing procedures should be minimized as much as possible as to avoid subtle changes in the spectroscopic profiles of certain metabolites that may occur and lead to a systematic error among the samples [17].
4. It is important to mention that each metabolite present in blood serum sample has its own degradation time according to its chemical nature and the temperature that it was stored. For instance, l-proline immediately undergoes chemical changes in serum samples when stored at  $-20\text{ }^{\circ}\text{C}$ , while other metabolites, such as, uric acid and cholesterol are stable at this temperature for a period of 3 months [18].
5. In analysis of tissues and some biofluids like urine, phosphate buffered saline (PBS) is commonly used as to provide pH stabilization and to simulate native conditions, thus keeping cells/fluids as close as possible to the physiological conditions [19, 20].
6. The magnetic field frequency of the NMR spectrometer is essential to a gain better sensitivity, so high-field NMR (greater than 500 MHz) provides a better resolution and high-quality information.
7. In  $^1\text{H}$  NMR spectra analysis for metabolomics, it is preferable to use the TBI probe in order to have higher sensitivity in analysis because the inner coil of this probe is designed for optimized  $^1\text{H}$  nucleus analysis.
8. It is important to collect the blood during the morning and after 12 h of fasting to ensure that other metabolites or xenobiotics present in the patient organism resulting from daily activities (to lunch, to take medications) do not interfere or mask the biological process in question.
9. Once fasting blood is collected, serum is obtained and the sample prepared, it is necessary to store it at an ultra-low temperature ( $-70$  or  $-80\text{ }^{\circ}\text{C}$ ). Do not have abrupt changes of conditions and do not unfreeze them more than two times before submitted to the NMR analysis [21, 22].
10. In the WATERGATE method, a frequency-selective RF pulse is applied, which gives a broad excitation in the small selected region that corresponds to an intensity reduction of

about 20 % plus some signals nearer to the solvent region are lost. In practice, these peaks correspond to anomeric carbons of carbohydrates, which is a method limitation [23].

11. Ideal water suppression takes into account the following criteria: high effectiveness (suppression of water signals in the range of  $10^5$  to  $10^9$  times more intense than metabolites peaks); high selectivity; and high efficiency to suppress peaks in a short time [24].
12. NMR serum sample analysis can sometimes be difficult to analyze due to high concentrations of proteins and lipids that can obscure and overlap the low-molecular-mass signals. To solve this problem, different pulse sequences have been used that suppress resonances in a spin-echo mode through phase modulation and relaxation effects [25]. A commonly used pulse sequence is Carr–Purcell–Meiboom–Gill (CPMG) [26] which suppresses signals of large molecules (peptides and lipids) improving spectral resolution and consequently improving multiplicity identification and peak assignment.
13. In metabolomics, it is not appropriate to integrate the signals of  $^1\text{H}$  NMR spectra, because there are small distortions in the baselines that do not allow reliable data quantification [27].
14. Besides lactate ( $\delta$  1.33, 3H, doublet,  $^3J=7$  Hz), some reference compounds can be used in aqueous medium, such as 3-(trimethylsilyl)propionic acid-*2,2,3,3-d*4 (TMSP-*d*4) referenced at  $\delta$  0.00 (9H, singlet) and 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt (DSS) (main signal at  $\delta$  0.00, singlet). In organic solvents, tetramethylsilane (TMS) is widely used [19, 28].
15.  $T_2$ -edited  $^1\text{H}$  NMR spectra should not be used as data for chemometrics due to fact that differences in the relaxation times of metabolites might invalidate their quantification. In  $T_2$ -edited  $^1\text{H}$  NMR spectra the intensities of peaks one variable of the PCA are altered, but chemical shifts and/or coupling constant values are not.
16. Sometimes it is necessary to employ Dynamic Nuclear Magnetic Resonance Spectroscopy for the search of complementary data that give information about fluxomics, studies about protein–ligand interactions, pharmacokinetics, thereby contributing to a greater understanding of biochemical mechanisms and processes involved in the biological system under study.

---

## Acknowledgments

We thank the *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP, São Paulo, Brazil) and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq, Brasília, Brazil) for financial supports (FAPESP Process No 2014/18938-8 and CNPq Process Number: 454234/2014-7) and fellowships.

## References

- Horgan RP, Kenny LC (2011) ‘Omic’ technologies: genomics, transcriptomics, proteomics and metabolomics. *Obstet Gynecol* 13:189–195
- Bonne NJ, Wong DTW (2012) Salivary biomarker development using genomic, proteomic and metabolomic approaches. *Genome Med* 4:82. doi:[10.1166/gm383](https://doi.org/10.1166/gm383)
- Kovac JR, Pastuszak AW, Lamb DJ (2013) The use of genomics, proteomics, and metabolomics in identifying biomarkers of male infertility. *Fertil Steril* 99:998–1007
- Oliver SG, Winson MK, Kell DB, Baganz F (1998) Systematic functional analysis of the yeast genome. *Trends Biotechnol* 16:373–378
- Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol* 48:155–171
- Villas-Bôas SG, Rasmussen S, Lane GA (2005) Metabolomics or metabolite profiles? *Trends Biotechnol* 23:385–386
- Pinto J, Domingues MRM, Galhano E, Pita C, Almeida MC, Carreira IM et al (2014) Human plasma stability during handling and storage: impact on NMR metabolomics. *Analyst* 139:1168–1177
- Emwas AHM, Salek RM, Griffin JL, Merzaban J (2013) NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. *Metabolomics* 9:1048–1072
- Bharti SK, Roy R (2012) Quantitative <sup>1</sup>H NMR spectroscopy. *TrAC Trend Anal Chem* 35:5–26
- Fourier JBJ (1822) *Théorie Analytique de la Chaleur*. Paris: Chez Firmin Didot, père et fils, p. 525
- Wold S (1987) Principal component analysis. *Chemom Intell Lab* 2:37–52
- Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab* 58:109–130
- Trygg J, Holmes E, Lundstedt T (2007) Chemometrics in metabolomics. *J Proteome Res* 6:469–479
- Verwaest KA, Vu TN, Laukens KA, Clemens LE, Nguyen HP, Gasse BV et al (2011) <sup>1</sup>H NMR based metabolomics of CSF and blood serum: a metabolic profile for a transgenic rat model of Huntington disease. *Biochim Biophys Acta* 1812:1371–1379
- Piotti M, Saudek V, Sklenár V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2:661–665
- Debruyne M, Engelen S, Hubert M, Rousseeuw PJ (2006) Robustness and outlier detection in chemometrics. *Crit Rev Anal Chem* 36:221–242
- Teahan O, Gamble S, Holmes E, Waxman J, Nicholson JK, Bevan C et al (2006) Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Anal Chem* 78:4307–4318
- Yin P, Lehmann R, Xu G (2015) Effects of pre-analytical processes on blood samples used in metabolomics studies. *Anal Bioanal Chem* 407:4879–4892
- Beckonert O, Keun HC, Ebbels TMD, Bundy J, Holmes E, Lindon JC et al (2007) Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc* 2:2692–2703
- Parab GS, Rao R, Lakshminarayanan S, Bing YV, Mochhalal SM, Swarup S (2009) Data-driven optimization of metabolomics methods using rat liver samples. *Anal Chem* 81:1315–1323
- Zheng P, Gao HC, Li Q, Shao WH, Zhang ML, Cheng K et al (2012) Plasma metabolomics as a novel diagnostic approach for major depressive disorder. *J Proteome Res* 11:1741–1748
- Rankin NJ, Preiss D, Welsh P, Burgess KEV, Nelson SM, Lawlor DA et al (2014) The emergence of proton nuclear magnetic resonance metabolomics in the cardiovascular arena as viewed from a clinical perspective. *Atherosclerosis* 237:287–300
- Adams RW, Holroyd CM, Aguilar JA, Nilsson M, Morris GA (2013) “Perfecting” WATERGATE: clean proton NMR spectra from aqueous solution. *Chem Commun* 49:358–360
- Tang H, Wang Y (2008) High-resolution NMR spectroscopy in human metabolism and metabolomics. In: Webb GA (ed) *Modern magnetic resonance – Part III: Applications in materials science and food science*. Springer, New York, pp 1623–1630
- Wevers RA, Engelke U, Heerschap A (1994) High-resolution <sup>1</sup>H NMR spectroscopy of blood plasma for metabolic studies. *Clin Chem* 40:1245–1250
- Meiboom S, Gill D (1958) Modified spin-echo method for measuring nuclear relaxation times. *Rev Sci Instrum* 29:688–691
- Smolińska A (2012) *Chemometrics and NMR spectroscopy for metabolomics analysis of neurological disorders*. Radboud University, Nijmegen. ISBN 978-94-6191-358-6
- Gebregiworgis T, Powers R (2012) Application of NMR metabolomics to search for human disease biomarkers. *Com Chem High T Scr* 15:595–610

# Chapter 25

## Lab-on-a-Chip Multiplex Assays

Harald Peter, Julia Wienke, and Frank F. Bier

### Abstract

Lab-on-a-chip multiplex assays allow a rapid identification of multiple parameters in an automated manner. Here we describe a lab-based preparation followed by a rapid and fully automated DNA microarray hybridization and readout in less than 10 min using the Fraunhofer in vitro diagnostics (ivD) platform to enable rapid identification of bacterial species and detection of antibiotic resistance. The use of DNA microarrays allows a fast adaptation of new biomarkers enabling the identification of different genes as well as single-nucleotide-polymorphisms (SNPs) within these genes. In this protocol we describe a DNA microarray developed for identification of *Staphylococcus aureus* and the *mecA* resistance gene.

**Key words** Bacterial infection, Early diagnosis, Point-of-care, Lab-on-a-chip, Antimicrobial drug resistance, Antibiotic resistance detection, MRSA detection, Genotyping, SNP detection, Automated DNA microarray hybridization

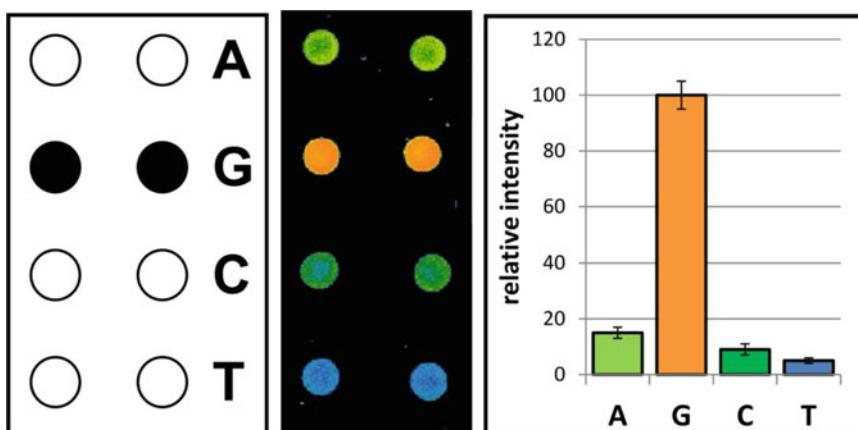
---

### 1 Introduction

There is an urgent need for rapid diagnostics in the field of infectious diseases where fast decisions need to be made for an effective therapy. Traditional microbiological methods take around 48 h for species identification and antibiogram results. The use of molecular tests built on multi-parameter lab-on-a-chip platforms are a good option to speed up the time taken between sampling and results. Here we describe a protocol which makes use of the Fraunhofer integrated lab-on-a-chip in vitro diagnostic (ivD) platform. The platform can be used to run multiplex immunoassays and serological assays to identify antibodies directly from blood samples [1]. In this protocol we demonstrate another type of application, consisting of a fully automated DNA microarray assay and readout, together with a lab-based preparation method. This protocol allows rapid bacterial species identification and the genotyping of relevant antibiotic resistance genes. The advantage of microarrays is the possibility to analyze a large amount of parameters at one time, making these assays particularly useful in the field of antibiotic resistance determination [2–7].

