Recent Trends in BIOCHEMISTRY

MEDDOCS
International

An Overview of Omics Approaches: Concept, Methods and Perspectives

Jeyachandran Sivakamavalli[1]*; Baskaralingam Vaseeharan[2]

[1]Fisheries Science Institute, Chonnam National University, Yeosu 59626, South Korea

[2]Department of Animal Health & Management, Alagappa University, Karaikudi-630003, India

Corresponding Author: Jeyachandran Sivakamavalli

Fisheries Science Institute, Chonnam National University, Yeosu 59626, South Korea

Email: dr.jsvalli@gmail.com

Published Online: June 15, 2020

eBook: Recent Trends in Biochemistry

Publisher: MedDocs Publishers LLC

Online edition: http://meddocsonline.org/

Copyright: © J Sivakamavalli (2020).

This chapter is distributed under the terms of Creative Commons Attribution 4.0 International License

Keywords: Proteins; Trancriptomics; Proteomics; Metabolomics

Abstract

Proteomics defines the post-genomic tests of the entire proteins (proteome) expressed by a genome in a cell, tissue or organ at a specified time under definite conditions. Protein expression is modulated at different levels from transcription to the maturation of the polypeptides produced by the translation of mature mRNAs. Metabolomics is a division of biochemistry deals with the biochemical pathways and its metabolites, these metabolites might be less than 1 KDa associated with omic technologies. This includes the wide range molecular level study which genomics, proteomics and metabolomics as a field of study is now firmly established as a functional genetics approach to understand the molecular complexity of life. Recently, metabolomics has been applied across a wide range of applications, including medical filed as diagnostic biomarker and etiological analysis in the field of medical treatment, quality control and quality forecast in the field of food products. Metabolomics represents the logical progression from large-scale analysis of RNA and proteins at the biological system. RNA and proteomics data analyses can't give us all information that occurred in the cell, while metabolomics tell us all data about what exactly happens in this cell.

Introduction

The genome of a typical bacterium comes as a single DNA molecule, which, if extended, would be about 2 mm long. (The cell itself has a diameter of about 0.001 mm). The DNA of higher organisms is organized into chromosomes-normal human cells contain 23 chromosome pairs. The total amount of genetic information per cell-the sequence of nucleotides of DNA-is very nearly constant for all members of a species, but varies widely between species not all DNA codes for proteins [1]. Conversely, some genes exist in multiple copies. Therefore the amount of protein sequence information in a cell cannot easily be estimated from the genome size. A single gene coding for a particular protein corresponds to a sequence of nucleotides along one or more regions of a molecule of DNA. The DNA sequence is collinear with the protein sequence. In species for which the genetic material is double-stranded DNA, genes may appear on either strand. Bacterial genes are continuous regions of DNA. Therefore the functional unit of genetic sequence information from a bacterium is a string of 3N nucleotides encoding a string of N amino acids, or a string of N nucleotides encoding a structural RNA molecule. Such a string, equipped with annotations, would form a typical entry in one of the genetic sequence archives.

Genomes

In eukaryotes the nucleotide sequences that encode the amino acid sequences of individual proteins are organized in a more complex manner. Frequently one gene appears split into separated segments in the genomic DNA. An exon (expressed

