MEDWIN PUBLISHERS
Committed to Create Value for researchers

Bioinformatics & Proteomics Open Access Journal
ISSN: 2642-6129

Proteomics: Applications in Disease Diagnosis to Develop Precision Medicine

Kavita Vasdev* and Sonali Sengupta

Department of Microbiology, University of Delhi, India

***Corresponding author:** Kavita Vasdev, Associate Professor, Department of Microbiology, Gargi College, University of Delhi, Siri Fort Road, New Delhi-110049, India, Tel: +91-9811873144; Email: atvikav@gmail.com

Mini Review

Volume 4 Issue 1

Received Date: December 30, 2019

Published Date: January 21, 2020

Abstract

Proteomics is a systems biology science that studies not only the protein components of a cell but also the interactions between them. Proteomics is a powerful tool to study disease progression at an individual level and its ability to provide information for diagnosis and therapy tailored to the individual manifestation of the disease is unlimited. It encompasses wet laboratory experiments such as 2-D gel electrophoresis and mass spectrometric techniques, but also computational analysis of biological data which include software such as GO, Cytoscape etc. The present review aims at studying the applications of proteomics in the diagnosis, Prognosis of disease progression during therapeutic intervention, and identifying therapeutic targets of the diseases, specifically neurodegenerative diseases like Alzheimer's, Parkinson's, Cardiovascular disease as Myocardial Infarction and systemic disease like Cancer- in order to develop precision medicine for applied personalized therapy

Keywords: Proteomics; 2-D Gel Electrophoresis; MS-Mass Spectrometry; Protein-Protein Interaction

Abbreviations: LRRK2: Leucine-Rich Repeat Kinase 2; ABPP: Activity-Based Protein Profiling; DIA: Data Independent Acquisition; DDA: Data Dependent Acquisition; PTMs: Posttranslational Modifications.

Introduction

The term "proteomics" was first used by Marc Wilkins in 1996 to denote the "Protein complement of a genome" [1]. Proteomics is one of the most significant methodologies to comprehend the gene function although; it is much more complex compared with genomic. Most of the functional information of genes is characterized by the proteome. Proteomics is the collective study of all measured proteins in cells of a given condition. It is a systems science that requires the understanding of the protein constituents and their expressions in a cell – but also the interplay of proteins, protein complexes, signaling pathways, and network modules as a whole for achieving biochemical functions [2,3]. Proteomics is crucial for early disease diagnosis and to monitor the disease development. Furthermore, it also has a vital role in drug development since proteins function as target molecules.

Physical Techniques in Proteomics

Data for proteomic studies in proteomics are often generated from high-throughput experimental platforms, e.g., two- dimensional(2D)gel liquid chromatography , coupled tandem mass spectrometers(LC–MS/MS), immunoassays, and multiplexed protein microarrays [4,5]. Thousands of proteins can be assayed simultaneously from complex biological samples [6]. This is done to measure the relative abundance of proteins or peptides in various biological conditions. Proteomics studies have been widely used to extract functional and temporal information in biological systems [7]. Protein–protein interactions measurements include the yeast two-hybrid (Y2H) system [8]. Stable isotope labeling of proteins/peptides can be carried out invitro (iTRAQ, TMT) and *invivo* (SILAC) followed by mass spectrometric analyses [9].

Computational Analysis of Proteomics Data

One application of computational methods in proteomics is mass spectra data handling and biomarker discovery. The primary goal of biomarker discovery is to find distinguishing features in peptide/proteomic profiles in healthy versus diseased subjects. The main challenge is the number of putative biomarkers discovered. Computational tools are available to circumvent such challenges 2008 [10]. The second application of computational methods in proteomics is data analysis in interactive proteomics. This consist of two arms i.e. pathway analysis and network analysis. Pathway analysis, aims to identify activated pathways or pathway modules from functional proteomic data. Biological pathway comprise of signaling pathways, gene regulatory pathways, and metabolic pathways. Network analysis, involves data analysis that build, overlay, visualize, and infer protein interaction networks from functional proteomics data [3].

Extensive knowledge bases have been published on biological pathways and network interactions which include STRING, HAPPI, BioGRID; Reactome databases [3]. For example, STRING is a search tool for retrieval of interacting genes/proteins while HAPPI is human annotated and predicted protein interaction database.