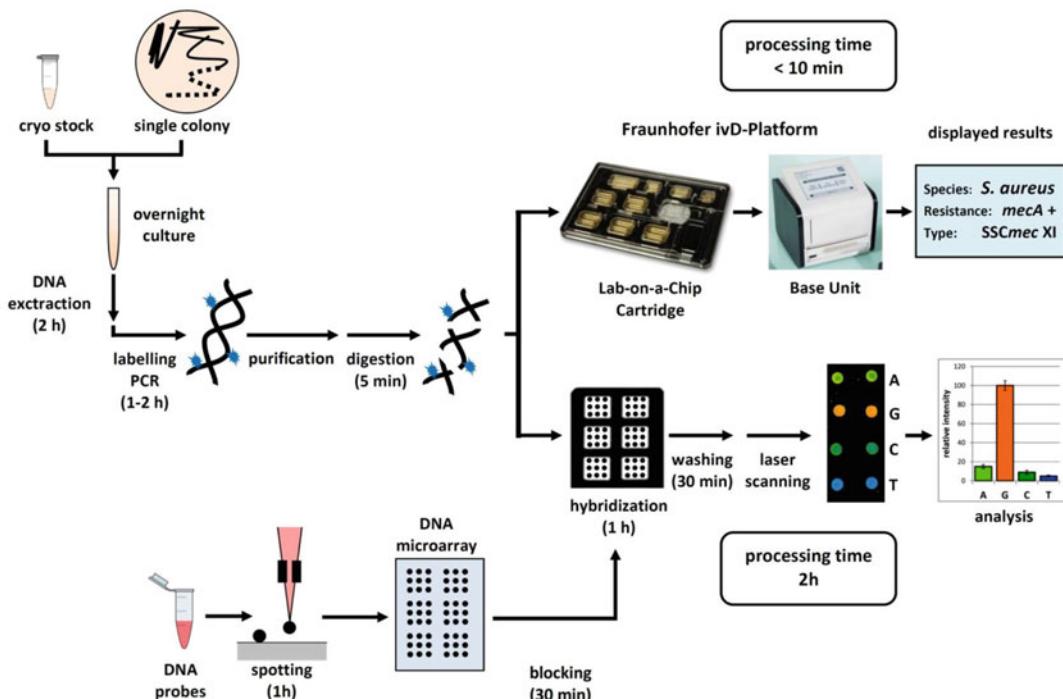
Nonetheless, there are still some disadvantages regarding current DNA microarray protocols that hinder the broad use of this approach in routine analysis. These include long hybridization times and complex laboratory based procedures, as well as the need for skilled operational personnel. The use of lab-on-a-chip based systems may help to overcome these limitations since they offer a rapid and automated solution, which integrates many of the lab-based procedures. The combination of microarrays with microfluidics constitutes an elegant solution to automate and speed up microarray hybridization. The ivD platform consists of a microfluidic cartridge and a base unit. The credit card sized cartridge contains all relevant elements necessary for DNA microarray hybridization. These are mainly reservoirs for all of the reagents, integrated pumping systems, the specific microarray, integrated temperature control for hybridization and an optical transducer to allow integrated sensing [1, 8, 9]. The base unit contains all necessary electronics to control the cartridge, an optical readout system to analyze the microarray directly after hybridization and a touch screen to control the assay and monitor the results. With this setup, hybridization times of less than 5 min can be achieved, yielding equivalent results as found by lab-based microarray hybridizations of 1 h. Therefore, the total time to generation of results is less than 10 min for hybridization, washing and readout.

Here, we describe all necessary steps from sample preparation to microarray analysis for a lab-based and lab-on-a-chip-based procedure. As an example, we use a species identification microarray for the detection of methicillin-resistant *Staphylococcus aureus* (MRSA). This microarray applies the principle of allele-specific hybridization and can therefore be used for identification of single-nucleotide polymorphisms (SNPs). It consists of four oligonucleotide probesets which differ only in one central nucleotide position, based on principles published previously [3, 4, 10]. Figure 1 shows

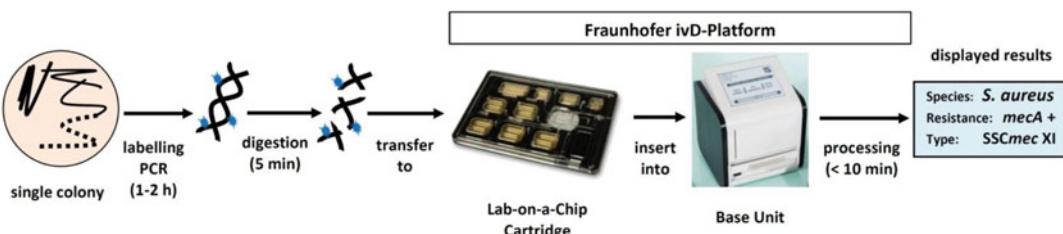


**Fig. 1** SNP detection principle used on DNA microarray. (Left) Array layout of a probe set in duplicate for SNP detection with the expected perfect match signal marked in black. (Middle) Fluorescent signals on the microarray (false color image). (Right) Quantified fluorescent signal intensities of one probe set, normalized to the perfect match signal

a SNP detection probe set as an example. The assay can be performed using conventional hybridization or automated using the Fraunhofer lab-on-a-chip system. The use of the latter will shorten the hybridization, washing and readout times from approximately 2 h to less than 10 min. Both procedures are listed in the protocol as shown in Fig. 2. Alternatively to this complete protocol, an even faster assay time can be achieved by carrying out the labeling PCR amplification directly from a colony, followed by DNA fragmentation and automated hybridization, washing, and readout using the Fraunhofer ivD platform (Fig. 3). Species identification and antibiotic resistance information can be obtained with this (fast) protocol directly from a colony with a PCR time of 60 min and an ivD



**Fig. 2** Schematic of the lab-based microarray protocol. For hybridization, washing, readout and data analysis, one can either use a manual variant (2 h processing time) or a lab-on-a-chip system, like the automated Fraunhofer ivD platform, allowing a processing time of less than 10 min



**Fig. 3** Schematic of the rapid, semi-automated ivD platform protocol, allowing a microarray analysis directly from a colony within 75 min (60 min PCR, 5 min fragmentation, and 10 min ivD platform)

step of 10 min (*see Note 1*). Our practical experience has shown that the results of the rapid protocol are comparable to those produced by the full protocol.

## 2 Materials

### 2.1 DNA Extraction

1. Müller-Hinton agar with sheep blood (Thermo Fisher Scientific Oxoid; Braunschweig, Germany).
2. 30 g/L CASO-Bouillon medium (ROTH; Karlsruhe, Germany).
3. Ethanol at -20 °C.
4. 2 mg/mL lysostaphin (Sigma-Aldrich; Taufkirchen, Germany).
5. 10 mg/mL ribonuclease A (Thermo Fisher Scientific; Darmstadt, Germany).
6. Enzymatic lysis buffer: 20 mM Tris base (pH 8), 2 mM Na-EDTA, 1.2 % Triton X-100.
7. Commercially available kit: DNeasy Blood & Tissue Kit, including buffers AW1 and AW2 (Qiagen; Hilden, Germany).

### 2.2 PCR Amplification and Labeling

1. Standard PCR reaction mixture for amplification of 23S rDNA, 16S rDNA, and *mecA* gene: 10 ng chromosomal DNA, 0.8× Taq buffer, 0.6 mM dNTPs (Qiagen), 0.6 mM 23S rDNA and 0.4 mM 16S rDNA/*mecA* oligonucleotide primers (Metabion; Martinsried, Germany), 3.75 U HotStar Taq polymerase (Qiagen), and 2 mM MgCl<sub>2</sub> in a final volume of 12.5 μL.
2. PCR labeling dNTP mixture for amplification of 23S rDNA, 16S rDNA, and *mecA*: 10 ng chromosomal DNA, 0.8× Taq buffer, dATP, dGTP, dCTP, and 0.36 μM dTTP/0.24 μM DY-647P1-aadUTP (Dyomics; Jena, Germany), 0.6 μM 23S rDNA, and 0.4 μM 16S rDNA/*mecA* oligonucleotide primers, 3.75 U HotStar Taq polymerase, and 2 mM MgCl<sub>2</sub> in a final volume of 12.5 μL.
3. Thermal cycler PTC200 with alpha block (MJ Research; St. Bruno, Canada) (*see Note 2*).

### 2.3 Agarose Gel Electrophoresis

1. 1 % agarose gel in 100 mM Tris-HCl (pH 8.3), 50 mM acetic acid, 1 mM EDTA (TAE) buffer containing peqGreen dye (peqLab; Erlangen, Germany).
2. GeneRuler 1 kb Ladder and 6× DNA Gel Loading Dye (Thermo Fisher Scientific).

### 2.4 PCR Purification

1. QIAquick PCR Purification Kit (Qiagen).
2. NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies; Wilmington, DE, USA) (*see Note 3*).

## **2.5 DNA Fragmentation**

1. 10× DNase buffer (Promega; Mannheim, Germany).
2. DNase: 1 U/ $\mu$ L (Promega).
3. DNase stop buffer: 20 mM EGTA (pH 8).
4. Agilent Bioanalyzer 2100 (Agilent Technologies; Palo Alto, CA, USA).

## **2.6 DNA Microarray Fabrication**

1. DNA spotting buffer: Nexterion Spot 2× (Schott; Jena, Germany).
2. Oligonucleotides: each 100  $\mu$ M (Metabion).
3. Epoxysilane slides (3D-Epoxy Glass Slides; PolyAn; Berlin, Germany).
4. Rinse buffer 1: 0.1 % Triton X-100.
5. Rinse buffer 2: 6 mM HCl.
6. Rinse buffer 3: 0.1 mM KCl.
7. Blocking solution: 0.4 M Tris base, 50 mM ethanolamine (pH 9).
8. Microarray spotter: sciFLEXARRAYER S11 (Scienion AG; Berlin, Germany) (*see Note 4*).

