



## **OMICS in Molecular Biology**

**Lekshmi R. S.**

*Ph.D. Scholar, Department of Plant Biotechnology, College of Agriculture, Kerala Agricultural University, Trivandrum*

---

### **Abstract**

*Molecular biology is the study of biology at a molecular level. The field overlaps with other areas of biology and chemistry, particularly genetics and biochemistry. Molecular biology chiefly concerns itself with understanding the interactions between the various systems of a cell, including the interactions between DNA, RNA and protein biosynthesis, the transcription factor as well as learning how these interactions are regulated. The scientific world is not the same again, and neither can society as the impact of 'omics' spreads from the laboratory to everyday life. This review can give a brief idea on different Omics in Molecular biology and its applications.*

*Keywords- Omics; genomics; proteomics; transcriptomics; metabolomics.*

---

### **1. INTRODUCTION**

The sum total of all genes in an individual organism is its genome. The study of all the genes in a cell is called genomics. The word "genome" was derived from the German word 'Genom', attributed to Hans Winkler. The term "genomics" was coined by Dr. Tom Roderick. Fred Sanger and his colleagues first sequenced the complete genome of a virus and a mitochondrion. They established the techniques of genome mapping, sequencing, and data storage and bioinformatics analysis in 1970-1980s. Determining the genomic sequence, however, is only the beginning of genomics. Once this is done, the genomic sequence is used to study the function of the numerous genes (functional genomics), to compare the genes in one organism with those of another (comparative genomics), or to generate the 3-D structure of one or more proteins from each protein family, thus offering clues to their function (structural genomics).

### **2. STRUCTURAL GENOMICS**

Describes the 3-dimensional structure of every protein encoded by a given genome. The principal difference between structural genomics and traditional structural prediction is that structural genomics attempts to determine the structure of every protein encoded by the genome, rather than focusing on one particular protein. With full-genome sequences available, structure prediction can be done more quickly through a combination of experimental and modeling approaches, especially because the availability of large number of sequenced genomes and previously solved protein structures allows scientists to model protein structure on the structures of previously solved homolog.

### **3. FUNCTIONAL GENOMICS**

Field of molecular biology that attempts to describe gene (and protein) functions and interactions. Functional genomics attempts to answer questions about the function of DNA at the levels of genes, RNA transcripts, and protein products. Functional genomics have a genome-wide approach, rather than "gene-by-gene" approach. The knowledge of full genomes has created the possibility for the field of functional genomics. The most important tools here are microarrays and bioinformatics.

#### **4. COMPARATIVE GENOMICS**

Study of the relationship of genome structure and function across different biological species or strains. Comparative genomics understands the function and evolutionary processes that act on genomes. Gene finding is an important application of comparative genomics. It is based on the genome mapping and sequencing, comparing structure of known genes and genome, to understand the gene function, mechanism of the expression and the evolution of species. Using the homology of the sequence and structure in coding which between human genome and model organism genome, to clone human disease genes, revealing mechanisms on gene function.

#### **5. EPIGENOMICS**

Epigenomics is defined as the study of inherited changes in phenotype or gene expression caused by mechanisms other than mutations in the underlying DNA sequence. Epigenetic markers include DNA methylation and modification of histone tails. In eukaryaotes, DNA can be modified by methylation of cytosine bases, while in prokaryaotes DNA methylation occurs primarily on adenosine bases. Aberrant or increased methylation has been correlated with gene silencing and the development of several cancers.

Histones are subject to several different covalent modifications, including methylation, acetylation, phosphorylation, ubiquitylation and sumoylation. Acetylation is the most commonly studied of these modifications and a strong correlation between histone acetylation and active transcription has been established. Conversely, many histone methylation events are correlated with transcriptional silencing. Different histone modifications likely function in differing ways. For example, acetylation at one position may have a different effect than acetylation at another position. Multiple modifications can exist simultaneously and are likely working together to influence chromatin stage and gene expression. The concept of multiple dynamic modifications regulating gene expression in a systematic and reproducible fashion is known as the histone code.

#### **6. METAGENOMICS**

Metagenomics is the study of metagenomes, genetic material recovered directly from environmental samples. The broad field may also be referred to as environmental genomics, ecogenomics or community genomics. While traditional microbiology and microbial genome sequencing and genomics rely upon cultivated clonal cultures, early environmental gene sequencing cloned specific genes (often the 16S rRNA gene) to produce a profile of diversity in a natural sample. Such work revealed that the vast majority of microbial biodiversity had been missed by cultivation-based methods. Recent studies use "shotgun" Sanger sequencing or massively parallel pyrosequencing to get largely unbiased samples of all genes from all the members of the sampled communities (Morozova et al., 2008). Because of its ability to reveal the previously hidden diversity of microscopic life, metagenomics offers a powerful lens for viewing the microbial world that has the potential to revolutionize understanding of the entire living world.