A complex protein network usually consists of thousands of proteins, which appears extremely messy on conventional network visualization platforms. The following are the software tools to visualize complex protein interaction networks which are divided into four categories, based on biological function and network topological interaction, (analysis of protein interactions based on tools from topology).These are : tools with basic functional information and little topological features (e.g., GO category analysis), tools with rich functional information and little topological features (e.g., GSEA), tools with basic functional information and rich topological features (e.g., Cytoscape), and tools with rich functional information and rich topological features (e.g., Pathway Express) [3].

Applications of Proteomics in Disease Diagnosis, Identification of Target Molecules for Therapeutics

Proteomics is crucial for early disease diagnosis, prognosis and to monitor the disease development. Furthermore, it also has a vital role in drug development since protein function as target molecules. In this review, we analyzed the role of proteomics in some of neurodegenerative, cardiovascular and systems disease –cancer.

Neurodegenerative Disease- Alzheimer's Disease

Alzheimer's disease is an irreversible, progressive brain disorder that slowly destroys memory and thinking skills, and, eventually, the ability to carry out the simplest tasks. In most people with Alzheimer's, symptoms first appear in their mid-60s.A systems-level study of disease-associated proteome changes in human frontal cortex of sporadic Alzheimer's disease patients was carried out. This involved an integrated approach that combines mass spectrometry-based quantitative proteomics, differential expression analysis, and co-expression network analysis. Analyses of human brain tissues from Alzheimer's disease patients and age-matched controls showed organization of the cortical proteome into a network of 24 biologically meaningful modules of co-expressed proteins. Of the above, 5 modules were positively correlated to Alzheimer's disease phenotypes with hub proteins that are up-regulated in Alzheimer's disease. Six modules were negatively correlated to Alzheimer's disease phenotypes with hub proteins that are down-regulated in Alzheimer's disease.

The above study revealed the dysregulation of multiple pathways and processes in Alzheimer's disease brain, including altered proteostasis, immune response, neuroinflammation, synaptic transmission, vesicular transport, cell signaling, cellular metabolism, cytoskeleton organization, and myelin-axon interactions.

Thus analysis of the proteome in Alzheimer's disease demonstrates new insights into disease pathogenesis so that novel candidates for future diagnostic and therapeutic development can be determined [11].

Neurodegenerative Disease- Parkinson's Disease

Parkinson's disease is a brain disorder that leads to shaking, stiffness, and difficulty with walking, balance, and coordination. Interactive Proteomics in Parkinson's disease has been investigated. In Parkinson's disease the protein alpha-synuclein is toxic by its tendency to aggregate. Alpha-synuclein interacting proteins have been identified by SILAC i.e. mitochondrial HSP70, endocytosis protein clathrin. Structural differences occur in in binding of synaptosomal proteins between monomers and oligomeric i.e. potential pathogenic alpha-synuclein. Differential proteomics revealed up-regulation of microtubule affinity regulating kinase MARK1 and MARK2 in cells exposed to alpha-synuclein. Mutations in leucine-rich repeat kinase 2 (LRRK2) account for up to 30% of inherited Parkinson's disease cases. Using

SILAC followed by IP-MS analyses, IP-MS analyses, actins and myosins have been identified as candidate interactors of LRRK2. Protein-protein interaction arrays have revealed RAB29, PAK6 as interaction factors for LRRK2. Substrate specific proteomics in Parkinson's disease have revealed outer mitochondrial membrane protein as a substrate for parkin an E3 ubiquitin ligase implicated in Parkinson's disease [12].

Cardiovascular Disease

Proteomic approaches have applications in cardiovascular medicine, specifically (i) - discovery of circulating protein biomarkers of heart disease from plasma samples (ii)-Identification of disease mechanisms and potential therapeutic targets in cardiovascular tissues in both pre-clinical models and translational studies (iii) targeted proteomics [13]. Seventy nine proteins were identified and validated by aptamer capture that increased during 24 hours after planned myocardial infarction, including several non-troponin structural heart proteins which were validated by mass spectrometry [14]. For stable coronary heart disease a proteome study was used to identify patients at risk for adverse outcomes based on an aptamer assay. A total of 9 proteins were associated with adverse outcome including troponin I, matrix metalloproteinase-12 and angiopoietin-2 [15]. Waldron et al. identified the TBX5 interactome in the developing heart to discover the interaction with the repressor complex Nurds, elucidating the mechanism by which TBX5 mutations can influence cardiac development and cause congenital heart disease [16]. Chiang DY, et al. recently investigated the interactome of the protein phosphatase 1 catalytic subunit identifying 78 interacting partners in the human heart. The results indicate increasing binding to PDE5A, in paroxysmal atrial fibrillation patients to impair proteins in electrical and cardiac remodeling. This finding has implications in the understanding and treatment of atrial fibrillation [17].

