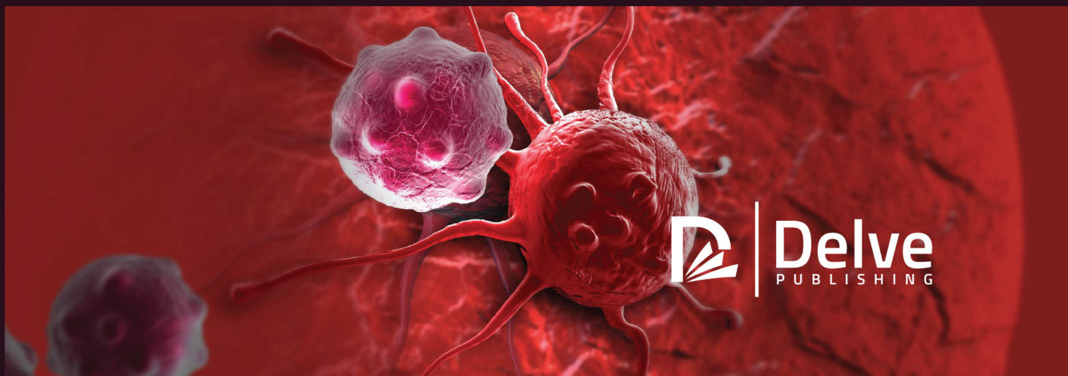
Proteomics in BIOMARKER IDENTIFICATION

Mohamed A. Selmy



PROTEOMICS IN BIOMARKER IDENTIFICATION

PROTEOMICS IN BIOMARKER IDENTIFICATION

Mohamed A. Selmy

www.delvepublishing.com

Proteomics in Biomarker Identification

Mohamed A. Selmy

Delve Publishing

2010 Winston Park Drive,

2nd Floor

Oakville, ON L6H 5R7

Canada

www.delvepublishing.com

Tel: 001-289-291-7705
001-905-616-2116
Fax: 001-289-291-7601

Email: orders@arclereducation.com

e-book Edition 2019

ISBN: 978-1-77361-799-2 (e-book)

This book contains information obtained from highly regarded resources. Reprinted material sources are indicated and copyright remains with the original owners. Copyright for images and other graphics remains with the original owners as indicated. A Wide variety of references are listed. Reasonable efforts have been made to publish reliable data. Authors or Editors or Publishers are not responsible for the accuracy of the information in the published chapters or consequences of their use. The publisher assumes no responsibility for any damage or grievance to the persons or property arising out of the use of any materials, instructions, methods or thoughts in the book. The authors or editors and the publisher have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission has not been obtained. If any copyright holder has not been acknowledged, please write to us so we may rectify.

Notice: Registered trademark of products or corporate names are used only for explanation and identification without intent of infringement.

© 2019 Delve Publishing

ISBN: 978-1-77361-082-5 (Hardcover)

Delve Publishing publishes wide variety of books and eBooks. For more information about Delve Publishing and its products, visit our website at www.delvepublishing.com.

ABOUT THE AUTHOR

Mohamed Selmy is a physician scientist and lecturer of Medical Biochemistry and Molecular Biology. He received his medical degree in 2004 from the faculty of medicine Suez Canal University in Egypt. He received his scientific training through many fellowships in United States, Spain and Italy. He received his PhD in Medical Biochemistry and Cancer biology in 2017 from the faculty of medicine Suez Canal University in Egypt. His fields of expertise comprises; cancer biology, cell signaling and molecular medicine. His work is continuously published in international peer reviewed journals.

TABLE OF CONTENTS

List of Figures ..... *xiii*
List of Tables ..... *xv*
List of Abbreviations ..... *xvii*
Preface ..... *xix*

Chapter 1 Biomarkers: An Overview 1
1.1. Introduction 2
1.2. Classification of Biomarkers 4
1.3. Types of Biomarkers Based on Different "Omics" 10
1.4. Biomarker Specifications 12
1.5. An Overview of Biomarkers In A Clinical Scenario 13
1.6. Characters of Ideal Biomarkers 15
1.7. Biomarkers or Clinical Endpoints 16
1.8. Merits And Demerits of Biomarkers As Surrogate Endpoints 17
1.9. Which Are The Biomarkers Used? 17
1.10. Gist of Significance of Biomarkers 19
1.11. Case Studies of Analysis of Biomarkers In A Clinical Context 20
1.12. Discovery and Validation of Biomarkers 21
1.13. Biomarkers In A Clinical Scenario: Approach 23
1.14. Valid Biomarkers and Clinical Research 24
1.15. Case Study of Biomarkers Detected By Metabolomics 26
1.16. Summary 29
References 31

Chapter 2 Proteomics: An Overview of Techniques Involved 35
2.1. Design of Biomarkers 36
2.2. Biobanking of Samples 36
2.3. Proteomics And Discovery of Biomarkers 37

2.4. A Quick Snapshot 38
2.5. Techniques For Biomarker Discovery Using Proteomics 39
2.6. The Mass Analyzers 42
2.7. Variations of Ms 43
2.8. Variations In Ms Technologies In Proteomics 44
2.9. Ief Fractionation Methods 46
2.10. 2D Gel Electrophoresis (2-De) 47
2.11. 2D Difference Gel Electrophoresis (2D-Dige) 50
2.12. Stable Isotope Labeling by Amino Acids In Cell Culture (SILAC) 52
2.13. Isotope-Coded Affinity Tag (ICAT) 54
2.14. Western Blotting 55
2.15. ^{18}O Stable Isotope Labeling 55
2.16. Isobaric Tagging Reagent For Absolute Quantitation (ITRAQ) 56
2.17. Ce-Ms 58
2.18. Stable Isotope Dilution Mass Spectrometry (SID-MS) 59
2.19. Microarrays 59
2.20. Multidimensional Protein Identification Technology (MUDPIT) 61
2.21. Label-Free Techniques 62
2.22. Another Proteomics For Biomarker Discovery 63
2.23. Bioinformatics and Proteomics And Biomarkers 65
2.24. Glimpses of Discovery of Biomarkers Using Proteomics 68
2.25. Case Studies: Biomarkers of Bladder Cancer Using Proteomics 68
2.26. Final Note 70
References 73

Chapter 3 Microbial Biomarkers Proteomics 79
3.1. Typing Of Microbes 80
3.2. Typing Methods 81
3.3. Phenotypic Methods 81
3.4. Genotypic Methods 85
3.5. Diagnosis Of Microbial Infections Using Biomarkers 88
3.6. What Is The Role Of A Biomarker From The Host? 88
3.7. Biomarkers And Antibiotics: A Scenario 90
3.8. Biomarkers That Are Routinely Analyzed 90
3.9. Biomarkers For The Future 91

3.10. What's New? ... 93
3.11. Proteome For Biomarkers Of Bacteria And Virus ... 94
3.12. Host–Pathogen Protein–Protein Interactions ... 96
3.13. Intact Proteins ... 98
3.14. Direct Interactions And Their Detections ... 100
3.15. Posttranslational Modifications That Are Regulated By Pathogen ... 100
3.16. Summary ... 102
References ... 103

Chapter 4 Plant Biomarkers Discovery by Proteomics ... 113
4.1. Introduction ... 114
4.2. Why Proteomics For Plant Biomarkers? ... 114
4.3. Approaches For Plant Biomarkers ... 116
4.4. Proteomics For Reactive Oxygen Species Research ... 119
4.5. Stress Responses In Two Plants: Proteomics For Biomarkers ... 119
4.6. Proteomics For Plants Infected With Virus: Biomarkers ... 122
4.7. The Advantages of Proteomics ... 122
4.8. Protein Microarrays For Biomarkers ... 126
4.9. Other Approaches ... 126
4.10. Interpreting The Data Using Bioinformatics ... 126
4.11. Proteomics For Food From Agriculture: Biomarkers: ... 127
4.12. Biomarkers For Biofuels ... 129
4.13. Biomarkers For Important Crops: Applications of Proteomics ... 130
4.14. Proteomics of Plants For Biomarkers: An Organ Wise Approach ... 133
4.15. Biomarkers: Proteomic Analysis of Seedlings of Crops ... 133
4.16. Proteomics And Biomarkers ... 138
4.17. Root Proteomics ... 138
4.18. Leaf Proteomics ... 140
4.19. Xylem And Phloem Proteomics ... 141
4.20. Summary ... 142
References ... 143

Chapter 5 Biomarkers For Animals by Proteomics ... 151
5.1. Introduction ... 152
5.2. Health of Farm Animals: Biomarkers ... 155
5.3. Proteomics And Veterinary Research ... 156

5.4. Proteomics Techniques For Animal Biomarker Discovery 157
5.5. Sample Preparation 161
5.6. Points to be Considered in the Terms of Animal Biomarkers 161
5.7. Veterinary Research and Biomarkers 162
5.8. Pathogen Proteomics.. 163
5.9. Avian Proteomics 163
5.10. Proteomics In Aquaculture 164
5.11. Parasite Proteomics 165
5.12. Animal Proteomics in Different Tissues 166
5.13. Overview of Serum Proteomics in Farm Animals........................... 168
5.14. Biomarkers Using Proteomics of Animal Products......................... 173
5.15. Case Study of Biomarkers: Targeting The Mosquito........................ 175
5.16. Scenario/Future For Animal Proteomics 177
5.17. Summary 179
References 180

Chapter 6 Clinical Biomarkers And Proteomics 185
6.1. Introduction... 186
6.2. Overview of Biomarkers 186
6.3. Historical Background 187
6.4. Role of Biomarkers In Drug Development....................................... 188
6.5. Brief Regulatory Aspects 189
6.6. Utility of Biomarkers In Clinical Studies.. 190
6.7. Biomarkers In Clinical Evaluation 190
6.8. Urine Biomarkers... 198
6.9. Serum Biomarkers.. 203
6.10. Classes of Biomarkers 213
6.11. Efficacy Biomarkers 217
6.12. Surrogate Biomarkers 217
6.13. Predictive Biomarkers 218
6.14. Pharmacodynamic (PD) Biomarkers 218
6.15. Prognostic Biomarkers 219
6.16. Discrepant Results And Its Major Impact on Clinical Trials 221
6.17. Lack of Traceability as a Major Source of Systematic Errors 225
6.18. Impact of Discrepant Results on Personalized Medicine 227

6.19. Mass Spectrometry And Marker Detection 230
6.20. Conclusion 234
References 235

Chapter 7 Biomarkers of Cancer Using Proteomics 243
7.1. Discovery of Biomarkers of Cancer 244
7.2. Types of Biomarkers 245
7.3. Diagnosis of Cancer By Using Proteomics 247
7.4. Proteomics Approaches Towards
The Early Detection And Diagnosis of Cancer 247
7.5. Proteomic Techniques In Cancer 252
7.6. Clinical Utility 262
7.7. Clinical And Analytical Necessities For Biomarker Presentation 262
7.8. Proteomic Approach In Different Types of Cancer 263
7.9. Challenges 271
7.10. Perspectives For Proteomics In Cancer 272
7.11. Summary 272
References 275

Chapter 8 Future Directions, Challenges and Solutions 285
8.1. Introduction 286
8.2. Challenges In Biomarkers For Proteomics 287
8.3. Proteomics vs. Genomics 293
8.4. Overcoming Challenges in Data Interpretation 293
8.5. Future Expectations 294
8.6. Points For Consideration 295
References 297

Index 299

LIST OF FIGURES

Figure 1.1. Biomarkers and disease

Figure 1.2. The 3 p type of biomarkers

Figure 1.3. Type 0 to 2 biomarkers

Figure 1.4. Biomarkers in nutrition

Figure 1.5. Biomarkers based on omics

Figure 1.6. Receiver operating characteristic (ROC) curve

Figure 1.7. Biomarkers types

Figure 1.8. Biomarker types

Figure 1.9. An ideal biomarker

Figure 1.10. PSA as a biomarker

Figure 1.11. Discovery

Figure 1.12. Use of biomarkers

Figure 1.13. Biomarkers in carcinogenesis

Figure 2.1. Design of biomarkers

Figure 2.2. Biobank

Figure 2.3. Proteomics in biomarker discovery

Figure 2.4. MS: Principle and scheme

Figure 2.5. Scheme of ESI-MS: Sample and solvent passed through charged needle

Figure 2.6. A scheme of MALDI-MS: the crystallized sample mixed with the matrix is excited by laser

Figure 2.7. Scheme of Tandem MS.

Figure 2.8. Scheme of 2DE: separation in 2 dimensions: pI and molecular weight

Figure 2.9. Scheme of 2D-DIGE

Figure 2.10. Scheme of SILAC

Figure 2.11. Scheme of iTRAQ

Figure 2.12. Microarrays for proteomics biomarkers

Figure 2.13. Scheme of MudPIT

Figure 2.14. Biomarkers from tissue culture and proteomics

Figure 3.1. Schematics of analysis of microbial biomarkers

Figure 3.2. Tools of proteomics for PPI

Figure 3.3. Overview of MS for proteomics for microbial biomarkers

Figure 4.1. Different biotic and abiotic stresses in plant

Figure 4.2. Proteomics in plant biomarkers: a scheme

Figure 4.3. Flow of proteomics in plant biomarkers

Figure 4.4. Proteomics for biomarkers and conventional breeding together can contribute to a sustainable economy by 2050.

Figure 4.5. Plant organ proteomics for biomarkers

Figure 4.6. Overview of select factors involved in stress in plants

Figure 4.7. Root proteomics

Figure 5.1. A schematic representation of the proteomics workflow

Figure 5.2. Approaches for biomarker discovery in animals

Figure 5.3. Overview of proteomics in animal biomarker

Figure 5.4. Various applications of biomarkers using proteomics in animals

Figure 5.5. An overview of biomarkers of animal products

Figure 5.6. Scheme of the study

Figure 6.1. Summary of different Salivary Markers

Figure 6.2. Scheme of MS in a clinical scenario

Figure 6.3. The chart for clinical biomarkers and the various techniques to aid in the process

Figure 6.4. A summary of biomarkers of clinical proteomics

Figure 8.1. Proteomics vs. Genomics

Figure 8.2. Increase in the tools applied in MS proteomics

LIST OF TABLES

Table 1.1. Biomarkers in HCC

Table 2.1. Proteomics and electrophoresis

Table 2.2. Databases for proteomics

Table 2.3. Annotated databases

Table 3.1. The use of Strem-1 in acute infections

Table 3.2. Clinical use of suPAR in acute infections

Table 3.3. Clinical use of pro-ADM in acute infections

Table 3.4. Clinical use of presepsin in acute infections

Table 4.1. Subcellular proteomic studies in crops under abiotic stress

Table 4.2. Techniques of proteomics for crops

Table 4.3. Selected proteomic studies on xylem and phloem sap

Table 5.1. Gene Ontology (GO) annotations

Table 5.2. The analysis of serum for biomarkers has impact on human health

Table 6.1. Basic Safety Biomarker Panels in Clinical Trials

Table 7.1. Some common cancer biomarkers that have acquired validation

Table 7.2. Cancer biomarkers and their characteristics

Table 7.3. Biomarkers for the diagnosis of ovarian cancer discovered through proteomic technology

Table 8.1. Proteomics vs. Genomics

LIST OF ABBREVIATIONS

2-DE	2-dimensional gel electrophoresis
2D-DIGE	2-dimensional difference gel electrophoresis
2DE-MS	Two-dimensional gel electrophoresis followed by mass spectrometry
APEX	Absolute protein expression
CE	capillary electrophoresis
CE- MS	capillary electrophoresis coupled to mass spectrometry
CEA	Carcinoembryonic antigen
CID	collision-induced dissociation
CM	Conditioned medium
CRC	Colorectal carcinoma
DAPA	DNA array to protein array
ELISA	Enzyme-Linked Immunosorbent Assay
emPAI	Exponentially modified PAI
ERLIC	Electrostatic repulsion-hydrophilic interaction chromatography
ESI	Electrospray Ionization
FASP	Filter aided sample preparation
FT	Fourier-transform
FTICR	Fourier transform ion cyclotron resonance
HPLC	High performance liquid chromatography
ICAT	Isotope-coded affinity tag
IEF	Isoeletric Focusing
Ig	Immunoglobulin
IMS	Ion Mobility Spectrometry
IPG	Immobilized pH gradient
iTRAQ	Isobaric tagging reagent for absolute quantitation
LC	Liquid chromatography
LC-MS	Liquid chromatography coupled to mass spectrometry
LTQ	Linear trap quadrupole

m/z	Mass to charge
MALDI	Matrix-assisted laser desorption ionization
MC CE	Microfluidic chip CE
MCE	Multicompartment Electrolyser
MRM	Multiple-reaction monitoring
MS/MS	Mass spectrometry/mass spectrometry, i.e., tandem mass spectrometry
MudPIT	Multidimensional Protein Identification Technology
NAPPA	Nucleic acid programmable protein array
PAI	Protein abundance index
pI	Isoelectric point
PMF	Protein mass fingerprinting
PPIs	Protein-protein interactions
PSA	Prostate specific antigen
Q	Quadrupole
QMS	Quadrupole mass spectrometer
ROC	Receiver operator characteristic
ROS	Reactive oxygen species
RP	Reversed phase
SCX	Strong cation exchange
SDS-PAGE	Sodium dodecyl sulphate-poly acrylamide gel electrophoresis
SELDI	Surface enhanced laser desorption/ionization
SELDI-TOF	Surface-enhanced laser desorption/ionization time of flight
SID-MS	Stable isotope dilution-mass spectrometry
SILAC	Stable isotope labeling by amino acids in cell culture
SRM	Selected Reaction Monitoring.
TOF	Time-of-flight

PREFACE

The term biomarker refers to a "characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention." (National Institutes of Health Biomarkers Definitions Working Group).

These biomarkers can serve as indicators of normal functioning or an abnormality that develops in a biological system. With biomarkers covering an entire set of organisms right from bacteria to plants, animals and humans, these indicators can reflect the physiological status of an organism. To illustrate an example, cancer is a feared disease and its diagnosis say for that of oral cancer is usually late. The identification of suitable biomarkers can facilitate the identification of a disease at an early stage so that appropriate medical intervention can be resorted to.

The term proteome is derived from PROTEins expressed by a genOME: hence it is a characterization of proteins expressed by an organism. There are several techniques to analyze the protein set of an organism such as chromatography, 2-dimensional electrophoresis to also chemical labels such as iTRAQ (Isobaric tags for relative and absolute quantitation) or ICAT (Isotope-coded affinity tag). There are also other techniques such as Stable isotope labeling by amino acids in cell culture (SILAC). Following separation most techniques use Mass Spectrometry to identify the proteins separated by the above-mentioned methods.

This book covers an Introduction to biomarkers and their importance followed by proteomics techniques to identify and discover such markers. The following chapters elucidate the discovery of biomarkers by proteomics in plants, animals, microbes as well as in clinical settings and in cancer. Finally, the scope and future of biomarkers is presented.

CHAPTER 1

BIOMARKERS: AN OVERVIEW

CONTENTS

1.1. Introduction 2
1.2. Classification of Biomarkers 4
1.3. Types of Biomarkers Based on Different "Omics" 10
1.4. Biomarker Specifications 12
1.5. An Overview of Biomarkers In A Clinical Scenario 13
1.6. Characters of Ideal Biomarkers 15
1.7. Biomarkers or Clinical Endpoints 16
1.8. Merits And Demerits of Biomarkers As Surrogate Endpoints 17
1.9. Which Are The Biomarkers Used? 17
1.10. Gist of Significance of Biomarkers 19
1.11. Case Studies of Analysis of Biomarkers In A Clinical Context 20
1.12. Discovery and Validation of Biomarkers 21
1.13. Biomarkers In A Clinical Scenario: Approach 23
1.14. Valid Biomarkers and Clinical Research 24
1.15. Case Study of Biomarkers Detected By Metabolomics 26
1.16. Summary 29
References 31

1.1. INTRODUCTION

Biomarker is a portmanteau of "biological marker" that is attributed as certain markers or indicators of a particular condition. In a medical scenario, these are not associated with symptoms as the latter are recognized by the patients or are present in a certain disease. Key examples are simple readings of blood pressure or pulse as well as molecules that are representative of a particular condition [1].

The first report of the use of biomarkers was in 1980: Isaakson in 1980 with the level of Nitrogen in urine as an indicator of the protein consumed in the diet [2].

There are several definitions of biomarkers that include:

*Alterations that are at the level of cells or molecules or biochemistry that can be quantified in a biological context and system such as cells or fluid*s [3].

*A property that is quantified and measured in order to indicate either optimal pathways in biology or pathogenesis or the response to a treatmen*t [4].

*A structure, substance, or pathway that is quantified in an individual to estimate the occurrence as well as result of a diseas*e [5].

A more comprehensive definition of biomarkers is:

*A quantification that is indicative of an association of a system of biology and a threat. This quantification can be at the level of molecules or cells or biochemistry or physiolog*y [6].

The above definition covers disease as well as the treatment, nutrition and chemicals in the environment.

Another definition is a molecule that is reflective of a normal system or otherwise or disease that is found in a tissue or fluid of the body The National Cancer Institute [7].

A 2013 definition goes like this:

*An article that is indicative of status of nutrition that covers both intake and its metabolis*m [8].

Thus, it can be seen that the definition of this term there are several aspects associated with it.

The concept can also be extended that encompasses other systems such as biological systems and pathways or pathogenicity or the response to a particular therapy [9].

The field of biomarkers has been defined as inclusive of technology as well as tools that offer insights into predicting, diagnosing, regressing, causes as well as the end result of a pathological condition. For instance, the use of measuring the nervous tissues can use analyzing samples such as cerebrospinal fluid or methods that do not have direct application rather imaging techniques. The use of biomarkers to analyze several samples of humans in a clinical scenario such as muscle, skin, blood, urine, brain, nerves and cerebrospinal fluid has been seen in several aspects in healthy as well as diseased samples.

They are aimed at the following while lacking recall bias:

- An analysis of disease;
- Causes of a disease;
- Measuring disease;
- Patterns of absorption;
- Metabolic pathways; and
- Pathways of several chemicals in a system.

Research has seen the role of biomarkers in not only diagnosing but also in the treatment of several diseases [10].

Biomarkers can distinguish between an individual with a disease and that of a healthy control. Few differences between individual with a disease and that of a healthy control can be attributed to a different profile of transcription, translation, posttranslational modifications as well as including germline or somatic mutations.

The biomarkers may include nucleic acids such as micro RNA or proteins or antibodies peptides. Certain profiles in terms of genomics, proteomics or metabolomics can include biomarkers. Their detection can be done in serum/ plasma/blood or excreted materials such as sputum/ urine/ stool or milk [7].

The potential of biomarkers in human diseases can be visualized by the Figure 1.1.

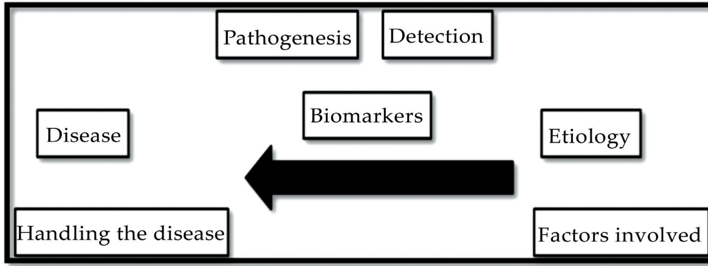


Figure 1.1. Biomarkers and disease.

Biomarkers can be subjected to epidemiological studies along with the role of history as well as disease prognosis.

Overall, they can be applied in several spheres such as:

- Showing the way to mechanisms associated in a disease
- Events of history of a disease
- Prediction of risk
- The processes between exposure and development of a disease
- They can decrease the effect of wrong classification of a disease and its potential risk/ exposure.
- The progression of disease
- Prognosis of a disease
- Response to a treatment/therapy
- They can be used to diagnose a particular manifestation
- They can identify diseases such as neurological disease at earlier stages to facilitate the timely treatment and prognosis [10].

1.2. CLASSIFICATION OF BIOMARKERS

The major types of biomarkers are **biomarkers of exposure** that find application in risk prediction, and **biomarkers of disease**, that can be used to screen, diagnose and examine the progression of a disease.

For diseases, biomarkers can decrease the extent of bias in epidemiological studies. They offer a direct quantification of an exposure/ effect instead of potential risk factors. This lessens the extent of misclassification of an exposure and its outcome that can avoid the reporting of erroneous results. The use of such biomarkers makes studies more sensitive and specific.

Molecular biomarkers can highlight disease susceptibility in patients. The use of a biomarker can aid in classification of a group based on the genotype/molecule associated with a disease instead of traditional methods such as pedigree [10, 12]. Such studies of susceptibility to disease are vital for the evaluation of disease risk across different populations.

When **Biomarkers of exposure** are looked at; it is seen that a disease or stress condition in plants or animals can result due to an exposure. A particular substance or chemical in an environment can serve as external biomarkers. The quantity of an element can be calculated using the levels of such chemicals in environment such as air or water.

A biomarker that is detected within body tissues or fluids can serve as an internal marker. The level of a substance in a tissue is reflective of the dose of exposure, for instance exposure to lead. The levels of lead can be measured in environment however, an accurate estimate can be obtained by measuring the levels of the metal in body samples such as hair or teeth.

Thus, though external biomarkers can highlight the exposure and cause of disease, internal biomarkers present a picture that is within tissues/organs.

Such internal biomarkers require the pharmacokinetic properties as different substances reach different targets on account of varied interactions. While chemicals such as organophosphate are located in body fluids such as urine or blood; halogenated hydrocarbons reach the adipose tissues.

Such biomarkers of exposure can either augment or stall the progression of a disease. A biomarker gives a picture of the quantity of substance that causes a disease over studies such as the exposure history. Thus, a combination of both internal as well as external factors can aid in the analysis of a particular disease.

In the field of pharmacologic properties of a particular substance that causes disease, biomarkers can offer valuable help in cross sectional analysis of disease models. The discovery of biomarkers in the field of neurological diseases like Alzheimer's disease requires stability that can be achieved by the use of serum samples from banks.

Antecedent biomarkers do not depend on exposures instead are found in an individual before the onset of a disease. In the case of epidemiological studies, the use of risk and tabulations can show the role of genetics and environmental exposures to study the role of a disease in an individual. For instance, degenerative diseases that have their onset in adulthood are a combination of genetics and exposures. Such biomarkers need not

necessarily be involved in the progression of a disease. In a particular pathway of disease, such biomarkers can serve as tools as they are not dependent on exposures. Depending on the role in disease development, various interactions can be measured.

For instance, biomarkers of genetic susceptibility can be very useful in neurological disorders. The role of variant alleles like APOE (apolipoprotein E) can serve as indicators of risk in pathogenesis of Alzheimer's disease [13]. Such biomarkers can highlight the role of genes and environmental exposures in a particular disease.

While some biomarkers can directly indicate disease, several **intermediate biomarkers** can be indirectly linked to causes. For instance, it could be linked to a factor (characterized or unknown) that causes a disease. This makes such biomarkers linked to a disease but is not a major cause in a disease progression pathway. A biomarker can be associated with an exposure that manifests in a disease. The tricky situation is presented by biomarkers that are linked to yet to be characterized factors of a disease. The association of such a biomarker and disease then is challenged.

Biomarkers that can show signs of a disease called **biomarkers of disease**, can facilitate early diagnosis of a disease at an early stage. Several samples such as serum/cerebrospinal fluid/urine can serve as samples for such analysis. The markers indicate a particular stage of a disease or subclinical manifestation. Surrogate manifestations of a disease can be used for screening samples to permit diagnosis.

Biomarkers of disease have several applications such as:

- Clinical trial targets
- Screen for patients who are at risk to develop a disease or are before the clinical manifestations of a disease.
- The history of the disease ranging from induction to its latent stage and subsequent detection
- Clinical trials or epidemiologic studies plagued with disease heterogeneity can be lowered.
- improvement in validity and precision.

The use of Diagnostic tests using such biomarkers can aid in the prevention of disease: either primary (prior to onset of symptoms) or secondary (early detection). Additional significance can be summarized below:

- Increase the level of detection of stress or disease

- The analysis of response to a particular treatment.
- The prediction of outcome of disease.
- Gauge the extent of disease.
- The prediction of risk of a disease [10, 11]

There is a system of classification of biomarkers into the following types:

Biomarkers of health/disease:

They reflect the stage of a disease or a phenotype of a disease. Example: PSA (Prostate specific antigen for health of prostate).

Biomarkers of dietary exposure:

They reflect the food intake inclusive of nutrients, non-nutrients and the pattern of diets.

Example: Urine Nitrogen that indicates protein.

1.2.1. Biomarkers of Nutritional Status

They reflect the metabolism of dietary intake as well as disease.

Example: homocysteine indicates diet as well as metabolism [2]

If cancer biomarkers are looked at, the biomarkers can range from proteins, metabolites or pathways such as angiogenesis, apoptosis, etc. to DNA or mRNA. These biomarkers can be produced either by cancer cells or the response of the tissues to cancer cells. The biomarkers can be used for the detection, progress as well as responses of cancer cells as well as role of chemicals such as carcinogens [14].

Biomarkers have been:

- single nucleotide polymorphisms
- Metabolites
- circulating nucleic acids
- gene expression products
- Gene variants
- Polysaccharides [15]

A biomarker can indicate a disease as well as facilitate quick detection or diagnosis that is applicable to several populations.

1.2.2. The 3 Classes of Biomarkers

1.2.2.1. Disease Specific Biomarkers

Biomarkers that are not specific to a disease but are ***associated with a morbidity or disease***.

Random biomarkers that arise due to variations in protocols of sample processing, control and test samples as well as the processing on Mass spectrometry [16].

Another classification of biomarkers includes: pharmacodynamic, prognostic and predictive (Figure 1.2):

Pharmacodynamic biomarkers are indicative of an association between a target and drug that cover adverse or beneficial effects.

Prognostic biomarkers are markers that are reflective of the potential path of a disease. They can predict the outcome of a particular disease in the absence of therapeutic intervention.

Predictive biomarkers are reflective of the likely response of a particular treatment regime. Predictive biomarkers aid in the assessment of response to treatment regime in contrast to prognostic markers that show the effect of a disease without intervention of drugs. Drug related biomarkers are reflective of the response or effectiveness of a drug on a particular system [15]:

According to the Biomarkers and Surrogate End Point Working Group [17] the following system of clinical biomarkers are described (Figure 1.3):

- Type 0 includes biomarkers of natural history of biomarkers clinical indices
- Type I is associated with the effect of the action of a drug and its mode of action.
- Type II is composed of the benefit conferred: includes surrogate end points [18]

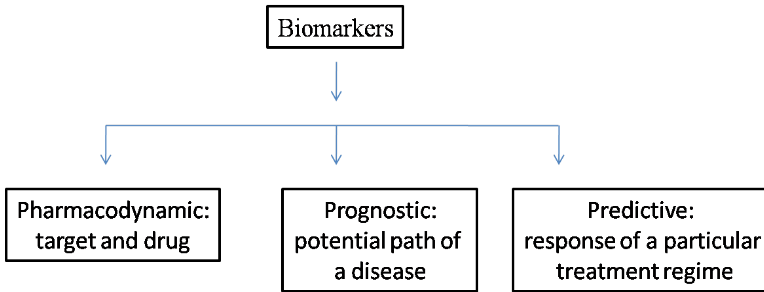


Figure 1.2. The 3-p type of biomarkers.

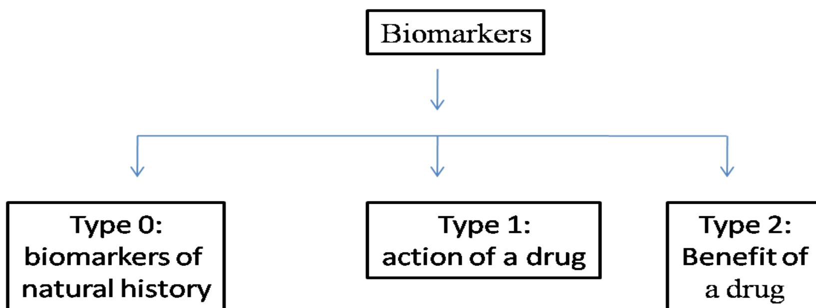


Figure 1.3. Type 0 to 2 biomarkers.

There is another system of classification that involves 4 groups of biomarkers in a *nutritional scenario*: recovery, concentration, replacement and predictive biomarkers (Figure 1.4).

1.2.2.2. Recovery Biomarkers

These examine metabolic balance between consumption and excretion to give a value of intake. These markers do not examine differences in metabolism between individuals and are absolute values. For example: total nitrogen/potassium in urine.

These biomarkers were studied in Observing Protein and Energy Nutrition (OPEN) Study by National Cancer Institute. Following food frequency questionnaire, the nitrogen content of urine was measured to assess effect of low fat diet on cancer and heart disease. The study revealed that obese and young women under reported the food frequency questionnaire.

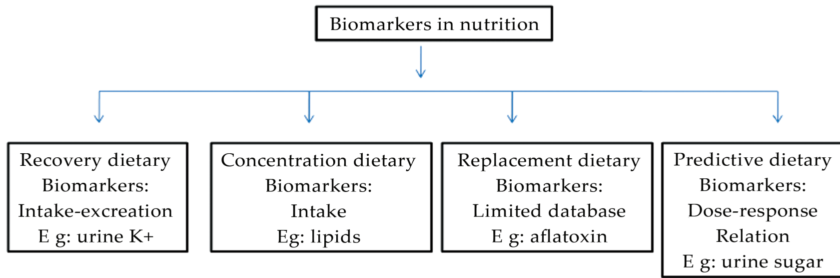


Figure 1.4. Biomarkers in nutrition.

1.2.2.3. Concentration Biomarkers

These examine the effect of personal habits such as diet/ smoking as well as metabolism and are not absolute. They are reflective of concentrations of a particular nutrient and effect on health. Examples include concentration of carotene or lipids

1.2.2.4. Replacement Biomarkers

These are similar to concentration biomarkers but lay emphasis on nutrients that are not covered in databases. Example: aflatoxins.

1.2.2.5. Predictive Biomarkers

These are sensitive and are reflective of the intake along with dose-response. Examples include: Urinary Sugars Biomarker: Fructose 24-hour analyzes [2, 19]

1.3. TYPES OF BIOMARKERS BASED ON DIFFERENT "OMICS"

The various omics technologies have cast light on several biomarkers: (Figure 1.5):

Genetic biomarkers:

They are polymorphisms connected with intake, metabolism or disease. Such markers are constant over time and can be detected from several samples such as fluids/ hair. The use of genotyping arrays can hasten the discovery of such markers.

Examples:

ALDH2 (aldehyde dehydrogenase) eliminates acetaldehyde that is generated from alcohol metabolism. Due to a Glu487Lys polymorphism, a reduction in metabolism of acetaldehyde is seen in the variants that causes flushed symptoms which in turn results in lesser alcohol consumed by 487Lys allele.

A 13910C>T polymorphism in lactase (*LCT*) gene results in lactase persistence in people of Europe. This manifests as discomfort following consuming milk that results in lesser use of milk products. This gene can serve as a biomarker for milk consumption.

Epigenetic biomarkers:

These are modifications to DNA to regulate expression. Methylation of the cytosine–phosphate–guanine(CpG) sequences has been explored as a marker of several diseases. Non-coding RNAs such as microRNA (miRNAs), long non-coding RNAs have also been explored as markers. miRNAs have been explored for their expression in cancers.

Transcriptomic biomarkers:

These involve analysis of transcription/ gene expression. Example: difference in expression following administration of a particular diet against controls.

Proteomics markers:

The analysis of protein sets constitutes this realm. Example: The proteins expressed in the presence or absence of a nutrient.

Lipidomic biomarkers:

They involve the lipid profile. Example: Lipid profile of a low fat/ high fat diet.

Metabolomic biomarkers:

They analyze the metabolome or metabolites. Example: NMR to study effect of a particular drug or nutrient [2].

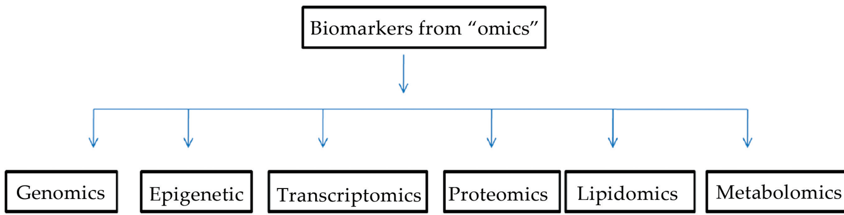


Figure 1.5. Biomarkers based on omics.

1.4. BIOMARKER SPECIFICATIONS

Specificity: This refers to the controls that lack the biomarkers.

Sensitivity: The individuals who test positive for a biomarker under study.

If a biomarker is to be used in diagnostics then it is essential that both sensitivity and specificity should be high.

The comprehensive ability of a marker can be defined by Diagnostic odd ratio (DOR) defined by the formula:

$$DOR = \frac{sensitivity}{1 - specificity} / \frac{1 - sensitivity}{specificity}.$$

A test can be summarized by defining the Likelihood ratio that is reflective of sensitivity and specificity. A positive test ratio (the odds of increase in disease when a test is positive) is shown by:

$$LR^{+} = \frac{sensitivity}{1 - specificity}$$

A negative test (the odds of decrease in disease when a test is positive) is shown by the formula

$$LR^{-} = \frac{1 - sensitivity}{specificity}.$$

Receiver operating characteristic (ROC): This is a curve that is indicative of sensitivity and specificity. The usefulness of a marker at various time points is measured as shown in the following figure**:** When the maximum area under a curve is seen: the test is good. In the Figure 1.6, the line close to diagonal is less efficient than the other.

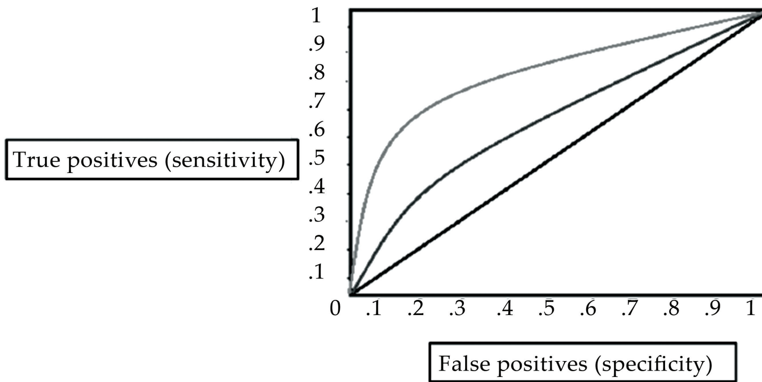


Figure 1.6. Receiver operating characteristic (ROC) curve.

1.5. AN OVERVIEW OF BIOMARKERS IN A CLINICAL SCENARIO

Stratification markers: the suitability of a drug for a particular patient.

Efficient markers: The appropriate dose for a patient that is efficient is selected.

Toxicity markers: they represent the side effect or adverse effects to which a patient is susceptible.

Surrogate endpoint markers represent end points such as survival or the overall feeling of a patient. (Figure 1.7).

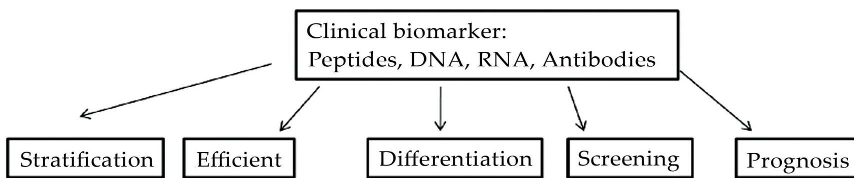


Figure 1.7. Biomarkers types [15].

A majority of ethical review boards dictate that people who test positive for a disease should attend follow up whether or not a disease is manifested or not. This also requires availability of treatment for such positive testing patients [10].

A few more terminologies:

A ***biomarker of exposure*** refers to an exogenous substance or a compound produced by the reaction of an exogenous chemical and a target molecule or cell. Such molecules can be located in body fluids or substances excreted. For example: nicotine or lead in saliva.

A ***biomarker of effect*** refers to an internal molecule that has been decreased as a result of interaction with an exposure compound. For instance, changes in lungs following tobacco smoke exposure.

The above two biomarkers have two layers:

- the first that refers to disease arising out of the effects of exposure of a harmful substance.
- the second is the beneficial effects such as treatment or recovery following exposure of a diseased condition to therapy.

A **biomarker of susceptibility** refers to individual factors that play a role post exposure such as genetic variations that induce susceptibility of individuals to a particular disease [20] (Figure 1.8).

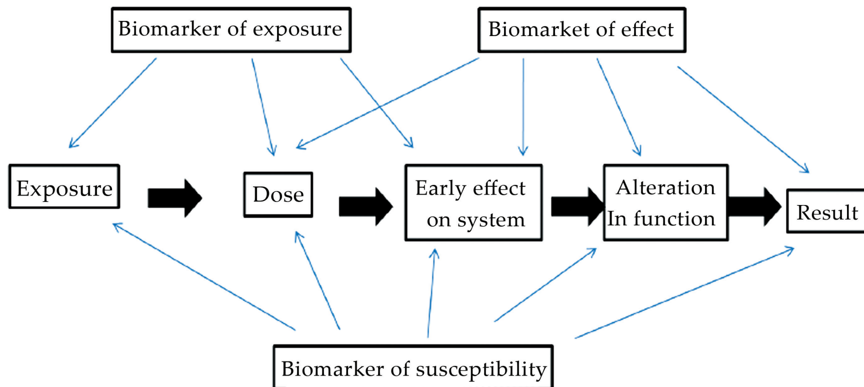


Figure 1.8. Biomarker types.

The three types of biomarkers and their relationships have been shown. The block arrows represent disease progression while the blue arrows show the effect of each type of biomarker on aspects of the disease following exposure.

Temporality can be used to classify biomarkers into the following classes:

Short-term: that are reflective of hours or days. Samples processed can include serum or urine.

Medium-term are reflective of weeks/months. Examples include adipose cells and red blood cells markers.

Long-term biomarkers are reflective of months/years. Sample analyzed include nails, teeth or hair [2, 21].

1.6. CHARACTERS OF IDEAL BIOMARKERS

- The biomarker should be classified as safe and should be elementary to measure.
- It should show consistency regardless of ethnicity or sex.
- Modification of the biomarker must be permitted.
- Follow up tests should be economical [22].

Using an example to study kidney toxicity, the following were highlighted as desirable characters of biomarkers:

- The visualization of the biomarker must be facilitated at an early stage before manifestations seen by histopathological examinations.
- It should also be an indicator following damage.
- It should have sensitivity.
- Correlation with the severity of damage must be shown by the marker.
- The biomarker must be available in outer tissues so that it can be measured in body fluids.
- The biomarker should possess stability to facilitate a lapse in time between sample collection and processing.
- It should possess translational properties across species.
- Localization is preferred; instead of an indicator of general damage it should highlight a particular specific region.
- The mechanism and biomarker requires association. The exact role of the biomarker in terms of its role in a pathway must be clear.
- It was suggested that as several characters are involved, a panel of biomarkers is favorable over single biomarkers [23]. (Figure 1.9).

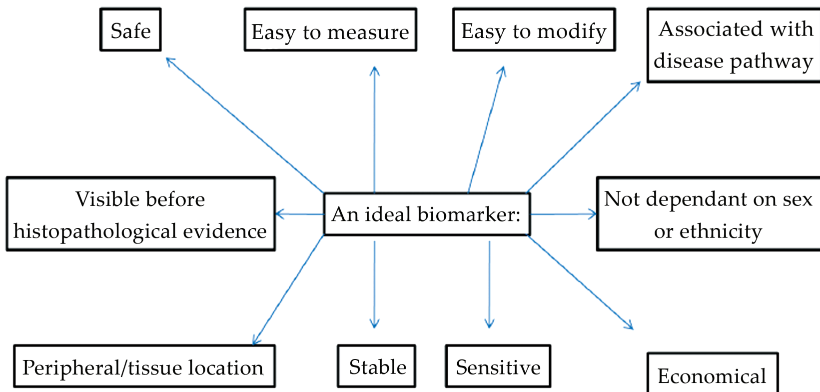


Figure 1.9. An ideal biomarker.

1.7. BIOMARKERS OR CLINICAL ENDPOINTS

Biomarkers refer to characters that can be subject to quantification in a pathway. It is not vital that biomarkers have a correlation with the symptoms or ease of a condition. However, clinical endpoints refer to the feeling or functioning or survival of a subject [4]. The focus here is on the symptoms or ease of a particular subject under scrutiny.

An important clinical endpoint for diseases such as AIDS is the survival of a subject. Other conditions may look at other manifestations such as nervous or heart issues or infections. These endpoints present a lucid picture that offers insight in the course of a disease as well as any requisite further medical course. However, apart from these clear endpoints there are other endpoints such as pain or breathing patterns that lack clarity.

The field of biomedical research looks at such clinical endpoints as main and only factors associated in research.

Biomarkers have been pitched in as surrogate endpoints in the fields of clinical trials. They are substitutes for such clinical endpoints; however, not all biomarkers can be pitched in as such substitute or surrogate endpoints. The use of biomarkers as such indicators requires validated data that shows such a biomarker is reflective of a particular health issue. It requires data that includes pathology or therapy epidemiology or physiological aspects to show the efficiency of the biomarker.

The use of biomarkers that have been tested statistically can also not be an integral component of a particular system that causes an observation/

disease. The presence of a biomarker as a pathway always can serve erroneous. In the case of diseases with multiple pathways, biomarkers can serve as indicators of an altered pathway (sometimes not connected to a disease pathway) in an indirect fashion.

Hence, such biomarkers can be a stand-in of such endpoints but cannot serve as a substitute.

1.8. MERITS AND DEMERITS OF BIOMARKERS AS SURROGATE ENDPOINTS

Two challenges recorded for the use of clinical endpoints:

- Clinical endpoints in certain conditions such as in cardiovascular conditions can be seen after long durations of time.
- Certain clinical endpoints (for instance, survival) can be challenged by ethical issues as well as the practicality of using such indicators.

1.8.1. On the Hand

The use of biomarkers can facilitate the use of smaller groups for a study. The use of such shorter durations can facilitate studies that can yield shorter times for the approval of a study. This can make the field of discovery of drugs more rapid. This can thus aid in efficient utilization of materials, time and funds.

The field of biomarkers can serve as interim evidence for a particular study till the entire data of clinical studies is revealed. The use of established biomarkers that can function as surrogate endpoints can reveal any unwanted effect on a study. This data can be used to appropriately alter or redesign the study [1].

1.9. WHICH ARE THE BIOMARKERS USED?

1.9.1. DNA Biomarkers

The increase in concentration of serum DNA is associated with several diseases such as auto immune diseases, infections, sepsis or auto immune diseases. Also mutations in several genes such as repair genes or tumor suppressor genes or oncogenes all are DNA used as biomarkers. For instance,

KRAS mutations are indicative of several metastatic cancers. Other genes such as *RAS* or *CDKN2A* as well as RB1 are all indicative biomarkers.

Certain SNPs are associated with the development of certain cancers such as lung, prostate and breast.

Mutations in DNA of mitochondria are biomarkers of disease such as colon or neck or head.

Other biomarkers include epigenetic modifications. For instance, CpG methylation and gene silencing has emerged as a target. Hyper methylated DNA has been detected in serum in lung/colorectal cancers or saliva in oral cancer cases. The variations in methylation can serve as a tool to differentiate between hyperplasia and cancer of prostate.

1.9.2. RNA Biomarkers

The expression of mRNA has been facilitated through the use of high through put technologies. The analysis of RNA biomarkers finds expression as finger prints or the profile of multiple genes. The analysis of profiles of several genes over single genes offers more accuracy however also requires advances in statistics as the use of several genes requires an additional effort in analysis.

The use of supervised algorithms in double hierarchical clustering that form clustered image maps have been of value.

Molecular subtypes were identified for the first time when the expression of RNA was analyzed in clinical breast cancer samples that were linked to survival. The added benefits of such an analysis included:

- Prediction of probability of metastasis.
- Increase in prognosis.
- Prediction of the grade of cancer.
- Prediction to neo-adjuvant therapy.

1.9.3. Protein Biomarkers

The analysis of the protein fingerprint has been shown to be more accurate than single proteins. Several methods shall be discussed in subsequent chapters such as DIGE (2D Difference Gel Electrophoresis), or MudPIT (Multi-dimensional protein Identification technology) have been shown to detect small quantities of protein (µg quantity).

Apart from protein levels other important aspects include posttranslational modifications such as phosphorylation or glycosylation. The use of reverse phase arrays has been valuable in the analysis of such phosphorylation pathways while laser capture microdissection has facilitated study of malignant cells and their expression [22].

1.10. GIST OF SIGNIFICANCE OF BIOMARKERS

Many biomarkers have become part of day-to-day parlance as well as in medical industry. Examples include the testing of Liver function tests for liver toxicity or prostate-specific antigen (PSA) for the analysis of status or health of prostate cancer.

PSA:

The status of prostate disease or cancer can be monitored using PSA: a common biomarker. (Figure 1.10).

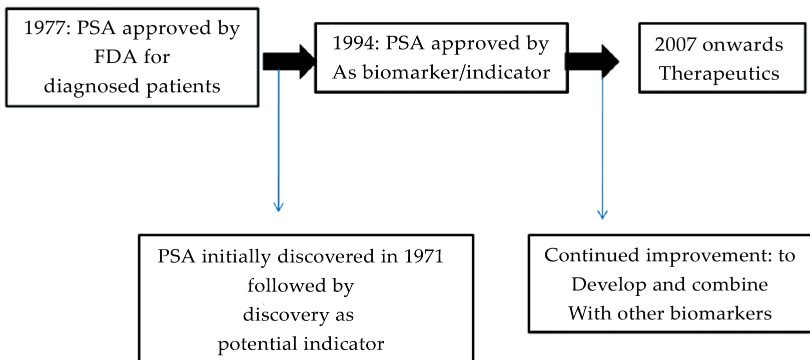


Figure 1.10. PSA as a biomarker.

Almost since 1970s, it has been 30 years since the biomarker has been put into use and applied in the field of biomarkers as seen in the above figure.

Cancer:

A host of biomarkers are being analyzed for oncology: a few of those discussed above in the section of "which are biomarkers used?." Another example is DNA resistance arrays that can estimate the resistance to a drug which in turn can facilitate the treatment of a patient. Breast cancer has shown a biomarker HER2: human epidermal growth factor receptor 2 that can indicate the type of tumor.

Vascular:

Several vascular diseases such as stroke or heart issues have been explored to scour for biomarkers. Matrix Assisted Laser Desorption Ionization/ Mass Spectrometry (MALDI/MS) and Electrospray Ionization Mass Spectrometry (ESI-MS) have aided in the discovery of 95 potential biomarkers from 731 proteins in cardiovascular analysis.

The role of such biomarkers is expected to make the field of scanning for such diseases less invasive and within shorter times that can facilitate quick treatment.

Neuroscience:

Clinical trials can involve a long time to address common issues such as Alzheimer's disease. The search for biomarkers that can be image can address such diseases including bipolar disorders, psychosis or schizophrenia [24].

1.11. CASE STUDIES OF ANALYSIS OF BIOMARKERS IN A CLINICAL CONTEXT

- The Genome Canada Initiative in Applied Human Health Genomics at the University of British Columbia looked at a biomarker project to look at transplantation. The use of biomarkers in the analysis of rejection of transplanted organs such as heart, liver or kidney can offer minimal invasive and quick analysis. Many peptides, mRNAs and metabolites have been identified as candidate biomarkers that can aid the field of clinical transplantation [25].
- The Biomarker Consortium a branch of Foundation for the National Institutes of Health (FNIH). This aims at bridging private and public sectors in several projects to hasten the process of biomarkers. The areas covered include:
  - Neuroscience: depression
  - Metabolic diseases.
  - Oncology.

Another branch is Genetic Association Identification Network (GAIN) [24, 26]

1.12. DISCOVERY AND VALIDATION OF BIO-MARKERS

The discovery and validation of biomarkers require precision and accuracy despite the cost and time involved. The validation of Type 0 biomarkers can be longitudinal: with studies against a reference standard in a population that is well defined. The validation of Type I biomarkers can be done parallel along with a drug while that of Type II biomarkers should encompass both the drug mode of action as well as the disease pathogenesis (Figure 1.11).

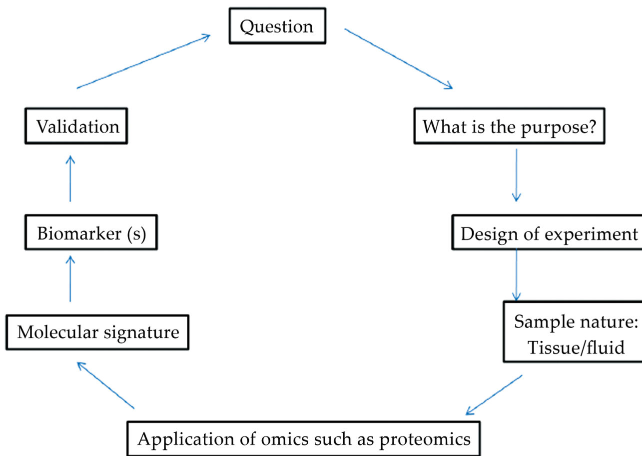


Figure 1.11. Discovery.

1.12.1. Question: What is the Biomarker Looked At?

Purpose: Are the biomarkers indicative of a condition? Are the biomarkers associated with a pathway?

Design of experiment and sample nature would depend on the above two criteria.

Molecular signature/ Biomarker: Is a single marker being targeted? Is a panel being looked at?

Validation is covered in the portion below.

Various factors involved in the assessment of biomarkers include:

- Quality of sample
- Variability of sample
- Design of experiment

- Optimal algorithms and modeling
- Capable technical platform
- Databases
- Cost benefits
- Knowledge assembly tools

1.12.2. Design of Experiment

This requires a clear understanding of the expected result of the biomarker. The quantity of samples to be analyzed should be statistically significant as well as the use of appropriate controls. The approach should be decided: targeted or global while the sample should be analyzed: whether it is a tissue or a body fluid.

1.12.3. Quality of Sample

Depending on the sample; the output of biomarkers would be proportional. Knowledge about sample quality as well as the source are vital. As samples possess variability, statistics and bioinformatics come into the picture.

1.12.4. Technology

This is the application of appropriate "omics" such as proteomics that is the scope of this book. An ideal platform is expected to be sensitive, accurate and as well as imaging and precise quantitation.

1.12.5. Bioinformatics

Data visualization tools can aid in the discovery of biomarkers. An added requirement is the necessity of free tools. Integration of data is attributed to, to name a few:

- Max Planck Institute (Heidleberg, Germany)
- Institute for Systems Biology (Seattle, WA, USA)
- Gene Network Sciences (Ithaca, NY, USA)
- Ingenuity Systems (Mountain View, CA, USA), as reviewed by Naylor [27].

1.13. BIOMARKERS IN A CLINICAL SCENARIO: APPROACH

Research to aid in increasing focus on patients can be facilitated by biomarkers. (Figure 1.12)

Collaboration:

The association of several companies/research institutes to study biomarkers aids in lessening the cost and promote the use of biomarkers. As the validation of biomarkers is not Intellectual property, the use of such collaborative projects can hasten the process of discovery.

Use of biomarkers:

The definition of a validated biomarker is one that has been measured in a framework that is scientific or is backed by evidence. The results should be explained in all aspects such as clinical, pharmacologic, physiologic and toxicologic [28].

Information based medicine:

Trials can be eased by the use of biomarkers as selection of patients who lack toxicity can rely on biomarkers. The use of adaptive trials where modifications are done along with experiments can aid in reduction of cost as ineffective conditions are eliminated.

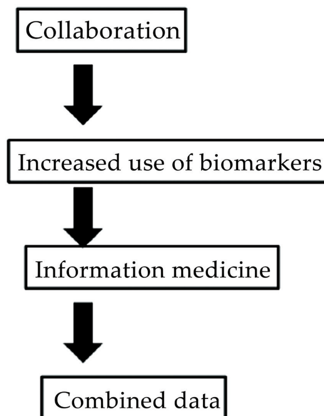


Figure 1.12. Use of biomarkers.

Combined data:

Across collaborations and partners the maintenance of operating conditions with cohesiveness can also contribute to optimum research in biomarkers [24].

1.14. VALID BIOMARKERS AND CLINICAL RESEARCH

The development of validated biomarkers can aid in the development of clinical research as highlighted below:

- Biomarkers can aid in the relation between dose and response
- They can highlight the events in progress and history of a disease.
- The various steps involved in disease from an exposure can be deduced.
- The various pathways and mechanisms between exposure and disease are characterized.
- There is reduction in wrong classification of a disease and risk due to use of validated biomarker.
- The risk can be assessed at individual and group levels.
- The studies such as effect modification can be analyzed.

A main lapse in biomarkers being successful is the inability to follow the rules of non-biological factors. The construction of pilot studies followed by effect of variation among individuals, stability and location of the biomarker can aid in efficiency [10].

1.14.1. Merits of Biomarkers

- The validated biomarkers can fulfill reliability.

The reliability is very vital in such studies. There can be compromises in reliability if there are alterations in staff or storage or procedure or conditions.

This requires the use of pilot studies as well as intra class correlation coefficients for consistent results.

- The process of disease can be revealed.
- The measurement is often objective.
- Standard practices such as objectives can be assessed.
- The measurements can be precise.

- The risk and disease can be homogenous as biomarkers can overcome bias.

As bias is an outcome in any study, the incidence can be avoided or decreased by:

- High rate of response from tests and controls
- Review of the study procedure and analyze the source of potential bias

1.14.2. Challenges with Biomarkers

- The cost associated with the technology can be expensive. For example, storage of all the kinds of cells as well as DNA isolated from them is costly.
- Certain samples can be difficult to handle: also, longevity of samples is a challenge with certain samples.

The storage of biomarkers dictates its stability. Serum samples banked can serve of help irrespective of the time. However, markers such as vitamins are sensitive to light leading to challenges in storage.

The fluids such as urine or blood can be advantageous while the use of tissues such as nervous tissue or cerebrospinal fluid is challenging to collect and possess an element of risk.

This challenge can be overcome by the use of Pilot studies.

- The studies of biomarkers in certain samples can be ethical issues.
- The processing of samples in the lab can give rise to errors. This can give rise to Intra individual variability.

The use of manuals that are well-organized procedures along with quality assurance can aid in lessening such errors of wrong measurement.

- Style of working of the workers can lead to errors. The use of appropriate SOPs or gloves by workers can influence the outcome. Newly appointed staff should be properly trained.
- Variability is a challenge. Due to differences in metabolism, there can be variations among individuals. The potential origins of variability during the search for a biomarker should be analyzed before the actual testing that would avoid erroneous grouping of a biomarker with a disease.

The variations between individuals should be accounted for such as the nutritional intake or details such as personal habits.

The apparatus used for collection of sample or transportation can cause an effect on the biomarker.

In order to lower the issue of confounding, various data such as sex, age, diet along with other internal and external factors should be accounted before the initiation of the study.

Point to note: To prevent erroneous association of a marker and a manifestation following are to be kept in mind:

- The use of appropriate false negatives and false positives
- Estimate of negative predictive power and positive predictive power
- receiver-operator characteristic (ROC) curves
- Clarity regarding the biomarker.
- Over interpretation of results can be lessened. For example a biomarker that is linked to a particular disease should not be used as a diagnostic tool till the biomarker is part of the disease.

To illustrate, APOE-ε4 allele is associated with Alzheimer's disease however it cannot be diagnostic as few individuals with the allele do not develop the disease [10].

1.15. CASE STUDY OF BIOMARKERS DETECTED BY METABOLOMICS

The analysis of metabolome was carried out to evaluate the role of a diet of Mediterranean origin: use of nuts and extra-virgin olive oil. The group was composed of patients who lacked diabetes. The urine of the control as well as those administered the diet was subjected to 1H NMR at year 1 and year 3.

It was found that the group that received Mediterranean diet showed expression of several metabolites such as:

- Creatine;
- amino acids such as glycine, proline;
- lipids such as suberic acid and oleic acid;
- Carbohydrates such as citrate, 3-hydroxybutyrate;
- Metabolites from microbes such as phenylacetylglutamine [2].

1.15.1. Novel Biomarkers

1.15.1.1. Longevity: A Case Study

An attempt was made to use biomarkers as a true indicator of age of the body that could facilitate an estimation of time left for an individual. A team in 2016 used Framingham Risk Score (FRS) to predict the risk of cardiovascular disease for a period of 10 years. Several markers were analyzed such as serum glucose levels, insulin-like growth factor-1 (IGF-1), Total/HDL ratio, free triiodothyronine (fT3), adiponectin, vitamin D3, ApoE, and leptine apart from BMI, APOE isoform, among others were tested.

It was reported that in females, specific lipids such as sphingomyelin and phosphatidylcholine can predict longevity while in males, IgG glycosylation measures and APOE234 genotype were the predictors of longevity [29].

1.15.1.2. Biomarkers for Nephrotoxicity: A Case Study

Using animal models, a set of biomarkers for nephrotoxicity was performed by Novartis. The compounds tested were 8 nephrotoxic ants (that could damage podocytes or induce oxidative stress) and 2 that targeted the liver (associated with cancer) to be used as controls over duration of between 14-21 days. The expression of genes was observed along with ELISA tests and traditional toxicology analysis. The observation of the various lesions in the kidney was then correlated to various biomarkers. A damage with cisplatin that displayed cell death was than analyzed for biomarkers that could reflect the damage and also the establishment of a correlation of level of damage and biomarker level.

Of the several biomarkers tested, the results are as follows for cisplatin:

- Serum creatinine showed a high level in animal models that were treated to high doses of the toxin.
- Blood urea nitrogen (BUN) also showed high levels in animals subjected to large quantity of the toxin.
- An effective biomarker reported was, Kim-1 (kidney injury molecule-1) that was present in both middle and high-grade toxin treatments.
- Another marker, urinary clusterin showed similar properties as above but the numbers of false negatives were more.

ROC (receiver operating characteristic) analysis of the above biomarkers revealed that maximum area was recorded by Kim-1 with the authors describing a “nearly perfect” curve.

For other nephrotoxicants, creatinine was not found as a good marker for damage to glomerulus. Such damage was found to be reflected by urinary proteins such as urinary β2-microglobulin and urinary cystatin C.

The team reported that several valid biomarkers were detected in the study. However, instead of a single marker, a battery of markers can serve as indicators of toxicity to kidney [23].

1.15.1.3. Biomarkers for a Deadly Disease: Case Study: Oral Cancer

Cancer presents a grim scenario with 5, 50,000 mortalities in a year in India mainly due to the delay in diagnosing the disease. The early diagnosis preferably before metastasis can aid this menace before it strikes or reaches an incurable condition. Here, the role of biomarkers (tissue or serum) can be significant especially in such a clinical scenario.

Biomarkers in this scenario can:

- Detect initial stages of malignancy.
- The various changes at molecular or genetic level in various stages of tumor progression.
- They can present the efficacy or toxicity of drugs such as chemo preventive agents.
- The prognosis can be boosted.

In carcinogenesis, there are several steps where biomarkers can enter the picture shown in Figure 1.13.

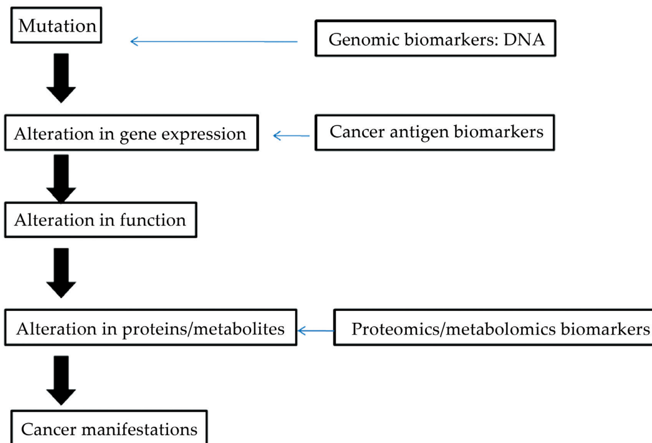


Figure 1.13. Biomarkers in carcinogenesis.

The use of saliva to search for biomarkers has been covered by several teams. The overall categories include:

- DNA, RNA and Proteins: to detect head and neck carcinomas.
- Viruses such as Epstein-Barr virus in carcinomas.
- Glycoproteins or Mucins to diagnose head and neck cancers.

Biomarkers are expected to lower the patient numbers and follow up time. These markers can show the efficiency and doses of chemo-preventive agents.

The validation of biomarkers can involve clinical trials, web-based electronic tools, literature studies as well as on the bench techniques.

The major limitations in the development of such biomarkers include:

- Limitation of validation or characterization.
- Limitation in selecting appropriate markers.
- Limitation of techniques applied.

Point to note: Several techniques are required to increase the presence of biomarkers amidst the several milieus of proteins and peptides in an individual [14].

1.16. SUMMARY

The role of biomarkers has been described as critical in the process of drug development. The role of biomarkers has received approval from FDA. Additionally, the role of biomarkers in research and clinical research and as surrogate end points is being promoted by FDA. The complete understanding of physiology is vital so that they can be used to replace clinical end points to avoid misclassification.

The applications of tools of molecular biology and omics such as proteomics can serve to make the field of biomarkers more strong and useful. These biomarkers can aid in the mechanisms of disease as well as trials and diagnosis along with disease management [13].

The role of biomarkers in the field of health such as carcinogenesis has shown much promise. In a disease like hepatocellular carcinoma (HCC) that has high incidence and low survival, several biomarkers have been identified that are summarized (Table 1.1) [22].

Table 1.1. Biomarkers in HCC

Marker	Phenomenon
Telomerase	Potential of regeneration
VEGF, angiotensin, IL-8	Angiogenesis
P53, c-myc, p21	Proliferation

The role of molecular biology and high throughput technologies can aid in the development of the field of biomarkers. There have been suggestions to invest time and finances in the field of biomarkers [22].

There are several types of biomarkers described in this chapter such as genomic, metabolomic, proteomic. There are Type 0, 1 and 2 biomarkers while several systems of classification have been proposed.

Reports indicate many **proteomic biomarkers** that can aid in improving the clinical scenario. For example, a panel of 273 peptides (classified by CE-MS) has been used for Chronic Kidney Disease (CKD). This panel could establish the progress of normoalbuminuria to macroalbuminuria in urine of diabetes type 1 and 2 patients [30].

REFERENCES

1. Strimbu, K, & Tavel, J. A. (2010). What are Biomarkers? *Current Opinion in HIV and AIDS*. 5(6), 463–466.
2. Corella, D., & Ordovás, J. M., (2015). Biomarkers: background, classification and guidelines for applications in nutritional epidemiology. *Nutricion Hospitalaria. 31*(Supl. 3), 177–188.
3. Hulka, B. S. (1990). *Overview of biological markers. In: Biological Markers in Epidemiology* (Hulka, B. S., Griffith, J. D., Wilcosky, T. C., eds.), pp. 3–15. New York: Oxford University Press.
4. Biomarkers Definition Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics. 69*, 89–95.
5. *WHO International Programme on Chemical Safety Biomarkers in Risk Assessment:* Validity and Validation. 2001.
6. *WHO International Programme on Chemical Safety Biomarkers and Risk Assessment:* Concepts and Principles. 1993.
7. Henry, N. L., & Hayes, D. F. (2012). Cancer biomarkers. *Molecular Oncology. 6,* 140–146.
8. Potischman N, & Freudenheim, J. L. (2013). Biomarkers of nutritional exposure and nutritional status: an overview. *Journal of Nutrition*. *133*, S3873–S3874.
9. Naylor, S. (2003) Biomarkers: current perspectives and future prospects. Expert Reviews in *Molecular Diagnosis*. *3*, 525–529.
10. Mayeux, R. (2004). Biomarkers: Potential Uses and Limitations. *NeuroRx*. *1*(2), 182–188.
11. Schulte, P. A. (1993). A conceptual and historical framework for molecular epidemiology. In: *Molecular Epidemiology: Principles and Practices* (Schulte, P. A., Perera, F. P. eds.), pp. 3–44. San Diego: Academic Press.
12. Merikangas, K. (2002) Genetic epidemiology: bringing genetics to the population-the NAPE Lecture 2001. *Acta Psychiatrica Scandinavica*. *105*, 3–13.
13. Mayeux R. (2003) Epidemiology of neurodegeneration. *Annual Reviews in Neuroscience*. *26*, 81–104.
14. Ali, S., Bansal, P., Bhargava, D., & Kunal, V. (2014). Biomarker: Validation, Characteristics, Pitfalls and Importance in Oral Squamous

Cell Carcinoma. *Journal of Evolution of Medical and Dental Sciences*. *4*(1), 88–94.

