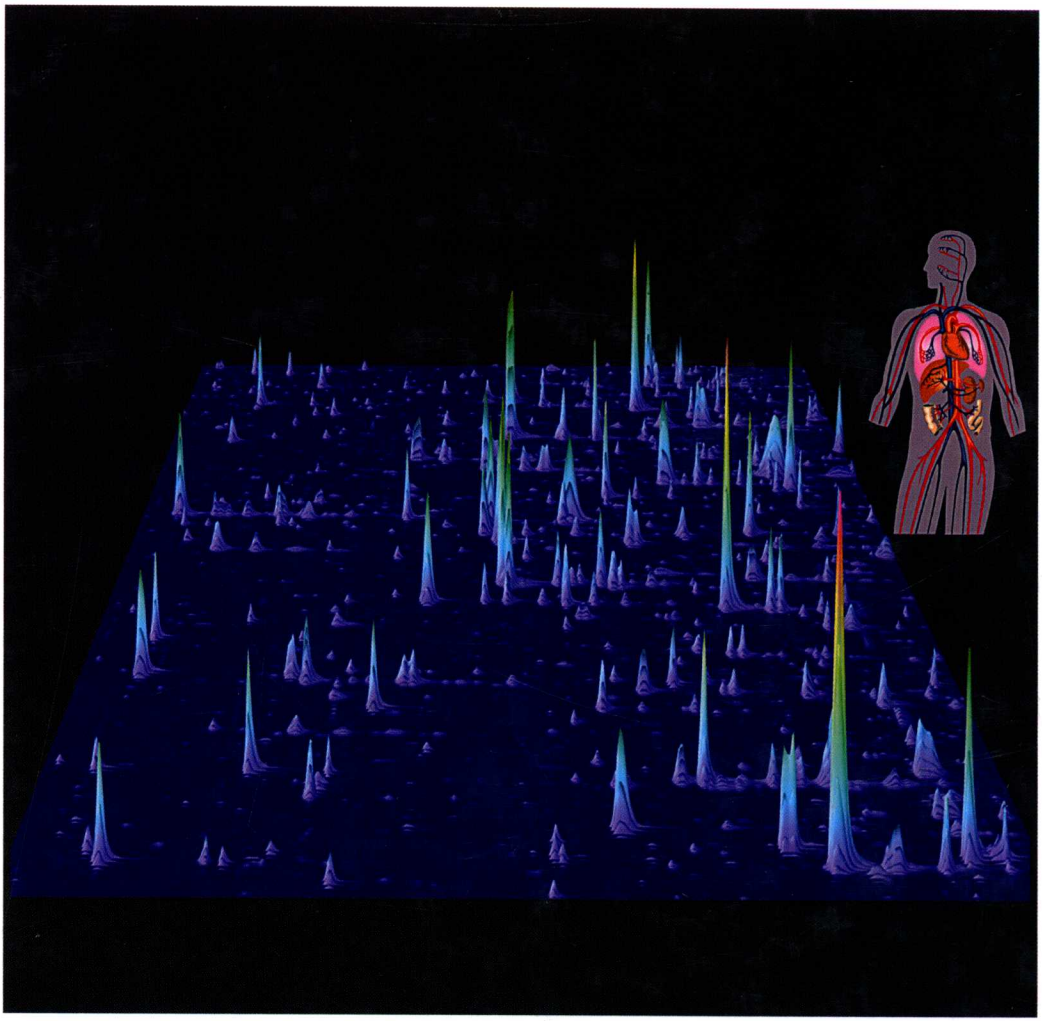
RSC Drug Discovery

Edited by Péter Horvatovich and Rainer Bischoff

Comprehensive Biomarker Discovery and Validation for Clinical Application



RSC Publishing

Comprehensive Biomarker Discovery and Validation for Clinical Application

Edited by

Péter Horvatovich and Rainer Bischoff
University of Groningen, Groningen, The Netherlands
Email: p.l.horvatovich@rug.nl; r.p.h.bischoff@rug.nl

RSC Drug Discovery Series No. 33

ISBN: 978-1-84973-422-6
ISSN: 2041-3203

A catalogue record for this book is available from the British Library

Published by The Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road,
Cambridge CB4 0WF, UK

