ECRONICON OPEN ACCESS

EC MICROBIOLOGY
Review Article

Emerging Trends in Microbial Genomics

Urvashi Vijay* and Sonal Gupta

Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research, Jaipur, Rajasthan, India

***Corresponding Author:** Urvashi Vijay, Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research, Jaipur, Rajasthan, India.

Received: December 22, 2018; **Published:** March 15, 2018

Abstract

The new era of 21st century marks the new age of biology generating rapid increment in molecular bioscience by revealing incredible potential to innovative and effective approaches of bioinformatics, genome sequencing, high-throughput genome-wide experimentation (~ 'omics') especially genomics, transcriptomics, proteomics and metabolomics. The omic's experiments have enabled in elucidating the molecular mechanisms in organisms. The organismal systems have major and minor components, with major being the tissues, organelles and cells and the minor being biomolecules. These components interact with each other giving a functional genomics perspective. For example, DNA-DNA, Protein- Protein, Protein-RNA interactions are found to be forming networks. Thus, systems biology and 'omics' based technologies together can provide a better understanding of complex biological systems of organisms. Through this review article we want to emphasize on emerging trends in the genomics of microbial community and further we have highlights on its application in other omics areas.

Keywords: *Genomics; Proteomics; Metabolomics; Omics; Microbial Genomics*

Introduction

There has been an exponential growth in the biological research from the time term genome has been coined by Hans Winkler, 1920. Deciphering the structure of DNA in 1970s had been another breakthrough, which expanded the girth of the discoveries. Human DNA approximately consists of 3.2 billion base pairs and with the advent of rapid sequencing techniques in the form of next generation sequencing, ca. 15 quadrillion bases were produced by the end of 2016 [1].

With the advent of high throughput technologies in molecular science, it is encouraging the researchers to study the phenomenon with a more holistic approach. Rather than few individual genes, it is being possible to study the whole genome of an organism. Cross referencing the findings of these genome based studies with other techniques is leading to the rise of proteomics, interactomics, metabolomics and now metagenomics.

Microbial Genomics and Its Implications in Medical Science

Microbial genomics as a disciple has placed a very important role in medical science. From the 16s sequencing to the longer read chemistry, there has been a tremendous need for ascertaining the role of microbes particularly bacteria, viruses in this field [2].

Our human body consists of prodigious richness in taxonomy, having hundreds and thousands of symbiotic species with their strains. Out of which 90% have not been cultivated in laboratory scale [3]. Genome microbes are generally very small in size and contains one or two chromosomes with viable number of plasmids as compared to the genome of plants and animals and are probably encoded with proteins and structural RNA's [4].

Microbial heterogeneity and adaptation are generally accompanied by loss of gene, reduction and rearrangement of genome, horizontal gene transfer and its duplication [5]. These processes have been successfully evident in studying human pathogens such as *Bordetella pertussis* – causative agent for whooping cough [6,7], *Yersinia pestis*–bubonic plague agent [8], *Mycobacterium leprae*- leprosy agent [9]. These all are generally associated with enteropathogenicity (Figure 1). Genome of *Mycobacterium leprae* provides us a dramatic form of reductive evolution. Bacterial chromosomes and plasmids are the clusters of contagious genes from 10,000 to 200,000 base pairs. These genomic islands enhance the fitness of recipient by providing new and evolutionary information about pathogenicity, drug resistance and catabolic functions. CRISPR (clustered regularly inter spaced short palindromic repeat) are the dramatic examples of genome evolution in bacteria and archae loci [10].