15. Drucker, E., & Krapfenbaue, K. (2013) Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalized medicine. *EPMA Journal*. *4*, 7.
16. Al-Tarawneh, S. K., Border, M. B., Dibble, C. F., & Bencharit, S. (2011) Defining Salivary Biomarkers Using Mass Spectrometry- Based Proteomics: A Systematic Review. *OMICS*. 15 (6), 353–361.
17. Atkinson, A. J., Colburn, W. A., & DeGruttola, V. G. (2001) Biomarkers and surrogate end points: preferred definitions and conceptual framework. *Clinical Pharmacological Therapy*. *69*(3), 89–95.
18. Frank. R., & Hargreaves, R. (2003) Clinical biomarkers in drug discovery and development. *Nature Reviews in Drug Discovery*. *2*(7),566–580.
19. Jenab, M., Slimani, N., Bictash, M., & Ferrari, P. (2009). Bingham SA. Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Human Genetics*. *125*, 507–525.
20. Chan, X. H., Huang, S., & Kerr, D. (2011) Biomarkers in clinical medicine *IARC Science Publication*. *163*, 303–22.
21. Potischman, N. (2003). Biologic and methodologic issues for nutritional biomarkers. *Journal of Nutrition*. 133, 875S–880S.
22. Kumar, M., Sarin, S. K. Biomarkers of diseases in medicine. Current Trends in Science. Platinum Jubilee Special 2009; 403–417. Availble at- www.ias.ac.in/pubs/splpubs/ pjubileebook/403.pdf.
23. Institute of Medicine (US) *Forum on Drug Discovery, Development, and Translation. Emerging Safety Science: Workshop Summary*. Washington (DC): National Academies Press (US); 2008. 7, Qualifying Biomarkers. Available from: https://www.ncbi.nlm.nih.gov/books/ NBK404.
24. McCormick, T., Martin, K., & Hehenberger, M. (2007). The evolving role of biomarkers. Focusing on patients from research to clinical practice. *IBM (Imaging) Biomarker Summit III IBM Life Sciences*. 1–24.
25. McManus, Bruce, MD, PhD, FRSC. “Biomarkers of Acute and Chronic Allograft Rejection in Heart, Kidney and Liver Recipients.” IBM (Imaging) Biomarker Summit III, January 2007

26. Mittleman, B. (2007). Public-Private Partnerships at the NIH. *IBM (Imaging) Biomarker Summit III*. IBM (Imaging) Biomarker Summit III, January 2007

27. Naylor, S. (2005) Overview of biomarkers in disease, drug discovery and development. *Drug Discovery World Spring*. 21–30.

28. U.S. Department of Health and Human Services. Guidance for Industry (2005). Pharmacogenomic Data Submissions.

29. Beekman, M, Uh, H. W., van Heemst D., et al. (2016). Classification for Longevity Potential: The Use of Novel Biomarkers. *Frontiers in Public Health*. *4*, 233.

30. Frantzi, M., Bhat, A., & Latosinska, A. (2014). Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker development. *Clinical and Translational Medicine. 3*, 7.

CHAPTER 2

PROTEOMICS: AN OVERVIEW OF TECHNIQUES INVOLVED

CONTENTS

2.1. Design of Biomarkers 36
2.2. Biobanking of Samples 36
2.3. Proteomics And Discovery of Biomarkers 37
2.4. A Quick Snapshot 38
2.5. Techniques For Biomarker Discovery Using Proteomics 39
2.6. The Mass Analyzers 42
2.7. Variations of Ms 43
2.8. Variations In Ms Technologies In Proteomics 44
2.9. Ief Fractionation Methods 46
2.10. 2D Gel Electrophoresis (2-DE) 47
2.11. 2D Difference Gel Electrophoresis (2D-DIGE) 50
2.12. Stable Isotope Labeling By Amino Acids In Cell Culture (SILAC) 52
2.13. Isotope-Coded Affinity Tag (ICAT) 54
2.14. Western Blotting 55
2.15. ^{18}O Stable Isotope Labeling 55
2.16. Isobaric Tagging Reagent For Absolute Quantitation (ITRAQ) 56
2.17. Ce-Ms 58
2.18. Stable Isotope Dilution Mass Spectrometry (SID-MS) 59
2.19. Microarrays 59
2.20. Multidimensional Protein Identification Technology (MUDPIT) 61
2.21. Label-Free Techniques 62
2.22. Another Proteomics For Biomarker Discovery 63
2.23. Bioinformatics And Proteomics And Biomarkers 65
2.24. Glimpses of Discovery of Biomarkers Using Proteomics 68
2.25. Case Studies: Biomarkers of Bladder Cancer Using Proteomics 68
2.26. Final Note 70
References 73

As much as 98% of drug targets are accounted for by proteins in the field of therapeutic intervention [1]. Thus, the role of proteomics in the field of discovery is biologically significant. The several aspects of proteomics include the discovery of biomarkers that can cover several fields such as plants/animals or microbes as well as clinical [2].

2.1. DESIGN OF BIOMARKERS

The search for biomarkers has been ongoing since several decades with several attempts. A main challenge is the absence of a gold standard against which a potential marker can be compared with. This necessitates a thorough and particular approach in order to develop as well as subsequently validate a biomarker. This can be coupled to extensive data regarding the associated fields as well as a lucid comprehension of the objective [3] (Figure 2.1).

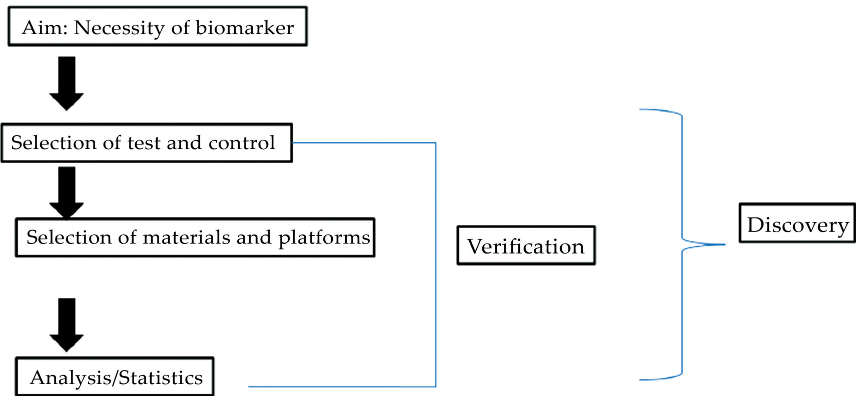


Figure 2.1. Design of biomarkers [4]

2.2. BIOBANKING OF SAMPLES

In order to assess for biomarkers, several samples of tissues or samples are essential, such a collection of such sample sources is known as biobank. (Figure 2.2).

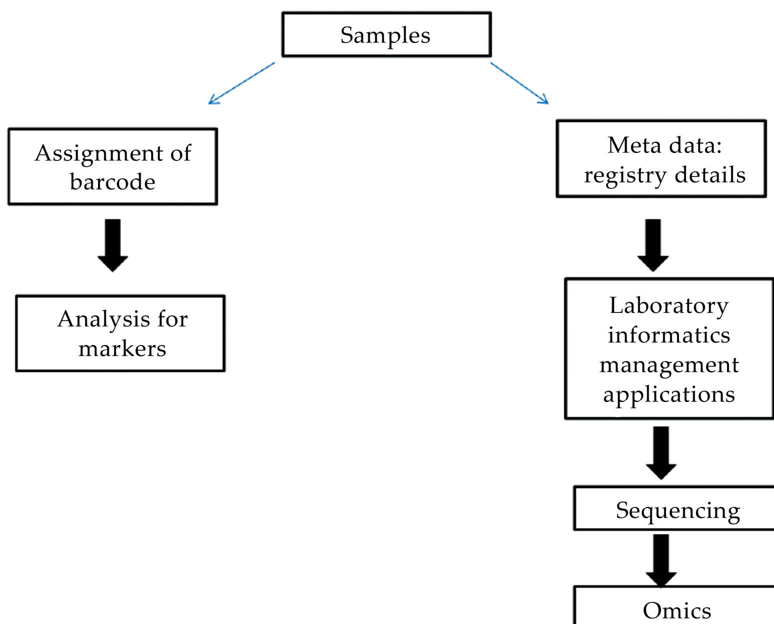


Figure 2.2. Biobank.

Such a biobank should:

- Have many samples with good quality;
- Access should be easy;
- The necessary data must be available;
- A unique ID number is important;
- Data base management for instance: laboratory informatics management applications (LIMS) [4].

2.3. PROTEOMICS AND DISCOVERY OF BIOMARKERS

A methodology in proteomics for the discovery of biomarkers is expected to be:

- Accurate;
- Well performing;
- Robust [5].

The discovery of biomarkers involves 3 stages: discovery, verification and validation at which proteomics can be applied at each stage. (Figure 2.3).

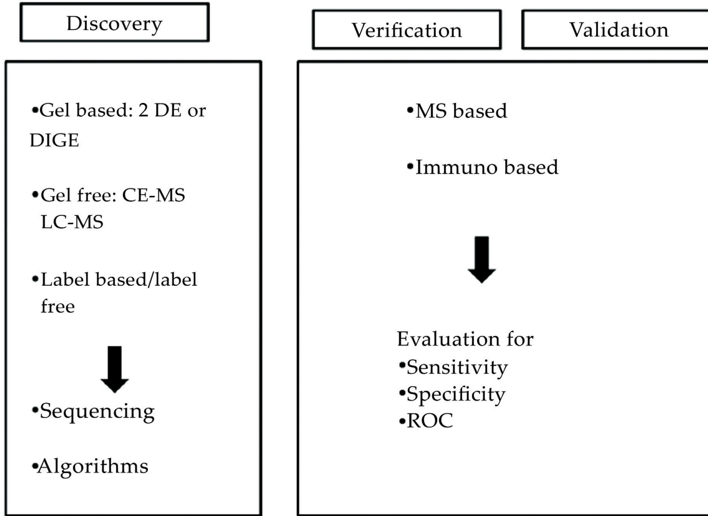


Figure 2.3. Proteomics in biomarkerdiscovery: The terms ROC have been presented in Chapter 1.

The discovery of biomarkers has been described with 2 approaches:

Knowledge based approach: Here the potential members of biomarkers are scoured for on the basis of the origin or steps involved in a disease.

Unbiased approach: Here the differences in protein expression between a control and test are used to search for biomarkers [6].

2.4. A QUICK SNAPSHOT

The realm of untargeted methods to approach biomarkers has seen the application of MS-based proteomics techniques. Certain challenges such as the quantity of samples analyzed as well as false discoveries. Less numbers of samples are not desired in the discovery phase as it can result in an over estimate of the numbers of biomarkers compromising sensitivity and specificity.

The approach to proteomics involves the separation of proteins using a gel based or gel free approach and a subsequent identification by MS.

Commonly used samples in a clinical scenario to analyze for biomarkers are body fluids such as cerebrospinal fluid, blood and urine. These liquids are complex and possess a range of proteins [4].

2.5. TECHNIQUES FOR BIOMARKER DISCOVERY USING PROTEOMICS

2.5.1. Mass Spectrometry (MS)

This technique is being pitched as a common method in the field of high-throughput proteomics. The instrument calculates mass of small size molecules by ionization of molecules. The ions are sorted o the basis of mass/charge (m/z) ratio in an electric field [2].

Whatever the technique of separation applied in proteomics, a mass spectrometer finds application in the identification of proteins [7]. The past several years have seen advancements in the realm of MS [8]. (Figure 2.4)

In many proteomic approaches, the use of MS has seen an increase. The mass spectrum generated is a plot of mass/charge (m/z) ratio and abundance of ions [9]

The major parts of MS are:

- ionization source

This generates ions from the sample molecules to form ionized analytes.

- mass analyzer

The ions generated are sorted on the basis of mass/charge (m/z) ratio. This part should possess:

- Accuracy of mass
- Sensitivity
- Resolution

With various advancements, proteomics can detect femtogram quantities of single proteins from mixtures [9]

- ion detection system

The ions are registered on the basis of mass/charge (m/z) ratio.

The analysis can involve steps such as:

- Ionization of proteins to generate gas-phase ions

- separation of ions on the basis of according to their mass to charge ratio
- detection of ions [7, 9, 10]

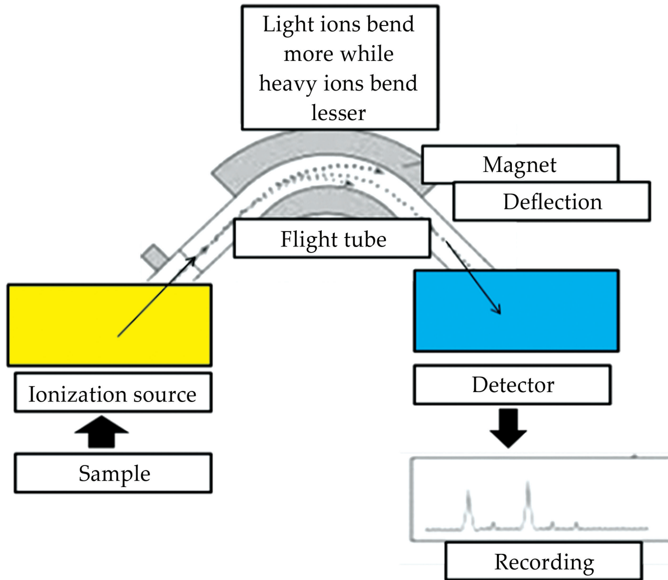


Figure 2.4. MS: Principle and scheme.

In techniques where gels are not used such as ICAT and MudPIT the MS is applied in a direct format. However, in techniques involving gel separation such as 2DE and 2D-DIGE the gel is subject to excision of proteins followed by enzyme digestion and then MS. The peptides derived are analyzed using several databases and search engines.

The ionization source uses protonation or deprotonation [9] and involves the use of following techniques.

2.5.2. Electrospray Ionization (ESI)

A high voltage needle generates a spray of charged analytes in the gas phase. As solution of analytes is converted to gas phase, several charges and ions are generated [7]. The injection through the fine capillary generates ions. This technique has found application in the fields of folding, interactions and structures of proteins [9].(Figure 2.5).

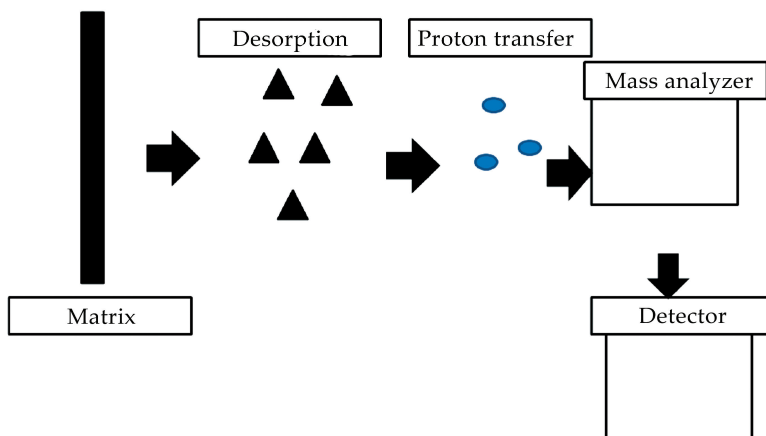


Figure 2.5. Scheme of ESI-MS: sample and solvent passed through charged needle.

2.5.3. Matrix Assisted Laser Desorption/Ionization (MALDI)

Here the sample and matrix are crystallized on a metal surface. Following laser excitation, the matrix is excited that causes the sample ions also to enter the gas phase. The development of "mass fingerprints" is a hallmark of this technique [7]. The matrix absorbs energy of laser that then transfers energy to sample to generate ions. The use of MALDI has facilitated the application of profiling of molecules [9]. (Figure 2.6)

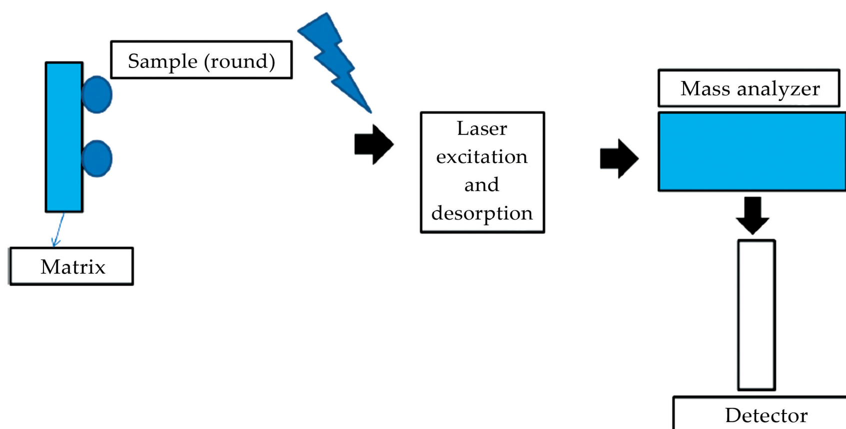


Figure 2.6. A scheme of MALDI-MS: the crystallized sample mixed with the matrix is excited by laser.

2.6. THE MASS ANALYZERS

2.6.1. Quadrupole Mass Analyzer

The analyzer utilizes electric field as sample ions pass a central axis surrounded by four parallel poles that are at equal distances. This Quadrupole has each pair of metal rods facing each other through which voltage of Radio Frequency is applied. Depending on the voltage, ions of a requisite mass-to-charge traverse and are detected while ions that follow an unstable trajectory are disintegrated. The technique has a high resolution coupled with accurate masses and good speed with linear outputs that are used in quantitative studies.

2.6.2. Ion Trap Analyzer

The technique uses a trap: a ring electrode and two end cap electrodes to capture ions. Following application of voltages, the ions of specified mass to charge ratios can be trapped on the device, dependent on the voltage applied. The ions out of range traverse trajectories that make them leave the analyzer. The technique in conjunction with an inert gas yields fragments for studies on structure.

2.6.3. Time of Flight (TOF) Analyzer

The ions produced are exposed to an increasing voltage to traverse an evacuated tube of known dimensions. Once the voltage is released, samples separate at a rate dependant on their mass to charge ratios. The technique is fast as well as has high sensitivity.

2.6.4. Fourier Transform-Ion Cyclotron Resonance (FT-ICR)

The ions generated enter a chamber subjected to electrical and magnetic fields that causes them to enter circular orbits. Excitation by an electrical field causes the production of a current per time that is transformed into orbital frequencies by Fourier transform which in turn is dependent on the mass-to-charge ratios. These analyzers can analyze a wide variety of mass with high resolution [11].

The ESI technique is combined with several mass analyzers such as quadrupole or ion trap while MALDI is generally paired with Time-of-flight (TOF) analyzer. The Fourier transform ion cyclotron resonance (FTICR) is a variation of the ion traps where a magnetic field is used to capture ions

instead of electrical fields. This technique has permitted the identification of zeptomoles (10^{-21}) of samples [2, 12].

2.7. VARIATIONS OF MS

2.7.1. Tandem MS

These instruments have facilitated the development in this field. There are two mass analyzers in which the first mass analyzer determines the mass of generated peptides. This is followed by selection of individual ions and breaking up by collision-induced dissociation (CID) that is analyzed by second mass analyzer. (Figure 2.7)

The mass analyzers may be same or different such as following examples:

- MALDI TOF-TOF with two TOF analyzers
- MALDI–Qq-TOF with first analyzer as quadrupole and second analyzer as TOF.

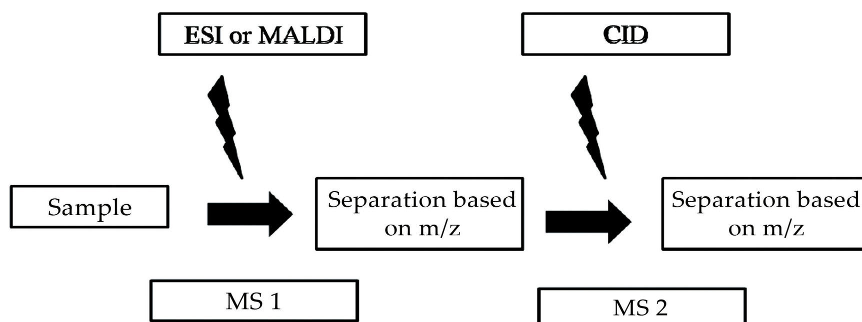


Figure 2.7. Scheme of Tandem MS.

These instruments can offer following features:

- The analysis in two steps facilitates determination of amino acid sequence to allow for more accuracy in identification of peptides.
- Complex samples can be used.
- Irrespective of the presence of sequence databases, the sequence of proteins can be deduced [2].
- While simple mass spectrometers calculate the mass; the determination of sequence of amino acids is performed by tandem mass spectrometers [7].

2.8. VARIATIONS IN MS TECHNOLOGIES IN PROTEOMICS

- Surface Enhanced Laser Desorption/Ionization (SELDI)

This has been reported as a versatile and strong approach and a branch of MALDI. The quest for biomarker (s) is enabled by this technique [2]. Instead of mixing sample with matrix as in MALDI, the sample here is placed on a chip. Subsequently, the chip is kept in a vacuum chamber to be ionized and subjected to detection [2]. The protein-chip arrays have several coatings such as affinity, anionic or cationic, hydrophobic or hydrophilic. Based on the property of the biomarkers or proteins of interest, they bind the surface followed by a wash step to remove unbound molecules [13, 14].

This technique has the following *merits*:

- Multiple samples can be analyzed in this approach to generate data points.
- It is high throughput [2, 13].
- It is reported as a tool for discovery of biomarkers where expression of proteins can be compared between a test and control [13].
- It is a versatile technique as the proteins coated on the chip can possess properties according to an experiment.
- It can use low volume of sample less than 10 µL [13].
- There is no requirement for protein concentration steps [13].

Challenges:

- An issue reported is reproducibility of results as well as challenges to sequence proteins that are of the discriminatory peaks [2].
- The importance of high operating standards as well as quality control can lessen the extent of bias in such studies [2].
- The standardization is a challenge as different factors such as surface coating to concentrations of salt and proteins can influence the proteome profile generated [13].

Case studies: Use of SELDI for biomarker discovery:

- In a study of LCM-microdissected prostate samples both controls and cancer samples, SELDI revealed alterations in proteome indicative of use of biomarkers [15].

- A 2006 study used SELDI to analyze cerebrospinal fluid (CSF) in patients with multiple sclerosis. A peak was identified in the patients that was later reported as an inhibitor of cathepsin B called cystatin C [16].
- LC-MS

Liquid chromatography (LC) involves separation of a protein using its properties of mass or affinity or hydrophobicity [17]. A mixture applied on capillary LC column can aid in the identification of proteins/ biomarkers that are abundant to a lesser extent.

Merits:

The technique has high resolution [13].

As hydrophobic proteins can be targeted the proteins not covered by 2DE can be revealed in this technique to make LC-MS complementary with such a technique as well as detailed evaluation of fluids such as urine for markers [13].

Challenge:

It consumes time;

It is subject to interference by substances such as salts [13].

Use of LC-MS in a quest for biomarkers:

A 2012 study evaluated urine samples of controls as well as IgA nephropathic patients to generate prognostic biomarkers as well as the classification of the disease(13)

- Ion mobility spectrometry (IMS)

This approach involves dispersion of peptide ions in gas phase to separate noise from species that are not abundant. The length of the run can be lowered by the application of this technique without affecting the throughput in LC based applications [7].

2.8.1. Interpretation of Results of MS

The spectrum generated by MS is subject to analysis to identify the tests. The technique of protein mass fingerprinting (PMF) can be an example. The use of a protease generates peptides of unique lengths that can facilitate the

identity of peptides. Though best results are facilitated by the use of a pure protein, the use of elementary mixtures can be facilitated with PMF.

The complexity of certain samples can be attributed to their modifications or isoforms. The development of non-specific peptides during digestion with proteases can add to the complexity.

Another challenge is potential modification of peptides during preparation of samples. For instance, oxidation of methionine's can be misinterpreted as leucine and isoleucine due to similar profiles of molecular mass [19].

Polymorphism in DNA can also generate the formation of different sequences of peptides to complicate the analysis [2].

2.8.2. Validation of MS Results

The validation of results following identification of samples using search programs is a next important step. The use of computational programs aids in boosting performance as well as an increase in accuracy to avoid false positives. For instance, a tool *Peptide prophet* uses probability scores to rank peptides [2].

2.9. IEF FRACTIONATION METHODS

The use of isoelectric focusing (IEF) has seen several applications as proteins can be subject to prefractionation on this technique. The use of such steps can aid in the easing out of the complexity of samples as well as the focus on samples with low abundance. This is a high-resolution electrophoresis technique that uses an in immobilized pH gradient (IPG) gel or ampholytes to separate peptides and proteins on the basis of isoelectric point (pI). Samples such as plasma or bacterial cultures have been applied to the technique. Following separation, the gel is cut and peptides are extracted and are of interest in proteomics.

This application has several applications in proteomics as summarized by Table 2.1.

Table 2.1. Proteomics and Electrophoresis

Technique	Principle	Pros	Challenges
Plat form 2D (Beckman Coulter)	Chromato focusing	Flexibility in samples	Low throughput
Multicompartment electrolyser (Proteome systems)	Using extremes i.e., basic or acidic isoelectric point, pI, the low abundance proteins are targeted	pI cutoff is accurate and large numbers of samples can be processed	Consistency
Rotofor (BioRad)	Solution-phase isoelectric focusing	Retention and large sample numbers	pI cutoff
Gradiflow system (Gragopore)	Membrane-based electrophoresis	Recovery	Sample capacity
Free flow electrophoresis (Becton Dickinson)	Charged analytes are separated	High load of samples and good recovery	Interference with MS due to buffers

The use of fractionation can enable samples to be applied on 2D electrophoresis.

Merits:

- It is micropreparative.
- High resolution
- Can separate up to 10,000 to 15,000 proteins
- The use of pI can be used in MS as a tool of additional identification [7].

2.10. 2D GEL ELECTROPHORESIS (2-DE)

In proteomics, this technique is reported as a primary tool [20]. The technique involves separation of proteins extracted on the principle of isoelectric point (pI) in a first step followed by a second electrophoresis that uses molecular weight [9].

Proteins that are extracted from a sample are rehydrated in a buffer to be first subjected to separation using pI immobilized pH gradient (IPG) strips followed by separation on SDS PAGE in a second dimension.

Protein spots are stained with appropriate dyes following which 2-DE software facilitates excision of differential spots that are digested with protease and subsequent MS for identification [9]. (Figure 2.8).

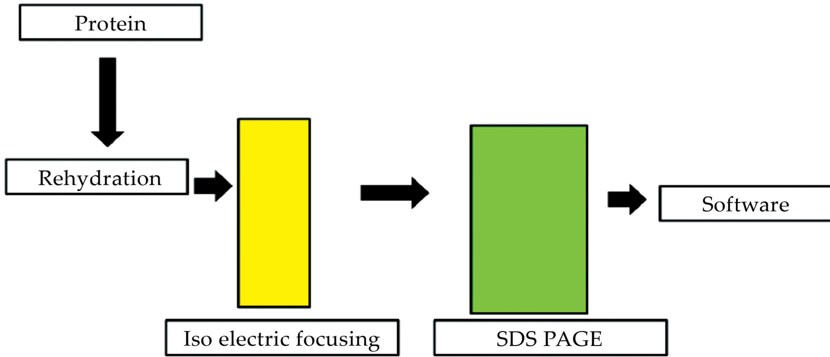


Figure 2.8. Scheme of 2DE: separation in 2 dimensions: pI and molecular weight.

Staining:

Coomassie Brilliant Blue or silver stains are commonly applied for the visualization of the protein spots [9].

Since this application is followed by MS; the sensitive stains should not be incompatible with subsequent steps. The sensitivity of silver and sypro ruby is similar but the latter is more reproducible. The use of sypro ruby can aid in a wider range and limited false positives [7].

Merits:

- The technique yields data on protein makers with respect to their pI, molecular weight as well as possible posttranslational modifications [7].
- It is efficient in the separation of complex protein mixtures such as tissue lysates [7].
- It can separate variants and modified proteins and finds extensive use [7].
- The identification of posttranslational modifications involves two approaches:

- The alteration in pI or molecular weight of a protein can be reflected due to a posttranslational modification. The shift in a spot can be indicative of a modification.
- The use of Western blotting where antibodies are applied can reveal the shifts in modifications.

Challenges:

- The technique is unable to detect low abundant samples [9].
- Ampholytes used lack stability at extremes of pH.
- Membrane proteins are hydrophobic and precipitate in this application or are insoluble [7, 9].
- Reproducibility is an issue [7]
- Streaking of spots are seen for basic proteins [7]
- The samples with extremes of pH (less than 3 or more than 10) are not sensitively detected [7].

Overcoming challenges:

- The extraction and solubility issues can use the application of treatment.
- A narrow range pH gradient can allow the increased separation of proteins that have similar molecular weights. The use of zoom gels can facilitate the separation of hydrophobic proteins [7].

2.10.1. Case Studies of Using 2-DE

A 2009 study examined the use of 2DE in the analysis of cancer samples. The technique showed 2000 unique spots; leukocytes of breast cancer patients showed the presence of 64 proteins that were expressed in a different format. These proteins were associated with key pathways of metabolism especially proteins associated with microfilaments that could be identified by this technique [21].

In another study in 2010, a research team reported 100 sets of proteins in breast cancer samples to serve as potential biomarkers [22].

These case studies illustrate the potential of 2DE in the detection of biomarkers.

2.11. 2D DIFFERENCE GEL ELECTROPHORESIS (2D-DIGE)

This technique is a development of 2-DE [9]. The technique involves the use of following sets:

- Control
- Test sample
- Internal control: (an equimolar combination of control and test)

The technique uses fluorophores such as Cy2, Cy3, or Cy5 to label the above tubes. Each of these fluorophores has different fluorescence that can be distinguished by the use of optical filters during the step of scanning [23]. These Cy dyes cyanine based dyes; that bind ε-amino group of lysine residues to on account of the N-hydroxysuccinimidyl group. The proteins labeled with different fluorescent dyes can be processed on a single gel [7]. (Figure 2.9)

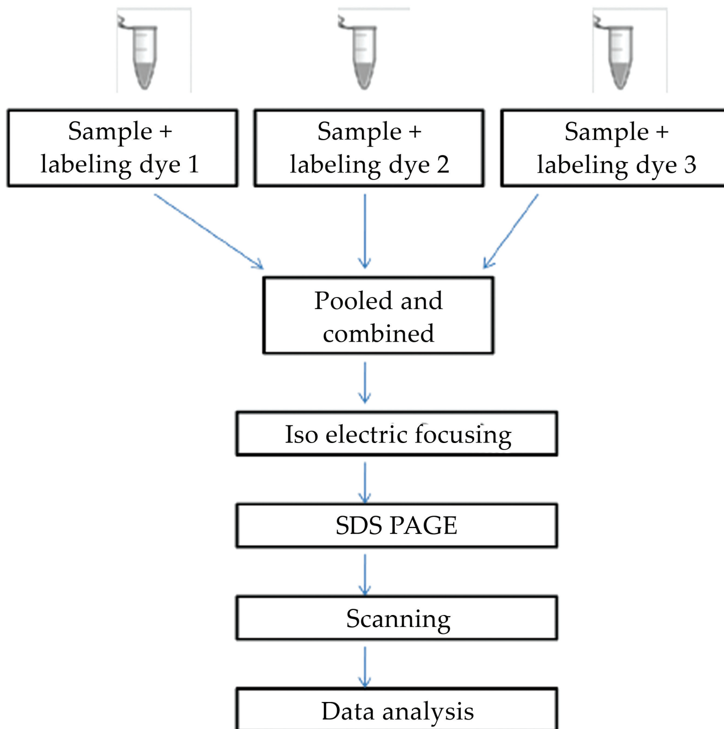


Figure 2.9. Scheme of 2D-DIGE.

The technique uses internal pool serves in the normalization of data. The use of a scanner aids several wavelengths to analyze the different wavelengths of the dyes;

- Cy2: 488 nm
- Cy3: 532 nm
- Cy5: 633 nm

Software's such as De-Cyder are specifically for the analysis of data generated by 2 DIGE [7].

Merits:

- The method facilitates the biomarker discovery.
- The technique offers an increased sensitive detection as well as linear data.
- The gel to gel variation is avoided as a single gel is used to analyze the test [9].
- Reproducibility is maintained [7].
- The steps subsequent to electrophoresis such as fixing and staining are not applicable [7].
- The dyes are sensitive and linear [7].
- The bias is reduced especially from variation of experiments.
- The accuracy of statistical data is enhanced by the use of the internal control in the samples.
- The protocol is straightforward.

Challenges:

- The fluorophores used are costly.
- Proteins that lack lysine are a challenge to label.
- The equipment used requires specialization [7].

2.11.1. Case Studies of DIGE Technique in Biomarkers

A 2011 study studied differences in protein expression between colorectal cancer of two types: invasive as well as noninvasive. The use of DIGE aided in the identification of carbonic anhydrase 2 as well as transgelin as markers of colorectal cancer that was also confirmed by western blotting and fluorescence-based quantitative polymerase chain reaction [24].

A 2012 study used 2D-DIGE along with MS that aided in the identification of biomarkers for prognosis in glioblastoma [25].

2.12. STABLE ISOTOPE LABELING BY AMINO ACIDS IN CELL CULTURE (SILAC)

In the quest for biomarkers, this is a quantitative technique in proteomics [26].

This technique uses the principle of labeled or heavy amino acids that are applied in the culture medium. These amino acids are incorporated into the cell during the cycle. Amino acids such as lysine as well as arginine are used for labeling.

The technique involves the culturing of cells to be tested (along with control populations). The test for example; tumor samples are cultured in a medium containing heavy isotope of an amino acid (mentioned above) while the controls are cultured in normal (light) isotopes. To achieve close to 95% labeling, five to seven passages are allowed to pass [27].

The samples are then combined in an equimolar ratio to be run on SDS-PAGE. These samples are digested with a protease like trypsin in solution or in the gel. MS analysis is the next step. The isotopes used for labeling can be discriminated in the MS analyzer. This is due to the shift in charge/mass ratio of the labeled amino acids against the unlabeled ones [9].

Thus, following culture of the cells in heavy/light amino acids for 6 generations yields maximum labeling. This is subjected to separation by electrophoresis followed by digestion and MS. The ratios of the heavy to light samples indicate the extent of biomarkers or their levels. With an increase in popularity in the technique, its applications are expected to increase [9]. (Figure 2.10).

Merits of SILAC:

- It is common for several types of cells [7].
- New biomarkers can be screened for based on the analysis of proteomes [7].
- It does not depend on the use of modifications by chemicals and is an in vivo technique [7].
- It also can reveal posttranslational modification as well as interactions among proteins [7].

- It does not require targeted analysis and is non -specific as all peptides are labeled [9].
- Accuracy is more [9].
- The technique is robust [9].

Challenges:

- It is challenging to apply the technique to tissue proteome in a direct format. To address this, heavy Nitrogen isotope based labeling on SILAC has been applied [7].
- It cannot be applied to autotrophs [9].
- The reagents are expensive [9].

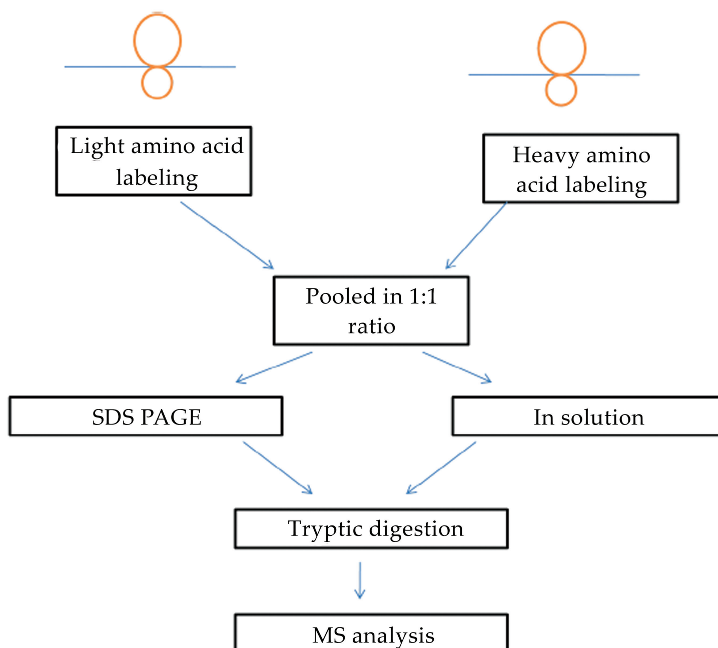


Figure 2.10. Scheme of SILAC.

2.12.1. Case Studies of SILAC

A 2012 study used analysis of SILAC to identify biomarkers in breast cancer cell lines followed by confirmation in tissue using immunohistochemistry. The biomarkers were reported as:

CRABP2 and IDH2: that indicates poor pgonosis

SEC14L2: biomarker of good prognosis

Thus, the study showed the utility of SILAC to identify biomarkers [28].

Apart from such in vitro applications, SILAC has been extended to in vivo by the inventor of the approach: Matthias Mann. In 2008, a team used heavy/ light isotopes of lysine-food for 4 generations of mice models. Following isolation of organs, the labeled amino acids were detected from the second generation [29]. Such a technique is pitched at analyzing functions of proteins by knockout of proteins in animal models. However, humans cannot be subjected to such analysis [9].

Advances in SILAC:

Super-SILAC is aims at improving the sensitivity of the SILAC technique by a combination of several cell lines. The use of multiple cell lines can serve as a larger base for the analysis of proteome against a single tissue used in SILAC. The technique has:

- More accurate
- Error is low [9].

A study in 2013 used super SILAC to analyze blood samples of breast cancer in different stages [30]. Using 11 cell lines indicative of breast cancer stages, proteomics revealed N-glycosylated proteins mainly of the secretory and membrane proteins with several biomarker candidates.

2.13. ISOTOPE-CODED AFFINITY TAG (ICAT)

The use of gel free methods is gaining popularity especially for quantitative techniques over traditional 2D gels. One of the common methods of chemical labeling is ICAT and is reported to be the first to use MS [31].

The technique employs labeling of protein samples using ICAT reagents: heavy or light. These reagents are made of a reactive thiol moiety followed by heavy linker and biotin to trap peptides. These reagents target the cysteine thiols that are then subjected to digestion by trypsin. The digested products are separated using multi-step chromatography. The use of MS can reveal the nature of the peptides as well as peak areas that reveal quantities of

putative test proteins. The data generated from MS of ICAT labels can ebe analyzed by programs such as Sashimi from the Institute of System Biology or proICAT from Applied Biosystems.

Merits of ICAT:

It can allow for the analysis and detection of proteins that are present at low concentrations

Challenges:

- Acidic proteins are a challenge
- The technique is selective for cysteine rich proteins [7].

2.14. WESTERN BLOTTING

The technique uses sample separation by SDS PAGE followed by immobilization of proteins on membrane. This is followed by use of antibodies that can distinguish non- comigrating proteins. The system can be scaled up with the use of a thousand antibodies being reported. The mobility can be analyzed by software to determine the protein.

Merits:

- The system is flexible.
- The function and verification of proteins can be done in a simplified manner.
- The antibodies used are well researched and immense data is available.

Challenge:

The technique cannot identify previously uncharacterized proteins [7].

2.15. ^{18}O STABLE ISOTOPE LABELING

The technique uses an exchange of Oxygen atoms at C-terminal carboxyl group of digested proteins with that of heavy oxygen (^{18}O). The use of

heavy water, i.e., $H_2^{18}O$ facilitates exchange of the 2 oxygen of carboxyl with the heavy oxygen atoms. This exchange causes a shift in mass of the peptides that can be used for detection as well as measurement of markers.

Merits:

- It does not need particular MS platform
- It is not specific for certain residues such as ICAT.
- It can be used on human samples
- The technique is efficient.

Challenges:

- The ^{18}O incorporation is non-uniform.
- Multiple samples cannot be subject to comparison(7)

2.16. ISOBARIC TAGGING REAGENT FOR ABSOLUTE QUANTITATION (ITRAQ)

An Isobaric Tagging Reagent for Absolute Quantitation (iTRAQ) label is the key in this quantitative technique.

The technique uses the isobaric iTRAQ labels composed of:

- Amine-reactive group that targets lysine chains
- Reporter (N-methylpiperazine)
- Balance group (Carbonyl)

Following sample digestion with trypsin, samples can be labeled with iTRAQ reagents. These are then pooled and then subjected to strong cation exchange (SCX) chromatography. Following the fractionation, the resultant molecules are subjected to MS. (Figure 2.11).

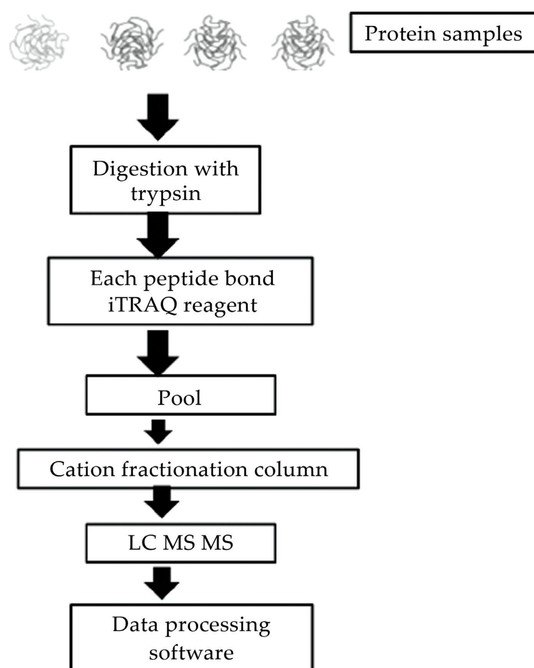


Figure 2.11. Scheme of iTRAQ.

The iTRAQ reagents have reporter tags that generate ions for each sample with different mass/charge ratios of values: 114, 115, 116, and 117. These differences facilitate the discrimination of sample ions by MS analysis [7, 9].

Merits:

- Multiplexing of samples [7].
- As the iTRAQ reagents are isobaric, peptides labeled differently appear as a single peak avoiding the overlap of peaks [7].
- The use of internal standard peptides can allow absolute quantification [9].

Challenge:

The digestion with enzymes increases sample complexity [7].

Reagents are costly [9].

2.16.1. Case Studies Using iTRAQ

A study in 2007 applied this technique to study the extent of tyrosine phosphorylation in breast cancer samples [32]. There were novel findings of progression of disease with the phosphorylation of TOLLIP (Toll interacting protein) as well as SLC4A7 that is a co-transporter of sodium bicarbonate.

Advances of iTRAQ:

As the process of SCX is taxing, the use of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) was developed. The use of such interactions of hydrophilic interactions as well as electrostatic repulsion aided in more coverage of proteome [33].

There are Eight-plex iTRAQ reagents that can allow a simultaneous run of 8 samples [9].

2.17. CE-MS

The technique offers high resolution as samples are subjected to migration across a capillary column filled with buffer. The analysis of separated peptides is facilitated by the application of either ESI: an online approach or MALDI: an offline approach. The former is pitched as a more favorable method due to interference from matrix and a sip in resolution of MALDI.

Merits:

- It favors analysis of molecules of low molecular weight: less than 20 kDa.
- It is sensitive.
- The technique is rapid.
- It is not hindered by compounds that interfere with LC.

Demerits:

- Sample loading is less.
- Reproducibility is compromised

(Reviewed by Kalantri et al., [13])

2.18. STABLE ISOTOPE DILUTION MASS SPECTROMETRY (SID-MS)

This technique is based on an absolute measurement of proteins. Amounts of known concentrations of isotope-labeled standards are used that have similar properties on a chromatograph as targets while the mass/charge ratio is different. The target proteins are analyzed for classic signatures that are then spiked with isotope-labeled standards. The area of peak of the chromatogram yields the amounts of samples.

Merits:

- The technique is precise.
- It yields linear data.

Challenges:

- Internal standards are to be arranged for that restricts target peptides.
- Triple quadrupole mass spectrometers are desired.

The technique has been used to analyze for biomarkers in cancer of prostate and pancreas while serving as a bridge between discovery and validation of biomarkers.

(Reviewed by Paul et al, 2013; 9).

2.19. MICROARRAYS

The use of DNA microarrays spawned the development of arrays of proteins to develop protein microarrays. They can facilitate the quest for a biomarker or several of them or even a proteome [2]. Protein microarrays involve a membrane or slide on which proteins are immobilized to bind ligands of choice [34]. (Figure 2.12).

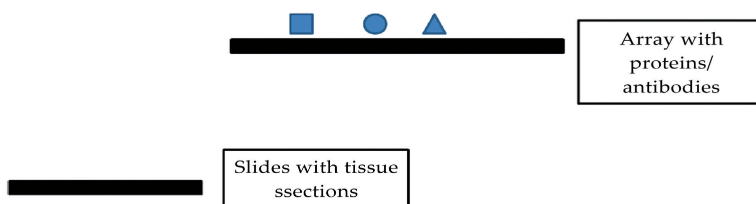


Figure 2.12. Microarrays for proteomics biomarkers.

The types of these microarrays include:

- Expression-based microarrays:

The expression of proteins in a sample is the target of this array. These can either be forward or reverse. In *forward phase*, several spotted molecules are coated on a solid surface. These baits could be antibodies or aptamers or even lysates to serve as baits for several protein samples [2].

The captured proteins can be detected by a *direct approach* by the use of labeled proteins before they are applied on the coated slide. This facilitates a comparison of expression of proteins between samples such as a screening done in 2003 for prostate cancer markers using sera [35].

Alternatively a *sandwich immunoassay* or indirect labeling can involve the use of a second antibody to target another domain. Despite the necessity of a second antibody, the technique is more in specificity and sensitivity. A 2003 study measured the level of receptor tyrosine kinase in cancer cells [36].

Reverse-phase protein microarrays involve analytes coated on a slide instead of baits at various concentrations that can aid in quantification. A 2003 study showed that Akt pathway played a key role in viability using these reverse phase systems [37] in ovarian and prostate cancer.

- Function-based microarrays:

This involves assignment of functions to proteins using measurement of products formed [2].

- Interaction-based microarrays

These are aimed at the interactions of molecules such as proteins association with other molecules [2].

Merits of microarrays:

- High-throughput [13].
- Sensitive [13].
- It is robust to analyze a global pattern on a single slide [7].

- The study of proteins can be benefitted by the use of nucleic acid programmable protein array (NAPPA) where a microarray is produced from cDNA molecules that have been captured [7].
- They are being addressed as the next “big thing” in proteomics as they can analyze low abundant biomarkers [7].

Challenges of microarrays:

- Specificity is an issue [13].
- Posttranslational modifications cannot be detected [13].

2.19.1. Antibody Microarray

The field of discovery of biomarkers can benefit from the use of immunoassays that can quantify the level of proteins. The substrates used can range from gels to beads to nylon and plastic. In direct labeling, an antibody micro array is treated with the sample proteins while in sandwich immunoassay proteins are loaded on a microarray and are detected with a second antibody [7].

2.20. MULTIDIMENSIONAL PROTEIN IDENTIFICATION TECHNOLOGY (MUDPIT)

The technique Multidimensional Protein Identification Technology (MudPIT) aims to circumvent the use of gels. The samples containing proteins are subject to proteolysis and then 2-dimensional liquid chromatography. The separated samples are then subject to tandem mass spectrometry. The use of appropriate computational programs can facilitate the dechipering of amino acid sequence [38].

The samples can be subjected to several formats of chromatography such as strong cation exchange (SCX) reversed phase (RP) high performance liquid chromatography (HPLC) if operated in a biphasic format. In case of a triphasic format; another RP is applied before SFX (7,39). (Figure 2.13)

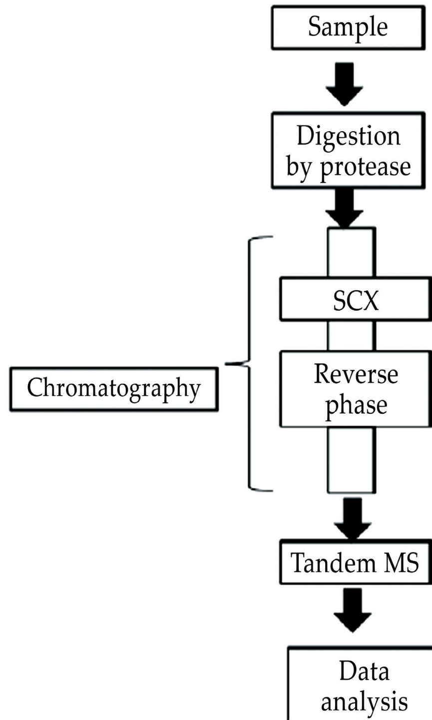


Figure 2.13. Scheme of MudPIT.

Merits:

- It is sensitive and reproducible [7].
- The proteins generated are exhaustive [7].
- Post-translational modifications can be studied using the correct design of experiment [38].

Challenges:

It is "bottoms up" approach that can yield several outputs that are challenging to analyze [39].

2.21. LABEL-FREE TECHNIQUES

The techniques discussed so far entailed the use of labeling reagents that require higher costs as well as higher sample levels and the potential of improper labeling. This has spawned the development of *shotgun proteomic*

technology that does not employ labels. This is pitched in as a label-free and high throughput for discovery of biomarkers.

The technology involves the basis of the concentration of a molecule under a peak of a chromatogram is proportional to the area under a peak. Following measurements of peak area and height of the chromatogram, data from tandem MS is counted for the spectrum. These techniques employ software such as Protein Lynx of Waters and Decyder MS of GE Healthcare. Initial methods of quantification of proteins used protein abundance index (PAI) and its later development called exponentially modified PAI (emPAI) that could estimate an absolute value.

Merits:

- Simple approach
- Less expensive

Challenges:

- If proteins possess similar peptides, the data may be redundant.
- The analysis of proteins that are less abundant is challenging.

Example of use:

A 2011 study evaluated the use of the label free technique in hepatocellular carcinoma. The application of label-free 2-D LC-MALDI MS aided in identifying identify N-glycoproteins as biomarkers [40]

2.22. ANOTHER PROTEOMICS FOR BIOMARKER DISCOVERY

Selected reaction monitoring (SRM):

This is tandem mass spectrometry that can analyze several targets from a digested protein. The use of triple quadrupole mass spectrometer as well as LTQ ion trap mass spectrometer in the tandem array can cast light on the signal strength. An added advantage is the absence of any immunological agent.

Accurate mass and time (AMT) tag:

This approach involves the application of LC-MS/MS on samples that have been fractionated. Each fraction can be subjected to an analysis that can consume time [7].

2.22.1. A Quick Round Up

Gel-based proteomics platforms for biomarker discovery

A combination of two dimensional gel electrophoresis (2-DE) and MScan target several proteins that are differentially expressed. There are 2 main steps involved: first isoelectric focusing (IEF) to separate the proteins on the basis of charge followed bySDS-PAGE to separate the proteins on the basis of molecular mass. The separation of 10,000 proteins has been achieved by two dimensional gel electrophoresis [41].

A second technique is Difference gel electrophoresis (DIGE) that involves the use of Cy2, Cy3, and Cy5: that are cyanine fluorescent dyes for labeling proteins. This is followed by pooling subsequent to the labeling of the standard, control and the test. The separation is then simultaneous; that facilitates lesser number of gels as variations that arise from the use of different gels. The sensitivity is 0.5 femto mol proteins but the main challenge is the cost [42].

Gel-free proteomics

This refers to “bottom-up” approach where peptides are derived naturally or through enzymes are run on tandem mass spectrometry (MS/MS). The major steps include:

- Ionization of peptide
- Separation of precursor ions on the basis of mass-to-charge ratio
- Fragmentation and analysis of resultant ions
- Analysis of data

Prior to sequencing; the peptides are labeled by several techniques:

- Isotope- coded affinity tags (iCAT) following digestion by trypsin.
- Quantification is facilitated by Isobaric tag: iTRAQ.
- SILAC involves the use of heavy amino acids.

The labels facilitate the samples to be differentiated and lower amounts of variations have been reported. A challenge with MS is false positives as there is a change to wrongly associate peptides. Such wrong identification can be corrected by the application of:

- decoy database;
- statistical models;
- scoring system;
- concatenated or chimeric database.

CE-MS refers to capillary electrophoresis coupled to mass spectrometry where proteomics is applied on proteins of low molecular weight. The separation of small proteins is achieved across an electric field on the basis of charge and size. There are several models of capillary coatings such bare fused silica capillary that is then followed by electrospray ionization (ESI) -MS. The coupling can be either sheathflow (large cohorts) or sheathless interface [4].

2.23. BIOINFORMATICS AND PROTEOMICS AND BIOMARKERS

2.23.1. Role of Bioinformatics in Proteomics

The vast repertoire of data generated from the field of proteomics can be benefitted by the application of bioinformatics. This can aid in generating matter that is of significance in the discovery of biomarkers in a relevant biological context.

Data management and mining:

This is a major part of high-throughput studies such as proteomics. Major databases that have aided the field of proteomics data are NCBI, SWISS-PROT and TrEMBL. The databases aid in the searching, storage and retrieval of data. There are also sequences from across species that can aid in analysis of a query.

In the context of proteomics, data mining requires that data generated by the technologies applied in proteomics are analyzed and interpreted. Tools of the databases mentioned above can be subjected to integration to generate appropriate outputs from generated proteomics data.

Predictive Bioinformatics in Proteomics:

A 2004 report highlighted the presence of close to 60% yet to be identified protein sequences that are a major challenge in such protein databases [43]. Here the role of predicative computational tools or predictive bioinformatics can aid the field of such yet to be annotated sequences. Several tools of databases such as SWISS-PROT can aid in:

- Localization of putative markers/ proteins.
- Functions can be deciphered before the wet lab tests.
- The structure and protein localization can be deciphered.

Pitfalls and overcoming them:

20–30% of genomes are constituted by membrane proteins [44], but only a fraction of such proteins have been studied. Proteins of the plasma membrane are pitched as novel markers and targets such as G-protein coupled receptors that can serve as novel targets. This necessitates the study of protein topology and protein structure and function. Based on the transmembrane segments, tools can predict protein topology. A challenge here is due to hydrophobic nature of such segments, there are possibilities of wrongly predicting N-terminal signal peptides.

To overcome such issues, a study in 2006 used five predictive computational methods that used different methods of prediction to analyze human proteome database. Another tool SignalP was used to avoid false positives as the tool can distinguish transmembrane segments and signal peptides. The study revealed 15% and 39% as the level of transmembrane segments with 13% as consensus [2, 43].

With the large quantity of data from such omics data [46], it is essential to look at the role of bioinformatics and computing.

The uses of bioinformatics can be summarized as follows:

- Novel resources that facilitate integration and analysis of data. Example: BioGrid.
- Classifying and organizing data that can facilitate the addition of new data. Example: Human Proteinpedia.
- Applying tools to generate matter in a biological context. Example: and Cytocape.

Few databases that serve as a reservoir of peptides/proteins are given Table 2.2.

Table 2.2. Databases for Proteomics

http://www.ebi.ac.uk/pride/	Proteomics Identifications Database
http://www.uniprot.org/	UniProt
http://prospector.ucsf.edu/prospector/mshome.htm	ProteinProspector
http://pepbank.mgh.harvard.edu/	PepBank
http://www.peptideatlas.org/	Peptide Atlas

There are several annotated databases that are useful in the field of proteomics that are summarized Table 2.3.

Table 2.3. Annotated Databases

Link	**Description**	**Role**
http://string.embl.de/	Retrieving Interacting Genes/Proteins	Protein-protein interactions
http://mint.bio.uniroma2.it/mint/	Molecular Interaction Database	
http://bio.informatics.iupui.edu/HAPPI/	Human Annotated and Predicted Protein Interaction Database	
http://www.hprd.org/	Human Protein Reference Database	
http://www.proteinlounge.com	Protein lounge	Signal Transduction
http://www.reactome.org	Reactome KnowledgeBase	
http://www.pantherdb.org	Protein ANalysis THrough Evolutionary Relationships	
http://www.biocarta.com/genes/index.asp	BioCarta Pathway Diagrams	
http://autosome.ru/HOCOMOCO/index.php	Homo Sapiens Comprehensive Model Collection (HOCOMOCO)	Pathways of regulation
http://www.pazar.info/	A Public Database of Transcription Factor and Regulatory Sequence Annotation	
http://wwwmgs.bionet.nsc.ru/mgs/gnw/trrd/	Transcription Regulatory Regions Database	

In addition to annotation, there are several computational tools that aid in the analysis of such generated data from proteomics that are given below:

- Reactome, BioCyc plugin, PathViso for mapping of pathways.
- Ingenuity Pathway Analysis, KEGG for analysis of pathways.
- MiMI, Bisogenet, iRefScape, PanGIA for mapping interactomes.
- KUPNetViz, GeneMania, NetworkAnalyzer for analysis of networks [4].

2.24. GLIMPSES OF DISCOVERY OF BIOMARKERS USING PROTEOMICS

Though there are advancements in the field of MS, reports indicate a covering of a portion of molecules. The presence of proteins at higher concentrations in a fluid can hinder the detection of several potential biomarkers that are present at lower concentrations. This requires the application of several methods to increase the scanning of low level protein candidates such as:

Reduction of sample complexity:

This involves the enrichment for desired proteins such as those of glycoproteins, phosphorylated, nuclear matrix and membrane. Immuno depletion can be applied to samples like plasma to lower the levels of high level proteins such as albumin or IgG.

A 2010 study applied affinity chromatography in the form of Multiple Affinity Removal System, Agilent Technologies, Inc. to plasma samples. The use of this fractionation aided in the elimination of 7 to 14 common proteins in the plasma. The use EF-LC-MS/ MS aided in the enriched 23 proteins present at low levels that comprised 5-6% of the biomarkers [47].

Few challenges include the removal of only abundant proteins, removal of proteins of interest and reproducibility.

A technique in 2013used an amalgamation of1D-PAGE, pIEF (peptide isoelectric focusing) and RP-HPLC to analyze the proteome at deeper level. The extract of nucleus of HeLa cells was subjected to a combination of 3 techniques against each of the technique applied singly. The numbers of peptides reported for only pIEF-LCMS/ MS: 31113 peptides and 3945 proteins against all the combined techniques 56228 peptides and 5260 proteins. Thus, a combination technique was useful to detect more novel peptides [48]. A combination of mix-bed ion-exchange chromatography was applied to MudPIT (multidimensional-protein identification technology) analysis of mammary tumor samples. The use of only the MudPIT revealed 1292 proteins while the integrated approach yielded 3084 proteins [49].

2.25. CASE STUDIES: BIOMARKERS OF BLADDER CANCER USING PROTEOMICS

To highlight the achievements of using proteomics in the field of proteomics the example of bladder cancer shall be used. Looking at the genitourinary

system, this cancer ranks second in occurrence as well as fatalities. Along with the lowered survival, diagnosis requires the use of invasive cystoscopy or cytology that is not invasive but is not sensitive. The use of proteomics has reported several potential biomarkers that can cast light on this disease.

A 2007 study involved differential gel electrophoresis of controls and samples that were positive for bladder cancer. MALDI-TOF MS revealed certain proteins that were expressed differently such as cytokeratin 1, Regenerative protein (Reg-1), prefoldin. The levels of cytokeratin 1 was found to be lesser in samples of bladder cancer [50].

A study in 2013 analyzed the proteome of bladder cancer samples using 2-DE that showed that cancer cells showed higher levels of cofilin. Immunohistochemistry and Western Blot showed the higher levels of cofilin in the cancer samples. An added study with phosphoylated Ser-3 of cofilin antibody showed more levels of phosphorylation of the cofilin in cancer samples. This was suggestive of cofilin as a biomarker whose presence and phosphorylation can be indicative of cancer [51].

A study in 2012 explored studied proteomics for biomarkers in controls as well as bladder carcinoma urothelium using iTRAQ labeling. Several proteins were detected of which 15 proteins such as B-cell receptor-associated protein 31, FK506 binding protein, DDX39 were present at higher levels in the cancer samples. Of these the levels of DDX39 was found to be lowered in cancer samples especially with progress in the grade of cancer. The application of siRNA technology involved transfection of T24 bladder cancer cell line with si-DDX39. There was an increase in the ability to invade that is suggestive of the role of this DDX39 as a biomarker [4, 52].

2.25.1. Tissue Culture/Proteomics and Biomarkers

Despite the usage of patient samples for the diagnosis of diseases or biomarkers; the use of cell lines or tumor lines can facilitate the analysis of biomarkers. Such cell lines have proteins that are secreted apart from the ones in the cell. The secreted proteins are known as secretome can be a source of biomarkers. Apart from cells, the secreted cell liquid or cells are subjected to protein separation techniques discussed previously. The proteins can be digested by protease in solution or in the gel to be followed by MS. Biomarkers can be detected from MS data [9]. (Figure 2.14).

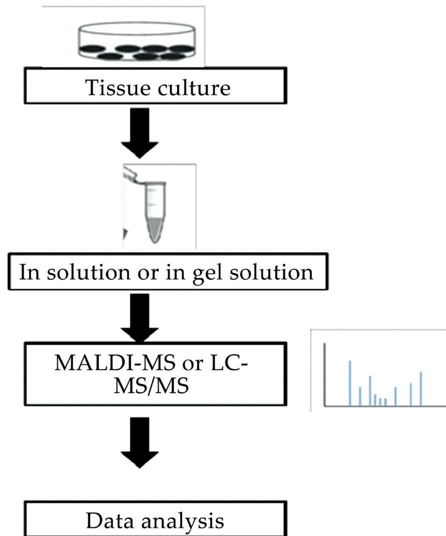


Figure 2.14. Biomarkers from tissue culture and proteomics.

2.26. FINAL NOTE

With several techniques of proteomics, biomarkers have seen discovery and the process is still ongoing with new reports emerging [53].

As previously seen, the analysis of several samples such as serum/plasma possess a dynamic proteome that offered several challenges [54]. The development of several techniques of fractionation as well as technologies such as mass spectrometers with increased resolution and precision [55] as well as assays based on arrays [56] are contributing to this field.

Proteomics is considered as offering more information in unveiling the field of biomarkers [53].

MS:

The discovery of biomarkers is termed as ideal when MS is applied with data dependent acquisition (DDA) [53]. While a major challenge includes preparation of sample; another challenge is the lengthy process of LC-MS/MS analysis of a single sample. In the case of proteins that are less abundant, the reproducibility and quantification are challenges [5].

The year 2012 saw the selection of MS as the technique for studying proteins [57].

These challenges require the application of data-independent acquisition approach. Here the discovery is facilitated by the application of wide proteome profiling without any targets to then be subjected to targeted MS. In the absence of antibodies or poor quality antibodies, MS is presented as a method of choice.

The use of MS to target several biomarkers in more than one sample yet needs optimization in the fields of:

- Linear calibration curves.
- An increase in range.
- Selecting the best ionized peptides.
- Calibration of amounts of internal standard.

These were attempted to be addressed by the development of termed Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH) strategy that aided in the identification of 15,000 peptides [58].

2.26.1. Affinity Proteome Profiling

Several sensitive techniques have been developed that could target a dynamic range with little volume of sample. The use of arrays of Antibody bead does not require the removal of highly abundant proteins such as albumin. However, challenges include possible cross reactivity of antibodies as well as preparation of sample.

An assay for bioanalysis requires a framework that aids its validity in a technical context. MS follows regulatory guidelines in terms of accuracy and coefficient of variations (CVs) [59]. In order to further qualification by regulations, the US FDA with National Cancer Institute and Clinical Proteomic Tumor Analysis Consortium aim at the establishment of a document to guide targeted MS covering:

- Stability of test.
- Reproducibility of test.
- Shelf time.
- Free thaw.
- Variability: intra and inter.

The use of internal standards as stable isotope-labeled proteins of full length before digestion was tested in a study published in 2013 that was found to estimate dystrophin in human muscle samples with high reproducibility [60].

In a nutshell, there are several techniques available for the analysis of biomarkers in proteomics. The use of technical advancements can facilitate a drop in costs as well as more technicalities [53].

REFERENCES

1. Drews, J. (2000). Drug discovery: a historical perspective. *Science*. *287,* 1960–1964.
2. Ahram, M., Petricoin, E. F. (2008). Proteomics Discovery of Disease Biomarkers. *Biomarker Insights*. *3,* 325–333.
3. Vlahou, A. (2013). Network views for personalized medicine. *Proteomics Clinical Applications*. *7,* 384–387.
4. Frantzi, M., Bhat, A., & Latosinska A. (2014). Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker development. *Clinical and Translational Medicine*. *3,* 7.
5. Tabb, D. L., Vega-Montoto, L., Rudnick, P. A., et al. (2010). Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *Journal of Proteome Research*. *9,* 761–776.
6. Frangogiannis, N. G. (2012). Biomarkers: hopes and challenges in the path from discovery to clinical practice. *Translational Research*. *159,* 197–204.
7. Chandramouli, K., & Qian P. Y. (2009). Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity. *Human Genomics and Proteomics*. 239204.
8. Domon, B., & Aebersold, R. (2006). Mass spectrometry and protein analysis. *Science*. *312*(5771), 212–217.
9. Paul, D., Kumar, A., Gajbhiye, A., Santra, M. K., & Srikanth, S. (2013). Mass Spectrometry-Based Proteomics in Molecular Diagnostics: Discovery of Cancer Biomarkers Using Tissue Culture. *Biomed Res Int. 2013:78313110.*
10. Mann, M., Hendrickson, R. C., & Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. Annual Review of Biochemistry. 70, 437–473.
11. Parasuraman, S., Anish, R., Balamurugan, S., Muralidharan, S., Kumar, K. J., & Vijayan, V. (2014). An Overview of Liquid Chromatography-*Mass Spectroscopy Instrumentation*. *5*(2), 47–55.
12. Belov, M. E., Gorshkov, M. V., Udseth, H. R., Anderson, G. A., & Smith, R. D. (2000). Zeptomole-sensitivity electrospray ionization--Fourier transform ion cyclotron resonance mass spectrometry of proteins. *Analytical Chemistry*. *72,* 2271–2279.

