

Network biology and applications

Neeru Redhu¹ and Zoozeal Thakur²

¹*Department of Molecular Biology, Biotechnology & Bioinformatics, College of Basic Sciences & Humanities, CCS Haryana Agricultural University, Hisar, India,* ²*Bacteriology Lab, National Research Centre on Equines, Hisar, India*

23.1 Introduction to biological networks

Biological networks are a mathematical depiction of an organization wherein subunits are linked into a whole and are found in physiological, ecological, and evolutionary studies. Nodes and edges are the basic units of biological networks wherein nodes correspond to the type of information being depicted in the network, which can be genes, proteins, metabolites, neuron, species, etc., and edges represent a pattern of interaction, for instance biochemical reaction, synapse, metabolic directionality, regulatory relationship, and energy transactions among others. Major types of biological network that have been widely investigated in detail include ecological network, gene regulatory network (GRN), gene coexpression network (GCN), protein–protein interaction network (PPIN), metabolic network, and cellular signaling networks (Sonawane, Weiss, Glass, & Sharma, 2019). Biological networks are dynamic, conserved, modular, and organized in a hierarchical manner (Wang & Zhang, 2007). They are robust to small perturbations but exhibit vulnerability at essential nodes (Azevedo & Moreira-Filho, 2015). They are constructed by using one or a combination of methods, which includes manual curation of scientific literature, high-throughput datasets, and computational predictions.

With the advent of high-throughput sequencing technologies along with advancement in bioinformatics tools and techniques, the spotlight is shifted from individual entities, such as genes or proteins, to large-scale interaction networks often referred to as -omes, that is, biome, interactome, genome, and proteome (D'Argenio, 2018). Network biology amalgamates biological omics data (genomics, proteomics, metabolomics) and biological interactome [protein–protein interactions (PPIs), gene–gene associations] and employs statistics, mathematical modeling as well as graph theory to decode significant patterns underlying the biological systems (Zhang & Itan, 2019). The amalgamation of both experimental and bioinformatics techniques has led to the generation and wide access to large-scale data sets that have formed the basis of the reconstruction of various biological networks for instance metabolic networks, regulatory networks, signaling networks, and PPINs (Covert, Knight, Reed, Herrgard, & Palsson, 2004). Network biology has augmented our understanding of intricate biological systems by decoding biological mechanisms and the pathogenesis of diseases and aided in the therapeutic intervention. In this chapter, we briefly discuss the general features of various kinds of biological networks and their network topological properties. In addition, experimental methodologies, biological databases, computational tools, applications, challenges, and future perspectives related to biological networks are also discussed.

23.2 Biological networks properties

In this section, topological metrics and properties of biological networks that are frequently probed during the analysis of biological networks to make predictions are discussed.

23.2.1 Path, average path length, and diameter

The path is defined as the sequence of nodes required to traverse from the source node to the target node, consequently decoding the structure of a graph. Whereas, the minimum distance required to reach two nodes is called the shortest path in a graph. On the other hand, the average path length is the mean of the distance among all the nodes of a graph and an indicator of network navigability (Perez & Germon, 2016). The average distance between nodes is dependent on the network size; the more the number of vertices the more is the distance between them. Interestingly, the average

path length is generally small for all empirical networks despite their huge size. The average path length is observed to be approximately 3 in a variety of analyzed metabolic networks and is independent of the organism type. On the other hand, the diameter of the graph is an overall indicator of the compactness of the network and describes the maximum distance linking a pair of nodes within a network. Even though real-world networks have short diameters and exhibit small world property and that may be considered advantageous for rapid information flow, several biological networks exhibit large diameters than the random networks. The relatively large diameter of biological networks can be owed to their modular nature (Klein, Marino, Sagot, Milreu, & Brilli, 2012).

23.2.2 Degree aka connectivity

Degree or valency aka connectivity refers to the number of linkages a node has with other nodes within the network. In the case of a directed network, there are two degrees, in degree and out degree, which refer to incoming and outgoing edges, respectively (Pavlopoulos et al., 2011).

23.2.3 Scale free

Networks wherein fraction of vertices that have degree k along with that obeys the power-law distribution $k^{-\alpha}$, where $\alpha > 1$ is considered to be scale free. Some versions of this “scale-free hypothesis” even require $2 < \alpha < 3$. Scale-free property is exhibited by most of the real-world networks, which include the worldwide web, power grid, collaboration network of film actors, food webs, neural, PPIN, and metabolic networks (Broido & Clauset, 2019).

23.2.4 Small world network

Small world networks are defined by high clustering coefficient and relatively small average path length, which is akin to a random network comprising of the same count of nodes and edges (Humphries & Gurney, 2008). These features provide regional specialization and efficient information transfer within the network. The high clustering coefficient depicts local interconnections and cliquishness. In reference to biological networks, these depict functional modules, such as a protein set involved in a common function, a gene set regulating a common biological process, or metabolic pathways’ interconnections. Small world network has found application in a variety of fields, which includes sociology, earth sciences, computing, as well as neural networks in the brain. Examples of small world networks are social networks, the internet, and biochemical pathways (Zhang et al., 2017).

23.2.5 Date and party hub

Han et al. (2004) first introduced the concept of “date” and “party” hubs in yeast interactome while amalgamating transcriptional profiling data with PPIN information. The concept is widely used in PPINs. Date hub interacts with various partners at different times and spaces and forms functional modules. They show limited coexpression with their partners. While party hubs show positive correlation expression with their partners and bind with them concurrently and generally work inside the functional modules (Agarwal, Deane, Porter, & Jones, 2010).

23.2.6 Network motifs

Network motifs are overrepresented and statistically significant patterns that are found in higher frequency in comparison to randomized networks. They were first detected and introduced in the GRN of *Escherichia coli* in 2002 by Shen-Orr, Milo, Mangan, and Alon (2002). Since similar network motifs have been observed in organisms ranging from bacteria to humans, they are considered as the significant unit with potential functional properties in the networks. The relative abundance of network motifs is positively correlated with the robustness of the network in response to little alternations. They are observed in both biological and nonbiological networks, for instance Internet, social network, signaling network, neuronal network, and ecological network among others (Wong, Baur, Quader, & Huang, 2012).

23.3 Types of biological networks

In this section, general features of different types of biological networks are discussed briefly.

23.3.1 Ecological networks

The pace at which the earth is changing has been never witnessed before (Sueur, Krause, & Farina, 2019). Species extinction is occurring at a rapid rate constantly, even greater than the five mass extinction events obtained from fossil records in earth history (Rothman, 2017). However, human well being is highly dependent on the services offered by ecosystems; thus global changes occurring in nature, such as climate change and species extinction, require our immediate concern. An ecological network is a depiction of the interactions of an ecosystem, wherein nodes are represented as species and edges as pairwise interaction (Delmas et al., 2019). The nature of interaction within the ecological network can either be trophic or symbiotic.

Mathematical network models simplify the intricacies of real-world networks by illustrating and examining the principles of ecology (Motta & Pappalardo, 2013). The ecological networks are used for depiction and comparison of real ecosystem structure, while network models examine the effect of network properties on network structure for instance ecosystem stability. Ecological network analysis (ENA) is a methodology employed for the investigation of organization, function, and the evolution of ecosystems and amalgamates modeling and analysis. Moreover, to track the movement of thermodynamically conserved energy or matter through the system, ENA is employed in network models (Borrett, Sheble, Moody, & Anway, 2018). In addition, ENA techniques have been employed for food web structure characterization, ecosystem status evaluation, tracing of biogeochemical cycles, and other socio-ecological systems.

The most important goal of ecology is to decipher the correlation between the complexity of ecosystems and their stability for instance the ability of ecosystems to returning to their functioning state after encountering perturbations, such as habitat fragmentation, climate change, and pollution among others (Malhi et al., 2020). Ecological networks are utilized to examine the correlation of ecosystem stability with various network properties. Earlier it was believed that complexity leads to a reduction in stability by the effect of perturbations for instance species loss or invasion by exotic species. Contrarily, other features have been detected that augment ecosystem stability and lessen the effect of indirect effects. For instance, the strength of interaction may decline with the increase in the number of linkages among species, curtailing the consequence of any interruption. Notably, extinction cascades are less likely to be found in various compartments of a network, as species loss effects are restricted to the original compartment.

To catalog life history, biotic interaction data, and other significant information for easy availability of data to researchers around the globe, few specialized databases have been constructed, such as Mangal-Ecological interaction database, IWDB interaction web database, and *GLOBI database* (Crea, Ali, & Rader, 2016). The ENA uncovers mechanisms underlying the functioning, stability as well as the resilience of ecosystems (Landi, Minoarivelo, Brännström, Hui, & Dieckmann, 2018). Moreover, investigation of Ecological networks is significant for biodiversity conservation, management of ecosystem services management, and prevention of ecosystem degradation (Meyer, Leempoel, Losapio, & Hadly, 2020).

23.3.2 Gene (genetic) regulatory network

Gene expression is the process of conversion of information stored in genes into functional gene products which can be mRNA or proteins via transcription and translation. Transcription factors that modulate transcription by acting as activators or inhibitors are also encoded by genes and are also regulated, thus forming a complex regulatory network (Mira, Teixeira, & Sá-Correia, 2012). Various steps of gene expression can be regulated to counteract extracellular cues starting from transcription to posttranslational modification (Kang & Han, 2011). Regulation of gene expression is the foundation of many cellular processes, including differentiation, cell signaling, development, morphogenesis, and adaptability of any organism. A GRN depicts interactions among molecular entities, including DNA, RNA, protein, and their complexes with one another as well as with other molecules in the cell to control the cellular gene expression levels of mRNA and proteins and that subsequently determines their function. Nodes in the GRNs represent either genes or their regulators, while edges depict either physical or regulatory relationships between the nodes (Fig. 23.1) (MacNeil & Walhout, 2011).

GRNs are directional as typically a regulator control the genes and in general vice versa is not observed. The bipartite nature of GRNs is due to the presence of two kinds of nodes, that is, genes and transcription factors which result in two types of degrees, that is, in degree and out degree. The *in degree* describes the number of transcription factor and gene interactions and obeys power law distribution, while *out degree* describes the number of genes that exhibits an association with a regulator and follows an exponential distribution. Some of the GRN nodes are highly linked in the network and are referred to as either “TF hubs” or “gene hubs” in accordance with their nature of the interaction, while the majority of nodes show a low degree in a regulated nested hierarchy (Zamal & Ruths, 2012). Notably, GRNs exhibit

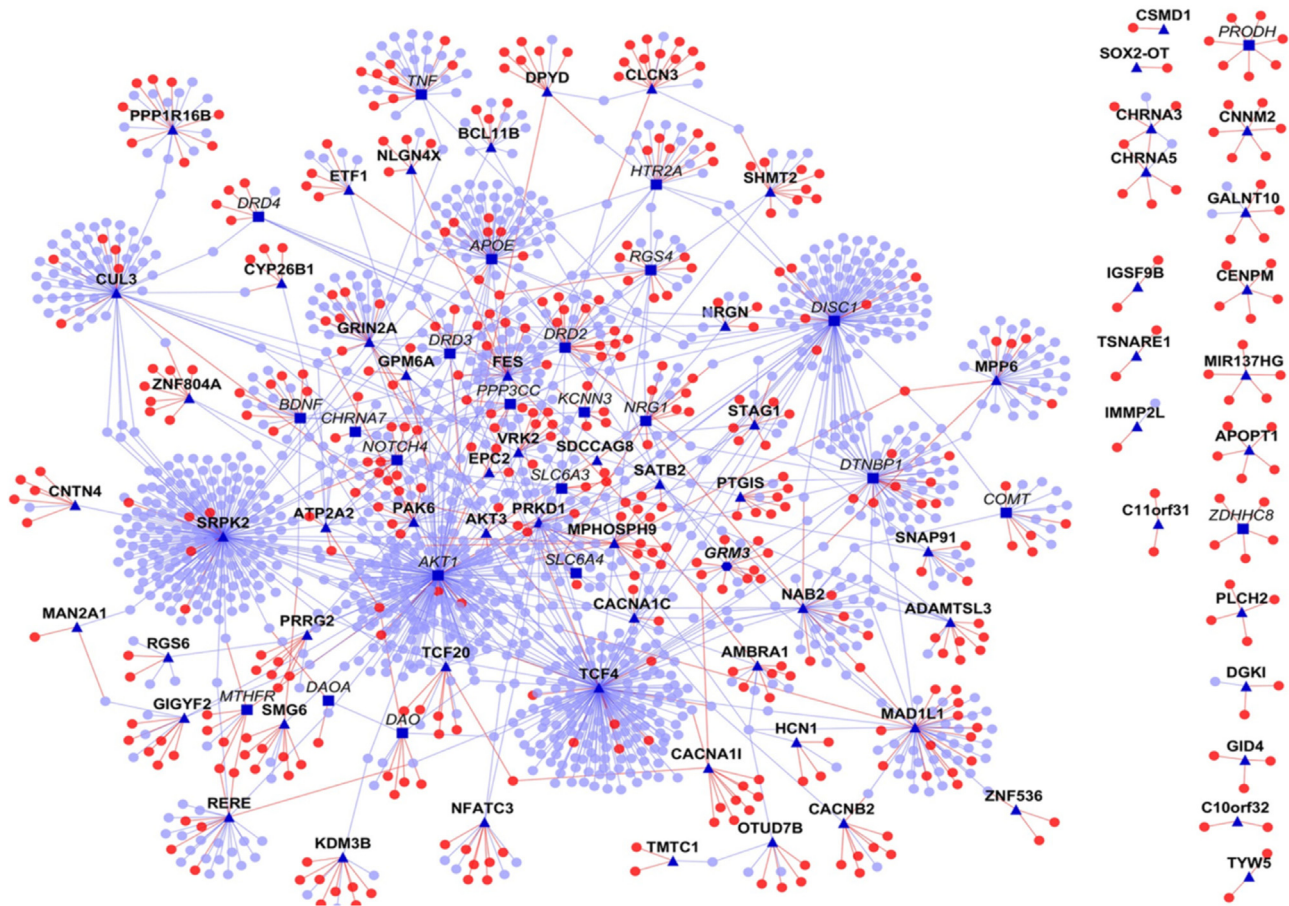


FIGURE 23.1 Figure depicts gene-transcription factor regulatory network of silicosis obtained by the integration of publicly available microarray data. Yellow round node, green-colored triangle node, and a pink-colored rectangular node represent a differential expression of genes, transcription factors, and transcription factor that function both as a regulator as well as are regulated, respectively. Adapted from Choudhari, J. K., Verma, M. K., Choubey, J., & Sahariah, B. P. (2021). Investigation of MicroRNA and transcription factor mediated regulatory network for silicosis using systems biology approach. *Scientific Reports*, 11(1), 1265.

robustness, stochasticity, redundancy, adaptability, and hierarchy in an organization. They also exhibit scale-free behavior and small world topology. Another significant feature of GRNs is the presence of certain repetitive subnetworks called network motifs. They are repetitive topological patterns for instance feed-forward loops (comprising of three nodes). This motif is the most widely found in the GRNs from *E. coli* to humans (MacNeil & Walhout, 2011).

