

# Life of Proteomics: Technologies, Challenges, Viewpoints and Future

Viswanath Vittaladevaram\*

Department of Biotechnology, Vydehi Institute of Biotech Sciences, Bangalore University, Bangalore, Karnataka, India

## ABSTRACT

Advances in the field of proteomics have created novel opportunities to carry out protein analysis and for biomarker discovery that serve as basis for diagnostic procedures. In recent years, Proteomics is integrated with bioinformatics tools to carry out data analysis. High-throughput proteomic analysis, comprised of several steps, represents data flow. Hence, there are several gaps in this data flow which requires careful analysis to realize the importance of data in biology. In this article, several challenges of proteomics in the field of bioinformatics, diagnostics and biomarker discovery were discussed. In addition, viewpoints that are relevant to proteomics and future prospects were highlighted.

**Keywords:** Proteomics; Bioinformatic tools; Biomarkers; High-throughput analysis

## INTRODUCTION

Proteomics is comprised of wide range of techniques that are used for identification and quantification of proteins in cells. Proteomics is used to determine three dimensional structures of proteins and their interactions. Genome expresses variety of proteins that are cell specific and it is termed as Proteome. Genomic research studies provided more information regarding activity of gene linked with disease but very limited information about proteins that undergo post translational modifications. Both genomics and proteomics enable to identify the drug targets and novel disease markers. This, in turn, helps in product designing, diagnosis, prevention and treatment of diseases [1].

Proteomics is mainly useful for identification, characterization and quantification of proteins on large scale. These changes in chemical and biological methods enable to better understand proteins and to address several questions. Proteomics is not only limited to protein identification and localization but also determines different activities, networks, interactions and functions of proteins. Multiple proteins are encoded by single gene through mechanism of translation and post-translational modifications [2]. Based on cellular environment and integrated networks single protein will perform different functions. All these possibilities form a proteome, comprised of proteins associated with genome which is complicated than genome [3].

Proteomics is used widely in different research settings like mechanism of understanding pathogenicity, vaccine production candidates, detection of markers for diagnosis and interpretation of disease specific functional protein pathways etc. Proteome characterize the functional information of different genes. Eukaryotic cells comprising of proteome is dynamic and complex.

Whereas in prokaryotes, it is responsible for mechanism of pathogens but due to different types of properties such as hydrophilicity, molecular size, hydrophobicity and quantity it is quite challenging to do in-depth analysis of proteins [4]. Application of proteomics is critical for early diagnosis of disease, development and prognosis. In addition, it also plays a major role in drug development as molecular targets.

## THE POWER OF PROTEOMICS

Proteomics is considered as the most efficient approach to unravel the complexity of gene function. The biological function of proteins and their expression levels are dependent on. mRNA levels, translation control and regulation on host cell. For this reason, proteomics is examined as pertinent dataset for characterization of biological system (Figure 1) [5].

Proteomics is the fast emerging field, as the instrumentation used for analysis is sensitive, accurate and affordable. Thus, it can be useful for precision and personalized medicine which improves the process of treatment, diagnosis and management of disease in future [6]. For protein analysis, different types of proteomics technologies are used. These are Conventional, Advanced, High-throughput and bioinformatics methods (Figure 2).