13. Kalantari, S., Jafari, A., Moradpoor, R., Ghasemi, E., & Khalkhal E. (2015). Human Urine Proteomics: Analytical Techniques and Clinical Applications in Renal Diseases. *Int J Proteomics. 2015:782798.*
14. Jia, J., & Zhang, L. (2012) Advance in proteomics research and application. *Journal of Animal and Veterinary Advances*. *11*(20), 3812–3817.
15. Petricoin, E. F., Ornstein, D. K., Paweletz, C. P., Ardekani, A., Hackett, P. S., Hitt, B. A., Velassco, A., Trucco, C., Wiegand, L., Wood, K., Simone, C. B., Levine, P. J., Linehan, W. M., Emmert-Buck, M. R., Steinberg, S. M., Kohn, E. C., & Liotta, L. A. (2002). Serum proteomics patterns for detection of prostate cancer. *Journal of National Cancer Institute*. *94,* 1576–8.
16. Irani, D. N., Anderson, C., Gundry, R., Cotter, R., Moore, S., Kerr, D. A., McArthur, J. C., Sacktor, N., Pardo, C. A., Jones, M., Calabresi, P. A., & Nath A. (2006). Cleavage of cystatin C in the cerebrospinal fluid of patients with multiple sclerosis. *Annals of Neurology*. *59,* 237–47.
17. Neverova, I., & Van Eyk, J. E. (2005). Role of chromatographic techniques in proteomic analysis. Journal of Chromatography B: *Analytical Technologies in the Biomedical and Life Sciences*.*815*(1–2), 51–63.
18. Kalantari, S., Rutishauser, D., Samavat, S., et al. (2013). Urinary prognostic biomarkers and classification of IgA nephropathy by high resolution mass spectrometry coupled with liquid chromatography. *PLoS ONE*. *8*(12) doi: 10.1371/journal.pone.0080830.e80830.
19. Johnson, R. S., Davis, M. T., Taylor, J. A., & Patterson, S. D. (2005). Informatics for protein identification by mass spectrometry. *Methods*. *35,* 223–36.
20. Meleady, P. (2011). 2D gel electrophoresis and mass spectrometry identification and analysis of proteins. *Methods in Molecular Biology.784,* 123–137.
21. Braun, M., Fountoulakis, M., Papadopoulou A et al. (2009). Down-regulation of microfilamental network-associated proteins in leukocytes of breast cancer patients: potential application to predictive diagnosis. *Cancer Genomics and Proteomics*. *6* (1), 31–40.
22. Cancemi, P., Di Cara, G., Albanese NN et al. (2010). Large-scale proteomic identification of S100 proteins in breast cancer tissues. *BMC Cancer.10,* 476.

23. Yan, J. X., Devenish, A. T., Wait, R., Stone, T., Lewis, S., & Fowler, S. Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of Escherichia coli. *Proteomics*. *2* (12), 1682–1698.
24. Zhang, J., Song, M. Q., Zhu JS et al. (2011). Identification of differentially-expressed proteins between early submucosal non-invasive and invasive colorectal cancer using 2D-dige and mass spectrometry. *International Journal of Immunopathology and Pharmacology*. *24*(4), 849–859.
25. Banerjee, H. N., Mahaffey, K., Riddick, E., Banerjee, A., Bhowmik, N., & Patra, M. (2012). Search for a diagnostic/prognostic biomarker for the brain cancer glioblastoma multiforme by 2D-DIGE-MS technique. *Molecular and Cellular Biochemistry*. *367*(1-2), 59–63.
26. Mann, M. (2006). Functional and quantitative proteomics using SILAC. *Nature Reviews Molecular Cell Biology*. *7*(12), 952–958.
27. Ong, S. E., Blagoev, B., Kratchmarova I et al. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular and Cellular Proteomics*. 1(5), 376–386.
28. Geiger, T., Madden, S. F., Gallagher, W. M., Cox, J., & Mann, M. (2012). Proteomic portrait of human breast cancer progression identifies novel prognostic markers.*Cancer Research*. *72* (9), 2428–2439.
29. Krüger, M., Moser, M., Ussar, S., et al. (2008). SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function.*Cell*. *134*(2), 353–364.
30. Boersema, P. J., Geiger, T., Wiśniewski, J. R., & Mann, M. (2013). Quantification of the N-glycosylated secretome by super-SILAC during breast cancer progression and in human blood samples. *Molecular and Cellular Proteomics*. *12*(1), 158–171.
31. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., & Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*. *17*(10).994–999.
32. Chen, Y., Choong, L. Y., Lin, Q., et al. (2007). Differential expression of novel tyrosine kinase substrates during breast cancer development. *Molecular and Cellular Proteomics*. *6*(12), 2072–2087.
33. Hao, P., Guo, T., & Li X et al. (2010). Novel application of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) in shotgun

proteomics: comprehensive profiling of rat kidney proteome *Journal of Proteome Research*. *9*(7), 3520–3526.

34. Hu, S., Loo J, A., & Wong, D. T. (2006). Human body fluid proteome analysis. *Proteomics. 6*(23), 6326–6353.
35. Miller, J. C., Zhou, H., Kwekel, J., Cavallo, R., Burke, J., Butler, E. B., The, B. S., & Haab, B. B. (2003). Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. *Proteomics*. *3,* 56–63.
36. Nielsen, U. B., Cardone, M. H., Sinskey, A. J., MacBeath, G., & Sorger, P. K. (2003). Profiling receptor tyrosine kinase activation by using Ab microarrays. *Proc Natl Acad Sci USA*. *100,* 9330–9335.
37. Wulfkuhle, J. D., Aquino, J. A., Calvert, V. S., Fishman, D. A., Coukos, G., Liotta, L. A., & Petricoin, E. F. (2003). Signal pathway profiling of ovarian cancer from human tissue specimens using reverse-phase protein microarrays. *Proteomics*. *3,* 2085–2089.
38. Schirmer, E. C., Yates, J.R. 3rd, Gerace, L., & Mud, PIT. (2003). A powerful proteomics tool for discovery.*Discovery Medicine*. *3*(18), 38–39.
39. Berri Trapero, Yumana & Villar Moreno, Rafael (accessed 2015). *Multidimensional Protein Identification Technology.* Department of molecular biology. Course: Fundamentals of Advanced Molecular Biotechnology.
40. Ishihara, T., Fukuda, I., & Morita, A. et al. (2011). Development of quantitative plasma N-glycoproteomics using label-free 2-D LC-MALDI MS and its applicability for biomarker discovery in hepatocellular carcinoma. *Journal of Proteomics*. *74*(10), 2159–2168.
41. Marcus, K., Joppich, C., May, C., Pfeiffer, K., Sitek, B., Meyer, H., & Stuehler, K. (2009). High resolution 2DE. *Methods Molecular Biology*. *519,* 221–240.
42. Minden, J. (2007). Comparative proteomics and difference gel electrophoresis. *Biotechniques*. *43,* 739–741.
43. Ahram, M., & Springer, D. L. (2004). Large-scale proteomics analysis of membrane proteins. *Expert Review in Proteomics*. *1,* 293–302.
44. Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Moleculer Biology*. *305,* 567–80.

45. Ahram, M., Zhou, L., Fang, R., & Al-Tawalbeh, G. (2006). Estimation of membrane proteins in the human proteome. *In Silico Biology*. *6,* 0036
46. Mayer, P., Mayer, B., & Mayer, G. (2012). Systems biology: building a useful model from multiple markers and profiles. *Nephrology Dialysis Transplantation*. *27,* 3995–4002.
47. Tu, C., Rudnick, P. A., Martinez, M. Y., Cheek, K. L., Stein, S. E., Slebos, R. J., & Liebler, D. C. (2010). Depletion of abundant plasma proteins and limitations of plasma proteomics. *Journal of Proteome Research*. *9,* 4982–4991.
48. Atanassov, I., & Urlaub, H. (2013). Increased proteome coverage by combining PAGE and peptide isoelectric focusing: comparative study of gel-based separation approaches. *Proteomics*. *13,* 2947–2955.
49. Zhang, L., Yao, L., Zhang, Y., Xue, T., Dai, G., Chen, K., Hu, X., & Xu, L. X. (2012). Protein pre-fractionation with a mixed-bed ion exchange column in 3D LC-MS/ MS proteome analysis. *Journal of chromatography. B, Analytical technologies in the biomedical and life* sciences. *905,* 96–104.
50. Orenes-Pinero, E., Corton, M., Gonzalez-Peramato, P., Algaba, F., Casal, I., Serrano, A., & Sanchez-Carbayo, M. (2007). Searching urinary tumor markers for bladder cancer using a two-dimensional differential gel electrophoresis (2D-DIGE) approach. *Journal of Proteome Research*. *6,* 4440–4448.
51. Chung, H., Kim, B., Jung, S. H., Won, K. J., Jiang, X., Lee, C. K., Lim, S. D., Yang, S. K., Song, K. H., & Kim, H. S. (2013). Does phosphorylation of cofilin affect the progression of human bladder cancer? *BMC Cancer*. *13,* 45.
52. Kato, M., Wei, M., Yamano, S., Kakehashi, A., Tamada, S., Nakatani, T., & Wanibuchi H (2012) DDX39 acts as a suppressor of invasion for bladder cancer. *Cancer Science*. *103,* 1363–1369.
53. Hathout, Y. (2015). Proteomic methods for biomarker discovery and validation. Are we there yet? *Expert Review of Proteomics*. *12*(4), 329–331.
54. Anderson, N. L., & Anderson, N. G. (2002). The human plasma proteome: history, character, and diagnostic prospects. *Molecular Cell Proteomics*. *1*(11), 845–867.

55. Pernemalm, M., & Lehtio, J. (2014). Mass spectrometry-based plasma proteomics: state of the art and future outlook. *Expert Review in Proteomics*. *11*(4), 431–448.
56. Ayoglu, B., Haggmark, A., Neiman, M., et al. (2011). Systematic antibody and antigen-based proteomic profiling with microarrays. *Expert Review in Molecular Diagnostics*. 11(2), 219–234.
57. Marx, V. (2013). Targeted proteomics. *Nature Methods*. 10(1), 19–22.
58. Selevsek, N., Chang, C. Y., & Gillet, L. C., et al. (2015). Reproducible and Consistent Quantification of the Saccharomyces cerevisiae Proteome by SWATH-mass spectrometry. *Molecular Cell Proteomics*. *14*(3), 739–749.
59. Addona, T. A., Abbatiello, S. E., Schilling, B., et al. (2009). Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nature Biotechnology*. *227*(7), 633–641.
60. Brown KJ, Marathi R, Fiorillo AA, Ciccimaro EF, Sharma S, Rowlands DS, Rayavarapu S, Nagaraju K, Hoffman EP, Hathout Y. (201). Accurate quantitation of dystrophin protein in human skeletal muscle using mass spectrometry. *J Bioanal Biomed. 2012 Dec 18;Suppl 7.*

CHAPTER 3

MICROBIAL BIOMARKERS PROTEOMICS

CONTENTS

3.1. Typing of Microbes 80
3.2. Typing Methods 81
3.3. Phenotypic Methods 81
3.4. Genotypic Methods 85
3.5. Diagnosis of Microbial Infections Using Biomarkers 88
3.6. What Is The Role of A Biomarker From The Host? 88
3.7. Biomarkers And Antibiotics: A Scenario 90
3.8. Biomarkers That Are Routinely Analyzed 90
3.9. Biomarkers For The Future 91
3.10. What's New? 93
3.11. Proteome For Biomarkers of Bacteria And Virus 94
3.12. Host–Pathogen Protein–Protein Interactions 96
3.13. Intact Proteins 98
3.14. Direct Interactions And Their Detections 100
3.15. Posttranslational Modifications That Are Regulated By Pathogen ... 100
3.16. Summary 102
References 103

3.1. TYPING OF MICROBES

Microbial typing is primarily cone to assess relation between microbial isolates. A comprehension of the nature of relationship between microbes can aid in the details of infections to design measures that can control the outbreak [1, 2]. The effectiveness of surveillance systems has greatly been enhanced by bacterial typing, which has also provided clues to public health control measures.

The epidemiology of infectious diseases have been studied using techniques such as serotyping or bio typing, phage typing as well as antibiogram, bio typing, serotyping. For some worldwide diffuse pathogens, serotyping, phage typing and antibiotic resistance patterns have historically provided data to be used for short-term epidemiological studies, assessment of epidemiological trends in well-defined geographical areas and comparison between different countries [1, 2]. However, these methods are of limited practical value in epidemiological investigations because they are too time-consuming, variable and labor intensive [2]. There are highly fastidious bacteria that cannot routinely be confirmed by cultured and result in infections that are communicable diseases result from infection by [1]. The diagnosis of certain diseases such as bacillary angiomatosis or known cat scratch disease is expensive and also very challenging [3]. As a result, DNA-based typing methods have emerged as important tools to study most microbial pathogen epidemiology. There are several methods for typing such as molecular tools.

The molecular typing methods most commonly used are the DNA-based, such as plasmid profiling, restriction endonuclease analysis of genomic and plasmid DNA, chromosomal DNA profiling using either pulse-field gel electrophoresis or polymerase chain reaction (PCR)-based methods, and southern hybridization analysis with the use of specific DNA probes [1, 3, 4]. A large number of microbial targets can be simultaneously assessed using the microarray technology [2, 3]. Whole genome sequencing (WGS) approach is another revolutionary approach (3,4).

There are a slew of protocols available that require a decision regarding which protocol can utilized that depends on conditions such as the reproducibility or lucid results or rapidity as well as the transport of the test across labs [5]. Finally, a specific method is favored over others on the basis of the suitability of the test to answer a specific epidemiological question.

3.2. TYPING METHODS

Typing is a term applied to the distinguishing bacterial strains on the criteria of genotype and phenotype. These typing methods are useful in understanding the pathogenesis of infection, hospital infection control, and epidemiological studies.

The application of typing in a clinical setting entails:

- determine whether the epidemiology is a result of a single infecting strain or multiple contaminants.
- determination of the nature of an infection: whether it is a relapse occurring as one strain or a cocktail of strains.

The typing systems can be evaluated as [6]:

- The test should be functional on a repertoire of microbes.
- For each analysis, the test should have clear results. Those isolates that give either a null or an uninterpretable result are called the Non-typeable isolates.
- Distinguishing between strains should be performed by the typing test.
- The results should be lucid and favorable to analyze.
- The results should not vary with time or repetition.

The typing methods are bifurcated as given in the following sections [7].

3.3. PHENOTYPIC METHODS

Phenotypic methods analyze properties such as size, staining, shape or antigens as well as specific biochemistry without the application of genomic contexts.

Challenges in this technique are:

- Any alterations in microbial gene expression, either spontaneously or in response to environmental stimuli.
- Cells may have distinct phenotypes/ phenotypic properties, which were genetically indistinguishable, if grown under variable conditions.
- A phenotype may arise out of aberrations such as mutations
- Inability to type many species.

3.3.1. Biotyping

The identification of microbes is on the basis of particular pathways of metabolism of an organism. Using manual or automated systems, an organism is studied for its morphology and habitat.

Examples of biotyping methods include:

- sugar fermentation tests
- pH range
- chemical tests
- amino acid reactions such as deamination
- agglutination of blood cells
- hydrolysis
- reactions with RBCs.

Advantages:

Include reproducible techniques which are relative ease of performance and interpretation.

Disadvantages:

Shits or aberrations such as mutations or altered gene expression can result in a difference in the tests of similar species.

3.3.2. Phage Typing

A standard set of bacteriophages are used to characterize strains on the basis of their pattern of resistance or susceptibility to these phages. The test exploits the binding of receptros and bacterial cell wall receptors. Isolates of *Salmonella* sps and *Staphylococcus aureus* are especially typed by this technique, and are referred as ‘phage types.’

Advantages:

- fairly reproducible
- discriminatory
- Easily interpreted.

Disadvantages:

- Many strains cannot be typed.
- Due to the test requirement of live phages, only few organizations possess the system.

3.3.3. Bacteriocine Typing

The term bacteriocine refers to peptides that are synthesized by a bacterial species that targets another species. This approach can test for bacteria such as *Yersinia pestis, E.coli* as well as *Pseudomonas aeruginosa.*

Advantages:

- fairly reproducible,
- discriminatory
- Easily interpreted.

Disadvantages:

- This technique is available only at specific centers because it is demanding.
- Many strains cannot be typed.

3.3.4. Serotyping

Strains of same species can differentially express certain antigenic determinants on their surface and are known as 'serotypes.' Antigenic variations may be exhibited by surface structures such as capsular polysaccharides, lipopolysaccharides, membrane proteins, fimbriae and flagella.

This approach involves application of techniques such as labeling with enzymes as well as fluorescence or agglutination of bacteria or latex.

Advantages:

- The test has good reproducibility
- can be easily interpreted
- easily performed

Disadvantages:

- There are challenges to serotype many strains.
- Some methods are technically demanding and depend heavily on good quality reagents.
- The reagents require arduous preparation.
- Additionally, the test is poorly discriminatory because of cross-reaction of antigens, large number of serotypes, and untypeable nature of some strains.

3.3.5. Antimicrobial Susceptibility Typing (Antibiogram)

The sensitivity of bacteria to different antibiotics is involved in this technique where resistance can be linked to outbreak of disease.

Advantages:

- Typing of most strains is allowed.
- Easy to perform
- Interpretation is simple
- Reproducibility is an advantage

Disadvantages:

Different strains may exhibit similar resistance pattern, thus, decreasing its distinguishing ability.

3.3.6. Protein Typing

The test relies on the protein profile of different strains of bacteria. There are several methods that can detect variations in the types and structures of the proteins expressed by bacteria. The technique involves extraction of the proteins or modified proteins followed by SDS-PAGE, staining and a comparison with other analysis.

Immuno blotting involves a transfer of products of electrophoresis on a nitrocellulose membrane and visualized using antibodies that are labeled with enzymes.

The epidemiological studies of *Clostridium* and *S. aureus* are frequently done using this method.

Advantages:

- Interpretation is not arduous.
- Good reproducibility

Disadvantages:

- Due to complexity in patterns, comparison across species is a challenge.
- Only few labs have access to the technique due to the apparatus and skill required.

3.3.7. Multilocus Enzyme Electrophoresis (MLEE)

The cell isolates are analyzed for a set of metabolic enzymes and their differences in the electrophoretic mobilities. Cell extracts are electrophoresed in starch gels and variations in the electrophoretic mobility, also referred to as 'electromorph,' are monitored. The differences typically reflect the charge of the protein that are altered due to amino acid substitution.

Advantages:

- Typing of most strains is allowed.
- Easy to perform
- Reproducibility is an advantage
- Interpretation is simple

Disadvantages:

- Only few labs have access to the technique due to the apparatus and skill required.
- The extent of discrimination is less.

3.4. GENOTYPIC METHODS

Genotypic techniques analyze the genetic material: DNA along with extrachromosomal material. These techniques are now routinely used in diagnostic laboratories. A main challenge is access only to few labs on account of the complex technique and high cost.

3.4.1. Plasmid Analysis

This involves the analysis of plasmids in a bacterial system.

Advantages:

- The method is fairly easy to interpret
- Typing of most strains is allowed.

Disadvantages:

- Samples without plasmids cannot be analyzed.
- The reproducible nature is affected as the migration on gels is different due to alternative forms of DNA such as linear or coiled.
- The technique is not favorable for cells with less than 3 plasmids.

3.4.2. Chromosomal DNA Restriction Endonuclease Analysis (REA)

Certain fixed sites of nucleotides are cleaved by restriction enzymes. The fragments generated are a characteristic of the sequence.

The microbial DNA is cut with chosen endonucleases that have frequent restriction sites on the DNA, to generate hundreds of ~0.5 to 50 kb long fragments. The fragments are then separated according to size on agarose gels by electrophoresis, stained and examined under UV light.

Advantages:

- The most important advantage is the all the strains can by typed
- The method is efficiently reproducible.

Disadvantages:

- The profile generated is very complex and consists of several bands that may not resolve or overlap, thus, making complicating the analysis.
- In addition, the plasmids also may be digested and lead to some pattern, further complicating the interpretation.

3.4.3. PFGE of Chromosomal DNA

The limitations of REA are dealt with in this technique. The change orientation of the electric field across a gel facilitates the resolution of large DNA.

Advantages:

- Interpretation is not difficult.
- Reproducibility

Disadvantages:

However, the process is costly, requires costly equipment and is labor intensive.

3.4.4. Southern Blot Analysis of RFLPs

This technique analyzes only certain specific products of restriction digestion. Following digestion of DNA by endonuclease the fragments are separated using electrophoresis. These separated fragments are then applied on a nitrocellulose membrane subject to detection by antibodies. These restriction fragment length polymorphisms (RFLPs) can reflect the detected fragments.

The digestion of highly conserved sequence such as 16S rRNA or 23S rRNA referred as ribotyping can indicate the evolutionary relationships.

If the sequences of organisms show less than 98% homology, they are classified as separate species and an identity of less than 93% can be classified as different genera.

Advantages:

- Typeable strains
- The technique is reproducible
- Easily interpreted

Disadvantages:

- The nature of probes dictates the nature of discrimination.
- The reagents are expensive
- The equipment requires skill

3.4.5. Nucleotide Sequence Analysis

The DNA (or RNA) nucleotide-base sequences enable genotype determination at highest precision. In the case of RNAs, they are often sequenced either by sequencing the DNA gene that gives rise to the RNA, or by converting the RNAs into DNA. It is possible to compare several isolates at one time by using PCR to amplify a known DNA segment and sequencing the amplified product.

Advantages:

- The results are reproducible
- Easily interpreted
- This technique can apply on all strains.

Disadvantages:

- The reagents are expensive.
- The equipment requires skill.

3.5. DIAGNOSIS OF MICROBIAL INFECTIONS USING BIOMARKERS

Biomarkers are becoming increasingly important tools within all areas of medicine. Potential applications of biomarkers in infectious diseases include distinguishing bacterial from nonbacterial infection, monitoring response to therapy, and predicting outcomes. Continued research into a number of noninvasive urinary, serologic, and genetic biomarkers will help clinicians with diagnosis, prognosis, and treatment.

3.6. WHAT IS THE ROLE OF A BIOMARKER FROM THE HOST?

The experts from U.S. National Institute of Health and the European Medicines Agency have issued regulatory definitions for biomarkers. A biomarker is an indicator of a metabolic or disease process that is measured in an objective context [8, 9].

There are two pathways in biomarkers:

- they permit diagnosis or prognosis.
- those used as a companion to treatment, to select patients who

may benefit from a specific therapy or used during follow-up of therapy as early predictors of efficacy or of treatment toxicity.

The ideal biomarker in infectious diseases:

Within the field of infectious diseases, a biomarker may be used for identifying a high-risk group or predisposing condition, as an aid to identification of the disease, or to direct therapy and stratify patients according to their specific risk factors, and/or as an aid to therapeutic management in order to avoid relapse of infection.

3.6.1. Acute Infections and the Role of Biomarkers

Biomarkers can help in assessing the severity of infection and making a decision for the best possible treatment approach including introducing or maintaining antibiotic therapy, and the site of care (i.e., hospital or ambulatory care, intensive or ward care).

While in the past few years, several potential biomarkers of infection have been described, the current trend is to use a combination of multiplex tests to measure several biomarkers simultaneously from a single biological sample.

For diagnostic purposes, the effectiveness of biomarkers is assessed for:

Specificity: probability of a negative test in unaffected patients,

Sensitivity: probability of a positive test among affected patients.

When assessing the operating characteristics of biomarkers, the following should be considered:

- The population characteristics which is under study and the non-infected, or the "control group."
- The terms of the gold standard used [10, 11].

3.6.2. The Limitations of Interpretations

When interpreting measurements of biomarker levels, several difficulties arise, especially in scenarios involving multiplex tests [12], thus increasing the volume of information generated. In the case of few biomarkers, precision can be compensated despite determination of threshold value.

The variability of measurements and subsequent interpretation may result from several factors and must be assessed and controlled for before providing an interpretation of the results:

- No standardization between different methods.
- Variations exist among members of a group or across groups.
- Factors that contribute before analysis (the storage tubes, transport media, time lapsed from sampling to analysis, etc.), and during analysis (reproducibility, precision, threshold of measurement, etc.).

3.7. BIOMARKERS AND ANTIBIOTICS: A SCENARIO

The study of serum of patients with disease has been studied to reveal close to 100 biomarkers [13–15]. However, a select make it to the market as the result interpretation is a challenge in this realm. There are limited numbers of biomarkers within the field of acute infection that are currently established or are of potential clinical interest.

3.8. BIOMARKERS THAT ARE ROUTINELY ANALYZED

Two biomarkers fulfill the above-mentioned selection criteria are:

C-Reactive protein (CRP):

A single, prospective, randomized, controlled trial performed in the 1990s in children is available [16]; other studies have compared an intervention group to historical controls [17, 18]. Although there are only few available studies confirming its usefulness, measurement of CRP is frequently done in children for determining and adjusting the therapy duration. However, studies indicate that the use of CRP cannot be recommended currently in adults as an aid to initiate or discontinue antibiotics, although in children, CRP may be used to help discontinue therapy, albeit, with limited evidence.

Procalcitonin (PCT):

This protein has been used to represent start or end of use of antibiotics in the sphere of controlled random trials in lower respiratory tract infections. Such results of meta analyzes are reported (19-21). Of these studies, four involved patients close to nine hundred in number in intensive care [22, 23]. There are reports of studies in children such as: 384 (age group of 1-36 months) with severe fever, 121 (neonates) with sepis in initial stages [24].

3.9. BIOMARKERS FOR THE FUTURE

The quest for biomarkers of prognosis and diagnosis is an incessant approach. A few of these markers appear to be promising in adults: Soluble urokinase-type Plasminogen receptor (suPAR), the soluble Triggering Receptor Expressed on Myeloid cells-1 (sTREM-1), Presepsin, and proadrenomedullin (ProADM) have given encouraging results.

Glimpse of the features:

- The sensitivity and specificity are satisfactory.
- The studies have been comprehensive.
- Accessible biomarkers.
- The use can be facilitated with increased studies in children.

3.9.1. sTREM-1

The TREM-1,a superfamily to which the sTREM-1 belongs, contributes to innate immunity and is a surface receptor of monocytes cells and mature polymorphonuclear. The expression of this receptor is high when bacterial and fungi are detected by phagocytic cells. This high expression causes an increased level of the soluble form in several fluids such as blood, broncho-alveolar lavage fluid and CSF.

The diagnostic and prognostic value of sTREM-1 have seen several clinical studies (25-27) (Table 3.1). The importance of this marker near the infection location such as CSF is more as compared to that of plasma.

3.9.2. suPAR

A common receptor of inflammation is soluble urokinase-type plasminogen activator receptor (suPAR) or CD87. The expression is constitutive to cells such as leucocytes and endothelium. However the molecule can be expressed in response to molecules such as Lipo Poly Saccharides (bacteria) as well as TNF α or IL-β (cytokines) and Vascular endothelial growth factor, Epidermal growth factor (growth factors).

The expression of suPAR is upregulated on epithelial cells, lymphocytes, smooth muscle cells and fibroblasts during the inflammatory and immune response, and tumor growth as well as metastatic tumor dissemination.

Measurements can be done by ELISA kits; and along with cytokines as a part of multiplex assays.

However, diagnostically, suPAR is of limited value. Its clinical value appears associated with its ability to identify patients at risk (Table 3.2) and might be of interest for the management of HIV patients receiving antiretroviral therapy [28], during the follow-up of patients who have nonpulmonary mycobacterial infection [29] and in children who have *Plasmodium falciparum* malaria [30]. The use of suPAR in antibiotic management in infected patients requires more studies (31-33)

Table 3.1. The use of Strem-1 in Acute Infections

Disease	Sample	Strem-1 importance
Meningitis	CSF	Diagnostic
Pneumonia	Plasma	
	BAL	
Sepsis, septic shock	Plasma	Prognostic
SIRS		

Table 3.2. Clinical use of suPAR in Acute Infections

Disease	Sample	suPAR clinical importance
Sepsis, septic shock	Plasma	Prognostic
		Diagnostic

A member of the CALC gene family, **pro-ADM Adrenomedullin (ADM)** is a 52-amino acids peptide that acts as cell proliferation, hormone regulation and embryogenesis mediator. ADM is synthesized by endothelial cells in order to maintain homeostasis and cause dilation of blood vessels. However, complete peptide are less stable than the pro-ADM, or Pro-hormone fragments. TRACE (Time-Resolved Amplified Cryptate Emission) can enable detection of the levels of these peptides. The immune response to bacterial or viral products stimulates proADM secretion.

The biomarker Pro-ADM is prognostically important (Table 3.3). This marker can aid in the severity of clinical pneumonia [34] as well as identification of ICU patients [35].

Table 3.3. Clinical use of pro-ADM in Acute Infections

Disease	**Sample**	**suPAR value clinically**
Pneumonia	Plasma	Prognostic
--	--	Diagnostic

3.9.2. Presepsin

It is a glycoprotein receptor, formerly also known as CD14, present on the surface of monocytes/macrophages. The receptor reacts with lipopolysaccharide (LPS) and LPS binding protein (LPB). This results in the initiation of toll-like receptor 4 (TLR4). The level can be assayed using chemi luminescence assay such as that offered by IngenW.

Presepsin is both sensitive and specific of the current 4 biomarkers analyzed, and may be helpful in differentiating SIRS from sepsis due to bacterial infection [36] (Table 3.4).

Table 3.4. Clinical use of Presepsin In Acute Infections

Disease	**Sample**	**Presepsin importance**
SIRS	Plasma	Diagnostic
Sepsis		
Sepsis, septic shock		Prognostic
SIRS		

Thus, these biomarkers may have role in future clinical developments and larger studies are required in this field to assess the impact of these biomarkers on the antibiotic therapy in patients.

3.10. WHAT'S NEW?

The latest discovery of potential candidate biomarkers includes **micro-RNAs** (miR). These are small molecules of nearly 20 nucleotides length, present in eukaryotic cells, and modulate posttranscriptional regulation, thereby acting as biologic regulators. They are ubiquitously found in kidney, liver, and lung. After binding to the cognate mRNA sequence, they regulate gene expression and their expression can be measured by quantitative PCR and RT-PCR.

The miR expression dysfunctions have been implicated in several human diseases (http://www.miR2Disease.org/), such as cardiomyopathy, various types of cancers ("oncomir"), or in central nervous system diseases.

The miR also play role in defense mechanisms against viral infections, to control the infections. A number of miR have been implicated in regulating viral mRNA such as hepatitis C virus, Epstein-Barr, herpes, and cytomegalovirus.

A role for miR has been suggested in bacterial infections, such as *M. tuberculosis*; this involves a regulation of virulence factors or a modulation of interactions of phagocytic cells with the bacterium or factors for resistance. The miR also plays a role in inflection of the inflammation attributed to infection with *H. pylori* [37], notably miR-155 [38].

It has been observed that the initially released miRNAs spectrum in blood and leucocytes of septic shock patients differs from that of control patients. The miR-150, miR-182, miR-342-5p; miR-150 three are most dysregulated and interfere with lymphocytic immune response development and can act as potential early diagnostic and/or prognostic markers [39].

The miR-223, miR-15a, miR-16, miR-122, miR-193*, and miR- 483-5p are other miRNAs may have a high probability of a poor outcome in septic shock patients [40].

However, before the use of these miRNAs for diagnosis, prognostication, or therapeutic decision, several more studies are needed to better understand their role in human biochemical and immunological processes.

Two main technological advances are in progress, including 1) the development of point-of-care testing, with the availability of miniaturized and portable machines, allowing rapid testing at the bedside, even for sophisticated measurements (e.g., flux cytometry), which have been confined to specialized laboratories up to recently; and 2) the development of new methods, including the analysis of gene expression (genomics), of ARN activation (transcriptome), of production of proteins (proteomics), of lipids (lipidomics), or of metabolites (metabolomics).

3.11. PROTEOME FOR BIOMARKERS OF BACTERIA AND VIRUS

The constant interaction between hosts and pathogens has been shaped as a result of millions of years of evolution and are one of the most intriguing aspects of life. These interactions provide information as to how the hosts

develop defense mechanisms against pathogens and how the pathogens overcome these defense mechanisms. The process of evolution has facilitated the co-existence of hosts and microbes certain pathogenic microbes induce fatal disorders. It is hence essential to understand the host–pathogen interactions as means to treat and prevent diseases that are infection-induced.

In the case of intracellular pathogens, host–pathogen interactions occur regularly at molecular level, throughout the time the pathogen is replicating in the host. While the replication strategies are different for different pathogens, their primary aim is to successfully propagate in the hot: to make an entry in the host cell, and use its cellular components for replication, and for spreading to the neighboring host cells. The replicative mechanisms of a pathogen enable the pathogen to circumvent immune functions of a host. (Figure 3.1).

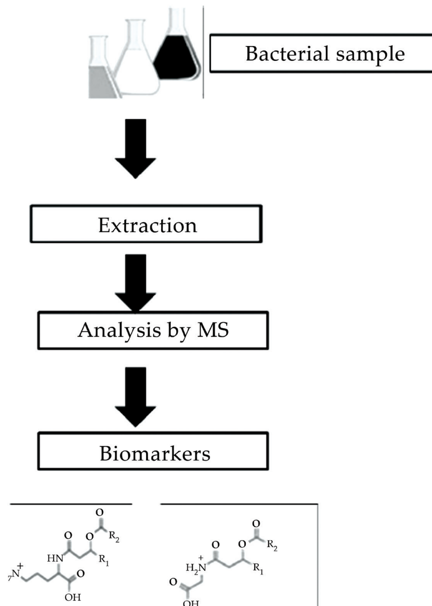


Figure 3.1. Schematics of analysis of microbial biomarkers.

The following questions still require solutions:

- The development of novel pathogens that dictate timely diagnosis as well as therapeutic intervention [41, 42].

- the emergence of drug-resistant microbes, and an urgent need for elucidation of host pathways that can be targeted and block the spread of pathogens.
- Several viral diseases are yet to see the development of drugs or vaccines [43].

The past decade has seen the emergence of ‘omic; approaches as pathbreaking tools for research and study of pathogen replication, host response, and disease progression pathways. The study of the protein components of biological systems, Proteomics, is being increasingly used for understanding the host–pathogen interactions [44]. The application of advanced tools aid in the sensitive analysis of such associations.

The amalgamation of proteomics with other methods, including other omic approaches, has increased the range of techniques to study pathogen infections. Proteomics can facilitate the study of:

- post-translational following infection.
- interaction of pathogens and hosts.
- changes in protein profiles following infection.

In addition, an integral part of proteomics are the bioinformatic tools that enhance several fold the ability to interpret large datasets.

3.12. HOST–PATHOGEN PROTEIN–PROTEIN INTERACTIONS

Once the pathogen enters the host cell, it must overcome host defense and reproduce to propagate. The normal host protein functions are either suppress or hijacked by the interaction of pathogen proteins with host proteins [44].

Protein-protein interactions (PPIs) are key in comprehending the process of infection apart from scouring for drug targets. The PPIs can be:

Direct: one protein physically interacting with another; or

Indirect: proteins interacting via other intermediate molecules.

The PPIs can provide mechanistic insights into the host–pathogen interaction network.

3.12.1. Immunoaffinity Purification with Mass Spectrometry (IP-MS)

This technique has been implemented the most in studies of host and pathogen [44]. In IP-MS, a protein of interest is isolated using either an antibody raised against the endogenous protein or by epitope-tagging the protein of interest and using an antibody against that epitope.

Following this, MS is used to identify the protein of interest and its co-isolated interacting proteins. An important advantage IP-MS in host -are that it enables unbiased detection of PPIs because the experiments can be performed in relevant cellular model systems [45].

The mechanisms involved in the replicative patterns of a virus along with alterations on the defense system of a host can be elucidated. The shifts in protein functions in a host can be characterized by IP-MS. The changes in host-pathogen can be studied through a combination of IP-MS with microscopy and fluorescence. Initially demonstrated for studying the RNA virus Sindbis [46], this approach was later applied to other viruses, such as the RNA virus respiratory syncytial virus [47] and the DNA viruses human cytomegalovirus (HCMV) and pseudorabies virus (PRV) [48, 49].

IP-MS studies have led to the discovery of numerous mechanisms through which HCMV modulates cellular processes, for which a vaccine or an effective antiviral treatment is still lacking), for example, activation of the mTOR pathway to suppress host stress response [50], inhibition of host sensing of viral DNA and immune signaling [51], or use of cellular trafficking pathways during maturation of infectious particles [48].

IPMS proved to be valuable, for example, in revealing that in influenza A viral RNA synthesis is repurposed by cellular nucleophosmin [52] and that the trafficking of PRV in neurons uses a host kinesin-3 motor [49]. Similarly, from the host perspective, IP-MS has helped to define mechanisms of cellular defense [53] and to distinguish protein domain-dependent interactions and functions for host antiviral factors [54]

Advantages:

In the case of viruses with very small genomes, it becomes very difficult to tag a viral protein with an epitope for purification, simultaneously keeping the virus replication competent. As a result, researchers use ectopic expression of tagged viral proteins beyond the infection to study probable viral-host PPIs that can be confirmed in vivo [55].

Close to 500 host–pathogen interactions were predicted from the interactome of all 18-human immunodeficiency virus (HIV) proteins [56, 57].

In terms of quantifying the identified host–pathogen interactions, most IP-MS studies have relied on label-free MS quantification (e.g., spectral counting), which is simple, versatile, and can be applied to any biological system. The proteins can be labeled as one of MS strategies, through the use of ***stable isotope labeling of amino acids in cell culture*** **(SILAC)**, while the peptides can be labeled by incorporating ***tandem mass tags*** **(TMT)** or other isobaric tags [58]. The technique of SILAC was extended to the realm of hepatitis C virus to limit false-positive in study of PPI (HCV) [59, 60].

The TMT has the capability of multiplexing (nearly 10 samples can be analyzed at once). However, label-free and isotopic labeling studies cannot be done in isolation, and several studies have successfully combined SILAC with label-freeIP-MS [61], thus providing valuable information about host–pathogen interactions dynamics.

Another limitation to IP-MS datasets is the presence of non-specifically interacting proteins that co-purify with the protein of interest. The infections can trigger changes in protein dynamics in a cell, leading to some background of nonspecific associations which may be somewhat different that in an uninfected cell. Here comes the need for control isolations, from which several available computer algorithms can use data to enable filtering false-positive PPIs [62].

The significance analysis of interactome (SAINT) [63] is an algorithm that eliminates interactions identified as low confidence interactions. The application of informatics can aid in refining associations such as non-specific relations can be filtered (CRAPome) [64].

Common resources for network visualization include STRING [65] and Cytoscape [66], and we point the readers to a protocol guiding users through IP-MS data analysis [67].

3.13. INTACT PROTEINS

A **Top-down MS analysis** can facilitate the analysis of a multi-complex protein [68] (Figure 3.2B).

When performed under non-denaturing conditions, this technique can preserve both the non-covalent interactions and the post-translational state of the proteins within the complex. This technique has been applied in

individual pathogen proteins, such as p7, the hepatitis C virus pore protein [69], followed by reconstitution in vitro of these pathogenic complexes, as was done in the case of Norwalk virus-like particles [70].

The analysis of molecules of high mass is a challenge yet, the scope of the analysis is slowly seeing an increase with advances in instruments.

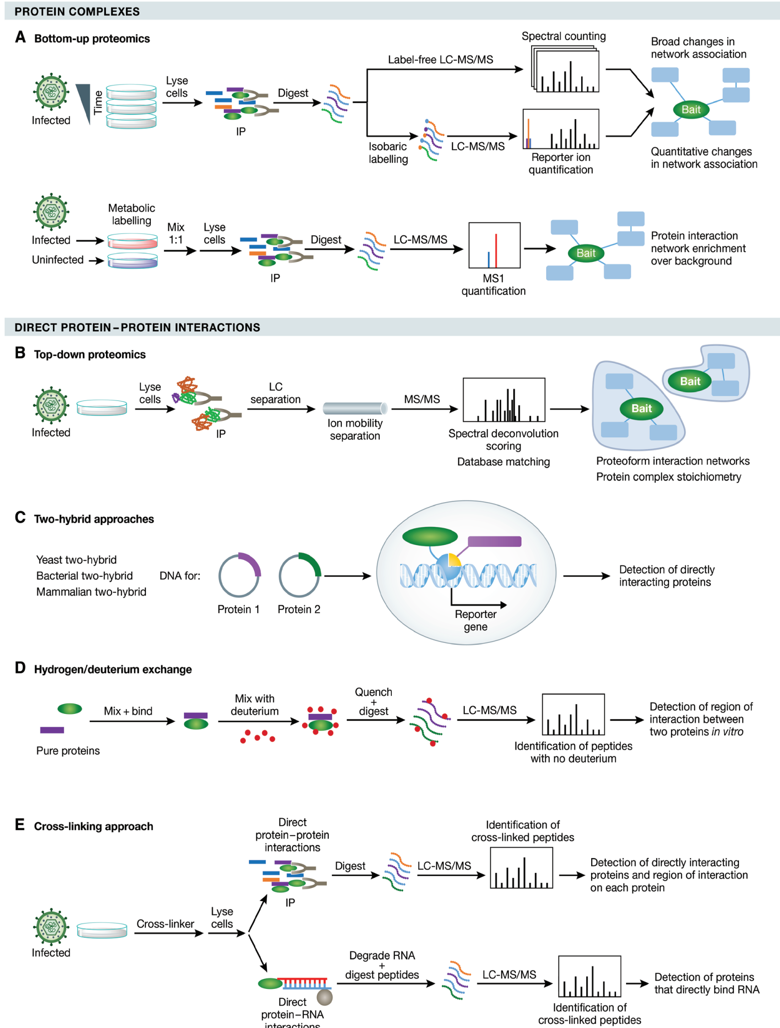


Figure 3.2. Tools of proteomics for PPI: A) Shot gun IP-MS: three approaches are covered from top to bottom the order is: Label free, Isobaric tags and SILAC. B) Top down MS to analyze for PPIs. C) Y2H to detect PPIs. D) Cross linker to analyze for PPIs. (Figure source: 71).

3.14. DIRECT INTERACTIONS AND THEIR DETECTIONS

While the methods discussed above provide unbiased detection of interactions (IP-MS) and information about the complex stoichiometry (top-down MS), these approaches are not able to classify PPIs as direct or indirect.

The yeast two-hybrid (Y2H) assay is a method of choice conventionally for detecting direct PPIs (72,73). The technique faces a challenge in false positives as proteins can be expressed that do not have a role in physiology. Moreover, several probable relevant interactions could be missed when pathogen proteins are expressed beyond an infection [74].

A cross-linking technology containing an affinity handle cleavable by the MS instrument improved the detection of cross-linked peptides and was effective for studying virus plant protein interactions and their surface topologies (75,76). Cross-linkers, in addition to the identification of direct PPIs, help in stabilizing nteractions that are transient or weak, enhancing their chances of identification, although the chances of non-specific associations increase. Cross linking generated a repertoire of human lung cells and *Acinetobacter baumannii*, (i.e., XLinkDB) [77].

Additionally, RNA–protein interactions were captured using *photo cross-linking*, to provide structural and stoichiometric information about the initiation of packaging of HIV viral genome [78].

MS with cross linking has been applied to the realm of RNA diseases such as [79] and dengue [80].

3.15. POSTTRANSLATIONAL MODIFICATIONS THAT ARE REGULATED BY PATHOGEN

Post-translational modifications (PTMs) facilitate the analysis of stability as well as interactions and localization of proteins.

Therefore, the progression and outcome of infection can be regulated by PTMs [81].

3.15.1. MS and PTMs

Proteomics can aid in the analysis of common post-translational modifications of cells. Several microbes such as bacteria or virus as well as protozoa have been targets to study PTMs [82, 83]. The PTMs studied include [84] (Figure 3.3)

- Phosphorylation
- Modification of histones
- Acetylation
- SUMOylations

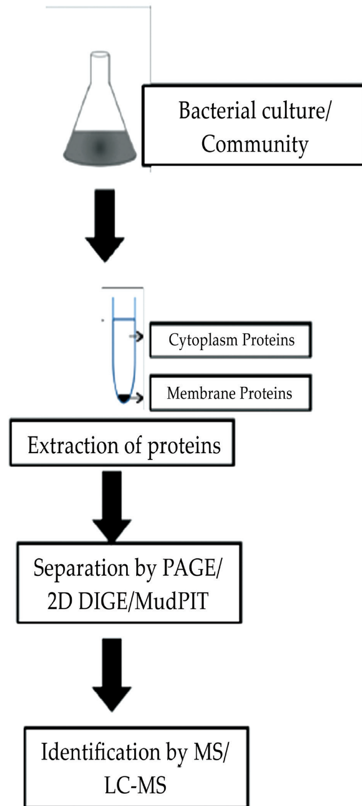


Figure 3.3. Overview of MS for proteomics for microbial biomarkers.

In studies involving PTM discovery, specific proteins are selectively enriched followed by identification of modified peptides. The use of antibodies can aid in the selection of a particular PTM by exploiting the affinity of the modified group such as phosphor groups via metal affinity [85].

But pathogen infection studies have not used these approaches frequently. The various types of PTMs as well as their regulation during pathogenesis still await gaps to be filled.

3.16. SUMMARY

The biomarkers can assess the clinical challenge of patient heterogeneity in disease phenotype. Molecular analysis reveals much evidence for microbial taxonomic membership and microbial products in association with infections, but their utility as clinical biomarkers is still underdeveloped.

The analysis of such data with huge amounts of data can be facilitated by the use of computational programs. Proteomics has aided in the analysis of pathogens in terms of the PTMs that are seen in host and pathogens.

An amalgamation of all the omics data can provide insights into the realm of microbiology with a focus on the design of targets for drugs.

REFERENCES

1. Tenover, F. C., Arbeit, R. D., & Goering, R. V. (1997). How to select and interpret molecular strain typing methods for epidemiological studies for bacterial infections: a review for healthcare epidemiologists. *Infection Control and Hospital Epidemiology*. *5,* 426–439.
2. Maccannell, D. (2013). Bacterial strain typing. *Clinical Laboratory Medicine*. *33*, 629–650.
3. Ehrlich, G. D., & Post, J. C. (2013). The Time Is Now for Gene- and Genome-Based Bacterial Diagnostics: "You Say You Want a Revolution." *The Journal of the American Medical Association*. 2013 Jul 15; doi: 10.1001/jamainternmed. 2013.7042.
4. Mcdade J. E., & Anderson B. E. (1998). Molecular epidemiology: applications of nucleic acid amplification and sequence analysis. *Epidemiologic Reviews*. *18,* 90–97.
5. Van Belkum, A. Tassios, P. T., Dijkshoorn, L. Haeggman, S. Cookson, B. Fry, N. K., Fussing, V. Green, J. Feil, E. Gerner-Smidt, P. Brisse, S. & Struelens, M. (2007). European Society of Clinical Microbiology and Infectious Diseases (ESCMID) Study Group on Epidemiological Markers (ESGEM). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clinical Microbiology and* Infection. Clin Microbiol Infect. Oct;13 Suppl 3:1-46.
6. Sabat, A. J., Budimir, A. Nashev, D. Sá-Leão, R. van Dijl, J. M., Laurent, F. Grundmann, H. & Friedrich, A. W. (2013). On behalf of the ESCMID Study Group of Epidemiological Markers (ESGEM). Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *EuroSurveillance*. *18*(4), pii=20380.
7. Singh, A. Goering, R. V., Simjee, S. Foley, S. L., & Zervos, M. J. (2006) Application of Molecular Techniques to the Study of Hospital Infection. *Clinical Microbiology Reviews*. *19*(3), 512–530.
8. Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., Schooley, R. T., Spilker, B. A., et al. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics*. *69,* 89–95.
9. De Gruttola, V. G., Clax, P. DeMets, D. L., Downing, G. J., Ellenberg, S. S., Friedman, L. Gail, M. H., Prentice, R. Wittes, J. & Zeger, S. L. (2001). Considerations in the evaluation of surrogate endpoints in

clinical trials. Summary of a National Institutes of Health workshop. *Controlled Clinical Trials*. *22,* 485–502.

10. Kaplan, J. M., & Wong, H. R. (2011). Biomarker discovery and development in pediatric critical care medicine. *Pediatrics and Critical Care Medicine*. 12, 165–173.
11. Schuetz, P. Christ-Crain, M. & Muller, B. (2009). Procalcitonin and other biomarkers to improve assessment and antibiotic stewardship in infections–hope for hype*? Swiss Medicine Weekly*. *139,* 318–326.
12. Dupuy, A. M., Kuster, N. Lizard, G. Ragot, K. Lehmann, S. Gallix, B. & Cristol, J. P. (2013). Performance evaluation of human cytokines profiles obtained by various multiplexed-based technologies underlines a need for standardization. Clinical Chemistry and Laboratory Medicine. 1–9.
13. Lever, A., & Mackenzie, I. (2007). Sepsis: definition, epidemiology, and diagnosis. *British Medical Journal* (Clinical research ed). *335,* 879–883.
14. Pierrakos, C., & Vincent, J. L. (2010). Sepsis biomarkers: a review. *Critical Care*. *14,* R15.
15. Standage, S. W., & Wong, H. R. (2011). Biomarkers for pediatric sepsis and septic shock. *Expert Review* of *Anti-infective* Therapy. *9,* 71–79.
16. Ehl, S., Gering, B., Bartmann, P., Hogel, J., & Pohlandt, F. (1997). C-reactive protein is a useful marker for guiding duration of antibiotic therapy in suspected neonatal bacterial infection. *Pediatrics. 99,* 216–221.
17. Bomela, H. N., Ballot, D. E., Cory, B. J., & Cooper, P. A. (2000). Use of C-reactive protein to guide duration of empiric antibiotic therapy in suspected early neonatal sepsis. *The Pediatric Infectious Disease Journal*. *19,* 531–535.
18. Jaswal, R. S., Kaushal, R. K., Goel, A., & Pathania, K. (2003). Role of C-reactive protein in deciding duration of antibiotic therapy in neonatal septicemia. *Indian Pediatrics.40,* 880–883.
19. Christ-Crain, M., Stolz, D., Bingisser, R., Muller, C., Miedinger, D., Huber, P. R., Zimmerli, W., Harbarth, S., Tamm, M., & Muller, B. (2006). Procalcitonin guidance of antibiotic therapy in community-acquired pneumonia: a randomized trial. *American Journal of Respiratory* and *Critical Care Medicine. 74,* 84–93.
20. Briel, M., Schuetz, P., Mueller, B., Young, J., Schild, U., Nusbaumer,

C., Periat, P., Bucher, H. C., & Christ-Crain, M. 2008 Procalcitonin-guided antibiotic use vs a standard approach for acute respiratory tract infections in primary care. *Archives of Internal Medicine. 168,* 2000–2007.

21. Burkhardt, O., Ewig, S., Haagen, U., Giersdorf, S., Hartmann, O., Wegscheider, K., Hummers-Pradier, E., & Welte T. 2010 Procalcitonin guidance and reduction of antibiotic use in acute respiratory tract infection. *European Respiratory Journal.36,* 601–607.
22. Nobre, V., Harbarth, S., Graf, J. D., Rohner, P., & Pugin J. 2008 Use of procalcitonin to shorten antibiotic treatment duration in septic patients: a randomized trial. *American Journal of Respiration and Critical Care Medicine.177,* 498–505.
23. Hochreiter, M., Kohler, T., Schweiger, A. M., Keck, F. S., Bein, B., von Spiegel, T., & Schroeder S. 2009 Procalcitonin to guide duration of antibiotic therapy in intensive care patients: a randomized prospective controlled trial. *Critical Care. 13,* R83.
24. Stocker, M., Fontana, M., El Helou, S., Wegscheider, K., & Berger, T. M. 2010 Use of procalcitonin-guided decision-making to shorten antibiotic therapy in suspected neonatal early-onset sepsis: prospective randomized intervention trial. *Neonatology.97,* 165–174.
25. Gibot, S., & Cravoisy, A. 2004 Soluble form of the triggering receptor expressed on myeloid cells-1 as a marker of microbial infection. *Clinical Medicine Research. 2,*181–187.
26. Bishara, J., Hadari, N., Shalita-Chesner, M., Samra, Z., Ofir, O., Paul, M., Peled, N., Pitlik, S., & Molad Y. (2007). Soluble triggering receptor expressed on myeloid cells-1 for distinguishing bacterial from aseptic meningitis in adults. *European Journal of Clinical Microbiology & Infectious Diseases.* 2007;*26,* 647–650.
27. Determann, R. M., Weisfelt, M., de Gans, J., van der Ende, A., Schultz, M. J., & van de Beek D. (2006). Soluble triggering receptor expressed on myeloid cells 1: a biomarker for bacterial meningitis. *Intensive Care Medicine 32,* 1243–1247.
28. Lawn, S. D., Myer, L., Bangani, N., Vogt, M., & Wood R. (2007). Plasma levels of soluble urokinase-type plasminogen activator receptor (suPAR) and early mortality risk among patients enrolling for antiretroviral treatment in South Africa. *BMC Infectious Diseases. 7,* 41.

29. Ostrowski, S. R., Ravn, P., Hoyer-Hansen, G., Ullum, H., & Andersen, A. B. (2006). Elevated levels of soluble urokinase receptor in serum from mycobacteria infected patients: still looking for a marker of treatment efficacy. *Scandinavian Journal of Infectious Diseases*. *38,* 1028–1032.
30. Perch, M., Kofoed, P., Fischer, T. K., Co, F., Rombo, L., Aaby, P., & Eugen-Olsen J. (2004). Serum levels of soluble urokinase plasminogen activator receptor is associated with parasitemia in children with acute Plasmodium falciparum malaria infection. *Parasite Immunology*. *26,* 207–211.
31. Koch, A., & Tacke F. (2011). Why high suPAR is not super–diagnostic, prognostic and potential pathogenic properties of a novel biomarker in the ICU. *Critical Care*. *15,* 1020.
32. Donadello, K., Covajes, C., Scolletta, S., Taccone, F. S., Santonocito, C., Brimioulle, S., Beumier, M., Vannuffelen, M., Gottin, L., & Vincent J. L. (2011). Clinical value of suPAR, a new biomarker. *Intensive Care Medicine*. *37,* S199.
33. Huttunen, R., Syrjanen, J., Vuento, R., Hurme, M., Huhtala, H., Laine, J., Pessi, T., & Aittoniemi J. (2011). Plasma level of soluble urokinase-type plasminogen activator receptor as a predictor of disease severity and case fatality in patients with bacteraemia: a prospective cohort study. *Journal of Internal Medicine. 270,* 32–40.
34. Fine, M. J., Auble, T. E., Yealy, D. M., Hanusa, B. H., Weissfeld, L. A., Singer, D. E., Coley, C. M., Marrie, T. J., & Kapoor, W. N. (1997). A prediction rule to identify low-risk patients with community-acquired pneumonia. *New England Journal of Medicine*. *336,* 243–250.
35. Huang, D. T., Weissfeld, L. A., Kellum, J. A., Yealy, D. M., Kong, L., Martino, M., & Angus, D.C. (2008). Risk prediction with procalcitonin and clinical rules in communityacquired pneumonia. *Annals of Emergency Medicine. 52*(48), 58–42.
36. Yaegashi, Y., Shirakawa, K., Sato, N., Suzuki, Y., Kojika, M., Imai, S., Takahashi, G., Miyata, M., Furusako, S., & Endo S. (2005). Evaluation of a newly identified soluble CD14 subtype as a marker for sepsis. *Journal of Infection and Chemotherapy*. *11,* 234–238.
37. Xiao, B., Liu, Z., Li, B. S., Tang, B., Li, W., Guo, G., Shi, Y., Wang, F., Wu, Y., Tong, W. D., et al. (2009). Induction of microRNA-155 during Helicobacter pylori infection and its negative regulatory role in the inflammatory response. *Journal of Infectious Diseases*. *200,* 916–925.

38. Vigorito, E., Kohlhaas, S., Lu, D., & Leyland, R. (2013). miR-155: an ancient regulator of the immune system. *Immunological Reviews*. *253,*146–157.

39. Vasilescu, C., Rossi, S., Shimizu, M., Tudor, S., Veronese, A., Ferracin, M., Nicoloso, M. S., Barbarotto, E., Popa, M., Stanciulea, O., et al. (2009). MicroRNA fingerprints identify miR-150 as a plasma prognostic marker in patients with sepsis. *PLoS One*. *4,*e7405.

40. Wang, H., Zhang, P., Chen, W., Feng, D., Jia, Y., & Xie L. (2012). Serum microRNA signatures identified by Solexa sequencing predict sepsis patients' mortality: a prospective observational study. *PLoS One*. *7,*e38885.

41. Morens, D. M., & Fauci, A. S. (2012). Emerging infectious diseases in 2012: 20 years after the institute of medicine report. *MBio*. *3,* e00494 12.

42. Malone, R. W., Homan, J., Callahan, M. V., Glasspool-Malone, J., Damodaran, L., Schneider Ade, B., Zimler, R., Talton, J., Cobb, R. R., Ruzic, I., Smith Gagen, J., Janies, D., Wilson, J., & Zika Response Working G. (2016). Zika virus: medical countermeasure development challenges. *PLoS Neglected Tropical Diseases*. *10,* e0004530

43. Rieder, F., & Steininger C. (2014). Cytomegalovirus vaccine: phase II clinical trial results. *Clinical Microbiology and Infections*. 20(Suppl. 5): 95–102

44. Lum, K. K., & Cristea, I. M. (2016). Proteomic approaches to uncovering virus-host protein interactions during the progression of viral infection. *Expert Reviews in Proteomics*. *13,* 325–340.

45. Greco, T. M., Diner, B. A., & Cristea, I. M. (2014). The impact of mass spectrometry-based proteomics on fundamental discoveries in virology. Annual Review of Virology. *1,* 581–604.

46. Cristea, I. M., Carroll, J. W., Rout, M. P., Rice, C. M., Chait, B. T., & MacDonald, M. R. (2006). Tracking and elucidating alphavirus-host protein interactions. *Journal of Biological Chemistry*. *281,* 30269–30278.

47. Wu, W., Tran, K. C., Teng, M. N., Heesom, K. J., Matthews, D. A., Barr, J. N., & Hiscox, J. A. (2012). The interactome of the human respiratory syncytial virus NS1 protein highlights multiple effects on host cell biology. *Journal of Virology*. *86,* 7777–7789.

48. Moorman, N. J., Sharon-Friling, R., Shenk, T., & Cristea, I. M.A.

(2010). targeted spatialtemporal proteomics approach implicates multiple cellular trafficking pathways in human cytomegalovirus virion maturation. *Molecular Cell and Proteomics*. *9,* 851–860.

49. Kramer, T., Greco, T. M., Taylor, M. P., Ambrosini, A. E., Cristea, I. M., Enquist, L. W. (2012). Kinesin-3 mediates axonal sorting and directional transport of alphaherpesvirus particles in neurons. *Cell Host Microbe. 12,* 806–814.
50. Moorman, N. J., Cristea, I. M., Terhune, S. S., Rout, M. P., Chait, B. T., & Shenk T. (2008).Human cytomegalovirus protein UL38 inhibits host cell stress responses by antagonizing the tuberous sclerosis protein complex. *Cell Host and Microbes. 3,* 253–262.
51. Li, T., Chen, J., & Cristea, I. M. (2013). Human cytomegalovirus tegument protein pUL83 inhibits IFI16-mediated DNA sensing for immune evasion. *Cell Host and Microbe*. *14,* 591–599.
52. Mayer, D., Molawi, K., Martinez-Sobrido, L., Ghanem, A., Thomas, S., Baginsky, S., Grossmann, J., & Garcia-Sastre A. (2007). Schwemmle M.Identification of cellular interaction partners of the influenza virus ribonucleoprotein complex and polymerase complex using proteomic-based approaches. *Journal of Proteome Research*. *6,* 672–682.
53. Diner, B. A., Lum, K. K., Javitt, A., & Cristea, I. M. (2015). Interactions of the antiviral factor interferon gamma-inducible protein 16 (IFI16) mediate immune signaling and herpes simplex virus-1 immunosuppression. *Molecular Cell Proteomics. 14,*2341–2356.
54. Diner, B. A., Lum, K. K., Toettcher, J. E., & Cristea, I. M. (2016).Viral DNA sensors IFI16 and Cyclic GMP-AMP synthase possess distinct functions in regulating viral gene expression, immune defenses, and apoptotic responses during herpesvirus infection. *MBio. 7,* e01553–16
55. Yamayoshi, S., Noda, T., Ebihara, H., Goto, H., Morikawa, Y., Lukashevich, I. S., Neumann, G., Feldmann, H., & KawaokaY. (2008). Ebola virus matrix protein VP40 uses the COPII transport system for its intracellular transport. *Cell Host Microbe.* 2008 Mar 13;*3*(3), 168–177.
56. Luo, Y., Jacobs, E. Y., Greco, T. M., Mohammed, K. D., Tong, T., Keegan, S., Binley, J. M., Cristea, I. M., Fenyo, D., Rout, M. P., Chait, B. T., & Muesing, M. A. (2016). HIV-host interactome revealed directly from infected cells. *Nature Microbiology*. *1,* 16068.
57. Jager, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., McGovern, K. E., Clarke, S. C., Shales, M., Mercenne, G., Pache, L., Li, K.,

Hernandez, H., Jang, G. M., Roth, S. L., Akiva, E., Marlett, J., Stephens, M., D'Orso, I., Fernandes, J., Fahey, M., & Mahon C. (2012). Global landscape of HIV-human protein complexes. *Nature. 481,* 365–370.

58. Bantscheff, M., Lemeer, S., Savitski, M. M., & Kuster B. (2012). Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Annals of Bioanalytical Chemistry. 404,* 939–965

59. Gerold, G., Meissner, F., Bruening, J., Welsch, K., Perin, P. M., Baumert, T. F., Vondran, F. W., Kaderali, L., Marcotrigiano, J., Khan, A. G., Mann, M., Rice, C. M., & Pietschmann T. (2015). Quantitative proteomics identifies serum response factor binding protein 1 as a host factor for hepatitis C virus entry. *Cell Reports. 12,* 864–878.

60. Auweter, S. D., Bhavsar, A. P., de Hoog, C. L., Li, Y., Chan, Y. A., van der Heijden, J., Lowden, M. J., Coombes, B. K., Rogers, L. D., Stoynov, N., Foster, L. J., & Finlay, B. B. (2011). Quantitative mass spectrometry catalogues Salmonella pathogenicity island-2 effectors and identifies their cognate host binding partners. *Journal of Biological Chemistry. 286,* 24023–24035

61. Joshi, P., Greco, T. M., Guise, A. J., Luo, Y., Yu, F., Nesvizhskii, A. I., & Cristea, I. M. (2013).The functional interactome landscape of the human histone deacetylase family. *Molecular Systems Biology*. 2013, *9,* 672.

62. Armean, I. M., Lilley, K. S., & Trotter, M. W. (2013). Popular computational methods to assess multiprotein complexes derived from label-free affinity purification and mass spectrometry (AP-MS) experiments. *Molecular Cell Proteomics. 12,* 1–13.

63. Choi, H., Larsen, B., Lin, Z. Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z. S., Tyers, M., Gingras, A. C., & Nesvizhskii, A. I. (2011). SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nature Methods. 8,* 70–73

64. Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J. P., St-Denis, N. A., Li, T., Miteva, Y. V., Hauri, S., Sardiu, M. E., Low, T. Y., Halim, V. A., Bagshaw, R. D., Hubner, N. C., AlHakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W. H., Goudreault, M., Lin ZY et al. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature Methods. 10,* 730–736.

65. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P.,

Kuhn, M., Bork, P., Jensen, L. J., & von Mering C. (2015). STRING v10, protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*. *43,* D447–D452.

66. Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico AR et al. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols*. *2,* 2366–2382.
67. Morris, J. H., Knudsen, G. M., Verschueren, E., Johnson, J. R., Cimermancic, P., Greninger, A. L., & Pico, A. R. (2014). Affinity purification-mass spectrometry and network analysis to understand protein-protein interactions. *Nature Protocols*. *9,* 2539–2554.
68. Toby, T. K., Fornelli, L., & Kelleher, N. L. (2016). Progress in top-down proteomics and the analysis of proteoforms. *Annual Review of Analytical Chemistry* (Palo Alto Calif). *9,* 499–519.
69. Konijnenberg, A., Bannwarth, L., Yilmaz, D., Kocer, A., Venien-Bryan, C., & Sobott F. (2015). Top-down mass spectrometry of intact membrane protein complexes reveals oligomeric state and sequence information in a single experiment. *Protein Science. 24,* 1292–1300.
70. Shoemaker, G. K., van Duijn, E., Crawford, S. E., Uetrecht, C., Baclayon, M., Roos, W. H., Wuite, G. J., Estes, M. K., Prasad, B. V., & Heck, A. J. (2010). Norwalk virus assembly and stability monitored by mass spectrometry. *Molecular Cell and Proteomics*. *9,* 1742–1751.
71. Jean Beltran, P. M., Federspiel, J. D., Sheng, X., & Cristea, I. M. (2017). Proteomics and integrative omic approaches for understanding host–pathogen interactions and infectious diseases. *Molecular Systems Biology*. *13*(3), 922. doi:10.15252/msb.20167062.
72. Blasche, S., Arens, S., Ceol, A., Siszler, G., Schmidt, M. A., Hauser, R., Schwarz, F., Wuchty, S., Aloy, P., Uetz, P., Stradal, T., & Koegl M. (2014). The EHEC-host interactome reveals novel targets for the translocated intimin receptor. *Scientific Reports*. *4,* 7531.
73. Calderwood, M. A., Venkatesan, K., Xing, L., Chase, M. R., Vazquez, A., Holthaus, A. M., Ewence, A. E., Li, N., Hirozane-Kishikawa, T., Hill, D. E., Vidal, M., Kieff, E., & Johannsen E. (2007). Epstein-Barr virus and virus human protein interaction maps. *Proceedings of the National Academy of Sciences USA*. *104,* 7606–7611

74. Monroe, E. B., Kang, S., Kyere, S. K., Li, R., & Prevelige PE Jr.(2010). Hydrogen/deuterium exchange analysis of HIV-1 capsid assembly and maturation. *Structure. 18,* 1483–1491.
75. Chavez, J. D., Cilia, M., Weisbrod, C. R., Ju, H. J., Eng, J. K., Gray, S. M., & Bruce, J. E. (2012).Cross-linking measurements of the Potato leafroll virus reveal protein interaction topologies required for virion stability, aphid transmission, and virus-plant interactions. *Journal of Proteome Research*. 2012,*11,* 2968–2981.
76. DeBlasio, S. L., Chavez, J. D., Alexander, M. M., Ramsey, J., Eng, J. K., Mahoney, J., Gray, S. M., Bruce, J. E., & Cilia M. (2015). Visualization of host polerovirus interaction topologies using protein interaction reporter technology. *Journal of Virology*. *90,* 1973–1987.
77. Schweppe, D. K., Harding, C., Chavez, J. D., Wu, X., Ramage, E., Singh, P. K., Manoil, C., & Bruce, J. E. (2015). Host-microbe protein interactions during bacterial infection. *Chemistry and Biology*. *22,* 1521–1530.
78. Kenyon, J. C., Prestwood, L. J., & Lever AM. (2015). A novel combined RNA-protein interaction analysis distinguishes HIV-1 Gag protein binding sites from structural change in the viral RNA leader. *Science Reports. 5,* 14369.
79. Lenarcic, E. M., Landry, D. M., Greco, T. M., Cristea, I. M., & Thompson, S. R. (2013). Thiouracil cross-linking mass spectrometry: a cell-based method to identify host factors involved in viral amplification. *Journal of Virology. 87,* 8697–8712.
80. Viktorovskaya, O. V., Greco, T. M., Cristea, I. M., & Thompson, S. R. (2016). Identification of RNA binding proteins associated with dengue virus RNA in infected cells reveals temporally distinct host factor requirements. *PLoS Neglected Tropical Diseases*. *10,* e0004921
81. Ribet, D., Cossart P.Pathogen-mediated posttranslational modifications: a re-emerging field. *Cell; 143,* 694–702.
82. Leach, M. D., & Brown, A. J. (2012). Posttranslational modifications of proteins in the pathobiology of medically relevant fungi. *Eukaryotic Cell. 11,* 98.
83. Bell, C., Desjardins, M., Thibault, P., & Radtke K. (2013). Proteomics analysis of herpes simplex virus type 1-infected cells reveals dynamic changes of viral protein expression, ubiquitylation, and phosphorylation. *Journal of Proteome Research*. *12,* 1820–1829.

84. Kulej, K., Avgousti, D. C., Weitzman, M. D., & Garcia, B. A. (2015). Characterization of histone post-translational modifications during virus infection using mass spectrometry-based proteomics. *Methods. 90,* 8–20.

85. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faergeman, N. J., Mann, M., & Jensen, O. N. (2005). Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Molecular Cell Proteomics*. *4,* 310–327.