Several experimental, mathematical, and in silico methods have been introduced for the detection, visualization, and analysis of interactions between genes and TFs. Experimental techniques for the detection of TF–gene interaction are chromatin immunoprecipitation (ChIP), DamID, and yeast one-hybrid (Y1H) system while regulatory relationships are detected by correlating transcriptional profiles of genes and putative regulators. The combination of both physical and regulatory relationships is required to generate whole and precise GRN's (MacNeil & Walhout, 2011; Sparks et al., 2016). Several database resources that have cataloged the gene expression data and facilitated the biological investigation of organisms, such as a mouse and bacteria, including Gene Expression Omnibus (GEO), MouseGene Expression Database, CollecTF, COLOMBOS, and Many Microbe Microarrays Database (M3D), are available for researchers. GRNs aid in uncovering differential gene expression, disease pathogenesis, and identification of biomarkers for disease diagnosis (Singh, Ramsey, Filtz, & Kioussi, 2018).

23.3.3 Protein–protein interaction network

PPIs play a significant role in almost every biological process, including growth as well as development, and their disruption leads to a variety of diseases (Waiho, Afiqah-Aleng, Iryani, & Fazhan, 2021). So, the identification of PPIs is

crucial for decoding the functioning and organization of cells as well as deciphering of underlying mechanisms of diseases (Kuzmanov & Emili, 2013). PPIs are classified as transient or stable interaction, obligate or nonobligate interaction, as well as homo- or heterooligomer interaction on the basis of persistence, stability, and interaction surface respectively. Transient interactions, such as protein kinases and nuclear pore importins, are the ones that drive the dynamic element of the interactome, while permanent interactions form stable complexes (Rao, Srinivas, Sujini, & Kumar, 2014).

PPIN is a mathematical depiction of the physical binding of proteins within the cell wherein proteins represent nodes and interactions are depicted by undirected edges (Rao et al., 2014). PPINs are usually depicted by undirected graphs and exhibit small world as well as scale free properties. They are the most widely analyzed networks among biological networks. PPINs have been constructed for several organisms, such as bacteriophages, bacteria, yeast, plants, and animals. Experimental techniques for PPI detection include yeast two-hybrid (Y2H) system for binary interactions, TAP (tandem affinity purification), and affinity purification among others (Rao et al., 2014). Owing to the increase of genome-scale PPI information of a number of species, a variety of in silico techniques that focus on the detection of protein complexes from PPINs have been developed. The current computational methodologies are categorized into three classes, that is, (1) cluster quality based, (2) node affinity based, and (3) ensemble clustering (Wu, Liao, & Liu, 2020).

Several databases for the archival of PPI information have been developed, such as STRING (Search Tool for the Retrieval of Interacting Genes/Protein), BioGRID (The Biological General Repository for Interaction Datasets), APID (Agile Protein Interaction DataAnalyze), MINT (Molecular INTeraction database), BIND (Biomolecular Interaction Network Database), Database of Interacting Proteins, and HPRD (Human Protein Reference Database) among others. Furthermore, several software applications have been introduced for the visualization and analysis of PPINs, such as ProViz, PINV, Cytoscape, PIANA, and others (Su, Morris, Demchak, & Bader, 2014). PPINs have been used for functional prediction, and evolutionary analysis as well as drug development for better disease outcomes (Yan, Zhang, Shen, Liang, & Hu, 2018). In addition, investigation of PPIN has revealed underlying disease mechanisms of diabetes, cancer, and various other genetic and neurodegenerative disorders (Li, Ivanov, et al., 2017; Li, Wang, & Gou, 2017). Moreover, PPIN analysis has provided insight into the understanding of pathogenesis, host–microbe interaction, and pathogen coinfection systems (Fig. 23.2).

23.3.4 Metabolic networks

Metabolism is the series of all chemical reactions catalyzed by enzymes that are required for the maintenance of living beings (Deberardinis & Thompson, 2012). A metabolic network is the absolute representation of interconnected physical as well as metabolic processes of a cell that forms a dynamic circuit and establishes its biochemical and physiological properties. These networks consist of the chemical reactions and metabolic pathways, along the crucial regulatory interactions that lead to these reactions (He, Murabito, & Westerhoff, 2016). Metabolic pathways are scale free and exhibit small world property. They are powerful tools for modeling metabolism as well as analysis (Angione, 2019).

Crosstalk occurring between different subparts of metabolic networks is highly significant. The molecular entities that arbitrate crosstalk conclude the overall robustness of the metabolic networks. Metabolic crosstalk characterization leads to a deeper understanding of the metabolic reactions occurring in the cell, and the information obtained can be extensively applied in basic as well as applied sciences (de Anda-Jáuregui, Guo, McGregor, Feldman, & Hur, 2019). The metabolic network can be modeled by several tools that include extreme pathways, elementary mode, minimal metabolic behaviors, and flux balance analysis. Owing to the technological advancements, the rapid increase in the biochemical reaction network of organisms ranging from *E. coli* to human beings has been observed (Pornputtpong, Nookaew, & Nielsen, 2015). A number of databases that catalog biological networks and are available online include Kyoto Encyclopedia of Genes and Genomes (KEGG), EcoCyc, BioCyc, and metaTIGER (Shameer et al., 2015). Reconstruction of metabolic networks and their simulation unravels molecular mechanisms of a particular organism as well as allows correlation of the genome with molecular physiology (Mei, Xu, Mei, Liu, & Wu, 2016). In addition, metabolic networks has been utilized to detect comorbidity patterns for instance obesity and diabetes can increase the risk for the presence of other diseases (Lee et al., 2008).

23.3.5 Cellular signaling network

Cell signaling involves the transfer of information between molecules of cell in response to external and internal cues. Disruption of these cellular pathways contributes to the pathogenesis of many human diseases and also affects the

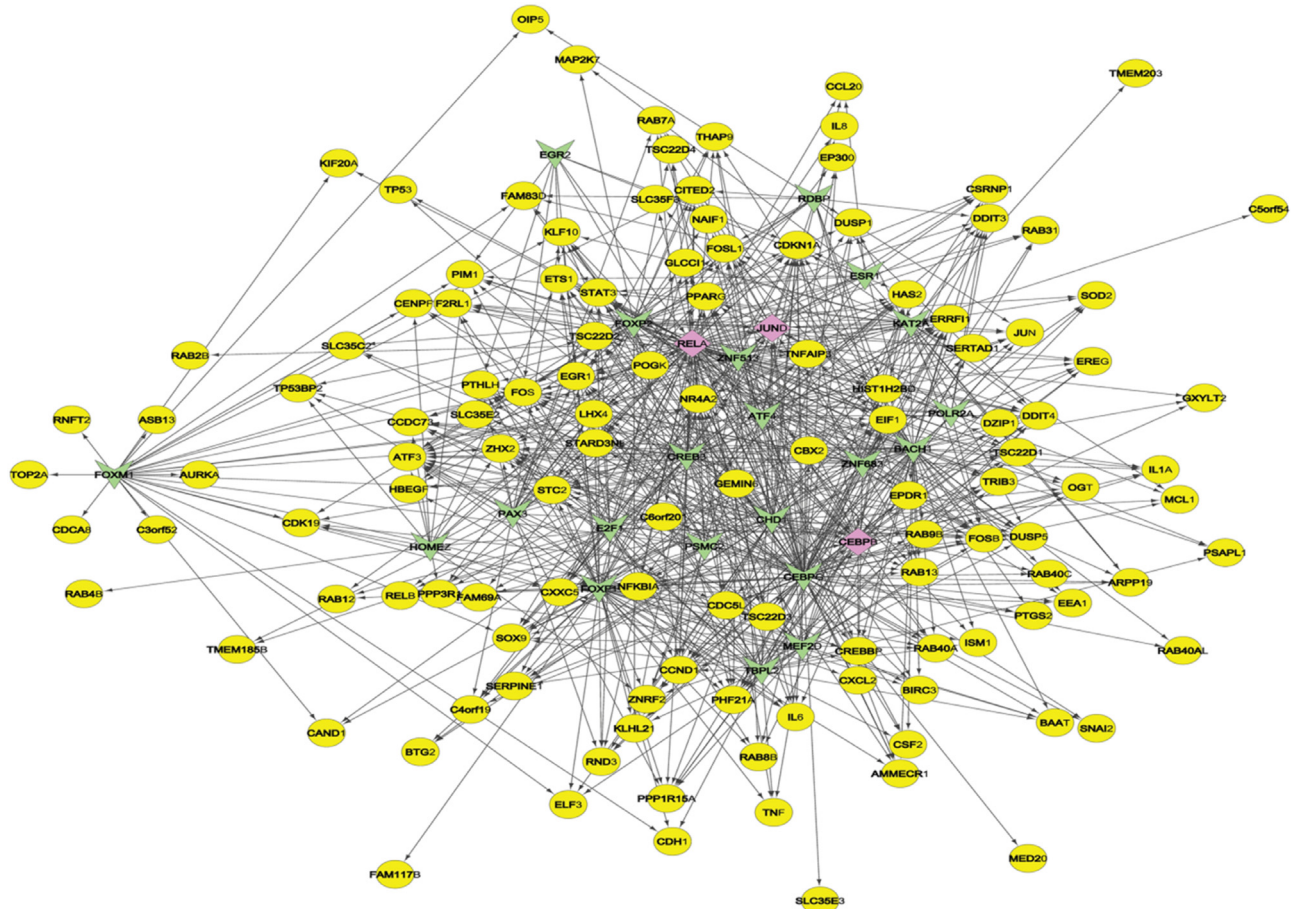


FIGURE 23.2 Protein–protein interaction network of Schizophrenia generated by Ganapathiraju et al. (2016) wherein genome-wide association study and already reported Schizophrenia genes were employed to generate an interactome that harbored 504 protein–protein interactions. Genes are depicted by nodes and protein–protein interaction by edges. *Dark blue, red, and blue nodes represent Schizophrenia-linked genes, novel interactors, and known interactors, respectively. Blue edges are reported interactions, while red edges are the novel interactions.* Genome-wide association study genes were depicted in *bold* and already reported Schizophrenia-associated genes in *italics* and one common gene is shown as both *italicized and bold*. Adapted from Ganapathiraju, M. K., Thahir, M., Handen, A., Sarkar, S. N., Sweet, R. A., Nimgaonkar, V. L., et al. (2016). *Schizophrenia interactome with 504 novel protein-protein interactions*. *Npj Schizophrenia*, 2, 16012.

pathobiology of various infectious agents, such as bacteria, viruses, and fungus. Targeting these cellular pathways is considered to be one of the strategies for improved therapeutic intervention. Thus a meticulous understanding of the molecular basis of signaling pathways is crucial. Cellular signaling networks depict the interaction of different cell signaling pathways that are detected by the amalgamation of both experimental as well as computational techniques (Papin, Hunter, Palsson, & Subramaniam, 2005). These networks receive and broadcast signals as well as process information. To deduce steps involved in the processing of information and input–output relationships in complex networks, computational models are used.

Signaling networks are understood by the combined usage of both graph theory and dynamical models. The application of graph theory facilitates understanding of organization and processing of information within a signaling network, whereas dynamical modeling can decode spatiotemporal changes in response to stimuli. Cell signaling networks are characterized by reaction hubs and hierarchical feedback modules, which in turn impart robustness and versatility to the networks (Xu & Lan, 2015). The emergent properties of signaling networks are ultrasensitivity, bistability, oscillatory behavior, robustness, and noise reduction ability. These characteristic properties provide the network the capability to overlook small or transient signals and augment signals that drive significant cellular processes. Several web-based databases, visualization, and mining tools have been constructed for cataloging and the visualization of cell signaling pathways, including VisANT, TRANSPATH, MiST, LitInspector, PID, and NET-SYNTHESIS among others. Owing to the availability of high-throughput technologies, condition-specific signaling networks can also be investigated.

23.3.6 Gene coexpression network

A GCN is an undirected graph, wherein each node depicts a gene, and a pair of connected nodes represents a significant coexpression relationship. GCNs are biologically significant as coexpressed genes are regulated via the common transcriptional regulatory program (Prieto, Risueño, Fontanillo, & De Las Rivas, 2008). Butte and Kohane in 1999 were the pioneers in introducing the concept of GCN as a relevant network and constructed the first gene expression network. They collected medical laboratory test data (e.g., blood sugar level, serum insulin level, hemoglobin level) regarding 5158 diseased individuals and estimated the Pearson correlation between the outcomes of each pair of tests considered and the pairs with higher correlation than a particular level were shown as connected in their first relevance network. Butte and Kohane (2000) later utilized this strategy and used mutual information as the coexpression measure and gene expression profiles for the construction of the first GCN.

