Chapter 37
Genomics, Proteomics, and Metabolomics

Pradip Chandra Deka
Sir Padampat Singhania University, India

ABSTRACT

Genomics could be viewed as the study of the randomness of DNA sequences. It may be possible to predict the structure of a gene product from the nucleotide sequences and thereby predict its function. The terms "structural genomics" and "functional genomics" were coined to denote the assignment of structure and function to a gene product, respectively. Proteomics focuses on the products of gene, which are basically proteins. Proteins are responsible for the development of phenotype, and proteomics is the bridge between genotype and phenotype. The transcribed mRNAs and their abundance are called transcriptome. Proteomics also deals with the interaction between proteins called intractomics. Metabolomics is concerned with identification, abundance, and localization of all the molecules excluding lipids and polysaccharides in the cell. In this chapter, the basic concepts and analysis of the genomic, proteomic, and metabolomics data for their practical utilization are discussed.

INTRODUCTION

Two important aspects of gene mapping are to identify mutants and establish linkage through appropriate crossing. In situations where the above mentioned methodology is difficult to apply, somatic hybridization and recombinant DNA technology was used to map DNA sequences to specific chromosomes. Initially, most of these sequences were not actually full-length genes but marker sequences such as restriction fragment length polymorphism (RFLPs), single nucleotide polymorphisms (SNPs) and other molecular markers. Once assigned to chromosomes, these markers were used in pedigree analysis to establish linkage between the markers and disease phenotypes for genetic disorders in human. The existing available techniques would be laborious, time consuming, and an insurmountable task.

The study of genomes by using a newly developed method called DNA sequencing has revolutionized the gene mapping in all organisms. The first sequencing of the 5400 nucleotide was made of the virus ΦX174. Sequencing of several other viruses was completed thereafter. But the technology was

DOI: 10.4018/979-8-3693-3026-5.ch037

slow and labor-intensive, limiting its uses to small genomes. It took about two decades to develop the computer-based automated DNA sequencing method, which was amenable for sequencing of large and complex genomes of the eukaryotes.

Advances in recombinant DNA technologies coupled with development of computer aided automated DNA sequencing methods has created a new area of research called genomics. Genomics is basically concerned with the analysis of nucleotide sequences of genes. This involves comparison of nucleotide sequences of the genes, and analysis of the succession of symbols in sequences. Initially attempts were made to elucidate the function of sequences whose functions is unknown, by comparing with the sequences of known function. It is based on the principle that similar sequences encode similar protein structures, and thus they perform similar functions. However, this principle may not be true universally. The second way is to compare sequences known to code for same protein in different organisms, in order to deduce phylogenic relationships. A third approach is to compare the sequences of healthy and diseased organisms, in an attempt to assign genetic causes to specific disease.

In its purest form, genomics could be viewed as the study of the randomness of DNA sequences. This endeavor is still inchoate, since the irregularities and their relation to function are not understood. However, it may be possible to predict the structure of a gene product from the nucleotide sequences. This can then be used to predict its function. Thus the term "structural genomics" and "functional genomics" were coined to denote the assignment of structure and function to a gene product, respectively.

Proteomics focuses on the products of gene, which are basically proteins. Usually, only 10 percent of the genes are actually translated into protein, in any given cell, under a given set of conditions. On the other hand, a given gene sequence can give rise to tens of different proteins, by varying the arrangements of the exons and by post-translational modifications. Thus, proteins are responsible for the development of phenotype, and proteomics is the bridge between genotype and phenotype.

At a particular epoch, the transcribed mRNAs and their abundance are called transcriptome, and all the translated proteins and their abundance or net rates of synthesis are called proteome. Usually there exist huge differences between the transcriptome and proteome. Separating and identifying the proteins from one another through different technique is an important step for generating primary data, which can be used for their analysis. For example, comparison of proteomes of diseased and healthy organisms may lead to identification of the molecular basis of a disease.

Proteomics also deals with the interaction between proteins called intractomics. Information about the affinity of each protein with every other protein in the cell, and non-protein material such as lipid bilayers, polysaccharides, RNA and DNA, constitute the primary data of intractomics. The investigation of protein glycosylation is called glycomics. Investigation of protein products is called metabolomics. Metabolomics is concerned with identification, abundance and localization of all the molecules excluding lipids and polysaccharides, in the cell. In this chapter the basic concepts and analysis of the genomic, proteomic and metabolomics data for their utilization in plant breeding have been discussed.

GENOMICS

The term genomics was first coined by Tom Rodrick in 1986 (Kusha 1998). Genomics involves sequencing and analysis of an organism's genome. The genome is the entire DNA content present within one cell of an organism. Genomics also involves the study of intra- and inter-allelic interactions such as