CHAPTER 4

PLANT BIOMARKERS DISCOVERY BY PROTEOMICS

CONTENTS

4.1. Introduction ... 114
4.2. Why Proteomics For Plant Biomarkers? ... 114
4.3. Approaches For Plant Biomarkers ... 116
4.4. Proteomics For Reactive Oxygen Species Research ... 119
4.5. Stress Responses In Two Plants: Proteomics For Biomarkers ... 119
4.6. Proteomics For Plants Infected With Virus: Biomarkers ... 122
4.7. The Advantages of Proteomics ... 122
4.7.8. Machinery of The Cell ... 125
4.8. Protein Microarrays For Biomarkers ... 126
4.9. Other Approaches ... 126
4.10. Interpreting The Data Using Bioinformatics ... 126
4.11. Proteomics For Food From Agriculture: Biomarkers: ... 127
4.12. Biomarkers For Biofuels ... 129
4.13. Biomarkers For Important Crops: Applications Of Proteomics ... 130
4.14. Proteomics of Plants For Biomarkers: An Organ Wise Approach ... 133
4.15. Biomarkers: Proteomic Analysis of Seedlings of Crops ... 133
4.16. Proteomics And Biomarkers ... 138
4.17. Root Proteomics ... 138
4.18. Leaf Proteomics ... 140
4.19. Xylem and Phloem Proteomics ... 141
4.20. Summary ... 142
References ... 143

4.1. INTRODUCTION

About 30,000 to 80,000 of the 352,000 plants documented in totality can be used by humans as food [1]. However, the use of agricultural practices of today is have brought down the extent of crops that are fit to be consumed by humans with 150 species being cultivated [2]. This lack of use of available resources can threaten security of food. Of the crops used as staple foods such as rice, wheat and corn; all are targets of stress conditions; this is a grim scenario as these constitute 60% of food [3]. For instance, 96% of *Triticum aestivum* L. of winter crop of Russia can be traced to 2 cultivars that is a clear indication of genetic erosion [4].

4.2. WHY PROTEOMICS FOR PLANT BIOMARKERS?

The term proteome has its roots from Proteins of a genome that can offer functional studies as well as the various modifications occurring following translation [5].

In order to look at tolerance of plants to stress research teams are actively pursuing studies in various angles. A technique that is showing an advantage includes proteome studies in order to identify several aspects in plants that will be covered in the subsequent portions.

The mechanisms by which plants achieve resilience especially in the field of changes in climate across the globe are vital today. Changes in weather have resulted in damage to the productivity of crops [6]. To illustrate this situation, the year 2012 saw a decrease of 13% in production of corn in the US in spite of having a highest amount of land cultivated [4]. The development of plants in fields or labs is subject to various external influences such as pathogens or temperature and water. These factors can negatively impact the growth and cultivation of crops.

The role of proteomics can be seen as important with the advances in instruments, platforms and bioinformatics tools. The field of proteomics can illustrate the following:

- Quantitatively present amount of proteins
- Identify interactions between proteins
- post-translational modifications can be studied
- the location of protein complexes can be identified

Proteomics can unveil the interactions between proteins as well as their modifications that can affect crop survival to various responses of plants to stress. Literature indicates an increase in the use of proteomics to study various pathways to study the molecules/ biomarkers involved in plant responses in models such as: sorghum, rice or *Arabidopsis.*

A 2012 study showed that in a crop scenario plants are exposed to a slew of stresses: biotic and abiotic. Studies in a lab can involve the effect of each stress on a plant growth while it was shown that more than one stress can affect a plant [7]. These multiple factors can involve multiple pathways and proteins that can be scoured for biomarkers. (Figure 4.1)

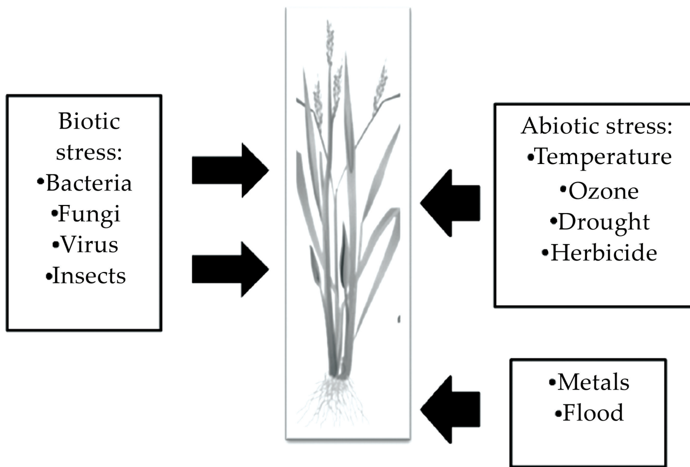


Figure 4.1. Different biotic and abiotic stresses in plant.

The role of proteomics in crop plants has illustrated several changes in stress involving a host of proteins and other molecules in the field [6]. Additionally, proteomics can be applied to specific organs as Organ-specific proteomics that can identify the effects of different stress in different parts.

Several plant genomes such as *Arabidopsis thaliana, Zea mays, Glycine max* to name a few have been sequenced. The identification of proteins and biomarkers can aid improvement of crops. The use of instruments and platforms of high accuracy can aid in the:

- Responses to stress
- Improving yield
- Boosting nutritional content

4.3. APPROACHES FOR PLANT BIOMARKERS

The analysis in proteomics involves several steps such as:

- The separation of proteins using liquid chromatography or 2DE or DIGE (all covered in chapter 2).
- Identification of proteins using MS platforms
- Study of post-translational modifications
- Evaluation of interactions among proteins
- Mapping of proteins to determine their function and structure
- Application of bioinformatics tools

Despite the benefits conferred by the use of gel free technologies in proteomics such as the level of advancements as compared to early techniques such as 2 DE there are few challenges. A challenge of 2-DE was the localization of more than one protein found in a band on a gel. However, the reagents used in this technique present advantages such as their ability to solubilize several hydrophobic proteins (associated with organelles). A limitation of gel free technologies is they can target proteomes that are more soluble and those that are less hydrophobic. (Figure 4.2).

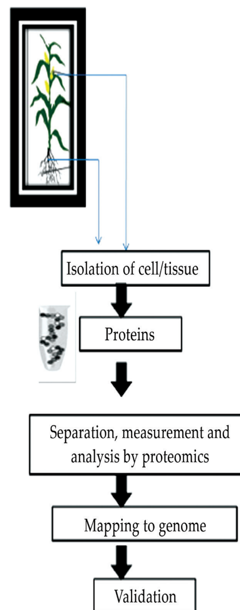


Figure 4.2. Proteomics in plant biomarkers: a scheme.

Applications such as multiplexed **selective reaction monitoring (SRM)** are emerging as powerful techniques in the analysis of proteomes. The technique can facilitate the study of low abundance protein markers. The validation of biomarkers is favored by the approach that can aid in marker-assisted breeding. A 2014 study used SRM using hybrid quadrupole-Orbitrap technology aided in large scale proteome analysis [8].

Another strategy presented includes ***SWATH MS*** was reported in 2012. It is a combination of targeted data extraction along with data-independent acquisition that is highly specific [9]. The approach involves the analysis of spectra of all ions in two dimensions according to the user. An advantage of this technique is re-mining of the data that is acquired in a retrospective and repetitive manner.

Label-free shotgun proteomics have been reported to analyze for markers/ proteins in a rapid manner from complex samples. It uses a mix of: LC–MS/MS. A 2013 study used this data independent (MSE) acquisition method to analyze allergens in wheat. The study identified gliadins and glutenin's in wheat apart from certain protein markers that cause asthma in bakers [10]. (Figure 4.3).

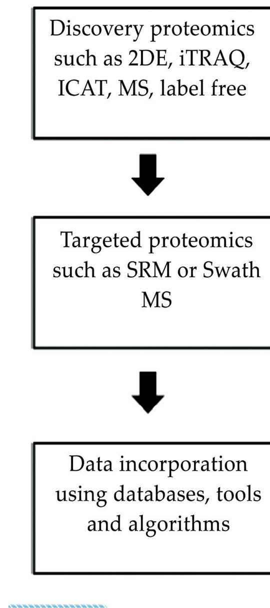


Figure 4.3. Flow of proteomics in plant biomarkers.

Proteomics has also been applied at the level of organelles that are involved in several biological processes. The application of proteomics to chloroplasts of wheat used two applications: LTQ-FTICR and MALDI-TOF/TOF and 2-DE and SDS–PAGE technique. This study revealed the role of proteins that are involved in abiotic stress of photosynthesis [11].

Other studies of organelles include:

- Cell wall of plants such as soybean, rice and maize
- Soybean plasma membrane was studied using LC–MS/MS that revealed the role of proteins in osmotic pressure.
- 2-DE Blue Native PAGE along with Q-TOF MS of pea mitochondria was analyzed for the effect of herbicide [6].

The field of wheat has received recent attention as previously the genome was limited. Despite the International Wheat Genome Sequencing Consortium of 2014, the annotation is still yet to be completed.

Several studies of wheat using proteomics are involved:

The coleoptiles of wheat were analyzed by 4-plex iTRAQ to look at the stress response to Hydrogen peroxide using UniProt database.

Label-free techniques were applied to study the effect of salinity stress using Swiss-Prot viridiplantae database.

Other examples are summarized Table 4.1.

Table 4.1. Subcellular Proteomic Studies in Crops Under Abiotic Stress [6]

Technique of proteomics	**Target**	**Result**
2DE and MALDI	Chloroplasts of *Zea mays* and *T. aestivum*	Several proteins identified as responses to salt
2DE and MS	Mitochondria of roots and shoots of *T. aestivum*	Stress related proteins were identified
Label free techniques	Nucleus of roots of *Glycine max*	Flood responsive proteins
I DE and LC MS /MS	Seeds of Hordeum vulgarae	Proteins associated with germination were identified

4.4. PROTEOMICS FOR REACTIVE OXYGEN SPECIES RESEARCH

Reactive oxygen species (ROS) are involved in growth, signaling and regulation of plants as well as plant interactions with environment. An alteration in the pathways that regulate the mechanisms that control ROS results in oxidative stress and subsequent cell death. Thus, the measurement of ROS damage requires the sensitive and precise measurement of ROS.

The challenges in measuring ROS are their short life and highly reactive property that limits their detection in a biological context. A common method to measure ROS is electron paramagnetic

Resonance coupled with spin trapping in plants. However this technique requires the incorporation of spin molecules whose toxic effects or permeability within plants yet require characterization.

The outcome of peroxidation of lipids can be an indicator of oxidative damage in plants. The peroxidation of lipids such as polyunsaturated fatty acids yields 4-hydroxynonenal (4-HNE) and malondialdehyde (MDA). A combination of liquid chromatography- MS could detect 2, 4-dinitrophenyl (DNP) hydrazine derivatives of these compounds up to pico mole ranges [12].

4.5. STRESS RESPONSES IN TWO PLANTS: PROTEOMICS FOR BIOMARKERS

Two cereal plants chosen to analyze proteome include barley: *Hordeum vulgare* and wheat: *Triticum aestivum; T. durum* as they form important crops. However, there are a host of factors that affect cereals such as:

Biotic: viruses and fungi such as leaf rust (*Puccinia triticina*) and Fusarium

Abiotic: minerals, cold, heat, salinity and drought

With the direct involvement of proteins in several responses, the analysis of proteome is a key. With an increase in the use of high-throughput proteomics, various factors have been analyzed. Several studies of proteomics have identified proteins that were differentially expressed that are summarized in the table 4.2:

Table 4.2. Techniques of Proteomics for Crops

Material	**Stress**	**Method of proteomics**	**Results**
Naked barley (*Hordeum vulgare*)	Pathogen: *Fusarium culmorum* and *F. graminearum*	2DE and MALDI-TOF nano LC-MS/MS	Upregulated proteins: • serpin • DNA-dependent RNA-polymerase • NBS-LRR (transcription regulation) • Dof zinc-finger protein Down regulated protein: ADP-glucose pyrophosphorylase
Common wheat	*Septoria tritici*	micro-column; LC-MS/MS	Phosphoproteins and CDPK, MAPK were higher in Stakado compared to Sevin
Common wheat–spring wheats	Osmotic stress (PEG-6000)	2DE MALDI-TOF/TOF	Upregulated proteins: • 26S proteasome, • V-ATPase A • GAPDH B Down regulated proteins: • RubisCO • LSU and SSU • GAPDH • AGPase • TPI
Common wheat–Thésée–grain	Heat:	2 DE MALDI-TOF	Upregulated proteins: • 20 kD sHSP, • 17 kD class II HSP; • HSP82 (HSP90 family); eEF-Tu, • V-ATPase subunit E Down regulated proteins: • starch biosynthesis enzymes granule-bound starch synthase, • glucose-1-phosphate adenyltransferase; β-amylase; β subunit ATP synthase
Barley (*Hordeum vulgare*)	Cold: 3°C	2D-DIGE MALDI-TOF	Upregulated proteins: • HSP70 • OEE1 (PsbO) Down regulated proteins: • eEF-Tu; GS1 • 2; UDP-glucose 6-dehydrogenase

Durum wheat (*T. durum*	Salinity: 100 mM NaCl	2DE MALDI-TOF	Up regulated proteins: • TPI • CPN60-β • RubisCO activase • carbonic anhydrase • osmolyte biosynthesis-related enzymes Down regulated proteins: • ALDO • PGK • RubisCO SSU, • OEE1 precursor • β-glucosidase, • ATP synthase CF1 α
Barley cv.	Drought: no watering for a week	2D-DIGE MALDI-TOF	Upregulated proteins • ABA-induced protein r40c1 • small G-protein Rab2 • Myb-like protein • 14-3-3 protein Down: • GST • GPX

NBS-LRR, nucleotide-binding site leucine-rich repeat protein; CDPK, calcium-dependent protein kinase; MAPK, mitogen-activated protein kinase; GAPDH B, glyceraldehyde-3-phosphate dehydrogenase B form; RubisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase; TPI, triose phosphate isomerase; TPI, triose phosphate isomerase; PGK, phosphoglycerokinase; GPX, glutathione peroxidase; GST, glutathione S-transferase

The study of the proteome has revealed several observations such as:

- Several kinases such as MAPK, CDPK are activated suggesting a role for phosphorylated proteins as markers.
- An increase in level of HSP70, HSP100, chopper chaperone and protein disulfide isomerase during heat stress.
- HSP90 was decreased during cold.
- ROS was decreased through a dip in photosynthesis; this is reflective of a dip in D1 and D2 (photosystem II reaction center) as well as RubisCO.
- Late embryogenesis-abundant (LEA) were increased in the presence of cold, salinity as well as wounds.
- Small glycine-rich RNA-binding proteins (sGRPs) levels were altered during cold.

- All the stress patterns increase the levels of makers associated with catabolism of glucose such as aconitase, glyceraldehyde-3-phosphate dehydrogenase, enolase as well as β subunit of CF1 complex.
- Pathogenesis-related (PR) proteins were up regulated in stress such as pathogens and cold/ salinity stress.
- There are differences in the proteomic profile of plants that are resistant to stress and those that are susceptible to stress. For instance, stress resistant plants show higher levels of ABA-responsive proteins and PR proteins.
- The tolerance to frost and cold was seen in wheat WCS120 and barley DHN5that are dehydrin are suggestive of the use of WCS120 and DHN5 as markers of frost tolerance.
- eIF5A2 factor (cell cycle switch and translation initiation) was found to be lower in *T. aestivum* × *Th. Ponticum* was lower in salt resistant and salt sensitive plants.

Thus, the role of proteomics has been found to indicate various novel biomarkers that can indicate tolerance to stress (such as above-mentioned FT marker)

(Reviewed by Kosová et al, 2014; 13)

4.6. PROTEOMICS FOR PLANTS INFECTED WITH VIRUS: BIOMARKERS

The comprehension of the mechanisms involved in the interactions of virus and plant cells can aid the development of strategies for therapeutic agents.

4.7. THE ADVANTAGES OF PROTEOMICS

- The levels of proteins cannot be calculated always from levels of mRNA as regulatory mechanisms may be at play or there may be post-translational modifications.
- Proteins that are subcellular can be located.
- Despite several half-lives, proteins can be scoured for using high throughput techniques [14].

Many studies use Two-dimensional electrophoresis (2DE) or 2D difference in gel electrophoresis (2D DIGE) to analyze isoforms of proteins that show a difference in expression following infection by virus. Another

approach is shotgun proteomics where following digestion of proteins with trypsin, liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS) is applied. An approach of co-immunoprecipitation and LC–MS/MS is also been reported [15].

4.7.1. 2D-Gel Electrophoresis and Mass Spectrometric Analysis

The application of 2-DE in combination with mass spectrometry casts a light on the protein profile on virus infected plants to be used potential biomarkers. This combination is pitched as unbiased in identifying proteins/ biomarkers that are differentially expressed. The entire scenario of virus-plant can be analyzed the technique. A challenge is the identification of preferentially highly abundant proteins against proteins present at lower levels [14].

4.7.2. Examples of This Approach

The effects of Rice yellow mottle Virus (RYMV) infection on rice plants was studied; rice was chosen as its genome is sequenced and is small. Following 1 hour postinoculation (hpi) with RYMV, two day postinoculation (dpi), 5 dpi and 7 dpi; the protein levels were analyzed by 2-DE and liquid chromatography-tandem mass spectrometric (LC-MS/MS). Proteins of rice were identified that were associated with stress response as well as translation and metabolism [16].

Nicotiana benthamiana (with N resistance gene) was used to study immunity using TMV infection. The use of 2-DE was applied to evaluate the proteins 0, 2, 8, 16 hrs post infection. The samples were also subjected to isobaric iTRAQ. Both the techniques showed an overlap of protein profiles. Certain proteins were decreased such as NbCRT2, NbCRT3, NbERp57 and NbP5 of which the first two were implicated in immunity of the plant [17].

Evidence also has been seen in the common pathways between virus infection and drought: aminopeptidases are antioxidant enzymes that are increased in tomato. The level of DIP-1 was higher in plants like water melon. PR proteins such as Beta-1,3-Endoglucanase (GLU) and Chitinase (CHI) were increased on viral infection in tomato upon infection with TMV [14].

In order to delve deeper into organelles; a study in 2004 studied the changes on photo-system II electron transport upon infection with Tobamo virus. Shifts in the levels of PsbO and PsbP were reported [18].

4.7.3. Changes in Metabolism as Biomarkers

Aminotransferases such as Glutamine synthase is decreased upon infection with Papaya meleiravirus; PMeV.

Enzymes ADP/UDP-glucose pyrophosphorylase in starch biosynthesis are increased in infection with Rice black-streaked dwarf virus.

Soybean mosaic virus caused a decrease in the levels of GAPDH while the levels of NADPH-specific isocitrate dehydrogenase were increased in a resistant variety of soybean.

The synthesis of amino acids was increased in chloroplasts upon infection of squash by Squash mosaic virus.

4.7.4. Chaperones as Biomarkers

Four homologs of Hsp70 were found along with Potato leaf roll virus and Rice yellow mottle virus.

In resistant and susceptible varieties of rice, the presence of a homolog of Hsp90 was found with Rice yellow mottle virus.

4.7.5. Reactive Oxygen Species ROS as Biomarkers

Potato leaf roll virus was found to precipitate along with several important enzymes involved scavenging of ROS.

The levels of catalase enzymes are altered upon infection of Zucchini yellow mosaic virus.

Mungbean yellow mosaic India virus that targeted resistant *V. mungo* saw an increase in the level of peroxiredoxins while this difference was not seen in susceptible strains.

4.7.6. Photosynthesis

Several key proteins such as phosphoenolpyruvate carboxylase, RuBisCO large subunit, ATP synthase sub units were precipitated along with Rice yellow mottle virus.

PsbP and PsbQ (enhancer proteins of Photosystems I and II), transketolase and ATP synthase sub units were found along with infection of Potato leaf roll virus.

Odonto glossum ring spot virus is precipitated along with RuBisCO and enzymes of photorespiration.

4.7.7. Enzymes as Biomarkers

Chitinases such as PR-3, 4, 8, and 11 are associated with resistance to fungi were found to be at a higher level upon infection with Pepper mild mottle virus.

Pepper mild mottle virus infection of pepper caused an increase in the level of beta-1,3-glucanases; such an observation was also seen in tomato infected with Cucumber mosaic virus.

A putative Oxalate oxidase was found to be decreased in papaya upon exposure to Papaya meleiravirus.

Infection by on resistant rice caused an increase in the levels of Glutathione-S-transferase upon exposure to Potato leaf roll virus as well as Rice yellow mottle virus.

4.7.8. Machinery of the Cell

A translation elongation factor was found along with Rice yellow mottle virus while close to 20 elongation or initiation factors are found along with Potato leaf roll virus.

E3 ubiquitin ligase as well as ubiquitin fusion protein was detected along with Potato leaf roll virus.

4.7.9. Cell Wall Factors as Biomarkers

Pectin methyl esterase was found to co purify along with Potato leaf roll virus.

Infection by Tomato chlorotic mottle virus caused an increase in the levels of Callase and callose synthase.

Sugarcane mosaic virus caused an increase in the level of cinnamyl alcohol dehydrogenase involved in synthesis of lignin in a resistant variety of sugarcane.

(Reviewed by Alexander and Ciliaa, 2016; 15).

4.8. PROTEIN MICROARRAYS FOR BIOMARKERS

As discussed in Chapter 2; a microarray is composed of a surface bearing several proteins that can identify potential biomarkers such as nucleic acids or proteins [14].

The interaction of Brome mosaic virus (BMV) with host was explored using FAST slides with host proteins. Using labels of Cy3 and Cy5; it was determined that several protein markers such as Pus4p (pseudouridine synthase) and App1p (actin patch binding protein) were instrumental in reducing the plus strand formation of the virus. The proteins also inhibited the spread of the virus across *Nicotiana benthamiana* plant [19].

4.9. OTHER APPROACHES

Most plant-virus proteomics has seen the application of 2-dimensional electrophoresis or 2-dimensional fluorescence difference gel electrophoresis (2D DIGE) and sometimes MS and spectral counting. The role of far Western analysis, MS or co-immunoprecipitation can cast light on protein interactions [15]

The membrane fractions of Tobacco mosaic virus and Brome mosaic virus were isolated and expressed with Tomato Mosaic virus replicase FLAG tag. This was subjected to affinity purification and co-purified proteins were identified. It was reported that this approach of co-immunoprecipitation using antibodies identified molecules such as Hsp70, eEIF-3, Tom proteins [20].

4.10. INTERPRETING THE DATA USING BIOINFORMATICS

The interpretation of the large quantities of data requires the role of bioinformatics to decipher the data. Examples of such data include:

REACTOME is a database that can report the pathways of the host that are correlated with viral proteins. These pathways of the host are composed of several organisms.

STRING is a database of interactions of proteins inclusive of genetics as well as biochemistry.

In a nutshell, the role of proteomics have aided in the identification of several markers of infection of plants by virus. The combination with genomic data can offer more insights on the association of host and pathogen. An amalgamation of cell biology and biochemistry with proteomics an integrated bioinformatics can aid in understanding of plant-pathogen interactions [14].

4.11. PROTEOMICS FOR FOOD FROM AGRICULTURE: BIOMARKERS:

The analysis of proteome for biomarkers can illustrate the nutritional quality of food.

A 2008 study evaluated biomarkers for selecting cultivars of wheat that were best developed for making of pasta. The starch as well as properties of of *Triticum turgidum* L. var. durum cultivars were analyzed using two-dimensional gel electrophoresis [21].

A team in 2005 analyzed the proteins that were key in influencing crumbling structure as well as controlling the stability of gas bubbles in dough. Following two-dimensional gel electrophoresis separation, matrix-assisted laser desorption/ionization-time of flight and quadrupole-time of flight was used to analyze dough liquor and dough liquor foam. There were 42 protein markers identified such as a major chunk of alpha-amylase/trypsin inhibitor followed by serpins, beta-amylase, tritin [22].

The role of proteomics in the analysis of bioactives to enhance the nutritional content of crops has also been reported. A Bioactive is a peptide that is formed during processing of food or during ripening or digestion. Several crops are sources of bioactives such as rice, wheat, soybean, maize and pumpkins among a few.

In the quality of food; it is important to look at stages such as ripening or the mechanisms after harvest. For instance, a 2011 study used gel based assays to scour for markers of withering of grapes after harvest. Two-dimensional differential in gel electrophoresis (2D-DIGE) proteomics of the Corvina variety of grapes revealed differences in the levels of soluble proteins. Of 90 proteins that were different in levels; MS aided in identifying 72 of them. The proteins involved in withering were found to be: 30%: stress and defense activity

- 25%: energy and primary metabolism
- 7%: remodeling of cytoskeleton (7%)
- 5% secondary metabolism [23].

The quality of commercial food items can be verified using proteomics for biomarkers. For instance, almond milk and orgeat syrup are two commercial beverages were analyzed using peptide ligand libraries. The almond product yielded 132 unique proteins while the orgeat syrup showed 14 proteins. The levels of proteins in the almond milk were close to what was displayed on the label. The orgeat syrups that were cheaper were found to lack proteins to suggest their lack of purity or the use of extracts/ chemicals instead of pure plants. Thus, such protein biomarkers can offer a glimpse into purity of proteins [24]. (Figure 4.4)

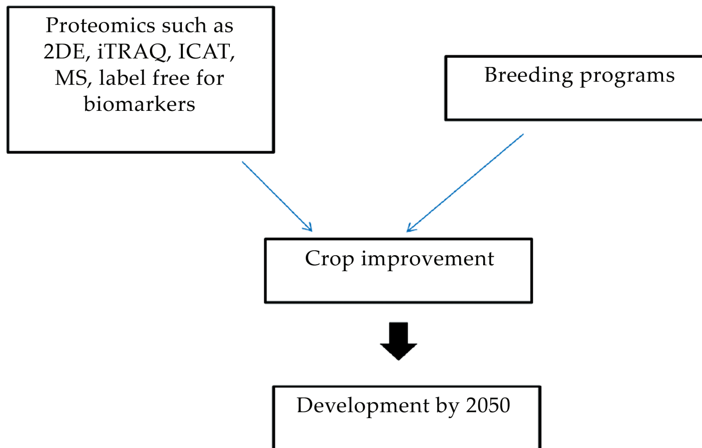


Figure 4.4. Proteomics for biomarkers and conventional breeding together can contribute to a sustainable economy by 2050 [5].

In another approach, in an attempt to look at more affordable cultivars of coffee; biomarkers were scoured for. Fingerprints of expressed proteins were studied using two-dimensional (2-D) maps of two species: *Coffea arabica* and *Coffea canephora.* Following bean milling in an inert atmosphere, following extraction in Trichloroacetic acid (TCA; 10% w/v) and beta-mercaptoethanol (0.07% v/v) in acetone followed by TCA in acetone. The proteins were solubilized and subjected to 2-D electrophoresis and stained with Sypro Ruby and analyzed by PDQuest analysis.

The differences in potential biomarkers between the two species are highlighted below:

- *Coffea canephora* showed sixteen exclusive proteins.
- *C. arabica* showed five exclusive proteins.
- There was an increase in expression of 8 proteins in *Coffea canephora* as compared with *C. Arabica.* [25].

4.11.1. Allergens in Food

The analysis of allergens in food has been attempted by various DNA based technologies. However certain food items such as egg whites lack DNA but still posses proteins that can cause allergies. For such cases, the role of biomarkers and proteomics can be a boon.

A team in 2002 analyzed rice (*Oryza sativa*): leaf, root, and seed tissue for the presence of allergens using two-dimensional gel electrophoresis followed by tandem mass spectrometry and multidimensional protein identification technology. The study revealed a total of 556 unique proteins of which the break up is:

Leaf:	348
Root:	199
Seed:	152

The seeds showed the presence of four allergenic compounds and were found to be of the family of α-amylase/trypsin inhibitor [26].

A 2006 study studied the differences between the allergic natures of the cereal *Zea mays* with a typical grass: *Phleum pretense*. The use of 2-DE immunoblotting involved the reaction with mAbs that targeted classified allergens. It was revealed that cross reacting allergens (Zea m 1 and Zea m 13) were found to react with IgE [27].

A2011 study described an approach of multiplexed liquid chromatography and triple-quadrupole tandem mass spectrometry to assay for allergens. The technique could assay for almond, peanut, milk, soy, egg and hazelnut. Using a standardized protocol of baking, bread was used to test for the seven allergic products with a detection range of 10 to 1000 μg/g [28].

4.12. BIOMARKERS FOR BIOFUELS

Biofuels are being pitched as clean fuels as well as their capacity to be used in place of fossil fuels. For instance, a study published in 2009 studied the proteome of xylem of Populous and reported close to 6000 proteins.

Several candidates reported include cellulose synthase, polygalacturonase associated with synthesis of cell wall [29].

4.13. BIOMARKERS FOR IMPORTANT CROPS: APPLICATIONS OF PROTEOMICS

1. Soybean

UV light:

Reduction in flavonoid levels increased the susceptibility of the leaves to UV-B light. The protein biomarkers here were identified as oxygen evolving enhancer protein 1 (OEE) of photosystem II as well as enzymes such as carbonic anhydrase.

Aluminum:

The differences in a sensitive and tolerant strain were evaluated by 2D-DIGE of root exposed to the metal. The tolerant strain showed markers such as malate oxidoreductase, malate dehydrogenase as well as antioxidants such as thioredoxin as well as isoflavone reductase and cysteine synthase.

2. Rice

Salinity:

The analysis of 2D between a salt tolerant and sensitive rice strain revealed biomarkers in the tolerant variety such as caffeoyl-CoA O-methyltransferase, and ASR1, associated with abscisic acid stress.

Deficiency:

A comparative analysis of species that were different in utilization of Nitrogen revealed that a sensitive strain showed altered levels of proteins associated with lack of the nutrient such as glutathione S-transferase, heat shock protein GSTF14, DegP2 protease and fibrillin-like protein.

3. Corn

UV light:

Differential in Gel Electrophoresis (DIGE) of maize species revealed the role of ferridoxin, ADP-glucose pyrophosphorylase small subunit, translation elongation factor Tu, among others that were key in improving tolerance to UV light.

Drought:

iTRAQ analysis revealed an increase in markers such as superoxide dismutase, ascorbate peroxidase, an initiation factor eIF3 and an elongation factor in mitochondria EF-TuM upon drough stress.

4. Banana

Drought:

The study of leaves treated with sorbitol to simulate water stress by Differential in Gel Electrophoresis (DIGE) revealed several markers to be at a higher level that were ROS detoxification proteins and NAD/NADH dehydrogenases.

5. Wheat

Salinity:

Salt tolerant varieties possessed proteins of the mitochondria such as elevated levels of serine hydroxymethyltransferase and glycine decarboxylase. Other markers identified include aconitase, manganese superoxide dismutase among others.

Drought:

The biomarkers of drought were identified using iTRAQ : heat shock proteins, calnexin, P-ATPase and potassium channels at increased levels in the sensitive strains.

6. Chickpea:

Drought:
A cultivar that was challenged with drought showed the expression of biomarkers such as ascorbate peroxidase, GSH peroxidase, superoxide dismutase.

7. Tomato

Salinity:

2DGE analysis of salt induced stress in a tolerant variety was seen as changes in the level of heat shock proteins and peroxidases.

8. Barley

Salinity:

Several biomarkers found in a tolerant variety as deciphered by quantitative proteomics (2DGE) revealed glutathione-S-transferase, subunit of plastocyanin photosystem I, sedoheptulose 1,7-bisphosphatase, phosphoribulokinase as well as oxygen-evolving enhancer protein

Drought:

DIGE showed certain markers in the tolerant variety such as Elongation factor EF2, metalloprotease, HSP 70 while a sensitive strain showed levels of lipoxygenase, leucine aminopeptidase as well as betaine aldehyde dehydrogenase were at elevated levels.

9. Peanut

Drought:

Tolerant variety had increased levels of Acetyl-CoA carboxylase, 1L-myo-inositol-1-phosphate synthase and Lipoxygenase with a dip in carbon assimilation proteins and oxygen evolving enhancer protein 2 as well as RubisCo.

4.14. PROTEOMICS OF PLANTS FOR BIOMARKERS: AN ORGAN WISE APPROACH

With each part of a plant playing key roles; the analysis of each organ can reveal several biomarkers to be used in crop improvement, pathogen resistance or resistance to stress.

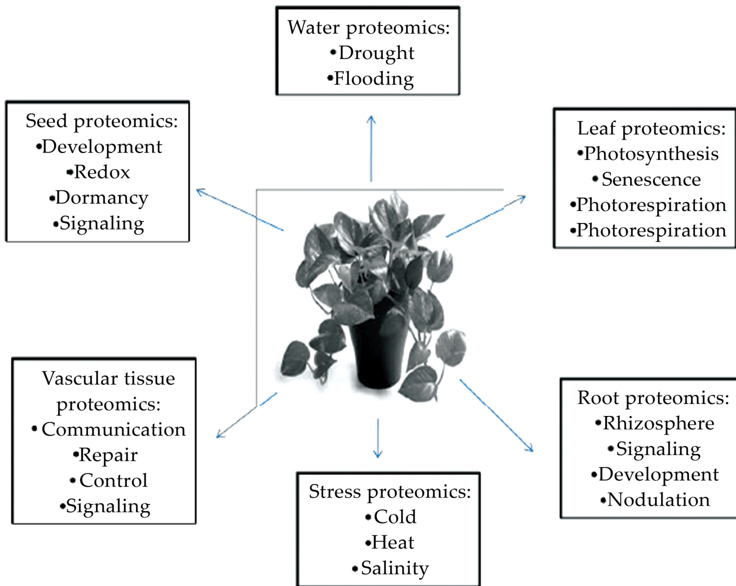


Figure 4.5. Plant organ proteomics for biomarkers.

4.15. BIOMARKERS: PROTEOMIC ANALYSIS OF SEEDLINGS OF CROPS

Germination is an intricate trait in plants which is responsible for the embryonic growth from the metabolically inactive seed to the next generation of plant; thus making germination in seeds an essential characteristic for the plant species to continue their race. Seeds revive with metabolism to germination on supplying them with the suitable environmental condition. Germination is influenced by many environmental factors and genes [30]. Research in physiology and genetics attest that the plant hormones gibberellins and abscisic acid have an important role in for regulating the seed dormancy followed by germination. Lately, studies in quantitative genetics and mutations have helped support additional and detailed genetic

dissection of the traits whilst aiding in the detection of novel components. Molecular strategies like proteome and transcriptome along with their expression are the innovative means to investigate the dormancy in seeds and its progression through germination [31]. In the initial growth phase, the crop seedlings are subjected to several abiotic conditions in the fields and that can result in crop failure or lower yields. Investigations in quantitative proteomics employed in wheat roots and soya bean seedlings were subjected to osmotic or flooding stresses. The studies depicted the metabolic pathways of flooding-responsive proteomes that responded to anoxia and the overload of water. Also, osmosis-related proteins were found to be reactive in stresses like cold, salinity and drought [32]. Ahsan and team performed proteomic investigations on the proteomes from soybean seedlings which are tissue-specific when exposed to heat stress. These studies signified the regular adaptive and defense means linked with the increased stimulation of various tissue-specific heat shock proteins and those participating in anti-oxidative resistance [33]. The presence of proteome in the early transformation in soybean root tip under flooding stress signified the role of calcium signaling in the initial responses. Exogenous calcium treatment studied in soybean roots post flooding stress revealed revitalizing outcome on proteins in hormone metabolism, protein synthesis and degradation, cell wall and DNA synthesis.

Salt-sensitive and salt-tolerant varieties of soybean [34] and wheat [35] were tested for proteomic analysis for salinity stress-responsive proteins found in the seedlings. Majority of the proteins related to salt-tolerance are recognized in wheat seedling roots.

Exposing wheat leaves for salt-stress tolerant genotype of durum wheat for proteomic investigations when made to undergo escalating salinity levels yielded chief transformation in proteins which are engaged in energy production, protein metabolism, primary metabolism, cellular defense and those leading to cell wall lignification's which permits significant growth recovery potential [36]. Proteomic investigations on soybean seedling leaves were tested for salt stress-responsive network of protein. Proteins accountable for redox homeostasis, accelerated proteolysis along with reduced activity of protein biosynthesis, enhanced ethylene biosynthesis, energy supply and impaired photosynthesis was also proposed [34].

Tolerance to salt stress during the initial seed growth and germination in durum wheat whole seeds and seed embryos along with its adjacent tissues were investigated by label-free quantitative analysis. This study

was performed to comprehend the function of ascorbate priming which occurs which boosts against salt stress. Results suggested pre-treatment with ascorbate avoids the consequences caused by salinity. This is done by altering profusion of proteins engaged in protein destination, storage categories and metabolism. This could be transformed by auxin, methionine and further metabolism in hormones and signaling systems [37]. These investigations shared details of the probable management strategy adopted by cellular activities in seedlings of wheat and soybean when exposed to salt-stress.

Ruan and team studied the germination phase in rice (*Oryza sativa* L.). In the initial growth phase of rice, when exposed to salt stress, the seedling roots, leaf blade and leaf sheath are all found to be extremely sensitive [38]. Studies by Yan, et al., showed that when rice roots were exposed to salt stress, there was disruption in the basic metabolism and enzyme activities, resulting in activation of energy production [39]. Delayed seed germination, reduced seed maturation and sluggish seedling development is observed in high salinity and this results in decline of rice growth and eventually decline in rice yield. Proteomic investigations on rice leaf blades, leaf sheath and roots were performed to decipher the abundance of salt-responsive proteins and results showed that these proteins altered according to the intracellular ion homeostasis due to extreme and incessant uptake of ions [40]. Research by Liu and team confirmed the relationship between the enzymes involved in metabolism and carbohydrates and the observation of elevated production of antioxidants which intervene in the continuation of homeostasis in cells [41]. Another study supported that through salt stress, proteins in rice seedling leaves expressed differentially and were played an essential role in the oxidative damage and photosynthetic metabolic processes [42].

Research in shoot stems of rice seedlings noticed the presence of an extracellular salt stress-responsive apo plastic protein network. This was suggestive of the particular proteins playing an important role in the initial sensitivity to salt stress like enzymes involved in intracellular equilibrium between ROS scavenging and ROS creation, carbohydrate metabolism and processing and degradation of proteins [43]. Transonic rice seedlings that over-expressed cyclophilin OsCYP2 confirmed their enhanced tolerance when put through salt stress. This experiment also proved an improved antioxidant enzyme activity and decreases in the peri oxidation of lipids which points at the role of OsCYP2, (which alters the behavior of antioxidant enzymes present at the translational level), handling oxidative damages [38]. Research on barley (Oryza sativa L.) confirmed it to be a salt-tolerant

crop and it reveals discrepancy in tolerance towards salt-stress. Proteomic analysis based on MS applications yielded results on salt-stress-responsive protein expression patterns and cultivar-specific. This study suggested that the proteins implicated in the expression of salt-tolerant genotype by in glutathione-detoxification of ROS whereas salt-sensitive genotype was by the proteins involved in iron uptake. The aim of the study was to find the function of protein when it is involved in ROS detoxification when exposed to salt-stress [44]. Studies on barley under salt stress also got to light numerous protein phosphatases and hormone-related kinases taking part in the defense mechanism alongside salinity. This study reveals essential data that will help improve the salt tolerance of cereals [45].

Proteomic research is also being carried out to study the effects of abiotic stress and environmental changes in wheat. This can be coupled with studies regarding water supply responsive proteins and their reaction to drought or extreme heat or cold or frost tolerance or heavy metal toxin stress [6]. Extended drought conditions have severe shock on the wheat crop yield. Access to moisture in the initial growing stage of wheat crop is also an essential criterion in the production of wheat.

Mature plant leaves of three varieties of wheat were investigated for the diversity illustrated by stress-responsive proteomes under drought intolerant and tolerant yields. An elevation was observed in proteins engaged in ROS scavenging. Also, down regulation was observed in proteins concerned with Calvin cycle and photosynthesis. These results were derived from all the three varieties of wheat along with the tolerant variety exhibiting major protein changes in the initial response and quantification in proteins acting in cell detoxification [46].

Alvarez and team studied the adaptation of two varieties of wheat to various environmental factors, drought sensitive and tolerant to estimate the protein expression patterns seen in effect to abscisic acid (ABA) on the proteins in roots. This study proved that the tolerant variety of wheat drastically had huge amounts of ABA-induced and ABA-responsive proteins, most of them responded to oxidation-reduction reactions and environmental stress, both of which play an essential part in adaptation to drought conditions [47]. Wheat seedlings were exposed to salicylic acid on induction to drought tolerance. This study detected the salicylic acid responsive protein interaction network that suggested resourceful defense systems, active anabolism, abundant energy supply and efficient photosynthesis [48]. Seedling leaves of two wheat bread cultivars were exposed to drought stress and investigated

using gel free TiO_2 phosohopeptide enrichment and label free relative quantitative approach; results of which showed trivial modifications in the phosphoproteome [49]. The majorly involved proteins were found to be engaged actively in stress/defense/detoxification, transcription/processing and drought tolerance, osmotic regulation and signal transduction.

Crop productivity is adversely affected by both abiotic and biotic stresses (Figure 4.6)

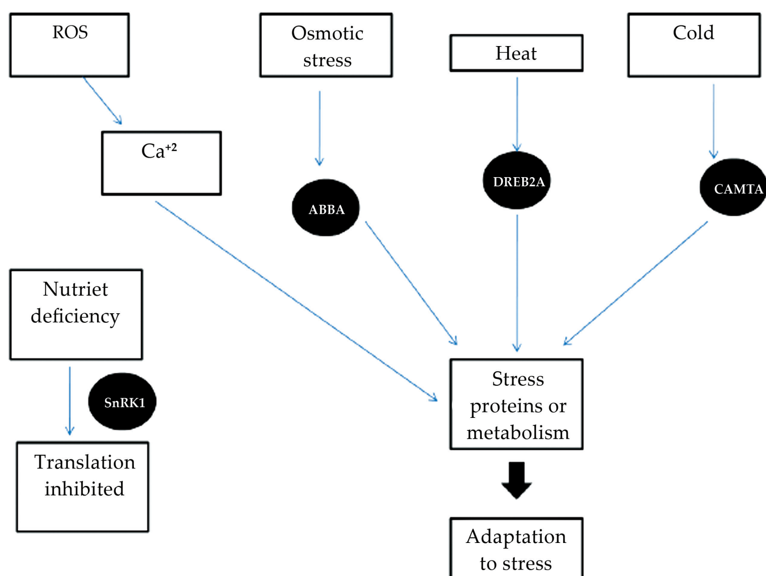


Figure 4.6. Overview of select factors involved in stress in plants.

Genomics-driven breeding of crops to achieve higher productivity and increase their tolerance to stress is dependent on the thorough understanding of mechanisms undertaken by plants to face climate changes. Adaptive measures have evolved in plants to survive these events. Exposure to these stresses brings changes in metabolism and thus leads to changes in growth of plant leading to effect on productivity. Various stresses faced by plants include drought, salinity, temperature stress, nutrient deficiency, UV-B radiation and pathogenic stress. Each type of stress can elicit different type of response in plants and some of these are enlisted below:

Salinization: Proteomics has been used to identify molecular pathways rendering salinity tolerance to many plants including cucumber, cotton and *Arabidopsis* [50, 51].

4.16. PROTEOMICS AND BIOMARKERS

Currently, proteomics has been widely employed in quantifying the differentially expressed genes under environmental stress in plants. In the light of post-transcriptional changes, translational changes and posttranslational changes, protein quantification lately is considered to be more reliable than transcript level changes. Many tissue specific and subcellular compartments of tissue specific proteins are identified under stress induced conditions.

Next step after specific protein identification is generation of biomarkers. To achieve this, streamlined bioinformatics analysis that reduces the rate of false positives and reliable validation of the high throughput data obtained are required. Western blot remains as a standard technique for validation of high-throughput techniques like 2DGE.

4.17. ROOT PROTEOMICS

Roots are considered to be an important organ for assessing abiotic stress. Roots are the organs which absorb water and nutrients from soil and transport them throughout the plant. During stress, structural and functional modifications are adopted by roots of the plants in an attempt to revive the balanced system. Some of the modifications observed include cell wall hardening, alterations in root length and metabolic changes [52].

Salinization/ soil salinity has emerged as a problem which affects the crop productivity. Seed germination, vegetative growth, flowering and fruit set are all affected by salt stress and thus results in economic loss. Plants affected by salt stress condition display osmotic stress, ion toxicity and reactive oxygen species (ROS) and other related damages. As adaptive measures to salt stress, accumulation of solutes and adjustment of ion concentration has been observed in plants. Many phytohormones and ROS-scavenging enzymes have been reported as salt tolerant measure in plants [53, 54]. Thus, timely assessment of risk due to salinity is required to save crop yield and quality. Some of the examples of use of proteomics in estimation of salinization response are discussed below:

- Li W. and colleagues had used isobaric tag for relative and absolute quantitation (iTRAQ)-based proteomic technique to unravel early differentially expressed genes (DEPs) from salt

treat Upland cotton (*Gossypium hirsutum* L.) roots [51]. One hundred and twenty-eight differentially expressed proteins were identified, majority of which had functions related to transcription, metabolism, membrane and transport, signal transduction and stress and tolerance. Validation of enhanced activity of some of the enzymatic proteins like superoxide dismutase (SOD), glutathione S-transferase (GST), malate dehydrogenase (MDH) was observed which also displayed increased protein levels under salt stress conditions. These candidate genes could be used to improve crop salt tolerance.

- Two-dimensional electrophoresis and liquid-chromatography-tandem mass spectrometry was used by Jiang Y. and colleagues to identify the differentially abundant protein in Arabidopsis roots exposed to 150 mMNaCl for either 6 h or 48 h. this study highlighted the importance of proteomic studies for identification of stress induced gene expression changes due to the poor correlation observed with the proteomic data obtained and the previously reported transcriptional level changes under the same conditions.
- iTRAQ analysis of radish (*Raphanussativus* L.) roots under exposure to different concentrations of NaCl (0,100 and 200 mM for 48 h) resulted in differential expression of proteins which were found to be enriched in processes like metabolic pathways and biosynthesis of secondary metabolites [55].

Recently, more advanced proteomic approaches have been developed to quantitate root proteomics. Cell level resolution for root proteomics has been achieved by development of laser capture microdissection (LCM) followed by gel-liquid chromatography-tandem mass spectrometry (GeLC-MS/MS)-based proteomics analysis [56]. Thin sections of root tips of tomato germinating seedlings were prepared. This was followed by identification of each cell type specific proteins on the section. (Figure 4.7)

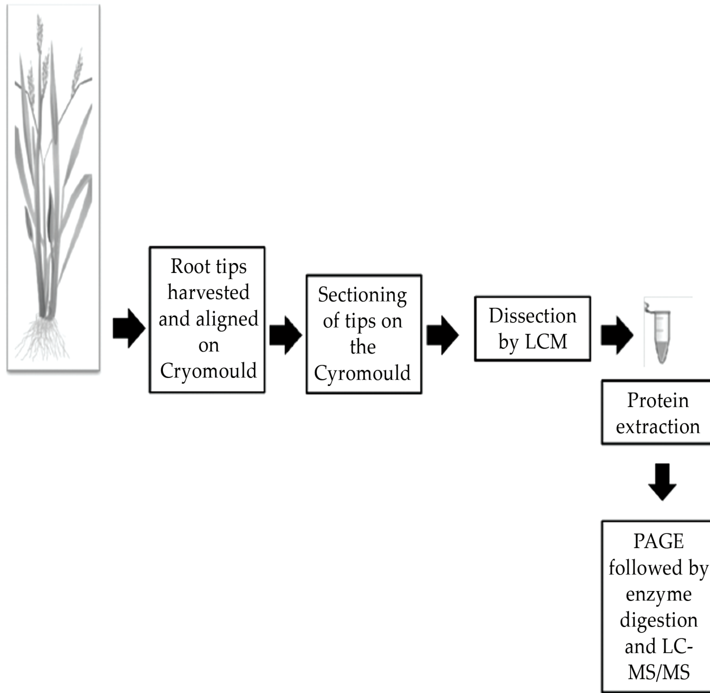


Figure 4.7. Root proteomics.

Proteomics case study: Root tissues are fixed and thin slices (10 µm) were obtained that were used to prepare single sections by Laser Capture Microdissection (LCM). Extracted proteins were subjected to liquid chromatography-tandem mass spectrometry (LC-MS/MS).

4.18. LEAF PROTEOMICS

Identification of alterations in protein levels of plant leaves due to both abiotic and abiotic stresses is demonstrated to highlight panel of biomarkers that can be assessed. A few studies are highlighted below:

- Drought stress was mimicked over 23 days of transpiration without watering for three-week old plants of rice (*Oryza sativa* L. cv CT9993 and cv IR62266) by Bennett J. and colleagues [57]. Differentially abundant proteins were identified from leaf extracts but only one protein; chloroplast Cu-Zn superoxide dismutase was found to behave in opposite fashion in both cultivars. 16 drought-responsive proteins were detected in stressed plants but

not in well-watered plants. S-like RNase homologue was found to be the most abundant protein which is upregulated in drought conditions. Moreover, four drought responsive mechanisms involving upregulation of actin depolymerizing factor, Rubisco activase and S-like RNase while downregulation of isoflavone reductase-like protein were unraveled.

- Fungus *Hemileiavastatrix*infects coffee plant (*Coffea arabica* L.) and causes coffee leaf rust (CLR) in these plants 58). This pathogen interacts with host cells and changes the metabolism to these cells to benefit it. Additionally, host immune suppression and gene expression changes are reported during such host-pathogen interactions. A proteomic study comparing the sensitive and resistant strain of *Coffea arabica* was undertaken and led to identification of glycohydrolases, proteases, and PR-proteins as proteins associated with pathogen response.
- Another example of leaf proteome alteration is by solar ultraviolet-B (UV-B) radiation on soyabean leaves. Xu et al. took two isolines different in their flavonoid content, leading to their difference in susceptibility to UV-B [59]. More proteins were altered in less flavonoid plants and most of them were involved in cellular energy reactions.

4.19. XYLEM AND PHLOEM PROTEOMICS

Xylem and phloem are part of the plant transport system, where xylem allows unidirectional water and nutrients flow, whereas phloem allows bidirectional flow of photosynthate and nutrients. Moreover, phloem sap is rich in molecules and also acts as conduit for exchange of information. Plants elicit a local and systematic stress response that can be transmitted through plasmodesmata (symplastic) and extracellular space (apoplastic). Few of the proteomics studies based on xylem and phloem sap are enlisted in Table 4.3.

- iTRAQ based proteomics approach had been used to identify seven hundred and forty-five genes as differentially expressed in cucumber phloem sap in response to salt. Salt stress sensitive and tolerant cultivars were used in the study and proteins involved in carbon fixation pathway were found to be decreased in sensitive cultivars while no significant protein level alterations were seen in tolerant cultivars indicating that stabilization of carbon fixation

and maintenance of energy and carbohydrates may render salt tolerance [50].

- Alterations in protein expression upon infection by *Fusarium oxysporum* in *Solanum lycopersicum* and *Brassica oleracea* were explored by Rep et al. [60] and Pu et al. [61], respectively. Proteins from infected and uninfected crops were compared and differentially expressed proteins were identified. Furthermore, *Fusarium* candidate effectors secreted in xylem were identified which could shed light on host pathogen interactions and hence pathogenesis.

Table 4.3. Selected proteomic studies on xylem and phloem sap

Species	Stress	Proteomics method	References
Solanum lycopersicum	*Fusarium oxysporum*	1D-DIGE	[60]
Brassica napus	*Verticillium longisporum*	1D-DIGE	[62]
Brassica oleracea	NaCl	2D-DIGE	[63]
Cucurbita maxima	Wounding	Isotope-coded protein labeling	[64]

4.20. SUMMARY

Approximately 30,000 to 80,000 of flowering plants in world offer edible parts to the world. It is believed that earlier 7000 species of plants were in cultivation, however at present only 150 species account for the plant products consumed [65]. Global climate changes results in drastic temperature fluctuations and also in rainfall. Aberrant rainfall patterns led to change in soil salinity. Moreover, due to limited availability of land available for crop production, agriculture may be done on nutrient poor soils. Biotic stresses like infection by pathogen also pose a serious threat to crop yield and quality. All these mentioned conditions demand a need for genetically improved crops which can tolerate combination of stresses present. This step may abrogate the decline in cultivated varieties and also the reduction in crop productivity observed due to the stresses. For the generation of stress tolerant varieties, identification of proteins uniquely differentially expressed in response to the stress are identified. These proteins may then serve as biomarkers and may prove helpful in improving the strain of crops to increase yield and make the existing species more tolerant of the stresses and competent to survive in poor soil conditions.

REFERENCES

1. Paton, A. J., Brumitt, N., Govaerts, R., Harman, K., Hinchcliffe, S., Allkin, B., & Lughadha, E. N. (2008). Towards target 1 of the global strategy for plant conservation: A working list of all known plant species-progress and prospects. *Taxon*. *57,* 602–611.
2. Esquinas-Alcázar, J. (2005). Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nature Review Genetics*. *6,* 946–953.
3. Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. *Nature Review Genetics*. *2,* 983–989.
4. Barkla, B. J. (2016). Identification of Abiotic Stress Protein Biomarkers by Proteomic Screening of Crop Cultivar Diversity. *Proteomes*. 4(3), 26.
5. Eldakak, M., Milad, S. I. M., Nawar, A. I., & Rohila, J. S. (2013). Proteomics: a biotechnology tool for crop improvement. *Frontiers in Plant Science*. *4,* 35.
6. Hu, J., Rampitsch, C., & Bykova, N. V. (2015). Advances in plant proteomics toward improvement of crop productivity and stress resistance. *Frontiers in Plant Science. 6,* 209.
7. Atkinson, N. J, & Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. *Journal of Experimental Botany*. *63,* 3523–3543.
8. Wu, C., Shi, T., Brown, J. N., He, J., Gao, Y., Fillmore, T. L., et al. (2014). Expediting SRM assay development for large-scale targeted proteomics experiments. *Journal of Proteome Research*. *13,* 4479–4487.
9. Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., et al. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular and Cellular Proteomics*. 11 O111.016717.
10. Uvackova, L., Skultety, L., Bekesova, S., McClain, S., & Hajduch, M. (2013). The MS(E)-proteomic analysis of gliadins and glutenins in wheat grain identifies and quantifies proteins associated with celiac disease and baker's asthma. *Journal of Proteomics*. *93,* 65–73.
11. Kamal, A. H., Cho, K., Kim, D. E., Uozumi, N., Chung, K. Y., Lee, S. Y., et al. (2012). Changes in physiology and protein abundance

in salt-stressed wheat chloroplasts. *Molecular Biology Reports. 39,* 9059–9074.

12. Shulaev, V., & Oliver, D. J. (2006). Metabolic and Proteomic Markers for Oxidative Stress. New Tools for Reactive Oxygen Species Research. *Plant Physiology. 141,* 367–372.
13. Kosová, K., Vítámvás, P., & Prášil, I. T. (2014). Proteomics of stress responses in wheat and barley—search for potential protein markers of stress tolerance. *Frontiers in Plant Science. 5,* 711.
14. Xu, K., & Nagy, P. D. (2010). Dissecting virus-plant interactions through proteomics approaches. Current Proteomics. 7(4), 316–327.
15. Alexander, M. M., & Ciliaa, M. (2016) A molecular tug-of-war: Global plant proteome changes during viral infection. *Current Plant Biology. 5,* 13–24.
16. Ventelon-Debout, M., Delalande, F., Brizard, J. P., Diemer, H., Van Dorsselaer, A., & Brugidou, C. (2004) Proteome analysis of cultivar-specific deregulations of Oryza sativa indica and, O. sativa japonica cellular suspensions undergoing rice yellow mottle virus infection. *Proteomics. 4,* 216–225.
17. Caplan, J. L., Zhu, X., Mamillapalli, P., Marathe, R., Anandalakshmi, R., & Dinesh-Kumar, S. P. (2009). Induced ER chaperones regulate a receptor-like kinase to mediate antiviral innate immune response in plants. *Cell Host Microbe. 6,* 457–469.
18. Perez-Bueno, M. L., Rahoutei, J., Sajnani, C., Garcia-Luque, I., & Baron, M. (2004). Proteomic analysis of the oxygen-evolving complex of photosystem II under biotic stress: Studies on Nicotiana benthamiana infected with tobamoviruses. *Proteomics. 4,* 418–425.
19. Zhu, J., Gopinath, K., Murali, A., Yi, G., Hayward, S. D., Zhu, H., & Kao, C. (2007). RNA-binding proteins that inhibit RNA virus infection. *Proceedings of the National Academy of Sciences. 104,* 3129–3134.
20. Osman, T. A., & Buck, K. W. (1997) The tobacco mosaic virus RNA polymerase complex contains a plant protein related to the RNA-binding subunit of yeast eIF-3. *Journal of Virology. 71,* 6075–6082.
21. De Angelis, M., Minervini, F., Caputo, L., Cassone, A., Coda, R., & Calasso, M. P., et al. (2008). Proteomic analysis by two-dimensional gel electrophoresis and starch characterization of Triticum turgidum, L. var. durum cultivars for pasta making. *Journal of Agriculture and Food Chemistry.* 56, 8619–8628.

22. Salt, L. J., Robertson, J. A., Jenkins, J. A., Mulholland, F., & Mills, E. N. (2005). The identification of foam-forming soluble proteins from wheat (Triticum aestivum) dough. *Proteomics*. 5(6), 1612–23.
23. Di Carli, T. M., Zamboni, A., Pè, M. E., Pezzotti, M., Lilley, K. S., Benvenuto, E., & Desiderio, A. (2011) Two-dimensional differential in gel electrophoresis (2D-DIGE) analysis of grape berry proteome during postharvest withering. *Journal of Proteome Research*. 10(2), 429–46.
24. Fasoli, E., D'Amato, A., Kravchuk, A. V., Citterio, A., & Righetti, P. G. (2011). In-depth proteomic analysis of non-alcoholic beverages with peptide ligand libraries. I: almond milk and orgeat syrup. *Journal of Proteomics*. *74,* 1080–1090.
25. Gil-Agusti, M. T., Campostrini, N., Zolla, L., Ciambella, C., Invernizzi, C., & Righetti, P. G. (2005). Two-dimensional mapping as a tool for classification of green coffee bean species. *Proteomics*. *5,* 710–718.
26. Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L., Deciu, C., Haynes, P. A., et al. (2002). Proteomic survey of metabolic pathways in rice. *Proc Natl Acad Sci USA*. *99,* 11969–11974.
27. Petersen, A., Dresselhaus, T., Grobe, K., & Becker, W. M. (2006). Proteome analysis of maize pollen for allergy-relevant components. *Proteomics*. *6,* 6317–6325.
28. Heick, J., Fischer, M., & Pöpping, B. (2011). First screening method for the simultaneous detection of seven allergens by liquid chromatography mass spectrometry. *Journal of Chromatography, A. 1218,* 938–943.
29. Kalluri, U. C., Hurst, G. B., Lankford, P. K., Ranjan, P., & Pelletier, D. A. (2009). Shotgun proteome profile of Populus developing xylem. *Proteomics*. 9 :4871–4880.
30. Arc, E., Galland, M., Cueff, G., Godin, B., Lounifi, I., Job, D., et al. (2011). Reboot the system thanks to protein post-translational modifications and proteome diversity: how quiescent seeds restart their metabolism to prepare seedling establishment. *Proteomics*. *11,* 1606–1618.
31. Koornneef, M., Bentsink, L., & Hilorst, H. (2002). Seed dormancy and germination. *Current Opinion in Plant Biology.* 5(1), 33–36
32. Nanjo, Y., Nouri, M. Z., & Komatsu, S. (2011). Quantitative proteomic analyzes of crop seedlings subjected to stress conditions; a commentary. *Phytochemistry*. *72,* 1263–1272.

33. Ahsan, N., Donnart, T., Nouri, M. Z., & Komatsu, S. (2010) Tissue-specific defense and thermo-adaptive mechanisms of soybean seedlings under heat stress revealed by proteomic approach. *Journal of Proteome Research*. *9,* 4189–4204.

34. Ma, H., Song, L., Shu, Y., Wang, S., Niu, J., Wang, Z., et al. (2012). Comparative proteomic analysis of seedling leaves of different salt tolerant soybean genotypes. *Journal of Proteomics*. *75,* 1529–1546.

35. Guo, G., Ge, P., Ma, C., Li, X., Lv, D., Wang, S., et al. (2012). Comparative proteomic analysis of salt response proteins in seedling roots of two wheat varieties. *Journal of Proteomics. 75,* 1867–1885.

36. Capriotti, A. L., Borrelli G. M., Colapicchioni V et al. (2014). Proteomic study of a tolerant genotype of durum wheat under salt-stress conditions. *Analytical and Bioanalytical Chemistry*. *406,* 1423–1435.

37. Fercha, A., Capriotti, A. L., Caruso, G., et al. (2014). Comparative analysis of metabolic proteome variation in ascorbate-primed and unprimed wheat seeds during germination under salt stress. *Journal of Proteomics*. *108,* 238–257.

38. Ruan, S. L., Ma, H. S., Wang, S. H., et al. (2011). Proteomic identification of OsCYP2, a rice cyclophilin that confers salt tolerance in rice (Oryza sativa, L.) seedlings when overexpressed. *BMC Plant Biology*. *11,* 34

39. Yan, S., Tang, Z., Su, W., & Sun, W. (2005). Proteomic analysis of salt stress-responsive proteins in rice root. *Proteomics. 5,* 235–244.

40. Abbasi, F. M., & Komatsu, S. (2004). A proteomic approach to analyze salt-responsive proteins in rice leaf sheath. *Proteomics. 4,* 2072–2081.

41. Liu, D. W., Ford, K. L., Roessner, U., et al. (2013). Rice suspension cultured cells are evaluated as a model system to study salt responsive networks in plants using a combined proteomic and metabolomic profiling approach. *Proteomics*. *13,* 2046–2062

42. Kim, D. W., Rakwal, R., & Agrawal G. K. (2005). A hydroponic rice seedling culture model system for investigating proteome of salt stress in rice leaf. *Electrophoresis*. *26,* 4521–4539.

43. Song, Y., Zhang, C., & Ge, W. (2011). Identification of NaCl stress-responsive apoplastic proteins in rice shoot stems by 2D-DIGE. *Journal of Proteomics*. *74,* 1045–1067.

44. Witzel, K., Weidner, A., & Surabhi, G. K. (2009). Salt stress-induced alterations in the root proteome of barley genotypes with contrasting

response towards salinity. *Journal of Experimental Botany*. *60,* 3545–3557.

45. Gao, R., Duan, K., & Guo, G. (2013). Comparative transcriptional profiling of two contrasting barley genotypes under salinity stress during the seedling stage. *International Journal of Genomics*. *139,* 822–835.
46. Ford, K. L., Cassin, A., Bacic, A. (2011). Quantitative proteomic analysis of wheat cultivars with differing drought stress tolerance. *Frontiers in Plant Science*. *2,* 44.
47. Alvarez, S., Roy Choudhury, S., & Pandey, S. (2014) Comparative quantitative proteomics analysis of the ABA response of roots of drought-sensitive and drought-tolerant wheat varieties identifies proteomic signatures of drought adaptability. *Journal of Proteome Research*. *13,* 1688–1701.
48. Kang, G., Li, G., Xu, W., et al. (2012). Proteomics reveals the effects of salicylic acid on growth and tolerance to subsequent drought stress in wheat. *Journal of Proteome Research*. *11,* 6066–6079
49. Zhang, M., Lv, D., Ge, P., et al. (2014). Phosphoproteome analysis reveals new drought response and defense mechanisms of seedling leaves in bread wheat (Triticum aestivum, L.). *Journal of Proteomics*. *109,* 290–308.
50. Fan, H., Xu, Y., Du, C., & Wu, X. (2015). Phloem sap proteome studied by iTRAQ provides integrated insight into salinity response mechanisms in cucumber plants. *Journal of Proteomics*. *125,* 54–67.
51. Li, W., Zhao, F., Fang, W., et al. (2015). Identification of early salt stress responsive proteins in seedling roots of upland cotton (Gossypium hirsutum L.) employing iTRAQ-based proteomic technique. *Frontiers in Plant Science*. *6,* 732.
52. Gowda, V. R. P., Akira Yamauchi, A. H., Shashidhar, H. E., & Serraj, R. (2011). Root biology and genetic improvement for drought avoidance in rice. *Field Crops Research*. *122,* 1–13.
53. Aghaei, K., Ehsanpour, A. A., & Komatsu, S. (2009) Potato responds to salt stress by increased activity of antioxidant enzymes. Journal of integrative plant biology *51,* 1095–1103.
54. Deinlein, U., Stephan, A. B., Horie T et al. (2014) Plant salt-tolerance mechanisms. Trends in Plant Science. *19,* 371–379.
55. Sun, X., Wang, Y., Xu, L. et al. (2017) Unraveling the Root Proteome

Changes and Its Relationship to Molecular Mechanism Underlying Salt Stress Response in Radish (*Raphanus sativus*, L.). *Frontiers in Plant Science*. *8,* 1192.

56. Zhu, Y., Li, H., Bhatti, S., Zhou, S., Yang, Y., Fish, T., & Thannhauser, T. W. (2016). Development of a laser capture microscope-based single-cell-type proteomics tool for studying proteomes of individual cell layers of plant roots. *Horticulture Research*. *3,* 16026.
57. Salekdeh, G. H., Siopongco, J., Wade, L. J., Ghareyazie, B., Bennett, J. (2002). Proteomic analysis of rice leaves during drought stress and recovery. *Proteomics*. *2,* 1131–1145.
58. Guerra-Guimaraes, L., Tenente, R., & Pinheiro, C. (2015). Proteomic analysis of apoplastic fluid of Coffea arabica leaves highlights novel biomarkers for resistance against Hemileia vastatrix. *Frontiers in Plant Science*. *6,* 478.
59. Xu, C., Sullivan, J. H., Garrett, W. M., Caperna, T. J., & Natarajan, S. (2008). Impact of solar ultraviolet-B on the proteome in soybean lines differing in flavonoid contents. *Phytochemistry. 69,* 38–48.
60. Rep, M., Dekker, H. L., Vossen, J. H., et al. (2002). Mass spectrometric identification of isoforms of PR proteins in xylem sap of fungus-infected tomato. *Plant Physiology*. *130,* 904–917.
61. Pu, Z., Ino, Y., Kimura, Y., et al. (2016). Changes in the Proteome of Xylem Sap in Brassica oleracea in Response to Fusarium oxysporum Stress. *Frontiers in Plant Science*. *7,* 31.
62. Floerl, S., Druebert, C., Majcherczyk, A., et al. (2008). Defence reactions in the apoplastic proteome of oilseed rape (Brassica napus var. napus) attenuate Verticillium longisporum growth but not disease symptoms. *BMC Plant Biology*. *8,* 129.
63. Fernandez-Garcia, N., Hernandez, M., Casado-Vela, J., Bru, R., Elortza, F., Hedden, P., & Olmos, E. (2011). Changes to the proteome and targeted metabolites of xylem sap in Brassica oleracea in response to salt stress. *Plant, cell and environment*. *34,* 821–836.
64. Gaupels, F., Sarioglu, H., & Beckmann, M. (2012). Deciphering systemic wound responses of the pumpkin extrafascicular phloem by metabolomics and stable isotope-coded protein labeling. *Plant physiology*. *160,* 2285–2299.

65. Esquinas-Alcazar, J. (2005). Science and society: protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nature reviews Genetics*. *6,* 946–953.