## **2.7 Microarray Hybridization, Washing and Scanning**

1. 20× saline sodium citrate buffer (SSC): 3 M sodium chloride, 0.3 M sodium citrate (pH 7.0).
2. Wash buffer 1: 2× SSC, 0.2 % SDS (freshly added).
3. Wash buffer 2: 2× SSC.
4. Wash buffer 3: 0.2× SSC.
5. ProPlate Multi-array system (Grace Bio-Labs; Bend, OR, USA).
6. ProPlate adhesive seal-strips (Grace Bio-Labs).
7. Hybridization oven or temperature controlled orbital shaker (*see Note 5*).
8. Laser scanner (*see Note 6*).
9. GenePixPro (Molecular Devices LLC; Sunnyvale, CA, USA) or alternative quantification software.

## **2.8 Automated Hybridization, Washing and Readout (Fraunhofer ivD Platform)**

1. Fraunhofer ivD cartridge (e.g., with MRSA-Detect microarray) (BiFlow Systems; Chemnitz, Germany).
2. Fraunhofer ivD platform (base unit).

## **2.9 Microarray Data Analysis**

1. Spreadsheet software such as Microsoft Excel (Microsoft, Redmond, WA, USA).

### 3 Methods

#### 3.1 Genomic DNA Extraction

1. Inoculate 5 mL of CASO-Bouillon with a single colony of *Staphylococcus* species and incubate at 37 °C overnight while shaking.
2. Pour 1.8 mL of the culture into a 2 mL Eppendorf tube and centrifuge at  $\geq 6000 \times g$  for 5 min at 4 °C.
3. Remove as much of the medium as possible and proceed immediately to the next step or freeze the cell pellet at -20 °C overnight if needed (see Note 7).
4. Suspend the pellet in 200 µL enzymatic lysis buffer containing freshly added 100 µg/mL lysostaphin and 100 µg/mL ribonuclease A.
5. Incubate the tube at 37 °C for 20 min until it becomes more transparent and viscous (see Note 8).
6. Add 25 µL proteinase K and 200 µL Buffer AL (provided in the Quiagen DNase kit), vortex, and incubate at 56 °C for 1 h.
7. Add 200 µL of ethanol at -20 °C and mix by inverting the tube several times.
8. Pour the whole suspension into the spin column provided in the kit and centrifuge at  $\geq 6000 \times g$  for 1 min at 4 °C.
9. Transfer the column to a new 2 mL Eppendorf tube, add 200 µL buffer AW1 and centrifuge at  $\geq 6000 \times g$  for 1 min at 4 °C.
10. Add 200 µL buffer AW2 and centrifuge at  $\geq 8000 \times g$  for 1 min at 4 °C.
11. Transfer column to a new 2 mL Eppendorf tube and centrifuge at  $14,000 \times g$  for 2 min at 4 °C.
12. Transfer column to a new 1.5 mL Eppendorf tube and elute the DNA by adding 50 µL water, incubating for 1 min at room temperature and centrifuging at  $\geq 6000 \times g$  for 1 min at room temperature (see Note 9).
13. Estimate the concentration of extracted DNA using a UV/Vis spectrophotometer and check for the correct size of about 20 kb by running an agarose gel (see Note 10).

#### 3.2 Standard PCR Amplification (See Note 11)

1. Carry out PCR to amplify as follows: hot start at 95 °C for 15 min and then 30 cycles of denaturation at 95 °C for 0.5 min, annealing at 50 °C for 0.5 min, and elongation at 68 °C for 1.5 min, followed by a final elongation step at 68 °C for 4 min (see Note 12).
2. For verification of PCR products, carry out an agarose gel electrophoresis on 1 % agarose gels in TAE buffer using 2 µL of the PCR product (see Note 13).

### **3.3 Labeling PCR Amplification**

1. Follow the protocol for a standard PCR amplification using the labeling dNTP mixtures (*see Note 15*).
2. Confirm amplification by agarose gel electrophoresis as above using 0.5  $\mu$ L of the PCR product.

### **3.4 Purification of PCR Products (See Note 14)**

1. Use the PCR Purification Kit to purify the labeled PCR product following the standard protocol and elute in 30  $\mu$ L water.
2. Analyze the purified labeled PCR products for DNA yield and fluorescent dye incorporation by spectrophotometry using the ND-1000 UV/Vis spectrophotometer.
3. Calculate the incorporation rate of the fluorescent dye as shown in the equation below (*see Note 16*).  

$$R = \text{DNA concentration } (\text{ng}/\mu\text{L}) \times 1000 / \text{dye concentration } (\text{pmol}/\mu\text{L}) \times 330 \text{ g/mol}$$
 (*see Note 17*).

### **3.5 DNA Fragmentation**

1. For each microarray hybridization use approximately 750 ng DNA generated from the PCR step and add 1 $\times$  DNase buffer and 0.6 U DNase (*see Note 18*).
2. Digest for 5 min ( $\pm 2$  s) at room temperature and stop the reaction by adding 6  $\mu$ L of DNase stop buffer and incubate at 65 °C for 10 min (*see Note 19*).
3. Keep the digested DNA solution at 4 °C until further use but do not store longer than a few hours.

### **3.6 Microarray Fabrication**

1. Dilute each oligonucleotide probe in 1 $\times$  DNA spotting buffer to a final concentration of 20  $\mu$ M and pipette 30  $\mu$ L of probe solution into designated wells of a 384-well microtiter plate.
2. Use a non-contact spotter to print the probes onto epoxy coated slides, setting the grid to 400  $\mu$ m distance between the spots (*see Note 20*).
3. In order to immobilize the probes after spotting, incubate the slides at 60 °C for 30 min in a drying oven (*see Note 21*).
4. To remove unbound probes, incubate the slides in rinse buffer 1 for 5 min, in rinse buffer 2 for 4 min, in rinse buffer 3 for 10 min and finally in water for 1 min at room temperature, with constant steering.
5. To deactivate free epoxy groups block the slides for 15 min in blocking solution at 50 °C and rinse for 1 min in water.
6. Dry the slides by centrifugation or using a flow of nitrogen.

### **3.7 Microarray Hybridization, Washing, and Image Acquisition**

1. For hybridization use the labeled and fragmented target DNA solution in a total of volume of 46  $\mu$ L, add 0.5  $\mu$ L of 0.05  $\mu$ M labeled hybridization control (*see Note 22*), dilute to a final 2 $\times$  SSC concentration in a total volume of 80  $\mu$ L.
2. Assemble the hybridization chamber on the slide and add the hybridization solution into the chamber, ensuring that no air bubbles form.

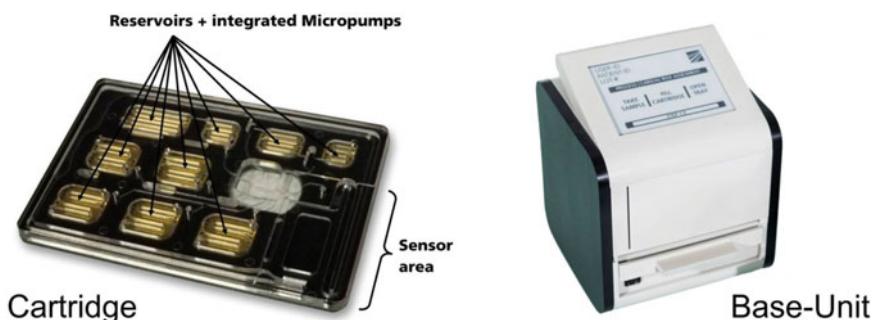
3. Seal the chamber with adhesive seal strips and incubate for 1 h at 48 °C while shaking (*see Note 23*).
4. Following hybridization, wash the slides in wash buffer 1, 2, and 3 each for 10 min at room temperature with constant steering, then dip the slides in water for 1 s and dry under a flow of nitrogen (*see Note 24*).
5. For fluorescence image acquisition, image the slides using a laser scanner at 635 nm and using appropriate PMT/gain settings depending on signal and background intensity (*see Note 25*).
6. Quantify the fluorescent signals with the software provided with the scanner.

**3.8 Automated Hybridization, Washing, and Readout (Fraunhofer ivD Platform)**

1. Prepare the hybridization solution as described above and add to reservoir 5 of the ivD cartridge (*see Note 26*) (Fig. 4).
2. Insert the cartridge into the ivD platform base unit and start the hybridization program (*see Note 27*).
3. The fluorescence image data can either be analyzed automatically within the base unit or exported for external analysis.

**3.9 Data Analysis (See Note 28)**

1. After image acquisition and fluorescent signal quantification, subtract the local background of each spot from the raw spot intensity value, and calculate the mean net signal intensity (NI) and standard deviation (SD) of the replicates.
2. Within each probe set, which is interrogating one mutation site, the probe with the highest signal intensity is termed perfect match (PM) and the remaining probes are marked as mismatch (MM).
3. In order to evaluate the performance of each probe set, calculate the ratios between the MM and PM signal intensities, as the relative signal intensity  $RI_{max(MM)}$  (*see Note 29*).



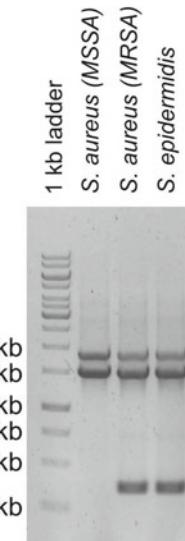
**Fig. 4** Fraunhofer ivD platform for fully automated DNA microarray hybridization and analysis. (*Left*) Lab-on-a-chip cartridge (size: 60 × 40 mm) with 9 reservoirs and integrated micropumps, microfluidic channels, thermal control elements, electronics and a sensor area for microarrays with up to 400 spots. (*Right*) Base-unit (size: 14 × 14 × 14 cm) for control, readout and data analysis of the lab-on-a-chip cartridge. Results can either be transferred to a computer or analyzed directly and presented on the display

4. The MM probe with the highest signal intensity is used for calculation of the relative signal intensity ( $RI_{max(MM)} = NI_{max(MM)} / NI_{PM}$ ) (see Note 30).
5. In addition to the RI value, the limit of detection (LOD) is used to evaluate performance and this is calculated based on the maximum signal intensity ( $NI_{max}$ ) obtained within each probe set based on a no template control (NTC) hybridization plus 3 times the highest standard deviation ( $LOD = NI_{max} + 3 \times SD_{max}$ ) (see Note 31).
6. In addition, the coefficient of variation (CV) is calculated for each set of replicate probes ( $CV = SD / NI_{PM}$ ).
7. Probe sets with a  $CV > 30\%$  of the signal intensity of PM should be flagged and excluded from analysis to ensure that only those probe signals with a high reproducibility are used for the analysis (see Note 32).