Tissue fibrosis is a hallmark of most cardiovascular disease and includes modification and deposition of the proteins in the extracellular matrix. Proteomics has been used to study the extracellular matrix providing insights into its biology and remodeling. Targeted proteomics focuses on a pre-determined group of proteins for e.g. extracellular matrix proteins. Prototypic peptides unique to the extracellular matrix proteins are quantified by selected reaction monitoring or multiple reaction monitoring. The targeted approach increases selectivity, sensitivity, accuracy and enables detection of cardiovascular disease biomarkers which include extracellular biomarkers of fibrosis [18].

Systemic Disease- Cancer

In the context of cancer proteomics can be used in the following ways to (i) discover for novel cancer biomarkers [19] (ii)fast-tracking molecular diagnostics in oncology using antibody based proteomics [20] (iii) Dissect tyrosine kinase signaling pathways in cancer i.e. functional proteomics [21] (iv) carry out activity-based protein profiling (ABPP) for biochemical pathway discovery in cancer. When ABPP is integrated with other large-scale profiling methods, it can provide insight into the metabolic and signaling pathways that support cancer pathogenesis and indicate new strategies for treatment [22]. (v) Perform molecular imaging by mass spectrometry .Combining image analysis with mass spectrometry techniques provides new possibilities for molecular pathology [23]. In one study of proteomes of 45 breast tumors, 9 represented from each of the 5molecular classifications i.e. basal-like, luminal A, luminal B, Her2-like, Normal-like, were analyzed. Analysis of whole proteome data recapitulated the current mRNA-based molecular classifications. The high-quality proteome profiles were used as a base to interpret multiple layers of systems measurements collected on the same tumors, including those of mRNA expression, genome copy -number alterations, single-nucleotide polymorphisms, phosphoprotein levels, and metabolite abundances. Independent layers of analyses reveal novel immuno-histochemical biomarker candidates. This helped to reliably stratify difficult-to-classify patients for treatment options, provide a proteome-based framework to assess prognosis for those straddling treatment class assignments, link immune cell infiltration and tumor extracellular matrix composition to prognosis, and connect molecular classification to metabolic phenotype. Furthermore, the depth and quality of proteome profiling enables discovery of neoantigens arising from tumor-specific variants of known proteins and regions of the genome previously thought to be noncoding.

In the same study genes included within prognostic mRNA panels have significantly higher than average mRNA-protein correlations. Gene copy number alterations were dampened at the protein-level; highlighting the value of proteome quantification for prognostication and phenotypic classification [24]. Proteomics has also been performed to identify the prognostic and predictive value of proteins/peptides in lung cancer. Biomarkers are measurable biological indicators found in cells, tissues, blood or other biological fluids that may be used for detection, diagnosis, treatment and monitoring cancer research by advanced quantitative proteomic approaches. Several biomarkers of different types of lung carcinoma have been identified such as cytokeratin, and heat shock proteins. Other predictive markers identified

are fatty acid binding protein-Heart (H-FABP). A mass spectrometric approach (8 peak mass spectrographic signature) towards lung cancer called Veristrat found that in the NCICBR.21 trials patients with advanced non-small cell lung cancer in the placebo with Veristrat "good signature " had a far better outcome of overall survival compared to "Veristrat poor. Using the Veristrat assay patients were observed to be benefiting from treatment with erlotinib compared to erlotinib. Thus Veristrat is a prognostic biomarker. The prognostic value of Veristrat test in advanced non-small cell lung cancer has been observed in studies with combinations of sorafineb or bevacizumab in combination with erlotinib [25]. Thirty three protein candidates from paired adenocarcinoma tissues with different extents of lymph node metastasis were identified by iTRAQ labeling technology coupled with 2-D -LC MS/MS. Six potential biomarkers were highly expressed in adenocarcinoma tissue compared to adjacent normal tissue. They are ERO1L, NARS, PABPC4, RCC1, RPS25 and TARS. ERO1L and NARS are positively associated with lymph node metastasis [26]. Triple SILAC quantitative proteomics identified exosomal proteins as biomarkers for non-small cell lung cancer. They are Integrin beta-1, basigin 4, SLC3A2, LAMP2 and CEACAM6 [27]. Comparative proteomic profiling across 23 non-small cell lung cancers revealed differences in protein expression harboring oncogenic KRAS and EGFR mutations. The above study provided information for identification of candidate therapeutic targets which carry out oncogenic processes driven by mutant Kras and EGFR proteins [28].