CHAPTER 5

BIOMARKERS FOR ANIMALS BY PROTEOMICS

CONTENTS

5.1. Introduction 152
5.2. Health of Farm Animals: Biomarkers 155
5.3. Proteomics And Veterinary Research 156
5.4. Proteomics Techniques For Animal Biomarker Discovery 157
5.5. Sample Preparation 161
5.6. Points To Be Considered In The Terms of Animal Biomarkers 161
5.7. Veterinary Research And Biomarkers 162
5.8. Pathogen Proteomics 163
5.9. Avian Proteomics 163
5.10. Proteomics In Aquaculture 164
5.11. Parasite Proteomics 165
5.12. Animal Proteomics In Different Tissues 166
5.13. Overview of Serum Proteomics In Farm Animals 168
5.14. Biomarkers Using Proteomics of Animal Products 173
5.15. Case Study of Biomarkers: Targeting The Mosquito 175
5.16. Scenario/Future For Animal Proteomics 177
5.17. Summary 179
References 180

5.1. INTRODUCTION

The proteomics is being referred to as the study of protein profiling of a given tissue and/or specimen fluid, which is identified as the proteome. The field is being recently acknowledged to be significantly important to numerous scientific areas, like veterinary science, human diseases and therapeutic applications, etc. However, the field is largely limited, despite its wide application; due to many reasons, which include cost of analysis, unavailability of authentic genomic data from many species and lack of knowledge as well as awareness about its applications. Accordingly, the chapter is intended to offer some valid examples of successful application of the said technology, especially in the most ignored areas, like animal production and health.

The chapter will as well provide insights into proteomics research based on farm animals, allowing directed progression of the field towards its significant contribution for better maintenance of farm animals.

The products obtained through farm animals as well as from the aquaculture industry, like meat, poultry, milk and milk products, etc. can be utilized as the best sources of proteins in food and are identified to be very health for human consumption. Moreover, the products are helpful in contributing towards the balanced diet programme in majority of the population, across the globe.

It is very surprising to note that these products, despite being important sources of protein supplements are neglected so far, by the most advanced protein technology; the proteomics.

This is facilitated by the formation of Farm Animal Proteomics (FAP) that deals with the impact of food on farm animals [1].

The proteomics has wide applications in the areas of health of animals as well as boosted production of products. In fact, the technology can as well be used to characterize:

- Associations between host and pathogen.
- The reproductive health.
- Assess the dynamics of their muscular growth; for quality production.

Moreover, proteomic investigations have as well found to have significant role in

- post-harvest modifications.
- identification in the fish muscle alterations [2].
- meat production in relation with safety of food [3].

However, it is important to increase the availability of expertise and early stage researchers in animal science, who are exposed to this technology; to exploit full range of applications of proteomic investigation. Thus, their aspirations towards knowledge and interest in benefitting animal science should be recognized and facilitated.

Currently, the science of proteomics has been advanced to the extent of separating and identifying the proteins from a complex mixture of biological sample.

The resolution of complex mixtures is done by:

- at the protein level, i.e., top down approach.
- after protein mixture is being digested into peptides, i.e., bottom down approach.

As explained in the figure below, for the first strategy two-dimensional gel electrophoresis (2DE) is applied after successful reduction and denaturation of the peptide fragments; in order to minimize protein-protein interactions, allowing maximized separation of protein subunits according to charge (isoelectric point) in the first dimension.

This is followed by separation on the basis of mass in a second dimension. The protein fragments are confirmed and evaluated with the help of staining through visible/colorimetric or fluorescent dyes, along with the application of dedicated software. Alternatively, protein spots can be separated with the help of modern techniques like pre-electrophoretic protein labeling with the help of fluorescent dyes, also known to be as fluorophores [4]. Furthermore, protein spots that are usually variable depending upon the treatment intensity and disease severity can be enzymatically separate into different peptide fragments, to allow easy separation on the basis of size and fragmentation. This can be achieved through MS analysis to maintain computer based searchable databases. (Figure 5.1)

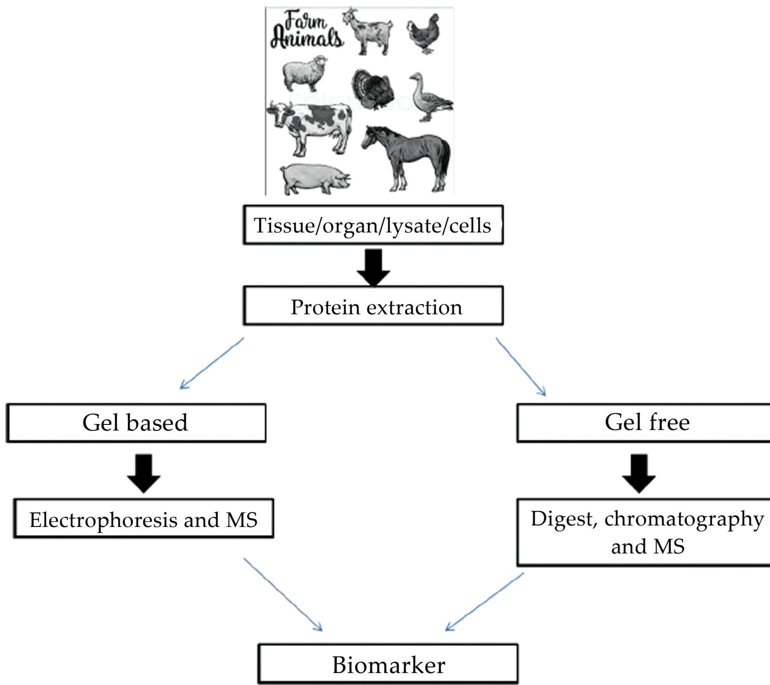


Figure 5.1. A schematic representation of the proteomics workflow.

The study of proteins in proteomics is being observed through one of the two approaches, like gel based separation and gel free separation. The gel based separation is being achieved through two dimension electrophoresis, which further allows digestion of particular peptide in multiple protein fragments and their separation; the protein is being quantified with the help of MS. Whereas, in the gel free approach, the entire protein extract is being digested with peptidases, especially trypsin and the fragments are being separated through chromatography, further to which the proteins of interest are being identified and quantified using high-throughput MS instrument.

Out of the two approaches, it is believed that the later approach is particularly suitable for species like cattle, pig, sheep, chicken and salmon; with reasonable coverage level in database.

In another approach using protease digestion, peptides generated from proteins are separated by chromatography. Generally, the technique of separation is often based upon the multidimensional set up along with different separation principles, like ion exchange chromatography, reverse phase chromatography and/or affinity chromatography, etc.

As far as the quantification of availability of protein is concerned, it can be achieved through MS depending upon isotopic as well as chemical labels that are introduced into the organism at the cellular level through labeled proteins and peptides, as well as label free mode that requires extensive reproducibility of results [5].

It should be understood that variations exist even between gel based as well as gel free approaches, which is generally based upon the question under investigation. These separation techniques are as well combined with the sample pre-fraction techniques to reduce sample complexity.

Studies have further elaborated how complementary results can be obtained through exploitation of both these strategies and have as well focused upon the importance of structural protein monitoring and modifications due to glycosylation, phosphorylation, etc.; besides analyzing the changes in the concentration of proteins [1].

5.2. HEALTH OF FARM ANIMALS: BIOMARKERS

Proteomics is to be emphasized as the advanced and emerging technology to facilitate sustainable production of different animal produces, without compromising quality as well as welfare. Although, currently more focus is being established on porcine and bovine species to identify biomarkers in order to predict the quality of food, animal condition as well as further detection and diagnosis of different infectious diseases; recent activities have as well diverted their attention towards poultry and fisheries. Through understanding the economic importance of the dairy farming; many of the studies related to structural as well as functional analysis of protein in bovine have been carried out to determine pathological attack and analyzing pathophysiological changes of mastitis and endometritis. The levels of stress in an animal can be facilitated by the proteomics in order to maintain the welfare of animals.

Over the years, there have been major changes associated with the animal production, such as but not limited to housing conditions, increased animal origin food production, forceful induction of animals in order to meet the increasing demands with a range of non-specific response elicited in the body, known to be as the stress response [6].

In dairy stress related to the cows is often being characterized as increased susceptibility to many infectious diseases like mastitis, John's

disease, salmonellosis and many other respiratory disorders; which can together affect production of animal products, fertility and overall health condition of an animal [7].

Mastitis has been studied using approaches of proteomics such as 2 D gel electrophoresis as well as MS or liquid chromatography along with tandem MS [1].

The techniques can be effective in analysis of systemic modifications of milk proteins during diseased condition, like cow mastitis with naturally occurring infectious attacks [1].

An increase in albumin and α-lactoglobulin, β-lactoglobulin in whey was seen in cows with the disease [8]. The study further suggests failure of blood milk barrier to maintain separation of serum protein with milk.

A study in 2007 identified several biomarkers from milk of cows suffering with mastitis. The markers were acute-phase protein serum amyloid A, cathelicidin-I as well as apolipoprotein A-I [9].

Through yet another study, a marker under the family of acute phase protein, named to be -1-acid-glycoprotein, has been identified in both the animals, i.e., with and without mastitis; this identification was done through whey samples obtained from both control as well as test groups animals that have been experimentally inoculated with *E. col*i [10].

5.3. PROTEOMICS AND VETERINARY RESEARCH

Majority applications in research look at the use of rodents such as mice that are reported as not suitable or appropriate. Diseases of animals can also affect health of humans as there are diseases borne by vectors or zoonosis [11].

Characters in a veterinary setting:

Based on the use of a biomarker, there are several characters of biomarker expected [12]. They are summarized below:

- Accuracy
- It should be changes in a disease setting.
- Sensitive for a particular condition/disease.
- Available in body fluids of animals such as plasma, urine, sweat and saliva.

- It should not be affected by unrelated disorders
- The estimation should be reliable.
- The biomarker should not have different concentrations across individuals in a population.
- Reproducibility of results
- Easy interpretation of results

Following identification, validation is of importance so that it can indicate a process or a disease. There are several sources of variations that should be accounted for in animal models of biomarkers that are summarized below:

Biological:

- Sex
- Age
- Diurnal variation
- Status of neuter
- Hormone profile
- Species of animal
- Handling of the animal

Analytical

- Sample used
- Collection of sample
- Sensitivity of assay
- Temperature of storage
- Time of storage [11].

5.4. PROTEOMICS TECHNIQUES FOR ANIMAL BIOMARKER DISCOVERY

A ubiquitous approach in the profiling in proteomics is 2-dimensional gel electrophoresis (2-DGE). The use of peptide mass fingerprinting (PMF) can facilitate the analysis of proteins that were separated by the technique. A challenge in this technique is a paucity of sequences that are completely annotated in genomes. The technique facilitates the study of proteins that are conserved however, a change in an amino acid can cause an alteration

in the peptide mass fingerprint. The use of liquid chromatography-tandem mass spectrometry (LC-MS/MS) can enable the identity of novel candidates with accuracy. (Figure 5.2).

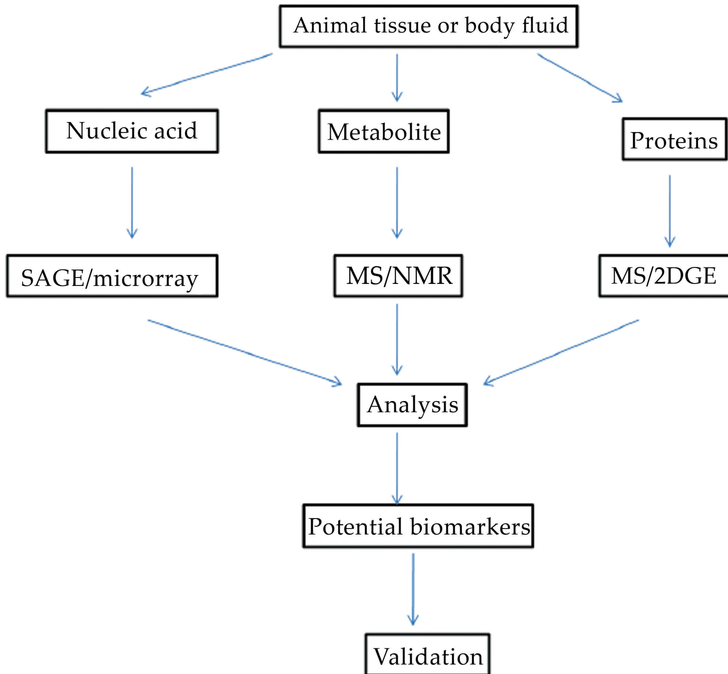


Figure 5.2. Approaches for biomarker discovery in animals.

The variation due to different gels can be minimized by the use of difference gel electrophoresis (DIGE). The use of 'Shotgun proteomics' can aid the study of complex protein mixtures using multi dimensional protein identification technology (MudPIT) based on MS [11].

5.4.1. Gel-Based Proteomics

The status of a protein can be facilitated by the use of methods based on gels. Proteins that have different functions can contain identical peptides; hence a gel based approach can yield information on cleaved/ uncleaved proteins. However, gel free approaches can yield s same pattern. The use of difference gel electrophoresis or difference in gel electrophoresis aids in the processing of 2-3 samples as well as statistics by the use of an internal standard.

The application of precast gels can facilitate the gel handling with ease except for the voluminous quantity of buffer required.

The gel based approach can aid in the study of:

- Levels of proteins at different conditions.
- PTMs can be detected.
- Complexes or native proteins can be separated.
- The employment of antibodies can facilitate the result validation.

5.4.2. Shotgun LC-MS

The approach uses discovery approach to study biomarkers using alternate methods using Label-based or label-free techniques. The label based methods can be applied when analysis can be applied for samples that involve enrichment or prefractionation. The label free techniques are suitable for studies involving multiple samples and require a standardised approach as well as quality control checks at the steps of:

- Protein extraction.
- Digestion of proteins.
- LC-MS [13].

During the analysis of MS: there are two approaches that include:

Data-dependent acquisition:

It involves fragmentation of each ion generated in tandem MS to obtain information about sequence. The hurdles in this approach are:

the speed of the tandem MS.

Reproducibility is compromised with loss of data when multiple assays are combined.

Data-independent acquisition:

In order to handle these challenges, this acquisition involves tandem MS in the absence of particular precursor. Thus, all the data can be acquired with no compromise in reproducibility and the lack of the bias introduced in the data dependency.

5.4.3. Targeted Proteomics in Farm Animal Research

The shotgun approach discussed above has certain reports of its challenges in analyzing proteins that are not abundant [14]. In milk samples of cows, key vital proteins such as cytokines or their receptors have not been detected by such shotgun approaches [15]. The validation of biomarkers necessitates the use of absolute quantification especially in such veterinary medicine.

The validation is being approached by Selected reaction monitoring (SRM). This technique is based on tandem MS analyzes to quantify the level of choice peptides of the protein of interest. The instruments use triple quadruple instruments as these analyzers are more selective and sensitive. The LC-SRM/MS analysis can run more than 100 peptides (corresponding to up to 50 peptides). This increases the base and range of the analysis as compared to the shotgun approach. The levels of proteins as absolute values can be measured by SRM that involves the introduction of peptides of heavy amino acids to measure the extent of such peptides in samples.

The examples of SRM in farm animals include:

Measure the levels of membrane proteins in globules of milk fat.

The response of proteins in the host of udder of cows in controls and cows challenged with lipopolysaccharide.

A source of observed peptides for this approach involves the use of The PeptideAtlas repository (www.peptideatlas.org).

- In this resource there are 1921 proteins that comprise proteomes of milk and mammary gland as well as immune cells and joints that are inflamed.
- PeptideAtlas contains more than 8000 peptides representative of more than 20 cells.
- In horses, more than 2600 proteins

5.4.4. Quantification concatemers (QconCAT)

The technique involves the design of a chimeric protein made of tryptic standard peptidesthat are ^{13}C-labeled Arg and Lys in order to measure the quantities of peptides. A 2012 study evaluated host response proteins in

mastitis in cows. The technique aided in the calculation of 17 of the proteins of cow proteins in a multiplex format without the use of antibodies [16].

5.5. SAMPLE PREPARATION

A major source in studies is plasma or serum; however, a major challenge is higher levels of proteins such as IgG or albumin. In order to remove these over expressed proteins, there are techniques such as:

Immuno-depletion removes proteins at high levels but a challenge is the cross reactivity of antibodies with that of animal species or Protein Equalizer Technology that can aid in overcoming the complex samples across species [11].

5.6. POINTS TO BE CONSIDERED IN THE TERMS OF ANIMAL BIOMARKERS

- The use of replicates aid to account for variability at technical and biological grounds.
- Studies in farm animals have smaller quantities of samples that require statistics.
- Quantity of sample is vital as well as source of the sample.
- The level of protein content per cell along with the cell number is required for cell cultures.
- The level of proteins in biofluids is to be quantified.
- Homogenous samples are essential for tissue samples.
- Proteomics can entail the use of microdissection for ideal samples.
- The use of inhibitors can suppress enzymes that are endogenous.
- Internal standards can aid in reproducible results in the form of housekeeping proteins or labeling agents or exogenous proteins.
- The application of chemical modification such as alkylation, labeling, reduction of disulfide or the use of proteases requires a prior step of denaturation or solubilization [1] (Figure 5.3)

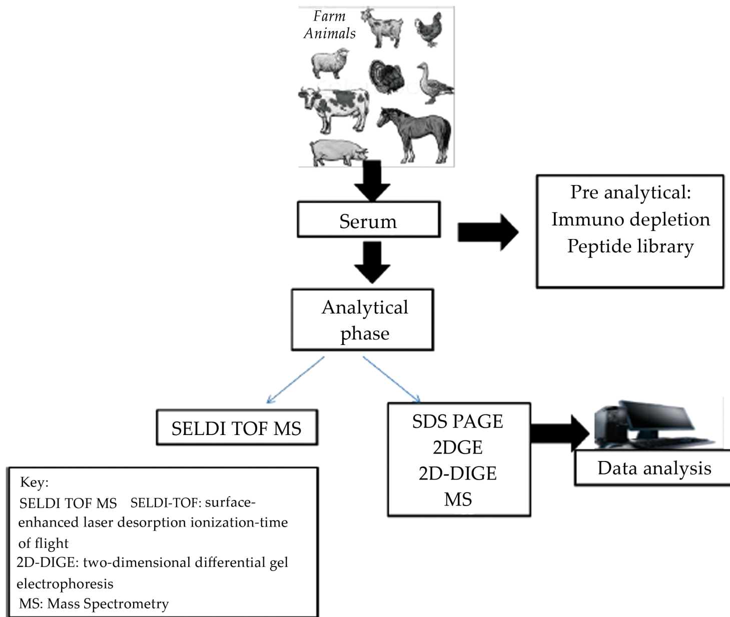


Figure 5.3. Overview of proteomics in animal biomarker.

5.7. VETERINARY RESEARCH AND BIOMARKERS

5.7.1. Bovine Peripartum Health Prognosis

The expression of proteins during the process of pregnancy as well as peripartum time in cows was studied in 2006. The serum samples of Friesian heifers that had been inseminated artificially were collected and subjected to gel electrophoresis and mass spectrometry. The time of calving showed alterations in the level of haptoglobin and orosomucoid/α1-acid glycoprotein.

The level of orosomucoid was lesser in cows with postpartum endometritis as compared to healthy controls. Hence, a potential biomarker of postpartum health in cows can be the concentration of serum orosomucoid [17].

5.7.2. Respiratory Health of Pigs

The pathogens that cause respiratory tract infections in pigs can persist in animals that are convalescing. A 2006 study evaluated the protein profile following infection with *Actinobacillus pleuropneumoniae*in pigs. The

bronchoalveolar lavage fluid (BALF) were analyzed by 2-DGE followed to analyze following infection of pigs. The samples that were infected showed several spots that were different as well as 8 proteins that were at different levels following three weeks of infection. Of these, 3 were identified as PR-39, calgranulin C and prophenin-2. Of these, the first was found to be at high concentrations in proportion to lesions formed in the lungs. Thus, it can be a potential biomarker of lung infection [18].

5.8. PATHOGEN PROTEOMICS

The technology can put novel insights in the pathogenic mechanism of bacterial infections, leading to animal diseases; accordingly, it can offer unique opportunity to study bacterial pathogenic proteome, during infection. In this regard, very limited number of data is available related to the proteomic analysis of pathogenic responses, during indications like clinical intramammary infections. A 2011 study reported 15 proteins that were expressed in isolates of *S. aureus* isolates [19].

2D electrophoresis was as well being applied to analyze and further detect virulent state of a bacterium species, *M.avium*subsp. Paratuberculosis. Direct comparative analysis when carried out, it was evident that 10 biomarkers can possibly be implied for disease confirmation and marker detection [20].

The application of advanced technologies of proteomics can facilitate the quick diagnosis of several diseases

Accordingly, similar to bovine species, a substantial amount of proteomic studies are being performed even on the porcine species. As a matter of fact, porcine species have proven their importance in being ideal animal models; as well as in effective and healthy meat production [1].

5.9. AVIAN PROTEOMICS

Along with analysis of pathogens affecting farm animals, it is as well very interesting to know about pathogenic infections of avian species. The interest can be variable such as the economic aspects, reduction in the impact of avian pathogenic diseases through identification of new markers, production of effective vaccines and investigation on some other avian diseases, like avian flu, etc. [1].

5.10. PROTEOMICS IN AQUACULTURE

Although, aquaculture was operational, since centuries; with the latest advancements and extensive upgradation of technology the industry is currently experiencing rapid growth in average seafood production, in the last 50 years. Moreover, due to increasing awareness about the benefits of fish related products, there is a growing demand of seafood consumption per person; to fulfill this rising demand, it is very crucial to improve production facility by improving growth rates, conversion of feed, directly into the muscles, resistance to the pathogenic attack, improvement in the reproductive age of the animals, etc. However, the industry has to as well look into environmental sustenance, as one of the primary challenges in its establishment. In this regard, application of proteomics can be offered to allow further improvement in animal nutrition, welfare and health management and screening of infectious diseases; in order to support the effective production of fish related products [21].

Traditionally, marine harvested resources were used as the primary source of nutrition to fish farming; however, recently, due to upgraded technology, the trend is moving towards the use of diet containing vegetable proteins and oil resources. However, it has been analyzed that the use of these products are impacting the growth rates and feeding efficiency of fishes; along with reducing the impact of marine based food source. In this regard, proteomics is contributing greatly towards better understanding of the pathophysiology as well as metabolic pathways that have been significantly affected due to dietary changes; as evident in species like:

- Gilthead seabream
- rainbow trout.
- Atlantic Salmon
- *Diplodus sargus*

Several pathways of metabolism of carbohydrate, lipid and proteins were studied. Fish mortality due to increased disease association is the leading cause of economic loss in aquaculture. The diseases can mainly be viral, pathogenic as well as parasitic in nature; significantly reducing the global production.

To further analyze the causative agents, several pathogenic detection systems have been employed like traditional, immunological, molecular, etc. and have been extensively studied; with an aim to be vaccination for disease prevention.

Fortunately, proteomics can mainly assist in solving the problem to the great extent, especially at the level of formulation and the development of new vaccines, along with accurate diagnosis of the disease. Recently, studies have demonstrated the isolation as well as analysis of proteomic profiling of the envelope proteins, of the pathogen Iridovirus, which is well known for its association with high mortality in cultured group as well as Southeast Asian farmed species [22].

Proteomics as well be noted as the extremely important tool in assessing fish welfare, through the development as well as implementation of novel aquaculture practices; thus, allowing better environment for farmed marine animals in order to optimize their efficiency to cope up with unavoidable challenges/stress, and effectively modify their state of welfare and health. It has been estimated in many clinical analyzes that the primary target organ to be analyzed is the liver. Since, it reflects the metabolic status of an animal; other than which, an easily retrievable body fluid like plasma can as well be focused.

The role of stress and proteome can involve association of handling, stress prior to slaughter or high stocking.

5.11. PARASITE PROTEOMICS

The field of parasitology has great applications to use many advanced proteomics technologies. As being evident, the host–parasite interaction is highly complex process, due to involvement of two genetically distinct multicellular biological systems. The investigations are being progressed into two directions. On one hand, the study is intentional to identify parasitic expression, which is especially tedious due to their different developmental stages, and secondly about understanding the host pathophysiology, by focusing on the complex dynamics of various host-parasitic immune interactions.

PTMs (post-translational modifications) are used by antigens of the parasites to manipulate the immune system of the host. PTM can facilitate the study of processes including development and differentiation [23].

Accordingly, the advancements in immunoproteomic are being helpful in offering new insights into the interaction of parasites through host perspectives, with respect to particular biomarker discovery.

5.12. ANIMAL PROTEOMICS IN DIFFERENT TISSUES

5.12.1. Liver Proteomics

With the liver playing an important role in metabolism as well as detoxification the proteomics of this organ is of key interest [24].

5.12.2. Studies Reported

Cattle show a disorder of metabolism namely ketosis with elevated levels of ketone bodies and low levels of glucose in the blood. This leads to mobilization of fat and proteins of muscles [24]. As a response to decreased feed intake, global proteome changes were visualized as Fifty-nine different proteins of which thirty five were reported as proteins of lipid metabolism. The markers identified included lower levels of β hydroxyacyl co-A dehydrogenase type-2, Acyl co-A acyl transferase and Acyl co-A dehydrogenase [25].

A 2012 study evaluated the changes in liver and mammary glands during lactation in cows. Biomarkers were identified using 2-DE and mass spectrometry to reveal Lactalbumin, β and κ-casein, β – Lactoglobulin A & B among others to be expressed in the mammary gland while the liver showed four times the level of Pyruvate carboxylase than the mammary gland. This is suggestive of the conversion role of liver in metabolic pathways such as conversion of glucose [26].

The breakdown of fats is mainly seen in the liver in chickens as opposed to adipose tissue in mammals. A study in 2008 studied the proteome of laying hens at various ages namely 0, 10, 21, 32 weeks. It was found that as the age of a hen increases, the levels of certain enzymes of glycolysis such as triose phosphate isomerase, enolase were decreased. The levels of Fatty acid synthase and malic enzyme were increased suggestive of changes in profiles with age i.e., a shift from glycolysis to lipogenesis [27].

A team in 2014 published the proteomics of fat and thin ducks. The analysis of proteins of liver of pekin ducks (*Anas platyrhynchos domestica*) revealed differences in 76 proteins of the liver. The lean ducks showed biomarkers such as breakdown of proteins causing lean meat. The fat ducks showed biomarkers such as protein catabolism, synthesis of ATP and glycolysis. This difference in biomarkers can be associated the quality of meat [28].

5.12.3. Quality of Meat

The properties of meat like tenderness or color can be linked to biomarkers. The processes of oxidation of proteins as well as degradation of proteins alter the proteins and introduce changes in meat tenderness. The nature of muscle proteins is reflected in the tenderness of meat. Hence, the analysis of biomarkers can reveal the quality of meat or the presence of fraud to detect the original species [24].

A study in 2006 evaluated the changes in muscle proteins in cows especially in the Semitendinosus (ST) and Longissimus Dorsi (LD) muscles immediately post slaughter and post a day of storage. Proteins such as lactolylglutathione lyase, cofilin, substrate protein of mitochondrial ATP-dependent proteinase, HSP-20KDa and SP-22 were altered in both muscles [29].

The conditions before slaughter influenced the type of biomarkers that were formed in animals. For instance, different conditions induce changes in proteomes by conditions before slaughter. Proteomics has revealed that the tenderness of pig meat was revealed by 6 proteins: 3 actin fragments with a heavy chain of myosin along with light chain II of myosin and triose phosphate isomerase of glycolysis were found to influence the tenderness [24].

Pigs with light or dark meats were studied for biomarkers using proteomics; the analysis of SM muscles revealed that 22 proteins were different between the two sets. The level of mitochondrial proteins was higher in the dark meat that showed more oxidative metabolism while the level of cytosolic proteins were more in light meat [30].

The eating quality of meat is reflected by the proteins in either fish or the animal hence the proteomics of meat can reflect the composition and hence quality of meat. Apart from the genetic factors such as calpain genes that influence meat tenderness; *ante*- and *post-mortem* features also impact the quality of meat [1].

The proteomics of fish and poultry are yet to see more advancements and studies [1]. Biomarkers could aid in the:

- Assessment of growth rates of different genetic lines of chicken.
- Distinguish between cooked and war meats of poultry.
- The status of meat; whether the product is local or exotic.

With increase in concerns of animal rights, biomarkers can aid in the analysis of stress or the condition of an animal/bird before its slaughter [31].

A study published in 2012 found that in chianina breed of beef was tender due to biomarkers such as enzymes of glycolysis that were at elevated concentrations [32]. Other markers of meat tenderness include:HSP27 and HSP70 family of proteins that are involved in inhibition of apoptosis as well as protection of myofibrils [33].

The application of protein mapping revealed that the alteration from muscle till it is consumed as meat is associated with meat quality. Post the animal slaughter, the quality of meat is influenced by changes that are seen.

5.12.4. Post-Mortem Storage of Meat: Biomarkers

The analysis of meat post its harvest has been the subject of study for several teams that are summarized below:

There is an increase on enzymes of glycoloysis post slaughter suggestive of aerobic metabolism for a while in muscle.

Inhibitors of apoptosis such as heat shock proteins (HSP) 27 and 70 are reduced that leads to cell death in muscle.

5.12.5. Processed Meat

Studies have shown that the addition of salt to meat (especially ham) is associated with close to 45 myofibrillar proteins of the muscles (*biceps femoris*) of ham. The release of such proteins influences the odor and taste of ham. Additionally, markers such as creatine kinase, myosin light chain and actin are released when Spanish dry-cured hams are ripened [1].

5.13. OVERVIEW OF SERUM PROTEOMICS IN FARM ANIMALS

Serum samples are dynamic with reports of ten orders [34] that are challenged by the use of instruments. Thus, the use of *pre-analytical techniques* plays an important role in such analysis to facilitate the differences in serum. Examples of such techniques in the field of animal proteomics are:

Protein immuno-depletion:

This involves the application of specific antibodies that target and remove proteins that are present at high concentrations in samples such as IgG, albumin, transferrin. Thus, the use of an antibody with high specificity can facilitate the removal of these highly abundant molecules with efficiency.

A challenge in this technique is the unavailability of antibodies of several species of animals as well as binding of antibodies with proteins to be studied.

Combinatorial peptide ligand libraries technique (CPLL):

This technique is based on enrichment of proteins. It decreases the range of concentration usually found in serum and normalizes the levels of low abundant and high abundant proteins. Challenges in this technique are the excluding of proteins that behave against a predicted format as well as not capturing the hydrophilic protein section.

Shotgun proteomics as previously discussed involves the application of MS and is a gel free approach. Samples of sera are digested with enzyme to then separate peptides on high resolution liquid chromatography (LC). The fragments are identified with tandem mass spectrometry (MS/MS) and the use of bioinformatics. Alternatively, the use of gel electrophoresis as one-dimensional (1-DE) and

Two-dimensional (2-DE) formats can facilitate the separation followed by MS and use of bioinformatics.

Quantitative proteomics that uses MS is of two types:

Absolute quantification: the amount of proteins is measured in a sample.

Relative quantification: the differences in expression are observed between different samples such as a control vs. diseased animals.

The techniques that have been applied for relative protein quantification in animal biomarkers include:

Gel-based platforms:

2-DE: comparison of samples.

Two-dimensional differential gel electrophoresis (2D-DIGE): a single gel is used for comparison of different samples.

Gel-free platforms:

ICAT and iTRAQ employ isotopes that label a product of enzymatic digestion that causes a shift in the mass of the labeled sample against the unlabeled sample that can be detected quantitatively on a MS instrument.

Label free methods use LC–MS/MS analysis of different samples and avoid the use of labels.

The field of animal proteomics has been facilitated with surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) in MS. This profiles potential biomarkers of disease between a control and diseased animal and is complementary to the 2 DE technique.

The use of databases is vital in the field of animal proteomics such as Gene Ontology (GO) that aims at the annotation of genes using sequence data or literature studies. Close to 4 lakhs of sequences have been annotated by GO Consortium (GOC); the use of the Uniprot-GOA program has the proteome databases of pig, cow and hens. (Table 5.1).

Table 5.1. Gene Ontology (GO) Annotations

Animal	Numbers of peptides
Pig	19,723
Chicken	14,238
Cow	20,032

Despite the incomplete annotation of several farm animals; database searches can enable homology or the use of reference proteomes. An *enrichment analysis* can facilitate the study of which products are under-expressed or over expressed using gene annotations. These analyzes can be installed or are web-based. (Figure 5.24)

Few such programs include: Ontologizer, gProfiler, Cytoscape [35]

Table 5.2. The Analysis of Serum for Biomarkers has Impact on Human Health and is Summarized by the Table 5.2 [35]

Animal	Proteomics technique	Result (s)
Sheep	2-DE	Immune response in sheep against *Staphylococcus aureus* antigens that cause mastitis.
	2-DE MALDI-TOF-MS	Antigens of *Staphylococcus aureus*
	SELDI-TOF-MS	Putative serum protein biomarkers that are reflective of responses against infectious diseases
	SELDI-TOF-MS 1-DE LC–MS/MS	Biomarkers for detection of paratuberculosis
	2-DE MALDI-TOF-MS	Biomarkers for early diagnosis and welfare of sheep
Cow	MALDI-TOF-MS 2-DE	Changes induced by *Besnoitia besnoiti*to characterize the infection
	iTRAQ MS	Biomarkers for looking at the diseases caused by *Mycobacterium bovis* or *Mycobacterium paratuberculosis*
	2D-DIGE MALDI-TOF-MS MS/MS	New biomarkers of stress in Bruna cows
	2D-DIGE LC–MS/MS	Biomarkers for diagnosis of Johne's disease
	MALDI-TOF-MS	Biomarkers to evaluate cows undergoing wrong treatments
	2-DE MS	Biomarkers of last phase of pregnancy and early postpartum in heifers for better management
Fish	2-DE MALDI-TOF-MS	Biomarkers of response to probiotics in feed in rainbow trout (*Oncorhynchus mykiss*, *Walbaum)*
	CE-MS	Stress levels measured by *N*-glycans level in serum of salmon (*Salmo salar*)
	2-DE MS/MS	Biomarkers of acute response of injury in loach
	2-DE LC–MS/MS	Changes in osmotic pressure measured by lysozyme and angiotensin carboxypeptidase
Pigs	2-D-DIGE MALDI-TOF-MS LC–MS/MS	Biomarkers of infection with classical swine fever virus
	2D-DIGE MS	Biomarkers of stress and conditions of housing

Horses	2D-DIGE MALDI-TOF-MS	Biomarkers for diagnosis and treatment of autoimmune uveitis
	MALDI-TOF-MS LC–MS/MS	Biomarkers of disease
	1-DE MALDI-TOF-MS MS/MS	Biomarkers of consumption of *Senecio jacobea*(a toxic weed found in hay)
Poultry	1-DE MALDI-TOF-MS	Characterization of avian pathogenic *Escherichia coli*
	2-DE LC–MS/MS	Biomarkers of resistance against avian pathogenic *Escherichia coli*
	2-DE MALDI-TOF-MS	Growth and laying biomarkers

2-DE: two-dimensional electrophoresis;

SELDI-TOF-MS, surface-enhanced laser desorption/ionization time-of-flight mass spectrometry;

MALDI-TOF-MS, matrix assisted laser desorption ionization-time of flight-mass spectrometric;

LC–MS/MS, liquid chromatography–tandem mass spectrometry;

1-DE, one-dimensional electrophoresis;

2D-DIGE, two-dimensional differential gel electrophoresis;

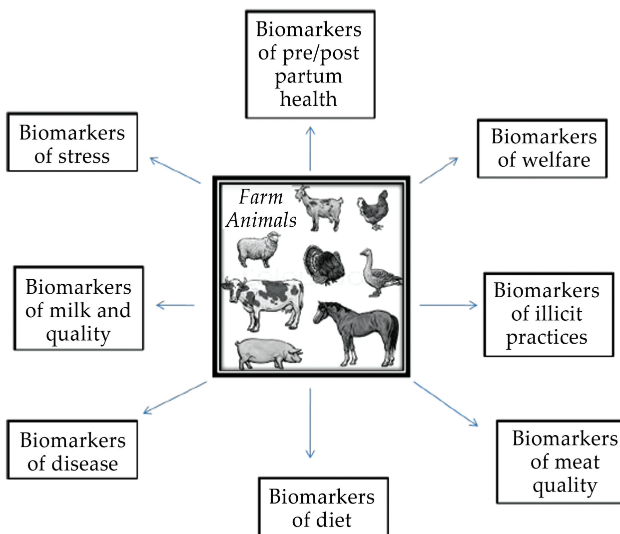


Figure 5.4. Various applications of biomarkers using proteomics in animals.

5.14. BIOMARKERS USING PROTEOMICS OF ANIMAL PRODUCTS

1. Serum

An analysis of serum of cows infected with mastitis was seen as an upregulation of vitronectin (inflammatory proteins) suggestive of its use as a marker for diagnosing the disease in cows [36].

Stress responses in cows were characterized using DIGE labeling, 2-DE and MALDI-MS. It was found that markers such as glutathione peroxidise, paraoxonase, α-HSG (acute phase protein) as well as cholesterol and cortisone in faeces can serve as biomarkers of health of cows under stress [37].

A study published in 2012 evaluated the changes in serum post intense exercise in horses. Proteomics revealed the biomarkers: proteins associated with modulation of immune responses, clotting and inflammation [38]

2. Tears

The constitution of tears of camels undergoes changes according to seasons especially in terms of VMO1 homolog and lactoferins. The former protein aids in the maintenance of health of the eye of camels given its habitat [24].

The analysis of dog tears by 2-DE and MS revealed that lactoferrin or lysozyme was present to a lesser extent as humans while an analog of lipocalin was detected as an allergen in dogs [24].

3. Milk

The analysis of major milk protein Caseins in horses revealed that the proteins were similar to that of human caesins suggestive of the use of equine milk in humans with allergy [39].

In an attempt to search for biomarkers of mastitis in cows; a team in 2004 analyzed whey using proteomics to reveal that the levels of α-lactalbumin and β-lactoglobulin that are the major elements of whey were decreased. However, the level of albumin and serotransferrin that are serum proteins were found in whey of cows with mastitis [40].

A study in 2013 evaluated the use of biomarkers using proteomic technologies of MALDI-TOF-TOF to check purity of milk. The technique

could identify powdered milk in liquid milk with a sensitivity of less than 1% in both processed and raw milk [41].

The detection of α_{S1}-CN variants could detect the various types of milk as well as adulteration in another 2013 publication. The use of electrospray ionisation quadrupole TOF analysis prior to casein dephosphorylation that could detect the breeds of cows (Mediterranean water buffalo and foreign breed milk) (42)

4. Milk Products

The use of biomarkers in the cheese industry can aid in the identification of quality and maturation of such dairy products. A study published in 2008 reported the use of β-CN fragment, lactoferrin and vitamin D-binding proteins as well as isoform B of β lactoglobulin as markers of yield of cheese [43].

A study published in 2012 reported that the role of decreased concentrations of phosphorylated forms of α_{S1}-CN form (α_{S1}-CN 8P)could indicate milk that did not coagulate or was poorly coagulated [44]. (Figure 5.5).

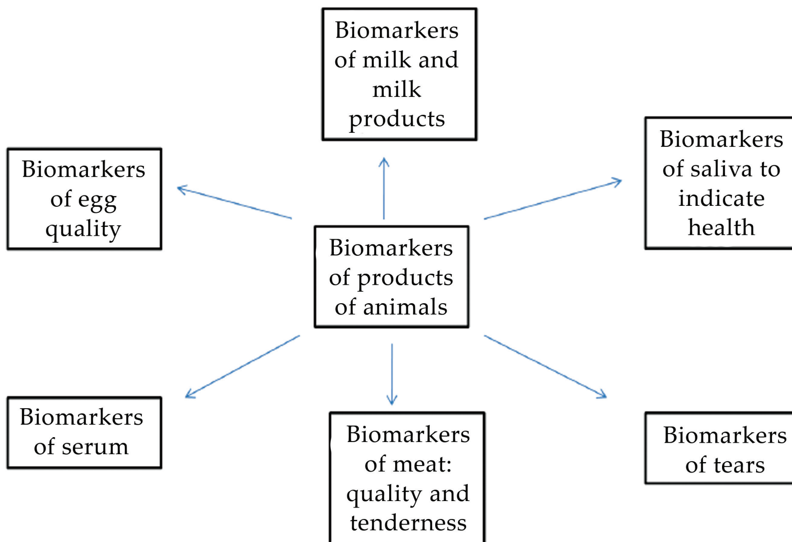


Figure 5.5. An overview of biomarkers of animal products.

5. Saliva

Due to the complex constitution of saliva as well as its production from the salivary glands as well as blood makes its analysis interesting especially for stress biomarkers in animals. The nutritional status of animals can be reflected as the salivary composition thus suitable biomarkers can indicate deficiencies such as:

Deficiency of ascorbate caused decrease in amylase in saliva of guinea pigs.

Deficiency of iron caused decrease in peroxidase in saliva of rats.

The consumption of diets can influence the proteome composition. The salivary protein amylase is not present in parotid saliva of ruminants. However the amylase is abundant in omnivores [24].

6. Egg

There is not much data available on the biomarkers of egg quality of chickens [1]. A combination of 2DE and LC/MS/MS on egg white proteins of different varieties [6] of eggs revealed a novel finding: a quiescence precursor protein [45].

In another study, a 2DE-proteomic approach examined the egg proteins during storage at a range of temperatures 4°C, 20°C and 37°C for half a month. The increase in temperature caused degraded albumin and a complex of lysozyme–ovalbumin was detected. The concentration of clusterin showed a dip with increasing temperature suggestive of its use as a biomarker of egg storage [46].

The analysis of eggshell cuticle by LC-MS/MS can be vital as the cuticle is a protective factor from the environment to serve as biomarkers of egg quality [1]. Two of the 47 proteins that were studied are anti-microbial: ovocalyxyn 32 and Kunitz-like protease inhibitor [47].

5.15. CASE STUDY OF BIOMARKERS: TARGETING THE MOSQUITO

Rationale:

Diseases caused by mosquitoes are a huge menace with malaria fatalities at 655,000 in 2010 as well as Chikungunya virus infection. A team in 2010 used proteomics to assess mosquito population age structure as the epidemiology

of such infections depends on time: for instance, 7–12 days for dengue and 9–14 days for malaria.

Traditional techniques to estimate age of mosquitoes entail the use of dissection of females to look at the reproductive status and estimate the age. However, the use of biomarkers can circumvent such techniques: aging biomarkers as protein profile expression.

Approach:

The application of 2-D DIGE analysis to *Ae aegypti* as it ages followed by mixed-effects model variance to select candidates of biomarkers. Following collection of protein lysates of Ae. aegypti females the head and thorax proteins samples were subjected to 2-D DIGE. The application of model fitting aided in addressing variation due to aging. 86 spots were located as well as 10 candidates of biomarkers that were subjected to in-gel trypsin digestion and Matrix Assisted Laser Desorption Ionization tandem Time of Flight (MALDI-TOF/TOF) mass spectrometry followed by database study of Ae. aegypti. (Figure 5.6).

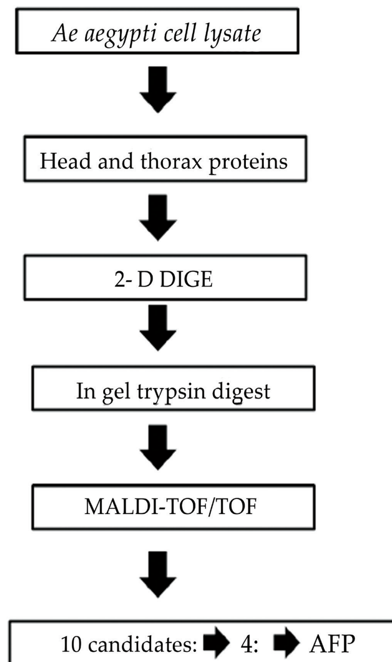


Figure 5.6. Scheme of the study.

Results:

The ten candidate aging biomarkers are as follows:

- Pyruvate kinase (PK)
- Q17LN8
- Paramyosin (Pm)
- Anterior Fat body Protein (AFP)
- 70 kDa protein matching heat shock (HSP)
- eukaryotic Initiation Factor 5A (eIF5A)
- Electron Transport Oxidoreductase (ETO): acidic and basic
- SOD2a (an acidic isoform Manganese superoxide dismutase (SOD2)

Of these 10 biomarker candidates, the levels of PK, ETOa, SOD2a and Pm increased along with age while that of ETOb, eIF5A, ADFa, Q17LN8, AFP and HSP showed a dip with age. Of these, vivd changes was recorded with AFP; the levels decreased by a factor of 10 between day 1 to day 5 followed by a further dip by tenfold up to day 9 beyond which it could not be detected.

The authors then narrowed down on 4 candidates ADF, eIF5A, Q17LN8 and AFP based on variance with age of which AFP could serve as a marker. The level of AFP decreases after 13 days; the age at which the mosquito can transmit dengue. Hence, the AFP marker could be used to identify the infectious mosquitoes in a population [48].

5.16. SCENARIO/FUTURE FOR ANIMAL PROTEOMICS

The various aspects of research presented in this chapter highlight the various advantages and studies involving biomarkers in animals.

5.16.1. Meta Proteomics

This is the study of proteomics of an environment or a community i.e., use of multiple biomarkers. This analysis can facilitate the study of markers from an ecosystem or habitat that can present the scenario of a larger community. Though a challenge of meta proteomics is requirements of advancements in the techniques of this approach. The technology is described as powerful to comprehend the microbiomes of farm animals to elucidate mechanisms

between the host and microbes. Thus, a combination of proteomics as well as meta proteomics can facilitate the opening of new vistas for diagnosis or prognosis.

The *advantages* of this approach are:

- The gastrointestinal details of ruminants can be assessed to scour for disorders.
- A picture of biomarkers of an environment can be revealed.
- The studies of microbial biomarkers in animals: change in several markers of body fluids of animals can aid in diagnosis of diseases or the effect of stress on an animal.

There have been several advancements in the field of animal proteomics with special skills required for skills such as bioinformatics in order to analyze the data from such studies. Certain reports in literature indicate that though the advancements of proteomics are less compared to genomics and transcriptomics in animals; the development of proteomics is expected.

The development of COST Action on FAP with an International forum can aid in the development of technology for analysis in farm animals. To circumvent the challenges in proteomics in farm animals, this COST action can aid in the access of the technology to groups working with animals. Despite the smaller numbers of labs involved in proteomics with sophisticated technology, collaboration can be facilitated.

To summarize in a nutshell, proteomic biomarkers have aided in the study of:

- Proteome maps of tissue/fluid in pigs and cows.
- The changes in meat as well as changes post-harvest of fish.
- Change in milk proteins in order to assess the quality of milk.
- The study and indicators of disease
- The alterations in smoke dried meat have been characterized.
- The muscle growth is monitored.
- The feed can be authenticated.
- Food safety such as authenticity of meat such as whether a sample is really cow/goat.
- Animal reproduction technologies.

In order to reap maximum benefit from the field of animal science, it is essential that proteomics can aid several fields including biomarker discovery; a shift from limited labs to a broader application [1].

5.17. SUMMARY

The use of data-independent techniques discussed in this chapter can facilitate the studies of animals using the sensitive platform of MS.

The realm of proteomics can be extended to the spheres of research as in:

- The processes occurring in the growth, development, production of animals.
- The interactions of an animals and parasite
- Shifts in enzymes or proteins that can serve as indicators or biomarkers
- Mechanisms of pathology and physiological processes

Despite the importance of genomics; the contribution to an aspect of animal studies is proteins as they influence milk and meat quality. The potential of proteomics and biomarkers in analysis of meat quality is immense. The scanning of biomarkers of disease and subclinical conditions are encompassed in its medical applications (Moore et al, 2007).

The welfare of farm animals can be analyzed by proteomic techniques to search for relevant biomarkers. This aspect is important as the safety of animals and minimum disease is essential for mitigation of financial loss. The use of pre-analytical sample treatments as well as sensitive platforms coupled to data processing place proteomics at a higher edge to scan for relevant biomarkers.

The platforms of proteomics can facilitate analysis of multiple samples as compared to traditional analytical techniques such as PCR or ELSA. Despite the initial heavy investment for these technologies, the use of these high throughput technologies can answer the search for biomarkers in animals (Jayshri et al, 2014).

REFERENCES

1. Almeida, A. M., Bassols, A., Bendixen, E., et al. (2015). Animal board invited review: advances in proteomics for animal and food sciences. *Animal*. *9*(1), 1–17.
2. Kjaersgard IVH, Norrelykke, M. R., & Jessen, F. (2006). Changes in cod muscle proteins during frozen storage revealed by proteome analysis and multivariate data analysis. *Proteomics. 6,* 1606–1618.
3. Paredi, G., Raboni, S., Bendixen, E., de Almeida, A. M., & Mozzarelli, A. (2012)..Muscle to meat' molecular events and technological transformations: the proteomics insight. *Journal of Proteomics*. *75,* 4275–4289.
4. Miller, I. (2012). Application of 2D DIGE in animal proteomics. *Methods in Molecular Biology*. *854,* 373–396.
5. Nikolov, M., Schmidt, C., and Urlaub, H. (2012).. Quantitative mass spectrometry-based proteomics: an overview. *Methods in Molecular Biology*. *893,* 85–100.
6. Blokhuis, H. J., Hopster, H., Geverink, N. A., Korte, S. M., & van Reenen, C. G. (1998). Studies of stress in farm animals. *Comparative Haematology International*. *8,* 94–101.
7. Nir O. (2012). What are production diseases, and how do we manage them? *Acta Veterinaria Scandinavica*. *98,* 21–32.
8. Hogarth, C. J., Fitzpatrick, J. L., Nolan, A., Young, F. J., Pitt, A., & Eckersall, P. D. (2004). Differential protein composition of bovine whey: a comparison of whey from healthy animals and from those with clinical mastitis. *Proteomics*. *4,* 2094–2100.
9. Smolenski, G., Haines, S., Kwan FYS, Bond, J., Farr, V., Davis, S. R., Stelwagen K & Wheeler, T. T. (2007). Characterization of host defence proteins in milk using a proteomic approach. *Journal of Proteome Research*. *6,* 207–215.
10. Boehmer, J., Bannerman, D., Shefcheck K & Ward, J. (2008).. Proteomic analysis of differentially expressed proteins in bovine milk during experimentally induced *Escherichia coli* mastitis. *Journal of Dairy Science*. *91,* 4206–4218.
11. Moore, R. E., Kirwan, J., Doherty, M. K., & Whitfield, P. D. (2007). Biomarker Discovery in Animal Health and Disease: The Application of Post-Genomic Technologies. *Biomarker Insights*. *2,*185–196.
12. Aronson, J. K. (2005). Biomarkers and surrogate endpoints. *Brit J Clin*

Pharmacol. *59,*491–4.

13. Martinez-Bartolome, S., Deutsch, E. W., Binz, P. A.,et al. (2013). Guidelines for reporting quantitative mass spectrometry-based experiments in proteomics. *Journal of Proteomics*. *95,* 84–88.
14. Malmstroem, J., Lee, H., & Aebersold, R. (2007). Advances in proteomic workflows for systems biology. *Current Opinion in Biotechnology*. *18,* 378–384.
15. Boehmer, J. L. (2011). Proteomic analyzes of host and pathogen responses during bovine mastitis. *Journal of Mammary Gland Biology and Neoplasia*. *16,* 323–338.
16. Bislev, S. L., Kusebauch, U., Codrea, M. C., Beynon, R. J., Harman, V. M., Rontved, C. M., Aebersold, R., Moritz, R. L., & Bendixen, E. (2012). Quantotypic properties of QconCAT peptides targeting bovine host response to Streptococcus uberis. *Journal of Proteome Research*. *11,* 1832–1843.
17. Cairoli, F., Battocchio, M., Veronesi, M. C., et al. (2006). Serum protein pattern during cow pregnancy: acute-phase proteins increase in the peripartum period. *Electrophoresis*. *27,*1617–25.
18. Hennig-Pauka, I., Jacobsen, I., Blecha, F., et al. (2006). Differential proteomic analysis reveals increased cathelicidin expression in porcine bronchoalveolar lavage fluid after an Actinobacillus pleuropneumoniae infection. *Vet Res*. *37,*75–87.
19. Wolf, C., Kusch, H., Monecke, S., Albrecht, D., Holtfreter, S., von Eiff, C., Petzl, W., Rainard, P., Broeker, B. M., & Engelmann, S. (2011). Genomic and proteomic characterization of *Staphylococcus aureus* mastitis isolates of bovine origin. *Proteomics. 11,* 2491–2502.
20. Hughes, V., Smith, S., Garcia-Sanchez, A., Sales, J., & Stevenson, K. (2007). Proteomic comparison of *Mycobacterium avium* subspecies paratuberculosis grown in vitro and isolated from clinical cases of ovine paratuberculosis. *Microbiology-SGM*. *153,* 196–205.
21. Rodrigues, P. M., Silva, T. S., Dias, J., & Jessen, F. (2012). PROTEOMICS in aquaculture: applications and trends. *Journal of Proteomics*. *75,* 4325–4345.
22. Zhou, S., Wan, Q., Huang, Y., Huang, X., Cao, J., Ye, L., Lim, T. K., Lin, Q., & Qin, Q. (2011). Proteomic analysis of Singapore grouper iridovirus envelope proteins and characterization of a novel envelope protein VP088. *Proteomics*. *11,* 2236–2248.

23. Frenal, K., Kemp, L. E., & Soldati-Favre, D. (2014)..Emerging roles for protein S-palmitoylation in toxoplasma biology. *International Journal for Parasitology*. *44,* 121–131.
24. Jayasri, K., Padmaja, K., Eswar Prasad, P. (2014). Proteomics in Animal Health and Production. *IOSR Journal of Agriculture and Veterinary Science*. 7(4), 50–56.
25. Kuhla, B., Albrecht, D., Kuhla, S., & Metges, C. C. (2009). Proteome analysis of fatty liver in feed deprived dairy cows reveals interaction of fuel sensing, calcium fatty acid and glycogen metabolism. *Physiol Genomics*. *37,* 88–98.
26. Rawson, P., Stockum C et al. (2012)., Metabolic proteomics of the liver and mammary gland during lactation. *Proteomics*. 4429–4435.
27. Lee, K. Y., Jung, K. C., Jang, B. G., Choi, K. D., Jeon, J. T., & Lee, J. H. (2008). Identification of differentially expressed proteins at four growing staes in chicken liver. *Asian Austras J Anim Sci*. 21(10), 1383–8.
28. Zheng, W. Chang, Z. W., Hou S et al. (2014). Unraveling molecular mechanistic differences in liver metabolism between lean and fat lines of Pekin duck (*Anas platyrhynchos domestica*): A proteomic study. *Journal of proteomics*. *98,* 271–288.
29. Jia, X., Hollung, K., Therkildsen, M., Hildrum, K. I., & Bendixen, E. (2006). Proteome analysis of early post-mortem changes in two bovine muscle types: *M.longissimus dorsi* and *M.semitendinosis*. *Proteomics*. *6,* 936–944.
30. Sayd, T., Morzel, M., Chambon, C., Frank, M., Figwer P., Larzul C et al. (2006). Proteome analysis of the sarcoplasmic fraction of pig semimembranosus muscle: implications on meat color development. *J Agric Food chemistry*. *54,* 2732–2737.
31. Hazard, D., Fernandez, X., Pinguet, J., Chambon, C., Letisse, F., Portais, J., Wadih-Moussa, Z., Remignon, H., & Molette, C. (2011). Functional genomics of the muscle response to restraint and transport in chickens. *Journal of Animal Science*. *89,* 2717–2730.
32. Alessandro, D., Marrocco, C., Rinalducci, S., Mirasole, C., Failla, S., & Zolla, L. (2012)..Chianina beef tenderness investigated through integrated Omics. *Proteomics*. 4381–4398.
33. Guillemin, N., Jurie, C., Cassar-Malek, I., Hocquette, J., Renand, G., & Picard, B. (2011). Variations in the abundance of 24 protein biomarkers

of beef tenderness according to muscle and animal type. *Animal. 5,* 885–894.

34. Anderson, N. L., & Anderson, N. G. (2002). The human plasma proteome: History, character, and diagnostic prospects. *Mol Cell Proteomics. 1,* 845–867.
35. Girolamo, F. D., D'Amato, A., Lante I, Signore F, Muraca, M., & Putignani, L. (2014). Farm Animal Serum Proteomics and Impact on Human Health. *Int J Mol Sci. 15,* 15396–15411.
36. Turk, R., Piras C et al. (2012). Proteomics of inflammatory and oxidative stress response in cows with subclinical and clinical mastitis. *Proteomics. 2012,* 4412–4428.
37. Ramell, A. M., Arroyo, L., et al. (2012). Proteomic analysis reveals oxidative stress response as the main adaptative physiological mechanism in cows under different production systems. *Proteomics.* 4399–4411.
38. Scoppetta, F., Tartaglia M et al. (2012). Plasma protein changes in horse after prolonged physical exercise: A proteomic study. *Proteomics.* 4494–4504.
39. Miranda, G., Mahé, M. F., Leroux, C., & Martin, P. (2004). Proteomic tools to characterize the protein fraction of Equidae milk. *Proteomics. 4,* 2496–509.
40. Hogarth, C. J., Fitzpatrick, J. L., Nolan, A. M., Young, F. J., Pitt, A., & Eckersall, P. D. (2004). Differential protein composition of bovine whey: a comparison of whey from healthy animals and from those with clinical mastitis. *Proteomics. 4,* 2094–100.
41. Calvano, C. D., Monopoli, A., Loizzo, P., Faccia, M., & Zambonin, C. (2013). Proteomic approach based on MALDI-TOF MS to detect powdered milk in fresh cow's milk. *Journal of Agricultural and Food Chemistry. 61,* 1609–1617.
42. Caira, S., Pinto, G., Balteanu, V. A., Chianese, L., & Addeo, F. (2013). A signature protein-based method to distinguish Mediterranean water buffalo and foreign breed milk. *Food Chemistry. 141,*597–603.
43. Wedholm, A., Moller, H., Stensballe, A., Lindmark-Mansson, H., Karlsson, A., Andersson, R., Andren, A., & Larsen, L. (2008). Effect of minor milk proteins in chymosin separated whey and casein fractions on cheese yield as determined by proteomics and multivariate data analysis. *Journal of Dairy Science. 91,*3787–3797.

44. Jensen, H., Poulsen, N., Andersen, K., Hammershoj, M., Poulsen, H., & Larsen, L. (2012). Distinct composition of bovine milk from Jersey and Holstein-Friesian cows with good, poor, or noncoagulation properties as reflected in protein genetic variants and isoforms. *Journal of Dairy Science*. *95,* 6905–6917.
45. Wang, J., Liang, Y., Omana, D. A., Kav, N. N. V., & Wu, J. (2012).. Proteomics analysis of egg white proteins from different egg varieties. *Journal of Agricultural and Food Chemistry*. *60,* 272–282.
46. Qiu, N., Ma, M., Zhao, L., Liu, W., Li,Y., & Mine, Y. (2012). Comparative proteomic analysis of egg white proteins under various storage temperatures. *Journal of Agricultural and Food Chemistry*. *60,* 7746–7753.
47. Rose-Martel, M., Du, J., & Hincke, M. T. (2012).. Proteomic analysis provides new insight into the chicken eggshell cuticle. *Journal of Proteomics*. *75,* 2697–2706.
48. Hugo, L. E., Monkman, J., Dave, K. A., Wockner, L. F., Birrell, G. W., Norris, E. L., et al. (2013). Proteomic Biomarkers for Aging the Mosquito *Aedes aegypti* to Determine Risk of Pathogen Transmission. *PLoS ONE*. 8(3): e58656.

CHAPTER 6

CLINICAL BIOMARKERS AND PROTEOMICS

CONTENTS

6.1. Introduction 186
6.2. Overview Of Biomarkers 186
6.3. Historical Background 187
6.4. Role Of Biomarkers In Drug Development 188
6.5. Brief Regulatory Aspects 189
6.6. Utility Of Biomarkers In Clinical Studies 190
6.7. Biomarkers In Clinical Evaluation 190
6.8. Urine Biomarkers 198
6.9. Serum Biomarkers 203
6.10. Classes of Biomarkers 213
6.11. Efficacy Biomarkers 217
6.12. Surrogate Biomarkers 217
6.13. Predictive Biomarkers 218
6.14. Pharmacodynamic (PD) Biomarkers 218
6.15. Prognostic Biomarkers 219
6.16. Discrepant Results And Its Major Impact on Clinical Trials 221
6.17. Lack of Traceability as a Major Source of Systematic Errors 225
6.18. Impact of Discrepant Results on Personalized Medicine 227
6.19. Mass Spectrometry And Marker Detection 230
6.20. Conclusion 234
References 235

6.1. INTRODUCTION

Since many years, one of the common equations of drug discovery has always been "one drug fits all" paradigm. However, in the recent years, a major shift has been experienced by today's pharma industry with the motif of administering right drug to the right patient, but in a right dose. The approach has as well been identified as the personalized medicine.

In this regard, identification of different biomarkers and their relativity with the expression of particular disease condition has been proposed by scientists across the globe; in order to analyze their critical role in personalized medicine. Accordingly, current chapter is intended to identify different biomarkers, critical for expression of various diseases, current state of biomarker development and applications; with closer insights into the latest technologies and their path towards clinical implementation.

6.2. OVERVIEW OF BIOMARKERS

Since decades, applications of biomarkers have been employed in the clinical medicine. In fact, a need of an hour is to streamline more accurate identification and screening technologies, within the specified period of time to allow faster diagnosis and effective implementation of relative therapeutic modules to the patients. The goal can indeed be achieved through the availability of analytical tools for faster assessment of biological parameters, which can as well be referred to be as the Biomarkers.

Biomarkers can be defined as the markers of certain diseases, which can be quantified and evaluated as indicators of pathophysiological condition of certain biological diseases in terms of their clinical as well as pharmacological responses towards therapeutic interventions of diseases [1].

One of the most commonly known biomarkers can be noted as the blood pressure, which allowed the discovery, development and application of antihypertensive compound. Furthermore, this particular biomarker has reached 9.2 billion $ industry in the US, in 2004.

Based on their applications in different diseases, these biomarkers can further be classified as:

- Antecedent biomarkers, which can be applied to propose the risk of illness.
- Screening biomarkers, to assess subclinical diseases, associated with the primary one.

- Diagnostic biomarkers, to recognize overt diseases.
- Staging biomarkers, which can relate disease severity.
- Prognostic biomarkers, which can predict future diseases, recurrence, suitable response to particular therapy and monitoring efficacy of the therapy.

6.3. HISTORICAL BACKGROUND

In an ancient time, therapies used to be offered only on the basis of physical examination and analysis of pumping efficiency of the body. Body fluids testing have been initiated around 6000 years ago through human urinary analysis [2].

However, it should as well be noted that prior to the era of Hippocrates, i.e., 460-370 BC; Babylonian, Egyptian and Far Eastern cultures were as well documented to be familiar with routine urine analysis. The analysis was proposed with an agenda of clinically confirming the conditions. In fact, some population used to follow a practice of allowing patient with serious ailment to breathe into sheep's nose; the animal used to be alternatively slaughtered to inspect its liver for further evidences. This particular diagnosis was based on the belief that the liver should be treated as the center of our organ system, controlling entire physiological processes. The modern understanding of metabolic importance of hepatic cells goes linear with the concept.

During an era of Ikhnaton and Cleopatra, recorded evidences are available in Egyptian literature stating the importance of diagnostic hormonal tests for indications, like pregnancy. The concept was later evolved by twentieth century scientists to be human chorionic gonadotropin hormone (hCG); in fact, the first bioassay was developed with the same hormone, by injecting woman's urine into an immature rat and looking for an estrous reaction, like follicular growth and ovarian observation for excessive blood supply. Another ancient diagnostic test that had been documented in Hindu cultures can be noted as analysis of the sweetness of urine and its ability to attract ants, to confirm diabetes mellitus [3].

Thus, application of biomarkers in the disease management has been exploited to a great extent in the last four years.

6.4. ROLE OF BIOMARKERS IN DRUG DEVELOPMENT

Since, these ancient eras, a transformative evolution has been galvanized to support; theranostic drug development, which employs targeted therapy to patients, after diagnostic testing. One of the primary goals in utilizing these markers is to compromise probable drug attrition, during various phases of its development, such as preclinical, clinical, etc.; in order to reduce overall cost of its development. Biomarkers are being used in various phases of drug development, right from initial phase to final discovery to the extent of market analysis, in postproduction phase. In early stages of these drug developments, biomarkers play a crucial role in evaluating the activity of specified drugs in animal models; verify the proposed mechanism of action, and proof of concept investigational analysis. The markers can as well be utilized to connect pre-clinical analysis with clinical pharmacological evidences. Moreover, in the later stages of drug development, marker analysis can be effective in confirming decisions, related to dose response and evaluation of optimal dosage regimen for desirable pharmacological effect, along with the patient's safety. In depth studies related to biomarkers have confirmed their effective applications in identifying particular patient population, more likely to respond to the drug, both in a positive as well as negative way.

Hence, these strategies towards utilization of patient's biomarkers to relate his/her genetic makeup and current medical profile, in order to offer him/her a 'personalized medicine' have been found to be tremendously helpful in offering more cost effective therapeutic solutions that are quick and target oriented; especially for diseases that currently have no effective treatments, like neurodegenerative disorders, cancer, etc.

In this regard, current technical market is full of variety of assays to analyze range of biomarkers for different purposes; these assays can be variable right from fit-for-purpose basis to final surrogate end point, for patient selection. However, considering the increasing demand of biomarkers and their effective therapeutic exploitation, it is very important to accelerate their production, with accurate assay validation; since, discrepancies at any point of time may trigger false decisions, posing threat to many lives.

Thus, it is very imperative to adopt advanced methodologies to mitigate discrepancies, in order to integrate multidimensional technology into patient centric models.

6.5. BRIEF REGULATORY ASPECTS

Due to the strong media publication and rising concern regarding false positive as well as negative results in early detection of certain cancers; government have authenticated certifications from two of the most prestigious authorities, like clinical laboratory improvement act and college of American pathologists. Further to this decision, congress has sanctioned Clinical Laboratory Improvement Amendments of 1988 to ensure accuracy as well as reliability in laboratory testing. The said legislation has been circulated to all the laboratories, under regulation, which may have included hospitals, independent and physician office laboratories, etc. thus allowing better performance of all types of biochemical analysis, like microbiology, serology, immunology, cytogenetical and serology, wherein body material isolated from human body is being tested for diseases, for the purpose of diagnosis, disease prevention and treatment of patients.

Accordingly, since 1988, both Food and Drug Administration (FDA) and the Centers for Disease Control and Prevention (CDC) have been working in coordination to authenticate diagnostic results, improve testing quality through continuous research and technological upgradation. Under this regulation, the authorities as well conduct surveys to authenticate laboratory test results through external evaluation scheme, under approved proficiency testing programmes. (PT).

It is further interesting to note that under tens of thousands of biomarker assays studied in many clinical analyzes; only 88 markers have been regulated so far through these regulatory authorities, consisting of 9 hematology markers, 17 general chemistry, 7 endocrinology, 15 toxicology and 9 microbiology markers.

Ideal biomarkers are:

- It should express great sensitivity, specificity and higher predictive value
- It should be safe and easy to measure.
- It is cost efficient to implement in routine drug discovery as well as development.
- It should be modifiable with effective outcome of the treatment.
- It should provide consistent outcome, across the gender and ethnic groups

6.6. UTILITY OF BIOMARKERS IN CLINICAL STUDIES

Minimization of side effects, through application of strong chemotherapy has been a major concern, within a segment of oncology. [4].

It has been estimated that every year from an average of 10, 000 new chemical entities, proposed for pharmacological testing and safety evaluation, only 10% that means 1000 markers would pass the criteria decided by regulatory authorities, and out of these 1000 markers, only 1 % would show combined results for both safety as well as efficacy. Furthermore out of 10 NDA submissions to the FDA, only 1 drug on an average passes the review process. [5]; this in turn cause post market stagnation, due to much lengthy process of cost analysis and process optimization. [6].