---

#### 4 Notes

1. For the rapid protocol, the DNA extraction can be omitted and a labeling PCR can be performed directly from a colony [11]. The subsequent purification can also be omitted.
2. Other thermal cyclers can be used but it is recommended to optimize the cycling conditions on a machine-by-machine basis.
3. Other spectrophotometers can be used.
4. Other spotters can be used but compatibility with other materials and reagents is essential.
5. We used the Array Plate Multi-Well Microarray Hybridization Station (ArrayIt, Sunnyvale, CA, USA).
6. We used the GenePix 4300A (Molecular Devices LLC).
7. Freezing improves the breaking down of the cells, especially of gram positive species.
8. If the suspension does not become transparent, add an additional aliquot of lysostaphin (100 µg/mL) and keep the tube at 37 °C until this occurs.
9. Heating the water for elution to 60 °C, as well as increasing the incubation time up to 5 min can improve the yield.
10. The expected yield can differ from 15 to 120 ng/µL. The DNA concentration can be increased by using less water for the elution (although there is a minimum amount of 30 µL).
11. This non-labeling protocol can be used during the establishment of a new single or multiplex PCR but can be omitted once a new PCR is established. Then the user can continue



**Fig. 5** Example gel electrophoresis image of three *Staphylococcus* strains (MRSA, MSSA, and *S. epidermidis* with *mecA*). Results are shown for a triplex PCR analysis for 23S rDNA (1840 bp), 16S rDNA (1485 bp) and *mecA* (364 bp)

directly with the labeling PCR step. In this example we use a triplex PCR to amplify the 23S rDNA, 16S rDNA, and *mecA* gene.

12. Once completed keep samples at 4 °C until further use or store at -20 °C.
13. The PCR reaction mix contains three primer pairs to identify the three genes: 23S rDNA, 16S rDNA, and *mecA* gene. The DNA fragments corresponding to each of the genes are shown in Fig. 5.
14. Purification of the PCR product is important to remove residual fluorescent dye.
15. Decreasing the salt concentration by using 0.8× Taq buffer can improve amplification of longer amplicons. In contrast, increasing the salt concentration can improve the yield of smaller amplicons and using an elongation temperature of 68 °C improves the amplification of longer fragments in multiplex PCRs [12].
16. The incorporation rate R is defined as the average distance of each labeled nucleotide calculated as shown the equation. For an efficient hybridization, the incorporation rate R should be in the range between 80 and 200.
17. The average mass of a nucleotide is 330 g/mol.
18. The enzymatic digestion of the amplified DNA molecules is necessary to improve the hybridization efficiency. The amount of DNase needed for digestion should be optimized if a new target

is to be amplified, although 0.1 U DNase for 125 ng DNA is a good amount to start with. The labeled PCR amplicons should be digested nonspecifically to a range of 20–200 nucleotides. The fragmentation efficiency can be monitored with an Agilent Bioanalyzer 2100 using a DNA 1000 LabChip kit.

19. This is for heat inactivation of the enzyme. For optimization purposes or development of a new fragmentation protocol the fragmentation efficiency can be monitored with an Agilent Bioanalyzer 2100 using a DNA 1000 LabChip kit.
20. We use the sciFLEXARRAYER for spotting and PolyAn slides for the array surface. Using the indicated settings, the spots will have a diameter of about 150  $\mu\text{m}$ .
21. After this stage, the slides can be stored for several months before use.
22. This is complementary DNA to the immobilized positive control probe.
23. Make sure the slides are protected from light during the hybridization as the fluorophores are light sensitive.
24. For the rapid protocol, the purification step can be omitted. The labeled Amplicon can then be directly fragmented using the standard protocol. Constant stirring of all solutions during washing is needed. Slides should not run dry between washing steps.
25. We use the GenePix 4300A scanner with the corresponding Cy5 filter, as the dye DY-647P1-aadUTP has similar spectral properties as Cy5.
26. All other reservoirs are already delivered with the necessary hybridization and wash buffers. If an empty cartridge is being used, each reservoir can also be filled with custom buffers and reagents individually.
27. After 5 min of hybridization the cartridge will automatically start with the washing and imagine acquisition and the results are obtained after 10 min.
28. The following protocol is recommended to analyze DNA microarrays containing genotyping SNP identification probe sets. The algorithms need to be adjusted for the specific DNA microarrays used.
29. The larger the relative difference between MM and PM signal, the better the discriminative power of the probe set.
30. Only probe sets that show a performance with  $\text{RI}_{\max(\text{MM})} < 0.7$  are used for the analysis. The use of this threshold has been shown to result in high quality discriminations [3, 4, 13].
31. Only probe sets with a perfect match signal intensity above the limit of detection ( $\text{NI}_{\text{PM}} > \text{LOD}$ ) should be used for analysis.

32. For convenience it is recommended to fully automate the algorithms described above (e.g., by using a spreadsheet program, using the raw quantification file from the quantification software as input).

## References

- Schumacher S, Nestler J, Otto T, Wegener M, Ehrentreich-Förster E, Michel D et al (2012) Highly-integrated lab-on-chip system for point-of-care multiparameter analysis. *Lab Chip* 12:464–473
- Antwerpen MH, Schellhase M, Ehrentreich-Förster E, Bier FF, Witte W, Nübel U (2007) DNA microarray for detection of antibiotic resistance determinants in *Bacillus anthracis* and closely related *Bacillus cereus*. *Mol Cell Probes* 21:152–160
- Peter H, Berggrav K, Thomas P, Pfeifer Y, Witte W, Templeton K et al (2012) Direct detection and genotyping of *Klebsiella pneumoniae* carbapenemases from urine by use of a new DNA microarray test. *J Clin Microbiol* 50:3990–3998
- Leinberger DM, Grimm V, Rubtsova M, Weile J, Schröppel K, Wichelhaus TA et al (2010) Integrated detection of extended-spectrum-beta-lactam resistance by DNA microarray-based genotyping of TEM, SHV, and CTX-M genes. *J Clin Microbiol* 48:460–471
- Strommenger B, Schmidt C, Werner G, Roessle-Lorch B, Bachmann TT, Witte W (2007) DNA microarray for the detection of therapeutically relevant antibiotic resistance determinants in clinical isolates of *Staphylococcus aureus*. *Mol Cell Probes* 21:161–170
- Weile J SRD, Bachmann TT, Susa M, Knabbe C (2007) DNA microarray for genotyping multidrug-resistant *Pseudomonas aeruginosa* clinical isolates. *Diagn Microbiol Infect Dis* 59:325–338
- Yu X, Susa M, Weile J, Knabbe C, Schmid RD, Bachmann TT (2007) Rapid and sensitive detection of fluoroquinolone-resistant *Escherichia coli* from urine samples using a genotyping DNA microarray. *Int J Med Microbiol* 297:417–429
- Streit P, Nestler J, Shaporin A, Schulze R, Gessner T (2016) Thermal design of integrated heating for lab-on-a-chip systems. Proceedings of the 17<sup>th</sup> International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), April 18–20, pp. 1–6
- Schumacher S, Ludecke C, Ehrentreich-Förster E, Bier FF (2013) Platform technologies for molecular diagnostics near the patient's bedside. *Adv Biochem Eng Biotechnol* 133:75–87
- Barl T, Dobrindt U, Yu X, Katcoff DJ, Sompolinsky D, Bonacorsi S et al (2008) Genotyping DNA chip for the simultaneous assessment of antibiotic resistance and pathogenic potential of extraintestinal pathogenic *Escherichia coli*. *Int J Antimicrob Agents* 32:272–277
- Jonas D, Speck M, Daschner FD, Grundmann H (2002) Rapid PCR-based identification of methicillin-resistant *Staphylococcus aureus* from screening swabs. *J Clin Microbiol* 40:1821–1823
- Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH (1997) Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques* 23:504–511
- Grimm V, Ezaki S, Susa M, Knabbe C, Schmid RD, Bachmann TT (2004) Use of DNA microarrays for rapid genotyping of TEM beta-lactamases that confer resistance. *J Clin Microbiol* 42:3766–3774

# Chapter 26

## Multiplex Smartphone Diagnostics

Juan L. Martinez-Hurtado, Ali K. Yetisen, and Seok-Hyun Yun

### Abstract

Increasing computing power in smartphones allows for their transformation into point-of-care diagnostic devices. Mobile medical diagnostic applications enable utilization of the processing capabilities of smartphones through their cameras. Hardware attachments or stand-alone versions of smartphone diagnostics have the capability to revolutionize quantitative readouts. Here, we describe a protocol for quantifying commercial colorimetric diagnostic tests with a stand-alone smartphone application. This approach can be used in the multiplexed analyses of biomarker readouts.

**Key words** Smartphone, Diagnostics, Mobile, Medical, Application, Quantitative assays

---

### 1 Introduction

Smartphone applications utilizing their built-in cameras have been developed for dermatology [1], microscopy [2, 3], ophthalmology [4], chemical analyses [5], and paper-based diagnostic devices [6–8]. Although they are normally optimized for photography, smartphone camera algorithms can be modified so that sufficient information can be extracted to enable their use as diagnostic tools. Here, we describe the use of a stand-alone smartphone application that compensates for lighting variability and fully utilizes the sensor capabilities of smartphone cameras. This mobile application is capable of analyzing multiple colorimetric tests allowing for multiplex molecular analyses in scenarios such environmental monitoring, veterinary screening, and medical diagnostics. Specifically, we describe the detailed methodology for a three-analyte commercial test (glucose, protein, and pH) that can be extrapolated for 12 or more analytes. The device uses test strips similar to commercial kits that are utilized for monitoring kidney and liver functions, or screening for diabetes [9]. Although these analytes are commonly measured in laboratories, there is an increasing demand for home

or workplace monitoring due to expanding healthcare expenditures. It is important for diagnosed patients or individuals at risk to continuously monitor their conditions to limit disease progression.

## 2 Materials (See Note 1)

### 2.1 Artificial Urine

1. Prepare artificial urine stock solutions by varying the concentrations of protein and glucose, and the pH level following previously reported protocols [10] and as shown in Table 1.
2. Prepare 200 mL of each stock solution in individual 10 mL vials.
3. The protein concentrations should be 0 mg/dL; 30 mg/dL; 100 mg/dL; and 500 mg/dL.
4. Adjust the pH values to 5, 6, 7, 8, and 9.

**Table 1**  
**Stock concentrations of reagents in artificial urine**

| Component                              | Quantity (g) | Concentration (mM) |
|--|--------------|--------------------|
| Peptone L                              | 37           | 1                  |
| Yeast extract                          | 0.005        | N/A                |
| Lactic acid                            | 0.1          | 1.1                |
| Citric acid                            | 0.4          | 2                  |
| Sodium bicarbonate                     | 2.1          | 25                 |
| Urea                                   | 10           | 170                |
| Uric acid                              | 0.07         | 0.4                |
| Creatinine                             | 0.8          | 7                  |
| Calcium chloride (2H <sub>2</sub> O)   | 0.37         | 2.5                |
| Sodium chloride                        | 5.2          | 90                 |
| Iron II sulphate (7H <sub>2</sub> O)   | 0.0012       | 0.005              |
| Magnesium sulphate (7H <sub>2</sub> O) | 0.49         | 2                  |
| Sodium sulphate (10H <sub>2</sub> O)   | 3.2          | 10                 |
| Potassium dihydrogen phosphate         | 0.95         | 7                  |
| Dipotassium hydrogen phosphate         | 1.2          | 7                  |
| Ammonium chloride                      | 1.3          | 25                 |
| Distilled water to 1 L                 |              |                    |
| Hydrochloric acid to specific pH       |              |                    |
| Sodium hydroxide to specific pH        |              |                    |

5. The glucose values should be 0 mg/dL; 50 mg/dL; 100 mg/dL; 300 mg/dL; and 1000 mg/dL.

---

### 3 Methods (See Note 2)

#### 3.1 Basic Protocol

1. Calibrate the smartphone using three test strips dipped into each of the 15 vials prepared as above.
2. Dip the test strip into artificial urine for 1 s and ensure that all test areas are wetted.
3. Wipe the strip against the edge of the recipient to remove excess liquid.
4. After 60 s reaction time, analyze the test strips through the smartphone application by selecting the appropriate type of test in the application menu.