region) is a stretch of DNA retained in the mature messenger RNA that a ribosome translates into protein. An intron is an intervening region between two exons. Cellular machinery splices together the proper segments, in RNA transcripts, based on signal sequences flanking the exons in the sequences themselves. Many introns are very long-in some cases substantially longer than the exons. Control information organizes the expression of genes. Control mechanisms may turn genes on and off (or more finely regulate gene expression) in response to concentrations of nutrients, or to stress, or to unfold complex programs of development during the lifetime of the organism. Many control regions of DNA lie near the segments coding for proteins. They contain signal sequences that serve as binding sites for the molecules that transcribe the DNA sequence, or sequences that bind regulatory molecules that can block transcription. Bacterial genomes contain examples of contiguous genes coding for several proteins that catalyse successive steps in an integrated sequence of reactions, all under the control of the same regulatory sequence. F. Jacob, J. Monod and E. Wollman named these operons. One can readily understand the utility of a parallel control mechanism. In animals, methylation of DNA provides the signals for tissue specific expression of developmentally regulated genes. Products of certain genes cause cells to commit suicide-a process called apoptosis. Defects in the apoptotic mechanism leading to uncontrolled growth are observed in some cancers, and stimulation of these mechanisms is a general approach to cancer therapy. The conclusion is that to reduce genetic data to individual coding sequences is to disguise the very complex nature of the interrelationships among them, and to ignore the historical and integrative aspects of the genome. Robbins has expressed the situation unimprovably: Consider the 3.2 gigabytes of a human genome as equivalent to 3.2 gigabytes of files on the mass-storage device of some computer system of unknown design. Obtaining the sequence is equivalent to obtaining an image of the contents of that mass-storage device. Understanding the sequence is equivalent to reverse engineering that unknown computer system (both the hardware and the 3.2 gigabytes of software) all the way back to a full set of design and maintenance specifications. 'Reverse engineering the sequence is complicated by the fact that the resulting image of the mass-storage device will not be a file-by-file copy, but rather a streaming dump of the bytes in the order they were entered into the device. Furthermore, the files are known to be fragmented. In addition, some of the device contains erased files or other garbage. Once the garbage has been recognized and discarded and the fragmented files reassembled, the reverse engineering of the codes can be undertaken with only a partial, and sometimes incorrect, understanding of the CPU on which the codes run. In fact, deducing the structure and function of the CPU is part of the project, since some of the 3.2 gigabytes are the binary specifications for the computer-assisted-manufacturing process that fabricates the CPU. In addition, one must also consider that the huge database also contains code generated from the result of literally millions of maintenance revisions performed by the worst possible set of kludge-using, spaghetti-coding, opportunistic hackers who delight in clever tricks like writing self-modifying code and relying upon undocumented system quirks.

Proteomes

The proteome project is a large-scale programme dealing in an integral way with patterns of expression of proteins in biological systems, in ways that complement and extend genome projects. What kinds of data would we like to measure, and what mature experimental techniques exist to determine them? The basic goal is a spatio-temporal description of the deployment of proteins in the organism. The rates of synthesis of different proteins vary among different tissues and different cell types and states of activity [2]. Methods are available for analysis of transcription patterns of genes. However, because proteins 'turn over' at different rates, it is also necessary to measure proteins directly. High-resolution Two-Dimensional Polyacylamide Gel Electrophoresis (2D PAGE) shows the pattern of protein content in a sample. Mass-spectroscopic techniques identify the proteins into which the sample has been separated [3].

In principle, a database of amino acid sequences of proteins is inherent in the database of nucleotide sequences of DNA, by virtue of the genetic code. Indeed, new protein sequence data are now being determined almost exclusively by translation of DNA sequences, rather than by direct sequencing of proteins. (Historically the chemical problem of determining amino acid sequences of proteins directly was solved before the genetic code was established and before methods for determination of nucleotide sequences of DNA were developed. [4] F. Sanger's sequencing of insulin in 1955 first proved that proteins had definite amino acid sequences, a proposition that until then was hypothetical.) Should any distinction be made between amino acid sequences determined directly from proteins and those determined by translation from DNA? First, we must assume that it is possible correctly to identify within the DNA data stream the regions that encode proteins. The pattern-recognition programs that address this question are subject to three types of errors: A genuine protein sequence may be missed entirely, or an incomplete protein may be reported, or a gene may be incorrectly spliced. Several variations on the theme add to the complexity: Genes for different proteins may overlap, or genes may be assembled from exons in different ways in different tissues. Conversely, some genetic sequences that appear to code for proteins may in fact are defective or not expressed. A protein inferred from a genome sequence is a hypothetical object until an experiment verifies its existence. Second, in many cases the expression of a gene produces a molecule that must be modified within a cell, to make a mature protein that differs significantly from the one suggested by translation of the gene sequence. In many cases the missing details of post-translational modifications-the molecular analogues of body piercing are quite important. Post-translational modifications include addition of ligands (for instance the covalently-bound haem group of cytochrome c), glycosylation, methylation, excision of peptides, and many others. Patterns of disulphide bridges-primary chemical bonds between cysteine residues-cannot be deduced from the amino acid sequence. In some cases, mRNA is edited before translation, creating changes in amino acid sequences that are not inferable from the genes