One of the cost effective solutions in this regard, can be noted as the bedside safety evaluation of a new drug, directly on the patients; which can generate translational biomarkers early in the research process to allow predictive evaluation of mechanism of action of a particular drug, in clinical studies. Some innovative trial designs have been proposed by FDA in 2004, with an idea to propose some encouraging guideline. In this regard, safety evaluation, predictive effectiveness and relative information should be provided between mechanism of action and clinical effectiveness of these biomarkers; which can be implemented through new imaging technologies. However, much is to be achieved through development and standardization of biological, statistical and upgraded methods; before their wide applications.

6.7. BIOMARKERS IN CLINICAL EVALUATION

Accordingly, clinical laboratory measurements can be treated as the most crucial component of drug studies to demonstrate safety as well as efficacy of the same. Accordingly, different biomarkers have been classified into five different classes, as per their diagnostic stages; which have been noted earlier. It should be noted that the clinical significance of these markers, depends upon their sensitivity, specificity, accuracy, reproducibility of results and reliability. Out of these five different classes, diagnostic and prognostic markers are identified to be as the most significant biomarkers; and hence it is very important to correlate the test outcome and the treatment application, as the lead domain of the prospective clinical studies. This warrants the importance of the diagnosis markers for the better clinical performance.

In the subsequent section, we have discussed different important biomarkers, as sources of indicators for local, systemic and infectious disorders.

6.7.1. Salivary Biomarkers

The buccal cavity is a complex environment, with multiple tissues and structures working in synchronized manner. However, these tissues have been colonized with many different bacteria and have been immersed in the salivary fluids, in order to perform dedicated unique task.

Earlier, saliva was only identified as the essential component of the digestive processes and one of the first points of contacts; which breaks down the food into lipids as well as starches. However, with successive technological development, understanding about the role of saliva has changed to a great extent. Studies have confirmed that it can be one of the important biomarkers with variety of molecular and microbial analytes. [7].

Many of these studies have been published in reputed journals to ascertain that these constituents can be referred to be as effective indicators for both local as well as systemic disorders.

These revelations are the basics of salivary diagnostics that have sparked interests of global scientists to identify salivary based markers for disorders, ranging from cancer to many infectious diseases.

6.7.2. Salivary Biomarkers for Cancer

"Saliva omics" is a wider compilation of technologies exploited to identify different types of markers, present in the saliva. To understand their mechanism of action, it is necessary to analyze them at level of genomics, epigenomics, transcriptomics, metabolomics, proteomics and microbiota levels.

6.7.2.1. Genome and Epigenome

Both human as well as microbial DNA is present in salivary genome. Studies have further confirmed that total DNA in saliva is approximately 24 ug, which can be related as 10 times lower than other counterparts of the body, such as blood. Formation of tumors is a multistep process, can be referred to be as the indication of cancer pathology, both at the molecular as well as genetic levels. Apart from tumorigenesis, detection of methylated DNA

at random locations can be noted as epigenetic indications to be associated with cancers. Diverse biomarker techniques are available to assay salivary fluids both at genetic as well as epigenetic levels.

6.7.2.2. Transcriptomics

The analysis related to salivary transcriptome is mainly centered towards mRNA and miRNA that are being detected in the cells of the oral cavity, placed at a distance to original cells. The transcriptome was originally fingerprinted by UCLA lab. Accordingly, they are emerging as the new regulators of diverse biological functions that have used crucial role in early detection of cancer genes and tumor progression. Globally, many different mRNA and micro-RNA are noted as markers of lung cancer [8], pancreatic cancers [9] and breast cancers. [10]

6.7.2.3. Proteomics

The proteomic analysis of salivary fluid can give idea of the entire protein content of the oral cavity, which can be related to around 2000 different proteins and peptides; which are responsible for many different biological functions of the oral cavity. Thus, proteomic analysis through saliva has distinct advantages to know about different salivary peptides, through non invasive as well as label free techniques.

Currently mass spectrometry is one of the most trusted techniques for salivary protein identification. Similarly, in cases such as lung and breast cancers, two-dimensional gel electrophoresis combined with mass spectrometry is found to be highly sensitive and specific to particular tests.

6.7.2.4. Metabolomics

It is a worldwide comprehensive biodata of the metabolic status of various pathophysiological processes, related to many metabolic disorders. Techniques are intended to measure the levels of endogenous metabolites to analyze different biomarkers. These endogenous metabolites can be referred to be as lipids, amino acids, peptides, vitamins, organic acids, thiols, etc.

By the year 2010, cancer specific markers have been identified in salivary metabolites. The said identification was confirmed with the study involving analysis of metabolites from oral, pancreatic, breast cancers and periodontal diseases. Apart from which, some other metabolites were confirmed to

differentiate oral squamous cell carcinoma and neurodegenerative dementia patients.

6.7.2.5. Microbiota

Advances in the sequencing allowed detection and confirmation of approximately 19,000 phylotypes in the oral cavity. Further evidences in this regard, have confirmed myriad of bacteria and other microorganisms leading to the oral diseases, like caries, periodontitis, as well as other systemic diseases including cancer.

Recently other scientists have as well demonstrated that identification of two bacterial markers through microarray and PCR techniques can distinguish pancreatic cancer patients from healthy subjects. [9]

Accordingly, the current status of different salivary diagnostic markers for various available cancers is noted herewith.

6.7.2.6. Brain Cancer

Globo can has published the report, according to which 139,608 and 1,16,605 new brain cancer cases were reported in both the sexes, by the end of 2012. [11] Out of these many cancer markers, glioblastoma has been detected to be the most frequent one, with the poor survival rates. [12]

It has been further estimated that oxidative stress should be the major factor in the development of brain cancer, due to its implications towards cellular growth and higher rate of proliferation. In fact, damage related to the oxidative stress is considered to be a pro-tumorigenic agent, due to the abnormal sensitivity of brain towards it, may be due to its higher rate of oxygen consumption, higher level of lipid content and comparatively low antioxidant defense system. According to the reported studies, antioxidant power due to ferric reduction and levels of protein thiol can be noted as the possible markers of oxidative stress and predictors of disease prognosis in saliva. Thus, it should be noted that when the mean salivary FRAP values, are significantly lower in patients, it can be an indication of benign as well as malignant brain tumors; thus, two of these markers, which can be reported to be as FRAP and Protein thiols are reported to be as the better indicators of stress related brain cancers.

Thus, these markers can be referred to be as the possible window of opportunity; allowing early detection for better therapeutic modules, in case of brain markers.

6.7.2.7. Pancreatic Cancer

Relative survival rate of five-year pancreatic cancer has been estimated to be of approximately 7%, as per the report submitted by Globo can; which, suggests the urgent need of novel diagnostic tests with latest biomarkers, for early detection of the said cancer. Several studies have evidently confirmed that mRNA and miRNAs can be referred to be as the most reliable and potential diagnostic as well as prognostic markers, in patient suffering with the said cancer. [13]

Interestingly, these markers have as well been detected in saliva, along with several other body fluids, like blood, stool, pancreatic juice and tumor itself. In one such study, logistic regression model was created with salivary transcriptomics biomarkers, like KRAS, MBD3L2, ACRV1, and DPMI; these markers were found to have 90% sensitivity and 95% specificity. The study further allowed discrimination between patients with and without expression of pancreatic cancer. [14]

Thus, this high sensitivity of the marker, suggest their successful application in performing screening tests for pancreatic cancer. Relatively, many new studies have been performed that have evidently proven important role of miRNA during the progression of pancreatic tumors, while promoting several biological pathways. In the current scenario they are considered to be the most reliable biomarkers, due to their better stability in many body fluids, easy detection system and more accurate estimation of tumor progression. Various studies have confirmed the great potential of these markers in early detection of the pancreatic cancer, with an average of 70 % specificity and 72% sensitivity. [15]

Other than these systemic revelations, where salivary miRNA detection system has been confirmed to be the reliable detection, for systemic pancreatic cancer; role of certain bacteria, as disease specific salivary biomarkers have as well been confirmed through various other studies. One of the studies have confirmed that expression levels of two microbial species, such as Neisseria elongata and Streptococcus mitis are significantly different in patients with and without pancreatic cancer. [16] When a regression model is being created, 96.4 % sensitivity and 82.1% specificity was confirmed in the levels of microbial species, named to be Granulicatella adiacens and S.

mitis was confirmed to be significantly different between both chronic as well as acute cancer.

6.7.2.8. Lung Cancer

Lung cancer is being identified to be the most lethal cancer of all; and is being treated with the most significant efforts and higher rate of awareness. In many such cases, early detection is found to be the key for better chances of survival and improved quality of life. However, so far available lung cancer screening tools are found to be invasive with lack of sensitive and specific markers. In this regard, many researchers across the globe have investigated different saliva specific biomarkers, which are found to be useful in allowing noninvasive faster detection system. [17]

In one such case study, a team of scientists have confirmed the importance of five biomarkers, detected in a logistic regression model, like *CCNI, EGFR, FGF19, FRS2, and GREB1;* the study further confirmed 93.75% sensitivity and 82.81% specificity, demonstrating the discriminatory power of the proposed biomarkers for faster detection. Scientists further analyze the markers through microarray, and confirm the specificity of the said markers to be relevant in lung cancer detection. [18]

Similarly, a proteomic study confirmed the importance of three different proteins, such as HP, AZGP1, and human calprotectin to be significantly higher in people with lung cancer, as compared to their healthy counterparts. A regression model, in the similar study confirmed 88.5% sensitivity and 92.3% specificity, with higher positive as well as negative predictive value; as an indication of better discriminatory power of the said protein biomarkers.

This high sensitivity and specificity of protein based biomarkers is an indicator of them to be useful in early and speedy detection of lung cancer, through salivary fluids.

Lately, studies have significantly proposed the use of multiplexible electrochemical sensor detection system to confirm expression of epidermal growth factor receptor mutation, directly from different body fluids, including human saliva. The ROC analysis further confirmed that the identification mutational changes in salivary DNA can as well be utilized as a good biomarker for disease monitoring and therapeutic modalities.

6.7.2.9. Gastric Cancer

Gastric cancer ranks third in fatalities in the world, after liver and lung cancer. Through conventional detection system, endoscopy is the only available option for confirming diseases, associated with gastric problems, including but not limited to benign and malignant diseases. Thus, there is an urgent need for quick detection system, which is beneficial in the context of fast and easy detection of benign and malignant diseases. However, diagnostic technique with reduced specificity and sensitivity of diagnostic techniques are to be upgraded for quicker identification of more reliable biomarkers, for early detection.

Much effort is being invested to search for the said biomarkers in different body fluids; however, so far only one study could confirm diagnostic biomarkers from saliva. The study has confirmed identification of four distinct salivary markers, which could be significantly be differential in expression in both gastric cancer as well as control group. The expression levels of these biomarkers were confirmed through mass spectrometry, wherein the sensitivity as well as specificity was found to be 95.65% and 100% respectively. The scientists could suggest possible role of these four biomarkers in salivary detection system, on the basis of higher level of expression. [19]

6.7.2.10. Oesophageal Cancer

It has been ranked as the eighth most common cancer with very poor survival rate; with higher incidents in men than in women. Although aberrant expression of miRNA was confirmed in many other body fluids, like serum, plasma and tumoral tissue; the studies as well discovered that five miRNAs, such as miR-144, miR-10b, miR-451, miR-486-5p, and miR-634 can be detected in saliva of patients, with oesophageal cancer. [20]

The study confirmed some of the miRNA as biomarkers, like miR-10b, miR-144, miR-21 and miR-451 with sensitivity in the range of 79.5%, 43.6%, 89.7% and 51.3% respectively; and 57.9%, 89.5%, 47.4% and 84.2% respectively. Another study confirmed up regulated expression of a marker protein, identified to be MiR-21; which was not significantly correlated with cancer stage, differentiation and nodal metastasis. In case of salivary samples, it is important to note that a team of scientist investigated the diagnostic value of a biomarker miR-144, in the salivary samples of patients with oesophageal cancer. These findings suggest that the proposed

biomarker is highly expressed in each and every sample; thus, confirming its crucial role as a genetic marker for faster detection.

6.7.2.11. Breast Cancer

Breast cancer can be referred to be as the most common cause of cancer, with higher incidents in women; especially in underdeveloped countries. Thus, the development of biomarkers for early detection is the need of an hour, to establish much safer detection systems. Some of the very early markers that were analyzed and confirmed for their safety and efficacy are Epidermal Growth Factor, p53, cathepsin D, c-erbB-2, CA15-3, etc.

Many scientists in this regard, have confirmed elevated levels of these proposed markers in salivary fluids, as compared to their healthy counterparts.

Other than some of the cancers mentioned herewith, there are various other salivary biomarker system having defined range of biomarkers for numerous cancers, like prostate cancer, ovarian cancer, leukemia, etc.

Thus, clinical significance of several salivary biomarkers has been studied so far in many malignancies; moreover, researchers could explore multiple other markers. (Figure 6.1)

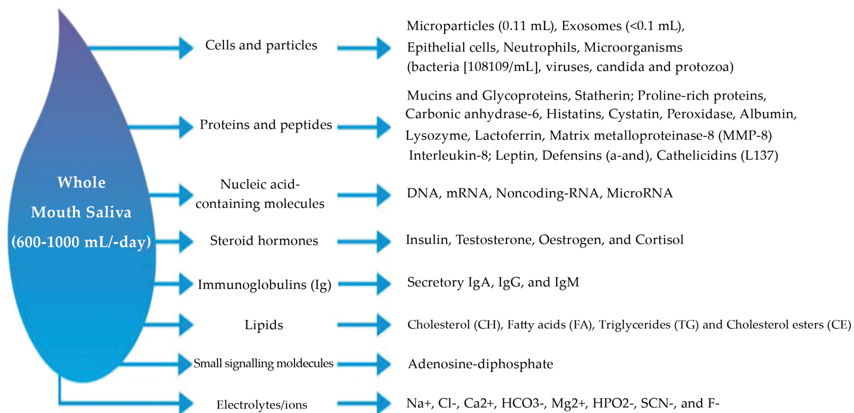


Figure 6.1. Summary of different Salivary Markers [Taken from Open access 4]. [21].

- Thus, salivary biomarkers can be summarized as:
- Salivary Transcriptomic Biomarkers, which can be strategically identified through microarray. The process can further be validated with quantitative PCR.
- Exosomes, are the small, right-side out cell-secreted vesicles of about 30-100 nm, multicellular bodies of plasma membrane. Recently, studies have confirmed that these bodies are responsible for altering cellular environment and salivary messenger RNA, found embedded in them can be the better sources, in order to understand the molecular basis of the disease.
- Salivary micro RNA, (miRNA) play several important functions like, cell to cell interactions, cellular differentiation, apoptosis, stress and immune response; thus, these markers can provide important disease specific information, in order to analyze drug specificity and drug related toxicity, in the context of therapeutic responsiveness.
- Cytokines are present in saliva, which include the interferon, tumor necrosis factors and the interleukins, etc. The availability of vast data base related to natural as well as synthetic cytokine, can be referred to know about cancer profiling of a patient.

Thus, in conclusion, salivary biomarkers can be noted as the most useful tool in diagnosis of variety of diseases, for faster, accurate diagnosis; and should be preferred for being non-invasive, uncomplicated, diagnostic tool.

6.8. URINE BIOMARKERS

Urine can provide lot of information of the body, like pregnancy, hormonal imbalance, aging, daily rhythms and disease condition, etc. mainly due to not being subjected to homeostatic mechanisms. Keeping this in mind, it is possible to utilize urine with identification of different urinary biomarkers, to track disease profile of a person.

Currently, scientists are mainly focusing their attention towards identification of urine biomarker for analyzing kidney disorders, due to their close connection. However, other diseases like brain disorders, cancer or may be some other metabolic disorders are largely neglected; as a matter of fact, identification of biomarkers from urine for these specified diseases is more significant, for being convenient, non-invasive and faster, hence should be worked out. Thus, in the subsequent paragraphs, we have summarized

proteomic as well as metabolomics profile of urine, towards their progressive application in assessing brain disorders, metabolic disorders and cancer specific diseases.

Recently, high throughput and high-sensitivity capacities of urine has been investigated in the field of biomarkers, especially in the proteomics and metabolomics profiling of the urine. It has been studied that certain proteins can be recovered through urine that have been infiltrated by blood or secreted through kidneys and/or urinary tract. Studies have confirmed that somewhere around six thousand proteins can be isolated from urine, allowing us to get an idea about the availability and complexity of proteomics, involved in urine.

Apart from being utilized for analysis of kidney disorders, these proteins can as well be explored to study variety of other diseases, like sleep apnea, eclampsia, cardiovascular disorders, etc. Furthermore, urinary metabolomics can well be assessed for marker analysis, due to the presence of small molecule metabolites; which can reflect about the state of the body to some extent and serve as the informative biomarkers for some other diseases. Hence, metabolomics profiles of the urine can be widely used for confirming markers related to acute kidney disorders, heart failure, liver cancer, breast cancer, etc.

Thus, the specific and stable biomarkers, identified for various body issues can be the key towards successful treatment regimen.

6.8.1. Major Depressive Disorders

It is being identified as the severe neuropsychiatric disorder, generally associated with mood swings, cognition, neuro-vegetative functions, delayed mental activity, etc. [22]

Through conventional mode of detection, so far the diagnosis of particular disorder was based on subjective symptomatic analysis, without any availability of effective measurements. Thus, biomarkers which can confirm assessment with a more reliable outcome will greatly be acknowledged. Several studies have been designed to fulfill the proposed motif, in one such studies; urine of nearly 42 subjects along with 28 control patients was analyzed using peptidomics method, specifically targeting the small polypeptides. Good diagnostic performance was being expressed with 90.5% sensitivity, 92.9% specificity and 91.4% accuracy, by five

different peptides. These markers serum albumin (*m*/*z* 1196.47), alpha-1-microglobulin/Bakunin precursor (AMBP, *m*/*z* 3222.17), heparin sulfate proteoglycan (HSPG, *m*/*z* 4640.35), and Apolipoprotein A-I (APOA1) (*m*/*z* 5072.14) were able to effectively segregate expression at molecular level, among diseased group as well as control group. [23]

As far as the metabolomics is considered, 82 subjects with drug naïve depressive disorders and 82 control patients were evaluated for biomarkers, using nuclear magnetic resonance. The analysis confirmed that metabolites associated with Kreb's cycle, intestinal microflora and tryptophan nicotinic acid pathways were significantly altered in patients, in comparison to control group. Out of these, different availability of metabolites, six metabolites, namely, sorbitol, uric acid, azelaic acid, hippuric acid, quinolinic acid and tyrosine were selected for quantitative analysis and confirmation; on the basis of these two studies, the markers for the said disorders have been confirmed. However, it is very necessary to validated them on large number of subjects to authenticate the same further, for effective clinical applications.

6.8.2. Bipolar Disorder

It is again one of the lifelong and debilitating psychological disorders, analyzed to be affecting more than 1% population, globally. [24]

In spite of being very common, scientists are not able to understand clearly its pathogenesis and underlying cause of concern. So far, accurate diagnosis of the same is lacking due to unavailability of definitive biomarkers; and hence, the diagnosis is solely on the basis of subjective identification of the disease symptoms.

Studies have confirmed that some of the metabolites, like azelaic acid and N-methylnicotinamide can be treated as the potential biomarkers for faster BD detection. The study analyzed urinary metabolites from 78 different subjects and the results were confirmed with 43 normal subjects; through combinational technology of NMR and GC-MS. Te study established, a panel of five different markers, namely to be azelaic acid, 2,4-dihydroxypyrimidine, β-alanine, pseudouridine, and α-hydroxybutyrate. The proposed panel had confirmed to have shown better accuracy; suggesting good integrative approach of the two methodologies to obtain comprehensive model of biomarker, for the said disorders. However, the study did not analyze influence of different anti-psychotic medicine on metabolomics profiling of the markers. [25]

It should as well be noted that some of the bipolar patients express gender specific markers, and hence another study confirmed different panel of biomarker for male specific disorders, such as α-hydroxybutyrate, choline, formate, and *N*-methylnicotinamide andα-hydroxybutyrate, oxalacetate, acetone, and *N*-methylnicotinamide for female specific disorder. The results were validated with their healthy counterparts. [26]

6.8.3. Autism Spectrum Disorders

The disorder can be referred to be as the umbrella term for multiple complex disorders, related to neuronal abnormality. The said disorder, if usually diagnosed within first few years of life, can be tackled in a more effective way; however, due to lack of any firm methodology and only availability of subjective discrimination, autism spectrum disorder, is generally not being diagnosed in a stipulated period of time.

The study estimated metabolomics profiling of children suffering with the said disorder and compared it with the control group, using techniques, such as NMR. During the study, it has been observed that there is an alteration in metabolites as well as co-metabolites, which have been associated with nicotinic acid metabolism. The study as well confirmed altered levels of several amino acids, like taurine, glutamate, and *N*-acetyl glycoprotein in ASD children. Higher concentration of several different organic sugars like 3-(3-hydroxyphenyl)-3-hydroxypropanoic acid, five-carbon sugars, and ribose, upon metabolomics based analysis through GC-MS. It has been further confirmed that there is marked reduction in the marker profiling of compounds like 1,2,3-butanetriol, and propylene glycol. Thus, the study indicated that these metabolites have higher potential to confirm ASD, within a stipulated period of time, without any need of invasive analysis.

6.8.4. Schizophrenia

It is a severe disorder, debilitating emotional abilities of a person and is characterized by diversions from reality to delusions. Through conventional treatments, there are as such no objective methodologies available to confirm the disorder, and hence, doctors are only relying upon subjective analysis of the disability. A study confirming metabolomics profiling of the markers from patient urine, could detect specialized neurotransmitter metabolites in the patient's urine. Their expression was similarly confirmed with the controlled subjects. The study observed increased level of concentration

of several metabolites like, glucosamine, glutamic acid and vanilmandelic acid, addition to which the concentration of creatinine, KG citrate, valine and glycine were as well being modified in urinary analysis of patients suffering with the said disorders.

Another study evidently declared expression of five serum metabolites like glycerate, eicosenoic acid, beta hydroxybutyrate, pyruvate and cystine, through combination of NMR and MS, along with one urinary metabolite. The model can distinguish schizophrenic patients with the normal patients, through accurate analysis. Additionally the study as well suggested an increased level of fatty acids and ketone bodies in the serum as well as urine samples; thus, promoting them to be good biomarkers for faster analysis. Studies have thus, evidently offered the foundation for supporting urinary analysis for the laboratory-based diagnostic tests to treat different types of brain and psychological disorders and also have identified some potential urinary biomarkers accordingly. Although, the recruitment of sample was relatively less in number and there is an increasing demand of large scale clinical analysis.

6.8.5. Neurodegenerative Disorders

Myriad of neurological indications like Alzheimer's disease (AD), multiple sclerosis (MScl), Parkinson's disease (PD), and transmissible spongiform encephalopathies (TSEs), etc. that have been associated with neuronal loss of brain and thus, can be characterized with loss of functional aspects, like mobility, responsiveness, intellectual ability, etc. However, it should be noted that faster detection and effective treatment can delay the progressive loss of neurons and hence debilitating effect of these diseases. Therefore, investigation of different effective markers is very demanding for the speedy detection, diagnosis, monitoring, and treatment of these diseases.

In diseases, studies have reported the screening of complete range of urinary metabolomics markers, using multiple animal models as well as upgraded technologies. In diseases like Alzheimer's, expression of top three markers were analyzed like 3-Hydroxykynurenine, homogentisate, and tyrosine using a combinational approach of NMR and MS; whereas other markers like, 1-methylnicotinamide, 2-oxoglutarate, citrate, urea, dimethylamine, trigonelline, and trimethylamine were detected as potential biomarkers for chronic disorders. The increased expression of peptide marker associated with oxidative stress, including 3-hydroxykynurenine, homogentisate, and allantoin, in mice model, proposed that oxidative stress

is a pre-symptomatic state of the disease. Thus, studies have identified and acknowledged 5-hydroxyindoleacetic acid, methionine, taurine, desaminotyrosine, and N1-acetylspermidine to be promising biomarkers. In fact, taurine is as well proposed in other neurodegenerative indications like Parkinson's disease and Multiple sclerosis. In case of PD, the study evidently confirmed presence of expression of almost 106 metabolites, with significantly different in their expression levels as compared to control groups. The results were confirmed through high performance liquid chromatography and LCMS. [27]

Thus, distinct expression of certain PD markers like hydroxylauroylcarnitine, phenylacetic acid, histidine, dihydrocortisol, and acetylserotonin have been acknowledged to be the potential markers for accurate detection of a disorder, which was altogether dependent upon subjective analysis so far. Thus, conclusively, urine can be exploited as the promising disease specific biomarker; which was largely ignored for disorders other than kidneys. However, the limited evidences available so far on urinary-based biomarker analysis of different diseases call for an urgent demand of using urine as a potential source of biomarkers.

6.9. SERUM BIOMARKERS

Information related to improved biomarker discovery can be achieved through proteomics technologies, due to recent development in techniques like electrophoresis, imaging techniques, protein labeling, etc. along with the development in protein and genomic bioinformatics. In this regard, several technologies related to proteomic analysis have been established, which may be noted as 2D PAGE electrophoresis, surface enhanced laser desorption/time of flight (SELDI-ToF), isotope coded affinity tags (ICAT), iTRAQ and other multidimensional protein identification technology can be exploited for identifying definitive markers in offering early therapeutic solution to cancer. Out of those listed herewith, 2D-PAGE and SELDI-ToF are the leading and the most acknowledged technologies, which can be applied on serum of cancer patients. In fact, other specified technologies, like can as well express great possibility for future marker discovery that are more authentic and reliable. Above all, two dimensional electrophoresis or 2D-PAGE can be referred to be as the most widely used marker analysis technique. [28].

Continuous monitoring of expression pattern of different proteins extracted from different tumor cells, through application of proteomics

technologies can offer wider opportunity to detect advanced biomarkers for faster detection and early diagnosis of cancer. Various biological specimen like cell Lysate, secretome, serum and plasma fluid can be better analyzed for their pathogenesis at molecular levels and identification as well as effective validation for disease associated proteins; using different techniques like 2DPAGE, 2D-DIGE, SELDI-ToF-MS technology, protein arrays, ICAT, iTRAQ and MudPIT have been used for differential analysis of various biological samples, including cell lysates, cell secretome (conditioned medium), serum, plasma, etc.

The first protein cancer marker reported to be carcinoembryonic antigen (CEA), was detected in 1965; through patients serum, for early identification of colorectal cancer.

Other than which, marker relevant for detection of prostate specific antigen for prostate cancer, CA-19 for colorectal cancer as well as pancreatic cancer, CA-15-3 for breast cancer and CA-125 for ovarian cancer are being discovered lately in the year 1970s. However, it should be noted that identification of these biomarkers is not every time effective for clinical situations. Like PSA is well standardized for its clinical relevance, but approximately one third of the patients have been reported to have elevated PSA levels; and have complained about undergoing unrelated medical procedures, due to not being diagnosed with malignant cancer in time. [29].

Moreover, many other types of cancers like lung cancers and skin cancers, etc. cannot be detected at a very early stage, due to lack of significant biomarkers at a very early stage. Thus, it is very imperative to identify firm biomarkers that are reliable in case of early detection of cancers, with the definitive predictive values.

6.9.1. Lung Cancer

Lung cancer is reported to be a major threat to the society, out of its two significant forms, i.e., non-small cells and small cells lung cancer, the former one accounts for about 75-85% of the total patients; and the later one accounts for about 15-25% of the total population. There is scarcity of satisfactory biomarkers for early detection of lung cancer; however, many different markers available like CEA, CYFRA SCC, NSE and ProGRP [30], etc. for differential detection and sub-typing. However, they offer a very low sensitivity as well as specificity towards the said indications and hence cannot be relied upon. Although, various potentially significant biomarkers can be expressed in serum as well as conditioned medium and the same have

been detected using proteomics. A study confirmed expression of PGP 9.5 and other auto antibodies in the sera of subjects suffering with lung cancer, and their expression level was confirmed through antibody based reactive assay and 2D-PAGE. [31]

In one other study, approximately 100-250 different protein markers were identified from serum samples of patients suffering with lung cancer, using 2D-LCMS/MS and 1D LCMS/MS respectively. [32] When these serum samples were analyzed with the help of techniques like 2DPAGE and immuno-affinity chromatography, markers like catalase, clusterin, ficolin, gelsolin, lumican, tetranectin, triosephosphate isomerase and vitronectin were identified and same was being separated and confirmed through ion exchange chromatography. [33]

In yet another study, the biomarker were identified with the help of SELDI proteomics pattern and are further assessed for their differentiation potential from healthy individuals; wherein three protein peaks were identified with 93.3% sensitivity and five protein markers like 11493, 6429, 8245, 5335 and 2538 Da were automatically detected from almost around 208 serum samples; out of which 158 were lung cancer patients and 50 are healthy individuals. The said pattern could obtain specificity of 91.4% for biomarkers. [34].

6.9.2. Breast Cancer

The traditional methods that can be employed to analyze and further predict survival, progress of metastatic diseases and a selection guide to primary therapy in patient with breast cancer were mostly dependent upon confirming anatomical staging of different biomarkers.

Thus, there is a need for discovering different proteomic biomarkers, with relevant accuracy and reliability. During proteomic analysis two proteins namely, up-regulated expression of HSP27 and down-regulated expression of 14-3-3 sigma have been confirmed through 2D-PAGE coupled with MALDI-TOF-MS technology [35].

When comparative analysis was carried out 97% of the patients did not show cancer and among the positive marker tests, all were found to be malignant for cancers. Thus, the study confirm specified marker to be 100% sensitive with 98% predictive values.

Another study, carried out the analysis of markers in 39 breast cancer patients as well as 35 control groups with the help of 2D-DIGE; the study further revealed positive expression pattern for markers like pro

apolipoprotein A-I, transferrin, and hemoglobin and negative expression pattern for markers, like apolipoprotein A-I, apolipoprotein C-III, and haptoglobin a2 [36].

However, in a study analyzing new set of samples a new set of biomarkers were reported to be a complement component of C3a (desArg) and C terminal-truncated form of C3a (desArg) [37].

In a different study, single biomarker Ca 15.3 has been established, using a combinational technology of SELDI/TOF, MS from patient's serum. The marker size was confirmed to be of 4.286 kDa and 4.302 kDa. [38]

The analysis further confirmed to have significantly improved detection of breast cancer, in a very early state of prognosis.

Another study confirmed detection of four different proteins, like CA1 (17.3 kDa), CA2 (26.2 kDa), CA3 (5.7 kDa), and CA4 (8.9 kDa) obtained from serum samples of 49 patients, suffering with breast cancer. Their expression levels have found to be up regulated as compared to their healthy counterparts. The study was confirmed through artificial neural networks and discriminant analysis [39]. The sensitivity of the marker, when analyzed was found to have detected 100 % and specificity was confirmed to be 90%.

Two genetic markers, namely BRCA-1 and BRCA-2, have been confirmed to be the most reliable biomarkers, in case of detecting tumor suppressors both in men as well as women [40]. The study confirmed the analysis of 15 serum samples, obtained through women suffering with breast cancer wherein BRCA-1 mutations were prominent. The proteomic analysis was performed with the help of SELDI/TOF from serum samples obtained from 15 serum samples with positive marker test, 15 negative marker tests and 16 control patients. The analysis confirmed that in case of carriers, sensitivity was measured to be 87% and specificity was around 87%; thus, confirming effective application of BRCA-1 in patients suffering with breast cancer. Conclusively, for process validation and further optimization of their application towards effective therapeutic decision, a multi-protein complex had been developed with the help of SELDI technology to correctly predict their therapeutic outcome.

6.9.3. Pancreatic Cancer

It is often being associated with poor prognosis, due to the fact that many patients don't express overt symptoms, until advanced stage of progression [41].

Currently CA 19-9 is the most acknowledged serum marker for pancreatic cancer, but its application was approved only for assessing treatment response and not towards the detection or for stage specific analysis. It has been statistically analyzed that almost in 80-90% of people with pancreatic cancer; the expression level of proposed marker is high, as compared to their healthier counterparts [42].

However, current ways to diagnose early stage of pancreatic cancer is found to be highly inefficient due to variety of reason. Thus, it is very crucial to identify potential serum markers for pancreatic cancer, which are fast, efficient and reliable in detection.

Accordingly, a study analyzed serum samples from 3 patients with confirm malignancies and 3 normal individuals, using 2D-DIGE coupled with MALDI/TOF/TOF-MS. During the study, scientists were able to isolate 24 unique up-regulated proteins and 17 unique down-regulated proteins in the patient sera [43].

Another study confirmed increased expression levels of apolipoprotein E, R-1-antichymotrypsin and inter-R-trypsin inhibitor in serum proteome analysis obtained through 20 patients with pancreatic cancer and 14 controls with the help of western blot analysis; these biomarkers were found to be 82.6% sensitive and 100% specific in pancreatic cancer diagnosis. In some other study, sera from 32 normal subjects and 30 healthy subjects was obtained and further analyzed using technologies like 2D PAGE; wherein 100% patients sensitivity was observed towards markers expressing pancreatic cancer.(44).

The study further confirmed identification of markers, like fibrinogen A; which was later confirmed to be an ideal marker for differential expression of positive marker from that of negative one. [45].

Sialylated glycoproteins, extracted from positive as well as negative subjects along with the extraction of highly abundant protein depleted serum samples. The study further confirmed that patient with positive pancreatic cancer, secretes different auto-antibodies in the body; accordingly, identification of increased level of expression of different tumor auto-antigen have been confirmed to have played a significant role in early diagnosis of cancer.

DEAD-box protein 48 (DDX48), which is highly prominent to eukaryotic initiation factor 4A, was confirmed to have secreted in 63.64% of patients with recently diagnosed pancreatic cancer. The results were compared with 1.9% of normal controls, using an antibody-based reactivity assay [46].

To confirm effective application of DDX48 as potential diagnostic biomarker, larger cohort of subjects was studies, containing almost 33.33% of pancreatic cancer patients, 10% of colorectal cancer patients, 6.67% of gastric cancer patients and 6.67% of hepatocellular cancer patients. The study further confirmed good discrimination between positive as well as negative expression of auto-antibodies.

Thus, conclusively, so far significantly higher progress rate has been maintained in case of advancements in proteomics technology, enabling scientists all over the world to better understand the underlying mechanism of disease and analyze the pathophysiology of the same. Accordingly, different potential cancer type specific biomarkers have been investigated from different body fluids like serum for faster detection of the disease or towards taking better therapy related decisions using advanced technologies, like 2D electrophoresis or SELDI. Technological advancements can enhance this potential for biomarker discovery in cancer related diseases in a much better way.

6.9.4. Biomarker of Aging

Modern society is more focused towards achieving common goals, such as healthy aging and well being. Due to which, the global population is experiencing major shift towards higher proportions of older people. It has as well been understood by the democracy that much of the burden associated with health related costs and social care in economically-developed countries is conglomerated in the last decade or two of human life. This understanding has created a need to focus on aging [47].

Research on healthy aging encompasses the incorporation of biological processes that are responsible for aging; along with the other reasons like socio-economic and environmental exposures, throughout life which promotes aging and therefore the risk of disability, age-related frailty and myriad of diseases. Scientists are focusing their attention towards development of interventions, which can alter the phenomenon of aging (48).

Several attempts were initiated to identify different markers of aging, since last couple of decades, however the complexity associated with aging phenotype [49] is found to be practically difficult to resolve.

Thus, despite lot of initial efforts [50], there is currently no availability of universally accepted biomarkers, which can confirm aging. This albeit

triggers a need of definitive assessment through biomarkers that are robust and validated for further analysis [51]

The American Federation for Aging Research (AFAR) in this regard, proposed different characteristics of biomarkers of aging, which can be noted as:

- Prediction of rate of aging, according to the total life span.
- Continuous assessment of basic process that are associated with the process of aging
- The analysis should be non invasive without harming the person, which may include a blood test or an imaging technique.
- Results should be compatible with humans as well as animals, like mice.

Biomarkers, which can fulfill all the above criteria, mentioned by AFAR are difficult to exist, moreover, several potential biomarkers of aging have been recognized in the past few years but none has proven suitable globally for predicting the extent of aging. [52]

Aging affects both at the structural as well as functional levels and, in the majority of body systems; it is generally being associated with the gradual functional loss. When achieved extensively on larger organs, these functional losses have been suggested to have profound effects, with physical as well as emotional impact on the individual and on family.

The section is further intended to locate and confirm range of objective biomarkers, which can be obtained through healthy individuals; because the healthy aging should be related to the functional maintenance for maximum time period.(53).

Having said that, analysis of important biomarkers which characterize and quantify deterioration in mean levels, during aging are to be focused. Most literature determines that healthy aging is living in a diseased free state, without any significant trigger of disease related conditions.

Accordingly, considering the hardship that is being faced in the current scientific world to standardize and define the role of different biomarkers, with projections to contain various groups of surrogate endpoints of important functions that are generally influenced by the aging method.

6.9.5. Biomarkers of Physical Capability

Biomarkers are quantified on the basis of person's ability to perform everyday physical as well as mental tasks. Their further analysis and detection is useful for estimation of current as well as future health [54]. According to the previous work done and other references, we tend to elect four sub domains, like locomotors function; strength; balance; and dexterity. Studies have confirmed that men can perform in a better way, in a later part of the life.

Poor performances are associated with higher mortality rates and lifelong disability [55]. In addition, lower levels of physical capability are associated with higher risk of cardiovascular disease (CVD), dementia, institutionalization and difficulties performing activities of daily living (ADLs) [56]

6.9.6. Biomarkers of Cognitive Function

Decrease in the ability of performing cognitive function may reduce independence and can be associated with a kind of neurodegenerative disorder, named to be dementia [56]. Although, still not clear, without proper evidentiary documents; it has been that the onset of cognitive decline is generally starting early in adulthood, for example from around 45 years of age or may be earlier than that in some other cases. [57] Studies have focused more widely on cognitive domains, in human aging accordingly; nine different domains have been identified together with some of the tests commonly used for their assessment, which can be noted as, executive function, processing speed and episodic memory. Upon further availability of assessment time, certain assessment tests on crystallized cognitive ability and non-verbal reasoning can be suggested.

Executive function is markedly reduced during aging [58], expression of an inverted U-shape pattern across the lifespan. Even processing speed declines progressively with age [59] and is associated with higher risk of mortality [60] cardiovascular disorders and respiratory diseases are as well being nourished, due to reduced oxygenated blood supply towards targeted organs [61]. In addition, aging of brain tissue is sensitive towards episodic memory, which is found to be reduced in individuals with mild cognitive impairment and neurodegenerative diseases [62].

6.9.7. Biomarkers of Endocrine Function

Imbalances linked with different hormonal secretions are frequently being associated with the age-related changes, especially in the endocrine system, and can as well impact poor health outcomes. This relative marker analysis can directly be linked with differential secretion of different sex hormones, the HPA axis, growth hormone IGF-1, adipokines, thyroid hormones as well as melatonin. Accordingly, various longitudinal studies have evidently shown that hormonal secretions like estrogen, testosterone, growth hormones like IGF-1 and DHEAS have expressed their connection with premature mortality and physical disability [63]. However, in some other biomarkers, the said connection with aging can be non significant, for example both IGF-1 indicate high mortality rate, both low as well as high. By the age of 30, DHEAS are reduced with higher mortality in older subjects with concurrent frailty. Hormone replacement studies suggest casual association of aging with hormones like testosterone and estrogen, along with the risk of physical frailty and bone health [64] Cortisol can as well be linked with age-related disease and disability, and abnormality in the secretion pattern of cortisol can be associated with increased BP, impaired glucose metabolism and increased incidence of Cardiovascular disorders along with certain other metabolic disorders, in men. [65].

6.9.8. Biomarkers of Immune Function

Since, the field of immunology is well progressed, recently incorporation of advanced technology is being facilitated for further study of age-related decline in immunity. The technique can be named as immuno senescence [66].

To further analyze age-related immune function and inflammatory factors, it is important to understand relative marker profiling of particular disease. Conventionally, the diagnostic field associated with the ability of longitudinal studies to further confirm immune cells with cellular function of mortality as well as with age related functional decline due to higher infection rate or vaccination response is lacking. [67].

Studies like octogenarian and nonagenarian are being carried out for further identification of specialized immune markers like T-cell phenotype, cytomegalovirus, serostatus and pro-inflammatory cytokine status with subsequent mortality rates. Studies have declared that upon further analysis of immune profiling, there has been increase in the expression levels immune

risk profile (IRP) (68), which is associated with mortality in those over 60 years [69].

The IRP limitations are much narrowed, as is does not acknowledge innate immune response like killer cell (NK cell) function, which are primary immune cells, linked with mortality and infectious rates.

The age related increase in the production of systemic inflammatory cytokines can be best studied aspect of inflammatory aging. [70].

Higher plasma concentrations of IL-6 and TNF-α are associated with gait and lower grip strength in older adults. Although some inflammatory aging is evident, centenarians can exhibit some fewer signs of aging of the immune system.

6.9.9. Sensory Functions As Potential Biomarkers of Aging

Sensory functions support independence, better communication pattern, thus allowing good experience of life. However, it has been well acquainted that loss of these functions is more prevalent in older adults than their younger counterparts. They may experience loss of audition and visual loss, loss of locomotion, etc. In the recent years, statistical analysis exhibited fairly rational relation with decreased visual ability with age, which may further reduce ability to perform routine tasks like reading, mobility and social interactions.

Decline in olfactory acuity declines with age, is more commonly exhibited in different studies, especially among men; and can be relatively an indicator of loss of brain integrity in older people. Amongst many early signs of preclinical indications, smell dysfunction is found to be more common, followed by neurodegenerative disorders like such as Alzheimer's disease and sporadic Parkinson's disease [71], can be associated with mortality in the National Social Life, Health and Aging Project [72]. Accordingly, studies have reviewed vision, audition, olfaction, gustation, vestibular function and pain. Many of these functions, without pain have been associated with reduced ability across the lifespan. Sometimes, it has as well been seen that cognitive and motor functions are being overlapped with sensory changes. However, the predictive value of marker expression can be questioned; as does the opportunity to modulate aging-related changes in sensory function through lifestyle or other interventions. Further evidence will be needed before sensory measures can be recommended with confidence as reliable markers of healthy aging.

In this regard, studies have confirmed a panel of markers that can be compared with healthy aging, which may prove to be useful to researchers in understanding limitations and further proposing different cross-sectional and longitudinal analysis. Several settings along with different clinical studies and analysis have acknowledged the need for well established biomarkers. Some prominent markers are as well selected and acknowledged to be in clinical practice and health related research. They are currently being referred as the strongest available markers, detecting age related issues; as their predictive value is being replicated in number of subjects, with defined specificity. The panel may include biomarkers like blood pressure, fasting glucose and HbA1C, bone mineral density, and blood lipids; each of which is considered to be disease-defining and disease specific. Although, markers appear to be predictive of biological age and of the rate of aging in younger healthy subjects; they have as well been shown to reflect subtle changes in age related processes

From the available evidence, it was not possible to rank the domains or sub-domains proposed nor to suggest how information from the various domains might be aggregated to provide a 'healthy aging' score – thus assuming validity and practical utility of the concept. However, combinations of some of these biomarkers appear to predict biological age and the rate of aging among young adults, as well as frailty, and further research in this area should help to identify whether the proposed biomarkers can be combined to produce an overall 'aging score' and the circumstances where in practical utility is being justified.

A further limitation of our work can mainly be attributed to fluctuations in the validity in older population, which appears to be robust in younger-old individuals. Thus approach can practically enhance research related to the same through comparing larger pool of data obtained from number of studies; which can in turn enhance research and healthy aging.

6.10. CLASSES OF BIOMARKERS

Safety and efficacy of different drugs in clinical laboratory can be essentially evaluated with the help of different classes of biomarkers. These markers can be classified as follows:

6.10.1. Safety Biomarkers

These are being categorized under the group of biomarkers, which can be applied to evaluate the most sensitive procedures to assess toxicity as well as clinical progression of the disease; before its practical engagement into costly phase III clinical investigations. This can be initiated through careful selection of required marker tests, generally at phase 1 and 2. It should be noted that this selection can largely variable, depending upon the specimen profile and pre-clinical toxicological evidences. [73].

In this regard, frequently monitored safety markers can play an important role in detecting functional aspects of that particular organ; apart from several other signs like physical examination, electrocardiogram, etc. This marker analysis can be done across different therapeutic arenas, in order to detect wide range of toxicity. These markers can further be classified depending upon the organ as follows:

6.10.2. Liver Safety Tests

A liver is very uniquely placed organ, with the capacity to perform diverse functions that are crucial for life. It is functional between gastrointestinal tracts and perform primary role in maintaining metabolic homeostasis of the body, along with being a first resort for drugs and other toxic elements. Some of the general tests that have been commonly studied in clinical evaluation can be noted as serum alanine transaminase (ALT), gamma-glutamyl transferase (GGT), aspartate transaminase (AST), alkaline phosphatase (ALP), and bilirubin.

Accordingly, changes in these safety markers can be detected depending upon clinical relevance of liver and its significant functional involvement. Out of these specified markers ALT is mainly detected in the cellular cytosol but AST can be isolated from mitochondria; thus, these physiological properties of ALT forced the enzyme to be released at a faster rate in case of hepatocellular injury. ALP and GGT are observed to be membrane-bound enzymes and are evidently detected to be increasing in conditions, which may cause biliary obstruction. The enzyme is detected to be moderately elevating during parenchymal cellular damage. The hepatobiliary origin is proposed to be the leading source of serum GGT, along with renal tissue being the secondary source of the same enzyme.

In conditions like fatty liver due to drug overload or toxic effects of alcohol, expression of GGT can be increased in microsomal structures of hepatic cells. Another important biomarker, associated with excretory

liver functions can be referred to be as Bilirubin. Studies have supported the fact that over expression of bilirubin in both conjugated as well as unconjugated form can be evident in case of disorders, like obstructive liver damage. Similarly, markers assessments can be achieved through albumin and prothrombin expression, based upon their alterations in the synthetic functions due to chronic hepatocellular damage.

According to the latest guidelines promoted by FDA, drug induced liver injury is identified to be as the hepatocellular damage, which can be confirmed, after increased expression of serum ALT or AST. It should be however noted that many drugs with the ability to cause increased in the secretion of serum amino transaminases (AT) activity cannot trigger progressive or severe DILI.

6.10.3. Renal Safety Tests

The human kidneys are essentially responsible for excretory functions as well, in addition to other homeostatic and endocrine functions. Through its important function, the undesirable end products and toxicants are released out of the body. In case of kidneys, markers like creatinine, blood urea nitrogen (BUN), and glomerular filtration rate (GFR) along with serum electrolytes like potassium, sodium, chloride, and bicarbonate have been proposed as the traditional renal toxicity tests in clinical trials. Along with other specified markers, Cystatin-c, β 2-microglobulin, uric acid, clusterin, N-acetyl-beta-dglucosaminidase, neutrophil gelatinase-associated lipocalin (NGAL), N-acetyl-β-dglucosaminidase (NAG), and kidney injury molecule-1 (KIM-1) are nephrotoxic biomarkers.

6.10.4. Hematology Safety Biomarkers

Bone marrow bears multiple toxicity effects, imparted by many different classes of drugs, which may as well include certain cytotoxic compounds. Further to which, toxic effects of these drugs can be reflected through their association with modifications in their other counterparts in peripheral blood. One of the most traditional safety indices, CBS (Complete blood count) includes analysis of total hemoglobin, hematocrit values, red blood cell count, mean corpuscular volume, mean hemoglobin, total leukocyte count, differential leukocyte count like neutrophils, lymphocytes, basophils, eosinophils, monocytes, and platelets [73].

6.10.5. Bone Safety Biomarkers

Bone can be referred to be as one of the most complex connective tissues, which constantly undergoes remodeling. The process includes different stages of development, like a degradation stage, resorption by the action of osteoblasts and a reformation of matrix. In this regard, conventionally two biomarkers were frequently exploited, which can be noted as serum calcium and inorganic phosphates.

6.10.6. Basic Metabolic Safety Biomarkers

Some of the other metabolic markers like triglycerides (TG),blood glucose, total cholesterol, low density lipoprotein cholesterol (LDLs), and high density lipoprotein cholesterol (HDL-c) can be commonly applied as the safety markers.

6.10.7. Other Specific Safety Biomarkers

Some of the action driven biomarkers, depending upon targeted organ or specimen of action can be noted as serum immunoglobulin levels, C-reactive protein (CRP), fibrinogen, thyroid stimulating hormone (TSH), thyroxine, testosterone, insulin, lactate dehydrogenase (LDH), creatine kinase (CK) along with its isoenzymes, cardiac troponin (cTn), and methemoglobin. These markers can be exploited in case of specific toxicities [74].

Table 6.1. Basic Safety Biomarker Panels in Clinical Trials

Sr. No.	**Markers**
01	Aspartate Trasnsaminase (AST)
02	Alanine Transaminase (ALT)
03	Alkaline Phosphatase (ALP)
04	Gamma Glutamyl Transferase (GGT)
05	Urea
06	Creatinine
07	Sodium
08	Potassium
09	Uric Acid
10	Glucose
11	Cholesterol
12	Total Protein

13	Albumin
14	Calcium
15	C-Reactive Proteins
16	Thyroxine
17	Thyroid Stimulating Hormone
18	Complete Blood Count with Platelet
19	Complete Urinary Analysis

6.11. EFFICACY BIOMARKERS

These markers are detected with the sole purpose of disease confirmation, which is very different with safety biomarkers. These markers can be utilized to propose positive clinical outcome of the treatment; in fact, it should be noted that higher the efficiency of a drug, more is the expression of efficacy biomarkers.

These markers can further be classified into Efficacy into different classes like surrogate, predictive, pharmacodynamic (PD), and prognostic biomarkers.

6.12. SURROGATE BIOMARKERS

A should be noted as the final endpoint referred in case of clinical trials as a laboratory or physical measurement and as an indicator of a drug's response towards the disease manifestation. The same surrogate end point can as well be extended as the clinical endpoint [75]. In fact, a clinical end point reflects the health status of the patient, along with disease status; and can usually be correlated with disease efficacy. The clinical end point can be relevant in case, if solid evidences are obtained from the efficacy as well as the regulatory purpose.

The surrogated biomarker can be used as a reference for effective assessment of safety as well as feasibility of drug. These markers are hugely effective, when utilized to know clinical endpoints or effective treatment outcomes. Example of surrogate biomarker can be referred to be as the blood pressure (BP) for stroke or myocardial infarction. Other examples of surrogate biomarkers are cholesterol, LDL-c, triglycerides, etc,

6.13. PREDICTIVE BIOMARKERS

These biomarkers can analyze patient populations depending upon their classification on the basis of responders and non-responders. They can as well be useful in making reliable predictions related to the effective drug outcome or for forecasting the extent of drug toxicity or effective outcome, in particular population. In this regard, Cytochrome P450-2D6 (CYP2D6) polymorphism in 1977 (76) and [77] was discovered to be the new markers. These markers can functionally describe drug exposure, variability with the clinical response, risk for adverse event, genotypic dosing, targeted drug delivery and disposition of genes along with supplementary work like, precaution, interaction, contraindications, patient counseling, nutritional management [78].

Biomarker can be integrated at any stage of drug development; this co-development is indeed important to support data associated with clinical studies, for both drug as well as test approval. Depending upon the regulatory requirements, it is essential to validate marker efficiency on different clinical samples. The table below in this regard, can be referred to identify few oncology drugs and IVD kits, which have been approved by Datamonitor, 2011. It has been estimated that simultaneous development of drug along with its diagnostic markers can facilitate the future trend of personalized medicine. In an ideal situation, a co-development of marker diagnostic test should be during phase III clinical analysis.

6.14. PHARMACODYNAMIC (PD) BIOMARKERS

These are the markers; responsible for elaborating how effectively a drug can be reacted with its targeted organ and impacts its pathophysiological pathway. Such types of biomarkers are responsible for elaborating proof of concept estimation of drug delivery and its mechanism of action, and hence are being identified as the markers of pharmacological response. This class can include majority of biomarkers, which can be obtained through early phases of drug discovery, such as preclinical, phase I, and, probably, phase II. Compared to pharmacokinetic measurements, these biomarkers can easily be helpful in determining the strength of dose and dose schedule, which is mostly suitable for its effective treatment outcome. The biomarker illustration depicts that identification of an intermediary signal can be considered as a signal; through which, the drug has identified its target organ; while the magnitude and the intensity of the signal can access the efficacy of the proposed interaction.

Thus, in a study the role of biomarkers in order to support the proposed mechanism of action has been evaluated through their contribution in clinical investigation; the study further revealed that in almost about 87 phase I oncology trials, the analysis of biomarker contributed to its role in mechanism of action in 39% of the trials, contributed towards the right selection of dosage for subsequent phase II studies in 13% and contributed towards the right selection of dosing schedule in phase II studies as 8%.

Thus, 19% of the patient's population provided being right patient population are reported to be benefitted through these biomarkers. Out of total biomarkers, 36.8% were detected through serum, 25.6% were identified through tumor tissue, 22.7% were detected through peripheral blood mononuclear cells, 3.7% were confirmed through normal solid tissue, and 0.2% was evident in cerebrospinal fluid (0.2%). Additionally, 10.9% of the total markers were confirmed through special in-vivo imaging.

As far as the non-imaging biomarkers were concerned, detection of proteins, cytokines, and enzyme activity in serum, CSF, or tissue Lysate, identification and confirmation of proteins by immuno-histochemistry (IHC), and quantification of expression of DNA/ RNA (79) can be applied.

6.15. PROGNOSTIC BIOMARKERS

Prognostic biomarkers are being employed to predict the relative risk of disease outcome, particular patient population, without any therapeutic involvement. For example, a positive marker detection test for particular population can survive longer or live improved quality of life than another counterpart, which tested negative.

- Some of the studies revealed the importance of expression of prognostic biomarkers, which can be noted as follows:
- Preoperative expression of CA125, can be referred as a prediction for expression of progressive metastatic disease in patients with uterine carcinoma [80]
- C reactive protein expression can be referred as a predictive risk factor in cardiovascular disorders. [81]
- C reactive protein is studied to be used as a predictive marker for disease free survival in case of breast cancer. [82]
- Serum LDH level can be referred for predictive analysis of survival of metastatic brain tumor cases. [83]
- The number of circulating tumor cells, have been evidently

confirmed to predict progression free survival in patients with metastatic breast as well as ovarian cancers. [84]

- HER-2 positive circulating tumor cells have been studied and confirmed as reliable prognostic markers in predicting metastatic breast cancer. [85]

6.15.1. Phases of Evaluation of Biomarkers

Thus, in order to ensure scientifically proven, validated and meaningful application of biomarkers, it is important to evaluate the process for better adoption and transparency. Accordingly, a five phase approach had been proposed by the early detection network committee of the National Cancer Research Institutes, in the year 2002. Accordingly the process can further be categorized as:

- Initialization of phase 1 analysis to identify markers and their further prioritization as per their diagnostic, prognostic and therapeutic value; which could further determine their adoption in routine clinical application.
- Phase II study involve designing, validation and standardization of diagnostic assay with the clear intention of clinical application. Further to which the validation can be done for better reproducibility of results.
- During phase III, the sensitivity as well as the specificity of the test can be evaluated for faster detection of disease, which have yet to be clinically identified.
- Phase IV study evaluates the sensitivity as well as specificity of the marker test, on a prospective cohort of subjects. It is during this phase, a false referral rate can be tested to evaluate the extent as well as characteristics of the disease, which can be detected.
- Phase V study, is used for evaluation of the overall benefits and the risk of the new diagnostic test on a screened population.

On the basis of their application in different phases, these biomarkers can as well be categorized on the basis of study phase. Thus, different types of biomarkers have specified application in different phases, which can be evaluated as follows:

6.15.2. Biomarkers in Phase I study

Pharmacodynamic markers are often focused in phase I study, due to assumptions that they may be able to provide proof of targeted interaction of drug with organ, in order to support drug selection and dosage evaluation for further study. Considering their utilization these markers can as well be referred to as exploratory markers. The phase I studies can further be initiated through sample analysis and/or images obtained prior to and after treatment. The data can as well be collected after comparison with an untreated control.

6.15.3. Biomarkers in Phase II studies

In phase II studies, the biomarkers can be applied to provide the evidences that the drug is suitably interacting with target organ in a pharmacodynamic assessment, which is similar to that of phase 1 study. The application of biomarkers can as well be initiated to evaluate its association with effective clinical outcome. The marker analysis is done to assess patient eligibility and to determine the dose response relationship across a small set of subjects as well as among more homogenous subject population.

6.16. DISCREPANT RESULTS AND ITS MAJOR IMPACT ON CLINICAL TRIALS

6.16.1. Types of Laboratory Errors

Result discrepancies can be a major threat to the concept of personalized medicine, despite of it being potentially beneficial to allow advancement of pharmaceutical research and development; as it may trigger false decision towards its therapeutic applications. Accordingly many different tools and strategies can be implemented to improve quality control procedures of the laboratory, including internal quality control (QC) procedures, external proficiency testing programmes, regulatory compliance, certification and accreditation, licensing, continuous medical educations, and the regulation of lab services. However, some errors need immediate attention, despite implementing all the quality measures in appropriate manner.

Pre analytical phase deals with the errors that are related to the laboratory tests. Upon further analysis, it has been observed that around 90% of the errors that have been currently encountered, within the entire diagnostic process are mainly remarkable during pre analytical phase. It is all the more surprising to note that measures implemented in order to visualize

positive trends towards the reduction of laboratory errors over the past decade are more significant in analytical phase, with limited or no impact in pre-analytical phase. These error rates have been found to have significant impact on production of incorrect results mainly due to sub-optimal or poor specimen quality.

Laboratory errors can be categorized as pre-analytical, analytical, and post-analytical.

6.16.1.1. Pre-analytical Errors

The errors occurring during the test, in between order and the analytical phase are identified to be as pre-analytical errors. These errors are mainly responsible for affecting sample integrity and its suitability for optimized analysis. Some of the most commonly noted pre-analytical variables, which may have negatively impact final analysis, can be noted as follows:

Pre-Preparatory Stage

- Diet and nutritional status
- Before sampling the patient has indulged into forced physical activity.
- Emotional disturbances prior to or during sampling.
- Faulty life style habits, like smoking, alcohol intake etc.
- Hormonal imbalance, like menstruation, pregnancy
- Intake of medications or supplements, which may have interfering effect with measured analytes.
- Unknown diagnostic interventions and clinical procedures other than prescribed one.

Phlebotomy-Related Errors

- Wrong identification of tests or incorrect timing of sampling.
- Improper posture of the patient during sampling
- Wrongly put or missing patient ID..
- Expired tube or anticoagulant usage during collection
- Contamination from skin wipes.
- Incorrect source of blood (arterial versus venous).
- Unavailability of enough quantity of blood may be due to collapsed vein.
- Incorrect dilution factor

- Improper mixing of anticoagulant which may produce partial clotting.
- Hemo concentration due to long application of tourniquet, especially at high pressure.
- Hemolysis due to usage of small needle, vigorous mixing, or pneumatic tube systems with many curves.
- Formation of air bubbles
- Improper collection techniques and/or non homogenous blood sample due to partial clotting during collection.
- Wrong order of blood draws. The recommended order should always be blood culture bottles, citrate, serum, heparin, EDTA, and then fluoride-containing tubes.

Sample processing, storage, and shipping-related errors

- Incorrect centrifugation temperature, speed, or time.
- Inadequate coagulation time for serum separation.
- Lag time between sampling and analysis.
- Necrotic tissue/fluid or wrong tissue collected via fine needle biopsy.
- Inappropriate acquisition and handling of fresh frozen tissue for RNA, proteins, or phosphoproteins analyzes.
- Incorrect sample storage temperature or wrong freezer, e.g., automatic freeze-thaw.
- Inappropriate fixation and embedding of tissue biopsies into paraffin blocks.
- Unsuitable shipping conditions.
- Wrong thickness of tissue sections for histopathology, IHC, FISH, etc.

These variables are observed to be due to faulty pre-analytical processes, mainly due to the lack of reliable guidelines. All available laboratory equipments should be well calibrated as per regulatory guidelines, moreover for the onsite reference there should be easy availability of clear, concise, illustrative lab manuals, well-trained phlebotomists.

Training related to good sampling procedures, like tissue biopsy should be carried out as per regulatory norms, along with onsite practical training, which can be referred to be as some of the mitigating tools for minimizing pre-analytical errors.

6.16.1.2. Post-Analytical Errors

Post-analytical error can be referred to be as the error occurring due to faulty specimen analysis.

- Wrong manual transcription or questionable interface between analyzing instrument and database.
- Improper documentation of test results
- Cross sharing of information
- Incorrect patient identification information entered at time of test.
- Failure to recognize and act on abnormal results, e.g., repetition of samples with unexpected results or panic values.

It should as well be noted that apart from these common error indices, certain uncommon, post-analytical errors can be very serious, especially when producing alarming values without verification, such as:

- Very low platelet count from a sample which was inappropriately collected or mixed.

To minimize errors, mentioned herewith, the central lab has to implement some effective process validation parameters for sample identification and acquisition, proper connection of testing device with the database, identification and process repletion, and probably recollection, of samples with unexpected abnormal values especially those with panic results. It is important for a pharmaceutical company to take proper measures to ensure error free post-analytical processing and reporting.

6.16.1.3. Analytical Errors

Due to combined efforts of lab professionals and in-vitro diagnostic partners and innovation of latest laboratory technologies along with the implementation of a number of quality control and quality assurance (QC&QA) check points, including internal (electronic) QC, liquid QC; many clinical laboratory errors due to faulty analytical processes have been significantly reduced over time. Random errors can affect sample run or may be few sample run within an analytic run; whereas the systemic error can affect all samples analyzed after an error has occurred until it is fixed.

It is to be noted that each assay related to biomarker analysis has a “default” imprecision; allowing value fluctuations from the same sample, when measured multiple times and final value is the average of all the observations or mean value.

Typically, a clinical lab considers an assay to be successful if results from a quality control samples are nicely distributed around the average and within "Average ± 2SD."

Random error can usually be attributed to faulty manual handling, like the pipetting of wrong volumes, formation of air bubbles, sample clotting, as well as improperly mixed samples. Whereas systemic run can affect particular run or just a part of a run, which sometime be extrapolated to few runs or major impact for longer duration of time, which may span the entire life of a testing device. In fact, short-term systematic error can be linked with inappropriately calibration of an instrument or inclusion of new lot of reagents.

6.17. LACK OF TRACEABILITY AS A MAJOR SOURCE OF SYSTEMATIC ERRORS

Thus the quality of range of laboratory results can be variably compromised, due to different factors, such as the lack of accountability; variation from method-to-method or platform-to-platform, or even between different reagent lots, lack of standardization and validation; or at least harmonization of test results. Upon proper validation of assays or technologies via method-to-method and lab-to-lab comparison; there are fair chances of minimizing errors due to lack of traceability aspects, variability due to reagent lots, etc.

Thus, precautions like avoiding lot wise variation of reagents can further assist in error reduction. The problem is not only limited to "sophisticated" biomarker assays, genotyping, etc.; but as well may impact some of the validated chemical assays, which have been in use since decades as standard practice. Apart from which, there can be different gravitational issues as well, like for one sample analyzed by more than 4,000 laboratories using different types of instruments and thromboplastin reagents produced by different companies and lots, the INR (international normalized ratio of prothrombin time) values may be variable between the range of 2.9 and 7.6. Similarly, when some different sample was analyzed for activated partial thromboplastin time (aPTT), thrombin time (TT), and anti-FXa assay (Heparin test) at different locations and through employment of different platforms and methodologies; the ratio of maximum to minimum reported results was varied between 4-fold, 40-fold, and 50-fold, respectively. The ration was observed to be four fold, for another sample analyzed with ALP and LDH.

In some reported cases, 47% variability was reported between minimum and maximum HDL-C results, the sample was on the contrary analyzed at different places; accordingly, whereas in yet another case, which had been approved as a standard procedure, with a targeted evaluation of up to 152 mg/dl. The result was analyzed on different platforms for LDL-c, the evaluation of result variability is between 120 and 202 mg/dl.

Thus, lack of traceability between variability of different lab platforms or methodologies, even for well established technologies like chemistry or immunochemistry analyzers employed by central labs, is primarily due to absence of accuracy towards meeting primary or secondary standards, in order to calibrate devices or methodologies across different brands.

Also, there is at present no existence of any "gold standard" device or methodology to use as a predicate even for well established lab analyzers. Although automation and innovation has improved outcome and efficiency of results, it can significantly promote bias between different devices and reagents.

Thus it is very important to revise legally marketed devices, to get faster approval from FDA (2011) substantial equivalence to (precedent devices).

Although, long term systematic source of error is generally neglected and is often reflected as a major issue, which may create total disconnect between clinicians offering clinical laboratory services on the one side, and drug developers on the other. The results can thus be variable and misleading, when compared with new reference range or cut off values. Thus, some of the samples like Warfarin (2.0 to 3.0 units) indicate slight and/or dramatic anti-coagulation that needs urgent medical intervention; according to the widely applicable INR therapeutic target range. Similarly, the difference between maximum and minimum results from the anti-FXa example can be more than 23- fold variable as compared to the unfractionated heparin (UFH) therapeutic range (0.35-0.70 U/ml). Following the National Cholesterol Education Program (NCEP) guidelines, a clinician may interpret the results from the LDL-c as example as near optimal, borderline high, high, or very high and will treat patient accordingly.

The issue can thus have undesirable effect on decision making towards effective therapeutic modalities, if absolute marker values have been used to compare the study outcome on analyzing drug's efficacy and/or toxicity, or applying biomarkers to connect between different drug candidates of a similar class of compounds.

It is to be understood that clinical studies worldwide may have major impacts, through different locations of biomarker labs, as it is very evidently common that these labs at different locations use different verticals interchangeably to analyze samples from similar trials. In general, a common drug development programme may employ almost a decade or more, thus in between the tenure it is quite possible for them to switch biomarker vendors and as well using different lots of reagents and calibrators. It is as well possible that during this time period, a company may experience major changes in the manual handling of the sample. These variables are never to be neglected, since it can result in discrimination of results, which may lead to wrong decisions making outcome incompatible, almost from different studies. In case of wrong paneling of these markers by certain labs, without testing efficacy or result interpretation, the drug may be wrongly labeled giving unnecessary troubles.

The discrepancies of results obtained from different laboratories could be even higher than those mentioned herewith, since collection of data was achieved through “well-controlled” laboratories for theoretically standardized tests used to manage patients’ marker analysis and health outcome in clinical trials. Thus, it is very crucial to approve biomarker globally, for rigorous data monitoring for better understanding of challenges and result interpretations.