In comparison to GRN, directionality and type of casual relationships among genes are not inferred by GCNs, while, in GRN, a directed edge can depict activation, inhibition, interaction, or transformation (Roy, Bhattacharyya, & Kalita, 2014). A significant feature of GRNs is the presence of modules or highly connected subgraphs that represents clusters of genes that have a similar function or are involved in a common biological process that causes many interactions among themselves (Ficklin & Feltus, 2011). GCNs are usually built by using gene expression patterns obtained by gene expression profiling techniques for instance microarray analysis or RNA-seq technology, wherein genes showing similar expression patterns across various conditions are looked for (Emamjomeh, Saboori Robat, Zahiri, Solouki, & Khosravi, 2017).

Several methods have been developed for the construction of GCNs. All of them essentially follow two basic steps, that is, (1) calculations of coexpression measure and (2) significant threshold selection. The expression value of a gene for different samples or experimental conditions is represented as a vector for the calculation of coexpression measure between a gene pair (Makrodimitris, Reinders, & van Ham, 2020). The most commonly used coexpression measures for the construction of GCNs are Euclidean distance, bivariate correlation, Spearman's rank correlation coefficient, and mutual information. Each of them has their own advantages and disadvantages (Song, Langfelder, & Horvath, 2012). The second crucial step of GCN construction is significant threshold selection, which may be performed by the employment of simple thresholding, Fisher's Z-transformation, clustering coefficient, and/or random matrix theory (McKenzie, Katsyv, Song, Wang, & Zhang, 2016). GCNs have been extensively utilized in the area of single cell sequencing, gene network reverse engineering, and plant biology (Lamere & Li, 2019).

23.4 Experimental methods in network biology

Different types of biological networks are generated from studies involving several experimental techniques. This section describes the popular methods employed in various interaction and network studies.

23.4.1 Microarray

The principle of microarray is hybridization. It was first demonstrated by Grunstein and Hogness in 1975 by creating random and unordered DNA spots, containing cloned *E. coli* plasmid fragments carrying DNA of interest, through lysis of bacterial colonies on nitrocellulose filters. Subsequent hybridization with radiolabeled DNA/RNA probe rapidly screened thousands of *E. coli* colonies (Grunstein & Hogness, 1975). The technique was automated to organize clones onto arrays as well as newer methods of array producing and fluorescent dyes usage for detection were incorporated during the 1990s (Lennon & Lehrach, 1991). Tens of thousands of targets can be studied simultaneously by microarray. Typically, it involves RNA isolation (at least two samples, i.e., reference and experimental) followed by reverse transcription to obtain cDNAs and their labeling with fluorescent dyes (e.g., cy3 or cy5); also known as target preparation. The samples are pooled and transferred onto a microarray chip for hybridization with probes attached to the chip followed by unhybridized cDNA removal by washing, scanning, and detection of differential expression among samples. The technique is used to generate gene expression profiles depicting changes in gene(s) expression caused by the specific treatment or conditions.

The microarray technique can be separated into two steps, that is, array fabrication and target preparation. The array fabrication consists of the synthesis and attachment of both arrays and probes. So, microarray can be classified based on the mode of array preparation (spotted, in situ synthesized, and self-assembled arrays) as well as types of the probe (Miller & Tang, 2009). The spotted arrays use a robotic spotter with multiple pins for DNA spotting on poly-lysine-coated glass slides. It generates very high-density DNA arrays, which are sensitive and cheaper, but the reproducibility of the spotted array is very low.

The in situ synthesized arrays (licensed to Affymetrix), developed by Fodor and coworkers in 1991, are light directed, spatially addressable chemical synthesis by combing photolabile protecting groups with photolithography directly on a solid substrate (Fodor et al., 1991). Another technique of in situ synthesis is inkjet printing technology (licensed to Agilent technologies), developed by Blanchard and team in 1996, and delivers four phosphoramidite bases to prepatterned glass slides with hydrophilic region surrounded by hydrophobic region (Blanchard, Kaiser, & Hood, 1996). Both Affymetrix and Agilent chips are primarily used in expression analysis, genotyping, and sequencing (Miller & Tang, 2009). The self-assembled array (licensed to Illumina), developed by David Walt in 1998, fabrication consists of DNA synthesis on polystyrene beads that are subsequently placed randomly on fiber optic array (Walt, 2000). On the basis of probes utilized in the synthesized arrays, the microarray can also be classified into several categories, such as DNA, protein, carbohydrate, chemical compound, and cellular among others.

23.4.2 Deep mRNA sequencing

Advances in sequencing technologies from 1st generation developed in 1977 by Sanger and group as well as Maxam and Gilbert to recently developed next-generation sequencing (NGS) technologies have proved significant for deciphering the life on earth. The NGS technologies are branded as “High Throughput Sequencing Technologies,” as they require DNA fragmentation and amplification before sequencing that results in the production of millions of reads (Kchouk, Gibrat, & Elloumi, 2017). One major drawback of NGS technology is that a single sequencing experiment might not be sufficient for the genome because of the presence of repetitive regions, lower representation in the amplified libraries, and sequencing errors (Tørresen et al., 2019). To overcome this, the sequencing is performed repeatedly; deep sequencing refers to sequencing a genomic region hundreds to thousands of times for identifying unique reads from every region of a sequence. Similarly, repeated RNA sequencing (transcriptome or mRNA) through NGS is known as deep mRNA sequencing. The sequence and frequency of RNA molecules specific to the cell, tissue, organ, or development stage can be obtained.

RNA-seq analyzes the gene expression patterns encoded within RNA by examining RNA quantity and sequences. The early RNA-seq studies mainly used Sanger sequencing technology, but being low-throughput, expensive, and error-prone in nature, a comprehensive understanding of transcriptome was not achieved. Only with the arrival of NGS, the cell transcriptome can be fully explored (Wang, Gerstein, & Snyder, 2009). Typically, RNA-seq workflow involves RNA isolation, cDNA preparation, cDNA library generation, cDNA sequencing, and RNA-seq data analysis. Subsequent to RNA isolation, their 3' poly-A tail was hybridized poly-T oligonucleotides that are covalently bound to a substrate (e.g., magnetic beads) for mRNA enrichment, selection, and reverse transcription. The NGS technologies produce millions of reads (short nucleotide sequences), which are aligned, assembled, and mapped to a reference genome to produce a whole transcriptome (Kukurba & Montgomery, 2015). After alignment, tools, such as Sailfish, RSEM, MISO, and BitSeq12 among others, can be used to quantify expression levels (Byron, Van Keuren-Jensen, Engelthaler, Carpen, & Craig, 2016).

To sum up, NGS-based RNA-seq investigates entire cellular RNAs including mRNA, rRNA, and tRNA aka transcriptome of the cell. Understanding the transcriptome is the bridge that connects the genome with its protein expression and phenotype. RNA-seq studies have helped the identification of specific genes, their expression levels, their activation, and deactivation leading to deeper insight into the cell biology and changes that may indicate disease (Sonawane et al., 2017). Currently, RNA-seq is predominately used for transcriptional profiling, RNA editing, SNP identification, and differential gene expression analysis.

23.4.3 Exome sequencing

Traditionally, the exome is referred to as the complete set of exons or entire protein coding regions of genes in the genome and its sequencing is known as exome sequencing or whole exome sequencing (WES) (Mahajan & McLellan, 2020). The exome only covers 1%–2% of the genome but could easily identify genetic variants that change protein sequences as well as nonprotein coding elements, such as microRNA and long intergenic ncRNA among others, at a lower cost than whole genome sequencing or NGS. WES has been widely applied in research and diagnostics, as several variants have been linked to genetic and other polygenic diseases, like Alzheimer's linked variant's have been identified. For instance, <2% human genome is exome, but nearly 85% of disease-related variants belong to it, making this method cost-effective compared to genome sequencing (Fan et al., 2020).

This technique involves library preparation (genome fragmentation and adapter ligation), target enrichment (selection of exonic regions or genomic regions of interest), sequencing, and data analysis. Several target-enrichment

strategies have been developed as (1) hybridization-based, (2) transposon-mediated fragmentation, (3) molecular inversion probes (MIPs), and (4) polymerase chain reaction (PCR) based (Kozarewa, Armisen, Gardner, Slatko, & Hendrickson, 2015). Hybridization-based approaches involve the ligation of adapters containing “index” or barcode bases that are unique to the sample. Capture protocols, a Hybridization-based strategy, involve hybridization of single-stranded probes complementary to the specific region to adapter-ligated PCR amplified DNA library (Pel et al., 2018). In-solution capture, a capture protocol, genomic regions of interest are targeted using a pool of custom probes synthesized and hybridized in solution to a fragmented genome sample (Querfurth, Fischer, Schweiger, Lehrach, & Mertes, 2012). NimbleGen (now Roche NimbleGen) is the first commercial hybridization-based enrichment kit (Sulonen et al., 2011). The transposon-mediated fragmentation approach also uses probes, but instead of shearing, end-repair, A-tailing, and ligation in a sequential manner, DNA fragmentation and adapter tagging take place simultaneously by transposase enzyme. The transposon ends are used for amplification; hence, the library preparation is significantly reduced by eliminating several steps. The Nextera, commercialized by Epicentre Biotechnologies and later acquired by Illumina, was based on this method (Kia et al., 2017). MIPs technique targets a specific genomic region using single-stranded probes (also called “padlock probes”) that have 20 nucleotides on their ends that is complementary to the ends of a target region which are connected by a linker sequence (Lau, Palanisamy, Trau, & Botella, 2014). The padlock probes hybridization with the complementary DNA takes place by their circulation and can be used to detect sequences with even 40 nucleotides. It was commercialized by Halo Genomics as HaloPlex, but currently, it is available with Agilent Technologies (Stoddard, Niemela, Fleisher, & Rosenzweig, 2014). PCR based approaches can be categorized as singleplex and multiplex. Singleplex PCR Target Enrichment Strategies utilizes standard primers for amplification; only after amplicons produced from different samples are pooled in equimolar amounts, the DNA library preparation is performed (Kozarewa et al., 2015). The microfluidics-based technology by RainDance Technologies uses emulsion to produce millions of droplets in which PCR amplification takes place. “Multiplexing PCR” uses multiple primer pairs simultaneously for amplification. TruSeq Amplicon developed by Illumina and Ion AmpliSeq developed by Life Technologies are its examples (Onda, Takahagi, Shimizu, Inoue, & Mochida, 2018).

23.4.4 ChIP-seq

ChIP-seq or ChIP-sequencing approach is used for detection of protein/DNA interactions by uniting ChIP with massively parallel DNA sequencing (Shah, 2009). It is primarily used for the identification and determination of chromatin-associated proteins, such as polymerases, transcription factors, and other transcriptional machinery. ChIP generates a library of specific regions in DNA bound to a protein of interest. The earliest full genome-wide maps consisting of RNA Polymerase II, histone variant H2A.Z, histone methylation, and the DNA-binding protein CTCF in human T cells using ChIP-seq were created in 2007 (Barski, et al., 2007). The technique has been utilized in uncovering of nucleosome architecture in yeast genes, functional variation in transcription enhancers of mice heart and forebrain, and development of genome-wide ChIP-seq among others (Akerberg et al., 2019).

The five crucial steps of ChIP are: (1) cross-linking of DNA/protein or RNA/protein by formaldehyde; (2) chromatin fragmentation either by enzymatic digestion or sonication; (3) ChIP was achieved by first increasing specific cross-linking using an antibody specific to the protein of interest followed by its incubation and centrifugation to obtain the immunoprecipitation; (4) DNA recovery and purification, accomplished by reversing the effect of cross-linking and later cleaning DNA with an extraction; and (5) DNA purification and quantification by qPCR or ChIP-on-chip (hybrid array) or ChIP sequencing. Purified DNA is ligated with adapters for sequencing that can provide precise information about genome-wide associations with respect to location on the chromosome at high resolution (Gade & Kalvakolanu, 2012). Several approaches, such as DNase-Seq and FAIRE-Seq, have been developed for determining active regulatory regions of nucleosome disrupted or nucleosome depleted in the genome (Winter, Song, Mukherjee, Furey, & Crawford, 2013).

ChIP-seq offers several advantages such as base-pair resolution, reduced noise due to the hybridization as well as distinct and biologically meaningful peaks. In addition, it provides better genome coverage, as the outcome is not limited by the fixed probe sequences on the array. Furthermore, the technique is significant for repetitive regions, heterochromatin, or microsatellites analysis.

23.4.5 Genome-wide bisulfite sequencing

Whole genome bisulfite sequencing (WGBS) is NGS based and aims to determine the status of the DNA methylation [5-methylcytosine (5mC)] within the genome, as it influences multiple cellular processes involving gene expression and chromatin remodeling (Feng, Zhong, Wang, & Jacobsen, 2020). The DNA methylation directly impacts cell

differentiation, selective X-chromosome remodeling, and transposable element suppression among others (Li et al., 2020). Its protocol involves sodium bisulfite treatment to genomic DNA followed by NGS procedure, that is, PCR amplification, library preparation, and sequencing. The sodium bisulfite modifies unmethylated cytosines into uracil, leaving methylated cytosines unaffected. Unmethylated cytosines are represented as thymines subsequent to sequencing; comparing sodium bisulfite-treated and untreated sequences ascertains methylated nucleotide (Li & Tollefsbol, 2011). With the aid of NGS, genome-wide identification of methylated nucleotides at single-nucleotide resolution can be performed since it can precisely estimate the ratio of methylated nucleotide rather than enrichment levels. However, repeated genome resequencing for every experiment is essential (Rauluseviciute, Drabløs, & Rye, 2019).