How hereditary information is stored, passed on, and implemented is perhaps the fundamental problem of biology. Three types of maps have been essential (see Box):

1. Linkage maps of genes

2. Banding patterns of chromosomes

3. DNA sequences

These represent three very different types of data. Genes, as discovered by Mendel, were entirely abstract entities. Chromosomes are physical objects; the banding patterns their visible landmarks. Only with DNA sequences are we dealing directly

with stored hereditary information in its physical form. It was the very great achievement of the last century of biology to forge connections between these three types of data. The first steps and giant strides they were indeed-proved that, for any chromosome, the maps are one-dimensional arrays, and indeed that they are collinear. Any schoolchild now knows that genes are strung out along chromosomes, and that each gene corresponds to a DNA sequence. But the proofs of these statements earned a large number of Nobel prizes. Splitting a long molecule of DNA-for example, the DNA in an entire chromosome-into fragments of convenient size for cloning and sequencing requires additional maps to report the order of the fragments, so that the entire sequence can be reconstructed from the sequences of the fragments. A restriction endonuclease is an enzyme that cuts DNA at a specific sequence, usually about 6 bp long. Cutting DNA with several restriction enzymes with different specificities produces sets of overlapping fragments. From the sizes of the fragments it is possible to construct a restriction map, stating the order and distance between the restriction enzyme cleavage sites. A mutation in one of these cleavage sites will change the sizes of the fragments produced by the corresponding enzyme, allowing the mutation to be located in the map. Restriction enzymes can produce fairly large pieces of DNA. Cutting the DNA into smaller pieces, which are cloned and ordered by sequence overlaps.

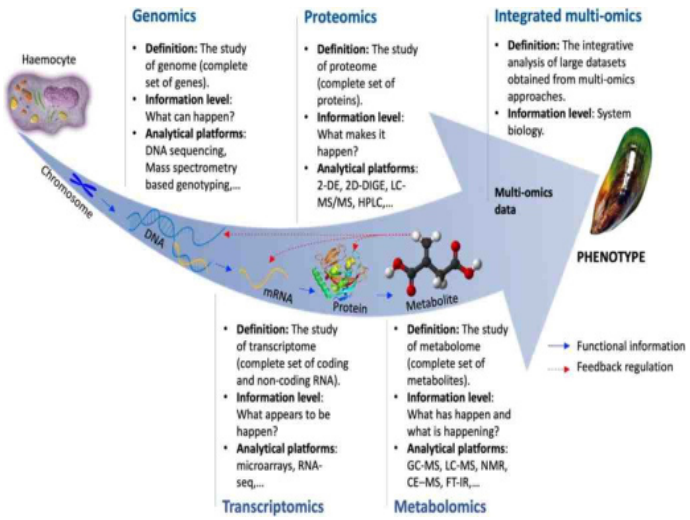

Figure 1: Depicts the integrated omics approaches (Genomics, Trancriptomics, Proteomics, Metabolomics) in eurkaryotic systems through analytical platforms.