Registered Charity Number 207890

For further information see our web site at www.rsc.org

Printed in the United Kingdom by CPI Group (UK) Ltd, Croydon, CR0 4YY

Comprehensive Biomarker Discovery and Validation for Clinical Application

RSC Drug Discovery Series

Editor-in-Chief
Professor David E. Thurston, *King's College, London, UK*

Series Editors:
Dr David Fox, *Vulpine Science and Learning, UK*
Professor Ana Martinez, *Instituto de Quimica Medica-CSIC, Spain*
Professor David Rotella, *Montclair State University, USA*

Advisor to the Board:
Professor Robin Ganellin, *University College London, UK*

Titles in the Series:

1: Metabolism, Pharmacokinetics and Toxicity of Functional Groups
2: Emerging Drugs and Targets for Alzheimer's Disease; Volume 1
3: Emerging Drugs and Targets for Alzheimer's Disease; Volume 2
4: Accounts in Drug Discovery
5: New Frontiers in Chemical Biology
6: Animal Models for Neurodegenerative Disease
7: Neurodegeneration
8: G Protein-Coupled Receptors
9: Pharmaceutical Process Development
10: Extracellular and Intracellular Signaling
11: New Synthetic Technologies in Medicinal Chemistry
12: New Horizons in Predictive Toxicology
13: Drug Design Strategies: Quantitative Approaches
14: Neglected Diseases and Drug Discovery
15: Biomedical Imaging
16: Pharmaceutical Salts and Cocrystals
17: Polyamine Drug Discovery
18: Proteinases as Drug Targets
19: Kinase Drug Discovery
20: Drug Design Strategies: Computational Techniques and Applications
21: Designing Multi-Target Drugs
22: Nanostructured Biomaterials for Overcoming Biological Barriers
23: Physico-Chemical and Computational Approaches to Drug Discovery
24: Biomarkers for Traumatic Brain Injury
25: Drug Discovery from Natural Products
26: Anti-Inflammatory Drug Discovery
27: New Therapeutic Strategies for Type 2 Diabetes: Small Molecules
28: Drug Discovery for Psychiatric Disorders
29: Organic Chemistry of Drug Degradation
30: Computational Approaches to Nuclear Receptors
31: Traditional Chinese Medicine
32: Successful Strategies for the Discovery of Antiviral Drugs
33: Comprehensive Biomarker Discovery and Validation for Clinical Application

How to obtain future titles on publication:
A standing order plan is available for this series. A standing order will bring delivery of each new volume immediately on publication.

For further information please contact:
Book Sales Department, Royal Society of Chemistry, Thomas Graham House, Science Park, Milton Road, Cambridge, CB4 0WF, UK
Telephone: +44 (0)1223 420066, Fax: +44 (0)1223 420247,
Email: booksales@rsc.org
Visit our website at www.rsc.org/books

PREFACE

Introduction to Biomarker Discovery and Validation

PÉTER HORVATOVICH*[a,b,c] AND RAINER BISCHOFF[a,b,c]

[a] Analytical Biochemistry, Department of Pharmacy, University of Groningen, A. Deusinglaan 1, 9713 AV Groningen, The Netherlands; [b] Netherlands Bioinformatics Centre, Geert Grooteplein 28, 6525 GA Nijmegen, The Netherlands; [c] Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, the Netherlands
*Email: p.l.horvatovich@rug.nl

In 2001, the biomarkers definitions working group of the National Institutes of Health in the United States defined a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention".[1] A large number of analytical techniques are available to discover and measure biomarkers in complex biological samples such as immunochemical assays, molecular imaging, molecular arrays and mass spectrometry. From these techniques, mass spectrometry excels to be the most versatile analytical instrument able to comprehensively determine the identity and absolute or relative quantity of biomolecules present in complex biological samples. Mass spectrometry, a fairly recent approach to discover and validate biomarkers, requires close collaboration between multiple disciplines such as biology, pathology, analytical chemistry and bioinformatics. This approach is currently gaining momentum in the clinical environment. Different types of biomarkers are needed to support early diagnosis of disease onset, to predict the outcome of therapy, to monitor the efficacy of treatment or to get an objective clinical measure of drug efficiency based on molecular signatures.