For earlier WGBS methods, the DNA shearing, ligation of methylated adapters, prior to sodium bisulfite treatment is a crucial step but these adapters are degraded during bisulfite conversion resulting in poor outcomes. A newly developed technique by Epicentre (an Illumina company) performs bisulfite conversion before adapter ligation. This also reduces the amount of sample (50–100 ng) required for sequencing (Kacmarczyk et al., 2018). Whole genome bisulfite sequencing is most popularly employed for acquiring comprehensive base-pair resolution and quantitative information about methylated cytosines in the genome, which further allows unbiased genome-wide DNA methylation profiling (Nordlund, 2020). The technique can easily provide insights into gene cell-fate commitment and reprogramming along with gene regulation (Zhou et al., 2019). The method is also highly significant for the recognition of novel epigenetic markers, possible targets for disease, and investigation related to cancer biopsies among others (Locke et al., 2019).

23.4.6 Yeast two-hybrid

Before the 1980s protein interactions were studied through biochemical techniques. The discovery of the Gal4 molecular structure in 1986 by Keegan and coworkers revolutionized proteomics. The Gal4, a transcriptional activator, consists of two domains, that is, N-terminal DNA-binding domain (DBD) and C-terminal activation domain (AD), which can interact noncovalently and constitute a fully functional protein. This noncovalently binding characteristic of Gal4 was exploited by Stanley Fields and Ok-Kyu Song in 1989 for developing the Y2H technique for monitoring and detecting protein–protein interactions in *Saccharomyces cerevisiae*. The technique involves the formation of constructs by the fusion of two target proteins with DBD and AD of Gal4, respectively, that is, 1st construct with the first target protein and Gal4 DBD (bait) and 2nd construct with the second target protein and Gal4 AD (prey). Expression of fusion proteins in yeast is followed by bait and prey interaction and reconstitution of functional Gal4, which then recruits RNA polymerase II for reporter gene transcription (Xing, Wallmeroth, Berendzen, & Grefen, 2016).

Several variants of Y2H have been developed since 1989, such as one hybrid, one–two hybrid, and three hybrid, to study the interaction between other macromolecules (Drees, 1999). The objective of an one-hybrid system is to study protein–DNA interactions by utilizing a single fusion protein containing both AD and DBD (Xu & Noyes, 2015). Unlike conventional Y2H, the DBD is selected against the promoter sequence of a reporter gene construct; for a positive selection procedure, the interaction of AD with the upstream activation sequence region results in reporter gene transcription (Stylen, Tournu, Tavernier, & Van Dijck, 2012). A three-hybrid system is designed to investigate RNA–protein interactions using a hybrid RNA molecule that serves as an intermediary between two noninteracting proteins through their RNA-binding domains (Hook, Bernstein, Zhang, & Wickens, 2005). The one–two hybrid system simultaneously employs one- and two-hybrid methods to investigate protein–protein and protein–DNA interaction. In addition, various other variants are developed to target either limitation of Y2H or to improve its sensitivity. Split-ubiquitin Y2H is one such method that enables the study of insoluble integral protein interactions, a limitation of traditional Y2H, by the construction of two fusion proteins, that is, bait and prey with C-terminal (Cub) and N-terminal (Nub) moieties of ubiquitin protein, respectively. In addition, the Cub moiety is combined with a transcription factor that can be cleaved off by ubiquitin-specific proteases. The bait–prey interaction causes the assembly of Nub- and Cub-moieties or functional split-ubiquitin, a target ubiquitin-specific protease, resulting in the transcription factor cleavage and initiation of reporter genes transcription (Li, Ivanov, et al., 2017; Li, Wang, et al., 2017).

In traditional Y2H, yeast *S. cerevisiae* is used as a model host for a two-hybrid analysis owing to its reduced cost, ease of culturing and manipulation via nonmolecular techniques, tolerance toward diverse conditions for instance harsh chemicals, as well as the presence of complete genome (Selim, El-Ghwas, Easa, & Abdelwahab Hassan, 2018). The Y2H technique is a powerful and widely used tool in proteomics. However, the technique has some limitations due to inherent characteristics of the host utilized which includes different codon usage, lack of post translational modifications in yeast system and also deficiency of proteins required for expression of proteins of bacterial or mammalian origin (Stylen, van Dijck, & Tournu, 2010). Several novel two-hybrid systems host, such as *Candida albicans*, *E. coli*, *Arabidopsis thaliana*, and *Bombyx mori* among others, have been recognized to address these issues (Schoeters, Munro,

d'Enfert, & Van Dijck, 2018). To study PPIs in native environment *C. albicans* two-hybrid (C2H) system was developed, as *C. albicans* translates the CUG codon into serine instead of leucine (Schoeters & Van Dijck, 2019).

23.4.7 Mass spectrometry-based proteomics

The mass spectrometry (Ms) utilizes magnetic fields and mass-to-charge ratio (m/z) for separating pure or complex, organic or inorganic, solid or liquid, or gaseous samples (Wilschefska & Baxter, 2019). Eugen Goldstein in 1886 first reported the presence of rays under low pressure in gas discharges that travels from anode to cathode (Paital, 2015). Later in 1899 Wilhelm Wien observed their deflection in presence of strong electric or magnetic fields. Modern Ms techniques are inspired by the inventions of Arthur Jeffrey Dempster in 1918 and F.W. Aston in 1919 (Sharma, 2013). The Ms studies involve the sample introduction, ionization, sorting/separation according to m/z , detection, and plotting of spectra (Rubakhin & Sweedler, 2010).

There are several sample introduction techniques, such as direct infusion, gas chromatography (GC), liquid chromatography (LC), high-performance liquid chromatography (HPLC), capillary electrophoresis (CE), direct insertion probe (DIP), and direct ionization utilized in Ms (Callejón-Leblic, García-Barrera, Grávalos-Guzmán, Pereira-Vega, & Gómez-Ariza, 2016). Direct infusion, DIP, and direct ionization are the simplest methods entailing sample pumping directly into the mass spectrometer by syringe with no prior separation step. Typically, the pure sample, that is, free of nonvolatile salts/buffers or detergents is used for identification, measurement, and analysis (Klampfl & Himmelsbach, 2015). The GC, LC, and HPLC utilize mobile and stationary phases to separate the chemical components using properties, such as ion-exchange, adsorption, or size among others. The mobile phase contains volatile, thermally stable molecules with <1250 Da MW and the stationary phase is a column coated with a solid adsorbent or a liquid. In GC, the mobile phase is gaseous due to inert carrier gas while in LC and HPLC, it is liquid as the sample is dissolved in liquid. CE method utilizes a capillary tube to separate liquid sample components according to their electrophoretic mobility under high voltage.

Within a mass spectrometer, the samples are passed through three components, that is, ion source, a mass analyzer, and a detector (Rubakhin & Sweedler, 2010). The ion source causes the ionization of the sample. Numerous techniques are developed for ionization, such as electron impact (EI), chemical ionization (CI), electrospray ionization (ESI), glow discharge (GD) ionization, ambient ionization inductively coupled plasma (ICP) ionization, and matrix-assisted laser desorption ionization (MALDI) among others (Miner & Beauchemin, 2020). EI is one of the oldest and widely employed for gas samples owing to its sharp peaks as an outcome; the samples are fragmented by colliding high energy electrons. Both EI and CI are employed for the GC–Ms analysis (Thibon et al., 2015). In CI, the sample is protonated, instead of fragmentation, by colliding it with protonated gas, such as ammonia, isobutane, or methane (Bouchonnet, Libong, & Sablier, 2004). ESI, widely used for liquid samples containing polar or large biomolecules, transforms samples into aerosols or fine mist by applying high electrical charge during ionization. ESI provides precise masses for whole proteins and complexes. MALDI technique encompasses sample and matrix mixing followed by pulse irradiation, resulting in adsorption or desorption of positive charge on the sample surface (Leopold, Popkova, Engel, & Schiller, 2018). GD and ICP methods use plasma sources, that is, ionized low pressure gas like argon after passing electric current, for ionization and employed for elemental analysis. The choice of ionization method depends on the physical makeup, size, polarity, and other properties of the sample (Wilschefska & Baxter, 2019). Ionization techniques can be divided into two categories as per the source of ions, that is, gas phase (e.g., CI and EI) and desorption (e.g., ESI and MALDI) (Siuzdak, 2004). Alternatively, the techniques can also be categorized into hard (e.g., EI) or soft (e.g., ESI, CI, and MALDI) ionization based on high and low residual energy imparted on samples, respectively (Amirav, Keshet, & Danon, 2015). High residual energy results in large degrees of systematic bond cleavage and molecules fragmentation, whereas low residual energy cause little or no molecular fragmentation.

The next component of the mass spectrometer is a mass analyzer that sorts ions after ionization as per m/z using either static/dynamic fields or electric/magnetic fields (Rubakhin & Sweedler, 2010). Popular mass analyzers typically used with mass spectrometers are quadrupole (Q), time of flight (TOF), magnetic sector, ion trap (IT), and Fourier-transform ion cyclotron resonance (Campuzano et al., 2020). Q mass analyzers utilize radio frequency and the electric field generated by four parallel metal rods to stabilize or destabilize the path of ions depending on their m/z ratio. Only ions with stable paths are passed and later detected by the detector. In Q mass analyzer, the range of ions detected can be widened or narrowed by altering the frequency or electric current. The resolution and selectivity of the quadrupole mass analyzer can be enhanced by adding a magnetic field (Haag, 2016).

TOF mass analyzer computes time taken by ions from the ion source to the detector; the difference is due to ion masses. It is a field-free (devoid of the electric and magnetic field) flight tube in which charged ions were pulsed and

accelerated by a voltage to pass through it. The ions with a smaller mass will move faster and detected earliest by the detector relative to the ions of a higher mass (Haag, 2016). IT mass analyzer traps ions by utilizing an oscillating electric field similar to quadrupole. But unlike quadrupole, IT detects trapped ions. Therefore IT mass analyzers can be considered as a variant of quadrupole mass analyzers. IT first trap every ion in stable orbits, which are sequentially ejected by ramping up electric current and causing instability in ions orbit. Currently, various types of ITs are being used, such as linear, cylindrical, toroidal, 3D IT, and orbitrap. For enhancing the sensitivity and resolution of Ms studies, sometimes more than one mass analyzers are employed in a tandem or hybrid manner. Some examples included triple quadrupole, TOF/TOF, quadrupole–TOF, and quadrupole(x2)–linear IT among others (Allen & McWhinney, 2019).

The last component of the mass spectrometer is the detectors that register the charge or current generated by the ions passing or their collision to the surface (Rubakhin & Sweedler, 2010). The detector provides data for quantitative estimation of each ion through spectrum (Urban, 2016). Several types of detectors are currently in use, such as Faraday cups, ion-to-photo detectors (photographic plates), micro- or multichannel plate, electron multiplier, array, and charge (inductive) detectors (Arevalo, Ni, & Danell, 2020). A few of the popular combinations employed for proteomics are GC–Ms (GC–Q–Ms, GC–QqQ–Ms, GC–Qit–Ms), LC–Ms (LC–ESI–QIT–Ms/MS, LC–LTQ–Orbitrap–Ms/MS), HPLC–Ms (HPLC–ESI–QqQ–IT–Ms/MS), CE–Ms (CE–QTRQP–Ms, CE–TOF–Ms), and ICP–Ms. The technology has been utilized for the identification of secretory proteins, translational proteins, and interaction with pathogens.

23.4.8 Flow and mass cytometry

Flow cytometry (FC) is employed for the detection and measurement of physical and chemical traits of a cell population. The first FC device was invented by Wallace H. Coulter using the Coulter principle in 1953. However, majority of today's flow cytometers are influenced by Mack Fulwyler invention (Robinson, 2013). Nowadays it is often applied for cell counting, sorting, and characterization, along with microorganism or biomarker detection, under diagnostics (Millán & Brunet, 2015). The methodology includes suspension of a sample in a fluid followed by its injection into the flow cytometer where focused laser beams precisely measure cells' optical density. Cells are mostly labeled with fluorescent markers; therefore the output is mostly obtained as wavelength band (Cossarizza et al., 2017).

Originally, fluorescence-based FC was known as “pulse cytophotometry” as per the first patent application in 1968 by Wolfgang Göhde from the University of Münster (Valet, 2003). Five components of the flow cytometer are flow cell, measuring system, detector, amplification system, and computer. The flow cell contains a fluid that carries cells through the light beam for sensing. The measuring system consists of lamps (xenon, mercury); high-power water-cooled lasers (krypton, argon); low-power air-cooled lasers (green-HeNe, red-HeNe, argon); and diode lasers (blue, violet, green, red) for the measurement of impedance or conductivity using resulting light signals (Telford, Hawley, & Hawley, 2003). The detector detects, measures, and converts analog forward-scattered light, side-scattered light, and dye-specific fluorescence signals to digital, which are amplified in either linear or logarithmic manner before computer processing.

A subsequent combination of FC with Ms led to the mass cytometry technique that utilizes elemental mass for determining cell's properties. Inductively coupled plasma and ToF Ms are employed in mass cytometry. Intracellular or extracellular metal-conjugated antibodies bound to antigens can be detected through mass cytometry. This is achieved by labeling cellular proteins with isotopically pure elements followed by their introduction to argon plasma for ionization; the resulting signals are then analyzed by the mass spectrometer. Unlike FC where broad emission spectrum fluorophores resulted in spectral overlap, the mass cytometry utilizes discrete isotopes; therefore no spectral overlap (Spitzer & Nolan, 2016).