In the past, the connections between chromosomes, genes and DNA sequences have been essential for identifying the molecular deficits underlying inherited diseases, such as Huntington disease or cystic fibrosis. Sequencing of the human genome has changed the situation radically. Given a disease attributable to a defective protein: If we know the protein involved, we can pursue rational approaches to therapy. If we know the gene involved, we can devise tests to identify sufferers or carriers. In many cases, knowledge of the chromosomal location of the gene is unnecessary for either therapy or detection; it is required only for identifying the gene, providing a bridge between the patterns of inheritance and the DNA sequence. (This is not true of diseases arising from chromosome abnormalities). For instance, in the case of sickle-cell anaemia, we know the protein involved. The disease arises from a single point mutation in haemoglobin. We can proceed directly to drug design. We need the DNA sequence only for genetic testing and counselling. In contrast, if we know neither the protein nor the gene, we must somehow go from the phenotype back to the gene, a process called positional cloning or reverse genetics. Positional cloning used to involve a kind of 'Tinker to Evers to Chance' cascade from the gene map to the chromosome map to the DNA sequence. Later we shall see how recent developments have short-circuited this process. Patterns of inheritance identify the type of genetic defect responsible for a condition. They show, for example, that Huntington disease and cystic fibrosis are caused by single genes. To find the gene associated with cystic fibrosis it was necessary to begin with the gene map, using linkage patterns of heredity in affected families to localize the affected gene to a particular region of a particular chromosome. Knowing the general region of the chromosome, it was then possible to search the DNA of that region to identify candidate genes, and finally to pinpoint the particular gene responsible and sequence it (see Boxes, Identification of the cystic fibrosis gene, and Positional cloning: Finding the cystic fibrosis gene.) In contrast, many diseases do not show simple inheritance, or, even if only a single gene is involved, heredity creates only a predisposition, the clinical consequences of which depend on environmental factors. The full human genome sequence, and measurements of expression patterns, will be essential to identify the genetic components of these more complex cases. initial (5') exon starts with a transcription start point, preceded by a core promotor site such as the TATA box typically ~30 bp upstream. It is free of in-frame stop codons, and ends immediately before a GT splice signal. (Occasionally a noncoding exon precedes the exon that contains the initiator codon.) Internal exons, like initial exons, are free of in-frame stop codons. They begin immediately after an AG splice signal and end immediately before a GT splice signal. The final (3') exon starts immediately after an AG splice signal and ends with a stop codon, followed by a polyadenylation signal sequence. (Occasionally a noncoding exon follows the exon that contains the stop codon.) All coding regions have nonrandom sequence characteristics, based partly on codon usage preferences. Empirically, it is found that statistics of hexanucleotides perform best in distinguishing coding from noncoding regions. Starting from a set of known genes from an organism as a training set, pattern recognition programs can be tuned to particular genomes. Accurate gene detection is a crucial component of genome sequence analysis. This problem is an important focus of current research.

Genomes of prokaryotes

Most prokaryotic cells contain their genetic material in the form of a large single circular piece of double-stranded DNA, usually less than 5 Mbp long. In addition they may contain plasmids. The protein-coding regions of bacterial genomes do not contain introns. In many prokaryotic genomes the protein-coding regions are partially organized into *operons*-tandem genes transcribed into a single messenger RNA molecule, under common transcriptional control. In bacteria, the genes of many operons code for proteins with related functions. For instance, successive genes in the trp operon of *E. coli* code for proteins that catalyse successive steps in the biosynthesis of tryptophan. Reaction step (1) trpE and trpD encode two components of anthranilate synthase. This tetrameric enzyme, comprising two copies of each subunit, catalyses the conversion of chorismate to anthranilate. Reaction step (2) The protein encoded by trpD also catalyses the subsequent phosphoribosylation of anthranilate. Reaction step (3) trpC encodes another bifunctional enzyme, phosphoribosylanthranilate isomerase-indoleglycerol-phosphate synthase. It converts phosphoribosyl anthranilate to indoleglycerolphosphate, through the intermediate, carboxy-

phenylaminodeoxyribulose phosphate. Reaction steps (4) and (5) trpB and trpA encode the subunits, respectively, of a third bifunctional enzyme, tryptophan synthase (an 2x2 tetramer). A tunnel within the structure of this enzyme delivers, without release to the solvent, the intermediate produced by the ubunit-endoleglycerolphosphate →indole to the active site of the subunit which converts indole →tryptophan. A separate gene, trpR, not closely linked to this operon, codes for the trp repressor. The repressor can bind to the operator sequence in the DNA (within the control region) only when binding tryptophan. Binding of the repressor blocks access of RNA polymerase to the promoter, turning the pathway off when tryptophan is abundant. Further control of transcription in response to tryptophan levels is exerted by the attenuator element in the mRNA, within the leader sequence. The attenuator region (1) contains two tandem trp codons and (2) can adopt alternative secondary structures, one of which terminates transcription. Levels of tryptophan govern levels of trp-tRNAs, which govern the rate of progress of the tandem trp codons through the ribosome. Stalling on the ribosome at the tandem trp codons in response to low tryptophan levels reduces the formation of the mRNA secondary structure that terminates transcription. The typical prokaryotic genome contains only a relatively small amount of noncoding DNA (in comparison with eukaryotes), distributed throughout the sequence. In *E. coli* only ~11% of the DNA is noncoding.