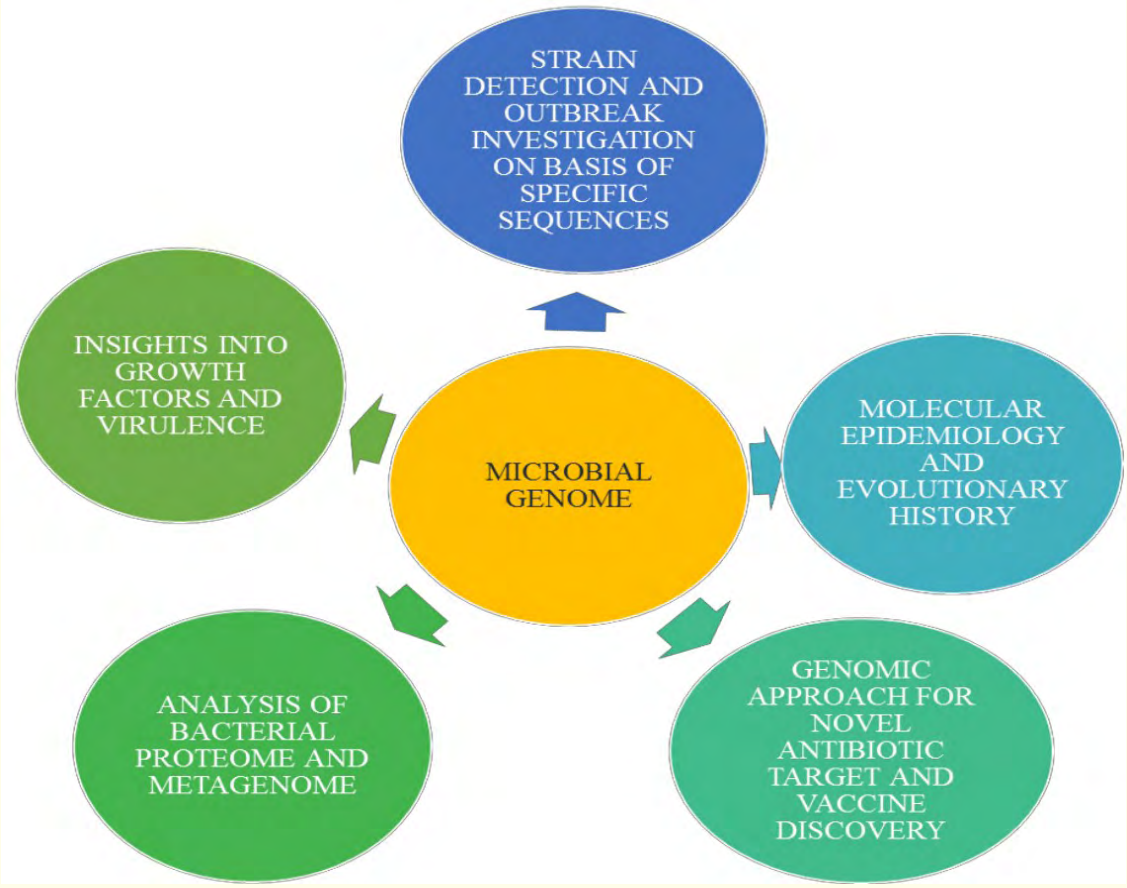

Figure 1: *Microbial genomics and its future implications [1].*

Molecular Epidemiology and its evolution – With the continuous increase in the population, large number of microbial genome sequences are also available which provides a better understanding to study the evolution of microbes, virus and human relationship in different aspects of mutations, recombination and selection of a variety. This information in resulting insights enhances the practical knowledge about the diagnostics, forensics investigations and vaccines development [11]. Yancy and his coworkers revealed about the quantification of uropathogenic bacteria through whole genome sequencing contrasting with *E. coli* bacteria which results in showing

that these bacteria's had long evolutionary history and diversity. Study also illuminated as extensive sequencing of uropathogenic bacteria will provide us the deeper knowledge of genetic signals, virulence factors and mechanism driving for epidemiology [12]. *Yersenia pestis* and *Human influenza* virus are the best examples which have been studied widely and can develop their vaccine.