The DVS Sciences, Inc., first commercialized cytometry by time of flight (CyTOF) in 2009 and later acquired Fluidigm in 2014; CyTOF2 and Helios (CyTOF3) are other commercialized cytometry technologies (Lee et al., 2019). Mass cytometry applications are diverse from the medical line of work, including immunology, hematology, and oncology to basic science, such as hematopoiesis, cell cycle, cytokine expression, and differential signaling response-related studies.

23.4.9 Live cell imaging

Live cell imaging, which started in the first decade of the 20th century, uses time-lapse microscopy for investigation of living cells with the goal of comprehending cellular dynamics (Baker, 2010). One of the first studies involving time-lapse microcinematographic films was the fertilization and development of the sea urchin egg by Julius Ries (Landecker, 2009). Since then, the development in microscopy permits researchers to study living cells in greater detail

but with less effort. The conventional contrast methods, such as differential interference or phase contrast, and fluorescence-based methods of microscopy are compatible with live-cell imaging. The phase contrast microscopy, introduced in the 1940s by Frits Zernike (Nobel Prize in 1953), was mainly used to observe unstained living cells; fluorescent microscopy was primarily for synthetic and organic fluorescent stains (Dunst & Tomancak, 2019). Other alternative techniques for living cells are Quantitative Phase Contrast microscopy and holotomography (HT) (Kim, Lee, Jung, et al., 2018; Kim, Lee, Youn, et al., 2018). Quantitative Phase Contrast microscopy is a noninvasive quantitative technique, which in combination with rotational scanning generates time-lapse 3D images at high resolution (Kim, Lee, Jung, et al., 2018; Kim, Lee, Youn, et al., 2018). On the other hand, HT is a most common, laser-based, nonphoto toxic technique that measures three-dimensional refractive index and is used to study/trace the fluorescent proteins in a microscopic sample of cells and tissues without the need for staining (Kim et al., 2017).

Live cell imaging has revolutionized cell biology by facilitating cell structure and process observation in real time, and over time leading to better insight into the operations of a cell. It has been used to study breast cancer cells among others (Ji et al., 2019). In addition, endocytosis, exocytosis, cellular integrity, protein trafficking, enzyme activity, and signal transduction can be monitored. It has been employed for understanding interaction and response to environmental cues along with monitoring molecules within alive animals. For instance, it traced neural stem and progenitor cells in adult mouse brains for 2 months (Pilz et al., 2018).

23.5 Resources for biological network-based studies

The recent advancement in the various omics field, such as genomics, transcriptomics, and proteomics among others, has produced a huge amount of biological data. This data expansion demands the development of bioinformatics resources which includes databases for storage, organization, processing, and easy access (Manzoni et al., 2018). Some of the important databases are discussed below.

23.5.1 Kyoto Encyclopedia of Genes and Genomes

KEGG database provides information related to genes, proteins, reactions, and pathways. Developed in 1995 with the primary objective of computational reconstruction of biological systems to comprehend functions and utilities from molecular level to organism to ecosystem level. The database contains information derived from the output of high-throughput experimental and sequencing technologies. The KEGG pathway, KEGG reactions, KEGG enzymes, and KEGG network are some of the sections at the KEGG database that provides information related to the biological networks. KEGG also provides mapping and annotation tools like KEGG Mapper, BlastKOALA, and GhostKOALA (Kanehisa, Sato, Kawashima, Furumichi, & Tanabe, 2016).

23.5.2 BioCyc Database Collection

The BioCyc is an assemblage of Pathway/Genome Databases (PGDBs) for deciphering genomes and metabolic pathways of other sequenced organisms. The BioCyc PGDBs are created using predicted operons, metabolic pathways along with information like protein features and Gene Ontology integrated from other databases. Tools hosted at the BioCyc website can be used for searching and visualization, analyzing omics data, and comparative genomics. BioCyc databases are arranged into three tiers according to the amount of manual attention and revision. Tier 1 PGDBs receive extensive manual attention and are updated regularly. Tier 1 includes MetaCyc containing comprehensive information related to metabolic pathways and enzymes from over 2063 organisms, and species-specific EcoCyc (*E. coli* K-12 substr. MG1655), HumanCyc (250 human metabolic pathways), AraCyc (*A. thaliana*), and YeastCyc (*S. cerevisiae*). The databases under Tier 2 PGDBs were generated computationally using the PathoLogic program and are moderately updated, while Tier 3 PGDBs were generated similar to Tier 2 but do not receive any manual updates. Currently BioCyc contains a collection of 18,023 PGDBs and displays individual and complete metabolic pathways (Caspi et al., 2018).

23.5.3 ENZYME

ENZYME database contains information related to the enzyme nomenclature, characteristics, and Enzyme Commission (EC) number. It is built with counsels of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB). The database is hosted and maintained at the EXPASy proteomics server of the Swiss

Institute of Bioinformatics. The search of specific entry not only leads to the reaction that is catalyzed but also provides link outs to other genes, enzymes, and literature databases, such as BRENDA, KEGG, and PUBMED (Gasteiger et al., 2003).

23.5.4 ExplorEnz: the enzyme database

ExplorEnz is a manually organized, peer-reviewed, and open-access database of the enzyme nomenclature. The database supplies a list, produced jointly by the Nomenclature Committee (NC) of IUBMB and IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN). Enzyme entries are authorized by a subcommittee of NC-IUBMB comprising members from Switzerland, UK, USA, France, Japan, Ireland, and Germany. Currently, over 7800 entries are listed on the database and recommended additions and amendments are accessible for public review, 4 weeks before their incorporation into the list. ExplorEnz database is created and is maintained for the IUBMB by Andrew McDonald of the School of Biochemistry and Immunology, Trinity College, Ireland.

23.5.5 Biochemical Genetic and Genomic/Biochemical Genetic and Genomic models

Biochemical Genetic and Genomic (BiGG) is a freely available noncommercial knowledge base of biochemical and genomic-scale reconstructed metabolic networks. The Systems Biology Research Group at the University of California, San Diego, is hosting and maintaining the BiGG database. Presently, the BiGG models, also freely available for non-commercial application, have replaced the original BiGG Database. Genes and metabolites in the BiGG models are linked to NCBI genome annotations and various external databases, such as PubChem and KEGG among others (Schellenberger, Park, Conrad, & Palsson, 2010).

23.5.6 STRING

STRING is a collection of known and predicted PPIs and their functional enrichment analysis. The interactions were derived using physical and functional associations obtained from literature, primary databases, and computational prediction. Interactions in STRING are drawn utilizing genomic context predictions, coexpression (conserved), high-throughput lab experiments, automated text mining, and information from other databases. STRING is part of the ELIXIR infrastructure along with databases, such as ArrayExpress, BioStudies, and IntAct among others (Szkarczyk et al., 2021).

23.5.7 metaTIGER

metaTIGER is a database of metabolic profiles and phylogenomic info generated from 2257 large phylogenetic trees of eukaryotes. metaTIGER uses SHARKhunt, a high-throughput genome annotation program for the prediction of enzymes and organisms' metabolic profiles. SHARKhunt program assesses genomic DNA sequence against enzyme profiles created by the alignment of conserved regions of genes of known function. PSI-BLAST and Hidden Markov Models are used for generating alignment for enzyme profiles as well as maximum-likelihood phylogenetic tree construction. metaTIGER can also be used to compare 10 organisms through table-based comparison. The organism-specific and comparative pathway images can be viewed interactively at metaTIGER (Whitaker, Letunic, McConkey, & Westhead, 2009). Several other databases are listed in Table 23.1.

23.6 Tools for network pathway analysis

The network pathway analysis involves the reconstruction of the network by assembling and integrating data generated from various experimental methodologies mentioned above. In addition, information from different databases like KEGG and BioCyc can be used to identify entire genes or proteins as well as reconstructions of the draft network for the organism of interest. For instance, ERGO, Pathway Tools, and ModelSEED first compiles raw data into pathways then to metabolic and nonmetabolic comprehensive network followed by mathematical simulation for their verification and refinement (Jing et al., 2014). The quality of the reconstructed network model relies on the accurate prediction of catalysis of biochemical reaction using protein sequence along with phenotype-sequence inference (Pitkänen, Rousu, & Ukkonen, 2010). Several tools have been developed for modeling and refinement of biological networks as mentioned below:

TABLE 23.1 Various biological databases and the type of information archived by them.

Database	Description	References
BambooNET	Coexpression networks with functional modules for Moso Bamboo	Ma et al. (2018)
CoryneRegNet 7	Analysis platform of corynebacterial gene regulatory networks	Parise et al. (2020)
<i>RegNetwork</i>	Transcriptional and posttranscriptional regulatory relationships for human and mouse	Liu (2015) and Liu, Wu, Miao, and Wu (2015)
TRRUST v2	Human and mouse transcriptional regulatory interactions reference database	Han et al. (2018)
GRNdb	Gene regulatory networks in diverse human and mouse conditions	Fang et al. (2021)
Abasy Atlas v2.2	Bacterial regulatory networks	Escorcia-Rodríguez, Tauch, and Freyre-González (2020)
APID	Agile Protein Interactomes DataServer	Prieto and De Las Rivas (2006)
MIPPIE	Integrated protein–protein interaction reference	Alanis-Lobato, Möllmann, Schaefer, and Andrade-Navarro (2020)
PICKLE 2.0	A human protein–protein interaction <i>meta</i> -database	Gioutlakis, Klapa, and Moschonas (2017)
Protein Interaction Network Analysis	A platform for protein interaction network construction	Cowley et al. (2012)

23.6.1 Pathway tools

A pathway tool is an informatics package, developed by Peter Karp and coworkers at the SRI International Bioinformatics Research Group, under the BioCyc database. Its module, such as PathoLogic and MetaFlux, utilizes an annotated genome for the prediction of possible metabolic reaction pathways and flux-balance analysis for making metabolic model, respectively. The complete network and its components are visualized using the Navigator module ([Karp et al., 2016](#)).

23.6.2 ERGO

ERGO toolkit primarily acquires and visualizes sequenced genomes for comparative genomics and systems biology. ERGO is popularly used for the functional annotation of genes and their integration into pathways. Furthermore, it can recognize mischaracterized genes, puzzling pathways, or previously unknown gene products. Developed by Integrated Genomics, it uses proprietary algorithms for creating a comprehensive microbial metabolic and nonmetabolic network ([Overbeek et al., 2003](#)).

23.6.3 KEGGtranslator

KEGGtranslator is a stand-alone tool for both the visualization and conversion of KGML formatted XML-files, that is, the KEGG database. Unlike other translators, it amends the missing components of partial reactions within the pathway. KEGGtranslator can provide output in SIF, SBML, BioPAX, SBGN, GML, GraphML, JPG, GIF, or LaTeX format among others ([Wrzodek, Dräger, & Zell, 2011](#)).

23.6.4 ModelSEED

ModelSEED is an online resource that can analyze, compare, create, and organize metabolic models of genome size. The network of metabolic and gene/protein reactions is automatically constructed by ModelSEED using various associations. Flux Balance Analysis can also be performed to produce a model of microbial metabolism using reactions

related to biomass composition. The output of the RAST annotation system can be directly used to construct a draft metabolic model by ModelSEED (Seaver et al., 2020).

23.6.5 Network Analysis Tools

Network Analysis Tools is an integrated suite of various algorithms that are utilized in biological networks for analysis. It provides complete analytical workflow in a stepwise manner. Its tools are used for creating, comparing, and analyzing network-based graphs and clusters. The cluster and path information further enrich the networks and subnetworks. The time taken for the complete execution of the workflow is approx. 1 h (Brohée, Faust, Lima-Mendez, Vanderstocken, & van Helden, 2008).

23.6.6 BioNetStat

BioNetStat is available with the Bioconductor package. It is user-friendly and used for comparison of correlation networks generated using probability distribution and features like centrality. As compared to the Gene Set Coexpression Analysis, its scoring is not affected by the increase of network numbers. The tool also identified previously unknown networks associated with signaling pathways (Jardim, De Siqueira Santos, Fujita, & Buckeridge, 2019).

23.6.7 OmicsNet

OmicsNet is a free web-based tool primarily used to create and visualize varied types of molecular interaction networks in 3D space. The tool utilizes and merges several lists of molecules of interest, that is, genes, transcription factors, microRNAs, proteins, or metabolites among others. The Web Graphics Library technology, a popular choice to deliver its 3D network visualization system in browsers. Various functions, such as coloring, topology, and enrichment analysis, are available within OmicsNet (Zhou & Xia, 2018).

23.7 Applications of network biology

Biological networks have been successfully employed in decoding biological mechanisms and disease etiologies as well as prediction of therapeutics interventions at both molecular and systems levels. The major applications of network biology are (1) detection and prioritization of gene-disease associations, (2) disease-associated subnetwork detection, and (3) target identification and drug discovery. In this section, we discuss the utility of network biology in the analysis of rare diseases, protein function determination, pathway determination, essential protein identification, and functional module detection. The methodologies employed in the below mentioned studies can also be utilized in the investigation of other common or complex diseases via modulation.

23.7.1 Applications in rare diseases

Several disease types including few rare diseases have been investigated at molecular levels via biological networks. Herein, we discuss few such studies of rare diseases for instance congenital hyperinsulinism (CH) and systemic sclerosis (SSc).

23.7.1.1 Congenital hyperinsulinism

CH is characterized by abnormally high levels of insulin secretion by pancreatic beta-cells in neonates and infancy. Lifelong implications of CH include hypoglycaemic brain injury and mental retardation in diseased individuals (Xu et al., 2019). Stevens et al. (2013) mapped nine functionally diverse genes that were considered to be linked with CH onto the BIOGRID interactome and observed a highly connected core subnetwork comprising of mapped genes. In addition, modularity was examined in the inferred network and associated pathways were analyzed (Stevens et al., 2013).