RSC Drug Discovery Series No. 33
Comprehensive Biomarker Discovery and Validation for Clinical Application
Edited by Péter Horvatovich and Rainer Bischoff

Published by the Royal Society of Chemistry, www.rsc.org

To perform successful biomarker research ending with validated assays that are suitable for clinical application requires collaboration between academic centers and industry with considerable investments of time and money. More importantly it requires meticulous research based on an experimental design, well-characterized biospecimen collections, targeted and comprehensive profiling analytical platforms, and the processing and statistical analysis of enormous sets of data as well as determined leadership to arrive at go/no go decisions throughout the process from discovery to final implementation. Last but not least, biomarker-based assays must provide sufficient benefit for patients and the population at large to be eligible for reimbursement by health-care providers.

This book presents some of the state-of-the-art methodology used by leading academic and industrial research groups worldwide to advance biomarker research from planning sample collection and storage in biobanks, to the initial comprehensive discovery study all the way to the requirements for approval of biomarker assays by regulatory authorities. It provides a comprehensive view of mass-spectrometry-based biomarker discovery in 12 chapters covering all parts of this process.

Introductory Chapters

Chapter 1 covers strategic and practical aspects related to biomarkers for translational and personalized medicine related to pharmaceutical drug development.

Chapter 2 shows the complex and laborious process to take a biomarker product from discovery to commercialization. This chapter will familiarize the reader with regulatory issues and give a general understanding of quality system requirements. The chapter gives an overview over the current regulatory environment at the US Food and Drug Administration (FDA) and the European Medicines Agency (EMEA) for approving a biomarker test for clinical use.

Chapter 3 describes approaches used to design experiments for clinical biomarker discovery, verification and validation, discussing protocols and good practices for sample collection, storage, and quality control. It presents the cardinal role of biobanks and human biospecimen collections in biomarker discovery and validation, discusses the crucial points of patient and control selection to design longitudinal and cross-sectional studies. Finally, this chapter gives guidelines on how to select patient cohorts and to stratify patient populations.

Sample Preparation and Profiling

Chapter 4 deals with approaches of biomarker discovery in easily accessible body fluids such as serum, urine, epithelial lining fluid and cerebrospinal fluid, discussing sample-preparation methods and problems related to sample stability. Sample prefractionation methods that try to cope with the large

dynamic concentration range of proteins in body fluids are described in this chapter.

Chapter 5 describes various profiling platforms that are based on liquid chromatography coupled to mass spectrometry (LC-MS) for proteomic biomarker discovery, verification and validation. Platforms used for biomarker discovery provide qualitative (identity of proteins) and quantitative information about the analyzed proteomics samples. A discussion of targeted proteomics platforms as applied in the postdiscovery verification and validation phases is part of this chapter. Targeted protein analysis by LC-MS/MS allows quantification of peptides directly in clinical samples and contributes to verification and validation of biomarker candidates from biological samples.

Chapter 6 deals with the use of affinity array-based technologies in biomarker discovery and validation. The chapter demonstrates the strength of this technology by the use of peptide arrays to profile kinase activity to characterize phosphorylation-mediated signal transduction. Several applications of peptide arrays for kinome analysis are discussed including specific examples in which arrays have contributed to the identification of disease-associated processes/biomarkers, provided therapeutic targets, and offered novel insight into complex biological processes.

Bioinformatics and Statistics

Chapter 7 describes the structure of data preprocessing pipelines used to extract quantitative information from label-free LC-MS data. The chapter provides an overview of the main steps based on the example of a "Threshold Avoiding Proteomics Pipeline". This chapter further presents data visualization techniques and quality-control methods to assess the accuracy of automated label-free LC-MS data processing.