The distribution of protein-coding genes over the genome of *E. coli* does not seem to follow any simple rules, either along the DNA or on different strands. Indeed, comparison of strains suggests that genes are mobile. The *E. coli* genome is relatively gene dense. Genes coding for proteins or structural RNAs occupy~89% of the sequence. The average size of an open reading frame is 317 amino acids. If the genes were evenly distributed, the average intergenic region would be 130 bp; the observed average distance between genes is 118 bp. However, the sizes of intergenic regions vary considerably. Some intergenic regions are large. These contain sites of regulatory function, and repeated sequences. The longest intergenic region, 1730 bp, contains noncoding repeat sequences. Approximately three-quarters of the transcribed units contain only one gene; the rest contain several consecutive genes, or operons. It is estimated that the *E. coli* genome contains 630-700 operons. Operons vary in size, although few contain more than five genes. Genes within operons tend to have related functions. In some cases, the same DNA sequence encodes parts of more than one polypeptide chain. One gene codes for both the and subunits of DNA polymerase III.

Translation of the entire gene forms the subunit and the subunit corresponds approximately to the N-terminal two-thirds of the subunit. A frameshift on the ribosome at this point leads to chain termination 50% of the time, causing a 1:1 ratio of expressed and subunits. There do not appear to be any overlapping genes in which different reading frames both code for expressed proteins. In other cases, the same polypeptide chain appears in more than one enzyme. A protein that functions on its own as lipoate dehydrogenase is also an essential subunit of pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase and the glycine cleavage complex. Having the complete genome, we can examine the protein repertoire of *E. coli*. The larges class of proteins are the enzymes-appoximately 30% of the total genes. Many enzymatic functions are shared by more than one protein. Some of these sets of functionally similar enzymes are very closely related, and appear to have arisen by duplication, either in *E. coli* itself or in an ancestor. Other sets of functionally similar enzymes have very dissimilar sequences, and differ in specificity, regulation, or intracellular location.

Several features of *E. coli's* generous endowment of enzymes give it an extensive and versatile metabolic competence, which allow it to grow and compete under varying conditions: It can synthesize all components of proteins and nucleic acids (amino acids and nucleotides), and cofactors. It has metabolic flexibility: Both aerobic and anaerobic growth are possible, utilizing different pathways of energy capture. It can grow on many different carbon sources. Not all metabolic pathways are active at any time, the versatility allows response to changes in conditions. Even for specific metabolic reactions there are many cases of multiple enzymes. These provide redundancy, and contribute to an ability to tune metabolism to va rying conditions. However, *E. coli* does not possess a complete range of enzymatic capacity. It cannot fix CO2 or N2. We have described here some of the *static* features of the *E. coli* genome and its protein repertoire. Current research is moving into dynamic aspects, to investigations of protein expression patterns in time and space.