#### 3.2 Experimental Conditions

1. Light source: white light or fluorescent.
2. Environmental conditions: 24 °C, 60 % relative humidity.

#### 3.3 Technical Variables

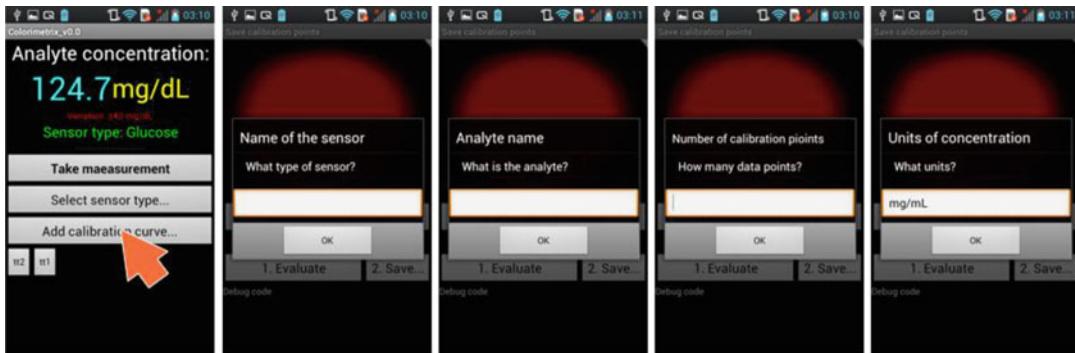
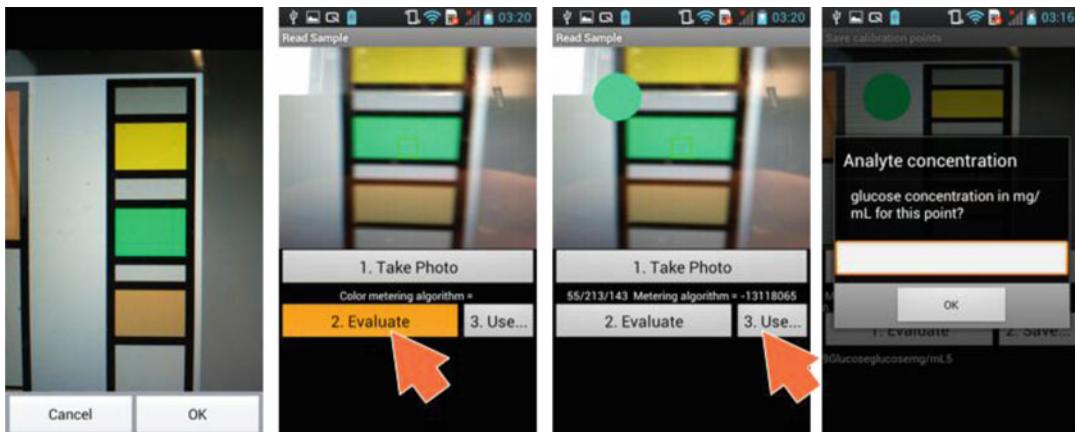
1. Phone model: Samsung Galaxy 5.
2. OS version: Android 2.4.
3. Mobile application: Colorimetrix v1.0 (Fig. 1).

#### 3.4 Calibration

1. Open the app and select “ADD CALIBRATION” (Fig. 2).
2. Follow the on-screen instructions and add information such as test name, analyte, units of concentration, and number of calibration points (see Note 3).



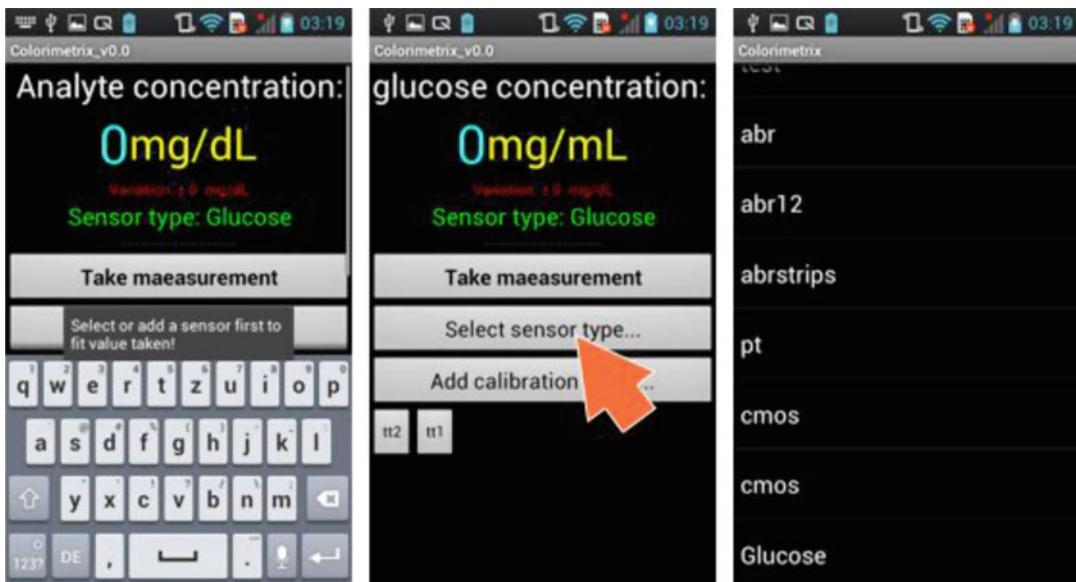
**Fig. 1** Ideal positioning of the smartphone and test strip. Ensure all test zones are well lit with the ambient light kept as constant as possible. To allow for multiplexing, increase the distance from the phone to the test strip until all areas are covered

**Fig. 2** Calibration steps 1 and 2**Fig. 3** Calibration steps 3–7

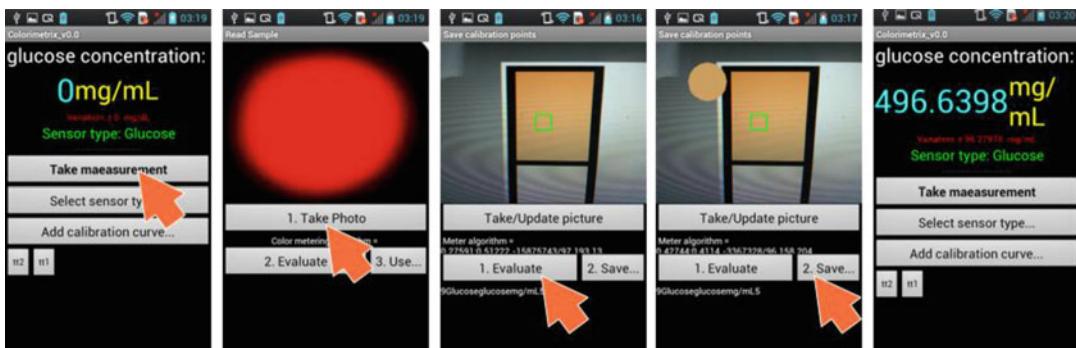
3. Proceed to record each calibration point by aligning the center of the capture screen to the test zone (Fig. 3) (*see Note 4*).
4. Take a photograph of the calibration assay (or reference chart) and approve the image (*see Note 5*).
5. Click “ANALYZE” and wait for a calibration circle to appear on the screen.
6. Click “USE” and enter the concentration value for the first point.
7. Repeat **steps 4–6** until all calibration points are saved.
8. Return to the main screen and click “SELECT CALIBRATION” and select the newly recorded calibration (Fig. 4) (*see Note 6*).

### 3.5 Measurement

1. Click “MEASURE” on the main screen to analyze the test region.
2. Take a photograph and approve the photograph.



**Fig. 4** Calibration step 8



**Fig. 5** Measurement, steps 9 and 10

3. Click “EVALUATE,” wait for circle to appear, click “USE” and record the value as displayed or access it from the database (Fig. 5) (*see Note 7*).
4. Repeat measurement for all test zones (*see Note 8*).
5. Example calibration data are shown in Tables 2, 3, and 4 (*see Notes 9 and 10*).

**Table 2**  
**Example calibration for pH test**

| Sample | Real value | Meas.1 | Meas.2 | Meas.3 | Instrument value |
|--------|------------|--------|--------|--------|------------------|
| 1      | 5          | 5.44   | 4.76   | 5.38   | N/A              |
| 2      | 6          | 5.77   | 6.32   | 6.43   | N/A              |
| 3      | 7          | 6.76   | 7.35   | 7.56   | N/A              |
| 4      | 8          | 8.44   | 8.36   | 7.64   | N/A              |
| 5      | 9          | 9.24   | 9.33   | 8.66   | N/A              |

N/A = not applicable

**Table 3**  
**Example calibration for protein test**

| Sample | Real value | Meas.1 | Meas.2 | Meas.3 | Instrument value |
|--------|------------|--------|--------|--------|------------------|
| 1      | 0          | 23     | 13     | 18     | N/A              |
| 2      | 50         | 14     | 56     | 67     | N/A              |
| 3      | 100        | 77     | 167    | 154    | N/A              |
| 4      | 500        | 567    | 435    | 534    | N/A              |
| 5      | 1000       | #      | #      | #      | N/A              |

N/A = not applicable

**Table 4**  
**Example calibration for glucose test**

| Sample | Real value | Meas.1 | Meas.2 | Meas.3 | Instrument value |
|--------|------------|--------|--------|--------|------------------|
| 1      | 0          | 8      | 24     | 3      | N/A              |
| 2      | 30         | 25     | 81     | 78     | N/A              |
| 3      | 100        | 156    | 63     | 189    | N/A              |
| 4      | 300        | 367    | 266    | 376    | N/A              |
| 5      | 1000       | 1078   | 922    | 945    | N/A              |

N/A = not applicable

#### 4 Notes

1. All solutions should be prepared with purified deionized water and analytical grade reagents, and stored and used at room temperature unless indicated otherwise. In this study, we prepared an artificial urine solution for practical purposes although native body fluids can also be used. Waste disposal regulations should be followed for used solutions and materials.