23.7.1.2 Systemic sclerosis

SSc is a complex, multisystem connective tissue disorder whose complete pathogenesis and progression is not understood. SSc affects the skin, internal organs, musculoskeletal system along the gastrointestinal tract and is characterized by fibrosis, vascular wall damage, inflammation, perturbation of immune response, and specific autoantibody presence.

The prevalence of the disease is 1 in 6500 adults and majorly affects females (female/male ratio 4:1) (Sobolewski et al., 2019). Multiple network strategies were employed to decode tissue-specific molecular patterns of SSc. A weighted GCN was constructed by microarray gene expression data acquired from 321 SSc patients related to four affected tissues, that is, skin, lung, esophagus, and peripheral blood, and consensus clustering methodology was employed. The investigation successfully detected disease-linked subnetworks that overlapped across the tissues, and a common pathogenic pattern related to the immune-fibrotic axis in multiple tissues was also observed, and that suggested the presence of profibrotic macrophages in the tissues of SSc-inflicted persons. Furthermore, the GIANT database was queried to extract the tissue-specific networks by the usage of the immune-fibrotic axis gene sets, and the subnetworks were identified by *igraph* R package. In addition, to decode the disparity of the immune-fibrotic connectivity between lung and skin, differential network analysis was performed by contrasting the detected lung and skin subnetworks, and distinct transcriptional programs were detected for activation of macrophages in the lungs from SSc inflicted persons (Taroni et al., 2017).

23.7.1.3 Application of human gene connectome to herpes simplex virus encephalitis

Herpes simplex virus encephalitis (HSE) is a rare disease that is associated with significant morbidity and death. The causative agent herpes simplex virus (HSV) affects the central nervous system, which leads to personality changes, altered mental function, and seizures (Bradshaw & Venkatesan, 2016). Genetic investigations carried out to decode the pathogenesis of HSE revealed *TLR3* deficiency in a fraction of HES-inflicted children (Lim et al., 2014). To detect more HSE responsible candidate genes, the human gene connectome (HGC), a network-based method, was proposed by Itan et al. for the identification of closely linked gene clusters that are centered around a given gene of interest (Zhang & Itan, 2019). HGC was constructed by retrieval of the direct human PPIs (14,129 genes and 328,391 connections) from the STRING database. Various topological matrices, such as gene-to-gene distance, shortest path as well as their analogous routes were computed for each possible pairs of genes within the network. At last, a gene-centered connectome for all studied genes based on HGC was constructed, while considering the parameters, such as shortest path, distance distribution, and *P*-value, as well as the closeness of a gene present at the periphery to the central gene of interest. Notably, 20 of the 21 reported HSE-associated genes in the *TLR3*-centered connectome were detected among the top 5% of the closet neighbors of *TLR3* which included *TRIF*, *TRAF3*, and *UNC93B1* among others (Lim et al., 2014; Zhang & Itan, 2019). The results obtained by computational investigation corroborated with experimental investigation performed to understand the pathogenesis of HSE (Zhang & Itan, 2019).

23.7.1.4 Application of vertex similarity to rare diseases

Vertex similarity (VS) method detects as well as ranks candidate disease genes of orphan disorders (OD) via calculation of VS score between nodes of a given network by using Edge-weighted equation and shortest-path-based equation (Zhu, Kushwaha, Berman, & Jegga, 2012). The edge-weighted equation is used to provide a similarity score for directly connected genes in the network, while also considering the neighborhood, whereas the shortest-path-based equation is utilized to compute in case of no direct connected pair of genes. As a result, for the disease under investigation with reported linked genes considered as seed genes, the VS method provides ranks to all other nodes in accordance with the computed VS scores with the seed genes. The human protein interaction network comprising of 11,765 proteins and 69,167 PPIs was constructed by amalgamating three databases, that is, HPRD (Nguyen, Gardiner, & Cios, 2011), BIND, and Reactome along with literature mining. VS method was tested on OD-associated genes (1598) corresponding to 172 ODs, via information retrieved from the Orphanet database. The leave-one-out cross validation was employed as the validation strategy by selecting a reported disease-linked gene for one OD as the target gene and adding it with 99 others that were selected randomly to create a test set comprising 100 genes. The top score genes identified by VS were also corroborated with the literature studies. VS methodology exhibited a success rate in the range from 43% to 68% (Zhang & Itan, 2019).

23.7.1.5 Application of DIGNiFI (Disease-Causing Gene Finder) to rare disorders

The DIGNiFI methodology identifies and provides ranks to disease-linked genes by calculating the topological similarity between genes by the usage of local and global properties (Liu, Yang, Lin, Simmons, & Lu, 2017). Similar to VS method, it computes the similarity between a pair of genes, that is, for both directly linked as well as indirectly associated genes, and provides a direct neighbor (DN) score that represents the local connectivity. In addition, it utilizes the local random walk (LRW) algorithm, for big and sparse graphs, for the detection of global network properties. Finally, a ranking of disease-linked genes is done by adding the DN and LRW scores. To test the method, a human PPIN was

constructed by amalgamating 9453 proteins and 36,867 PPIs from the HPRD database (Keshava Prasad et al., 2009) along with 128 ODs and 1184 associated genes extracted from the Orphanet database. DIGNiFI performed better than the VS method and many other methodologies that utilize topological properties for the ranking of disease-associated genes for rare disorders, by the employment of same validation approach (Zhang & Itan, 2019).

23.7.2 Determination of protein function

One of the challenges of a postgenomic era that requires to be addressed is the determination of protein function also known as functional annotation. Owing to the increased accessibility of PPINs for several organisms, several *in silico* methodologies had been presented for the determination of the function of proteins via the usage of high-throughput datasets (Peng, Wang, Peng, Wu, & Pan, 2017). A network flow-based algorithm for the elucidation of protein function prediction has been developed by Nabieva, Jim, Agarwal, Chazelle, and Singh (2005) that uses structural information of PPINs. In addition, they reported that by taking into account the reliability of the source of PPI, prediction performance improves significantly. The reliability score for an experimental source of data is calculated as the fraction of PPI pairs from the source that share a minimum of one common function, and further reliability of a PPI is also assessed. Similarly, Chua, Sung, and Wong (2006) assessed the reliability of PPIs and predicted functions of proteins in two steps by using the strategy proposed (Nabieva et al., 2005). The functional similarity between two proteins was computed by using both local topologies of the PPINs and reliability of the experimental sources. For a given protein, based on functional similarity, a weight was first assigned to each of its direct and indirect neighbors. Next, each function was given a score based on the weighted frequency of their neighbors. Contrarily, Hu et al. (2011) employed both PPIN and biochemical properties of proteins to elucidate functions of proteins in the mouse. For a given query protein, both neighboring proteins, as well as the interaction weights, were taken into account (Peng et al., 2017).

23.7.3 Pathway determination

The binding of extracellular signaling molecules to the receptors placed on the surface of cells or inside the cell that leads to a series of events within the cell to invoke a response is referred to as the signal transduction pathway (Peng et al., 2017). Several strategies have been reported to detect signal transduction pathways from PPINs while taking into account the PPIs' reliability. Homologous pathways can be searched for a linear query pathway in a PPI network by using the QPath algorithm. The method involves the assignment of confidence scores to PPIs by the usage of a logistic based regression scheme. The detected homologous pathways are given a score based on variables, such as insertions and deletions, in protein sequences against the query pathway as well as confidence score of interactions and sequence similarity between their query proteins and the component proteins. Missing pathway segments in a network between two biologically important proteins can be determined by using the PathFinder method. The methodology takes into account false-positive as well as false-negative PPIs. False-positive PPIs are removed by combining information, such as microarray gene profiling, protein localization, and sequence information with the network. On the other hand, information regarding protein families is utilized to detect false-negative PPIs.

Pathways exhibit orientation, while PPIs do not have an orientation (Peng et al., 2017). To elucidate high confidence pathways and orient PPI, Gitter, Klein-Seetharaman, Gupta, and Bar-Joseph (2011) presented three approximation algorithms based on either probabilistic assignments or weighted Boolean satisfiability solvers. Two weighting schemes were employed for the construction of weighted PPINs. The first weighting method increases the weights of PPINs that are presented in several databases. The second weighing scheme provides a score based on the type of experiments carried to determine the PPI, which is similar to the one by Nabieva et al. (2005). With a similar goal of the orientation of PPIs and elucidation of pathways, a genetic algorithm has been utilized by Nguyen, Vu, Tu, and Bui (2015). Wherein reliability of PPIN is computed based on experimental types utilized to detect PPI as well as the number of different experiments that involve such PPIs.

23.7.4 Essential protein identification

Essential proteins play a vital role in the existence and development of living beings (Zhao & Lei, 2019). Consequently, they are considered significant in pathology investigation as well as in drug development (Chiliza, Pillay, & Pillay, 2017). Various *in silico* techniques have been presented to detect essential proteins within PPINs while also considering the false positives PPIs. To determine essential proteins, several weighting methodologies have been presented for the construction of weighted PPI networks by taking into account false positives. PPIs have been weighted

by Li, Wang, Wang, and Pan (2010) by combining a model based on logistic regression as well as by calculating function similarity between proteins. The proposed method significantly improved the performance of investigated network centrality measures in detecting essential proteins on employment on yeast PPIN. By combining gene expression data with network topology, Tang, Wang, Zhong, and Pan (2014) used Pearson's correlation coefficient and edge-clustering coefficient for weighing of PPIs and to identify essential proteins on the weighted PPINs. Taking into account that essential proteins tend to be conservative in nature, Peng et al. (2012) united the edge-clustering coefficient as well as protein orthology information weighing of PPIs.

23.7.5 Functional modules' identification

Identification of protein complexes from PPIN aids in a better understanding of the function of proteins inside the cell. Performance of protein complex identification methodologies improves considerably with the integration of reliability information of PPIs (Peng et al., 2017). Asthana, King, Gibbons, and Roth (2004) constructed a weighted PPI network by using two metrics, that is, semantic similarity and semantic interactivity, to compute PPIs' reliability on the basis of GO annotation. Next, a flow-based modularization was applied to detect shared modules within the weighted PPINs. Sharan et al. (2005) utilized a method to compute PPI reliability for each when detecting common subnetworks via network alignment algorithm. To identify functional modules, Chou and Cai (2006) used the gene expression microarray profiles information to provide a confidence score for each of the PPIs in the yeast PPIN. Krogan et al. (2006) exploited a machine learning approach for assigning probabilities to PPIs identified by TAP-Ms by combining the Ms scores. The high confident PPIs were then clustered into protein complexes (547) by the usage of the Markov clustering method. To weight the PPIs, Lubovac, Corne, Gamalielsson, and Olsson (2007) exploited the GO semantic similarity between proteins and identified functional modules on the weighted PPI network. Liu, Wong, and Chua (2009) employed an iterative scoring method to estimate the PPIs reliability and presented a maximal cliques clustering method to identify protein complexes in weighted PPI networks. The performance of several clustering algorithms was shown to be improved by weighted PPI networks by Kritikos, Moschopoulos, Vazirgiannis, and Kossida (2011). PE measure was employed by Zaki, Efimov, and Berenguères (2013) to assess the reliability of PPI data. In addition, the researchers also identified protein complexes based on the concept of weighted clustering coefficient from the PPIN dataset.

23.8 Challenges and future perspective

The investigations of biological networks are crippled by several challenges. As discussed in this chapter, network biology investigations are vastly varied with respect to the type of data, data collection methods and preprocessing, statistical methods, mathematical models, and the aim of the specific investigations (Zhang & Itan, 2019). The most analyzed networks are the PPINs; nevertheless, these networks are static, deficient of spatial as well as temporal information, and incomplete in terms of coverage and interaction quality and consist of false-positive interactions (Snider et al., 2015). One of the challenges faced in developing a methodology for PPI prediction is to find good quality data set of multimeric proteins. The dataset required to train predictors should be from known structural complexes in the PDB, contrarily few structures in PDB are multimeric proteins. The relative scarcity of good quality datasets required for training prediction methods is one of the challenges. de Vries and Bonvin (2008) have enlisted 19 different testing and training sets for the 22 different predictors that they use. Owing to the extensive nature of PPIN, the visualization tools encounter both performance and graphical challenges. Moreover, the availability of annotations of nodes and edges adds further complexity to the issue (Agapito, Guzzi, & Cannataro, 2013).

23.8.1 Pseudo temporal ordering

The majority of the expression data that we utilize for network analysis is static in nature and retrieved from methodologies, such as RNA-seq (Saint-Antoine & Singh, 2020). Ideally, dynamic data for each gene are required to get a holistic view of cellular machinery (Levy & Vogel, 2021). Few methods have been presented that utilize static data to generate "pseudo time series" data (Reid & Wernisch, 2016). Integration of pseudo temporal ordering in network inference has been illustrated by Sanchez-Castillo, Blanco, Tienda-Luna, Carrion, and Huang (2018). They inferred GRN by using a collection of 48 genes, and 442 expression profiles originating from mouse embryos. The static expression profiles were derived from single-cell qPCR, and the MOLO algorithm was utilized to generate pseudo time series. Finally, biological insights were drawn with reference to cell differentiation in mice from the constructed GRN. In addition, a computational experiment was also performed to demonstrate how disparities in the temporal order can influence

the properties of the network inference. Moreover, pseudo-temporally ordered expression data from zebrafish was employed for deriving network inference (Sanchez-Castillo et al., 2018). Pseudo-temporal ordering has been advantageous in few studies. However, the assumptions employed in the pseudo-temporal ordering algorithms have been criticized which includes fate transitions of cells that are smooth and continuous. Therefore, more investigation is required to address the limitations of these algorithms (Saint-Antoine & Singh, 2020).