Eukaryotic genome

In eukaryotic cells, the majority of DNA is in the nucleus, separated into bundles of nucleoprotein, the chromosomes. Each chromosome contains a single double-stranded DNA molecule. Smaller amounts of DNA appear in organelles-mitochondria and chloroplasts. The organelles originated as intracellular parasites. Organelle genomes usually have the form of circular doublestranded DNA, but are sometimes linear and sometimes appear as multiple circles. The genetic code by which organelle genes are translated differs from that of nuclear genes. Nuclear genomes of different species vary widely in size (see pages 68-69). The correlation between genome size and complexity of the organism is very rough. It certainly does not support any preconception that humans stand on a pinnacle. In many cases differences in genome size reflect different amounts of simple repetitive sequences. In addition to variation in DNA content, eukaryotic species vary in the number of chromosomes and distribution of genes among them. Some differences in the distribution of genes among chromosomes involve translocations, or chromosome fragmentations or joinings. For instance, humans have 23 pairs of chromosomes; chimpanzees have 24. Human chromosome 2 is equivalent to a fusion of chimpanzee chromosomes 12 and 13. The difficulty of chromosome pairing during mitosis in a zygote after such an event can contribute to the reproductive isolation required for species separation. Other differences in chromosome complement reflect duplication or hybridization events. The wheat first used in agriculture, in the Middle East at least 10,000-15,000 years ago, is a diploid called *einkorn* (*Triticum monococcum*), containing 14 pairs of chromosomes. *Emmer* wheat (*T. dicoccum*), also cultivated since Palaeolithic times, and *durum* wheat (*T. turgidum*), are merged hybrids of relatives of einkorn with other wild grasses, to form tetraploid species. Additional hybridizations, to different wild wheats, gave hexaploid forms, including *spelt* (*T. spelta*), and modern common wheat *T. aestivum. Triticale*, a robust crop developed in modern agriculture and currently used primarily for animal feed, is an artificial genus arising from crossing durum wheat (*Triticum turgidum*) and rye (*Secale cereale*). All these species are still cultivated-some to only minor extents-and have their individual uses in cooking. Spelt, or *farro* in Italian, is the basis of a well-known soup; pasta is made from durum wheat; and bread from *T. aestivum*. Even within single chromosomes, gene families are common in eukaryotes. Some family mem-

bers are paralogues-related genes that have diverged to provide separate functions in the same species. (Orthologues, in contrast, are homologues that perform the same function in different species. For instance, human and globin are paralogues, and human and horse myoglobin are orthologues.) Other related sequences may be pseudogenes, which may have arisen by duplication, or by retrotransposition from messenger RNA, followed by the accumulation of mutations to the point of loss of function.

Human genome

In February 2001, the International Human Genome Sequencing Consortium and Celera Genomics published, separately, drafts of the human genome. The sequence amounted to~3.2 x 109 bp, thirty times larger than the genomes of *C. elegans* or *D. melanogaster*. One reason for this disparity in size is that codingsequences form less than 5% of the human genome; repeat sequences over 50%. Perhaps the most surprising feature was the small number of genes identified. The finding of only about 20,000-25,000 genes suggests that alternative splicing patterns make a very significant contribution to our protein repertoire. It is estimated that~50% of genes have alternative splicing patterns. The published human genome is distributed over 22 chromosome pairs plus the X and Y chromosomes. The DNA contents of the autosomes range from 279 Mbp down to 48 Mbp. The X chromosome contains 163 Mbp and the Y chromosome only 51 Mbp. The exons of human protein-coding genes are relatively small compared to those in other known eukaryotic genomes. The introns are relatively long. As a result many protein-coding genes span long stretches of DNA. For instance, the dystrophin gene, coding for a 3685 amino acid protein, is 2.4 Mbp long. A comparison of the genomes of yeast, fly, worm and human revealed 1308 groups of proteins that appear in all four. These form a conserved core of proteins for basic functions, including metabolism, DNA replication and repair, and translation. These proteins are made up of individual protein domains, including singledomain proteins, oligomeric proteins, and modular proteins containing many domains (the biggest, the muscle protein titin, contains 250-300 domains.) The proteins of the worm and fly are built from a structural repertoire containing about three times as many domains as the proteins of yeast. Human proteins are built from about twice as many as those of the worm and fly. Most of these domains appear also in bacteria and archaea, but some are specific to (probably, invented by) vertebrates (see Table). These include proteins that mediate activities unique to vertebrates, such as defence and immunity proteins, and proteins in the nervous system; only one of them is an enzyme, a ribonuclease To create new proteins, inventing new domains is a unusual event. It is far more common to create different combinations of existing domains in increasingly complex ways. This process can occur independently, and take different courses, in different phyla. Gene duplication followed by divergence is a mechanism for creating protein families. For instance, there are 906 genes pseudogenes for olfactory receptors in the human genome. These are estimated to bind~10,000 odour molecules. Homologues have been demonstrated in yeast and other fungi (some comparisons are odorous), but it is the need of vertebrates for a highly-developed sense of smell that multiplied and specialized the family to such a great extent. Eighty percent of the human olfactory receptor genes are in clusters. Compare the small size of the globin gene cluster (page 90), which did not require such great variety.

