Chapter 14
Transcriptomics to Metabolomics:
A Network Perspective for Big Data

Ankush Bansal
Jaypee University of Information Technology, India

Pulkit Anupam Srivastava
Jaypee University of Information Technology, India

ABSTRACT

A lot of omics data is generated in a recent decade which flooded the internet with transcriptomic, genomics, proteomics and metabolomics data. A number of software, tools, and web-servers have developed to analyze the big data omics. This review integrates the various methods that have been employed over the years to interpret the gene regulatory and metabolic networks. It illustrates random networks, scale-free networks, small world network, bipartite networks and other topological analysis which fits in biological networks. Transcriptome to metabolome network is of interest because of key enzymes identification and regulatory hub genes prediction. It also provides an insight into the understanding of omics technologies, generation of data and impact of in-silico analysis on the scientific community.

INTRODUCTION

The appearance of majority of disease processes cannot be explained by modification in one gene, but involves the involvement of synchronized genes associated with same function using various network analyses. Since the advent of molecular biology, extensive progress has been made in the quest to grasp the mechanisms that lie beneath human disease. Consequently, to be unbeaten in drug development, scientific community must shift its focus from individual genes that carry disease-associated mutations towards a numerous gene association perspective of disease mechanisms.

No doubt that genomic sequencing resulted in high advancement in prognosis and personalized medicine but still leaves the elementary questions pertaining to genotype-phenotype relationships unresolved

(Vidal, Cusick, and Barabási, 2011). The contributory changes that connect genotype-phenotype remain usually unidentified, especially for complex attributes like gene loci and cancer related mutations. Though, when mutations are identified, it's still unclear how a perturbation function works and correlates with splicing, post translational modifications and other autocrine and paracrine cellular communications. Till date there is no justified answer discovered for functional association which merge the knowledge and construct a bridge between genotype to phenotype. To connect the dots of the unorganized information at genomic, proteomic level and find out their association with function variation we need to opt interdisciplinary approaches. Even after genome sequencing, there was need to align data with respect to condition variation. Also, analysis of multiple genes can be done at same point using differential gene expression analysis, which ultimately can give limelight to function prediction, co-expressed gene identification and functional validation (Zhu et al. 2012). Next generation sequencing based RNA-seq transcriptomic analysis answered almost all the questions as it was the need of decade.

During the past decade, tremendous development has been made in terms of speed, read length, and throughput, along with a sharp reduction in per-base cost. Together, these advances democratized NGS and paved the path for the development of a outsized number of novel NGS applications in basic science as well as in translational research areas such as clinical diagnostics, agrigenomics, and forensic science. The evolution of NGS and significant improvements in sequencing technologies and library preparation protocols benefited the way as expected to resolve from enormous bio-fields. But, transcriptome data has its own disadvantages like abnormality in transcripts alignment, annotation and abundance values in terms of TPM and FPKM, which turns the scientific community to think about alternative to resolve such big data.

As quantity of data increases, the variable associated with function also seems impossible to calculate through manual curation. A quantitative description of a complex system is intrinsically restricted by our capability to estimate the system's interior state from experimentally accessible outputs. Although the simultaneous measurement of all internal variables, such as all metabolite concentrations in a cell, offers a complete description of a system's state, in practice however experimental access is limited to only a subset of variables, or sensors. A system is called observable if its interior state could be reconstructed from its output. Here, we adopted a graphical approach derived from the dynamical laws that govern a system to determine the subset of variables that are obligatory to reconstruct the full interior state of a complex system. Before, scientist has applied this approach to biochemical reaction systems, and found that the identified sensors are not only essential but also enough for observability. The developed approach can also identify the optimal sensors for target or partial observability, helping us reconstruct selected state variables from appropriately chosen outputs, a pre-requirement for optimal biomarker design. Observability plays key role in complex systems; these results offer avenues to systematically discover the dynamics of a extensive range of natural, technological and socioeconomic systems.

Merely, combining big data complexity with network theory can assist in study of genotype-phenotype characteristics and identification of hub nodes may predict the impact of gene silencing method inclusion with disappearance with disease phenotype. Various studies which reflect the impact of random networks reflected that biological networks do not reciprocate the property of randomness. Biological network follow power law and hence regulation of biological and chemical pathways have patterns. Although, various software and tools are available in market, but optimization of these tools to resolve complexity of biological problems remains problematic till date.

Basic structural and spectral parametric analysis of complex networks gives an insight towards the study of network model behavior using metabolic networks, transcription regulatory networks, protein-

Related Content

Algae as Superfood

Shital Uddhav Giri, Namdev Gopal Krishna Hadapad, Aditya Akhadeand Parth Bhilare (2023). *Biomass and Bioenergy Solutions for Climate Change Mitigation and Sustainability (pp. 129-147).*

www.irma-international.org/chapter/algae-as-superfood/314361

Genetically Engineered Microorganisms for Bioremediation Processes: GEMs for Bioremediaton

Stephen Rathinaraj Benjamin, Fabio de Limaand Ashok K. Rathoure (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications (pp. 1607-1634).*

www.irma-international.org/chapter/genetically-engineered-microorganisms-for-bioremediation-processes/228686

What Influences the Growth of Canadian Biotechnology Firms?

Catherine Beaudryand Joël Levasseur (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications (pp. 1795-1825).*

www.irma-international.org/chapter/what-influences-the-growth-of-canadian-biotechnology-firms/228694

De Novo Motif Prediction Using the Fireworks Algorithm

Andrei Lihuand tefan Holban (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications (pp. 1069-1085).*

www.irma-international.org/chapter/de-novo-motif-prediction-using-the-fireworks-algorithm/228658

Paratexts and Documentary Practices: Text Mining Authorship and Acknowledgment from a Bioinformatics Corpus

Nicholas M. Weberand Andrea K. Thomer (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications (pp. 597-624).*

www.irma-international.org/chapter/paratexts-and-documentary-practices/228641