Journal of Biotechnology and Bioengineering
Volume 4, Issue 4, 2020, PP 15-17
ISSN 2637-5362

Multi-Omics Data Integration and Systems Approaches in Biology

Prabir K. Mandal[1]*, Noor K. Islam[2] and Anita Mandal[3]

[1]Department of Mathematics & Sciences,

[2]Department of Social & Behavioral Sciences,

[3]Department of General Studies, Edward Waters College, 1658 Kings Road, Jacksonville, FL 32209, USA

***Corresponding Author:** Prabir K. Mandal, Ph.D., Professor& Chair, Department of Mathematics & Sciences, Edward Waters College, Jacksonville, Florida 32209, USA*

ABSTRACT

Omics aims at the collective characterization and quantification of pools of biological molecules that translate into the structure, function and dynamics of an organism. The exponential advances in the technologies and informatics used to generate and process large biological data sets (omics data) are promoting a critical paradigm shift in the study of biomedical study. The mechanism responsible for generating omics information have seen considerable progress in recent years. The advent of new technologies is capable of transforming the level of information and have placed the integration of multiple omics data sets within the reach of scientists. We are witnessing the rise of inter-disciplinary data integration strategies to support a better understanding of biological systems. After completion of genome sequencing, most of the biological questions remained unanswered.

Keywords: *Genomics, metabolomics, omics, proteomics, sociomics, transcriptomics*

INTRODUCTION

Technologies that measure some characteristic of a large family of cellular molecules, such as genes, proteins or small metabolites have been named by appending the suffix “omics”. Omics refers to the collective technologies used to explore the roles, relationships and actions of various types of molecules that make up the cells of an organism. Omics technologies have a broad range of applications and provide the tools needed to look at the differences in DNA, RNA, proteins and other cellular molecules between species and among individuals of a species. These technologies include genomics (study of genes and their function), proteomics (study of proteins), metabolomics (study of molecules involved in cellular metabolism), transcriptomics (study of mRNA), glycomics (study of cellular carbohydrates), lipidomics (study of cellular lipids) in a non-targeted and non-biased manner. This can also be referred to as high-dimensional biology and the integration of these techniques is called systems biology (Kell 2007; Westerhoff and Palsson 2004). Systems biology is an integrative discipline connecting the molecular components within a single biological scale and also among different scales to physiological functions and organismal phenotypes through quantitative reasoning, computational models and high-throughput experimental technologies (Tavassoly et al 2018). When a gene is expressed it results in the production of a messenger RNA and ultimately a particular protein. The DNA sequences of a gene that code for a protein are called exons, and they are interspersed with DNA called introns, which do not code for proteins. Proteins are key players in cellular function. Unlike the static nature of cell’s genes, proteins are constantly changing to meet the needs of the cell. Each gene is a linear stretch of DNA, which is transcribed into messenger RNA (transcription) and then translated by ribosomes into amino acid chains that make up protein (translation). A single-omics technique (e.g., transcriptomics) detects biomolecules of one type, but the integration of multiple omics data sets promises a substantial improvement in detecting biomolecules of multiple types (Canzler et al 2020).The use of multiple omics techniques is becoming increasingly popular in all facets of biology. Scientists and clinicians can now begin to attempt investigation of any

individual dysregulations occurring within the genomic, transcriptomic, miRnomic, proteomic and metabolomic levels thanks to advancing wet-lab technologies such as mass spectrometry, quantitative polymerase chain reaction (qPCR) and next generation sequencing, and detailed bioinformatics suites. The interplay between the wet and dry lab with specific clinical expertise not only is a main current component of translational medicine, but also is enabled by systems medicine (Ayers and Day 2015).

Techniques Currently in Use

Genomics is the study of organism's whole genomes. MicroRNA (miRNA) is a recently discovered class of small non-coding RNAs. Cells use miRNA to regulate amount of protein synthesized by a gene by the mechanisms of translational inhibition and mRNA destabilization. Genomics and proteomics research has been advanced through the development of experimental techniques that increase throughput, such as microarrays. High Throughput Screening consists of assays developed to produce and analyze many individual data points or results in one experiment. The human (*Homo sapiens*) genome contains 3.2 billion bases (Baltimore 2001) and an estimated 30,000-40,000 protein-coding genes. These make up the coding regions (1-2% of the entire genome), while the remaining 98-99% (non-coding regions) holds structural and functional relevance (International Human Genome Sequencing Consortium, 2004, Venter et al 2015). Traditionally, genes have been analyzed individually but microarray technology has advanced substantially in recent years. DNA microarrays measure differences in DNA sequences between individuals and the expression of thousands of genes can be analyzed simultaneously (Horgan and Kenny 2011).The epigenome is the supporting structure of genome, including protein and RNA binders, alternative DNA structures and chemical modifications on DNA. The proteomics involves the applications of technologies for the identification and quantification of overall proteins present content of a cell, tissue or an organism. Unlike the genome, which is fixed for most cells, the proteome is dynamic and proteomics is one of the most significant methodology to comprehend the gene function. The transcriptome is the total complement of ribonucleic acid (RNA) transcripts in a cell and consists of coding (1-4% messenger) and non-coding (>95% ribosomal, transfer, small nuclear, small interfering, micro and long-non-coding) RNAs (Berg and Stryer 2002, Mattick and Makunin 2006). Metabolomics is an emerging omics science involving the comprehensive characterization of metabolites and metabolism in biological systems. has already entered the clinic, with applications in newborn screening. It can be used to determine relative and absolute amounts of sugars, lipids, amino acids, organic acids, nucleotides, steroids, drugs and environmental constituents from a wide variety of sample types including primary cells, cell lines, tissues, biofluids, entire organisms and diverse geo-climatic environments (Misra et al 2019).Lipidomics is the entire complement of cellular lipids, including the modifications made to a particular set of lipids, produced by an organism. Bioinformatics is the application of computational technology concerned with the acquisition, storage, analysis, and dissemination of biological data, most often DNA and amino acid sequences. Bioinformatics uses computer programs for a variety of applications, including determining gene and protein functions, establishing evolutionary relationships, and predicting the three-dimensional shapes of proteins (Bayat 2002).A major application of bioinformatics is in the fields of precision medicine (health care techniques) and preventive medicine (to prevent influenza, cancer, heart disease and diabetes). Precision oncology would greatly benefit from actionable knowledge gained from multi-omics assays (Nicora et al 2020). Bioinformatics is used for transcriptome analysis where mRNA expression levels can be determined. All of life is social, from genes cooperating to form organisms, to animals cooperating to form societies. Omic approaches offer exceptional opportunities to solve major outstanding problems in the study of how sociality evolves (sociomics). First, omics can be used to clarify the extent and form of sociality in natural populations. This is especially useful in species where it is difficult to study social traits in natural populations, such as bacteria and other microbes. Second, omics can be used to examine the consequences of sociality for genome evolution and gene expression (Ghoul et al 2017).