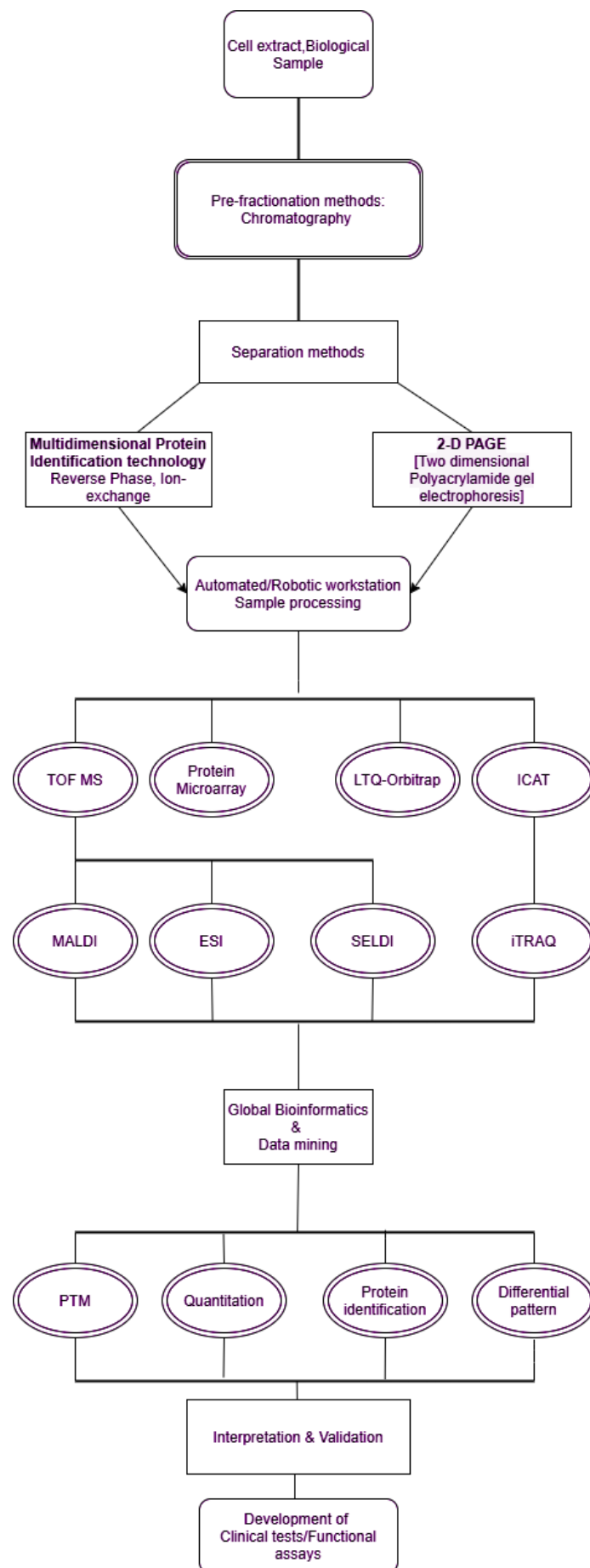
Proteomics technologies are used to generate large amount of data which is integrated and stored to search data for easy accessibility. Proteomics relies on DNA and protein sequence related databases which enable to track the variations in expression of proteins [7]. Proteome databases, comprising protein databases are species specific and can be easily accessible.

**Correspondence to:** Viswanath Vittaladevaram, Department of Biotechnology, Vydehi Institute of Biotech Sciences, Bangalore University, Bangalore, Karnataka, India, E-mail: viswa.vittal@gmail.com