#### **7. APPLICATIONS OF GENOMICS**

##### **7.1 Genomics in human health:**

With the advancement in genomics there is a vast improvement in the diagnoses of diseases. As sequencing techniques developed the whole genome of large number of organisms have been sequenced. Human genome is fully sequenced and annotated and so it is easy to identify genetic pre-disposition to various genetic diseases. Biomarkers have been identified for

various genetic disorders in human. Genomics helps in drug discovery, that are tailored to a person's genetic information and this is called 'Personalized medicine'.

A full genome scan could shed light onto a person's genetic risk for certain diseases. Encourage one to make lifestyle changes if they are found to carry genes that could cause an increased risk for disease when linked to environmental factors. Companies like '23 and me' are striving to make facilities for scanning one's genome for identifying the biomarkers associated with various genetic diseases. Full genome information is useful if one is a carrier for a genetic disease. If one's partner is a carrier too, embryos could be screened for the defect, or get the children tested and treated for the disease early. Knowledge of certain genetic characteristics will help doctors tailor medicine to individual patients.

Whole genome sequencing is expensive for every day clinical needs. Human Genome Project cost was 3 billion USD. Now the cost has come down to 5495 USD by 'Gene by Gene Ltd'. Many companies are striving towards the goal of finding a way to sequence an entire genome in 1000 USD. They are also aiming to bring that facility within hospital settings.

### **7.2 Genomics in Agriculture:**

Sequencing of whole genome of crop plants facilitates identification of agronomically important and novel genes for important traits. This helps to create more nutrient rich and safe food. Help to identify production systems that can increase yield in particular soils and climatic conditions and reduce losses due to pests, diseases. To develop plants resistant to various biotic and abiotic stresses.

### **7.3 Farm animal genomics:**

In animal husbandry field genomics has been applied for sequencing whole genome and EST sequences. Genomics has been applied on various areas like xenotransplantation, increasing livestock productivity & bioengineering. RNAi based techniques has been developed as a means to improve the carcass value and meat quantity in livestock industry.

### **7.4 Environmental genomics:**

Create more sustainable energy sources such as biofuels. Environmental genomics helps to make special instruments called biosensors for detection of various pollutants and could take measures to control it. Bioremediation of oil spills, and can also monitor biodiversity and identify new species, making every species count.

## **8. TRANSCRIPTOMICS**

The study of the complete set of RNAs (transcriptome) encoded by the genome of a specific cell or organism at a specific time or under a specific set of conditions is called transcriptomics. It is the subfield of functional genomics which focuses on the gene expression. It is considered as the "intermediate phenotype" as it is very close to the genotype in the pathway that determines the final phenotype of an organism.

### **8.1 Transcriptome sequencing**

Gene annotation, discovering novel transcriptional units and alternatively spliced forms, mapping ESTs to a known reference genome or transcriptome for comparative analysis, discovery of fusion transcripts (e.g., cancer), detection of genetic variations, including SNPs and insertions/deletions, annotation of coding SNPs, discovery of transcript isoforms, identification of regulatory RNAs, characterization of splice junctions, determination of the relative abundance of transcripts.

## **9 APPLICATIONS OF TRANSCRIPTOMICS**

### **9.1 Application in human health**

- Facilitates analysis of minute amounts of transcripts, by RNA amplification and generation of a cDNA array enriched for genes expressed in stem cells studies.
- Single cell transcriptome analysis
- Live cell imaging and lineage tracing deciphers full gene expression network of individual cells in embryos and adults and to study diseases.

### **9.2 Application in the field of agriculture**

- Differential gene expression studies and transcript profiling from tissues subjected to biotic & abiotic stress
- SNP and gene expression variation data from RNA-seq data sets - applicable to species with no genomic resources.
- Provide markers to accelerate breeding by marker-assisted selection.
- Associative transcriptomics, uses transcriptome sequencing to identify and score molecular markers representing variation in both gene sequences and gene expression.

## **10 PROTEOMICS**

The word proteomics is PROTEins expressed by a genOME. Proteomics is the set of all proteins expressed in a cell, tissue or an organism or it can be explained as the protein complement of the genome. Proteomics is the (science and technology) process of studying the proteome. Protein sequence and structure can't be directly deduced from gene sequence or from transcript sequence. There occur different manipulations at post transcriptional level itself, like alternate splicing, trans splicing etc. At post translational level also there are different modifications that hinder a direct understanding of the protein structure and function like phosphorylation, glycosylation, acetylation, methylation etc. The proteome is highly dynamic and it changes from time to time in response to different environmental stimuli. Even the same cell may make different sets of proteins at different times, or under different conditions.