Today through symbiotic activity of pathogen we are having a complete and better understanding of how microbes cause disease, host adaptation, pathogen emergence and cause epidemic in humans. Four themes of pathogen with respect to virulence are: 1. Horizontal gene transfer 2. Symbiotic and avirulent species of pathogen which generally have same virulence associated gene (causes of disease) 3. Large diversity of genes associated with the mechanisms of virulence and 4. Genome reduction and pseudogene formation [13]. From the last 5 decades, human microbiota has received a great attention in studying human health and disease. The alteration in human microbe helps us to explain not only the absence and presence of disease but also the association with other organisms. Crohns disease is a leading example of this concept as communities of pathogen and SARS (Severe Acute Respiratory Syndrome) has been studied [14]. Through shot gun sequencing various known viruses such as adenovirus, calcivirus, norovirus, rotavirus and many more has been revealed. These approaches highlight on understanding microbial diversity in terms of healthy sequencing as well as for distinguishing genuine microbial sequences with their errors [15]. Thus, characterization of pathogens through sequenced based approaches enables us to design and develop sensitive and specific diagnostic assays and also insights the cultivation of pathogens. Mutagenicity evaluation of health centers waste water has been studied by advanced mutant strains of *Salmonella typhimurium* in Jaipur health centers. The study evaluated health centers waste water causing frameshift mutations and base pair substitutions. Screening of health centers wastewater using prokaryotic assay *Salmonella* Ames test and sos-chromotest and eukaryotic assay chromosome aberration's used for quick assessment of genotoxicity studies prior to its discharge [16,17]. Complete genome sequence provides acuity to deeply understand and provides accurate information regarding metabolic defects of particular bacteria about its free growth medium, growth factors which can enhance its productivity as well as resistivity towards diseases. Genome sequencing insight the blueprints of particular bacterial and viral components and disruption in sequences leads to growth inhibition and death. It also enables the drug susceptibility and resistance profiles, synthetic capabilities of microbes which provide us the accurate clue of their antigenic repertoire. Furthermore, this information can be vulnerable for vaccine designing and immunoprophylactic interactions. Genome approaches as reverse vaccinology has given a new movement in microbial genomics which facilitates us to develop gene cloning and protein expression. Purifying these proteins can be used for studying the genome 3D structure and designing structural drugs which inhibits their activity such as new drugs has been developed for treating *Schistosomiasis mansoni* causing disease [18].

Proteomics and Metabolomics: Concepts and Perspectives

Omics" is the commonly used suffix for a variety of technically advanced methods that provides ample of data. Its technologies are mainly aimed at the detection of genes (genomics), mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in a specific biological sample. As proteomics involves identification of proteins, metabolomics involves detection and quantification of small biomolecules, especially metabolites [19].

The terms "proteome" and "metabolome" were first mentioned in the published scientific literature in 1996 and 1998, respectively [20]. The range of proteins (in proteomics) and metabolites (in metabolomics) that can be detected depends on the choice of protocol and statistical method adopted [21].

The complete set of proteins in a cell is referred to as proteome and the study of protein structure and function expressed by an organism is known as proteomics. The proteome is highly dynamic and it changes with time in response to varied environmental stimuli. It is an indication of both genes and the environment and involves biomarker discovery as proteins are ubiquitously affected in disease and disease response, which is reflected in many protein disease biomarkers already available (e.g. CA125 and alpha-fetoprotein) [22]. The final goal of proteomics is to understand the structure and function of proteins and their interaction with life processes [23].

Based on aim of the study various proteomic approaches has been discovered (Figure 2). Data obtained through the approaches mentioned above together with bioinformatics, provide significant information on biological processes [23,24].

Figure 2: *Proteomic approaches towards aim of study.*

It has been proposed that the pillars of proteomics are structural proteomics, MS-based proteomics, proteome arrays, clinical proteomics, and proteome bioinformatics [25]. The primary considerations in a proteomic experiment include sample pre-fractionation (using gel-based or chromatography techniques), sample purification, protein quantification and protein digestion to reduce the complexity of the target biological sample that can be used either at the protein level or at peptide level [22]. Typical experiments include affinity separation methods, 1 or 2-dimensional gel electrophoresis and chromatographic separation [24].

Since early 1900s, MS research has come a prolonged way from being a chemistry instrument to a key biological interpretation tool that enables identification of proteins and peptides and can be complementary to nuclear magnetic resonance (NMR) for molecular protein structure prediction. The introduction of soft ionization methods, such as MALDI and ESI are regarded as the key to success of MS in life science research and are considered as most commonly used ionization methods for MS proteomics [25].