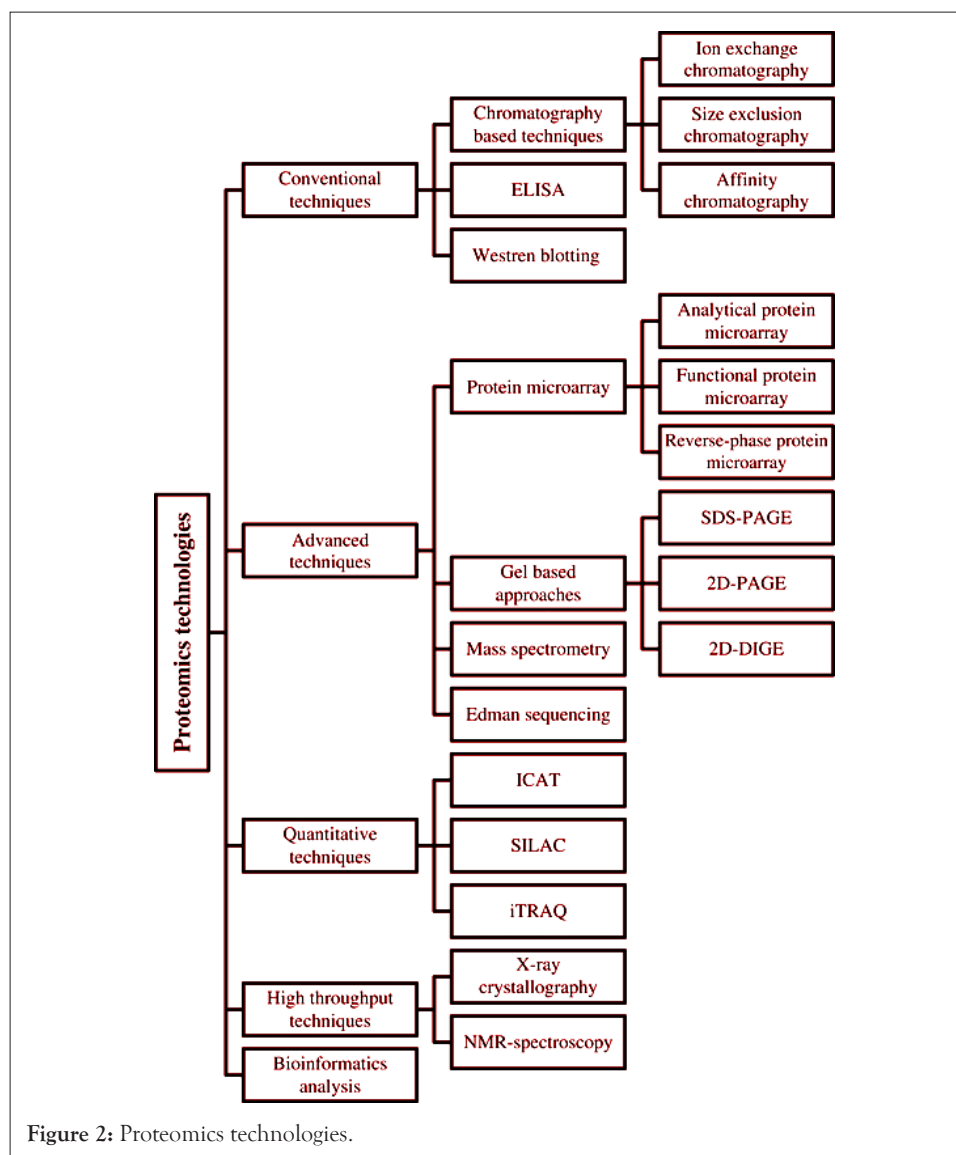
**Received:** March 29, 2021; **Accepted:** April 12, 2021; **Published:** April 19, 2021

**Citation:** Vittaladevaram V (2021) Life of Proteomics: Technologies, Challenges, Viewpoints and Future. J Proteomics Bioinform. 14:533.

**Copyright:** © 2021 Vittaladevaram V. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



**Figure 1:** Proteomics workflow. **Abbreviations:** TOF MS: Time-of-flight mass spectrometry; LTQ: Linear Ion Trap Quadrupole; ICAT: Isotope-Coded Affinity Tag; MALDI: Matrix-Assisted Laser Desorption/Ionization; ESI: Electrospray Ionization; SELDI: Surface-Enhanced Laser Desorption/Ionization; iTRAQ: Isobaric Tags For Relative And Absolute Quantification; PTM: Post-Translational Modification.