### **10.1 Importance of proteomics:**

Scientific study of proteins generates insight on how they affect cell processes. Conversely, this study also investigates how proteins themselves are affected by cell processes or the external environment. Proteins provide intricate control of cellular machinery, and are in many cases components of that same machinery. They serve a variety of functions within the cell, and there are thousands of distinct proteins and peptides in almost every organism. This great variety comes from a phenomenon known as alternative splicing, in which a particular gene in a cell's DNA can create multiple protein types, based on the demands of the cell at a given time.

The goal of proteomics is to analyze the varying proteomes of an organism at different times, in order to highlight differences between them. Put more simply, proteomics analyzes the structure and function of biological systems. For example, the protein content of a cancerous cell is often different from that of a healthy cell. Certain proteins in the cancerous cell may not be present in the healthy cell, making these unique proteins good targets for anti-cancer drugs. The realization of this goal is difficult; both purification and identification of proteins in any organism can be hindered by a multitude of biological and environmental factors.

- Differential proteomics, phosphoproteomics, glycoproteomics, degradomics, structural proteomics, interaction proteomics and metabolomics of plants are the recent field of study.
- To construct databases of protein sequences, modifications, natural products and metabolites
- For proteomic and metabolic data mining with the aid of bioinformatics tools
- To discover novel metabolic networks and signaling transduction pathways,
- Find specific protein targets and biomarkers in the plant physiological processes.

### **11. APPLICATIONS OF PROTEOMICS**

- Proteomics studies the relationships between changes in proteome composition and stress acclimation in plants and animals.
- Proteomics studies host – pathogen interaction
- Proteomics studies plant response to abiotic stress factors
- Proteomics studies cell signaling pathways

#### **11.1 Clinical proteomics:**

- Detection of functional dysregulation of protein interactions related with diseases.
- Application of proteomic at the bedside, for condition like cancer.
- To detect cancer earlier.
- To discover the protein targets and imaging biomarkers, and finally to tailor the therapy to the patient.

### **12. METABOLOMICS**

“Systematic study of the unique chemical fingerprints that specific cellular processes leave behind”. The metabolome - collection of all metabolites in a biological cell, tissue, organ or organism, which are the end products of cellular processes. Metabolic profiling can give an instantaneous snapshot of the physiology of a cell. One of the challenges of systems biology and functional genomics is to integrate proteomic, transcriptomic, and metabolomic information to give a more complete picture of living organisms.

### **13. IMPORTANCE OF METABOLOMICS**

The metabolome comprises the complete set of metabolites, the non-genetically encoded substrates, intermediates, and products of metabolic pathways, associated to a cell. By representing integrative information across multiple functional levels and by linking DNA encoded processes with the environment, the metabolome offers a window to map core attributes responsible for different phenotypes (Camacho et al., 2005). Given increasing demand to quantitatively identify the metabolome and understand how trafficking of metabolites through the metabolic network impact cellular behavior, metabolomics has emerged as an important complementary technology to the cell-wide measurements of mRNA, proteins, fluxes, and interactions (e.g. protein-DNA). Metabolomics is already a powerful tool in drug discovery and development and in metabolic engineering. While maintaining these strengths, the field promises to play a heightened role in systems biology research, which is transforming the practice of medicine and our ability to engineer living organisms (Broadhurst et al., 2006).

Small molecules counts. >95% of all diagnostic clinical assays test for small molecules 89% of all known drugs are small molecules. 50% of all drugs are derived from pre-existing metabolites. 30% of identified genetic disorders involve diseases of small

molecule metabolism. Small molecules serve as cofactors and signaling molecules to 1000's of proteins

#### **14. APPLICATIONS OF METABOLOMICS**

Toxicology testing, clinical trial testing, fermentation monitoring, food & beverage tests, nutraceutical analysis, drug phenotyping, water quality testing, petrochemical analysis, genetic disease tests, nutritional analysis, clinical blood analysis, clinical urinalysis, cholesterol testing, drug compliance, transplant monitoring, MRS and CS imaging

#### **15. EVALUATION OF GENETICALLY ENGINEERED CROPS USING TRANSCRIPTOMICS, PROTEOMICS AND METABOLOMIC PROFILING TECHNIQUES**

A major principle and guiding tool for the food safety assessment of genetically engineered (GE) crops is the concept of "substantial equivalence" according to principles outlined in OECD, 2006 and FAO. GE crop-derived foods and feeds are compared with their counterparts from parental or near isogenic lines to identify differences and evaluated with respect to safety for humans and animals as well as nutritional quality. Allows a holistic search for alterations in GE crops at different biological levels (transcripts, proteins, metabolites). These methodologies include cDNA microarrays, microRNA fingerprinting, proteome, metabolome, and toxicological profiling. The term "omics" in relation to food and feed safety appeared for the first time in 2005.