23.8.2 Multiple data sources

Another challenge faced in network biology is the amalgamation of data obtained from different methods to make a calculation. Gene expression data have been combined with DNA methylation array data as well as copy number variant data and successfully applied to breast cancer data (Yuan et al., 2019). On the other hand, Liang, Young, Hung, Raftery, and Yeung (2019) utilized an integrated approach and used multiple sources of data, such as gene expression profiles, ChIP-seq data, gene ontologies, pathway data, and genome binding data, for the construction of GRN and reported improved performance. To investigate cellular differentiation and/or pathology for instance of cancer, it is necessary to investigate some alteration in the GRN for which prior structural knowledge is also required.

23.8.3 Combination of algorithms

It has been long believed that a combination of different algorithms could increase the accuracy of network inference. Empirical confirmation by the computational experiment of the thought has been done by Hill et al. (2016) wherein they combined results derived from randomly selected network inference and assessed the accuracy of results utilizing the area under the ROC curve. The results illustrated a general trend in which the accuracy of results increased with the addition of more algorithms. Similarly, an in silico experiment was carried out wherein top scoring algorithms were combined and that showed higher accuracy in comparison to individual algorithms (Marbach et al., 2012). Although there are a lot of studies that show combining algorithm increases accuracy but still more investigation is required to find the best combination strategy (Saint-Antoine & Singh, 2020).

In recent years, various biological networks have been increasingly utilized to investigate large data sets (Yu, Kim, Xiao, & Hwang, 2013), which may be credited to the reduced cost of large-scale gene expression profiling, for instance, RNA-seq, and to more accessibility of tissue-specific data obtained from perturbation experiments as well as greater sample sizes, which are requisite for successful differential coexpression analyses (Anamika, Verma, Jere, & Desai, 2016). We anticipate better algorithms, integrated databases, and experimental techniques to be applied extensively in the future to attain a holistic view of cellular machinery.

23.9 Conclusion

Biological networks have become the focal point in many research fields and are being investigated by both experimental and computational methodologies. Biological network analysis provides a systematic understanding of the organization as well as the functioning of cellular machinery. The knowledge obtained by network biology can be utilized for analysis and effective intervention to cellular behavior. Biological networks are robust, modular, and dynamic in nature. In this chapter, we have provided general features of various types of biological networks, that is, ecological network, gene expression network, cell signaling network, protein-protein interaction network, metabolic network, and gene coexpression network. In addition, we have discussed various experimental techniques, such as Microarray, Deep mRNA sequencing, Exome sequencing, ChIP-seq, and genome-wide bisulfite sequencing among others, that are required for network construction.

To archive, mine, and perform the analysis of interaction data of biological entities, several databases, such as KEGG, BioCyc Database Collection, ENZYME, ExplorEnz—The Enzyme Database, STRING, and metaTIGER among others, have been constructed. Furthermore, for construction, analysis, and visualization of biological networks, several network analysis tools have been developed, such as ERGO, KEGGtranslator, ModelSEED, and Network Analysis Tools. Notably, both experimental and computational techniques required for interaction prediction, network construction, visualization, and analysis are crippled with limitations. Biological network analyses have deepened our understanding of disease mechanisms, the pathogenesis of various diseases, regulatory mechanisms, and host-pathogen interaction. Moreover, it has helped in biodiversity conservation, drug identification, as well as pathway elucidation. However, various aspects of network biology require better algorithms and techniques for increased accuracy and better biological insights.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Agapito, G., Guzzi, P. H., & Cannataro, M. (2013). Visualization of protein interaction networks: Problems and solutions. *BMC Bioinformatics*, 14 (Suppl. 1), S1.
- Agarwal, S., Deane, C. M., Porter, M. A., & Jones, N. S. (2010). Revisiting date and party hubs: Novel approaches to role assignment in protein interaction networks. *PLoS Computational Biology*, 6(6), e1000817.
- Akerberg, B. N., Gu, F., VanDusen, N. J., Zhang, X., Dong, R., Li, K., et al. (2019). A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. *Nature Communications*, 10, 4907.
- Alanis-Lobato, G., Möllmann, J. S., Schaefer, M. H., & Andrade-Navarro, M. A. (2020). MIPPIE: The mouse integrated protein-protein interaction reference. *Database: The Journal of Biological Databases and Curation*, 2020, baaa035.
- Allen, D., & McWhinney, B. (2019). Quadrupole time-of-flight mass spectrometry: A paradigm shift in toxicology screening applications. *Clinical Biochemist Reviews*, 40(3), 135–146.
- Amirav, A., Keshet, U., & Danon, A. (2015). Soft Cold EI—Approaching molecular ion only with electron ionization. *Rapid Communications in Mass Spectrometry*, 29(21), 1954–1960.
- Anamika, K., Verma, S., Jere, A., & Desai, A. (2016). *Transcriptomic profiling using next generation sequencing—Advances, advantages, and challenges*. *Next generation sequencing—Advances, applications and challenges*. IntechOpen. Available from <https://doi.org/10.5772/61789>.
- Angione, C. (2019). Human systems biology and metabolic modelling: A review—from disease metabolism to precision medicine. *BioMed Research International*. Available from <https://doi.org/10.1155/2019/8304260>.
- Arevalo, R., Ni, Z., & Danell, R. M. (2020). Mass spectrometry and planetary exploration: A brief review and future projection. *Journal of Mass Spectrometry*. Available from <https://doi.org/10.1002/jms.4454>.
- Asthana, S., King, O. D., Gibbons, F. D., & Roth, F. P. (2004). Predicting protein complex membership using probabilistic network reliability. *Genome Research*, 14(6), 1170–1175.
- Azevedo, H., & Moreira-Filho, C. A. (2015). Topological robustness analysis of protein interaction networks reveals key targets for overcoming chemotherapy resistance in glioma. *Scientific Reports*. Available from <https://doi.org/10.1038/srep16830>.
- Baker, M. (2010). Cellular imaging: Taking a long, hard look. *Nature*, 466(7310), 1137–1140.
- Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., . . . Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4), 823–837.
- Blanchard, A. P., Kaiser, R. J., & Hood, L. E. (1996). High-density oligonucleotide arrays. *Biosensors and Bioelectronics*, 21, 20–24.
- Borrett, S. R., Sheble, L., Moody, J., & Anway, E. C. (2018). Bibliometric review of ecological network analysis: 2010–2016. *Ecological Modelling*, 382, 63–82.
- Bouchonnet, S., Libong, D., & Sablier, M. (2004). Low-pressure chemical ionization in ion trap mass spectrometry. *European Journal of Mass Spectrometry*, 10(4), 509–521.
- Bradshaw, M. J., & Venkatesan, A. (2016). Herpes simplex virus-1 encephalitis in adults: Pathophysiology, diagnosis, and management. *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics*, 13(3), 493–508.
- Brohée, S., Faust, K., Lima-Mendez, G., Vanderstocken, G., & van Helden, J. (2008). Network analysis tools: From biological networks to clusters and pathways. *Nature Protocols*, 3(10), 1616–1629.
- Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, 10(1), 1017.
- Butte, A. J., & Kohane, I. S. (2000). Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. In *Pacific symposium on biocomputing* (pp. 418–429).
- Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D., & Craig, D. W. (2016). Translating RNA sequencing into clinical diagnostics: Opportunities and challenges. *Nature Reviews Genetics*, 17(5), 257–271.
- Callejón-Leblic, B., García-Barrera, T., Grávalos-Guzmán, J., Pereira-Vega, A., & Gómez-Ariza, J. L. (2016). Metabolic profiling of potential lung cancer biomarkers using bronchoalveolar lavage fluid and the integrated direct infusion/gas chromatography mass spectrometry platform. *Journal of Proteomics*, 145, 197–206.
- Campuzano, I. D. G., Nshanian, M., Spahr, C., Lantz, C., Netrojjanakul, C., Li, H., et al. (2020). High mass analysis with a fourier transform ion cyclotron resonance mass spectrometer: From inorganic salt clusters to antibody conjugates and beyond. *Journal of the American Society for Mass Spectrometry*, 31(5), 1155–1162.
- Caspi, R., Billington, R., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., et al. (2018). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*, 44(D1), D471–D480.
- Chiliza, T. E., Pillay, M., & Pillay, B. (2017). Identification of unique essential proteins from a Mycobacterium tuberculosis F15/LAM4/KZN phage secretome library. *Pathogens and Disease*, 75(1), ftx001.
- Chou, K. C., & Cai, Y. D. (2006). Predicting protein-protein interactions from sequences in a hybridization space. *Journal of Proteome Research*, 5 (2), 316–322. Available from <https://doi.org/10.1021/pr050331g>.

- Choudhari, J. K., Verma, M. K., Choubey, J., & Sahariah, B. P. (2021). Investigation of MicroRNA and transcription factor mediated regulatory network for silicosis using systems biology approach. *Scientific Reports*, *11*(1), 1265.
- Chua, H. N., Sung, W. K., & Wong, L. (2006). Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. *Bioinformatics (Oxford, England)*, *22*(13), 1623–1630.
- Cossarizza, A., Chang, H. D., Radbruch, A., Akdis, M., Andrä, I., Annunziato, F., et al. (2017). Guidelines for the use of flow cytometry and cell sorting in immunological studies. *European Journal of Immunology*, *47*(10), 1584–1797.
- Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J., & Palsson, B. O. (2004). Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, *429*(6987), 92–96.
- Cowley, M. J., Pinese, M., Kassahn, K. S., Waddell, N., Pearson, J. V., Grimmond, S. M., et al. (2012). PINA v2.0: Mining interactome modules. *Nucleic Acids Research*, *40*(D8625).
- Crea, C., Ali, R. A., & Rader, R. (2016). A new model for ecological networks using species-level traits. *Methods in Ecology and Evolution*. Available from <https://doi.org/10.1111/2041-210X.12471>.
- D'Argenio, V. (2018). The high-throughput analyses era: Are we ready for the data struggle? *High-Throughput*, *7*(1), 8.
- de Anda-Jáuregui, G., Guo, K., McGregor, B. A., Feldman, E. L., & Hur, J. (2019). Pathway crosstalk perturbation network modeling for identification of connectivity changes induced by diabetic neuropathy and pioglitazone. *BMC Systems Biology*, *13*(1), 1.
- de Vries, S., & Bonvin, A. (2008). How proteins get in touch: Interface prediction in the study of biomolecular complexes. *Current Protein and Peptide Science*, *9*(4), 394–406.
- Deberardinis, R. J., & Thompson, C. B. (2012). Cellular metabolism and disease: What do metabolic outliers teach us? *Cell*, *148*(6), 1132–1144.
- Delmas, E., Besson, M., Brice, M. H., Burkle, L. A., Dalla Riva, G. V., Fortin, M. J., et al. (2019). Analysing ecological networks of species interactions. *Biological Reviews*, *94*(1), 16–36.
- Drees, B. L. (1999). Progress and variations in two-hybrid and three-hybrid technologies. *Current Opinion in Chemical Biology*, *3*(1), 64–70.
- Dunst, S., & Tomancak, P. (2019). Imaging flies by fluorescence microscopy: Principles, technologies, and applications. *Genetics*, *211*(1), 15–34.
- Emamjomeh, A., Saboori Robat, E., Zahiri, J., Solouki, M., & Khosravi, P. (2017). Gene co-expression network reconstruction: A review on computational methods for inferring functional information from plant-based expression data. *Plant Biotechnology Reports*, *11*, 71–86.
- Escorcia-Rodríguez, J. M., Tauch, A., & Freyre-González, J. A. (2020). Abasy Atlas v2.2: The most comprehensive and up-to-date inventory of meta-curated, historical, bacterial regulatory networks, their completeness and system-level characterization. *Computational and Structural Biotechnology Journal*, *18*, 1228–1237.
- Fan, X., Ping, L., Sun, H., Chen, Y., Wang, P., Liu, T., . . . Chen, X. (2020). Whole-exome sequencing of discordant monozygotic twin families for identification of candidate genes for microtia-atresia. *Frontiers in Genetics*, *11*, 568052. Available from <https://doi.org/10.3389/fgene.2020.568052>.
- Fang, L., Li, Y., Ma, L., Xu, Q., Tan, F., & Chen, G. (2021). GRNdb: Decoding the gene regulatory networks in diverse human and mouse conditions. *Nucleic Acids Research*, *49*(D1), D97–D103.
- Feng, S., Zhong, Z., Wang, M., & Jacobsen, S. E. (2020). Efficient and accurate determination of genome-wide DNA methylation patterns in *Arabidopsis thaliana* with enzymatic methyl sequencing. *Epigenetics and Chromatin*, *13*, 42.
- Ficklin, S. P., & Feltus, F. A. (2011). Gene coexpression network alignment and conservation of gene modules between two grass species: Maize and rice. *Plant Physiology*, *156*, 1244–1256.
- Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., & Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science (New York, N.Y.)*, *251*(4995), 767–773.
- Gade, P., & Kalvakolanu, D. V. (2012). Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity. *Methods in Molecular Biology*, *809*, 85–104.
- Ganapathiraju, M. K., Thahir, M., Handen, A., Sarkar, S. N., Sweet, R. A., Nimgaonkar, V. L., et al. (2016). Schizophrenia interactome with 504 novel protein-protein interactions. *Npj Schizophrenia*, *2*, 16012.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*(13), 3784–3788.
- Gioutlakis, A., Klapa, M. I., & Moschonas, N. K. (2017). PICKLE 2.0: A human protein-protein interaction meta-database employing data integration via genetic information ontology. *PLoS One*, *12*(10), e0186039.
- Gitter, A., Klein-Seetharaman, J., Gupta, A., & Bar-Joseph, Z. (2011). Discovering pathways by orienting edges in protein interaction networks. *Nucleic Acids Research*, *39*(4), e22.
- Grunstein, M., & Hogness, D. S. (1975). Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(10), 3961–3965.
- Haag, A. M. (2016). Mass analyzers and mass spectrometers. *Advances in Experimental Medicine and Biology*, *919*, 157–169.
- Han, H., Cho, J. W., Lee, S., Yun, A., Kim, H., Bae, D., et al. (2018). TRRUST v2: An expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Research*, *46*(D1), D380–D386.
- Han, J. D. J., Berlin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., . . . Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, *430*, 88–93. Available from <https://doi.org/10.1038/nature02555>.
- He, F., Murabito, E., & Westerhoff, H. V. (2016). Synthetic biology and regulatory networks: Where metabolic systems biology meets control engineering. *Journal of the Royal Society Interfac*, *13*(117).
- Hill, S. M., Heiser, L. M., Cokelaer, T., Linger, M., Nesser, N. K., Carlin, D. E., et al. (2016). Inferring causal molecular networks: Empirical assessment through a community-based effort. *Nature Methods*, *13*(4), 310–318.