## CHALLENGES

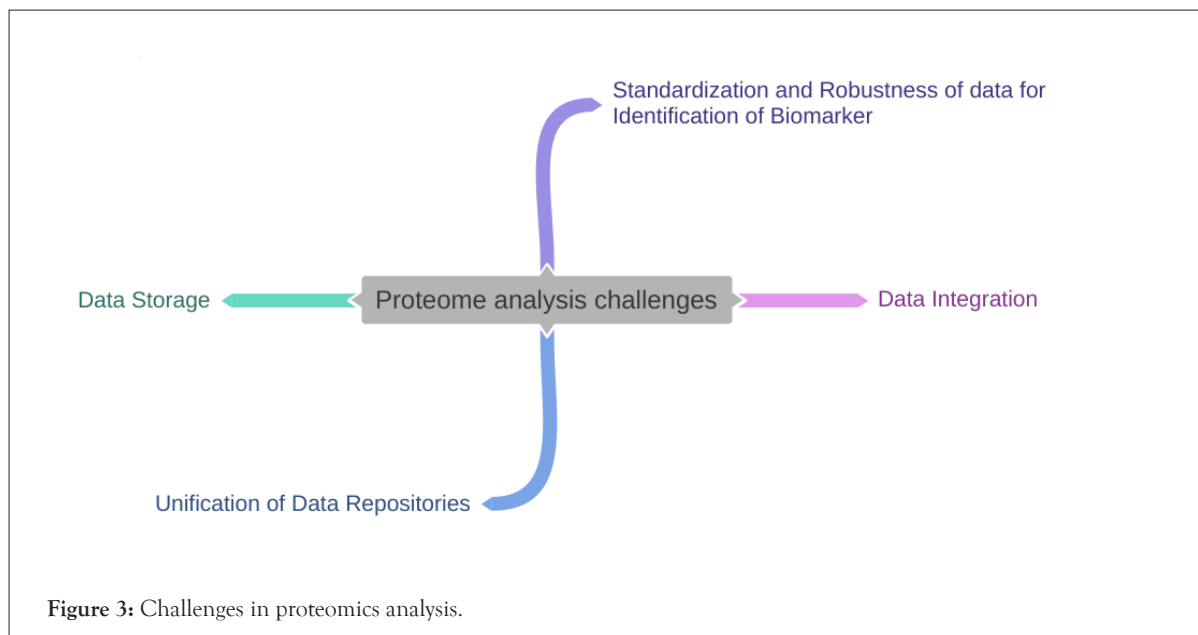
The growth of proteomics represents both challenges and solutions. Some of the challenges are poor stability of proteins making it difficult to analyse and not being inherent while studying nucleic acids it is difficult to carry out protein studies. Few species which are less abundant, it is very difficult to carry out detection as amplification of proteins is not possible similar to DNA [8]. There are specific issues with Tandem mass spectrometry and proteomics such as databases relevant to protein sequence which will either exclude isoforms or will be considered in text messages. Due to this, proteomics based analysis, carried out for identification of alternative splice isoforms related protein products, and may lead to incorrect protein and peptide identification. This is because of presence of peptides which degenerates by corresponding to more than one specific protein present in sequence database [9]. Other problems include absence of standardization when dealing with proteins that are annotated in various places. Due to lack of protein naming standards and identifiers, it will present a challenge for integration of annotations from various heterogeneous sources biologically for proteomic data interpretation [10].

Advances in profiling, regulation and analysis of protein has been observed in humans and organisms like yeast. But however application of proteomics in plant research is slow and not advanced

yet. This is due to complex nature of genome and information related with protein sequence [11]. However, it is accessible to conduct further studies using proteomics in plant research and some of the challenges include [12].

- Comprehensive protein identification, prevalence in tissue region and isoforms.
- Characterization of cellular and biochemical functions of protein.
- Protein regulation analysis and its association with other regulatory networks.

Several challenges associated with data storage, integration robustness needs to be addressed while understanding proteomics data relevant to specific disease and cell dynamics. Due to presence of confounding factors the proteomics data is not more often used in precision medicine [13]. Since there is increase in accessibility to Proteomics techniques it will lead to increase in volume of data generated and storage capacity is lacking to retain the data generated in the form of output. Though cloud storage offers solution there are issues related with sensitivity and privacy of data. Omics data for example, RNA-Seq express in raw data format doesn't require any storage. Raw data that is discarded will result in creating more space for data storage but question is will it be a reality with proteomics (Figure 3) [14].



**Figure 3:** Challenges in proteomics analysis.

A combination of proteomics data which is new and existing together with metadata and omics, is a challenging task which needs to be addressed. Upon successful data integration will enable to study the disease and other systems for which there are no standardized methods [15]. Some of the challenges associated with data integration include inconsistency with reference to datasets and annotation including pipelines related with output analysis. To overcome this standardized and consistent approach should be used for data collection and recording purposes [16].