Discussion

Though access to large-scale omics datasets has revolutionized biology and led to the emergence of systems approaches to advance our understanding of biological process, omics strategies still provide many challenges. The technology and the software are still evolving and mapping the human proteome and

metabolome is still ongoing. Systems biology strives to unravel the enormous complexity of a cross-disciplinary environment. Multi-omics data generated for the same set of samples can provide useful insights into the flow of biological information at multiple levels and thus can help in unraveling the mechanisms underlying the biological condition of interest. There are a few publicly available databases that provide multi-omics data sets of patients. The data repository are (1) The Cancer Genome Atlas (TCGA), (2) Clinical Proteomic Tumor Analysis Consortium (CPTAC), (3) International Cancer Genomics Consortium (ICGC), (4) Cancer Cell Line Encyclopedia (CCLE), (5) Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), (6) TARGET and (7) Omics Discovery Index (Subramanian et al 2020).Genomic medicine is an emerging medical discipline that involves using genomic information and is making an impact in the field of oncology, pharmacology and infectious diseases. Metabolomics is increasingly being used to diagnose disease, identify drug targets, customize drug treatments and monitor therapeutic outcomes. Integrative approach using multi-omics data is a powerful strategy to decipher the mechanistic details of the information flow in a cell, and each analysis can generate tera- to peta-byte sized data files on a daily basis. Despite existing challenges, omics methodologies along with bioinformatics to help systemic study of disease. Systems medicine will prove to become one of the mainstays in the way future research will be carried out. There is a need to generate a multi-layer omics focused design using up-to-date omics techniques to further evaluate the contribution of individual omics layers..

REFERENCES

[1] Ayers D and Day PJ (2015) Systems Medicine: The Application of Systems Biology Approaches for Modern Medical Research and Drug Development. MolBiolInternat., 698169

[2] Baltimore D (2001) Our genome unveiled. Nature, 409, 814-816.

[3] Bayat A (2002) Bioinformatics. BMJ, 324 (7344) 1018-1022.

[4] Berg JMTJ and Stryer L (2002) Biochemistry. New York: W H Freeman.

[5] Canzler S, Schor J, Busch W, Schubert K, Rolle-Kampczyk UE, Seitz H, Kamp H, Bergen MV, Buesen R and Hackermuller J (2020) Prospects and challenges of multi-omics data integration in toxicology. Archieves of Toxicololgy, 94, 371-388.

[6] Ghoul M, Andersen SB and West SA (2017) Sociomics: Using Omic Approaches to Understand Social Evolution. Trends in Genet, 33 (6), 408-419.

[7] Horgan RP and Kenny LC (2011) 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. The Obstetrician &Gynaecologist, 13, 189-195.

[8] International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature, 431, 931-945.

[9] Kell DB (2007) The virtual human: towards a global systems biology of multiscale, distributed biochemical network models. Iubmb Life, 59, 689-695.

[10] Mattick JS and Makunin IV (2006) Non-coding RNA. Hum Mol Genet, 15, R17-29.

[11] Misra BB, Langefeld C, Olivier M and Cox LA (2019) Integrated omics: tools, advances and future approaches. J MolEndocrinol, 62 (1), R21-R45.

[12] Nicora G, Vitali F, Dagliati A, Geifman N and Bellazzi R (2020) Integrated Multi-Omics Analyses in Oncology: A review of Machine Learning Methods and Tools. Frontiers in Oncology, 10, 1-11.

[13] Subramanian I, Verma S, Kumar S, Jere A and Anamika K (2020) Multi-omics Data Integration, Interpretation, and Its Application. Bioinformatics and Biology Insights, 14, 1-24.

[14] Tavassoly I, Goldfarb J and Iyengar R (2018) Systems biology primer: the basic methods and approaches. Essays Biochem, 62 (4), 487-500.

[15] Venter JC, Smith HO and Adams MD (2015) The sequence of the human genome. Clin Chem, 61, 1207-1208.

[16] Westerhoff HV and Palsson BO (2004) The evolution of molecular biology into systems biology. Nature Biotechnol, 22, 1249-1252.

Citation: *Prabir K. Mandal et.al, "Multi-Omics Data Integration and Systems Approaches in Biology", Journal of Biotechnology and Bioengineering, 4(4), 2020, pp 15-17.*