Chapter 8 describes important aspects of statistical analysis of the extremely complex biomarker discovery data leading to a selected set of highly discriminating peptides and proteins that can be followed-up in a subsequent validation phase. The main concepts of statistical validation to identify potential biomarker candidates are discussed. The presented statistical methods and concepts are described without the use of intricate mathematics, in a way that scientists involved in biomarker research with a general statistical or informatics background can follow. Techniques to avoid false positives are discussed and practical advice is given, including recognition of situations where no reliable biomarker candidate can be identified.

Chapter 9 describes a novel bioinformatics approaches to place biomarker candidates in biological context by exploring the molecular interaction networks and pathways that help in understanding the biological relevance of the discovered proteins. This chapter provides a detailed overview of bioinformatics and statistical approaches based on a case study related to plasma protein markers for breast cancer.

Discovery and Validation Case Studies, Recommendations

Chapter 10 presents a comprehensive study to discover, verify and validate urinary biomarkers for the early detection of bladder cancer. The chapter demonstrates the integration of a discovery study using stable isotope labeling on low numbers of pooled samples followed by the validation of a limited number of candidates with Multiple Reaction Monitoring (MRM) and ELISA approaches using a large set of individual samples. This chapter gives an example of the development of a biomarker panel rather than an individual biomarker.

Chapter 11 demonstrates the power of current mass-spectrometry-based biomarker discovery and validation approaches on the example of developing a biomarker panel to improve the prediction of near-term myocardial infarction. This case shows the power of a pooling strategy in combination with stable isotope labeling and 2-dimensional liquid chromatography coupled to mass spectrometry for biomarker discovery followed by validation with multiplexed immunoassays and MRM.

The last chapter, *Chapter 12*, provides a comprehensive overview of the bottlenecks and challenges of current protein biomarker research and provides guidance and recommendations to avoid common pitfalls. The challenges faced during protein biomarker development are multifactorial from the initial phase of sample collection and study design to final clinical validation comprising financial aspects and regulatory approval. This chapter helps researchers to recognize these potential challenges and help them to plan their study in a way that produces reliable results that can be critically interpreted.

Biomarker research underwent enormous technological and methodological developments during the last decade, a trend that continues today. Biomarker research covers many technological platforms and aspects that are not possible to include in one book. For example, this book does not include discussions on the bioinformatics of mass-spectrometry-based protein identification, analytical methods revealing protein isoforms or the tissue and cellular localization of proteins with optical- or mass-spectrometry-based imaging techniques. This book will aid practitioners ranging from specialists working at biobanks or at regulatory authorities, statisticians, bioinformaticians or researchers working in academic or industrial environments on biomarker discovery and validation projects.

Reference

1. Biomarkers Definitions Working Group, *Clinical Pharmacology and Therapeutics*, 2001, **69**, 89–95.

Contents

Introduction

Chapter 1 Introduction: Biomarkers in Translational and Personalized Medicine **3**
Chanchal Kumar and Alain J. van Gool

Summary 3
1.1 Introduction 3
1.2 Biomarkers 6
1.3 Biomarkers in Pharmaceutical Drug Development 7
1.3.1 The Pharmaceutical Research and Development Process 7
1.3.2 Biomarker-Based Decisions during Early-Phase Pharmaceutical Drug Development 10
1.3.3 The Use of Biomarkers in Late-Phase Pharmaceutical Drug Development 13
1.3.4 Fit-for-Purpose Biomarker Applications 14
1.4 Biomarkers in Translational Medicine 14
1.4.1 Translational Medicine 14
1.4.2 Translational Medicine in Drug Discovery and Development 15
1.4.3 Animal Models and Biomarkers in Translational Medicine 16
1.5 Biomarkers in Personalized Medicine 17
1.5.1 Personalized Medicine 17
1.5.2 Impact of Personalized Medicine 19
1.5.3 Molecular Profiling Toolbox and Personalized Medicine 21
1.5.4 Biomarkers in Personalized Medicine 24

RSC Drug Discovery Series No. 33
Comprehensive Biomarker Discovery and Validation for Clinical Application
Edited by Péter Horvatovich and Rainer Bischoff