Identification of biomarkers requires protein targets, which are reliable and appropriate is also a major challenge during extraction from high-throughput data. The challenges are due to the generation of irrelevant features and noisy which mask true indicators; and when coupled with biological samples, that are inherent, will exhibit heterogeneity [17]. For this reason, it will be a challenging task to isolate biomarkers that are specific and robust. To overcome this, different types of data should be incorporated that will promise to make specific biomarker selections across large biological datasets [18].

It is important to have central access point for sharing data. But currently, due to lack of unified data resources, the researchers are facing complications to gather data in wider context. There are numbers of data repositories exist, of which few are with restricted access, and others are available for free of cost. For this reason, there is a need to unify the data associated with proteome from public repositories. This helps to provide resources to back up the valuable data. The costs associated with usage of proteomics in clinical research laboratories are quite expensive. The other challenge is to integrate proteomics data with metabolomics and genomics for further carrying out functional analysis [19].

Lack of metadata and experimental data is another challenge for reusing biological data. This leads to annotation related problems for data relevant to proteomics. For this reason, a balance between data sharing and metadata required needs to be maintained [20]. There are possible chances for future challenges such as limited access to data relevant to clinical research data which requires controlled resources to access proteomics data.

## VIEWPOINT

The significance of Proteomics cannot be overstated as the term proteome refers to proteins, which are present within cell and depicts structure, produce energy and allow movement, reproduction and communication. Proteins are regarded as structural and functional units of cell, in which protein complement is dynamic and genetic information is static in nature [21]. Proteomics allows to discover new biomarkers for diagnosis and to carry out drug discovery for identification of molecular targets. There are wide range of applications identified using protein markers for the purpose of prognostic and clinical uses and also further use in measuring the efficacy and speed of clinical trials [22].

Differential proteomics allows drawing correlations between state of disease progression and different range of proteins produced within cell. Proteomics enables analysis of series of molecular events that are clinically relevant, thus assure promise in drug discovery research. Advanced bioinformatics in combination with proteomics technologies are used for identification of disease related molecular signatures based on protein related pathways and cascade of signalling events [23]. Proteomics based research provides abundant information complementing with genetic information provided by genomics based research. Proteomics is regarded as powerful tool for progression and further development of functional genomics; and both, together will play a major role in biomedical field which will have tremendous significance in developing therapeutic and diagnostic products in near future [24,25].

## FUTURE OF PROTEOMICS

Proteomics is widely used for isolation of proteins and their display based on criteria followed by identification of proteins. Advances in instrumentation and other analytical tools is required to get a complete human proteome map. Proteomics in combination with genomics would enable researchers to monitor expression levels of protein targets to characterize the organism behaviour at molecular level. Proteomics is useful to carry out characterization of proteins, understand post-translational modifications, protein-protein

interactions and identification of biomarkers for diseases.

## CONCLUSION

Recent advances in proteomics technologies, in combination with bioinformatics tools facilitate the discovery of pathways and data analysis relevant to proteomics. Bioinformatics tools are designed to facilitate the pathway discovery and functional annotation from proteomic based data on large scale and one such example is iProXpress. This will allow answering complex biological questions from multiple data resources. The combination of bioinformatics tool such as high-throughput genomics and proteomics with iProXpress will speed up the process of pathway prediction and other functional relationships, thereby, providing support for genome based studies for knowledge based discovery and generation of hypothesis.