Challenges and Future Prospects for Proteomics

In the context of cancer, the grand challenge is to decipher the cancer proteome .The large amounts of data that can currently be collected through proteomics allow the total definition of cancer sub-proteomes. This reveals the alterations in signaling and developmental pathways in cancer. Thus discovery of predictive biomarkers and the annotation of the cancer genome based on proteomic findings can be carried out. However there remains a considerable requirement for organized collaborative efforts to efficiently mine the cancer proteome [29]. Since there is a difference in proteome expression, activity or turnover in various drug doses, a future application of proteome analysis is, to provide better drug candidates in stage II and III clinical trials. This is based on the relation between drug dose and proteome profile that functions as a critical endpoint for candidate progression.

Proteomic analysis of a patient sample can reveal both unique host proteins and specifically bacterial or viral proteins resulting in a more accurate diagnosis and treatment. The success of a given therapeutic plan varies among patients being more or less effective for particular subpopulations of patients and for different stages of disease. Therefore, monitoring the efficacy of antibiotics or antivirals after administration could quickly detect resistance or ineffectiveness of a particular treatment. In addition to monitoring the processing, activation, and clearance of therapeutics could provide physicians with real time functional information about the success of the treatments administered. Proteomic analysis of serial patient samples could potentially detect activation or breakdown of a therapeutic agent allowing physicians to monitor the therapeutic dose. The next generation of proteomics involves data independent acquisition (DIA) compared to traditional data dependent acquisition (DDA) methods. In DIA every peptide in the sample in fragmented and analyzed in a single experiment compared to random subsets in DDA methods. In DIA novel algorithms have been developed providing a more complete list of proteins i.e. approximately 10,000 proteins in a single shot [30].

The future of personalized proteomics lies in precision medicine. The areas are biomarker monitoring, diagnosis and improved therapy selection. Most known biomarkers come from proteins present at high levels in the serum or tissue, e.g., albumin, CRP, OVA1, etc. The challenge for precision medicine is to develop new biomarkers making use of the exquisite sensitivity of mass spectrometry for detection of biomarkers at lower abundance. These biomarkers may include posttranslational modifications (PTMs), metabolites, or metabolic flux. In addition, identifying the appropriate molecular isoform from a complex mixture is necessary for several existing biomarkers. The sensitivity of mass spectrometry in the detection of addition of a phosphate, fucosyl group, ubiquitin, or glycosylation simulation, etc. makes mass spectrometry ideal for determining the relative abundance of these isoforms in disease association. Personalized proteomic analysis of the HLA loci, minor antigen loci, and peptide repertoire of potential donor/recipient pairs could reduce Graft versus Host disease [31]. Thus proteomics has tremendous application in the future in development of precision medicine. With more and more studies using better technology and big data analysis of proteomes better drug targets can be identified and their correlation with diseases can redefine future of precision medicine leading to personalized health care.

References

1. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, et al. (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev 13: 19-50.

2. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, et

al. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292(5518): 929-934.

3. Wu X, Hasan MA, Chen JY (2014) Pathway and network analysis in proteomics. J Theor Biol 362: 44-52.

4. Altelaar AF, Munoz J, Heck AJ (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 14(1): 35-48.

5. Kingsmore SF (2006) Multiplexed protein measurement: technologies and applications of protein and antibody arrays. Nat Rev Drug Discov 5(4): 310-320.

6. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422(6928): 198-207.

7. Blagoev B, Ong SE, Kratchmarova I, Mann M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. Nat Biotechnol 22(9): 1139-45.

8. Ito T, Tashiro K, Kuhara T (2001) [Systematic analysis of Saccharomyces cerevisiae genome: gene network and protein-protein interaction network]. Tanpakushitsu Kakusan Koso 46(16): 2407-2413.

9. Chahrour O, Cobice D, Malone J (2015) Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. J Pharm Biomed Anal 113: 2-20.

10. Cannataro M (2008) Computational proteomics: management and analysis of proteomics data. Brief Bioinform 9(2): 97-101.