Published by the Royal Society of Chemistry, www.rsc.org

1.6 Trends 25
1.6.1 Importance of Biomarkers 25
1.6.2 Globalization 26
1.6.3 High-Content Biomarker Discovery 27
1.6.4 Biomarker Combinations 27
1.7 Challenges 28
1.7.1 Interpretation of High-Content Biomarker Discovery Technologies 28
1.7.2 Biomarker Validation 28
1.7.3 Robust Biomarker Tests 29
1.7.4 Biospecimen 29
1.7.5 Disconnected Biomarker Development Pipeline 30
1.7.6 Translate Interindividual Findings in Personalized Biomarker Tests 30
1.7.7 Cost Reimbursement 30
1.7.8 Handling of Personalized Data 31
1.8 Opportunities 31
References 33

Chapter 2 Introduction: Regulatory Development Hurdles for Biomarker Commercialization: The Steps Required to Get a Product to Market 40
Melissa A. Thompson

2.1 Introduction 40
2.2 Regulatory Commercialization Path Options 41
2.3 Product Classification 43
2.3.1 Research Use Only (RUO) 44
2.3.2 Investigational Use Products (IUO) 44
2.3.3 Laboratory Developed Tests (LDT) 44
2.3.4 *In Vitro* Diagnostics (IVD) 47
2.3.5 IVD Kit 48
2.3.6 510(k) Clearance 48
2.3.7 PMA 50
2.3.8 CE Marking 51
2.3.9 Companion Diagnostics 54
2.4 Supporting Clinical Utility 55
2.5 Infrastructure and Other Considerations for Commercialization 56
2.5.1 Instrument Platform/Reagent Manufacturer Selection Challenges 56
2.5.2 Software 57

2.6 Quality Systems 59
2.6.1 What is Involved? 60
2.6.2 Elements of a Quality-Management System 61
2.6.3 Quality System General Requirements 61
2.7 Design Control 63
2.7.1 Product Development Life-Cycle Process 64
2.7.2 Who is Involved? 66
2.7.3 Reimbursement 68
2.8 Summary 68
Resources 69
Applicable Domestic and International Regulations 69
List of Important Internet Resources 70
Documents for EMEA Submission 70

Chapter 3 Introduction: The Cardinal Role of Biobanks and Human Biospecimen Collections in Biomarker Validation: Issues Impeding Impact of Biomarker Research Outcomes 73
Pascal Puchois, Lisa B Miranda and Alain van Gool

3.1 Introduction 73
3.2 Navigating the Biobanking–Biomarker Collaborative Landscape 75
3.2.1 Crucial Considerations in Planning Biobank–Biomarker Research Collaborations 75
3.2.2 Biobanking Challenges that Impede Impact of Biomarker Research Outcomes 79
3.2.3 Effect of Process-Chain Impediments on Impact of Biospecimen Collection Quality 85
3.3 Recommendations: Reducing Disparity of Impact and Lag in Outcomes of Biomarker Research 88
3.3.1 Technical and Scientific Recommendations for Biomarker Scientists 88
3.3.2 Technical and Scientific Recommendations for Biobankers 93
3.3.3 Joint Recommendations for Biomarker Scientists and Biobankers 98
3.4 Utilization of Biobank Samples for Biomarker Discovery and Development 100
3.4.1 Biomarker Discovery 100
3.4.2 Biomarker Development 102
3.4.3 Biobanking–Biomarker Collaborations 103
References 105

Sample Preparation and Profiling

Chapter 4 Sample Preparation and Profiling: Biomarker Discovery in Body Fluids by Proteomics **113**
N. Govorukhina and R. Bischoff

4.1 Introduction 113
4.2 Samples 114
4.2.1 Serum 114
4.2.2 Urine 122
4.2.3 Epithelial Lining Fluid (ELF) 122
4.2.4 Cerebrospinal Fluid (CSF) 125
4.3 Conclusion 128
Acknowledgements 129
References 129

Chapter 5 Sample Preparation and Profiling: Mass-Spectrometry-Based Profiling Strategies **136**
Yeoun Jin Kim and Bruno Domon