## REFERENCES

- Amiri-Dashatan N, Koushki M, Abbaszadeh HA, Rostami-Nejad M, Rezaei-Tavirani M. Proteomics applications in health: Biomarker and drug discovery and food industry. *Ir J Pharm Res.* 2018;17(4):1523-1536.
- Boonen K, Hens K, Menschaert G, Baggerman G, Valkenburg D, Ertaylan G, et al. Beyond genes: Re-identifiability of proteomic data and its implications for personalized medicine. *Genes.* 2019;10(9):682.
- Cappadona S, Baker PR, Cutillas PR, Heck AJ, van Breukelen B. Current challenges in software solutions for mass spectrometry-based quantitative proteomics. *Amino Acids.* 2012;43(3):1087-1108.
- Chen C, McGarvey PB, Huang H, Wu CH. Protein bioinformatics infrastructure for the integration and analysis of multiple high-throughput "omics" data. *Adv Bioinform.* 2010;423589.
- Cho WC. Proteomics technologies and challenges. *Genomics Proteomics Bioinformatics.* 2007;2(2):77-85.
- Dupree EJ, Jayathirtha M, Yorkey H, Mihasan M, Petre BA, Darie CC, et al. A critical review of bottom-up proteomics: The good, the bad, and the future of this field. *Proteomes.* 2020;8(3):14.
- Ebhardt HA, Root A, Sander C, Aebersold R. Applications of targeted proteomics in systems biology and translational medicine. *Proteomics.* 2015; 15(18): 3193-3208.
- Plebani M. Proteomics: The next revolution in laboratory medicine? *Clinica Chimica Acta International J Clin Chem.* 2005;357(2):113-122.
- Gregorich ZR, Ge Y. Top-down proteomics in health and disease: Challenges and opportunities. *Proteomics.* 2014; 14(10):1195-1210.
- Haymond A, Davis JB, Espina V. Proteomics for cancer drug design. *Expert Rev Proteomics.* 2019;16(8):647-664.
- Hortin GL, Jortani SA, Ritchie JC, Jr Valdes R, Jr Chan DW. Proteomics: A new diagnostic frontier. *Clin Chem.* 2006;52(7):1218-1222.
- Hu ZZ, Huang H, Wu CH, Jung M, Dritschilo A, Riegel AT, et al. Omics-based molecular target and biomarker identification. *Met Mol Biol (Clifton, N.J.).* 2011;719:547-571.
- Khal Khal E, Rezaei-Tavirani M, Rostamii-Nejad M. Pharmaceutical advances and proteomics researches. *Ir J Pharm Res.* 2019;18(Suppl1):51-67.
- Qoronfleh MW. Proteomics: The next frontier-a biotech perspective. *J Biomed Biotechnol.* 2004;3(3):121-123.
- Savino R, Paduano S, Preianò M, Terracciano R. The proteomics big challenge for biomarkers and new drug-targets discovery. *Int J Mol Sci.* 2012;13(11):13926-13948.
- Shi T, Song E, Nie S, Rodland KD, Liu T, Qian WJ, et al. Advances in targeted proteomics and applications to biomedical research. *Proteomics.* 2016;16(15-16):2160-2182.
- Vaudel M, Verheggen K, Csordas A, Raeder H, Berven FS, Martens L, et al. Exploring the potential of public proteomics data. *Proteomics.* 2016;16(2):214225.
- Minde DP, Dunker AK, Lilley KS. Time, space, and disorder in the expanding proteome universe. *Proteomics.* 2017;17(7):1600399.
- Perez-Riverol Y, Alpi E, Wang R, Hermjakob H, Vizcaino JA. Making proteomics data accessible and reusable: Current state of proteomics databases and repositories. *Proteomics.* 2015;15(5-6):930-949.
- Deswal R, Gupta R, Dogra V, Singh R, Abat JK, Sarkar A, et al. Plant proteomics in India and Nepal: Current status and challenges ahead. *Physiol Mol Biol Plants: International J Fun Plant Biology.* 2013;19(4):461-477.
- Swan AL, Mobasher A, Allaway D, Liddell S, Bacardit J. Application of machine learning to proteomics data: Classification and biomarker identification in postgenomics biology. *Omics J Integ Biol.* 2013;17(12):595-610.
- Tabb DL. Quality assessment for clinical proteomics. *Clin Biochem.* 2013;46(6):411-420.
- Wilmes P, Heintz-Buschart A, Bond PL. A decade of metaproteomics: Where we stand and what the future holds. *Proteomics.* 2015;15(20):3409-3417.
- Manes NP, Nita-Lazar A. Application of targeted mass spectrometry in bottom-up proteomics for systems biology research. *J Proteom.* 2018;189:75-90.
- Hongzhan H, Shukla HD, Cathy W, Satya S. Challenges and solutions in proteomics. *Current Genomics.* 2007;8(1):21-28.