#### **16. OMICS AND BIOINFORMATICS**

To understand the behavior of cells, tissues, organs, and the whole organism at the molecular level, methods such as systems biology, bioinformatics, as well as the computational tools are used (Perco et al., 2006). Bioinformatics tools are needed to analyze and make sense of the biological data. *Bioinformatics* is the science of managing and analyzing biological data using advanced computing techniques. Bioinformatics tools include computational tools that mine information from large databases of biological data.

#### **17. COMPUTATIONAL BIOLOGY**

Bioinformatics and databases of biological information can be used to generate "maps" of cellular and physiological pathways and responses (Mochida et al., 2011). This integrative approach is called *computational biology*. Bioinformatics is used to abstract knowledge and principles from large-scale data, to present a complete representation of the cell and the organism, and to predict computationally systems of higher complexity, such as the interaction networks in cellular processes and the phenotypes of whole organisms.

#### **18. SYSTEMS BIOLOGY**

Integration of data from all levels of complexity (genomics, proteomics, metabolomics, and other molecular mechanisms) into a systems view of biological and pathological processes. The goal is to create overall computational models of the functioning of the cell, multicellular systems, and ultimately the organism. These in silico models will provide virtual test systems for evaluating the toxic responses of cells, tissues, and organisms. Compounds will be tested in simulation studies before being applied to cells and tissues to obtain comparative results and validation of the system.



## 19. NATIONAL HUMAN GENOME RESEARCH INSTITUTE

Began as the National Center for Human Genome Research (NCHGR), which was established in 1989 to carry out the role of the National Institutes of Health (NIH) in the International Human Genome Project (HGP). The HGP was developed in collaboration with the United States Department of Energy and begun in 1990 to map the human genome. In 1993, NCHGR expanded its role on the NIH campus by establishing the Division of Intramural Research to apply genome technologies to the study of specific diseases. In 1996, the Center for Inherited Disease Research (CIDR) was also established (co-funded by eight NIH institutes and centers) to study the genetic components of complex disorders. In 1997 the United States Department of Health and Human Services renamed NCHGR the National Human Genome Research Institute (NHGRI), officially elevating it to the status of research institute - one of 27 institutes and centers that make up the NIH. NHGRI is now translating the results of HGP into advances in medicine. They ensure that results and discoveries remain freely available in the public domain. **Genetic variation mapping project**, or **HapMap**, will speed the discovery of genes related to diseases. Project ENCODE is trying to create a comprehensive encyclopedia cataloguing the identity and precise location of protein encoding and the non protein encoding genes within the genome.

## 20. CONCLUSION

Researchers are trying to develop genomics through various modern techniques. Research institutes all over the world are working on genomics and they are focusing on various aspects for the betterment of mankind. This requires extensive coordination and public investment to ensure that results and discoveries remain freely available in the public domain. In the next ten years, genomic science will have an increasing impact beyond the laboratory and computer programme. It is likely that the technologies and methods of genomic science will be common place in hospitals and test laboratories, that they will contribute greatly to improvement of mankind.

## BIBLIOGRAPHY

- [1] Anderson, S. 1981. Sequence and organization of the human mitochondrial genome. *Nature*. **290**: 457–465.
- [2] Brent, M. R. 2007. How does eukaryotic gene prediction work?. *Nature Biotech.* **25**: 883–885.
- [3] Broadhurst, D. I. and Kell, D. B. 2006. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics*. 2(4): 171-196.
- [4] Camacho, D., Fuente, A. and Mendes, P. 2005. The origin of correlations in metabolomics data. *Metabolomics*. 1(1): 53-63.
- [5] Dunwell, J. M., Moya-Leon, M. A. and Herrera, R. 2001. Transcriptome analysis and crop improvement. *Biol. Res.* 34(3): 153-64.
- [6] Gevaert, K. and Vandekerckhove, J. 2000. Protein identification methods in proteomics. *Electrophoresis*. 21: 1145-1154.
- [7] Gorg, A., Weiss, W. and Dunn, M. J. 2004. Current two-dimensional electrophoresis technology for proteomics. *Proteomics*. 4(12): 3665-3685.
- [8] Gregory, T. R. 2005. Synergy between sequence and size in large-scale genomics. *Nature Rev. Genet.* **6**: 699–708.
- [9] Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P. and Bancroft, I. 2012. Associative transcriptomics of traits in the polyploidy crop species *Brassica napus*. *Nature Biotech.* 30: 798-802.
- [10] Hossain, Z., Nouri, M. Z. and Komatsu, S. 2012. Plant cell organelle proteomics in response to abiotic stress. *J. Proteome Res.* 11(1): 37-48.
- [11] Komatsu, S. and Hossain, Z. 2013. Organ-specific proteome analysis for identification of abiotic stress response mechanism in crop. *Front. Pl. Sci.* 4: 71-78.
- [12] Kosova, K., Vitamvas, P., Prasil, I. T. and Renaut, J. 2011. Plant proteome changes under abiotic stress- contribution of proteomics studies to understanding plant stress response. *J. Proteomics*. 74(8): 1301-1322.