6.18. IMPACT OF DISCREPANT RESULTS ON PERSONALIZED MEDICINE

Targeted therapeutic products can be attained expected safety as well as efficacy, in case of **companion** diagnostics (CDx). So, the threat CDx failure is proportional to the risk of wrongly using therapeutic products. Reportedly, gene transcripts like BCR-ABL (leukemia biomarker and approved CDx) can be analyzed by different hospitals, approximately 140 in number, and the results were observed to be non-comparable with each other, where the number of transcripts reported from 6 CLIA-certified, reliable labs (two commercial and 4 cancer institutions) varied by more than 2Logs. Inducing of common primers/reagents/calibrators, although hard to achieve, improved comparability. With unfavorable effect of inaccurate test results on patient management, if the rate of target mutation is relatively, small the use of loosely validated assays may spoil a trial outcome and impose a wrong go/no-go decision. For instance, supposing the rate percentage of a mutant gene is 20%, and the rate of response to a therapeutic agent is 70% and 10%

in patients with mutant and wild type (WT) genes, respectively, such as tyrosine kinase inhibitors in mutant and WT EGFR, if the biomarker assay is 90% sensitive and 90% specific, which can be considered acceptable or good by some professionals, it would have two implications:

1. In the clinical trial, the efficacy signal will be thinned, as instead of the two arms (WT and M) being cleanly separated (100% WT and 100% M in the corresponding arms) and the efficacy in the M arm clearly demonstrated, the signal in the M arm would be diluted by the carryover from the WT falsely identified as M. In this case, the average efficacy signal, or overall survival, will be less than the 100% M if identified by the 100% but Discrepant Results May Have a Major Impact 419 specific assay. The indication signal in the WT arm would erroneously increase because of the carryover from M falsely identified as WT because of the 10% false negative, but as the majority is still WT, the impact is not as substantial as in the M arm. In this example, the average efficacy signal would be 0.52 and 0.12 in the M and WT arms instead of 0.70 and 0.10 had the assay been 90% sensitive and 90% specific versus 100% and 100%. This indicates that the ratio of efficacy signal (in M arm to WT arm) would be reduced from 7.0 to 4.4. By means of the same model, the efficacy signal in the target population of any given drug, such as Herceptin, changes by reduction in efficacy signal owing to an assay's low performance and the average overall survival for patients with high levels of HER2 and control arm would change from 16 and 11.8 months (Roche, 2010) to 14.7 and 11.9 months, respectively, decreasing the efficacy signal from 1.4 to 1.2.
2. If the biomarker is used as a CDx to qualify a patient for treatment after drug approval, two out of the 20 M will not be given the drug (10%), and 8 out of 80 subjects with WT (10%) will be wrongly treated with the drug. 8. It is therefore concluded that there is no uncertainty that in drug development biomarkers can play a vital role as gear to monitor drug toxicity, prove a compound mechanism of action, prove the notion for which a drug will be used, and predict efficacy and toxicity.

In Biomarker Hypothesis, drug development and personalized medicine seem to be the prospect of drug industry. Still, there is imminent risk to drug development despite the enormous enhancement in biomarkers laboratories' level of quality. A majority source of error is the discrepant

results from different laboratories or even from the same lab employing different platforms or methodologies which implies lack of standardization, even for well trusted safety biomarker assays. Although this source of error is commonly overlooked and is often aggravated by disconnect between clinical laboratory services on the one side, and clinical guidelines, clinicians, and drug developers on the other. Even for trusted standardized tests from "well-controlled" laboratories, with consequent impact on drug developers' decisions and patient management including personalized medicine approach, the results from the same sample can vary substantially.

Therefore, it can cause a higher expense rate for pharmaceutical companies to function and maintain in-house laboratories if its assets are underutilized, due to a global shortage of good laboratory and QA professionals, a consequential difficulty in acquiring and maintaining laboratory certification and licensure, and rapid development of technologies. So this results into the alluring option of outsourcing the lab services.

This brings the hour of contract research organization or academic institution laboratory services, with reduced overhead and operating costs and availability of new technologies to pharmaceutical companies. Since this greater automation has, in general, enhanced laboratory performance over the last decade, it is also a 'double-edged sword.' Resulting that many suppliers applying oversimplifying technology and electronically locking-out laboratories from using competitors' reagents or independent calibrators so as to increase sales and profits by the increase in usage of automation combined with consolidation of instrument/reagent/calibration manufacturers.

As a result, laboratories have become deeply dependent on suppliers for their quality and are frequently forced to revolutionize methods, instruments, calibrators and reference ranges at the whim of suppliers. This has been further compounded by many laboratories attempting to cut costs by reducing experienced and educated laboratory professionals (doctoral level and even master's level) who have the knowledge to maintain stable calibration and optimal accuracy and precision.

For the fact, some of the laboratories have gone further by reducing bench level personnel from 4-year degree certified medical technologists to 2-year associate degree laboratory technologists or lower. implement a road map to fix the major challenge in biomarker laboratory; lack of traceability between different technologies. In the meantime, it still remains the responsibility of drug developers to ensure that a biomarkers lab has accurate tools and skills

to analyze samples from a clinical trial, the assay validation is at the level of the verdict to be made, and that biomarker data are properly interpreted.

6.19. MASS SPECTROMETRY AND MARKER DETECTION

There are two most important factors that will drive unbeaten mass-spectrometry based biomarker discovery studies:

Firstly, implementations of strict experimental design constraints to ensure the biomarkers that are discovered reflect pathophysiology and not analytical artifacts. It is extremely essential to recognize that after surveying hundreds or thousands of features as is common in mass-spectrometry based analysis that false discoveries will be prepared. Utilization of contemporary approaches to false discovery rate correction such as the q-value correct the significance of findings based on the underlying distribution, and tends to overcorrect less than older methods such as Bonferroni correction.

A conclusion is taken through about the matrix to be examined, after ensuring strict experimental design. There are enumerable options and the opinion primarily depending on investigator preference. The 'options' are inclusive of examination of primary patient tissue, patient cerebrospinal fluid, animal model, patient urine, cell culture, and others. In comparison, the ideal biomarker would be present in high concentrations in a patient blood, some investigations, particularly for tumor biomarkers, metabolic categorization of tumor tissue or focus on protein. This workflow works under the hypothesis that a protein or metabolite found highly enriched in tumor may be secreted into the blood. As biomarker studies high throughput screens, analytical sensitivity may lack to detect the protein isoform or metabolite in blood without targeted enrichment or targeted mass spectrometry analysis also leads to direct analysis where tissue lends itself by using new atmospheric ionization detection mechanisms, such as LAESI or DESI. By using these, it provide additional information in the way of spatial resolution in mass spectrometry imaging experiment, but sometimes lack the sensitivity or precision of conventional protein mass spectrometry.

The specimens either tissue or liquid, could be analyzed for either proteins or metabolites and therefore the emphasis of protein biomarker discovery is still on elucidating disease specific protein isoforms. A primary challenge with unknown small molecule metabolite biomarker discovery is elucidating the structure, especially considering the possibility of isobaric

compounds.

To leverage contemporary technology, detection of labile modifications compared to the historical collision-activated dissociation is improved by application of ETD as a fragmentation method. Using iTRAQ, Robust protein quantification usually requires isobaric tag labeling—for peptides, on the other hand glycans could be quantified using QUANTITY where QUANTITY is the most fresh method described for macromolecule detection and providing insight into glycan modification.

Targeted analysis shall typically leverage precise internal standards, consecutively will require commercial availability as well as a priori hypothesis generation. Metabolite detection may be accomplished using a targeted approach by generating a library of MRMs associated with known standards or by untargeted analysis. Therefore using an internal standard infusion may help with both quantification as well as mass accuracy since untargeted analysis has more difficult considerations regarding quantification.

With appropriate researches, abundant potential clinical biomarkers should be generated with the design and utilization of cutting-edge enhancements to transitional mass-spectrometry biomarker discovery workflows. On the basis of mechanistic likelihood success will be vetted by chemists, biologists, and clinicians. This switch to the clinic will be transparent and elaborated by an intended use and only implemented after both analytical and pre-analytical requisites by performing laboratories.

6.19.1. Biomarker and Genomic Techniques

Proteomic research first came to the fore with the prologue of two-dimensional gel electrophoresis. At the turn of the century, proteomics study has been progressively more applied to cancer research with the wide-spread beginning of mass spectrometry and proteinchip. An intense interest in applying proteomics to foster a greater and better understanding of cancer pathogenesis, develop new tumor biomarkers for diagnosis, and premier detection using proteomic portrait of samples. The dramatically reduce mortality has helped in early cancer detection. The thermostable fractions of serum samples from patients with ovarian, uterus, and breast cancers, as well as samples from benign ovarian tumor were analyzed using two-dimensional gel electrophoresis (2-DE) combined with matrix-assisted laser desorption/ ionization time-of-flight (MALDI-TOF)/TOF MS. Of them, alpha- 1-acid glycoprotein and clusterin were expressly down-regulated in breast cancer, whereas transthyretin was decreased specifically in ovarian

cancer. Conventional 2DE method will persist contributing significantly in serum biomarker identification; the gelfree techniques such as LC-MS/MS and SELDI-TOF are expected to greatly facilitate the serum biomarker discovery process with augmented sensitivity, high-throughput and automation.

Recognition and identification of genomic biomarkers for cancer prediction is of great practical value leading to better understanding of cancer genetics, more precise prediction of tumor behaviours and rational treatment selection. Usage of gene expression data generated from microarrays for biomarker selection is very exigent as per high dimensionality and gene cluster structure, where the clusters consist of correlated genes or genes in the same pathway. This application implies development of novel clustering penalized methods for genomic biomarker selection in cancer studies and also investigating their applications in cancer classification and survival studies.

The explicit aims of this study comprise: expansion of effective clustering penalized methodologies for biomarker selection at both the cluster level and the within-cluster gene level. Following approaches are useful: Supervised Adaptive Group Lasso- SAGLasso and Group Bridge Lasso-GBL. Properties of the proposed approaches, including computational algorithms and asymptotic, will be investigated; second one is Classification analysis using proposed penalized approaches, where the outcome of interest denotes cancer status or response to therapy. Logistic classification and ROC based classification will be considered; third Cancer survival analysis using proposed penalized approaches, where the outcome is censored event time such as time to collapse in cancer patients. Especially we will consider Cox and AFT models; and fourth Intensive empirical studies of the proposed approaches using various cancer genomic data. Extensive numerical studies will be used to evaluate the proposed approaches under different clustering schemes and compare with existing approaches. The proposed clustering penalized approaches are expected to produce parsimoneous predictive models and properly account for the gene cluster structure. The associations of cancer outcomes with both gene groups and individual genes, and are expected to behave superior than existing approaches in terms of biomarker selection and predictive model building.

The current progress of proteomics levels up novel avenues for cancer-related biomarker discovery since adopting high throughput proteomic approaches to multiplexed set-ups, provides a minimally invasive screening

procedure, targeting non-fractionated biological fluids, such as blood, has proven to be challenging. In recent years, the technology has made significant progress. A presumptuous way resides that proteome as the global representative of all biological processes, takes place in cancer cells, then the breakthrough of specific biomarkers in the midst of such biological complexity appears difficult in the absence of ultra-high resolution analytical techniques for quantitative measurement of tens to hundreds of thousands of components, and robust data acquisition and analysis techniques to efficiently and reliably process these large datasets. Current progress in proteomics has been largely due to recent developments in mass spectrometry (MS)-based technologies

Particularly, new techniques for the ionization of proteins and peptides, such as matrix-assisted laser desorption-ionization (MALDI) and electro spray ionization (ESI) combined with time-of-flight (TOF), as well as new hybrid mass spectrometers, are now becoming the tools of choice for protein characterization, accompanied by dramatic improvements in bioinformatics tools for analysis of complex datasets. In addition, powerful multidimensional chromatographic and sample labeling techniques have been developed to further benefit from the improvements in mass spectrometry and advancements that have been highly recognized by the scientific community to include two mass spectrometrists, Drs. John B. Fenn and Koichi Tanaka as co-recipients (with the developer of NMR Dr. Kurt Wüthrich) recipient of the 2002 Nobel Prize for chemistry [34–36]. The standard proteomic approach for biomarker research consists of isolation of cell proteins from clinical specimens (tissue or biological fluids such as serum, ascites, saliva, etc.), digestion with proteases such as trypsin, and separation of the resulting mixture by two-dimensional (2D) electrophoresis or liquid chromatography (LC). The desired spots (2D) or protein fractions (LC) are isolated, digested, and peptides are separated by LC and depending on the sample complexity, the low-molecular weight fractions may be further fractionated by ion-exchange chromatography. The peptides are then subjected to electrospray or MALDI mass spectrometry (MS) or MS/ MS analysis for qualitative and quantitative.

Current clinical and pathological markers disappointingly predict early disease development and response to treatment. Standard diagnostic methods, includes of tissue histopathology are now shifting rapidly toward molecular diagnosis due to the rapid progress in proteomic instrumentation. This effective technology can promote different types of proteins and their post transcriptional modifications, in case of disease conditions; which

in turn can accelerate progress towards analyzing novel diagnostic and predictive tools to track many diseases and patient specific individualistic treatments.

6.20. CONCLUSION

Early detection of a clinical condition of a patient that would otherwise present healthy is the current largest demand, which has still not fulfilled. They can be referred as cancer or cardiovascular diseases as well as certain other parameters, which are completely dependent upon the clinical decisions pertaining to the effective improved clinical outcomes.

The field of biomarker discovery, based upon MS-outcome is so far being governed by camps of pessimism and cautious optimism. However, the field has advanced significantly to the extent of offering good platform for better marker discovery in order to reduce false discovery. Although, initial application can be quite limited due to higher cost and poor diagnostic accuracy, as compared to nucleic acid sequencing approach; the limited access of shotgun discovery of drugs and development through proteomics should not discourage aspiring clinicians or scientists involved in biomarker discovery, from adopting mass spectrometry technology as a biomarker discovery tool.

Mass Spectrometry has been commonly employed, with better success rates in field like cancer research and/or cardiovascular disorders, using different analytical paradigm. Thus, success of MS application can be focused in the clinic, in spite of its much-deserved respite and over-pessimism in the field.

Thus, major improvement and advancements can be expected from ideal MS-based biomarker analysis.

REFERENCES

1. Kyle, S., & Jorge, A. (2010). What are biomarkers*? Curr Opin HIV AIDS*. 5(6): 463–466.
2. Armstrong, A., Garrett-Mayer, E., Yang, Y., de Wit, R., Tannock, I., & Eisenberger, M. A. (2007). contemporary prognostic nomogram for men with hormone-refractory metastatic prostate cancer: a TAX327 study analysis. *Clincal Cancer Research*. *13,* 6396–6403.
3. Winsten, S. (1969). The skeptical chemist. *Clinical Chemistry*. *15,* 737–744.
4. Walker, I., & Newell, H. (2009). Do molecularly targeted agents in oncology have reduced attrition rates? *National Revolutionary Drug Discovery*. *8,*15–816.
5. Network Science. (2011) Drug Development: The Short Story-Attrition Rate, In: Network Science Corporation. in http://www.netsci.org/courseware/Drugs/intro/slide06.html
6. FDA: Innovation or Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products. Challenges and Opportunities Report. March 2004, In: FDA. Available from: http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm077262.htm
7. Armstrong, J. A. (2012). Urinalysis in Western culture: A brief history. *Kidney International*. *71,* 384–387
8. Zhang, L., Xiao, H., Zhou, H., et al. (2012). Development of Transcriptomic biomarker signature in human saliva to detect lung cancer. *Cellular Molecular Life Science*. *69,* 3341–3350.
9. Farrell, J. J., Zhang, L., Zhou, H., et al. (2012). Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer. *Gut*. *61,*582–588
10. Zhang, L., Xiao, H., Karlan, S., et al. (2010). Discovery and preclinical validation of salivary transcriptomic and proteomic biomarkers for the non-invasive detection of breast cancer. *PLOS ONE*. *5,* 1531
11. Globocan. Available online:http://globocan.iarc.fr/Default.aspx.
12. Ray, A., Manjila, S., Hdeib, A. M., et al. (2015). Extracranial metastasis of gliobastoma: Three illustrative cases and current review of the molecular pathology and management strategies. *Molecular Clinical Oncology*. *3,*479–486.

13. Frampton, A. E., Krell, J., Jamieson, N. B., et al. (2015). microRNAs with prognostic significance in pancreatic ductal adenocarcinoma: A meta-analysis. *European Journal of Cancer*. *51,*1389–1404.
14. Zhang, L., Farrell, J. J., Zhou, H., et al. (2010). Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. *Gastroenterology*. *138,*949–957.
15. Xie, G., Lu, L., Qiu, Y., Ni Q et al. (2015). Plasma metabolite biomarkers for the detection of pancreatic cancer. *Journal of Proteomic Research*. *14,* 1195–1202
16. Zhang, L., Farrell, J. J., Zhou, H., et al. (2010). Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. *Gastroenterology*. *138,* 949–957.
17. Xiao, H., Zhang, L., Zhou, H., et al. (2012). Proteomic analysis of human saliva from lung cancer patients using two-dimensional difference gel electrophoresis and mass spectrometry. *Molecular Cellular Proteomics*. *11,* M111.
18. Zhang, L., Xiao, H., Karlan, S., et al. (2010). Discovery and preclinical validation of salivary transcriptomic and proteomic biomarkers for the non-invasive detection of breast cancer. *PLoS ONE*. *5,* 1531.
19. Wu, D. L., Zhang, W. H., Wang, W. J., et al. (2008). Proteomic evaluation of urine from renal cell carcinoma using SELDI-TOF-MS and tree analysis pattern. *Technology of Cancer Research Treatment*. *7,*155–160.
20. Xie, G., Zheng, X., Qi, X., et al. (2010). Metabonomic Evaluation of melamine-induced acute renal toxicity in rats. Journal of Proteome Research. *9,* 125–133.
21. Wu, J., & Gao, Y. (2015). Physiological conditions can be reflected in human urine proteome and metabolome. *Expert Revolutionary Proteomics*. *12,* 623–636.
22. Hill, R. J., Chopra, P. & Richardi, T. (2012). Rethinking the psychogenic model of complex regional pain syndrome: somatoform disorders and complex regional pain syndrome. *Anesthesiology & Pain Medicine*. *2,*54.
23. Wang, Y., Chen, J., Chen, L., et al. (2014). Urinary peptidomics identifies potential biomarkers for major depressive disorder. *Psychiatry Research*. *217,*25–33.
24. Merikangas, K. R., Jin, R., He, J. P., et al. (2011). Prevalence and

correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of General Psychiatry*. *68,* 241–251.

25. Xu, X. J., Zheng, P., Ren, G. P., et al. (2014). 2,4-Dihydroxypyrimidine is a potential urinary metabolite biomarker for diagnosing bipolar disorder. *Molecular Biosystems*. *10,* 813–819.
26. Chen, J. J., Huang, H., Zhao, L. B., et al. (2014). Sex-specific urinary biomarkers for diagnosing bipolar disorder. *PLoS One*. *9,* e115221
27. Lang, A. E., & Lozano, A. M. (1998). Parkinson's disease. First of two parts. *New England Journal of Medicine*. 1998;*339,* 1044–1053.
28. Barlesi, F., Gimenez, C., Torre, J. P., Doddoli, C., Mancini, J., Greillier, L., Roux. F., & Kleisbauer, J. P. (2004). Prognostic value of combination of Cyfra 21-1, CEA and NSE in patients with advanced non-small cell lung cancer. *Respiratory Medicine*. *23 98*(4),357–362.
29. Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L., Deciu, C., Haynes, P. A., Hays, L., Schieltz, D., Ulaszek, R., Wei, J., Wolters D & Yates, J. R., 3rd. (2002). Proteomic survey of metabolic pathways in rice. *Proceedings of National Academy of Science USA*. *99*(18) 11969–11974.
30. Ochnio, J., Roginska, E., Kwiek, S., & Rowinska-Zakrzewska, E. (1984). Value of the determination of carcinoembryonic antigen (CEA) in the serum for the diagnosis of lung cancer; evaluation of its extent and prognosis in conservatively treated patients. *Pneumonology Polska*. *52*(7), 313–319.
31. Becker, S., Cazares, L. H., Watson, P., Lynch, H., Semmes, O. J., Drake, R. R., & Laronga, C. (2004). Surfaced-enhanced laser desorption/ ionization time-of-flight (SELDI-TOF) differentiation of serum protein profiles of BRCA-1 and sporadic breast cancer. *Annals of Surgical Oncology*. *11*(10), 907–914.
32. Fujii, K., Nakano, T., Kanazawa, M., Akimoto, S., Hirano, T., Kato, H., & Nishimura, T. (2005). Clinical-scale high-throughput human plasma proteome analysis: lung adenocarcinoma. *Proteomics*. *5*(4), 1150–1159.
33. Okano, T., Kondo, T., Kakisaka, T., Fujii, K., Yamada, M., Kato, H., Nishimura, T., Gemma, A., Kudoh, S., & Hirohashi, S. (2006). Plasma proteomics of lung cancer by a linkage of multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis. *Proteomics*. *6*(13), 3938–3948.

34. Yang, S. Y., Xiao, X. Y., Zhang, W. G., Zhang, L. J., Zhang, W., Zhou, B., Chen, G., & He, D. C. (2005). Application of serum SELDI proteomic patterns in diagnosis of lung cancer. *BMC Cancer*. *5,* 83.
35. Rui, Z., Jian-Guo, J., Yuan-Peng, T., Hai, P., & Bing-Gen, R. (2003). Use of serological proteomic methods to find biomarkers associated with breast cancer. *Proteomics*. *3*(4), 433–439.
36. Huang, L. J., Chen, S. X., Huang, Y., Luo, W. J., Jiang, H. H., Hu, Q. H., Zhang, P. F., & Yi, H. (2006). Proteomics-based identification of secreted protein dihydrodiol dehydrogenase as a novel serum markers of non-small cell lung cancer. *Lung Cancer*. *54*(1), 87–94.
37. Li, J., Zhang, Z., Rosenzweig, J., Wang, Y. Y., & Chan, D. W. (2002). Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clinical Chemistry*. *48*(8), 1296–1304.
38. Li, J., Orlandi, R., White, C. N., Rosenzweig, J., Zhao, J., Seregni, E., Morelli, D., Yu, Y., Meng, X. Y., Zhang, Z., Davidson, N. E., Fung, E. T., & Chan, D. W. (2005). Independent validation of candidate breast cancer serum biomarkers identified by mass spectrometry. *Clinical Chemistry*. *51*(12), 2229–2235.
39. Hu, Y., Zhang, S., Yu, J., Liu, J., & Zheng, S. (2005). SELDI-TOF-MS: the proteomics and bioinformatics approaches in the diagnosis of breast cancer. *Breast*. *14*(4), 250–255.
40. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*. *266*(5182), 66–71.
41. Sakorafas, G. H., Tsiotou, A. G., & Tsiotos, G. G. (2000). Molecular biology of pancreatic cancer; oncogenes, tumor suppressor genes, growth factors, and their receptors from a clinical perspective. *Cancer Treatment Reviews*. *26,* 29–52.
42. Steinberg, W. (1990). The clinical utility of the CA 19–9 tumor associated antigen. *The American Journal of Gastroenterology*. *85,* 350–355.
43. Yu, K. H., Rustgi, A. K., & Blair, I. A. (2005). Characterization of proteins in human pancreatic cancer serum using differential gel electrophoresis and tandem mass spectrometry. *Journal of Proteome Research*. *4*(5), 1742–1751.

44. Bloomston, M., Zhou, J. X., Rosemurgy, A. S., Frankel, W., MuroCacho, C. A., & Yeatman, T. J. (2006). Fibrinogen gamma overexpression in pancreatic cancer identified by large-scale proteomic analysis of serum samples. *Cancer Research*. *66*(5), 2592–2599.
45. Zhao, J., Simeone, D. M., Heidt, D., Anderson, M. A., & Lubman, D. M. (2006). Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: application to pancreatic cancer serum. *Journal of Proteome Research*. *5*(7), 1792–1802.
46. Xia, Q., Kong, X. T., Zhang, G. A., Hou, X. J., Qiang, H., & Zhong, R. Q. (2005). Proteomics-based identification of DEAD-box protein 48 as a novel autoantigen, a prospective serum marker for pancreatic cancer. *Biochemical and Biophysical Research Communication. 330*(2), 526-532.
47. Vaupel, J. W. (2010). Biodemography of human aging. *Nature*. *464,* 536–542.
48. Langie, S. A., Lara, J., & Mathers, J. C. (2012). Early determinants of the aging trajectory. Best Practice and Research. *Clinical Endocrinology and Metabolism*. *26,* 613–626.
49. Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell. 153,* 1194–1217.
50. Baker, G. T. 3rd, & Sprott, R. L. (1988). Biomarkers of aging. *Experimental Gerontology*, *23,*223–239.
51. Johnson, T. E. (2006). Recent results: biomarkers of aging. *Experimental Gerontology*. *41,* 1243–1246.
52. Martin-Ruiz, C., Jagger, C., Kingston, A., Collerton, J., Catt, M., Davies, K., et al. (2011). Assessment of a large panel of candidate biomarkers of aging in the New castle study. *Mechanism of Aging and Development*. *132,* 496–502.
53. Kuh, D., Cooper, R., Hardy, R., Richards, M., & Ben-Shlomo, Y. (2014) editors. *A life course approach to healthy aging.* Proc Nutr Soc. May; 73(2): 237–248.
54. Cooper, R., Kuh, D., Hardy, R. (2010). Objectively measured physical capability levels and mortality: systematic review and meta-analysis. *Bio Medical Journal*. *341,* 4467.
55. Cooper, R., Kuh, D., Cooper, C., Gale, C. R., Lawlor, D. A., Matthews, F., et al. ()2011. Objective measures of physical capability and

subsequent health: a systematic review. *Age Aging*. *40,*14–23.

56. Salthouse, T. (2012). Consequences of age-related cognitive declines. *Annual Review of Psychology*. *63,* 201–226.
57. Singh-Manoux, A., Kivimaki, M., Glymour, M. M., Elbaz, A., Berr, C., Ebmeier, K. P., et al. (2012). Timing of onset of cognitive decline: results from Whitehall II prospective cohort study. *Bio Medical Journal*. 344.
58. Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: mechanisms of change. *Trends in Cognitive Science*. *10,*131–138.
59. Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*. 2000;*54,* 35–54.
60. Deary, I. J., & Der, G. (2005). Reaction time, age, and cognitive ability: longitudinal findings from age 16 to 63 years in representative population samples. *Aging Neuropsychology and Cognition*. *12,*187–215.
61. Shipley, B. A., Der, G., Taylor, M. D., & Deary, I. J. (2007). Association between mortality and cognitive change over 7 years in a large representative sample of UK residents. *Psychosomatic Medicine*. *69,*640–650.
62. Pause, B. M., Zlomuzica, A., Kinugawa, K., Mariani, J., Pietrowsky, R., & Dere, E. (2013). Perspectives on episodic-like and episodic memory. *Frontiers in Behavioral Neuroscience*. 2013;*7,* 33.
63. Mazat, L., Lafont, S., Berr, C., Debuire, B., Tessier J-F, Dartigues J-F, et al. (2001). Prospective measurements of dehydroepiandrosterone sulfate in a cohort of elderly subjects: Relationship to gender, subjective health, smoking habits, and 10-year mortality. *Proceedings of National Academy of Science*. *98,*8145–8150.
64. MacLean, C., Newberry, S., Maglione, M., McMahon, M., Ranganath, V., Suttorp, M., et al. (2008). Systematic review: comparative effectiveness of treatments to prevent fractures in men and women with low bone density or osteoporosis. *Annals of Internal Medicine*. *148,*197–213.
65. Rosmond, R., Wallerius, S., Wanger, P., Martin, L., Holm, G., & Bjorntorp, P. (2003). A 5-year follow-up study of disease incidence in men with an abnormal hormone pattern. *Journal of Internal Medicine. 254,*386–390.

66. Whisler, R. L., & Newhouse, Y. G. (1985). Immunosenescence of the human B cell system: impaired activation/proliferation in response to autologous monocytes pulsed with Staph protein A and the effects of interleukins 1 and 2 compared to interferon. *Lymphokine Research*. *4,*331–337.
67. Fulop T, Franceschi C, Hirokawa K. Handbook on Immunosenescence: Basic understanding and clinical application. Springer Publishing. 2009.
68. Pawelec, G., Ferguson, F. G., & Wikby, A. (2001). The SENIEUR protocol after 16 years. *Mechanism of Aging & Development*. *122,* 132–134.
69. Wikby, A., Månsson, I., Johansson, B., Strindhall, J., & Nilsson, S. (2008). The immune risk profile is associated with age and gender: findings from three Swedish population studies of individuals 20–100 years of age. *Biogerontology*. *9,* 299–308.
70. Franceschi, C., Bonafe, M., Valensin, S., Olivieri, F., De Luca, M., Ottaviani, E., et al. (2000). Inflamm-aging. An evolutionary perspective on immunosenescence. *Annals of New York Academy of Science*. *908,* 244–254.
71. Doty, R. L. (2012). Olfactory dysfunction in Parkinson disease. *Nature Review Neurology*. *8,* 329–339.
72. Pinto, J. M., Wroblewski, K. E., Kern, D. W., & Schumm, L. P., & McClintock, M. K. (2014). Olfactory dysfunction predicts 5-year mortality in older adults. *PLoS One*. *9,* e107541.
73. Craig, A. (2004). Stephens' Detection of New Adverse Drug Reactions, Fifth Edition, John Wiley & Sons, Ltd. 2004.
74. Reagan, W. J. (2010). Troponin as a biomarker of cardiac toxicity: past, present, and future. *Toxicologic Pathology*. *38*(7),1134–7.
75. Lonn, E. (2001). The use of surrogate endpoints in clinical trials: focus on clinical trials in cardiovascular diseases. *Pharmacoepidemiology*. Drug Safety, *10*(6), 497–508.
76. Mahgoub, A., Idle, J. R., Dring, D. G., Lancaster, R., & Smith, R. L. (1977). Polymorphic hydroxylation of debrisoquine in man. *Lancet*. *2,* 584–586.
77. Tucker, G. T., Silas, J. H., Iyun, A. O., Lennard, M. S., & Smith, A. J. (1977). Polymorphic hydroxylation of debrisoquine in man. *Lancet*. *2*, 718.

78. FDA. (2011). Table of Pharmacogenomic Biomarkers in Drug Labels. *In: FDA Website*.
79. Goulart, B. H., Clark, J. W., Pien, H. H., Roberts, T.G., Finkelstein, S., & Chabner, B. A. (2007). Trends in the Use and Role of Biomarkers in Phase I Oncology Trials. *Clinical Cancer Research*. *13,* 6719–6726.
80. Gupta, D., Gunter, M. J., Yang, K., Lee, S., Zuckerwise, L., Chen, L. M., Goldberg, G. L., & Huang, G. S. (2011). Performance of serum CA125 as a prognostic biomarker in patients with uterine papillary serous carcinoma. *International Journal of Gynecological Cancer*. *21*(3), 529–34.
81. Ridker, P. M., Danielson, E., Fonseca, F. A. et al. (2008). Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *New England Journal of Medicine*. *359*(21), 2195–207.
82. Allin, K. H., Nordestgaard, B. G., Flyger, H., & Bojesen, S. E. (2011). Elevated pre-treatment levels of plasma C-reactive protein are associated with poor prognosis after breast cancer: a cohort study. *Breast Cancer Research*. *13*(3), R55.
83. Eigentler, T. K., Figl, A., Krex, D., Mohr, P., Mauch, C., Rass, K., Bostroem, A., Heese, O., Koelb, O, Garbe, C., & Schadendorf, D. (2011). Number of metastases, serum lactate dehydrogenase level, and type of treatment are prognostic factors in patients with brain metastases of malignant melanoma. *Cancer*. *117*(8), 1697–703.
84. Hayashi, N., Nakamura, S., Tokuda, Y., et al. (2012). Prognostic value of HER2-positive circulating tumor cells in patients with metastatic breast cancer. *International journal of clinical oncology*. *17*(2), 10.1007/s10147-011-0260-0.

CHAPTER 7

BIOMARKERS OF CANCER USING PROTEOMICS

CONTENTS

7.1. Discovery of Biomarkers of Cancer 244
7.2. Types of Biomarkers 245
7.3. Diagnosis of Cancer by Using Proteomics 247
7.4. Proteomics Approaches Towards The Early Detection and Diagnosis of Cancer 247
7.5. Proteomic Techniques In Cancer 252
7.6. Clinical Utility 262
7.7. Clinical And Analytical Necessities For Biomarker Presentation 262
7.8. Proteomic Approach In Different Types of Cancer 263
7.9. Challenges 271
7.10. Perspectives For Proteomics In Cancer 272
7.11. Summary 272
References 275

7.1. DISCOVERY OF BIOMARKERS OF CANCER

Biomarkers are a wide range of subcategory medical signs that can be considered as an indicator of biological process. It could be something happening normally in a body or during the development of a disease or in response to a particular medicine in a patient undergoing treatment. For example, a cancer biomarker can either be a substance or a process symptomatic for the existence of cancer in a body. The World health organization's definition of biomarkers incorporates them as "almost any measurement reflecting an interaction between a biological system and a potential hazard, which may be chemical, physical or biological. The measured response may be functional and physiological, biochemical at the cellular level, or a molecular interaction" [1]. Cancers being heterogeneous in nature, their biology is highly complex, thus, complicating the clinical conditions. Techniques like mass spectrometers and high-resolution need to speed up the detection and confirmation of novel biomarkers [2].

Mass spectrometry (MS) based strategies and technology in proteomics to investigate oncoproteome led to the advent in generation of large records of potential biomarkers [3]. Most of the biomarkers are either at the level of detection or under validation. The MS-based proteomic steps towards biomarker discovery is the key in the amalgamation of several oncoproteomic and oncogenomic data that helps realize cancer biology with the increasing array of proteins and genes compiled databases [4]. Profound investigations were conducted using several human biospecimens which included cell lines, blood components (mononuclear cells, plasma and serum), urine, saliva, cerebrospinal fluid and tissue biopsies [5]. They can be manufactured by both; cancer and normal cells; however, they are abnormally regulated in cancer.

Biomarkers offer powerful and active means to consider a spectrum of diseases with its relevance in analytical epidemiology, screening, diagnosis and prognosis. Due to altered levels of biomarkers obtained in samples, they offer classification of risk factors in a disease and also provide information regarding the primary pathogenesis. Biomarkers reveal entire range of ailments from the beginning to terminal stages [6]. Proteomics forms the basis in classifying the cancer based on the origin of the tumor and its various stages as the disease progresses; this is based on their molecular source. Molecules like PTKs (tyrosine kinases) along with their substrates are being discovered as suitable biomarkers for molecular categorization.

Specific profiling and classification of the protein content obtained from compound samples is identified by 2D-PAGE and MS techniques [7].

Cancer being complex and diverse results in high mortality globally. This resulted in the need for investigating biomarkers largely for the early diagnosis of cancer [5]. Biomarkers exist in various cellular and molecular forms such as in DNA, proteins, circulating tumor cells, etc. Extensive clinical research is made in the direction of calculation and projection of cancer. In oncology, they help in pre-detection of cancer in people and to determine the prognosis, evaluate the severity of the disease and also help in determining how a patient would react to the treatment thus helping in determination of optimal treatment approach [8]. Well characterized, specific biomarkers help calculate the applicable clinical outcomes through an array of treatments.

7.2. TYPES OF BIOMARKERS

7.2.1. Biochemical/Molecular Biomarkers

These are the genetic molecules determined in the bodily tissues or fluids. In cancer, these biomarkers are generally gene products like proteins. They are not restrained to a single molecule and may have an array of several biochemical entities which together provide a biochemical inscription [9]. The following are four forms of molecular biomarkers:

- *Genomic Biomarkers:* Based on the examination of the DNA profiles, generally SNPs (single nucleotide polymorphisms) which assist to identify variations in the genomic DNA. Scientists have discovered that all cancers confirm somatic DNA mutations that form in the DNA of individual cells all through a person's life. Somatic DNA being present only in the tumor cell DNA, it acts as a specific biomarker which can be identified and traced [10].
- *Proteomic or Protein Biomarkers:* These are based on examination and studies of the protein profile. They are discovered by means of technologies that analyze multiple proteins simultaneously like in mass spectrometry (MS) and protein microarray. In spite of several proteomic biomarkers been reported only OVA1 is approved of FDA-clearance [2].

- *Metabolic Biomarkers:* They are based on the examination and observation of metabolites that are the intermediates and the outcome of metabolism. They offer impending advantages in specificity and sensitivity [11].
- *Transcriptomic biomarkers:* These biomarkers are based on the assessment of RNA expression reports [12].

7.2.2. Physiological Biomarkers

These are involved with the functional processes occurring in the body. For example, blood flow in parts of the brain affected by a stroke is a prospective indicator to determine success in treatment. Use of physiological biomarkers will increase with the advancement in imaging techniques.

7.2.3. Anatomic Biomarkers

These are related to the anatomy of an organism. They comprise of the structural differences of the various organs like the kidneys or brain. For example, the size of specific brain structures in context to one another stands as a biomarker for Huntington disease.

Based on the different stages when these biomarkers are used, they can further be classified into the following six types:

- ***Risk assessment:*** Predisposition biomarkers help in determining the inclination of a patient towards cancer, informing them of a future risk in development of cancer. For example, presence of BRCA 1 & 2 genes signifies that the patient stands an increased receptiveness to breast cancer. Thus, genetic analysis is a supportive testing tool in case of a family history which indicates the risk of inherting a risk for a health condition.
- ***Detection or screening:*** These biomarkers are the indicators of the presence of cancer in the body. The body responds to tumors by producing and discharging antibodies or sheds serum proteins along with circulating DNA fragments and tumor cells in the bloodstream.
- ***Diagnosis:*** Diagnostic biomarkers can help in determining the base origin of tumor thus it can help in confirming cancer from biopsy samples. This assists in an earlier detection of a disease. For example, metamorphosis in CFTR gene indicates the presence of cyctic fibrosis in a newborn.

- ***Prognosis:*** Prognostic biomarkers offer information regarding the patient's result thus indicating how a disease would manifest and develop after being diagnosed. It is observed that certain breast cancers are more antagonistic than others and prognostic biomarkers help in determining the cancer which may develop quickly and metastasize. For example, Oncotype Dx test inspects 21 genes which help in the conclusion that breast cancer may surface in the patient post preliminary treatment. This aids in taking the decision if the patient needs to continue chemotherapy.
- ***Prediction:*** Predictive biomarkers help in predicting the response of a particular patient to the treatment given as all patients diagnosed with the same disorder may not have the same biological mechanism of the disease. These biomarkers also help in determining the optimal drug dose without causing an undesirable side effect. As breast cancer is a heterogeneous disease, various cancers respond differently to same treatment. This outlook of the personalized biomarkers shows potential in cancer and rare genetic mutation.
- ***Monitoring:*** Biomarkers help in predicting and monitoring the recurrence of cancer post treatment. For example, CA 27,29, CEA levels in biological fluids during follow up determine the later stages of observation in breast cancer patients [13].

7.3. DIAGNOSIS OF CANCER BY USING PROTEOMICS

Proteomics is the study of extensive, inclusive and worldwide study of a given proteome for comprehending the information in cells in normal conditions as well as with alterations in the cells. Proteomics aims at procuring ad deciphering in-depth knowledge of quantity of protein, its alterations and various classifications along with their interactive patterns and the various pathways they follow [14].

7.4. PROTEOMICS APPROACHES TOWARDS THE EARLY DETECTION AND DIAGNOSIS OF CANCER

To offer healthier outcomes for remedial intervention, it is essential that the cancer is detected in its earlier stage. This increases the chances of survival

of patients and with the need of less intense modes of treatment. Majority of the screening and analytical tools for cancer detection is deficient in yielding specific and/or sensitive reports and at times they may be invasive. Proteomics can be defined as the study of all proteins present in a biological system [15].

To characterize proteins to study and detect cancer in early stages, examination at molecular levels of the disease need to be done. This can be followed by further investigating the protein networks along with their interactions with cells and their interactions with cancer cells and tumor microenvironment. Proteomics holds promise in studying this biochemistry for further investigations in cancer. Proteomes are found to be reflective in the original state of cells along with tissues. It has been found to be accurately reliable to function as tumor biomarkers to diagnose and treat cancer [7].

One of the major objectives of proteomics is the identification of biomarkers for diseases such as cancers. Proteomes in cells characterizes all potential gene products. Genomes are simpler than proteomes and hence any protein may exist in numerous forms varying contained in a cell or different cells. Modifications in them can be derived from translation, regulatory, post-translational and degradative progressions, affecting the protein structure, its functions, localization and turnover [16]. As a result, protein biomarkers could possibly be more specific with respect to the cancer type along with the status as compared to gene based biomarkers. Studies have shown that protein based biomarkers obtained from body fluids offer an easy access for diagnostic applications.

Development in proteomic strategies such as identification of proteins, separation and their quantification allows possibility to analyze proteins rigorously and also aids in better observation and knowledge of their functions. Thus, a complete overview of expressed proteins can lead in early diagnosis of diseases, specially cancers, treatment and its prognosis. In the process of discovering biomarkers for diseases, 2-D gel electrophoresis, protein microarrays and MS techniques along with bioinformatics are proving to be the essential and powerful techniques for the identification of proteins [15].

Blood is an important biomarker as it is easy to procure and can be used for investigating routine blood chemistry measurements in patients. Most of the biomarkers present in biopsy samples from cancer tissues are found to be present in the blood as well. Blood includes fragments of circulating protein that are produced in the diseased tissue and its respective microenvironment.

As blood meets all the demands of a biomarker (being specific, invasive early diagnosis and easy to monitor cancer post therapy), blood samples are extensively used for the discovery of various cancers [17]. It is observed that the proteins in urine samples are comparatively more stable than those found in blood samples, thus, making it a better biomarker than blood. Though it is better biomarker, it is essential to ensure the steps of urine collection and storage along with following the uniform protocol throughout the trials using urine as a biomarker. This care is essential as cleavage of proteins may form due to the presence of specific proteases found in the patient urine samples [18].

Proteomic expertise promises to investigate new clinical biomarkers to detect and identify cancer in its earlier stages. This field of technology also promises to discover new remedial targets procured from easily accessible bio-samples. The scheme shows possibility for understanding cancer biology, making it easier to select the right medical assessment for patients. Proteomic advancements used for the identification of proteins and investigating their interactions and roles are found to be well established. However, protein expression profiling used for the discovery of biomarkers and its approval are lagging behind due to reproducibility issues making it a drawback for making application in medicine for cancer treatment [19].

Genome sequencing investigations leads to the study of proteins (bio-molecules) which are translated from the genes, controlling the cellular processes, malignancy and also the progression of a disease. Owing to several cellular mechanisms like phosphorylation, proteolytic cleavage, glycosylation and acetylation, the human proteome encompasses more than 5 lac proteins [20] in compared to 22 thousand protein coding genes [21]. Proteins have the interconnectivity in signaling pathways and it also responds to the stimuli like diseases and the treatment [22]. Additionally, it has been proved that alterations in genes can be related to cancer. Hence, cancer can be related and detected as a proteomic disease as linked to the post-transcriptional steps [23].

Though there is an extensive application of mammographic screening, breast cancer (BC) has the highest incidences due to lack of detection in preliminary stages. Mammography screening, though has helped reduce mortality in BC patients, it can only be implied when there is any physical, visible abnormality noticed in the mammary tissue due to the existence of tumor cells. Confirmation based upon biopsy reports helps in making decision for the removal of the tumor. BC is highly complex

and has varied heterogeneity in individual tumors. This makes the mode of treatment decision very crucial. The standard means for detection is the pathological report which has indicators like tumor size, metastases and the immunohistochemical appearance of the main proteins like progesterone receptor (PR) and HER2 and estrogen receptor (ER) [24, 25].

Due to lack in early detection in the early stages of cancer, it is difficult to monitor the recurrence of cancer using the traditional means. In order to decrease the mortality and reoccurrence of tumors, techniques that help in early diagnosis of cancer need to be implemented. One such technique is the study of body fluids through proteomes. Progress in proteomics mostly in mass spectrometry (MS) has led to the widened exploration in the field of using bio fluids and the study of proteins in normal (control) and diseased conditions [26]. Human urine is found to be one such important bio fluid for the studies in clinical proteomics. This urinary proteome is unpredictably complex and proves to be useful in the discovery of biomarkers for various disorders. Investigation of urine samples with MS led to the emergence of detection and quantification of several unique peptides and proteins [27]. Proteome studies using urine samples have proved to be informative and a number of novel markers to diagnose cancer and monitor its progression are detected [28].

Proteomic investigations using urine samples, is a non-invasive technique for the identification of the biomarkers for BC patients. To detect the urinary proteins in different stages of cancer progression in BC patients, techniques like liquid chromatography tandem mass spectrometry (LC-MS/MS) can be employed. Proteins ECM1 (extracellular matrix protein 1), MAST4 (microtubule associated serine/threonine kinase family member 4) and filaggrin were detected from urine of BC patients which are present in initial BC tissues, thus attesting the link of these proteins and BC. This proved considerable differences in the different stages of BC and the reservoir of protein biomarkers present for the recognition of different stages of cancer [25].

Beretov and others employed LC-MC/MC technique to quantify and characterize the different expressions of urinary proteins in various stages in BC patients. They used Progenesis software for comparing the protein expression in all samples and identified 166 proteins [25]. The 59 major urinary proteins were categorized based on their sub-cellular locations. Protein locations showed that 6% grouped as others comprised of cell

organelle, mitochondrial, nuclear, or unknown sub-cellular basis; 18% cytoplasmic, 24% membrane-associated, 52% of the proteins are secreted. Most of these essential BC proteins are membrane-linked and secreted in nature, which could either be found in host or tumor with the presence of disease; 37 distinctive circulating proteins were expressed in specific stages of BC. DCIS (which progresses to IBC), a noninvasive procedure helps in the identification of biomarkers thus helping in the early action to prevent IBC emergence [25]. Consequently, the detected BC biomarkers have a great deal of clinical significance. BC cell lines stand as preclinical mold that characterizes the various breast tumor subtypes. When the identified biomarkers were analyzed by Western blot technique, the result was outstanding with an increased expression in three biomarkers – MAST4, filaggrin and ECM1; making it evident that urine biomarkers can be identified with BC.

Proteomic outcome from studies reveal that the proteins identified in BC urine samples are involved in the commencement of LXR/RXR and acute-phase response pathways that are dynamic through inflammation and/or act as an input of the immune reaction to cancer. Other pathways comprise of manufacture of reactive oxygen species (ROS) and nitric oxide in macrophages along with IL12 signaling and manufacture in macrophages. Research by Antalis and team suggest that the migration of BC cells depends on the accessibility of cholesterol esterification and exogenous lipids; cholesterol being the structural constituent of the cell membrane and also of the proliferating cancer cells. Thus, cancer cells may have an amplified demand for cholesterol. The mutated cells are assumed to increase the lipid biosynthesis and consumption of cholesterol from the bloodstream [29]. Studies also suggest environment rich in LDL-cholesterol to promote BC sequences by initiating the main signaling pathways and altering the cell behavior. The LDL-cholesterol signaling pathway was responsible for inducing BC invasion and proliferation [30].

Growing substantiation highlights that the immune response of a person plays an essential role in BC. Several network pathways are implicated in response to BC. Hence, the immune response urinary proteins can be used as the perfect BC biomarkers to diagnose and monitor the progression of the disease. The relationship between these recognized pathways from the urine samples and BC is intricate as well as biologically significant [31].

Existing screening and diagnosis that are based on proteins for cancer are based on the measurement of the presence of serum markers. The table below depicts the most common types of cancer and their respective biomarkers which show altered levels in bio-fluids produced by either the under benign or non-tumor tissues [19]. (Table 7.1)

Table 7.1. Some Common Cancer Biomarkers That Have Acquired Validation

Biomarker	**Cancer**
Carcino Embryonic Antigen (CEA)	Colorectal cancer
Prostate-specific Antigen (PSA)	Prostate cancer
Carbohydrate antigen 19-9 (CA29-9)	Pancreatic cancer
Alpha-Fetoprotein (AFP)	Hepato Cellular Carcinoma (HCC)
Cancer Antigen 125 (CA-125)	Ovarian Cancer
C16α-hydroxylation, CA15-3, HSP90A and PAI-1	Breast cancer

Protein CA-125 levels in the blood are prominent only in half of women who are in the early stages of ovarian cancer. They also have low specificity as benign conditions like pregnancy and endometriosis may elevate the CA-125 levels [32, 33]. Abnormal levels of AFP expression are seen in only two-thirds of HCC patients [34]. These facts emphasize on the necessity of discovery of sensitive and precise tumor markers.

7.5. PROTEOMIC TECHNIQUES IN CANCER

7.5.1. Preparation of Proteins for Detection of Biomarkers

Development in proteomic techniques like protein separation, amplification and detection made it probable to analyze proteins and study them in detail and also helped in learning about their functions [35]. Thus, a complete overview of the deciphered and articulated proteins may lead in the direction of superior and enhanced diagnosis and treatment in cancer. 2D gel electrophoresis, MS and protein microarrays along with bioinformatics strategies together result in the excellent research for identification of proteins which is an important part in the detection of cancer biomarkers.

7.5.2. 2D Gel Electrophoresis (2DE)

Proteomic investigation using 2DE for the separation of proteins based on the uniqueness of two distinctive proteins, charge and size has been the traditional pathway followed. The use of immobilized pH gradients and superior bioinformatics have enhanced the comparability and reproducibility of the technique, however, low throughput continues to be a serious obstacle for 2DE to become a practice for clinical laboratories [16]. 2DE has great potential for research, for example, 2DE applied in ovarian cancer investigations assisted in the detection of dissimilarity between cancer and normal sample proteomes [35]. Introducing fluorescent two-dimensional differential in-gel electrophoresis (2D-DIGE), aids in the shortcoming of 2DE, enabling accuracy in protein separation, identification and quantification. 2D-DIGE is currently an essential device for the cancer biomarker discovery, including ovarian cancer [36].

7.5.3. Mass Spectrometry (MS)

MS is the major technique which aids in the cancer biomarker discovery, which also determines the accurate charge of proteins their mass and further helps in the identification of the protein profiles or definite precursor proteins. MS mechanism comprises of three basic components: the source that generates ions for identification, the mass analyzer that detects the ions depending on their mass-to-charge ratios and the detector that enumerates ions identified by the analyzer [37]. Recently, MS tools have improved: become extremely sensitive and the extreme machinery has been improved for the investigation of bio-molecules, specially proteins and peptides. Currently, this strategy is highly sensitive for picomole to femtomole range which is determines the detection of cancer biomarkers like small polar molecules, proteins, oligonucleotides and peptides along with post-translationally modified proteins like phosphoproteins and glycoproteins. On the other hand, MS does not help in the identification of profoundly glycosylated and high molecular weight proteins [37].

Two of MS-based proteomic techniques frequently used for the discovery of novel biomarkers are:

Matrix-assisted laser desorption and ionization time-of-flight (MALDI-TOF) and surface-enhanced laser desorption and ionization time-of-flight (SELDI-TOF). MALDI-TOF aids in the identification of nano to pico molar quantities of proteins. This procedure produces a co-precipitate of UV-light absorbing matrix and the protein which is under investigation. Laser irradiates the co-precipitate followed by the acceleration of the ionized bio-molecules in an electric field, entering the flight tube. Inside the tube the molecules are segregated depending on their mass by producing individual and distinct signals [37]. SELDI-TOF apparatus engages protein chip application, chromatography and MS derived investigative results. It involves confining proteins on a resin 'chip' which dissociates and separates proteins depending on their hydrophobicity, charge and other characteristics. The detected proteins are later assayed by means of the technique of desorption ionization time-of-flight mass spectroscopy [38]. This technique confines proteins of significance on a protein chip array straight from the original source without the need to prepare the sample previously, thus, making this technique with great potential for clinical applications [16]. This extremely sensitive procedure makes it possible to identify thousands of proteins which are attached to a single location in minuscule concentrations.

The two MS procedures hold benefit of having the capacity to be flexible to high-throughput situations whilst detecting the several parts of the proteome which includes the low molecular weight proteins which were not capable of being studied straightforwardly in the past [39]. Scientists have emphasized on the significance of reproducibility and the reliability of MS techniques for protein profiling. Findings from studies in tissues from ovarian cancer and normal specimen concluded that distinctions in the proteomic profiles exposed in a research were owing to the processing of sample and not in the principal biology of cancer [40]. Summary of studies regarding reproducibility of MALDI profiling of proteins and current steps to enhance its analytical presentation along with pre-fractionation strategies, automated sample processing, pre-structured target exteriors standard matri co-crystallization, improved MALDI-TOF MS instrument components, quality controlled samples, internal standards algorithms for normalization and peak detection and replicate measurements. Albrethsen, suggested additional assessment and optimization of MALDI-TOF MS prior to is use for regular analysis [41]. These studies affirm the quality evaluation of improved MS proteomic technology is essential in the stream of proteomics.

7.5.4. Protein Microarrays

Identical to gene arrays, protein microarrays have surfaced as a capable procedure for the analysis of large quantity of proteins at high-throughput level and their alterations example their phosphorylation status. These microarrays are used to outline the proteome of cells by the application of antibody-antigen interactions [39]. The arrangement of protein microarrays is divided into two main classes:

- reverse-phase arrays (RPAs): the lysates of cells are arrayed and investigated with antibodies.
- forward-phase arrays (FPAs): the antibodies are arrayed and investigated with cell lysates [41, 42]

Unlike FPAs, RPAs do not oblige in tagging of cellular protein lysates and also and comprises of a sensitive high-throughput stage for pathophysiologic studies, marker screening and therapeutic monitoring [35]. RPAs also are capable of investigating signaling pathways by means of minute number of cells separated using laser capture micro-dissection (LCM) or cultured cells procured from human tissues obtained during clinical trials. Conversely, RPAs has limitation for the need of specific antibodies. RPA approach is used in ovarian cancer to research on the disease progression and profile signaling pathway, identifying the therapeutic target whilst suggesting the suitable prognostic indicators [42].

Biomarkers for cancer can be detected by proteomic technologies like ROMA and phi, both of which have been approved by the FDA.

ROMA (risk of ovarian malignancy algorithm) has a section of two proteins: human epididymis protein 4 and CA 125. Hellström, *et al., detected* the human epididymis protein 4 to be over expressive mRNA in ovarian cancer by the application of cDNA microarray analysis [43]. phi includes a group of three investigations: free PSA, total PSA and pro-2PSA. PSA isoforms were revealed by characterization and protein purification techniques for example by liquid chromatography.

ROMA and OVA1 are known to enhance the activity of CA 125 to predict the occurrence of ovarian cancer in patients while phi enhances the activity of total PSA and free PSA in the detection of prostate cancers whilst preventing unwanted biopsy. At times, the result could suggest non-disease-related artifacts to be present in the patient samples, hence, precautions should be taken for designing the clinical research [2].

7.5.5. Nanotechnology

Nanotechnology produces techniques and tools of creation ranging from 1-100 nanometers. These devices hold the capacity to notably increase the standards of proteomics. This can be concluded by observing its benefits like the ability to investigate low abundance targets, reaching the target protein *in vivo* throughout biological and physical barriers and the ability to maintain a technique to translate the protein biomarker discovery for innovative diagnostic and therapeutic examinations. Nanowire arrays are used to measure biosensors which calculate the miniscule amounts of protein biomarkers present in the given biological fluid [44, 45].

7.5.6. Bioinformatics

Database designing, data comparison and interoperability, predictions based on the protein structures, protein and gene expression investigations, modeling for different biology systems, data modeling and ontologies and vocabularies are the roles played by bioinformatics in the study of cancer biomarkers/proteomics [44]. This technology can enhance the accuracy and eminence of the proteomic studies if it is applied from the earlier stages of cancer by developing, investigation and refining the results statistically. It also aids in the investigations of accurate experimental protocols and provides details for the minimum necessary sample size along with population specifications derived from each investigation. This technique aids in understanding of precise hypothesis trying protocols along with the execution of practical algorithms [45]. Bioinformatics allows scientists from various streams of research to exchange ideas; understand and interpret the obtained data by applying developed information retrieval systems also allowing visualization of data [47]. Thus, bioinformatics is highly essential to amalgamate data from several technologies and investigations to obtain through perceptive of the fundamental biological events. This is achieved by using and developing the data analysis based on network [48].

7.5.7. Biospecimens

Scientific research relies completely on the quality of biospecimens which would be utilized for measuring protein and genetic expressions along with connection of data by way of clinical results. Cancer diagnosis and management is initiated with the diagnostic biopsies. This is trailed by the removal of tumors by surgeries. These are the prospects for procuring the necessary biospecimens. Recently, less invasive techniques for collecting

serum and biological fluids are been applied along with the analytical techniques like GC/LC-MS. To procure accurate and reproducible data, trial sample assortment, conservation and preparation for all biological replicates helps to avoid obtaining false results [45].

7.5.8. Label Free Techniques

Labeling techniques consists of intricacy like high expenditure on reagents, requirement of larger sample size along with incomplete labeling. To prevail over these issues, scientists are applying MS based label free shotgun techniques in proteomics. This technique offers high throughput while leading to discovery of novel potential biomarkers. Label free strategy uses the basis of assumption that the peak area of a peptide in the chromatogram is directly proportional to its concentration [49]. This technique is based on:

- measurements of ion intensity alterations like peak heights or peptide peak areas observed in chromatogram
- spectral counting in the MS/MS analysis

Label free techniques are lately being implemented for the ultimate quantification along with relative quantification of proteins or peptides. Earlier, protein abundance was predicted employing protein abundance index (PAI). In later years, this got replaced by exponentially modified PAI (emPAI), routinely used for deciding the absolute protein abundance. Another technique that uses modified way of spectral data expressed absolute protein expression (APEX) profiling to detect the absolute protein concentration.

Protein Lynx from Waters, SIEVE from Thermo Electron and Decyder MS from GE Healthcare are a few software's that are available for label free investigations. It is useful in the discovery of candidate biomarker using clinical samples. Researchers identified N-glycoproteins as the probable biomarkers for hepatocellular carcinoma [50] and also detected differentially expressed proteins in samples of K562 human erythroleukemia cells [49]. This technique shows potential for shotgun quantitation, is cheap, simple and less complicated for analysis. This technique is found to be semiquantitative, less precise for short proteins and low abundance [45]. While applying this technique for quantitation, the correlation of MS/MS spectra with a protein is an approximate value due to the errors resulting from false identification. Low abundance proteins may be present in the sample even if the spectral count be zero. Also, large proteins may create additional tryptic digests

resulting in extra spectral count. Indication for a given peptide is presided by factors like ionization in electrospray and efficiency of fragmentation. Hence, MS/MS spectral count that accounts for the detection of a protein may only be used as an indicator of its profusion in the respective sample [51].

MS technique is restricted to high dynamic proteins traced in the serum samples. Hence, before detecting biomarker proteins, the proteomes in the biological fluids have to be separated by precise characteristics like, glycoproteome enrichments or glycoproteome, hydrophilicity or hydrophobicity via strong anion exchange/strong cation exchange (SCX/SAX), ion charges, molecular weights, normal or reverse phase chromatography, etc. Glycoproteome examinations in the biological fluids are found to have great advantages in the discovery of cancer biomarker. Almost half of the serum proteins are recognized to undergo glycosylation [35]. The glycoprotein glycosylation status, their forms and degrees are found to be altered by disease conditions including cancers. Actually, glycosylation outline patterns have been documented as trait in epithelial cells [52].

Each protein, post preparation can be examined by various process for identification, verification and proper validation. The proteins are identified by MS techniques. Detected biomarkers from the samples are then confirmed by ELISA or Western blot techniques. Validation for these detected biomarkers first needs to be performed in sera from patients with early stage cancer. Later on, several biomarkers should be tested and approved depending on their capability to diagnose asymptomatic patients before the detection of cancer by physical examination [53]. Assays which would allow for simultaneous assays for many biomarkers in small volumes should be available for trails. The Luminex LabMap technology merges the standard technique of a sandwich immunoassay with a fluorescent bead-based technology, thus allocating multiplex and individual analysis for nearly 100 diverse analytes using small (50 µl) serum samples to measure numerous markers [35].

Proteins are involved in the occurrence of cancers such that they contribute to the formation of tumors, their progression and metastasis, hence the information of each molecule and deciphering its signaling pathway can help in the identification and characterization of proteins involved in a particular cancer. This also can aid in suggesting in designing personalized smart drugs and also in offering combined therapies [54]. The

fundamental process for characterization of proteins involves two main steps. First is to separate the proteins using two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and later using MS for identifying the proteins present in a given sample [7]. The proteins disintegrate based on their respective physiochemical properties followed by the application of MS technique for the desired proteins. Also, surface-enhanced laser desorption/ ionization time-of-flight (SELDI-TOF)-MS forms the basis for analysis of serum and plasma samples. Reverse-phase protein arrays, isotope-coded affinity tag technology and antibody microarrays are promising substitutes for proteomic modus. Though 2D-PAGE stands out to be the classical technique, it is slow, not in power for screening large sample sizes. However, it still aids as a resourceful technique to diagnose several cancers in humans. Expression proteomics is the study where the samples are screened for differences occurring in the protein patterns existing between control and tumor samples. Thus, proteomics assists the progression in cancer research as follows:

- Improvement of molecular detection techniques or biomarker discovery for diagnosing cancer:
- Proteomics offers detailed analysis of molecular pathology of cell-signaling in cancer.
- Facilitating the integration of diagnosing cancer and the therapeutic phase of cancer through drug target.
- Enhance the classification of cancer
- Toxiproteomics which can aid in the progression of secure therapies for cancer by investigating and identifying the poisonous effects caused by anticancer drugs in the initial phase/stage of cancer.
- Assists in monitoring patients through the various stages of cancer.

Early diagnosis is essential for the prevention of cancer and molecular biology been studies by proteomics, helps identify cancer at a premature stage and also offers manageable modes of treatment. The biomarkers CA-125 and PSA, though helpful are still not as sensitive and specific to detect the early stage of the disease. Majority of times, cancer is discovered only after the cancer cells have begun their attack on tissues.

Cancer has an intricate molecular pathway and the gene expressions can be altered due to gene mutations, life style or changes in environmental conditions. As several genes and other factors are drawn in cancer, the means

of developing cancers is different. Identification of the molecular pathways and biomarkers helps in prescribing cancer therapy [7]. A few biomarkers for cancer along with some characteristics are presented in table 7.2. To observe the molecular alterations which create malignant and phenotypic changes, proteomic techniques to investigate protein expression, enzyme activity and modification are been employed.

Also identification of the key proteins along with their altered regulatory role gives a new direction to the evolutionary process of caner/tumor cells exposing new phenotypes and functions. Proteomics is also used to expound the signaling pathways and molecular mechanics which lead the progression of cancer [55]. (Table 7.2)

Table 7.2. Cancer Biomarkers and Their Characteristics

Tumor biomarkers	**Cancer and bio-fluid**	**False positive cancer conditions**	**Clinical**
Thyroglobulin	Thyroid (Blood)	Prescribed after thyroid is removed to evaluate treatment	Determines reoccurrence
Her-2/neu	Breast (Breast tissue)	Oncogene present in multiple copies in 20-30% of invasive breast cancer	Determines prognosis and guides treatment
CA-125 (Cancer antigen 125)	Ovarian (Blood, Urine)	Elevated levels with endometriosis, some other benign diseases and conditions; not recommended as a general screen	Helps to diagnose, monitor treatment, and determine reoccurrence
PSA (Prostate specific antigen), total and free	Prostate (Blood, Urine)	Elevated levels in benign prostatic hyperplasia, prostatitis and with age	Screen for and help diagnose, monitor treatment, and determine reoccurrence
AFP(Alpha-feto protein)	Liver, germ cell cancer of ovaries or testes (Blood)	Elevated during pregnancy	Help diagnose, monitor treatment, and determine recurrence
CEA (Carcino-embryonic antigen)	Colorectal, lung, breast, thyroid, pancreatic, liver, cervix, and bladder (Blood)	Elevated in other conditions such as hepatitis, COPD, colitis, pancreatitis and in cigarette smokers	Monitor treatment and determine recurrence
CA 19-9 (Cancer antigen 19-9)	Pancreatic, sometimes colorectal and bile ducts (Blood)	Also elevated in pancreatitis and inflammatory bowel disease	Stage disease, monitor treatment, and determine recurrence

CA 15-3 (Cancer antigen 15-3)	Breast cancer and others, including lung, ovarian (Blood, Urine)	Also elevated in benign breast conditions; doctor can use CA 15-3 or CA 27.29 (two different assays for same marker)	Stage disease, monitor treatment, and determine recurrence
Calcitonin	Thyroid medullary carcinoma (Blood)	Also elevated in pernicious anemia and thyroiditis	Help diagnose, monitor treatment, and determine recurrence
B2M (Beta-2 microglobulin)	Multiple myeloma and lymphomas (Blood)	Present in many other conditions, including Crohn's disease and hepatitis; often used to determine cause of renal failure	Determine prognosis

(Reference: Azodi, et al., 2013; 7).

7.5.9. Toxic Proteomics

Toxicoproteomics is a new scientific technique which merges bioinformatics with proteomic technologies. It is progressed by using qualitative and quantitative proteomic approaches along with its application in toxicology research [56]. Toxicoproteomics identifies proteins and the biological trails which get affected due to toxicants, unfavorable environmental and chemical exposures. Cancer can be induced through the various chemical contaminants present in food, water, air and workplace which are termed as carcinogens [57]. Toxiocproteomics permits to examine the body's response to particular toxicant, which help in determining the pathways followed by the toxicants in carcinogenesis. Studies in the topic affirm that adopting the proteomic strategies which allows for the early detection of biomarker, saves time and expenditure compared to the earlier means adopted for testing carcinogenicity [58]. Additionally, as this helps in early determination of the toxic effects caused by the intake of anticancer drugs, it helps in finding and adopting safer cancer therapies [59].

7.5.10. Monitoring Patient Health

Following diagnosis of cancer, it is essential to observe the patients to the treatment administered to them. Hence, after commencing treatment, testing serum samples with proteomic techniques can help in determining if the patient is responding and also for predicting the therapeutic efficacy. This monitoring will check if the tumor has become resistant to the treatment, which will need modifying the treatment, also called as responder profiling.

Multilevel arrays, which can detect high and low abundance proteins in the same sample, which are also super sensitive, can be designed. Merging of proteomic and protein microarray techniques can represent a new prototype for identifying diseases and also monitoring the disease response to the prescribed therapy [7]. The main goal of follow up sessions with patients is to identify the metastatic disease at the earliest rather than initiating treatment when the patient is symptomatic, thus, improving the chances for survival. Ultimately, both, genomics and proteomics are essential for managing the health of cancer patients by means of designing and tracking of personalized therapy revolutionizing cancer management.

7.6. CLINICAL UTILITY

The measure of the utility of a test leading to the improved patient outcome along with cost-effective care is termed as clinical utility. This is indicative of the pros and cons of the tests on a single patient along with the society as a whole [60]. Though the revelation of clinical utility is not considered for clearance by the FDA, yet significant numbers of decision makers are demanding it prior to compensating the test cost or intervention [61]. It is chiefly reviewed by random control trials, being less prone to being bias. But these investigations require comparatively large sample size, are costly along with being equipped with ethical disputes, and hence are not mostly reasonable [60].

7.7. CLINICAL AND ANALYTICAL NECESSITIES FOR BIOMARKER PRESENTATION

Biomarkers applied for the diagnosis of neoplastic disorders are usually considered as Class III devices, whilst those used for prognosis or monitoring of cancer are commonly considered as Class II devices. The impairment caused to the patient through wrong prognosis of a disease is usually considered low compared to that from wrong diagnosis. Immunohistochemistry protein biomarker assays are an exception to these as weak factors determines if an immunohistochemistry technique is class I, II, or III and is clearly demarcated as a component of the regulation 21 CFR 864.1860, "Immunohistochemistry reagents and kits" [61].

Standardization and validation of the pre-analytical protocol in Protein biomarker assays are essential for the reproducibility of the assay. The

assay developer should describe the clinical cutoff and algorithm in protein biomarker assays [61].

7.8. PROTEOMIC APPROACH IN DIFFERENT TYPES OF CANCER

7.8.1. Lung Cancer

It is the most common cancer in the world with a low survival rate. With the applications of 2D-PAGE and MS, 20 promising biomarkers for cancer were identified [62]. Another investigation comparing normal and lung cancer sample was performed using Label-free quantitative LC-MS/MS which resulted in 62 proteins expressed differentially helping in distinguishing the normal from affected samples [63]. Recent outcome signified proteomic investigation of endo-bronchial lesions may assist the diagnosis of lung cancer. This mechanism also helps in monitoring of high-risk people for lung cancer in observation, supervision and chemoprevention trials [64].