Metabolomics

Metabolomics, an omics' science encompasses the analysis of compounds with low molecular weight, i.e. less than 2000Da in biological samples especially cells, tissues and body fluids. Among all, tissue analysis is the most powerful approach for studying localized and specific responses to stimuli. Metabolomics emerged as a supplement to genomics, transcriptomics and proteomics. The metabolome acts as a cascade between omics and the phenotype. It helps in developing novel biomarkers for diagnostic purposes to detect and predict the disease condition well in advance by proper phenotyping/genotyping and also describes the physiological status of an organism. In addition, it provides insights into biochemical mechanisms underlying diseases, their modulation by drugs and helps in identification of novel metabolic pathways towards advancement of personalised medicine approaches in future [28]. With advent of techniques like Mass spectrometry (MS), high- resolution nuclear magnetic resonance spectroscopy (NMR), ultra-performance liquid chromatography, high-performance liquid chromatography (HPLC), Fourier transform infrared (FTIR) spectroscopy, capillary electrophoresis and new chemoinformatics and bioinformatics tools for data acquisition, analysis and integration of metabolic profiling has become more effective and reliable. Furthermore, recent technological advances make it simple for quantitative and qualitative metabolite assessment within biological systems and to understand complex interactions in biological systems, with high-throughput evaluation of many metabolites

[29]. Though NMR is valued for detecting molecular structures, it has relatively limited sensitivity and measures only few metabolites i.e. less than 100 whereas; MS platforms can measure huge number of chemicals. To improve the analysis, platforms such as above can be coupled with different detectors, configurations and separation techniques to pre-process the samples [30]. All the information of metabolites characterised so far are archived in Human Metabolites Database (HMDB: http://www.hmdb.ca/). At present, human metabolome consists of more than 41,000 small-molecule metabolites and are deposited in HMDB (version 3.5). Another database for metabolites in dietary components named FooDBis also available. Cancer Cell Metabolism Gene DataBase (ccmGDB) is another comprehensive annotation database for cell metabolism genes in cancer, available at http://bioinfo.mc.vanderbilt.edu/ccmGDB and supports useful resource for cancer research areas [26]. Metabolomics analysis is broadly arranged into two generalized experimental strategies: targeted and non-targeted. Targeted analysis is based on data hypothesis-driven research, which requires a set of pre-defined metabolites. This analysis is highly sensitive and selective. The quantification and characterization of the metabolites are focused on selected known metabolites and some related etiological pathways. This analysis uses defined internal standards and applies multiple reaction monitoring (MRM) intended for the absolute quantification in which the combination of chromatography and tandem MS are indispensible. The data is scanned for specific compounds through the reference library and subsequently characterized and quantified. The advantage of this approach is that there is prior knowledge about the identity of the measured metabolites. While non-targeted analysis is a metabolic profiling that provides a hypothesis-free analysis of metabolites. The prominent advantage of this analysis is the identification of unknown metabolites that have not previously been quantified and characterized. The identification of new metabolites should be further validated in a targeted approach [31,32]. Still there is a need for the development of robust, sensitive, high-throughput, low cost analytical technologies for further advancement of metabolomics [34].

Metabolomics studies have been focused on the identification of metabolites associated with many diseases including cancer, inborn errors of metabolism and cardiovascular diseases. Metabolomic assessment can be done using cells, bio fluids or tissues. There has been an extensive rise in the application of mass spectrometry and NMR based metabolomics for early disease detection, therapy monitoring, and finally reaching the goal of personalized medicine. As a multidisciplinary science, it includes combined aspects of biology, chemistry, and mathematics and requires analytical techniques such as chromatography, spectroscopy and mass spectrometry, coupled with univariate and multivariate data analysis methods. These fingerprinting approaches are frequently combined with multivariate analysis, to attain the widest possible coverage, in terms of the type and number of compounds analysed by using several analytical methods [27].

Various computational tools are available to process and interpret the results obtained from generation of complex data sets of metabolomic analysis.