11. Zhang Q, Ma C, Gearing M, Wang PG, Chin LS, et al. (2018) Integrated proteomics and network analysis identifies protein hubs and network alterations in Alzheimer's disease. Acta Neuropathol Commun 6(1): 19.

12. Li Y, Cookson MR (2019) Proteomics; applications in familial Parkinson's disease. J Neurochem 151(4): 446-458.

13. Lam MP, Ping P, Murphy E (2016) Proteomics Research in Cardiovascular Medicine and Biomarker Discovery. J Am Coll Cardiol 68(25): 2819-2830.

14. Ngo D, Sinha S, Shen D, Kuhn EW, Keyes MJ, et al. (2016) Aptamer-Based Proteomic Profiling Reveals Novel Candidate Biomarkers and Pathways in Cardiovascular Disease. Circulation 134(4): 270-285.

15. Smith JG, Gerszten RE (2017) Emerging Affinity-Based Proteomic Technologies for Large-Scale Plasma Profiling in Cardiovascular Disease. Circulation 135(17): 1651-1664.

16. Waldron L, Steimle JD, Greco TM, Gomez NC, Dorr KM et al. (2016) The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation. Dev Cell 36(3): 262-275.

17. Chiang DY, Lebesgue N, Beavers DL, Alsina KM, Damen JM (2015) Alterations in the interactome of serine/threonine protein phosphatase type-1 in atrial fibrillation patients. J Am Coll Cardiol 65(2): 163-73.

18. Barallobre-Barreiro J, Lynch M, Yin X, Mayr M (2016) Systems biology-opportunities and challenges: the application of proteomics to study the cardiovascular extracellular matrix. Cardiovasc Res 112(3): 626-636.

19. Swami (2010) M Proteomics: A discovery strategy for novel cancer biomarkers. Nat Rev Cancer 10(9): 597.

20. Brennan DJ, O'Connor DP, Rexhepaj E, Ponten F, Gallagher WM (2010) Antibody-based proteomics: fast-tracking molecular diagnostics in oncology. Nat Rev Cancer 10(9): 605-617.

21. Kolch W, Pitt A. (2010) Functional proteomics to dissect tyrosine kinase signalling pathways in cancer. Nat Rev Cancer 10(9):618-629.

22. Nomura DK, Dix MM, Cravatt BF (2010) Activity-based protein profiling for biochemical pathway discovery in cancer. Nat Rev Cancer 10(9): 630-638.

23. Schwamborn K, Caprioli RM (2010) Molecular imaging by mass spectrometry--looking beyond classical histology. Nat Rev Cancer 10(9):639-646.

24. Johansson HJ, Socciarelli F, Vacanti NM, Haugen MH, Zhu Y, et al. (2019) Breast cancer quantitative proteome and proteogenomic landscape. Nat Commun 10(1): 1600.

25. van der Wekken AJ, Hiltermann TJ, Groen HJ (2015) The value of proteomics in lung cancer. Ann Transl Med 3(3):29.

26. Hsu CH, Hsu CW, Hsueh C, Wang CL, Wu YC, et al. (2016) Identification and Characterization of Potential Biomarkers by Quantitative Tissue Proteomics of Primary Lung Adenocarcinoma. Mol Cell Proteomics 15(7): 2396-2410.

27. Clark DJ, Fondrie WE, Yang A, Mao L, (2016) Triple SILAC quantitative proteomic analysis reveals differential abundance of cell signaling proteins between normal and lung cancer-derived exosomes. J Proteomics 133: 161-169.

28. Grundner-Culemann K, Dybowski JN, Klammer M, Tebbe A, Schaab C, et al. (2016) Comparative proteome analysis across non-small cell lung cancer cell lines. J Proteomics 130: 1-10.

29. Hanash S, Taguchi A. (2010) The grand challenge to decipher the cancer proteome. Nat Rev Cancer 10(9): 652-660.

30. Meyer JG, Schilling B (2017) Clinical applications of quantitative proteomics using targeted and untargeted data-independent acquisition techniques. Expert Rev Proteomics 14(5): 419-429.

31. Ritari J, Hyvärinen K, Koskela S, Niittyvuopio R, Nihtinen A, et al. (2019) Computational Analysis of HLA-presentation of Non-synonymous Recipient Mismatches Indicates Effect on the Risk of Chronic Graft-vs.-Host Disease After Allogeneic HSCT. Front Immunol 10: 1625.