2. To test clinical urine samples, instructions from the test provider should be followed as described in the methods section.
3. Record a minimum number of five calibration points. The application allows for recording any number of points but it performs optimally with five calibration points. Additional calibration points will increase the accuracy of the measurements, which are limited by the sensitivity of the colorimetric test.
4. The mobile application uses a 100 pixel area per test zone and it should be entirely covered by the test zone color.
5. The calibration photograph can be retaken if outside of the test area as many times as needed.
6. For repeat measurements, it is not necessary to recalibrate the application. Recorded calibrations can be selected directly from the main menu.
7. For multiplexed assays, select “MULTIPLEXING” and align test zones with multiplex test zones on the capturing screen. Autodetection can also be used to determine the test zones and capture the images.
8. Versions of the application allow selection of other types of analytes for different commercial test strips. Such multiplex tests may target applications such as drug abuse or assimilation, urine adulteration, nutritional factors, and hormone levels.
9. It is suggested that multiple measurements should be performed to validate the use of the smartphone application.
10. Each triplet allows for accuracy comparison of the smartphone application to standard techniques. The presented mobile medical application can also be applied for quantifying photonic crystal arrays and holographic sensors [11–21]. The described application has been designed for laboratory use and for clinical testing. The next stage will be to seek regulatory approval for each test and its intended use [22].

## References

1. Kroemer S, Frühauf J, Campbell TM, Massone C, Schwantzer G, Soyer HP et al (2011) Mobile teledermatology for skin tumour screening: diagnostic accuracy of clinical and dermoscopic image tele-evaluation using cellular phones. *Br J Dermatol* 164:973–979
2. Breslauer DN, Maamari RN, Switz NA, Lam WA, Fletcher DA (2009) Mobile phonebased clinical microscopy for global health applications. *PLoS ONE* 4:e6320
3. Smith ZJ, Chu K, Espenson AR, Rahimzadeh M, Gryshuk A, Molinaro M et al (2011) Cellphone-based platform for biomedical device development and education applications. *PLoS ONE* 6:e17150
4. Pamplona VF, Mohan A, Oliveira MM, Raskar R (2010) Dual of Shack-Hartmann optometry using mobile phones. *Frontiers in Optics, Optical Society of America, Rochester, NY, paper FTuB4*.doi:[10.1364/FIO.2010.FTuB4](https://doi.org/10.1364/FIO.2010.FTuB4)
5. Zhu H, Mavandadi S, Coskun AF, Yaglidere O, Ozcan A (2011) Optofluidic fluorescent imaging cytometry on a cell phone. *Anal Chem* 83:6641–6647
6. Martinez AW, Phillips ST, Whitesides GM (2008) Three-dimensional microfluidic devices fabricated in layered paper and tape. *Proc Natl Acad Sci USA* 105:19606–19611
7. Wang S, Zhao X, Khimji I, Akbas R, Qiu W, Edwards D et al (2011) Integration of cellphone

- imaging with microchip ELISA to detect ovarian cancer HE4 biomarker in urine at the point-of-care. *Lab Chip* 11:3411–3418
8. Pollock NR, Rolland JP, Kumar S, Beattie PD, Jain S, Noubari F et al (2012) A paper-based multiplexed transaminase test for low-cost, point-of-care liver function testing. *Sci Transl Med* 4:152ra29
  9. Yetisen AK, Akram MS, Lowe CR (2013) Paper-based microfluidic point-of-care diagnostic devices. *Lab Chip* 13:2210–2251
  10. Martinez AW, Phillips ST, Butte MJ, Whitesides GM (2007) Patterned paper as a platform for inexpensive, low-volume, portable bioassays. *Angew Chem Int Ed Engl* 46:1318–1320
  11. Yetisen AK, Volpatti LR, Humar M, Kwok SJ, Pavlichenko I, Kim KS et al (2016) Photonic hydrogel sensors. *Biotechnol Adv* 34:250–271
  12. Yetisen AK, Naydenova I, Vasconcellos FC, Blyth J, Lowe CR (2014) Holographic sensors: three-dimensional analyte-sensitive nanostructures and their applications. *Chem Rev* 114:10654–10696
  13. Yetisen AK, Montelongo Y, Qasim MM, Butt H, Wilkinson TD, Monteiro MJ, Yun SH (2015) Photonic nanosensor for colorimetric detection of metal ions. *Anal Chem* 87:5101–5108
  14. Yetisen AK, Montelongo Y, Vasconcellos FC, Martinez-Hurtado JL, Neupane S, Butt H et al (2014) Reusable, robust, and accurate laser-generated photonic nanosensor. *Nano Lett* 14:3587–3593
  15. Yetisen AK, Butt H, Vasconcellos FC, Montelongo Y, Davidson CAB, Blyth J et al (2014) Light-directed writing of chemically tunable narrow-band holographic sensors. *Adv Opt Mater* 2:250–254. doi:[10.1002/adom.201300375](https://doi.org/10.1002/adom.201300375)
  16. Yetisen AK, Qasim M, Nosheen S, Wilkinson TD, Lowe CR (2014) Pulsed laser writing of holographic nanosensors. *J Mater Chem C* 2:3569–3576. doi:[10.1039/C3TC32507E](https://doi.org/10.1039/C3TC32507E)
  17. Tsangarides CP, Yetisen AK, Vasconcellos FC, Montelongo Y, Qasim MM, Lowe CR et al (2014) Computational modelling and characterisation of nanoparticle-based tuneable photonic crystal sensors. *RSC Adv* 4:10454–10461
  18. Martinez-Hurtado JL, Lowe CR (2014) Ammonia-sensitive photonic structures fabricated in nafion membranes by laser ablation. *ACS Appl Mater Interfaces* 6:8903–8908
  19. Martínez-Hurtado JL, Davidson CA, Blyth J, Lowe CR (2010) Holographic detection of hydrocarbon gases and other volatile organic compounds. *Langmuir* 26:15694–15699
  20. Martinez-Hurtado JL, Lowe CR (2015) An integrated photonic-diffusion model for holographic sensors in polymeric matrices. *J Membr Sci* 495:14–19. doi:[10.1016/j.memsci.2015.07.064](https://doi.org/10.1016/j.memsci.2015.07.064)
  21. Martinez-Hurtado JL, Akram MS, Yetisen AK (2013) Iridescence in meat caused by surface gratings. *Foods* 2:499–506. doi:[10.3390/foods2040499](https://doi.org/10.3390/foods2040499)
  22. Yetisen AK, Martinez-Hurtado JL, Vasconcellos FC, Simsekler MCE, Akram MS, Lowe CR (2014) The regulation of mobile medical applications. *Lab Chip* 14:833–840

# Chapter 27

## Development of a User-Friendly App for Assisting Anticoagulation Treatment

Johannes Vegt

### Abstract

Blood coagulation time is an important factor to consider for postoperative and cardiac disorder patients who have been prescribed anticoagulant coagulant medications. This chapter describes a patient self-management system for assessment of blood coagulation times and determining appropriate anticoagulant dosages using a test strip device and the Coagu app. This app can also be used as a patient reminder of treatment times and to monitor treatment and effects over time.

**Key words** Blood coagulation, Clotting cascade, Anticoagulant test strips, Coagu app, Cardiac disease

---

### 1 Introduction

Coagulation is the process in which blood changes from a liquid to a gel. The mechanism involves a cascade of reactions in which enzyme precursors are successively and rapidly activated to catalyze the next reaction, ultimately resulting in cross-linking of fibrin polymers and clot formation [1]. The extrinsic coagulation cascade ensues in the following sequence: tissue damage → activation of tissue factor → activation of factor VII → activation of factor X → cleavage of prothrombin to produce thrombin → cleavage of fibrinogen to produce fibrin → and polymerization of fibrin. Anticoagulants are often used in medicine to help prevent blood clots from forming and to reduce the risk of heart attack, stroke, and blockages in arteries and veins [2].

When anticoagulants are prescribed for a patient, it is essential to carry out periodic monitoring of the time it takes the blood to clot using the international normalized ratio (INR) to adjust the dose as necessary [3]. Many patients are advised to carry out the relevant checks themselves using coagulation test strips combined with a suitable reader [4]. Based on these results, the appropriate dosage of anticoagulant should be administered to achieve the required target range. This management process would be facilitated by the use of Coagu app, developed by Appamedix UG in Berlin, Germany [5].

In 2013, the Coagu app was recognized for its usability by the International Design Centre Berlin and as an example of a particularly user-friendly product on the international Funkausstellung (IFA) [6]. The app draws on a universal design, which means that it can be used by all ages alike. Currently, it is now used by patients in over 70 countries. The evidence shows that greater use of self-monitoring offers clinical and patient benefits and is likely to result in reductions in heart attacks and strokes caused by blood clots [7].

## 2 Materials

1. Smartphone or tablet (*see Note 1*).
2. Coagu app.
3. Test strips and meter (*see Note 2*).

## 3 Methods

### **3.1 Preparation: Determination of INR**

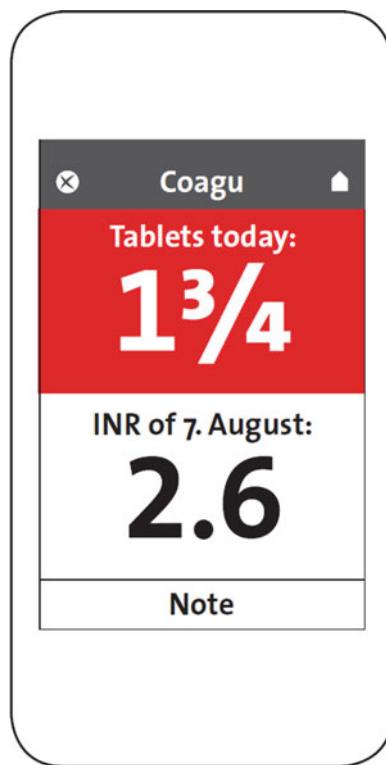
1. Insert the test strip into the coagulation meter (Fig. 1).
2. Lightly pierce the tip of a finger.
3. Immediately apply the resulting blood drop to the test strip.
4. After 60 s, the measured value appears on the meter display.