Transcriptomics

Transcriptomics is the study of transcriptomes, which are the sum of all coding and non-coding RNA transcripts produced by the genome during development or under specific circumstances [5]**.** Hence, a transcriptome provides a snapshot of the total transcripts present in a cell at a given time, which reflects the genes that are actively expressed [6]. Transcriptomics uses several different techniques, including Serial/Cap Analysis of Gene Expression (SAGE/CAGE), Expressed Sequence Tag (EST), Suppression Subtractive Hybridization (SSH), microarrays and RNA sequencing (RNA-Seq) to measure the expression of genes in distinct cell populations that are affected by different treatments, diseases or environmental factors at different time points. Some of these techniques, such as SAGE, CAGE, and EST (based on Sanger sequencing of cDNA or EST) are no longer used for transcriptomics analyses. SSH is a Polymerase Chain Reaction (PCR-based approach) which amplifies differentially expressed cDNAs (complementary DNAs) fragments. SSH offers many advantages in profiling gene transcripts, so it is still a widely used method for identification of gene expression [7]. DNA microarrays are a set of microscopic DNA spots which are arrayed on a solid substrate [8]. Since microarrays can only detect sequences homologous to what is on the array, prior genomic knowledge of the organism of interest (e.g., annotated genome sequence,and ESTs) is required to generate the probes of the array [6]. Due to the relatively low cost and high throughput, microarrays remain a reliable tool in model organisms, where highly standardized platforms have been developed over the years [9]. RNA-Seq is a revolutionary tool that uses recently developed deep-sequencing technologies combined with computational methods for both mapping and quantifying transcriptomes [6]. RNA-Seq methodologies have several advantages (e.g., higher throughput, sensitivity, accuracy, long read lengths, no need of prior knowledge of the organism's genome) and have been rapidly adopted over hybridisation-based approaches as the dominant transcriptomics technique in the last two decades [10].

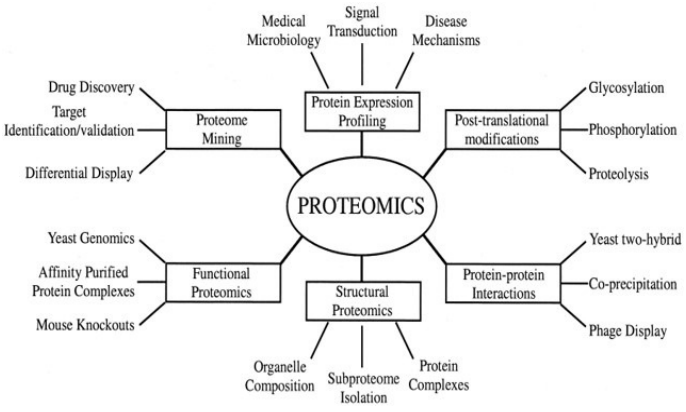

Figure 2: Demonstrates the proteomics applications in various fields.

Metabolomics

Metabolomics is the scientific study of the set of metabolites within biological samples (e.g., tissues, body fluids, entire organisms), which is called the metabolome. Metabolomics is one of the newest omics, and has been rapidly growing in the last decade. Since metabolites are endproduct of gene expression and cell activity, they are highly sensitive to environmental change. Thus, metabolomics can represent a physical snapshot of what is actually happening in the organism at a given time [11]. To this end, metabolomics is applied to characterize en-

dogenous metabolic changes in biological samples within different environmental conditions and biomarkers involved in these processes. Metabolomics uses various analytical platforms, such as Infrared Spectroscopy (IR), raman spectroscopy, Nuclear Magnetic Resonance (NMR) and many Mass Spectrometry (MS) techniques, including Directinfusion Mass Spectrometry (DI-MS), Matrix-Assisted Laser Desorption/ Ionization Mass Spectrometry (MALDI-IMS), Capillary Electrophoresis-Mass Spectrometry (CE-MS), Gas Chromatography-Mass Spectrometry (GC-MS), and Liquid Chromatography-Mass Spectrometry (LC-MS). With sufficient high throughput and resolution capabilities, NMR and MS are the most widely applied analytical tools [12].

