6
Bioinformatics and Omics

Timothy Craig Allen and Philip T. Cagle

Introduction

The term *genomics* originated in 1920 to describe the complete set of chromosomes and their associated genes; however, it has been in the past decade that the use of omics—genomics, transcriptomics, and proteomics—and bioinformatics has led to dramatic advances in the understanding of the molecular and genetic bases of disease.[1–29] This chapter briefly reviews the subject, with subsequent chapters providing more details on specific technologies.

Bioinformatics

Bioinformatics has become an essential part of omics research and requires unique practical and analytical skills for appropriate results interpretation. Bioinformatics uses computers and statistics to perform extensive omics-related research by searching biologic databases and comparing gene sequences and protein data on a vast scale to identify sequences or proteins that differ between diseased and healthy tissues or between different phenotypes of the same disease.[30–37] The techniques used in omics are called *high throughput* because they involve analysis of very large numbers of genes, gene expressions, or proteins in one procedure or combination of procedures. The vast amounts of data generated by these high-throughput studies typically require computers for analysis and comparison of differences between diseased and physiologic cells and tissues, a key feature of bioinformatics. Omics and bioinformatics are used not only for studying genes and signaling pathways involved in human diseases but also for identifying potential targets of therapy and the design of therapeutic drugs.

Omics

Omics—a suffix signifying the measurement of the entire complement of a given level of biologic molecules and information—today encompasses a variety of new technologies that can help explain normal and abnormal cell pathways, networks, and processes via the simultaneous monitoring of thousands of molecular components.[6,7]

Genomics

Genomics provides platforms for the study of genomes and their genes, including haplotyping and single nucleotide polymorphism detection by investigating single nucleotide polymorphisms (SNPs) and mutations using high-throughput genome sequencing techniques such as high-density DNA microarrays/DNA (oligonucleotide) chips.[7,38–44] The base sequence of the genes of the human mitochondrial genome was completed in 1981, and in 2003 the base sequence of the genes of the entire human genome was completed.[45–47] The human genomic sequence data from the International Human Genome Sequencing Consortium can be mined using tools that are now publicly available.[48] Public databases can also be mined for SNPs, human mitochondrial genomes, and other human DNA polymorphic markers.[49–54]

Transcriptomics

Also termed *functional genomics*, transcriptomics provides information about the expressions of individual genes at the messenger RNA (mRNA) level and correlates patterns of expression with biologic function.[7,55–64] A variety of techniques have been developed to investigate gene expression. These techniques include serial analysis of gene expression (SAGE), suppression subtractive hybridization (SSH), differential display (DD) analysis,

RNA arbitrarily primer–PCR (RAP-PCR), amplified restriction fragment-length polymorphism (AFLP), total gene expression analysis (TOGA), and use of internal standard competitive template primers in a quantitative multiplex RT-PCR method [StaRT-(PCR)], restriction endonucleolytic analysis of differentially expressed sequences (READS), differential screening (DS), high-density cDNA filter hybridization analysis (HDFCA), and gene expression microarrays.[65–67]

Proteomics

Proteomics is discussed in detail in Chapter 13. Proteomics investigates individual protein concentrations present in a biologic system and studies the structural, functional, and regulatory roles of proteins in the cell and in pathways, including how and where they are expressed.[7,68–86] Because gene function is ultimately performed by the proteins transcribed from the genes and mRNA, proteomics is essential to comprehend the actual definitive functioning of a gene or pathway.[3,4] Fortunately, proteomics can be performed on surgical specimens, including needle biopsy tissue, cytology specimens, serum, and other fluids.[3,4] Frequently, two-dimensional gel electrophoresis (two-dimensional polyacrylamide gel electrophoresis) and mass spectrometry are employed in proteomics to initially fractionate the groups of proteins in a specimen. Proteins in a given sample fraction may later be identified using fingerprinting or sequence tag techniques.[3,4]

Metabolomics

Metabolomics (also termed *metabonomics*) involves study of metabolic profiles by investigating the compounds in a process and the characterization and quantification of small organic molecules in either circulatory or cell tissue systems.[7,87–90] Nuclear magnetic resonance and mass spectrometry are techniques frequently used in metabolomics.[7,87–90]

DNA Microarrays

DNA microarrays are employed to simultaneously screen for the presence or expression of large numbers of genes.[91–98] Suppression subtractive hybridization selectively amplifies target cDNA fragments (differentially expressed genes) and suppresses nontarget DNA.[99,100] Serial analysis of gene expression is used for global analysis of gene expression and provides a comprehensive qualitative and quantitative expression profiles of virtually every gene in a cell population or tissue, and SAGE libraries from different cells and tissues have been created.[101–103]

Conclusion

Rapid medical advances in the laboratory have not necessarily translated into rapid treatment advances. However, continuing omics research to understand the conceptual framework of disease—including disease progression and treatment response—along with improved and more efficient bioinformatics tools to analyze the great amounts of data originating from omics investigations may permit future diagnostic, prognostic, and therapeutic benefit to arise from these technologies.[6]

References

1. Biron DG, Brun C, Lefevre T, et al. The pitfalls of proteonomics experiments without the correct use of bioinformatics tools. Proteomics 2006; Sept 22; [Epub ahead of print].
2. McKusick VA. Genomics: structure and functional studies of genomes. Genomics 1997; 45:244–249.
3. Palagi PM, Hernandez P, Walther D, Appel RD. Proteome informatics I: bioinformatics tools for processing experimental data. Proteomics 2006; Sept. 22; [Epub ahead of print].
4. Lisacek F, Cohen-Boulakia S, Appel RD. Proteome informatics II: bioinformatics for comparative proteomics. Proteomics 2006; Sept. 22; [Epub ahead of print].
5. Maojo V, Martin-Sanchez F. Bioinformatics: towards new directions for public health. Methods Inf Med 2004; 43: 208–214.
6. Bilello JA. The agony and ecstasy of "OMIC" technologies in drug development. Curr Mol Med 2005; 5:39–52.
7. Morel NM, Holland JM, van der Greef P, et al. Primer on medial genomics part XIV: introduction to systems biology—a new approach to understanding disease and treatment. Mayo Clin Proc 2004; 79:651–658.
8. Provart NJ, McCourt P. Systems approaches to understanding cell signaling and gene regulation. Curr Opin Plant Biol 2004; 7:605–609.
9. Wheelock AM, Goto S. Effects of post-electrophoretic analysis on variance in gel-based proteomics. Expert Rev Proteomics 2006; 3:129–142.
10. Debouck C, Metcalf B. The impact of genomics on drug discovery. Annu Rev Pharmacol Toxicol 2000; 40:193–207.
11. Ghosh D. High throughput and global approaches to gene expression. Comb Chem High Throughput Screen 2000; 3:411–420.
12. Hanke J. Genomics and new technologies as catalysts for change in the drug discovery paradigm. J Law Med Ethics 2000; 28(4 Suppl):15–22.
13. Harris T. Genetics, genomics, and drug discovery. Med Res Rev 2000; 20:203–211.
14. Rudert F. Genomics and proteomics tools for the clinic. Curr Opin Mol Ther 2000; 2:633–642.
15. Merrick BA, Bruno ME. Genomic and proteomic profiling for biomarkers and signature profiles of toxicity. Curr Opin Mol Ther 2004; 6:600–607.