5.1 Introduction 136
5.2 LC-MS-Based Proteomics Applied to Biomarker Discovery 138
5.2.1 Sample Preparation: Prefractionation and Enrichment 138
5.2.2 Peptide Separation by Liquid Chromatography 143
5.3 MS-Based Discovery Platforms: Unsupervised Profiling 145
5.3.1 Data-Dependent Acquisition (DDA) 145
5.3.2 Data-Independent Acquisition (DIA) 146
5.4 MS-Based Discovery Platforms: Supervised Profiling 146
5.4.1 Surrogate Identification by Accurate Mass and Elution Time 147
5.4.2 Selective Identification Using Inclusion List 148
5.4.3 Hypothesis-Driven Discovery Proteomics 151
5.5 Example: Directed Proteomics and Enrichment Strategies Applied to a Colon Cancer Stem-Cell Biomarker Study 155
5.6 Perspectives 156
Acknowledgement 157
References 157

Chapter 6 Sample Preparation and Profiling: Probing the Kinome for Biomarkers and Therapeutic Targets: Peptide Arrays for Global Phosphorylation-Mediated Signal Transduction 162
Jason Kindrachuk and Scott Napper

6.1 Introduction 162
6.2 Understanding Complex Biology through Phosphorylation-Mediated Signal Transduction 163
6.3 Kinome *vs.* Phosphoproteome 164
6.4 Peptide Arrays for Kinome Analysis 168
6.4.1 Generation and Application of Peptide Arrays 169
6.4.2 Peptide Arrays: Phosphoproteome or Kinome Analysis? 169
6.5 Recent Advances 170
6.5.1 Customized Arrays 170
6.5.2 Technical Advances 176
6.6 Understanding Biology 176
6.6.1 Inflammation 177
6.6.2 Infectious Diseases 177
6.6.3 Stress 180
6.6.4 Impact of Glucocorticoids on Insulin Signaling 181
6.6.5 Lupus 181
6.6.6 Angiotensin II-dependent Hypertensive Renal Damage 182
6.6.7 Cancer 183
6.7 Validation of Results 186
6.8 Remaining Challenges 187
6.8.1 Data Statistics and Mining 187
6.9 Conclusions 188
References 189

Bioinformatics and Statistics

Chapter 7 Bioinformatics and Statistics: LC-MS(/MS) Data Preprocessing for Biomarker Discovery 199
Péter Horvatovich, Frank Suits, Berend Hoekman and Rainer Bischoff

7.1 Introduction 199
7.2 Quantitative Preprocessing Workflow for Single-Stage LC-MS(/MS) Data 203
7.3 Example Workflow: The Threshold Avoiding Proteomics Pipeline 209

7.4 Performance Assessment of Quantitative LC-MS Data Preprocessing Workflows 214
7.5 Summary and Future Trends 221
References 222

Chapter 8 Bioinformatics and Statistics: Statistical Analysis and Validation 226
Huub C. J. Hoefsloot

8.1 Introduction 226
8.2 Terminology 228
8.3 Validation Strategies 229
8.3.1 Crossvalidation 229
8.3.2 Model Assessment Using Random Data 230
8.3.3 Multiple Testing Corrections 230
8.3.4 Data Augmentation 231
8.4 Types of Biomarkers and Biomarker Panels 231
8.5 Discovery and Statistical Validation of The Biomarkers 232
8.5.1 How to Find a Type I Biomarker 232
8.5.2 Validation for Type I Biomarker 232
8.5.3 How to Find a Type II Biomarker 233
8.5.4 How to Find a Biomarker Panel 233
8.6 Selection of the Right Classification Method 236
8.7 How to Put the Pieces Together? 237
8.8 Assessing the Quality of the Biomarker Panel 238
8.9 What to Do if no Biomarker Panel is Found? 240
8.10 Conclusions and Recommendations 240
References 241

Chapter 9 Bioinformatics and Statistics: Computational Discovery, Verification, and Validation of Functional Biomarkers 243
Fan Zhang and Renee Drabier

9.1 What is a Biomarker? 243
9.2 What is Biomarker Research? 244
9.3 Biomarker Workflow Overview 244
9.4 Types of Biomarkers in Clinical Application 246
9.5 Computational Methods for Biomarker Discovery, Verification, and Validation 247
9.5.1 Performance Measurements 247
9.5.2 Receiver Operating Curve 247
9.5.3 T-test 248
9.5.4 Fisher's Exact Test 248
9.5.5 Neural Network 249
9.5.6 Bayesian Network 252