Data processing and statistical analysis

After the data is uploaded, mass spectral peaks are picked, realigned, annotated and instrumental and chemical noise, are removed; thus, providing only the biologically relevant information. Online tools such as XCMS Online, Devium Web Metabo Analyst, and many others are available for these kinds of analysis.

Metabolite identification and databases

Initial presumed metabolite identifications can be made based on the accurate mass to charge ratio (m/z) of the mass spectral ion. This is aided using comprehensive metabolite databases such as METLIN (metabolite mass spectral database), HMDB (Human metabolome database), MassBank and GMD (Golm metabolome database). During match identification by a particular m/zratio or tandem mass spectrometry fragmentation pattern is failed in silico prediction tools, helps to provide further deep insight into metabolite identification. A recent innovation, ion mobility mass spectrometry, the rotationally averaged cross collisional section (CCS), offers another level of metabolite identification, and databases containing CCS information are currently in the early stages of development. Databases already have been set up on METLIN to help in assigning molecular structure to a metabolite and correlation to phenotypes. Another database MIDAS (Metabolite identification via database searching) can be used for metabolite identification by matching measured tandem mass

spectra (MS/MS) against the predicted fragments of metabolites in a database [36]. Watson highlights about the molecular identification of metabolite's in database having major challenges chromatography than mass spectroscopy and further emphasized to use mass spectroscopy for the separation of isomeric compounds to obtain clear fragmentation pattern [37].

Biological interpretation network modelling and pathway mapping tools aid in understanding the roles that metabolites play in relation to each other and in biological aberrations. The established and comprehensive metabolic network resources are Kegg, Recon1 and Biocyc and there are several other recently developed programs that use novel methods to find pathway connectivity, as well as assisting in metabolite identification. These include mummichog and metabolite set enrichment analysis (MSEA). Furthermore, omics scale big data integration with stable isotope metabolomics reveals interlinking between metabolites and their associations with genes and proteins [33,35].

Combined study is required to analyze jointly different metabolomics or proteomics data as to improve the overall understanding of the system [17] (Figure 3).