7.8.2. Breast Cancer

Breast cancer accounts for being the number one cancer-related mortality in women. 2D-PAGE and MS investigations analyzed the changes in proteomics infiltrating cancer whilst comparing normal breast tissue. The proteins expressed were recognized as cell defense proteins, structural and folding proteins, enzymes engaged in glycolytic metabolism of energy and homeostasis and proteins engaged in cell mobility and its cytoskeleton [65]. Proteomic studies with 2D-PAGE, MS, Antibody arrays and immunoblotting investigated proteome procured from the interstitial fluid and adipose cells from the mastectomy specimens of BC patients. Overall, 359 distinctive proteins were analyzed, which included hormones, growth factors and cytokines, implicated in a range of biological procedures such as cell communication and signal transduction, protein metabolism, energy metabolism, maintenance and/or, cell growth, immune response, regulation of nucleobase, nucleoside, and nucleic acid metabolism, apoptosis and transport [66]. An additional research suggested, breast tumors deficient in the estrogen receptor-α (ER-α) contains amplified frequency of resistance to therapy and inferior clinical diagnosis. PGRMC1 (progesterone receptor membrane component) phosphorylation possibly can play a role in the clinical differentiation that uphold breast tumors of contradictory ER status. Several studies by 2D-DIGE and MS resulted in detection of myosins,

β-actin, and numerous proteins concerned with actin filament dynamics and organization; which are the ligand dependent multiprotein complex [67].

7.8.3. Colon/Colorectal Cancer

Colorectal cancer (CRC) is found to be the second most life-threatening cancer in the world [68]. Proteomic research was performed using biopsies of whole tissue, cell lines and epithelial cells obtained from colorectal origin. Results achieved were 408 proteins expressed differentially; 83% of which were obtained in a single study [69]. Attestation at proteomic level with large-scale transcriptomic methods was possible in only 25% of the proteins. This suggested that gene expression investigations at proteomic and transcriptomic methods are low in human CRC. Samples achieved and controlled sampled were analyzed by ELISA. This technique revealed 70% sensitivity and 83% specificity for α-defensin in colorectal cancer [70]. Results aim at the concept of integrating serum protein analysis with tissue transcriptomic data as a technique to determine serum biomarkers. By means of SELDI technique, defensin isoforms were affirmed to be prominent in serum samples procured from colon cancer patients and in protein extracts from CRC [71]. A study united 2D-PAGE and SELDI-MS and different levels in PACAP protein, flavin reductase, hnRNP A1, NDKB (NM23H2), calgizzarin, smooth muscle protein 22 and cyclophilin A. Further, immunohistochemical investigation of subcellular localization and tissue distribution of some of the differentially expressed proteins confirmed variation in the sub-cellular protein distribution [72]. Proteomics supportive of one-dimensional gel electrophoresis along with nanoliquid chromatography tandem demonstrated proteome differences between differentiated tumor cells and colon cancer stem cells. Pathway analysis illustrated that "cell death" regulation is extremely different among the two cell types. Fascinatingly, BIRC6 was one of the top up-regulated proteins which actually belong to the class of inhibitor of proteins responsible for apoptosis. BIRC6 is a significant mediator of cancer stem cell resistance in opposition to oxaliplatin and cisplatin. Targeting inhibitors of apoptosis proteins could assist eliminate colon cancer stem cells. This study suggests that discrimination of colon cancer stem cells is escorted by the altered regulation of cell death pathways [72].

7.8.4. Skin Cancer

Skin cancer is the most widespread cancer worldwide of which Melanoma is the most fatal. It ranks 5th and 6th widespread cancer in males and females respectively. The American Cancer Society predicted it to be responsible for around 76,380 cases in the United States in 2016. It is skin growth due to various causes and equivalently variable degrees of cancer. Extreme exposure to UV radiation, arsenic compounds, coal, and immune suppression are the risk factors for melanoma. The analysis and treatment of melanoma is difficult owing to the high rate of metastases. The high frequency of mutations which correspond to the therapy pressures or environmental stress lead to intra-tumoral and/or inter-tumoral heterogeneity makes it hard to assign a single treatment strategy [73]. Overcoming these obstacles leads to successful treatment strategies for melanoma. Previously, chemotherapy was the standard option to treat severe or metastatic melanomas; however, due to the higher efficacy of immunomodulatory (immunostimulatory and immunoinhibitory) antibodies, they are replacing the traditional therapies. Combination of chemotherapy, radiation and immunotherapeutic is seen to have created greater efficacy in melanoma patients. CD137 (41BB), CD134 (OX40), CD40 and CD28 are the common immunostimulatory antibodies used. The onset of proteomics has led to the discovery of various diagnostic and prognostic melanoma biomarkers, satisfying a critical need. Improvement in protein fractionation and analysis techniques has helped in the advancement of proteomics to analyze complex protein samples procured from melanoma patients.

Ipilimumab monotherapy performed at some point in phase II clinical trial executed on patients having metastatic melanoma was not as effective as the amalgamation therapy of ipilimumab and GMCSF (granulocyte macrophage colony-stimulating factor) which proves to be advanced in terms of safety and efficacy [74]. Proteomics is further employed to guide for drug design and also the improvement in treatment algorithms. Therapies based on proteomics are attested to be successful in the investigations performed on melanoma along with the drugs used for curing the cancer [73].

Basal cell carcinoma is the most frequent types of non-melanoma skin cancers in individuals. Studies signify that the role of precise genes in the skin cancer is altered. The specific regulators of cellular proliferation and feasibility along with ARF/p53, Sonic Hedgehog, p16INK4A/CDK4/Rb, Ras/Raf and NF-B pathways are affected [75]. New modalities intended to aim these precise proteins may signify promising approaches towards therapy

for human skin cancers. Since it uses proteomics, an extensive variety of protein profiles has been broadly assembled via this technology. Proteomics using stem cells from human epidermis and experimented with MALDI-Q-TOF MS and MS/MS methods recognized differentially expressed proteins (both, up and down-regulated) after being treated with arbutin. Some of these proteins correlated with p53 tumor suppressor leading to cell apoptosis thus playing an important role in suppression of cancer growth [75].

7.8.5. Renal Cancer

The first investigation for renal carcinoma cancer (RCC) proteome was done by comparing the normal renal against cancer type tissues of kidneys applying 2D-PAGE in ten patients diagnosed with renal cancer. Results showed 43 out of 2789 separated polypeptides were obtained from gel comparision, N-terminal sequencing, immunodetection and amino acid analysis. Four polypeptides were absent in RCC; ubiquinol cytochrome c reductace (UQCR) and mitochondrial NADH-ubiquinone oxido reductase complex I. The later biomarker established the role of mitochondrial abnormality in RCC as it helped attest mitochondria as a drug target in RCC. Another study result showed that the over expression of heat shock 27 protein was detected as a significant biomarker by applying PAGE separation, Western blotting and MS immune-detection methods and the outcome was validated by immunohistochemistry application on tissue sections [76]. A recent proteomic study revealed the expression levels of galectin-1 (Gal-1), profilin-1 (Pfn1) and 14–3-3 zeta/delta (14–3-3ζ) altered in RCC patients. The grouped analysis of the altered expression proteins suggested that protein expression profile for metastatic RCC in non-aggressive and aggressive RCC is diverse [77]. It appears that proteomic pattern and discovery of biomarkers have an important function in the diagnosis and therapeutic characteristics of RCC.

7.8.6. Prostate Cancer

It is held responsible to be second in the list for cancer mortality in males. There is rigorous need for research in this cancer as its screening is outstanding due to the inadequacy of the prostate-specific test performed for the early detection of prostate cancer antigen (PSA) [78]. Studies suggest that androgen-deprivation therapy reduces the symptoms in a majority of patients. Conversely, over a period of time, the patients develop tumors

which are independent of androgen and eventually turn fatal in nature. David Ornstein and group investigated on the analysis of laser capture micro-dissected (LCM) human prostate cancer and also *in vitro* prostate cell lines. This study employed the malignant and normal tissues from prostatectomy tissue samples procured by LCM and the proteins detected by 2D-PAGE. Various proteins revealed different expressions, which also included the famous prostate biomarker PSA (prostate specific antigen). The remainder proteins were detected to be as numerous as PSA protein. This indicated the introduction of significant alterations in protein expressions associated with this cancer by 2D-PAGE analysis of LCM identified cells. Deciphering these proteins enables a probability for the research of novel biomarkers associated with prostate cancer, which could be used as therapeutic target points or as diagnostic probes [79]. Later research suggested proteomics model in serum was used as an indicator for the identification of prostate cancer.

An innovative pattern was assembled by merging MS spectra and bioinformatics for the detection of prostate cancer. This proteomic pattern envisaged 36 of 38 patients with prostate cancer while177 out of 228 patients were accurately distinguished and characterized of possessing benign states. The specificity of prostate cancer was recorded to be 71% in men with slightly elevated PSA levels. If the validation is approved, this serum proteomic model could be applied to decide if there is need to perform a biopsy with men having elevated PSA levels [80]. Studies by Hood et al. studied the patterns in proteomics in paraffin-embedded tissue of prostate cancer. Group spectral investigation of benign prostate hyperplasia (BPH) and prostate cancer guided towards the identification of several more biomarkers like macrophage inhibitory cytokine-1 and prostatic acid phosphatase [37].

As per recent findings in proteomics analysis, many proteins are known to become up and or down regulated in prostate cancer. OXCT1, BDH1, ACAT1 and HMGCL are the proteins and their expressions are observed to have increased *in vitro* investigations [81]. Another research by Bigot, et al., concluded the role of diethylstilbestrol (DES), its action in inhibition of prostate cancer along with the detection of alterations in proteins. Analysis by 2D-DIGE revealed DES-induced expression modifications for 14 proteins [82].

7.8.7. Ovarian Cancer

Ovarian cancer leads in the rate of mortality, almost 4.2%, caused from gynecological cancers accounting for cancer mortality and hence its early detection is of prime importance for survival of patients. Owing to lack of early stage symptoms, poor prognosis is observed in ovarian cancer. For early diagnosis of the cancer highly sensitive and specific (99.6% of positive value) biomarkers is highly essential as the biomarker CA-125 proves to be of inadequate help. With the aid of proteomics, several biomarkers are been observed to help in the diagnosis but they lack validation. Most of the biomarkers for the investigation in ovarian cancer can be found in urine of patients. Proteomic investigations to diagnose ovarian cancer can be done in two ways:

- Proteomic pattern diagnostics/serum proteomic profiling by MS

This technique is based on the compound MS differences detected between proteomic patterns of the provided samples which could be with or without any cancer which has also been deciphered by bioinformatics. The discovery of novel biomarkers using MS is based on

- Low-molecular-weight of serum proteomes which contain an array of biomarker information to be explored.
- A pattern followed by various biomarkers which may enclose biased information as opposed to just a single biomarker over an outsized heterogeneous group of patients.

This technique helps in obtaining novel biomarkers and is also an extremely sensitive diagnostic approach for early stages in cancer diagnosis. On the other hand this technique has been objected for quality control, reproducibility and standard operating protocols while collecting samples, handling and sampling.

- Integrative/ alternative proteomic approach.

It identifies novel, single biomarkers and following the successive progress in new assays. It follows the regular pathway where samples are prepared, separated based on glycosylation, detected biomarkers are tested with ELISA and later validated [83].

Amidst the several identified biomarkers, inter-alpha-trypsin inhibitor heavy chain H4, imitate protease profiles of certain cancers [83]. Transferrin,

another biomarker, is known to be linked with systemic inflammations along with several other non-neoplastic conditions [84]. They are not procured directly from ovarian cancer and are also not specific markers for the cancer. Their implication and specificity for ovarian cancer needs to be studied extensively. Ovarian cancer is also heterogeneous and highly complex in nature.

Hence, to detect the different stages of the disease using a single biomarker would not yield high sensitivity and specificity. Studies performed merging biomarkers is observed to develop the sensitivity of CA-125 in ovarian cancer [85]. Achieving high sensitivity through biomarkers in combination is detected to be linked with the decline in specificity [86]. Studies also decipher the fact that remarkable sensitivity and specificity of the biomarkers can be achieved by studying small number of samples (a small number of cases of stage I disease) exclusive of individual validation [83]. (Table 7.3).

Table 7.3. Biomarkers for the Diagnosis of Ovarian Cancer Discovered Through Proteomic Technology [83]

Samples and Techniques Applied	**Identified Biomarker**	**Regulated in Cancer**	
		Up regulated	**Down regulated**
SELDI-TOF MS (serum)	Haptoglobin-derived alpha subunit	Yes	No
nanoLC/ESI-TOF, FT-ICR MS (serum)	Fibrinopeptide-A	Yes	No
2DE, MALDI-TOF MS (serum)	Haptoglobin-1 precursor	Yes	No
SELDI-TOF MS (serum)	apolipoprotein A1	No	Yes
	truncated form of trans-thyretin	No	Yes
	cleavage fragment of inter-α-trypsin inhibitor heavy chain H4	Yes	No
2DE, MALDI-TOF MS (serum)	Haptoglobin	Yes	No
	Transferring	No	Yes
SELDI-TOF MS (plasma)	amyloid A1	Yes	No

SELDI-TOF MS (serum)	Transthyretin	No	Yes
	Beta-hemoglobin	Yes	No
	Apolipoprotein AI	No	Yes
	Transferring	No	Yes
2DE, SELDI-TOF MS (urine)	Glycosylated eosinophil-derived neurotoxin (EDN)	Yes	No
	COOH-terminal osteopontin fragments	Yes	No
2DE & MALDI-TOF MS (tissue)	NM23-H1	Yes	No
	annexin-1	Yes	No
	protein phosphatase-1	Yes	No
	proteasome alpha-6	Yes	No
	NAGK (N-acetyl glucosamine kinase)	Yes	No
	ferritin light chain	Yes	No
MALDI-TOF MS (tissue)	PA28 (Reg-alpha fragment)	Yes	No
2D-DIGE, lectin blot, MALDI-TOF MS (serum)	Afamin	No	Yes
SELDI-TOF MS (serum)	amyloid A1	Yes	No
1DE & 2DE; LC-MS/MS (serum)	Catabolic fragments of complement factors	Yes	No
	von Willebrand factor	Yes	No
	PEBP1 (RKIP)	Yes	No
	EMILIN2	Yes	No
SELDI-TOF MS (urine)	collagen alpha 1 (III) fragment	Yes	No
	fibrinogen beta NT fragment	Yes	No
	fibrinogen alpha fragment	Yes	No
2DE & MALDI-TOF MS (tissue)	hsp27	Yes	No
	hsp60	Yes	No
	mitochondrial short-chain enoyl-CoA hydratase	Yes	No
	Prohibitin	Yes	No
	prx-III	Yes	No
	prx-II	No	Yes

FT-ICR MS: Fourier-transform ion cyclotron resonance mass spectrometry; LC-MS/MS: liquid chromatography-mass spectrometry; MALDI-TOF MS:

matrix-assisted laser desorption ionisation time-of-flight mass spectrometry; SELDI-TOF MS: surface-enhanced laser desorption/ionization time-of-flight mass spectrometry; nanoLC/ESI-TOFMS: nano-liquid chromatography/ electrospray ionization time-of-flight mass spectrometry; 1DE: one dimensional gel electrophoresis; 2DE: two dimensional gel electrophoresis; 2D-DIGE: two-dimensional differential in-gel electrophoresis.

The choices of diagnosis are very limited in women with high risk of ovarian cancer from family history or in those showing the presence of BRCA1 and BRCA2 mutations in genes. Since there is incapability to identify the early onset of ovarian cancer, careful monitoring is the lone method to be relied upon. Also, the examination of CA 125 measurement along with the trans-vaginal ultrasound scans in women with increased risk of the cancer was found to be futile in the early detection of tumors [18].

A combination of tests may be essential to confirm the specificity and sensitivity to detect ovarian cancer for its early detection. The capability to utilize urine samples for the detection of ovarian cancer may be of beneficial as compared to the invasive approaches to gain samples like blood [18].

7.9. CHALLENGES

Though cancer biomarker research is progressing extensively using proteomic technology, the critical evaluation of results suggested evident inadequacy and uncertainty with regards to reproducibility of results in identification of proteins and also their validation. Validating the novel biomarkers is the most challenging phase in clinical proteomics. Critical concern also includes the reproducibility and specificity of the biomarker along with the steps of collection and handling the samples, experimental design, precise controls and data analysis. In addition, the results need to be tested if it is a cancer-specific phenomenon or some general metabolic disorder or an inflammatory response [87]. MS obtained data is validated by the use of antibodies present in ELISA or Western blot. However, these techniques may require large mass of specific antigen to be identified. Lack of antibodies which are specific to small peptide fragments which don't cross react with the longer/parental peptide should also be considered [88]. MS proteomics uses two approaches in cancer: protein identification and pattern recognition. Both the approaches require bioinformatics and high-computing systems for the evaluation of the massive data produced by proteomic techniques. The identified biomarkers should be reproducible in different sections of populations and by various laboratories [19].

Cancer being heterogeneous, considerable amount of time and research is needed to detect proteome alterations whilst translating them into applications in the clinics [14]. Testing smaller sample size is achievable with the techniques at hand, however to run investigations with hundreds or large sample sizes to obtain data is not well-suited. Preparing a large sample at a given point of time for a lengthy experiment may not deliver reliable results, indicating a substitution between reliable data and sample size. Classifications of known compounds that are based on data search are not established to yield accurate matches. Thus, classification of unknown bio-molecules with no accurate match in the database becomes a meticulous task specially while performing untargeted investigations which can have several possibilities. Thus, identification of the biomarkers is yet a exigent task in the study of cancer proteomics [45].

7.10. PERSPECTIVES FOR PROTEOMICS IN CANCER

Even though enormous progress is noticed in the MS techniques along with the advancement in developing standardized experimental procedures for separation, enrichment and amplification of proteins, research in proteomics is restricted by bioinformatics and other available technology tools for the identification and study of proteins. The heterogeneity in diseases are the major hurdles to conquer in proteomics due to the compound nature of the human proteome, the excess of protein iso-forms and the remarkable range of protein concentration in the specimens [89]. Taking into account the recent metamorphosis phenomenon in the proteomic stream by counting joint and inter-disciplinary efforts, some potential perspectives are anticipated for the near future. This includes rewiring a few previous steps, enhancing new protocols and merging proteomics with "omics" targeting more biological strategies [19].

7.11. SUMMARY

Proteome analysis has materialized as an influential tool for investigating highly intricate samples from tumor patients. Proteomic analysis has helped in the detection of novel biomarkers that has proved to be an important finding in the early detection of cancers helping in the prevention of its onset in suspected individuals. They are also helpful in monitoring the progression of the disease in patients with cancer along with therapeutic efficacy in the

patients post tumor removal surgery. Biomarkers further help in analysis of resistance system and treatment-related toxicity in patients. This makes biomarkers like personalized and specific mode of treatment for patients. Scientists' investigation for biomarkers in breast cancer has considered MS technique as the best for the analysis of metabolic biomarkers in urine. MS is a dominant procedure that offers a wide range of investigation of proteins. Procuring urine samples is an effortless and non-invasive procedure. This makes it the model technique to obtain samples from patients for clinical management to detect and examine biomarkers. These have a prospective application in the detection of BC. The BC proteins detected offer additional insight into the compound signaling pathway connections occurring through the progression of BC. LC-MS/MS investigating technique performed on procured urine samples from both; healthy and cancer patients offered a comparative study indicating a panel of notably distorted urinary proteins that are copious in invasive and pre-invasive BC which were not discovered earlier in either urine or any biological specimen. Proteins from urine indicate the presence of BC and can be supportive of direct and pathology tests for the final conclusion of occurrence of the disease. These urine biomarkers need to be calculated in BC tissues to examine for expansive use in cancer examination and prevention.

BC cell lines are models for pre-clinical research and they identify the various types and subtypes of breast tumors. Proteins obtained from urine samples are potential and indicates the occurrence of BC during preliminary screening. They can also be grouped with pathology testing and direct physical investigations for the confirmed end diagnosis. The biomarkers identified by Beretov, et al., can be employed for the early diagnosis and also monitoring the progress of BC from urine samples (Beretov, et al., 2015).

The application of proteomics has achieved momentum owing to its accuracy, speed, sensitivity, and throughput, along with the improvement of availability of influential analytical tools and software. However, the reproducibility of proteome analysis possesses concern and it hampers the attestation of biomarker identification due to lack of large samples from a large group of patients. Furthermore, several candidate biomarkers revealed may turn out to be indirectly related to tumor biology. Inadequacy in serum recognition thresholds also proves an obstacle. Conversely, more and more sophisticated and sensitive protein quantification procedures capable of detecting femtomolar protein concentrations are developed. Progress in proteomics will contribute to a better perception of drug mechanism in action to reaction in diverse therapies [90]. This will further help to guide

designs for next gen-cancer remedies and also in employing personalized medical progress.

Proteomics allows scientists to investigate an array of biospecimens to study their protein contents and evaluate their concentrations. More research needs to be performed in MS, nanotechnology, protein microarrays and bioinformatics to enhance the proteomic technology in cancer research [19].

Each approach in investigating cancer has its own merits and demerits mostly those in regards with specificity and sensitivity. Proteomics is not a full-fledged 100% reliable technique with drawbacks like lack of recognition of the low abundance proteins like regulatory, signal transduction and receptor proteins. Also, membrane proteins which comprise of nearly 40% of the combined cellular proteins are stubborn to be separated by proteomic methods [7]. Several drawbacks can be resolved by other techniques like chromatography and multidimensional electrophoresis. Apart from the drawbacks, proteomics would still be the priority technique for investigation of molecules associated with diseases such as cancer.

Through the years, thousands of prospective novel cancer biomarkers are known to be reported in research literature; however, the FDA has approved and given clearance to only a few cancer biomarkers. The only proteomic biomarker to have been cleared by the FDA for in vitro diagnostic multivariate index assay is OVA1 [2].

Proteome analysis has been a potent tool for investigating extremely complex samples obtained from cancer patients. Proteomics has played an essential function to identify several biomarkers linked with the early prognosis of the cancer to its progression, metastasis, efficacy in therapy, body's resistance against treatment and also in toxicity due to treatments using drugs and chemotherapy. Proteomics is gaining momentum owing to the analytical tools and availability of software along with the throughput, speed, sensitivity and accuracy in the study of investigating melanoma biomarkers. In spite of all the above factors, LDH remains to be the only biomarker used clinically for melanoma. Large patient samples are essential for further investigation of the cancer to check for reproducibility and validation. Further research in protein quantification techniques to detect femtomolar concentrations of protein is been done. These advancements in proteomics will provide an enhanced perception of drug mechanism of stimuli to its response through various therapies. These studies will further direct the next generation cancer therapeutics, helping in providing patients with personalized medicines [73].

REFERENCES

1. WHO International Programme on Chemical Safety Biomarkers and Risk Assessment: Concepts and Principles. (1993). Retrieved from http://www.inchem.org/documents/ehc/ehc/ehc155.htm.
2. Li, D., & Chan, D. W. (2014). Proteomic cancer biomarkers from discovery to approval: it's worth the effort. *Expert Review in Proteomics*. *11*(2), 135–136
3. Luna Coronell, J. A., Syed, P., Sergelen, K., Gyurján, I., & Weinhäusel, A. (2012). The current status of cancer biomarker research using tumor-associated antigens for minimal invasive and early cancer diagnostics. *Journal of Proteomics. 76,* 102–115
4. Baak, J. P., Janssen, E. A., Soreide, K., & Heikkilae, R. (2005). Genomics and proteomics--the way forward. *Annals of Oncology*. *16*(2), ii30–44
5. Tan, H. T., & Lee, Y. H., Chung MC (2012) Cancer proteomics. *Mass Spectrometry Reviews*. *31*(5), 583–605.
6. Mayeux, R. (2004). Biomarkers: Potential uses and limitations. *NeuroRx*. *1*(2), 182–188.
7. Azodi, MZ, Tavirani, M. R., Mortazavian, A., Vafee, R., Tavirani, M. R., Zali, H., & Kashani, M. S. (2013). Application of Proteomics in Cancer Study. *American Journal of Cancer Science*, 2116–113.
8. Kyle Strimbu, Jorge, A., & Tavel, M. D. (2010). What are Biomarkers? *Current Opinion on HIV and AIDS*. *5*(6), 463–466.
9. Xu, H., Nie, Z., Zhang, Y., Li, C., Yue, L., & Yang, W., et al. (2014). Four sulfur mustard exposure cases: Overall analysis of four types of biomarkers in clinical samples provides positive implication for early diagnosis and treatment monitoring. *Toxicology Reports*. *1,* 533–543
10. Richard, S. (2011). Genomic biomarkers in predictive medicine. An interim analysis. *EMBO Molecular Medicine. 3,* 1–7.
11. Montgomery, J. E., & Brown, J. R. (2013). Metabolic biomarkers for predicting cardiovascular disease. *Vascular Health Risk Management*. *9,* 37–45.
12. Lee, Y. H., Kim, J. H., Zhou, H., Kim, B. W., & Wong, D. T. (2012). Salivary transcriptomic biomarkers for detection of ovarian cancer: for serous papillary adenocarcinoma. *Journal of Molecular Medicine*. *90*(4), 427–434.

13. Lauro, S., Trasatti, L., Bordin, F., LAnzetta, G., Bria, E., Gelibter, A., Reale, M. G., & Vecchione, A. (1999). Comparison of CEA, MCA, CA 15–3 and CA 27–29 in follow-up and monitoring therapeutic response in breast cancer patients. *Anticancer Research. 19*(4C), 3511–3515.
14. Wulfkuhle, J. D., Paweletz, C. P., Steeg, P. S., Petricoin EF 3rd, & Liotta, L. (2003). Proteomic approaches to the diagnosis, treatment, and monitoring of cancer. *Advances in Experimental Medicine and Biology*. *532,* 59–68.
15. Zhang, B., Barekati, Z., Kohler, C., et al. (2010). Proteomics and Biomarkers for Ovarian Cancer Diagnosis. *Annals of clinical and laboratory science*. *40*(3), 218–225.
16. Plebani, M. (2005). Proteomics: the next revolution in laboratory medicine? *Clinicia Chimica Acta*. *357,* 113–122.
17. Omenn, G. S. (2006). Strategies for plasma proteomic profiling of cancers. *Proteomics*. *6,* 5662–5673.
18. Chambers, A. F., & Vanderhyden, B. C. (2006). Ovarian cancer biomarkers in urine. *Clinical Cancer Research*. *12,* 323–327.
19. Khadir, A., & Tiss, A. (2012). Proteomics Approaches towards Early Detection and Diagnosis of Cancer. *Carcinogene Mutagene*. *14,* 002.
20. Banks, R. E., Dunn, M. J., Hochstrasser, D. F., Sanchez, J. C., & Blackstock, W. (2000). Proteomics: new perspectives, new biomedical opportunities. *Lancet*. *356,* 1749–1756
21. Pertea, M., & Salzberg, S. L. (2010). Between a chicken and a grape: estimating the number of human genes. *Genome Biology*. *11,* 206.
22. Anderson, N. L., & Anderson, N. G. (2002). The human plasma proteome: history, character, and diagnostic prospects. *Molecular Cell Proteomics*. *1,* 845–867.
23. Li, G., Xiao, Z., Liu, J., Li, C., Li, F., et al. (2011). Cancer: a proteomic disease. *Sci China Life Sciences*. *54,* 403–408.
24. Lakhani, S. R., Ellis, I. O., Schnitt, S. J., Tan, P. H., & van de Vijver, M. J. (2012). WHO Classification of Tumors of the Breast Fourth Edition. Lyon*: IARC WHO Classification of Tumors*, 2012.
25. Beretov, J., Wasinger, V. C., Millar EKA, Schwartz, P., Graham, P. H., & Li, Y. (2015). Proteomic Analysis of Urine to Identify Breast Cancer Biomarker Candidates Using a Label-Free LC-MS/MS Approach. *PLoS ONE. 10*(11), 1–20.

26. Yates, J. R., Ruse, C. I., & Nakorchevsky, A. (2009). Proteomics by mass spectrometry: approaches, advances, and applications. *Annual Review of Biomedical Engineering*. *11,* 49–79.
27. Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V., & Mann, M. (2006). The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biology*. *7,* R80.
28. Linden, M., Lind, S. B., Mayrhofer, C., Segersten, U., & Wester, K. (2012). Proteomic analysis of urinary biomarker candidates for nonmuscle invasive bladder cancer. *Proteomics*. *2,* 135–144.
29. Antalis, C. J., Uchida, A., Buhman, K. K., & Siddiqui, R. A. (2011) Migration of MDA-MB-231 breast cancer cells depends on the availability of exogenous lipids and cholesterol esterification. *Clinical and Experimental Metastasis*. *28,* 733–741.
30. dos Santos, C. R., Domingues, G., Matias, I., Matos, J., & Fonseca, I. (2014). LDL-cholesterol signaling induces breast cancer proliferation and invasion. *Lipids in Health and Disease*. *13,* 16.
31. Coffelt, S. B., de Visser, K. E. (2015). Immune-mediated mechanisms influencing the efficacy of anticancer therapies. *Trends in Immunology*. *36,* 198–216.
32. Sarandakou, A., Protonotariou, E., & Rizos, D. (2007). Tumor markers in biological fluids associated with pregnancy. *Critical* Reviews in *Clinical Laboratory Sciences*. *44,* 151–178.
33. Bagan, P., Berna, P., Assouad, J., Hupertan, V., & Le Pimpec Barthes, F. (2007). Value of cancer antigen 125 for diagnosis of pleural endometriosis in females with recurrent pneumothorax. *European Respiratory Journal*. *31,* 140–142.
34. Sun, Y., Mi, W., Cai, J., Ying, W., & Liu, F. (2008). Quantitative proteomic signature of liver cancer cells: tissue transglutaminase 2 could be a novel protein candidate of human hepatocellular carcinoma. *Journal of Proteome Research*. *7,* 3847–3859.
35. Zhang, B., Barekati, Z., Kohler, C., et al. (2010). Proteomics and Biomarkers for Ovarian Cancer Diagnosis. *Annals of Clinical and Laboratory Science*. *40*(3), 218–225.
36. Jackson, D., Craven, R. A., Hutson, R. C., Graze, I., Lueth, P., Tonge, R. P., Hartley, J. L., Nickson, J. A., Rayner, S. J., Johnston, C., Dieplinger, B., Hubalek, M., Wilkinson, N., Perren, T. J., Kehoe, S., Hall, G. D., Daxenbichler, G., Dieplinger, H., Selby, P. J., & Banks, R. E. (2007)

Proteomic profiling identifies afamin as a potential biomarker for ovarian cancer. *Clinical Cancer Research*. *13,* 7370–7379.

37. Hood, B. L., Malehorn, D. E., Conrads, T. P., & Bigbee, W. L. (2009) Serum proteomics using mass spectrometry. *Methods in Molecular Biology*. *520,* 107–128.
38. Engwegen, J. Y., Gast, M. C., Schellens, J. H., & Beijnen, J. H. (2006) Clinical proteomics: searching for better tumor markers with SELDI-TOF mass spectrometry. *Trends in Pharmacological Sciences*. *27,* 251–259.
39. Ardekani, A. M., Akhondi, M. M., & Sadeghi, M. R. (2008) Application of genomic and proteomic technologies to early detection of cancer. *Archives of Iranian Medicine*. 11(4): 427–34.
40. Baggerly, K. A., Morris, J. S., Edmonson, S. R., & Coombes, K. R. (2005) Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *Journal of National Cancer Institute*. *97,* 307–309.
41. Albrethsen, J. (2007). Reproducibility in protein profiling by MALDI-TOF mass spectrometry. *Clinical Chemistry*. *53,* 852–858.
42. Sheehan, K. M., Calvert, V. S., Kay, E. W., Lu, Y., Fishman, D., Espina, V., Aquino, J., Speer, R., Araujo, R., Mills, G. B., Liotta, L. A., Petricoin, E. F., 3rd, & Wulfkuhle, J. D. (2005). Use of reverse phase protein microarrays and reference standard development for molecular network analysis of metastatic ovarian carcinoma. *Molecular and Cellular Proteomics*. *4,* 346–355.
43. Hellström, I., Raycraft, J., Hayden-Ledbetter, M., et al. (2003). The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma. *Cancer Research*. *63*(13), 3695–3700.
44. Ray, S., Reddy, P. J., Jain, R., et al. (2011). Proteomic technologies for the identification of disease biomarkers in serum: advances and challenges ahead. *Proteomics*. *11,* 2139–61.
45. Karimi, P., Shahrokni, A., & Ranjbar, M. R. (2014). Implementation of proteomics for cancer research: past, present and future. *Asian Pacific Journal of Cancer Prevention*. *15*(6), 2433–2438.
46. Li, J., Zhang, Z., Rosenzweig, J., Wang, Y. Y., & Chan, D. W. (2002). Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clinical Chemistry*. *48,* 1296–1304.

47. Nezami Ranjbar, M. R., Zhao, Y., Tadesse, M. G., Wang, Y., & Ressom, H. W. (2013). Gaussian process regression model for mormalization of LC-MS data using scan-level information. *Proteome Science*. *11,* 13.
48. Gortzak-Uzan, L., Ignatchenko, A., Evangelou, A. l., et al. (2007). A proteome resource of ovarian cancer ascites: integrated proteomic and bioinformatic analyzes to identify putative biomarkers. *Journal of Proteome Research. 7,* 339–51.
49. Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., et al. ()2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular and Cellular Proteomics*. *4*(10), 1487–1502.
50. Ishihara, T., Fukuda, I., Morita, A., et al. (2011). Development of quantitative plasma N-glycoproteomic using label-free 2-D LC-MALDI MS and its applicability for biomarker discovery in hepatocellular carcinoma. *Journal of Proteomics*. *74*(10), 2159–2168.
51. Mirza, S. P., & Olivier, M. M. (2008) Methods and approaches for the comprehensive characterization and quantification of cellular proteomes using mass spectrometry. *Physiological Genomics*. *33*(1), 3–11.
52. Sheng, P., Teresa, A., Brentnall, & Ru, C. (2016) Glycoproteins and glycoproteomics in pancreatic cancer. *World Journal of Gastroenterology*. *22*(42), 9288–9299.
53. Bast RC Jr, Badgwell, D., Lu, Z., Marquez, R., Rosen, D., Liu, J., Baggerly, K. A., Atkinson, E. N., Skates, S., Zhang, Z., Lokshin, A., Menon, U., Jacobs, I., & Lu, K. (2005) New tumor markers: CA125 and beyond. *International Journal of Gynecological Cancer*. *15*(3), 274–281.
54. Sayed, S. (2010). Daoud. *Cancer proteomics from bench to bed.* Humana Press*, Dec 9, 2010.*
55. Mishra, N. C. (2010). *Introduction to proteomics*. Canada: Wiley; 2010. ISBN: 978-0-471-75402-2
56. Wetmore, B. A., & Merrick, B. A. (2004). Invited review: Toxicoproteomics: Proteomics applied to toxicology and pathology. *Toxicologic pathology*. *32,* 619–642.
57. Meyerson, M., & Carbone, D. (2005). Genomic and proteomic profiling of lung cancers: Lung cancer classification in the age of targeted therapy. *Journal of clinical oncology*. *23,* 3219–3226.

58. Rajapakse, J. C., Duan, K. B., & Yeo, W. K. (2005). Proteomic cancer classification with mass spectrometry data. *American Journal of Pharmaco Genomics*. *5,* 281–292.

59. Daoud, S. S. (2008). *Cancer Proteomics from Bench to Bed* 59, 2008. ISBN 978-1-59745-169-7.

60. Van den Bruel, A., Cleemput, I., Aertgeerts, B., Ramaekers, D., & Buntinx, F. (2007). The evaluation of diagnostic tests: Evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. *Journal of Clinical Epidemiology*. *60,* 1116–1122.

61. Anna K Füzéry, Joshua Levin, Maria M Chan, & Daniel W Chan. (2013). Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges. *Clinical Proteomics*. *10,* 13.

62. Bergman, A. C., Benjamin, T., Alaiya, A., Waltham, M., Sakaguchi, K., Franzén, B., Linder, S., Bergman, T., Auer, G., & Appella, E. (2000). Identification of gel-separated tumor marker proteins by mass spectrometry. *Electrophoresis*. *21,* 679–686.

63. Lu, Q. Y., Yang, Y., Jin, Y. S., Zhang, Z. F., Heber, D., Li, F. P., Dubinett, S. M., Sondej, M. A., Loo, J. A., & Rao, J. Y. (2009). Effects of green tea extract on lung cancer a549 cells: Proteomic identification of proteins associated with cell migration. *Proteomics*. *9,* 757–767.

64. Somiari, R. I., Somiari, S., Russell, S., & Shriver, C. D. (2005). Proteomics of breast carcinoma. *Journal of Chromatography B. 815,* 215–225.

65. Celis, J. E., Moreira, J. M. A., Cabezón, T., Gromov, P., Friis, E., Rank, F., & Gromova, I. (2005). Identification of extracellular and intracellular signaling components of the mammary adipose tissue and its interstitial fluid in high risk breast cancer patients toward dissecting the molecular circuitry of epithelial-adipocyte stromal cell interactions. *Molecular & Cellular Proteomics*. *4,* 492–522.

66. Pan, J., Chen, H. Q., Sun, Y. H., Zhang, J. H., & Luo, X. Y. (2008). Comparative proteomic analysis of non-small-cell lung cancer and normal controls using serum label-free quantitative shotgun technology. *Lung*. *186,* 255–261.

67. Reymond, M. A., Steinert, R., Kähne, T., Sagynaliev, E., Allal, A. S., & Lippert, H. (2004). Expression and functional proteomics studies in colorectal cancer. *Pathology-Research and Practice*. *200,* 119–127.

68. Safaei, A., Sobhi, S., Rezaei-Tavirani, M., & Zali, M. R. (2013). Genomic and epigenetic instability in colorectal cancer. *Iranian Journal of Cancer Prevention*. *6,* 54–63.
69. Albrethsen, J., Bøgebo, R., Gammeltoft, S., Olsen, J., Winther, B., & Raskov, H. (2005). Upregulated expression of human neutrophil peptides 1, 2 and 3 (hnp 1–3) in colon cancer serum and tumors: A biomarker study. *BMC Cancer*. *5,* 8.
70. Bertin, S., Samson, M., Pons, C., Guigonis, J. M., Gavelli, A., Baqué, P., Brossette, N., Pagnotta, S., Ricci, J. E., & Pierrefite-Carle, V. (2008). Comparative proteomics study reveals that bacterial cpg motifs induce tumor cell autophagy in vitro and in vivo. *Molecular & Cellular Proteomics*. *7,* 2311–2322.
71. Droin, N., Hendra, J. B., Ducoroy, P., & Solary, E. (2009) Human defensins as cancer biomarkers and antitumor molecules. *Journal of Proteomics*. *72,* 918–927.
72. Van Houdt, W. J., Emmink, B. L., Pham, T. V., Piersma, S. R., Verheem, A., Vries, R., Fratantoni, S. A., Pronk, A., Clevers, H., & Rinkes, I. H. M. B. (2011). Comparative proteomics of colon cancer stem cells and differentiated tumor cells identifies birc6 as a potential therapeutic target. *Molecular & Cellular Proteomics*. *10*(12), M111.011353.
73. Mittal, P., & Jain, M. (2016). Proteomics: An Indispensable Tool for Novel Biomarker Identification in Melanoma. *Journal of Data Mining Genomics Proteomics*. *7,* 204.
74. Kwek, S. S., Kahn, J., Greaney, S. K., Lewis, J., Cha, E., et al. (2016). GM-CSF and ipilimumab therapy in metastatic melanoma: Clinical outcomes and immunologic responses. *Oncoimmunology*. *5,* e1101204.
75. Kasparian, N. A., McLoone, J. K., & Meiser, B. (2009). Skin cancer-related prevention and screening behaviors: A review of the literature. *Journal of Behavioral Medicine. 32,* 406–428.
76. Sarto, C., Valsecchi, C., Magni, F., Tremolada, L., Arizzi, C., Cordani, N., Casellato, S., Doro, G., Favini, P., Perego, R. A., Raimondo, F., Ferrero, S., Mocarelli, P., & Galli-Kienle, M. (2004). Expression of heat shock protein 27 in human renal cell carcinoma. *Proteomics*. *4,* 2252–2260.
77. Masui, O., White, N. M., DeSouza, L. V., Krakovska, O., Matta, A., Metias, S., Khalil, B., Romaschin, A. D., Honey, R. J., & Stewart, R. (2013). Quantitative proteomic analysis in metastatic renal cell

carcinoma reveals a unique set of proteins with potential prognostic significance. *Molecular & Cellular Proteomics*. *12,* 132–144.

78. Wang, X., Yu, J., Sreekumar, A., Varambally, S., Shen, R., Giacherio, D., Mehra, R., Montie, J. E., Pienta, K. J., Sanda, M. G., Kantoff, P. W., Rubin, M. A., Wei, J. T., Ghosh, D., & Chinnaiyan, A. M. (2005). Autoantibody signatures in prostate cancer. *New England Journal of Medicine*. *353,* 1224–1235.
79. Ornstein, D. K., Gillespie, J. W., Paweletz, C. P., Duray, P. H., Herring, J., Vocke, C. D., Topalian, S. L., Bostwick, D. G., Linehan, W. M., & Petricoin, E. F. (2000). Proteomic analysis of laser capture microdissected human prostate cancer and in vitro prostate cell lines. *Electrophoresis*. *21,* 2235–2242.
80. Petricoin, E. F., Ornstein, D. K., Paweletz, C. P., Ardekani, A., Hackett, P. S., Hitt, B. A., Velassco, A., Trucco, C., Wiegand, L., & Wood, K. (2002). Serum proteomic patterns for detection of prostate cancer. *Journal of the National Cancer Institute*. *94,* 1576–1578.
81. Saraon, P., Cretu, D., Musrap, N., Karagiannis, G. S., Batruch, I., Drabovich, A. P., van der Kwast, T., Mizokami, A., Morrissey, C., & Jarvi, K. (2013). Quantitative proteomics reveals that enzymes of the ketogenic pathway are associated with prostate cancer progression. *Molecular & Cellular Proteomics*. *12,* 1589–1601.
82. Bigot, P., Mouzat, K., Lebdai, S., Bahut, M., Benhabiles, N., Tassin, G. C., Azzouzi A. R., & Cussenot, O. (2013). Quantitative proteomic determination of diethylstilbestrol action on prostate cancer. *Asian Journal of Andrology*. *15*(3), 413–420.
83. Zhang, Z., Bast RC Jr, Yu, Y., Li, J., Sokoll, L. J., Rai, A. J., Rosenzweig, J. M., Cameron, B., Wang, Y. Y., Meng, X. Y., Berchuck, A., Van Haaften-Day, C., Hacker, N. F., de Bruijn, H. W., van der Zee, A. G., Jacobs, I. J., Fung, E. T., & Chan, D. W. (2004). Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Research*. *64,* 5882–5890.
84. Ahmed, N., Oliva, K. T., Barker, G., Hoffmann, P., Reeve, S., Smith, I. A., Quinn, M. A., & Rice, G. E. (2005). Proteomic tracking of serum protein isoforms as screening biomarkers of ovarian cancer. *Proteomics*. *5,* 4625–4636.
85. Yurkovetsky, Z. R., Linkov, F. Y., D, E. M., & Lokshin, A. E. (2006). Multiple biomarker panels for early detection of ovarian cancer. *Future Oncology*. *2,* 733–741.

86. Bast RC Jr, Badgwell, D., Lu, Z., Marquez, R., Rosen, D., Liu, J., Baggerly, K. A., Atkinson, E. N., Skates, S., Zhang, Z., Lokshin, A., Menon, U., Jacobs, I., & Lu, K. (2005). New tumor markers: CA125 and beyond. *International Journal of Gynecological Cancer*. *15*(3), 274–281.
87. Diamandis, E. P. (2004). Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Molecular Cell Proteomics*. *3,* 367–378.
88. Lee, C. J., Ariztia, E. V., & Fishman, D. A. (2007). Conventional and proteomic technologies for the detection of early stage malignancies: markers for ovarian cancer. *Critical Reviews in Clinical Laboratory Science*. *44,* 87–114.
89. Schiess, R., Wollscheid, B., & Aebersold, R. (2009). Targeted proteomic strategy for clinical biomarker discovery. *Molecular Oncology*. *3,* 33–44.
90. Saldanha, G., Elshaw, S., Sachs, P., Alharbi, H., & Shah, P., et al. (2009). microRNA-10b is a prognostic biomarker for melanoma. *Modern Pathology*. *29,* 112–121.

CHAPTER 8

FUTURE DIRECTIONS, CHALLENGES AND SOLUTIONS

CONTENTS

8.1. Introduction..286
8.2. Challenges In Biomarkers For Proteomics 287
8.3. Proteomics vs. Genomics...293
8.4. Overcoming Challenges In Data Interpretation 293
8.5. Future Expectations...294
8.6. Points For Consideration 295
References...297

8.1. INTRODUCTION

The term proteomics refers to the study and the evaluation of the entire complement of proteins in a cell or tissue or an organism [1]. While there are 26,000–31,000 protein encoding genes in the human genome [2], the figure corresponding to of human protein products, including splice variants and posttranslational modifications (PTMs) stands at a million [3].

Proteomics can have several applications that include:

- The profile of proteins.
- Comparison of analysis of samples.
- Analyze protein–protein interactions.
- Study of posttranslational modifications.

The study of biological samples is challenging as

- There are complex samples such as body fluids or serum or tissue sections.
- One cell is capable of dynamism in terms of concentration of proteins as well as the quantity (proteins can be present between copy numbers of one to a lakh). (Reviewed by Chandramouli and Qian [4]).

Though several proteins have been identified as biomarkers, as they have turned up in several research works, figures say that these molecules represent only a fraction. The other challenges include analysis, evaluation and processing of the proteomics data is challenging. Another challenge is analysis of LC/MS and LC-MS/MS data that involves several steps.

These challenges can be overcome by the following of optimal practices during preparation of sample:

- The complex samples can be processed into simple ones that enable the enrichment of samples that have low concentration.
- The techniques of mass spectrometry instruments that are state of the art technology.
- The application of a lot of processing of data and analysis.

Several techniques have been discussed in the chapters such as

- Gel-based applications: one-dimensional and two-dimensional poly acryl amide gel electrophoresis
- Gel-free high throughput screening technologies such as multidimensional protein identification technology, isotope-

coded affinity tag ICAT; SILAC; isobaric tagging for relative and absolute quantitation (iTRAQ), Shotgun proteomics and 2DE DIGE as well as protein microarrays.

- High throughout analysis techniques western blot assays, multiple reaction monitoring assay and label-free quantification of high mass resolution LC-MS data.

Advancements in proteomics methods have been reported using either bottom-up display and bottom-up identification. These techniques offer advantages such as:

- Faster analysis
- Increased sensitivity
- More analysis of proteomes

These techniques are expected to supersede traditional techniques such as 2D gels that use top-down approach.

The use of both bottom-up display and bottom-up identification has permitted detection of several markers (proteins) in cells as well as their organelles.

8.2. CHALLENGES IN BIOMARKERS FOR PROTEOMICS

8.2.1. Proteome of the Membrane

Integral membrane proteins are 20–30% of the genes in an organism have important roles in a cell. The main challenge of these proteins is their tendency for aggregation as well as precipitation. 2D electrophoresis is unsuitable for the integral membrane proteins. The reason is this technique depends on cleavage of target residues such as lysine and arginine that are not present in trans -membrane helices.

The analysis of membrane proteins has been attempted by techniques such as:

The solubilization of membranes with acids or detergents or organic solvents. These reagents are suitable for the later steps of protein digestion for analysis by Mass spectrometry. For instance, SDS and subsequent isotope-coded affinity tag (ICAT) labeling [2] or another approach involved application of formic acid with cyanogen bromide [1]. These techniques facilitated the identification of membrane proteins.

The use of high pH and proteinase K has been researched for the study of such membrane proteins while the former facilitates membranes to form sheets the latter digests the hydrophilic portions of the proteins of membrane [5].

Membrane proteins can be solubilized by detergents such as decaethylene glycol mono hexadecyl [6].

The use of other techniques for membrane proteins followed by mass spectrometry has been suggested such as:

Nanoflow chromatographic techniques.

Multidimensional liquid chromatography.

Membrane proteins subjected to native electrophoresis [4].

The quantity of the molecules that interact can be estimated by the use of methods that involve Subcellular fractionation. The use of lipid rafts or portions that resist the application of detergents serve as techniques as an alternative to the use of cells or other intact sources [7].

The use of chemical tagging along with such methods of physical separation can aid in the separation of desired proteins of interest to enhance the specificity of the process [8]. The use of chemical tags on the extracellular portion of plasma proteins can prevent mixing of other materials from within the cell used. An example of such a tag is the use of biotin tags that bind to the extracellular portion of cell membranes [4].

8.2.2. Use of Biomarkers of Serum

The term complex has been used to describe the status of serum. With several thousands of proteins occupying serum [9], it can be a source of biomarkers as it interacts with several parts of the body on account of its property of being circulated. The use of serum for biomarkers especially of disease is being pitched in as promising.

The term “most complex human proteome” (9) has been applied to serum. The term is applied as each protein varies in levels especially a few that at low concentrations. Additionally, the proteins show variation among members of a population apart from different modifications that are seen among individuals [4].

The analysis of proteomics requires the removal of albumin that is present at elevated levels in the serum in a range of 35 to 50 mg/ml [4]. This can facilitate the identification of other candidates present at lower levels. The proteins that are present at higher levels can serve as a hindrance to the

identification of other biomarkers that are present at lower concentrations. The absence of albumin from the picture can remove other lipoproteins or hormones or cytokines that interact with the protein. Other proteins that present a challenge are immunoglobulins or antibodies.

The removal of high concentration proteins of serum such as albumin can rely on the techniques such as:

- isoelectric trapping.
- immunoaffinity columns.
- peptide affinity chromatography.
- dye-ligand chromatography.

The immunoglobulins can be removed by affinity chromatography technique that uses the affinity of the protein to protein G or A.

The use of chromatography using protein G or Heparin can serve as pretreatment approach for serum that is both effective as well as affordable.

The use of columns of on mRP-C18 using immune-depletion and reversed-phase separation of plasma can facilitate steps of MS.

The use of antibodies can offer help such as polyclonal antibody that bind the proteins that are present at high levels in body fluids such as serum. The advantage is the use of columns to package such antibodies. An example is high-specificity polyclonal antibodies (MARS). These bind the 6 proteins that are present at high levels and that too as one step.

Another affinity system is Human-14 multiple affinity removal column that removes or lowers the levels of the 14 most common proteins that are found at high concentrations.

The use of SELDI-TOF MS has been reported for the analysis of serum for biomarkers. The technique uses forms of chromatography such as cation exchange or anion exchange and hydrophobicity or hydrophilic nature in the form of arrays. The technique uses small quantities of serum samples (5–10 μl) that can be loaded on an array. Following a wash to remove substances that do not bind, the use of time-of-flight mass spectrometry can reveal the biomarkers. The system offers advantages such as:

- Rapidity of analysis.
- It can be operated as high throughput.
- The less quantity of samples.

The technique has been reported for several samples such as serum.

The use of techniques for fractionation can facilitate the reduction of challenges associated with the "complex" proteome tag of serum.

The techniques include;

- liquid chromatography.
- SELDI.
- electrophoresis.

For instance, several proteins were identified as biomarkers such as apolipoprotein-A1, glutathione peroxidase-3, transthyretin in lung tumor bearing mice studies [11].

In another study, the use of stable isotope labeling along with tandem mass spectrometry in serum of pancreatic cancer samples facilitated the identification of 1065 proteins. Of these more than 10% were found to be at higher levels in the cancer samples [12].

8.2.3. Sample Collection

The use of techniques to collect clinical samples is essential in such studies. Such samples have to be maintained for quality as well as careful handling to avoid issues such as bias that can complicate the results.

The factors that influence the samples include:

The processing temperature and time

The compound used in the sample collection tubes

Changes in temperature such as freeze-thaw

Conditions of storage

Hemolysis

It is vital that the tests and controls are to be handled at identical conditions as well as treatments right from the initial steps to the final analysis.

This requires the employment of standard operating procedures by a representative working group to detect any discrepancy that can be communicated to quality control. The introduction of variability in the processing and handling of samples can affect the reproducibility and efficiency of analysis. This warrants the formulation of standard operating procedures for different samples for analysis of biomarkers [4].

8.2.4. Analysis of MS in Proteomics

The use of MS in proteomics facilitates the analysis of systems at a molecular level. The use of proteomics however is lesser than compared to genomics. Several reasons and arguments have been presented in this regard.

MS made an appearance in the previous millennium. This was followed by a "The one hour yeast proteome," where researchers presented the potential of high-performance MS along with chromatographic technique facilitated the analysis of the proteome of yeast. [13].

To summarize, the technique has the attributes of:

- Robustness
- High throughput
- Reproducibility (reviewed by Sidoli et al. [14])

Overview:

The mass-to-charge ratio of a sample that has been ionized in gas phase is calculated by a mass spectrometer that is used for determining mass of the molecule. The use of MS in proteomics has resulted in special expertise for different aspects of the filed such as interactions between proteins as one specialization and structural studies as another branch. This division has not been seen in genomics due to the use of similar instruments however different arms of proteomics need different setup.

A mass spectrometer can detect samples that ionize, and offer the following benefits:

High resolution (>400,000 mass)

High speed (12–20 Hz)

High sensitivity (<attomol)

High mass accuracy (<1 ppm) [14]

There are several roles of MS in proteomics such as:

- Establishing identity of proteins as well as their posttranslational modifications
- Structure studies
- The combination of labeling and MS can generate data of quantities/turnovers (reviewed by Sidoli et al, 2016; 14)
- Associations of proteins and other substances [15].

MS: the bright side in proteomics

The sensitivity of MS permits analysis of: [16]

- Protein interactions
- Posttranslational modifications
- Stoichiometry
- Levels/quantities of proteins

8.2.5. Challenges in MS

While genomics that involves nucleotide sequences can be subjected to amplification, such as facility is challenging in proteins. DNA can be subject to amplification, analysis can be done up to minute levels.

The tools of bioinformatics that are used in bioinformatics are available with ease. The tools used in proteomics are proprietary that is another major challenge.

A challenge is the "reality" of the sample. As mass spectrometers have a sensitivity of less than attomoles, the threshold of noise vs. signal is a challenge to define. The quantitative nature of MS associates the concentration of a sample with its signal. With other molecules possessing similar masses the development of noise can complicate results.

Samples used can be complex that possesses analytes of varying concentrations. In genomics, it is possible to analyze till the molecular level that is a challenge in proteomics. The complexity of sample can make the interpretation of data a challenge.

The increased sensitivity can report false positives that call for discretion.

Table 8.1. Proteomics vs. Genomics

Proteomics	**Genomics**
The amplification of proteins is under research.	Nucleotide s equences can be amplified.
Bioinformatics tools are proprietary	Bioinformatics tools are free and available.
Sample complexity can be challenging	Analysis to genes is possible
Lower reproducibility as chromato-grams are difficult to interpret.	DNA libraries reads offer more confidence in interpretation.

The change in the number of protein molecules can affect the results of MS as it is a quantitative technique. The numbers of proteins detected in a specialized cell like muscle are lower than that detected in stem cells. The difference arises due to the increased numbers of specific proteins in muscles that suppress other proteins as compared to undifferentiated cell.

8.3. PROTEOMICS VS. GENOMICS

However, it was recorded based on number of published research articles that the studies that used genomics were more than that of proteomics [14] as shown in the following Figure 8.1.

Figure 8.1. Proteomics vs. Genomics [14].

8.4. OVERCOMING CHALLENGES IN DATA INTERPRETATION

- Several tools of bioinformatics have been developed for MS. The identification of proteomics data is facilitated by several tools. (Figure 8.2)

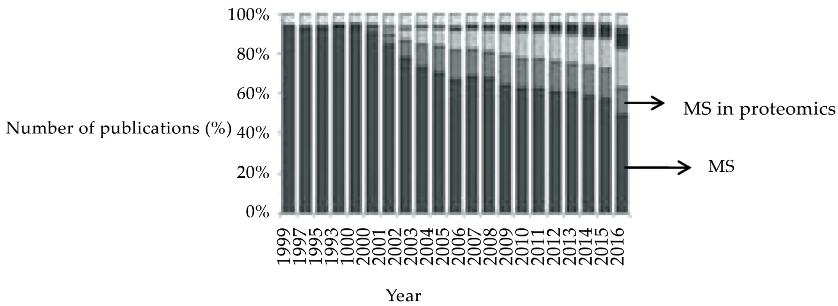


Figure 8.2. Increase in the tools applied in MS proteomics

- The combination of gas and liquid chromatography along with MS can aid the analysis especially in the reduction of false positives.
- The accuracy of quantification of proteins has increased due to computational advances.
- There are data-independent acquisition methods that facilitate the signal selection as the chromatograms of both precursor and fragment ions can be developed by such methods.

This has facilitated the change from semi-quantitative to quantitative tag of proteomics.

8.5. FUTURE EXPECTATIONS

The extraction of certain information is challenging in the realm of MS despite the advancements in technology and analysis. Proteins that have similar sequences but different folding yield one signal on MS that causes a limit on the applications of the technique in folding.

Tandem MS or MS/MS can aid in the identification of samples as the fragmentation of a molecule indicates its identity. Samples with similar mass can be distinguished despite similar mass.

This challenge has been addressed by the use of ion mobility. This technique employs the use of a gas that creates friction within a tube where the molecules are headed. A protein that is unfolded has a larger area of cross section that causes its increased retention as compared to a folded

protein (Kanu et al., 2008). Thus, irrespective of similar mass, samples can be analyzed and resolved.

The field of *imaging* is increasing in the realm of MS. The analysis of samples involves the development of a pixelated image; each pixel representing a spectrum. This aids in the application of extraction of the spectrum of specific ions.

Such developments have seen much hype in the scientific community. The 2016 American Society for Mass Spectrometry annual conference saw one discussion on electrospray, matrix-assisted laser desorption/ionization, and mass analyzers, three presentations on ion mobility and 4 on imaging.

It is expected by certain researchers that proteomics will soon be taught in courses across universities. One MS is expected to be used in a department in the coming decade as the potential of MS still requires analysis and more study. To quote, "it is safe to assume that the best has yet to come" [14].

8.6. POINTS FOR CONSIDERATION

The future can see the importance of techniques that avoid gels such as 18O stable isotope labeling or iTRAQ as well as MudPIT.

Another tool is the use of prefractionation of samples especially for complex samples such as serum.

The use of iTRAQ has been reported to facilitate the analysis of lysine rich peptides through MALDI ionization.

The use of Protein microarrays is pitched as high throughput. They facilitate the analysis of several markers across several samples. The technique can serve in a diagnostic or prognostic scenario. It can facilitate the analysis of large amounts of proteins for their properties or reactions. It can offer data on the functions of unknown candidates as well as new functions to established candidates.

The use of SILAC labeling offers few difficulties in application in a lab scenario. Another feature is the use of MSQuant that is an open analysis tool of the results. However, the major challenge is the use of MS that is not available to all labs.

The analysis of such large quantities of data generated in proteomics requires the development of a tool that can offer:

- Retrieving of the data requested by a user in an automatic fashion.
- The use of tools that permit analysis of function and 3D modeling.
- A graphical interface that permits analysis of data obtained from experiments
- Customized database
- An updating tool that changes as with each passing day, new tools are reported.

Though the development of such a tool is challenging; the amalgamation of tools in a linear pipeline can facilitate the construction of such a tool.

The analysis of biological markers is facilitated by an application of approaches to enrich target candidates or use of fractionation. The use of more advanced technologies can enable an increased analysis of more levels of proteomes in the fields of research, diagnosis as well as biotechnology research and development.

The fast development in techniques of proteomics in terms of their technology, sensitivity as well as the extent of coverage points to a leap in the use of these techniques in the analysis of biomarkers [4].

REFERENCES

1. Wilkins, M. R., Sanchez, J. C., Gooley, A. A., et al. (1995). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and Genetic Engineering Reviews*. *13*,19–50.
2. Baltimore, D, (2001). Our genome unveiled. *Nature*. *409* (6822)814–816.
3. Wilkins, M. R., Gasteiger, E., Gooley, A. A., et al. (1999). High-throughput mass spectrometric discovery of protein post-translational modifications. *Journal of Molecular Biology*. *289*(3), 645–657.
4. Chandramouli, K., & Qian, P, Y. (2009). Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity. *Human Genomics and Proteomics*. 2009 Dec 8;2009. pii: 239204.
5. Wu, C. C., MacCoss, M. J., Howell, K. E., & Yates, J. R., III. (2003). A method for the comprehensive proteomic analysis of membrane proteins. *Nature Biotechnology*. *21*(5), 532–538.
6. Luche, S., Santoni, V., & Rabilloud, T, (2003). Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics*. *3*(3), 249–253.
7. Jeong, J. A., Lee, Y., Lee, W., et al. (2006). Proteomic analysis of the hydrophobic fraction of mesenchymal stem cells derived from human umbilical cord blood. *Molecules and Cells*. *22*(1), 36–43.
8. Zhang, H., Yan, W., & Aebersold, R, (2004). Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Current Opinions in Chemical Biology*. *8*(1), 66–75.
9. Anderson, N. L., & Anderson, N, G. (2002). The human plasma proteome: history, character, and diagnostic prospects. *Molecular and Cellular Proteomics*. [11],845–867.
10. Anderson, N. L., Polanski, M., Pieper, R., et al. (2004). The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Molecular and Cellular Proteomics*. *3*(4), 311–326.

11. Chatterji, B., & Borlak, J, (2009). A 2-DE MALDI-TOF study to identify disease regulated serum proteins in lung cancer of c-myc transgenic mice. *Proteomics*. *9*(4), 1044–1056.
12. Yu, K. H., Barry, C. G., Austin, D., et al. (2009). Stable isotope dilution multidimensional liquid chromatography-tandem mass spectrometry for pancreatic cancer serum biomarker discovery. *Journal of Proteome Research*. *8*(3), 1565–1576.
13. Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S., & Coon, J, J. (2014). The one-hour yeast proteome. *Molecular and Cellular Proteomics*. *13*, 339–347.
14. Sidoli, S., Kulej, K., & Garcia, B, A. (2016). Why proteomics is not the new genomics and the future of mass spectrometry in cell biology. *Journal of Cell Biology*. 201612010
15. He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D. C., Wilusz, J. E., Garcia, B. A., & Bonasio, R, (2016). High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Molecular Cell*. *64*, 416–430.
16. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., & Selbach, M, (2011). Global quantification of mammalian gene expression control. *Nature*. *473*, 337–342.
17. Kanu, A. B., Dwivedi, P., Tam, M., Matz, L., & Hill Jr, H, H. (2008). Ion mobility-mass spectrometry. *Journal of Mass Spectrometry*. *43*, 1–22.

INDEX

A

abscisic acid (ABA) 136
Acinetobacter baumannii 100
allergens 117, 129, 145
Alzheimer's disease (AD) 202
American Federation for Aging Research (AFAR) 209
aminopeptidases 123
Aminotransferases 124
Ampholytes 49
analytes 39, 40, 47, 60
antibiogram 80
antibiotic therapy 89, 93, 104, 105
antigenic determinants 83
aquaculture 152, 164, 165, 181
Arabidopsis 115, 137, 139

B

bacillary angiomatosis 80
bacteriocine 83
bilirubin 214, 215
biobank 36, 37
biological diseases 186
biomarkers 36, 37, 38, 39, 44, 45, 49, 52, 53, 54, 59, 61, 63, 65, 68, 69, 70, 71, 72, 73, 74, 76
biotyping 82
Brome mosaic virus (BMV) 126
bronchoalveolar lavage fluid (BALF) 163
buccal cavity 191

C

cardiovascular disease (CVD) 210
cerebrospinal fluid (CSF) 45
Chitinases 125
collision-induced dissociation (CID) 43
Colorectal cancer (CRC) 264
C-reactive protein (CRP) 216
Cucumber mosaic virus 125

Cytokines 198

D

difference gel electrophoresis (DIGE) 158
differentially expressed genes (DEPs) 138
dormancy 133, 145

E

eclampsia 199
electrophoresis 80, 84, 86, 87
electro spray ionization (ESI) 233
endometritis 155, 162
equimolar ratio 52
erythroleukemia cells 257
estrogen receptor (ER) 250
Exosomes 198

F

Farm Animal Proteomics (FAP) 152
femtogram 39
fluorophores 50, 51, 153
Food and Drug Administration (FDA) 189
forward-phase arrays (FPAs) 255
Fourier transform ion cyclotron resonance (FTICR) 42

G

gene expression 81, 82, 93, 94, 108, 110
Genomic Biomarkers 245
genomics 191
glioblastoma 52, 75, 193
glycoproteome 258

H

heat shock proteins (HSP) 168
high performance liquid chromatography (HPLC) 61
Hordeum vulgare 119, 120
human biospecimens 244
human chorionic gonadotropin hormone 187
hydrophobic 116

I

immobilized pH gradient (IPG) 46, 47
immune risk profile (IRP) 212
immuno-histochemistry (IHC) 219
Integral membrane proteins 287
intracellular pathogens 95
ionization of molecules 39
Isobaric Tagging Reagent for Absolute Quantitation (iTRAQ) 56
isoelectric focusing (IEF) 46, 64
isoforms 46
isoleucine 46
isotope-coded affinity tag (ICAT) 287

L

laboratory informatics management applications (LIMS) 37
Late embryogenesis-abundant (LEA) 121
leucine 46
lipopolysaccharide 93
liquid chromatography (LC) 169
liquid chromatography-tandem mass spectrometry (LC-MS/MS) 158

M

mass fingerprints 41
mass spectrometer 39, 63
Mass Spectrometry (MS) 39
mastitis 155, 156, 161, 171, 173, 180, 181, 183
Matrix Assisted Laser Desorption/ Ionization (MALDI) 41
metabolomics 191, 199, 200, 201, 202
microarrays 59, 60, 61, 76, 78
microbial analytes 191
Microbial typing 80

N

Nanotechnology 256
National Cholesterol Education Program (NCEP) 226
Nicotiana benthamiana 123, 126, 144
nucleophosmin 97

O

oncogenomic data 244
Ovarian cancer 268, 269, 276

P

Parkinson's disease (PD) 202
peptide mass fingerprinting (PMF) 157
Pharmacodynamic markers 221
phenotype 81, 102
Phenotypic methods 81
phosphoproteome 137
photosynthate 141
plasmids 86
polymerase chain reaction (PCR) 80
posttranslational modifications (PTMs) 286
prefractionation 295
proficiency testing programmes 189, 221
Progenesis software 250
progesterone receptor (PR) 250
prostate cancer antigen (PSA) 266
protein mass fingerprinting (PMF) 45
Protein-protein interactions (PPIs) 96
proteome 114, 117, 119, 121, 127, 129, 134, 141, 143, 144, 145, 146, 147, 148
proteomics 36, 37, 38, 39, 46, 47, 52, 54, 59, 61, 64, 65, 67, 68, 69, 70, 72, 74, 75, 76, 77, 78, 286, 287, 288, 291, 292, 293, 294, 295, 296, 298

Q

Quadrupole 42

R

Reactive oxygen species (ROS) 119
renal carcinoma cancer (RCC) 266
Restriction Endonuclease Analysis (REA) 86
restriction fragment length polymorphisms (RFLPs) 87
reverse-phase arrays (RPAs) 255
ribotyping 87
Rice yellow mottle Virus (RYMV) 123

S

Schizophrenia 201

Selected reaction monitoring (SRM) 160
serotyping 80
Staphylococcus aureus 82
Strains 83

T

tandem mass tags (TMT) 98
thrombin time (TT) 225
Time-of-flight (TOF) 42
Toxicoproteomics 261, 279
transcriptome 134
transcriptomics 191, 194
translation elongation factor 125, 131
transmissible spongiform encephalopathies (TSEs) 202
Triticum aestivum 114, 119, 145, 147
tumorigenesis 191
Typing 81, 82, 83, 84, 85, 86

U

unfractionated heparin (UFH) 226

W

Whole genome sequencing (WGS) 80