9.5.7 Support Vector Machine 252
9.5.8 Crossvalidation 253
9.6 System Biology Approaches for Biomarker Discovery, Verification, and Validation 254
9.6.1 Literature Search 254
9.6.2 Crossvalidation of Multiple Studies 255
9.6.3 Pathway Analysis 255
9.6.4 Interassociation Analysis 257
9.7 Case Study: Breast Cancer Plasma Protein Biomarker Discovery and Verification by Coupling LC-MS/MS Proteomics and Systems Biology 260
9.7.1 Study Design 261
9.7.2 Biomarker's Statistical Discovery 261
9.7.3 Literature Search 262
9.7.4 Pathway Analysis and Gene Ontology Categorization of Significant Proteins 262
9.7.5 Crossvalidation of Multiple Studies 264
9.8 Conclusion 264
Acknowledgements 265
References 265

Discovery and Validation Case Studies, Recommendations

Chapter 10 Discovery and Validation Case Studies, Recommendations: A Pipeline that Integrates the Discovery and Verification Studies of Urinary Protein Biomarkers Reveals Candidate Markers for Bladder Cancer 271
Yi-Ting Chen, Carol E. Parker, Hsiao-Wei Chen, Chien-Lun Chen, Dominik Domanski, Derek S. Smith, Chih-Ching Wu, Ting Chung, Kung-Hao Liang, Min-Chi Chen, Yu-Sun Chang, Christoph H. Borchers and Jau-Song Yu

10.1 Introduction 272
10.1.1 Importance of Bladder Cancer 272
10.1.2 Other Methods of Diagnosis 272
10.1.3 Other Urinary Biomarker Studies for Bladder Cancer 273
10.1.4 Protein-Based Biomarker Discovery 274
10.1.5 Introduction to iTRAQ for Biomarker Discovery 275
10.1.6 Introduction to MRM for Biomarker Discovery and Verification 276
10.1.7 Biomarker Discovery Pipeline 277

10.2 Experimental 278
10.2.1 Materials 278
10.2.2 Sample Preparation for iTRAQ and MRM-MS 280
10.2.3 LC-ESI MS/MS Analysis of iTRAQ-Labeled Peptides by LTQ-Orbitrap Pulsed-Q Dissociation 281
10.2.4 AB 4000 Qtrap Mass Spectrometry for MRM-MS Samples 282
10.2.5 Verification of MRM-MS Results 284
10.2.6 Statistical Analysis 285
10.2.7 Functional Annotation and Network Analysis of Differential Proteins 286
10.3 Results and Discussion 286
10.3.1 iTRAQ for Candidate Selection 286
10.3.2 MetaCore™ Analysis of Biological Networks Associated with Differentially Expressed Proteins in Urine 287
10.3.3 Verification of iTRAQ-Discovered Biomarkers in a Larger Number of Individual Samples Using Western Blot Analyses 290
10.3.4 Verification of Additional Biomarkers in a Larger Number of Individual Urine Samples Using MRM-MS 292
10.3.5 Comparison of iTRAQ and MRM Results and the Biomarker Discovery Pipeline 307
10.4 Conclusions 307
Acknowledgements 308
References 308

Chapter 11 Discovery and Validation Case Studies, Recommendations: Discovery and Development of Multimarker Panels for Improved Prediction of Near-Term Myocardial Infarction 315
Peter Juhasz, Moira Lynch, Manuel Paniagua, Jennifer Campbell, Aram Adourian, Yu Guo, Xiaohong Li, Børge G. Nordestgaard and Neal F. Gordon

11.1 Introduction 315
11.2 Methods 317
11.2.1 Study Design 317
11.2.2 Discovery Proteomics – 8-plex iTRAQ 2D-LC MS/MS 318
11.2.3 Multiplex Immunoassays (Luminex 200) 319
11.2.4 Multiple-Reaction Monitoring (MRM) Analysis 319
11.2.5 Data Analysis 321