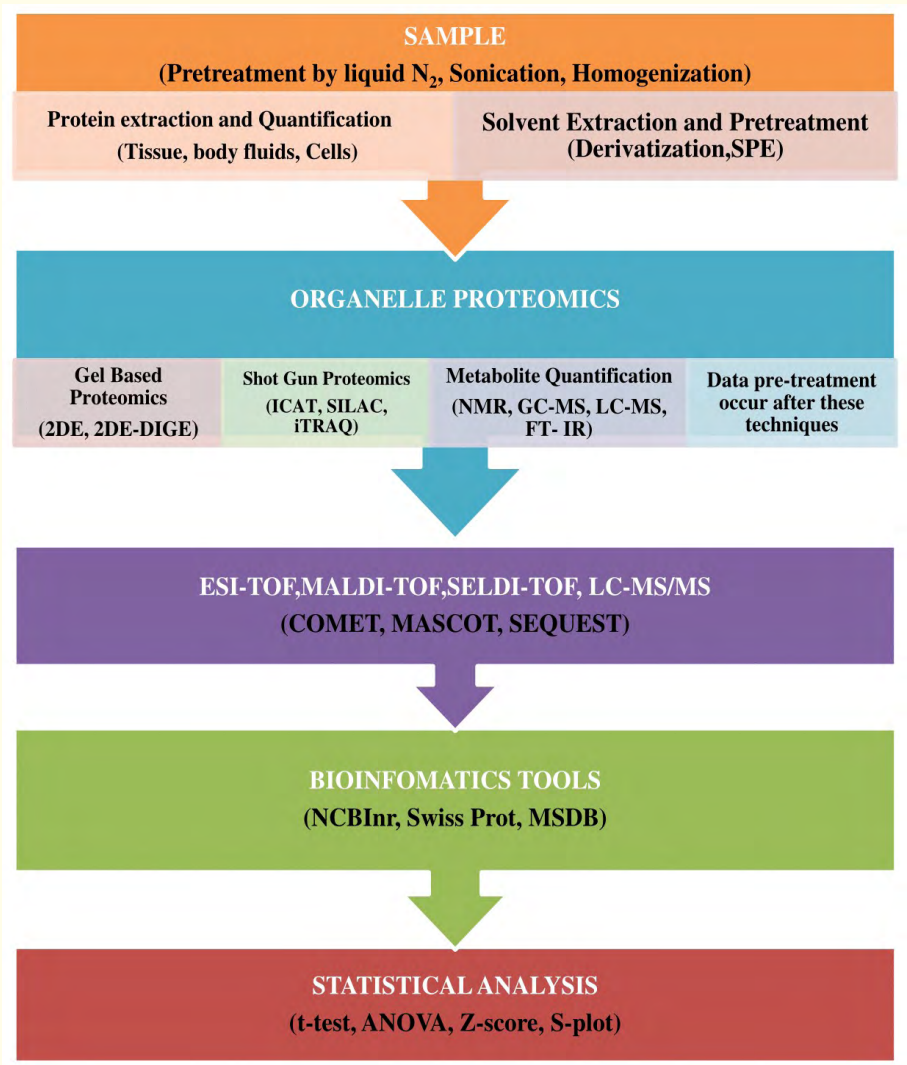

Figure 3: *The proteome and metabolome together represent an instant picture of a biological system.*

Future Perspectives

In coming years, the techniques of genome and proteomics sequencing may become routine matter in diagnostics research. With the development of genome sequencing in modern science facilitates technical knowledge to decipherate, manipulate and track different aspects of microbes. Thoughtfully to realize the benefits of genomics in medical science requires deeper understandings of how disease cause, pathogen escalate and identify its molecular determinants and develop vaccines for ameliorating.

Conclusion

The analysis of genomic and metagenomics of microbes play a vital role in several aspects of research such as cataloguing and understanding of microbial and viral diversity in human body. Genome sequencing has created a renaissance in microbial genomics and has also altered the way to study the infectious diseases.

Acknowledgement

The authors would like to gratefully acknowledge the help and enrich support of Dr. Prashanth Suravajhala in writing the manuscript.

Bibliography

1. Feero W., *et al*. "Microbial Genomics and Infectious Diseases". *New England Journal of Medicine* 365.4 (2011): 347-357.
2. Pallen M., *et al.* "High-throughput sequencing and clinical microbiology: progress, opportunities and challenges". *Current Opinion in Microbiology* 13.5 (2010): 625-631.
3. Dethlefsen L., *et al.* "An ecological and evolutionary perspective on human–microbe mutualism and disease". *Nature* 449.7164 (2007): 811-818.
4. Doolittle R. "Biodiversity: Microbial genomes multiply". *Nature* 416.6882 (2002): 697-700.
5. Fraser-Liggett C. "Insights on biology and evolution from microbial genome sequencing". *Genome Research* 15.12 (2005): 1603-1610.
6. Cummings C., *et al.* "Bordetella Species Are Distinguished by Patterns of Substantial Gene Loss and Host Adaptation". *Journal of Bacteriology* 186.5 (2004): 1484-1492.
7. Preston A., *et al.* "The bordetellae: lessons from genomics". *Nature Reviews Microbiology* 2 (2004): 379-390.
8. Parkhill J., *et al.* "Genome sequence of Yersinia pestis, the causative agent of plague". *Nature* 413.6855 (2001): 523-527.
9. Cole S., *et al.* "Massive gene decay in the leprosy bacillus". *Nature* 409.6823 (2001): 1007-1011.
10. Horvath P and Barrangou R. "CRISPR/Cas, the Immune System of Bacteria and Archaea". *Science* 327.5962 (2010): 167-170.
11. Baker S., *et al.* "Navigating the future of bacterial molecular epidemiology". *Current Opinion in Microbiology* 13.5 (2010): 640-645.
12. Yancy Lo., *et al.* "Whole-genome sequencing of uropathogenic Escherichia coli reveals long evolutionary history of diversity and virulence". *Infection, Genetics and Evolution* 34 (2015): 244-250.
13. Haas B., *et al.* "Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans". *Nature* 461.7262 (2009): 393-398.
14. Peterson D., *et al*. "Metagenomic Approaches for Defining the Pathogenesis of Inflammatory Bowel Diseases". *Cell Host and Microbe* 3.6 (2008): 417- 427.