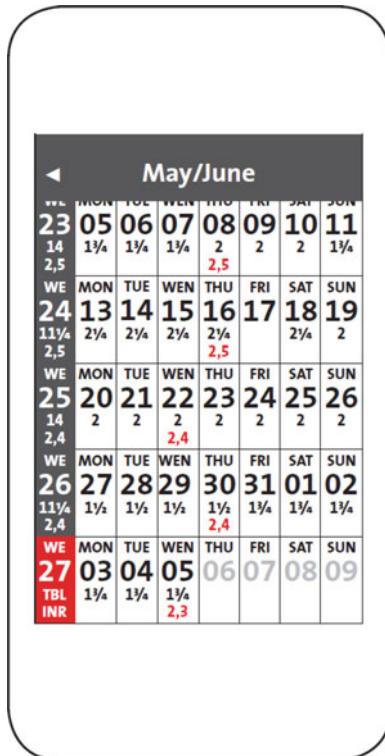
**Fig. 1** (A) The figures of the last INR (International Normalized Ratio) measurement gradually fade away over seven days. This reminds the patient to take the next measurement. (B, C) Lightly pierce the top of a finger. (D) Immediately apply resulting blood drop to the test strip that is in the measurement device (*see Note 2*). (E) After 60 s, the measured value appears on the meter display. (F) The patient enters the INR Value, using a number picker and stores it. (G) The measured and saved Value appears with the actual date on the start page of the Coagu app

**3.2 App Usage:**  
**Determination  
of Anticoagulant  
Dosage (Fig. 1)**  
**(See Note 3)**

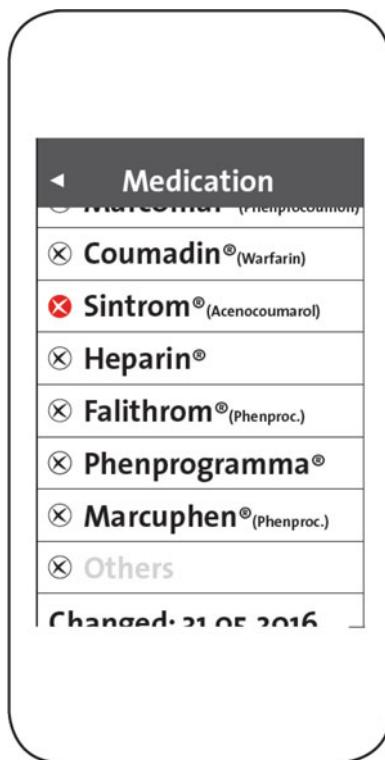
1. Open the Coagu app on the smartphone or tablet.
2. First use: each patient may configure the app for their specific needs including target INR range, drug varieties, and dosage (*see Note 4*).
3. Input the measured value for storage and display in the calendar and histogram of the app (Figs. 2 and 3) (*see Note 5*).
4. The patient determines which medication and what dosage he should take based on the current INR reading (Fig. 4).
5. Measurement history: the entered values are stored and can be visualized as a trend over a period of up to 6 months in a histogram (Fig. 5) (*see Note 6*).
6. Notifications: the patient is reminded up to twice a day via a notification which occurs as an audible and visual signal advising them to set the time for taking their medication or some other actions (*see Note 7*).
7. Patient comments: the patient may add a comment on the entry each day individually (*see Note 8*).



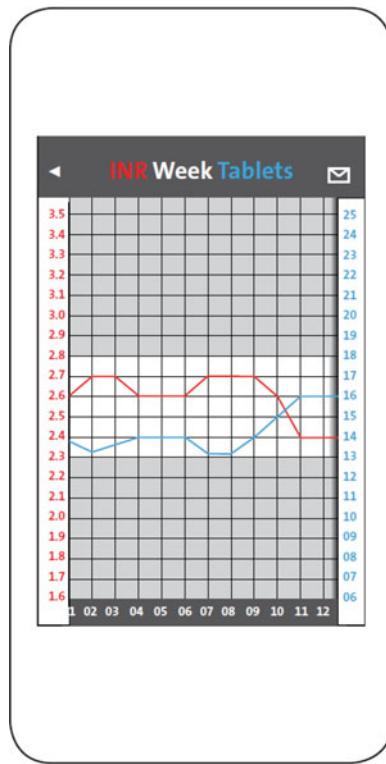
**Fig. 2** Anticoagulation patients have to take their medication every day. The app reminds them of the dose to be taken that day. The patient confirms having taken the tablets with a tap on the display. This is then registered in the histogram and in the calendar



**Fig. 3** Calendar showing the input daily INR values. Red values indicate a high reading



**Fig. 4** The patient chooses a prescribed medication from a list manually. A confirmation that the medication has been taken is registered automatically in the calendar. With a tap on the calendar, a daily summary appears (not shown). Here, the entered notes can be read or new notes can be entered



**Fig. 5** The histogram shows the measured INR values and medication use and relates these over time. Tendencies can be seen over a period of 6 months

---

#### 4 Notes

1. iOS (Apple) and Android (GooglePlay) smartphones and tablets can be used. It should be noted that there is no connection between the developer of the Coagu App Johannes Vegt (Appamedix UG i.Gr.) and the producer of the CoaguChek measurement device (Hoffmann-La Roche AG).
2. A lancing device comes with the CoaguChek instrument, although other sources can be used.
3. Data protection is paramount to the functioning of the app. For the Coagu app, the data on the smartphone or tablet is being managed by the patient alone and it is not stored on any external server or as cloud data. The patient makes the decision about data sharing. It should be noted that in its present form the Coagu app is not a medical product. Since the app is sold worldwide, this would require testing and approval by the regulatory authorities in the relevant countries. The escalation of the app to medicinal product would be possible if a health insurance company accepts liability.

4. The target range is the maximum and minimum amount INR values that patients receive from their doctor.
5. The average age of the audience is about 60 years. Accordingly, relatively large touch-sensitive surfaces were created to compensate for possible motor inaccuracies. In addition, larger font sizes are used at crucial points to compensate for potential visual impairments.
6. Here the user can see the correlation between drug dosage and reading.
7. For example, if the medication is to be interrupted, this can be listed in the app.
8. The comment is stored in the calendar and is available for 6 months. Through the app, the patient can also contact their physician by phone, SMS, or email. Technically, it would be possible for the patient to share their information with their doctor so that their history could be updated, if required. This is only possible in cooperation with the health insurance and higher authorities.

## References

1. Luchtman-Jones L, Broze GJ Jr (1995) The current status of coagulation. *Ann Med* 27:47–52
2. Chakrabarti R, Das SK (2007) Advances in antithrombotic agents. *Cardiovasc Hematol Agents Med Chem* 5:175–185
3. Jespersen J, Hansen MS, Dyerberg J, Ingerslev J, Jensen G, Jørgensen M et al (1991) Standardized prothrombin time determinations and optimal anticoagulant therapy. *Ugeskr Laeger* 153:355–360
4. van den Besselaar AM, Meeuwisse-Braun J, Schaefer-van Mansfeld H, van Rijn C, Witteveen E (2000) A comparison between capillary and venous blood international normalized ratio determinations in a portable prothrombin time device. *Blood Coagul Fibrinolysis* 11:559–562
5. <http://www.coagu.com/en/>
6. [http://www.inr-austria.at/index.php?article\\_id=6](http://www.inr-austria.at/index.php?article_id=6)
7. Tideman PA, Tirimacco R, St John A, Roberts GW (2015) How to manage warfarin therapy. *Aust Prescr* 38:44–48

## **Part IV**

### **Future Directions**

# **Chapter 28**

## **Multiplex Biomarker Approaches to Enable Point-of-Care Testing and Personalized Medicine**

**Paul C. Guest**

### **Abstract**

This chapter describes how current and future innovations driven by application of multiplex biomarker techniques can help in earlier and more efficacious treatment of patients, suffering from the world's most devastating and costly diseases. The application of new miniaturized biosensors and transducers will enable point-of-care testing by facilitating analysis of a single drop of a blood within the time span of a visit to the doctor's office. It is anticipated that the scoring algorithms used with future tests will incorporate both biochemical and clinical data, resulting in specific profiles for each patient or tested subject to enable personalized medicine approaches.

**Key words** Disease, Multiplex biomarkers, Lab-on-a-chip, Smartphone apps, Personalized medicine

Diseases such as cardiovascular disorders, type 2 diabetes, obesity, cancer, and mental health disorders can affect people of both sexes at different ages and seriously impair medical health, quality of life, social well-being, and productivity, with a significant negative impact on society and the economy. According to the World Health Organization (WHO), the global burden of noncommunicable diseases is expected reach approximately exceed 30 trillion US dollars within the next 20 years and account for almost half of the global GDP [1]. Given the urgent medical need and the importance of counteracting these negative effects, it is now vital that point-of-care testing gains wider acceptance on the marketplace. The existing methods are simply not working and may also be equally draining on the economy due to the large-scale instrumentation required along with long processing times and the delay in receiving, communicating, and acting upon diagnostic results. One means of overcoming these issues which has been emerging over the last decade is through clinically validated lab-on-a-chip (LOC) approaches. LOCs provide many advantages over existing methods, such as the need for lower biosample volumes, less waste, lower fabrication and reagent costs, improved process control due to faster system response times, and compactness due to integration

and high parallelization of functionality [2, 3]. Most importantly, this translates to a reduced waiting time for the results and tests can even be performed right there in the doctor's office.

One emerging LOC approach involves a new way to diagnose and manage HIV infections. Around 40 million people are infected with HIV in the world today, yet scarcely more than one million of these receive the correct anti-retroviral treatment. In fact, most of the people with HIV have never even been tested for the disease. Currently, measuring the circulating levels of CD4 positive lymphocytes in a person's blood is the best way to determine if they have HIV and this can also be used for tracking the infection. This is typically achieved using a technique called flow cytometry that is not available in most developing countries where the presence of HIV is disproportionately high since the instrumentation is large, expensive, and complicated and requires trained technicians in the operation and interpretation of data. Recently, a company called ClonDiag developed a LOC device that employs similar static image analysis and counting of CD4 positive cells as in flow cytometry but in a compact and transportable package that does not require extensive laboratory training [4]. Furthermore, it requires only 25 µL of blood and can deliver results within only 20 min. Another LOC device that was developed more recently consists of a printed flexible plastic microchip that can capture and quantify viruses such as HIV in several types of biosamples, including blood through an electrical sensing approach [5]. Another group has developed tuberculosis diagnostic LOC device that is 96% accurate for detection of tuberculosis [6]. The device uses an immunofluorescence-based microtip sensor that can detect tuberculosis complex cells in sputum within 30 min. Concentration mechanisms based on flow circulation and electric field are combined at different scales to concentrate target bacteria in 1 mL samples onto the surfaces of microscale tips. Multiplex antibody-based biomarker tests have also been developed on LOC devices that are the size of a credit card [7]. One specific application is the detection of prostate cancer using either surface-enhanced RAMAN scattering [8] or voltage-based [9] readouts. In general, the procedure involves the application of a blood drop into a chamber in the card, followed by the insertion of the card into a small table top analyzer. The diagnosis is then read out as a "score" in less than 15 min. One of the anticipated major benefits of all of these LOC tests is that the rapid diagnosis will help to cut down on waiting times for results of laboratory tests which can often take several days or even weeks using the customary methods. Furthermore, these devices can be manufactured to contain a universal serial bus (USB) that would enable connection to a computer and transmission of the data to other devices such as smartphones via near-field communication.

Point-of-care testing now encompasses a variety of devices ranging from larger table-top instruments to implanted, wearable, and handheld equipment. The handheld systems typically consist of disposable strips incorporated into a cassette that allows addition of the sample, conduction of the test, and signal generation that is usually interpreted visually or through an inexpensive reader. There are now such devices that can be used to indicate the presence of a heart attack, diabetes, blood clotting capacity, some infectious diseases, or even to measure the levels of substances such as alcohol and drugs of abuse. The most familiar example is most likely the Clearblue Digital Pregnancy test® produced by SPD Swiss Precision Diagnostics GmbH [10] which gives both a readout and an indication of the number of weeks since conception.