Conclusions

Classification of transcriptomes, proteomes and metabolomes of blood cells in response to pathogenic infections and other environmental stressors have revealed valuable information regarding the mechanisms that drive the innate immune system in response to stress challenges, as well as insights regarding complex host-pathogen-environment interactions across vertebrate as well as invertebrate species. Tranciptomics results reveals the number of coding and non coding transcripts in the biological system, many immunological studies such as host pathogen interactions and stress conditions this transcriptomic approaches also brings detailed understanding about the transcripts. Similarly, proteomics and metablomics research favors to identify the changes in the physiological status of the organism, including specific molecular pathways involved in these processes. Furthermore, the differently expressed molecules have been identified through these omics studies can be used as candidate biomarkers with applications in breeding selection programs, disease diagnosis and environmental monitoring. However, despite these significant biotechnological advances, the application of omics tools in medical research is currently hindered by several challenges and bottlenecks. In this contribution, this chapter deals with the major advances, current perspectives and future directions of three main omics (transcriptomics, proteomics and metabolomics) to their application in various platforms. The emergent field of omics and their applications has significantly enhanced our understanding of the functions of blood cells in response to microbial infections and environmental stress. Omic approaches unraveled the new good model system for host pathogen interactions and many other applications. Though, future growths of omics applications possess the lack of extensive libraries and databases (e.g., genomics databases across a range of species) and/or require complex data processing and interpretation of results. Discovery of unknown transcripts, proteins and metabolites still presents a problem for data interpretation, as well as the relative high costs of analyses and instrument accessibility.

References

1. Primrose SB, Twyman RM. Principles of Gene Manipulation and Genomics S EVENTH EDITION. © 2006 Blackwell Publishing.
2. Brown, Terence A. “Genome Anatomies.” Genomes. 2nd Edition., U.S. National Library of Medicine, 1970 edeno micro. 1970.
3. Sivakamavalli J, Vaseeharan B. Enzymatic elucidation of haemocyanin from Kuruma shrimp Marsupenaeus japonicus and its molecular recognition mechanism towards pathogens. Journal of Biomolecular Structure and Dynamics. 2015; 33: 1302-1314.
4. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of molecular biology. 1975; 94: 441-448.
5. Sheng J, Zheng X, Wang J, Zeng X, Zhou F, et al. Transcriptomics and proteomics reveal genetic and biological basis of superior biomass crop Miscanthus. Scientific reports. 2017; 7: 1-3.
6. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies, PLoS computational biology. 2017; 13.
7. Rusaini OL. Suppression subtractive hybridization in penaeid prawns: An approach in identifying diseases and differentially expressed genes. Journal of Aquatic Marine Biology. 2018; 7: 00177.
8. Romanov M, Ravi KV, Chan ES. Inventors: Oracle international corp, assignee. Temporally-correlated activity streams for conferences. United States patent US 8,812,510. 2014.
9. Wang JW, Czech B, Weigel D. miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. Cell. 2009; 138: 738-749.
10. Su LT, Li JW, Liu DQ, Zhai Y, Zhang HJ, et al. A novel MYB transcription factor, GmMYBJ1, from soybean confers drought and cold tolerance in Arabidopsis thaliana. Gene. 2014; 538: 46-55.
11. Nguyen TV, Alfaro AC, Young T, Ravi S, Merien F. Metabolomics Study of Immune Responses of New Zealand Greenshell™ Mussels (Perna canaliculus) Infected with Pathogenic Vibrio sp. Marine Biotechnology. 2018; 20: 396-409.
12. Nguyen TV, Alfaro AC, Young T, Merien F. Tissue-specific immune responses to Vibrio sp. infection in mussels (Perna canaliculus): A metabolomics approach. Aquaculture. 2019; 500: 118-125.