15. Wang D., *et al.* "Viral discovery and sequence recovery using DNA microarrays". *PLoS Biology* 1 (2003): E2-E2

16. Vijay U., *et al*. "Mutagenicity Evaluation of Health Center Wastewater with the Mutant TA 100 and TA 102 Strains of Salmonella typhimurium". *Inventi Impact: Pharm Biotech and Microbio* 4 (2014): 195-198.

17. Vijay U., *et al.* "Evaluating Genotoxicity of Treated Wastewater from Health Centers with Special Reference to their Mutagenicity". *EC Microbiology* 12.2 (2017): 83-96.

18. Sette A and Rappuoli R. "Reverse vaccinology: developing vaccines in the era of genomics". *Immunity* 33.4 (2010): 530-541.

19. Blanchet L., *et al.* "Fusion of metabolomics and proteomics data for biomarkers discovery: case study on the experimental autoimmune encephalomyelitis". *BMC Bioinformatics* 12.1 (2011): 254.

20. Semba RD. "Proteomics and metabolomics: the final frontier of nutrition research". *Sight and Life* 29.1 (2015): 71-76.

21. Applying proteomics and metabolomics for studying human skeletal muscle with a focus on chronic trapezius myalgia.

22. Pocket K No. 15: 'Omics' Sciences: Genomics, Proteomics, and Metabolomics.

23. Gowda G., *et al.* "Metabolomics-based methods for early disease diagnostics". *Expert Review of Molecular Diagnostics* 8.5 (2008): 617-633.

24. Ramalingam A., *et al*. "Proteomics and Metabolomics: Two Emerging Areas for Legume Improvement". *Frontiers in Plant Science* 6 (2015).

25. Monti M., *et al.* "Functional proteomics". *Clinica Chimica Acta* 357 (2005): 140–150.

26. Cristea I. "Proteomics techniques and their application to hematology". *Blood* 103.10 (2004): 3624-3634.

27. Wright P., *et al.* "A review of current proteomics technologies with a survey on their widespread use in reproductive biology investigations". *Theriogenology* 77.4 (2012): 738-765.e52.

28. Kim P., *et al.* "ccmGDB: a database for cancer cell metabolism genes". *Nucleic Acids Research* 44.D1 (2015): D959-D968.

29. Kasture VS., *et al.* "Metabolomics: current technologies and future trends". *International Journal of Research and Development in Pharmacy &Life Sciences* 2.1 (2012-13): 206-217.

30. Ramautar R., *et al.* "Human metabolomics: Strategies to understand biology". *Current opinion in chemical biology* 17.5 (2013): 841-846.

31. Zhang A., *et al.* "Metabolomics for Biomarker Discovery: Moving to the Clinic". *BioMed Research International* (2015): 1-6.

32. Li S., *et al.* "High-Resolution Metabolomics". *Biological Research for Nursing* 18.1 (2016): 12-22.

33. Abu Bakar M., *et al.* "Metabolomics - the complementary field in systems biology: a review on obesity and type 2 diabetes". *Molecular BioSystems* 11.7 (2015): 1742-1774.

34. Skibsted S., *et al.* "Bench-to-bedside review: Future novel diagnostics for sepsis- a systems biology approach". *Critical Care* 17.5 (2013): 231.

35. Johnson C., *et al*. "Metabolomics: beyond biomarkers and towards mechanisms". *Nature Reviews Molecular Cell Biology* 17.7 (2016): 451-459.

36. Wang Y., *et al.* "MIDAS: A Database-Searching Algorithm for Metabolite Identification in Metabolomics". *Analytical Chemistry* 86.19 (2014): 9496-9503.

37. Watson D. "A Rough Guide to Metabolite Identification Using High Resolution Liquid Chromatography Mass Spectrometry in Metabolomic Profiling in Metazoans". *Computational and Structural Biotechnology Journal* 4 (2013): e201301005.

Volume 14 Issue 4 April 2018