Growth in the LOC diagnostic and monitoring systems reflects patient preferences for being seen in a doctor's office or in a clinic rather than a central laboratory. This not only requires solutions that use lower sample volumes with reduced effort and time spent carrying out the tests, and lower costs, it also calls for an increase in connectivity solutions. Furthermore, massive companies such as Apple and Google have shown an escalating interest in the diagnostic market. This has been mainly driven by opportunities to connect an app result with a diagnostic answer via smart software. Likewise, pharmaceutical companies now consider that low cost and time-effective biomarkers are essential for decision-making in clinical trials and drug discovery efforts. Therefore, an important requirement of new point-of-care devices involves incorporation of mobile communication and internet capabilities so that the associated data can be formatted for ease of presentation and interpretation by the users. There has also been an increased move to incorporate networked computing to aid in such functions as disease prediction, diagnosis, prognosis and even for monitoring medication compliance. This may lead to the development of bio-profiles or biomarker fingerprints for individuals that combines genomic, proteomic, transcriptomic, metabolomic, and/or imaging data with patient physiometrics and histories.

In the twenty-first century, mobile phones have become virtually ubiquitous, with an estimated six billion subscriptions at the end of 2011 [11]. There has now been a convergence in the entire technological concept which has resulted in development of the smartphone. This combines general voice and text messaging services with computation functions to support applications (apps), sensors, as well as wireless internet accessibility and connectivity with other smart devices. These latter features make the smartphone an attractive and user-friendly platform for health and disease management [12, 13]. Basic mobile phone-based interventions have already shown promise and have led to improved outcomes in a variety of health conditions, diseases, and control of certain habits. A review of clinical trials that incorporated health care

interventions assisted through smartphone apps indicated improvements in 61 % of the outcomes such as better attendance of patients at appointments, faster diagnosis and treatment, improved communication [14]. In addition, the benefits included behavioral improvements such as cessation of smoking and better medication compliance, as well as clinical factors such as better blood sugar control, decreased symptoms of asthma, and lower behavioral stress levels.

Along with the theme of this book, there have now been developments of multiplex biomarker tests on hand-held and smartphone-based devices, which use a 3D-printed opto-mechanical interface to illuminate each reaction well on a multiwell plate [15]. The resulting images can be transmitted to databases for analysis and the results can be returned to the user within 1 min. This mobile platform has already been tested with a successful outcome in a clinical setting using multiplex assays for mumps, measles, and herpes simplex I and II virus immunoglobulins using the smartphone camera optics function for accumulation and transmission of the readouts. Other optical-based smartphone detection assays include the measurement of nucleic acids by PCR [16], avian flu virus subtyping by immunoassay [17], detection of kaposi's sarcoma by solar thermal PCR [18], hemoglobin and HIV levels by immunoassay [19], prostate-specific antigen using enzymatic amplification, and a chromogenic substrate [20]. It is easy to imagine that other LOC and smartphone-based system will be developed for other important diseases such as various types of cancer, cardiovascular conditions, diabetes, and mental disorders.

In conclusion, the movement has been toward improving patient care through the initial study of various diseases using multiplex biomarker approaches and then deploying these technologies as hand-held microfluidic devices linked with smartphone apps for ease of use in the clinical setting. This is currently our best approach of achieving a paradigm shift personalized medicine. This will allow persons to be treated based on their individual biomarker profiles rather than as one of many with a particular disease using a standard blockbuster drug. In addition, the use of multiplex biomarker tests on LOC and/or handheld devices that are capable of distinguishing disease subtypes may be useful for rapid identification of patients who are most likely to respond to specific medications either alone or in combination with other drugs. This general approach has been used increasing in the field of cancer treatment. For example, the measurement of human epidermal growth factor receptor 2 (HER2) at both the gene and transcript level can be used to identify those breast cancer patients who are most likely to benefit from treatment with Trastuzumab (also known as Herceptin), which was developed by Genentech in the 1990s [21]. It obvious that such an approach could result in more effective treatment of patients with fewer side effects and, thus, a decreased proportion of patients who decide to discontinue medication because of severe side effects.

## References

1. [http://www3.weforum.org/docs/WEF\\_Harvard\\_HE\\_GlobalEconomicBurdenNonCommunicableDiseases\\_2011.pdf](http://www3.weforum.org/docs/WEF_Harvard_HE_GlobalEconomicBurdenNonCommunicableDiseases_2011.pdf)
2. Pawell RS, Inglis DW, Barber TJ, Taylor RA (2013) Manufacturing and wetting low-cost microfluidic cell separation devices. *Biomicrofluidics* 7:056501. doi:[10.1063/1.4821315](https://doi.org/10.1063/1.4821315)
3. Yager P, Edwards T, Fu E, Helton K, Nelson K, Tam MR et al (2006) Microfluidic diagnostic technologies for global public health. *Nature* 442:412–418
4. Ermantraut E, Bickel R, Schulz T, Ullrich T, Tuchscheerer J (2011) Device and method for the detection of particles. USPTO Patent US8040494. Company: Clondiag GmbH
5. Shafiee H, Kanakasabapathy MK, Juillard F, Keser M, Sadasivam M, Yuksekka M (2015) Printed flexible plastic microchip for viral load measurement through quantitative detection of viruses in plasma and saliva. *Sci Rep* 5:9919. doi:[10.1038/srep09919](https://doi.org/10.1038/srep09919)
6. Kim JH, Yeo WH, Shu Z, Soelberg SD, Inoue S, Kalyanasundaram D (2012) Immunosensor towards low-cost, rapid diagnosis of tuberculosis. *Lab Chip* 12:1437–1440
7. Schumacher S, Nestler J, Otto T, Wegener M, Ehrentreich-Förster E, Michel D et al (2012) Highly-integrated lab-on-chip system for point-of-care multiparameter analysis. *Lab Chip* 12:464–473
8. Gao R, Cheng Z, deMello AJ, Choo J (2016) Wash-free magnetic immunoassay of the PSA cancer marker using SERS and droplet microfluidics. *Lab Chip* 16:1022–1029
9. Parra-Cabrera C, Samitier J, Homs-Corbera A (2016) Multiple biomarkers biosensor with just-in-time functionalization: application to prostate cancer detection. *Biosens Bioelectron* 77:1192–1200
10. Johnson S, Cushion M, Bond S, Godbert S, Pike J (2015) Comparison of analytical sensitivity and women's interpretation of home pregnancy tests. *Clin Chem Lab Med* 53:391–402
11. <http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats/a#subscribers>
12. Klasnja P, Pratt W (2012) Healthcare in the pocket: mapping the space of mobile-phone health interventions. *J Biomed Inform* 45: 184–198
13. Ventola CL (2014) Mobile devices and apps for health care professionals: uses and benefits. *P T* 39:356–364
14. Krishna S, Boren SA, Balas EA (2009) Healthcare via cell phones: a systematic review. *Telemed J E Health* 15:231–240
15. Berg B, Cortazar B, Tseng D, Ozkan H, Feng S, Wei Q et al (2015) Cellphone-based handheld micro-plate reader for point-of-care testing of enzyme-linked immunosorbent assays. *ACS Nano* 9:7857–7866
16. Liao SC, Peng J, Mauk MG, Awasthi S, Song J, Friedman H (2016) Smart cup: a minimally-instrumented, smartphone-based point-of-care molecular diagnostic device. *Sens Actuators B Chem* 229:232–238
17. Yeo SJ, Choi K, Cuc BT, Hong NN, Bao DT, Ngoc NM et al (2016) Smartphone-based fluorescent diagnostic system for highly pathogenic H5N1 viruses. *Theranostics* 6: 231–242
18. Snodgrass R, Gardner A, Jiang L, Fu C, Ceserman E, Erickson D (2016) KS-Detect – validation of solar thermal PCR for the diagnosis of Kaposi's sarcoma using pseudo-biopsy samples. *PLoS One* 11:e0147636
19. Guo T, Patnaik R, Kuhlmann K, Rai AJ, Sia SK (2015) Smartphone dongle for simultaneous measurement of hemoglobin concentration and detection of HIV antibodies. *Lab Chip* 15:3514–3520
20. Barbosa AI, Gehlot P, Sidapra K, Edwards AD, Reis NM (2015) Portable smartphone quantitation of prostate specific antigen (PSA) in a fluoropolymer microfluidic device. *Biosens Bioelectron* 70:5–14
21. Demonty G, Bernard-Marty C, Puglisi F, Mancini I, Piccart M (1997) Progress and new standards of care in the management of HER-2 positive breast cancer. *Eur J Cancer* 43: 497–509

# INDEX

## A

### Algorithm

- clinical score ..... 90, 93, 103, 104, 108, 109, 111, 112, 116, 118
- machine learning ..... 90, 104, 111, 119
- patient data ..... 51, 93, 117, 118
- statistics ..... 50, 104, 108

## B

### Biomarker

- genomics ..... 5, 275, 313
- metabolomics ..... 5, 46, 47, 49–51, 313
- proteomics ..... 5, 7, 28, 41–49, 57, 313
- transcriptomics ..... 5, 313

### Biosample

- blood ..... 161–167
- brain ..... 235, 241
- cells ..... 86
- cerebrospinal fluid ..... 67
- heart ..... 313
- plasma ..... 13, 67
- serum ..... 13, 67
- spittle ..... 13
- tissue ..... 6, 69, 209

## C

### Cytomics

### Drug discovery

- clinical trial ..... 5, 313
- efficacy ..... 8, 14
- FDA ..... 4
- imaging ..... 313
- marketing ..... 3
- side effect ..... 27
- stratification ..... 69
- surrogate biomarker ..... 8

## F

### Flow cytometry

- ..... 76, 80, 88, 91, 312

## L

### Lab-on-a-chip

- immunoassay ..... 283
- micrfluidics ..... 284, 290
- microarray ..... 284, 290

## M

### Metabolomics

- mass spectrometry ..... 8, 11
- NMR ..... 276

## P

### Personalized medicine

- ..... 15, 311–314
- Proteomics ..... 7, 41, 48, 57–71, 149, 150, 152–158, 205–220, 267–273

### 2D gel electrophoresis

- 2D-DIGE ..... 41, 48, 205–212

### mass spectrometry

- ITRAQ ..... 267–273
- label-free ..... 57–71
- MS/MS ..... 57–71
- shotgun ..... 7, 41, 58, 60–61, 69, 70, 267–273
- SILAC ..... 149, 150, 152–158
- SRM ..... 213–220
- multiplex immunoassay ..... 169–175

## S

### Smartphone

- app ..... 30, 303–308
- tablet ..... 304–307

### DNA/RNA

- microarray ..... 5, 13, 106, 135, 143, 283, 284, 287, 290
- RT-PCR ..... 131
- SNP analysis ..... 143, 284