- [13] Lage, K., Karlberg, E. O., Storling, Z. M., Olason, P. I., Pedersen, A. G., Rigina, O., Hinsby, A. M., Tumer, Z., Pociot, F., Tommerup, N., Moreau, Y. and Brunak, S. 2007. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature Biotech.* 25: 309-316.
- [14] Liotta, L. A., Kohn, E. C., Petricoin, E. F. 2001. Personalized molecular medicine. *Clinical Proteomics.* 286(18): 2211-2214.
- [15] Mehta, D., Klengel, T., Conneely, K. N., Smith, A. K. Altmann, A., Pace, T. W. Rex-Haffner, M., Loeschner, A., Gonik, M., Mercer, K. B., Bradley, B., Muller-Myhsok, B., Ressler, K. J. and Binder, E. B. 2013. Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proc. Nat. Acad. Sci.* 110(20): 8302-8307.
- [16] Mochida, K. and Shinozaki, K. 2011. Advances in omics and bioinformatics tools for systems analyses of plant functions. *Pl. Cell Physiol.* 52(12): 2017-2038.
- [17] Morozova, O. and Marra, M. A. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics.* 92(5): 255-264.
- [18] Neha, S. and Harikumar, S. L. 2013. Use of genomics and proteomics in pharmaceutical drug discovery and development : a review. *Int. J. of Pharm. and Pharm. Sci.* 5(3): 24-28.
- [19] Perco, P., Rapberger, R., Siehs, C., Lukas, A., Oberbauer, R., Mayer, G. and Mayer, B. 2006. Transforming omics data into context: Bioinformatics on genomics and proteomics raw data. *Electrophoresis.* 27(13): 2659-2675.
- [20] Pray, L. A. 2008. Eukaryotic genome complexity. *Nature Education.* 1(1): 96.
- [21] Reenan, R. 2005. Molecular determinants and guided evolution of species-specific RNA editing. *Nature.* 434: 409-413.
  
- [22] Ricroch, A. E., Berge, J. B. and Kuntz, M. 2011. Evaluation of genetically engineered crops using transcriptomic, proteomic, and metabolomic profiling techniques. *Pl. Physiol.* 155: 1752-1761.
- [23] Stower, H. 2012. Plant genomics: Associative transcriptomics. *Nature Rev. Genet.* 13: 597-603.
- [24] Tang, F., Lao, K. and Surani, M. A. 2011. Development and applications of single-cell transcriptome analysis. *Nature Methods.* 8: 1557-1568.
- [25] Turner, F. S., Clutterbuck, D. R. and Semple, C. A. M. 2003. POCUS: mining genomic sequence annotation to predict disease genes. *Genome Biol.* 4(11): 74-82.
- [26] Van Straalen, N. I. & Roelofs, D. 2006. *Introduction to Ecological Genetics.* New York, Oxford University Press.
- [27] Wirta, V. 2006. Mining the transcriptome – methods and applications. Royal institute of Technology. School of Biotechnology.
- [28] Wulfschuhle, J. D., Liotta, L. A. and Petricoin, E. F. 2003. Proteomic applications for the early detection of cancer. *Nature Rev. Genet.* 3: 267-275.
- [29] Zhang, H., Han, B., Wang, T., Chen, S., Li, H., Zhang, Y. and Dai, S. 2012. Mechanisms of plant salt response: insights from proteomics. *J. Proteome Res.* 11(1): 49-67.
- [30] Zhu, Y., Zhu, G., Guo, Q., Zhu, Z., Wang, C. and Liu, Z. 2013. A comparative proteomic analysis of *Pinelliaternata* leaves exposed to heat stress. *Int. J. Mol. Sci.* 14(10): 20614-20634.