- Hook, B., Bernstein, D., Zhang, B., & Wickens, M. (2005). RNA-protein interactions in the yeast three-hybrid system: Affinity, sensitivity, and enhanced library screening. *RNA (New York, N.Y.)*, *11*(2), 227–233.
- Hu, L., Huang, T., Shi, X., Lu, W. C., Cai, Y. D., & Chou, K. C. (2011). Predicting functions of proteins in mouse based on weighted protein-protein interaction network and protein hybrid properties. *PLoS One*, *6*(1), e14556.
- Humphries, M. D., & Gurney, K. (2008). Network “small-world-ness”: A quantitative method for determining canonical network equivalence. *PLoS One*, *3*(4), e0002051.
- Jardim, V. C., De Siqueira Santos, S., Fujita, A., & Buckeridge, M. S. (2019). BioNetStat: A tool for biological networks differential analysis. *Frontiers in Genetics*, *10*, 594.
- Ji, K., Sameni, M., Osuala, K., Moin, K., Mattingly, R. R., & Sloane, B. F. (2019). Spatio-temporal modeling and live-cell imaging of proteolysis in the 4D microenvironment of breast cancer. *Cancer and Metastasis Reviews*, *38*(3), 445–454.
- Jing, L. S., Shah, F. F. M., Mohamad, M. S., Hamran, N. L., Salleh, A. H. M., Deris, S., & Alashwal, H. (2014). Database and tools for metabolic network analysis. *Biotechnology and Bioprocess Engineering*, *9*, 568–585.
- Kacmarczyk, T. J., Fall, M. P., Zhang, X., Xin, Y., Li, Y., Alonso, A., & Betel, D. (2018). “Same difference”: Comprehensive evaluation of four DNA methylation measurement platforms. *Epigenetics and Chromatin*. Available from <https://doi.org/10.1186/s13072-018-0190-4>.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1), D457–D462.
- Kang, M. K., & Han, S. J. (2011). Post-transcriptional and post-translational regulation during mouse oocyte maturation. *BMB Reports*, *44*(3), 147–157.
- Karp, P. D., Latendresse, M., Paley, S. M., Krummenacker, M., Ong, Q. D., Billington, R., et al. (2016). Pathway tools version 19.0 update: Software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*, *17*(5), 877–890.
- Kchouk, M., Gibrat, J. F., & Elloumi, M. (2017). Generations of sequencing technologies: From first to next generation. *Biology and Medicine*, *9*(3). Available from <https://doi.org/10.4172/0974-8369.1000395>.
- Keshava Prasad, T. S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., et al. (2009). Human protein reference database—2009 update. *Nucleic Acids Research*, *37*(Database issue), D767–D772.
- Kia, A., Gloeckner, C., Osothprarop, T., Gormley, N., Bomati, E., Stephenson, M., et al. (2017). Improved genome sequencing using an engineered transposase. *BMC Biotechnology*, *17*, 6.
- Kim, G., Lee, M., Youn, S. Y., Lee, E. T., Kwon, D., Shin, J., et al. (2018). Measurements of three-dimensional refractive index tomography and membrane deformability of live erythrocytes from *Pelophylax nigromaculatus*. *Scientific Reports*, *8*, 9192.
- Kim, K., Park, W. S., Na, S., Kim, S., Kim, T., Do Heo, W., & Park, Y. (2017). Correlative three-dimensional fluorescence and refractive index tomography: Bridging the gap between molecular specificity and quantitative bioimaging. *Biomedical Optics Express*, *8*(12), 5688–5697.
- Kim, Y. S., Lee, S., Jung, J., Shin, S., Choi, H. G., Cha, G. H., et al. (2018). Combining three-dimensional quantitative phase imaging and fluorescence microscopy for the study of cell pathophysiology. *Yale Journal of Biology and Medicine*, *91*(3), 267–277.
- Klampf, C. W., & Himmelsbach, M. (2015). Direct ionization methods in mass spectrometry: An overview. *Analytica Chimica Acta*, *26*(890), 44–59.
- Klein, C., Marino, A., Sagot, M. F., Milreu, P. V., & Brilli, M. (2012). Structural and dynamical analysis of biological networks. *Briefings in Functional Genomics*, *11*(6), 420–433.
- Kozarewa, I., Armisen, J., Gardner, A. F., Slatko, B. E., & Hendrickson, C. L. (2015). Overview of target enrichment strategies. *Current Protocols in Molecular Biology*, *112*, 7.21.1–7.21.23. Available from <https://doi.org/10.1002/0471142727.mb0721s112>.
- Kritikos, G. D., Moschopoulos, C., Vazirgiannis, M., & Kossida, S. (2011). Noise reduction in protein-protein interaction graphs by the implementation of a novel weighting scheme. *BMC Bioinformatics*, *12*, 239. Available from <https://doi.org/10.1186/1471-2105-12-239>.
- Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., et al. (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*, *440*, 637–643.
- Kukurba, K. R., & Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harbor Protocols*, *2015*(11), 951–969.
- Kuzmanov, U., & Emili, A. (2013). Protein-protein interaction networks: Probing disease mechanisms using model systems. *Genome Medicine*, *5*, 37. Available from <https://doi.org/10.1186/gm441>.
- Lamere, A. T., & Li, J. (2019). Inference of gene co-expression networks from single-cell RNA-sequencing data. *Methods in Molecular Biology (Clifton, N.J.)*, *1935*, 141–153.
- Landecker, H. (2009). Seeing things: From microcinematography to live cell imaging. *Nature Methods*, *6*, 707–709.
- Landi, P., Minoarivelo, H. O., Brännström, Å., Hui, C., & Dieckmann, U. (2018). Complexity and stability of ecological networks: A review of the theory. *Population Ecology*, *60*, 319–345.
- Lau, H. Y., Palanisamy, R., Trau, M., & Botella, J. R. (2014). Molecular inners on probe: A new tool for highly specific detection of plant pathogens. *PLoS One*. Available from <https://doi.org/10.1371/journal.pone.0111182>.
- Lee, B. H., Kelly, G., Bradford, S., Davila, M., Guo, X. V., Amir, E. A. D., et al. (2019). A modified injector and sample acquisition protocol can improve data quality and reduce inter-instrument variability of the helios mass cytometer. *Cytometry Part A*, *95*(9), 1019–1030.
- Lee, D. S., Park, J., Kay, K. A., Christakis, N. A., Oltvai, Z. N., & Barabási, A. L. (2008). The implications of human metabolic network topology for disease comorbidity. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(29), 9880–9885.
- Lennon, G. G., & Lehrach, H. (1991). Hybridization analyses of arrayed cDNA libraries. *Trends in Genetics*, *7*(10), 314–317.
- Leopold, J., Popkova, Y., Engel, K. M., & Schiller, J. (2018). Recent developments of useful MALDI matrices for the mass spectrometric characterization of lipids. *Biomolecules*. Available from <https://doi.org/10.3390/biom8040173>.

- Levy, E. D., & Vogel, C. (2021). “Structuromics”: Another step toward a holistic view of the cell. *Cell*, *184*(2), 301–303.
- Li, M., Wang, J., Wang, H., & Pan, Y. (2010). Essential proteins discovery from weighted protein interaction networks. In M. Borodovsky, J. P. Gogarten, T. M. Przytycka, & S. Rajasekaran (Eds.), *Bioinformatics research and applications. ISBRA 2010. Lecture notes in computer science* (p. 6053). Berlin/Heidelberg: Springer. Available from https://doi.org/10.1007/978-3-642-13078-6_11.
- Li, Q. P., Wang, S., & Gou, J. Y. (2017). A split ubiquitin system to reveal topology and released peptides of membrane proteins. *BMC Biotechnology*, *17*, 69.
- Li, S., Lund, J. B., Christensen, K., Baumbach, J., Mengel-From, J., Kruse, T., et al. (2020). Exploratory analysis of age and sex dependent DNA methylation patterns on the X-chromosome in whole blood samples. *Genome Medicine*, *12*, 39.
- Li, Y., & Tollefsbol, T. O. (2011). DNA methylation detection: Bisulfite genomic sequencing analysis. *Methods in Molecular Biology*, *791*, 11–21.
- Li, Z., Ivanov, A. A., Su, R., Gonzalez-Pecchi, V., Qi, Q., Liu, S., et al. (2017). The OncoPPI network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies. *Nature Communications*, *16*(8), 14356.
- Liang, X., Young, W. C., Hung, L. H., Raftery, A. E., & Yeung, K. Y. (2019). Integration of multiple data sources for gene network inference using genetic perturbation data. *Journal of Computational Biology*, *26*(10), 1113–1129.
- Lim, H. K., Seppänen, M., Hautala, T., Ciancanelli, M. J., Itan, Y., Lafaille, F. G., et al. (2014). TLR3 deficiency in herpes simplex encephalitis: High allelic heterogeneity and recurrence risk. *Neurology*, *83*(21), 1888–1897.
- Liu, G., Wong, L., & Chua, H. N. (2009). Complex discovery from weighted PPI networks. *Bioinformatics (Oxford, England)*, *25*(15), 1891–1897.
- Liu, X., Yang, Z., Lin, H., Simmons, M., & Lu, Z. (2017). DIGNiFI: Discovering causative genes for orphan diseases using protein-protein interaction networks. *BMC Systems Biology*, *11*, 23. Available from <https://doi.org/10.1186/s12918-017-0402-8>.
- Liu, Z.-P. (2015). Reverse engineering of genome-wide gene regulatory networks from gene expression data. *Current Genomics*, *16*(1), 3–22.
- Liu, Z. P., Wu, C., Miao, H., & Wu, H. (2015). RegNetwork: An integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database*, *2015*, bav095.
- Locke, W. J., Guanzon, D., Ma, C., Liew, Y. J., Duesing, K. R., Fung, K. Y. C., & Ross, J. P. (2019). DNA methylation cancer biomarkers: Translation to the clinic. *Frontiers in Genetics*. Available from <https://doi.org/10.3389/fgene.2019.01150>.
- Lubovac, Z., Corne, D., Gamalielsson, J., & Olsson, B. (2007). *Weighted cohesiveness for identification of functional modules and their interconnectivity*, . *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 4414, pp. 185–198). Springer. Available from https://doi.org/10.1007/978-3-540-71233-6_15.
- Ma, X., Zhao, H., Xu, W., You, Q., Yan, H., Gao, Z., & Su, Z. (2018). Co-expression gene network analysis and functional module identification in bamboo growth and development. *Frontiers in Genetics*. Available from <https://doi.org/10.3389/fgene.2018.00574>.
- MacNeil, L. T., & Walhout, A. J. M. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research*, *21*(5), 645–657.
- Mahajan, M. C., & McLellan, A. S. (2020). Whole-exome sequencing (WES) for illumina short read sequencers using solution-based capture. *Methods in Molecular Biology*, *2076*, 85–108.
- Makrodimitris, S., Reinders, M. J. T., & van Ham, R. C. H. J. (2020). Metric learning on expression data for gene function prediction. *Bioinformatics (Oxford, England)*, *36*(4), 1182–1190.
- Malhi, Y., Franklin, J., Seddon, N., Solan, M., Turner, M. G., Field, C. B., & Knowlton, N. (2020). Climate change and ecosystems: Threats, opportunities and solutions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1794). Available from <https://doi.org/10.1098/rstb.2019.0104>.
- Manzoni, C., Kia, D. A., Vandrovцова, J., Hardy, J., Wood, N. W., Lewis, P. A., & Ferrari, R. (2018). Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Briefings in Bioinformatics*, *19*(2), 286–302.
- Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., et al. (2012). Wisdom of crowds for robust gene network inference. *Nature Methods*, *9*, 796–804.
- McKenzie, A. T., Katsyov, I., Song, W. M., Wang, M., & Zhang, B. (2016). DGCA: A comprehensive R package for Differential Gene Correlation Analysis. *BMC Systems Biology*, *10*, 106. Available from <https://doi.org/10.1186/s12918-016-0349-1>.
- Mei, J., Xu, N., Mei, J., Liu, L., & Wu, J. (2016). Reconstruction and analysis of a genome-scale metabolic network of *Corynebacterium glutamicum* S9114. *Gene*, *575*(2), 615–622.
- Meyer, J. M., Leempoel, K., Losapio, G., & Hadly, E. A. (2020). Molecular ecological network analyses: An effective conservation tool for the assessment of biodiversity, trophic interactions, and community structure. *Frontiers in Ecology and Evolution*. Available from <https://doi.org/10.3389/fevo.2020.588430>.
- Millán, O., & Brunet, M. (2015). *Flow cytometry as platform for biomarker discovery and clinical validation. General methods in biomarker research and their applications* (pp. 141–164). Springer. Available from https://doi.org/10.1007/978-94-007-7696-8_16.
- Miller, M. B., & Tang, Y. W. (2009). Basic concepts of microarrays and potential applications in clinical microbiology. *Clinical Microbiology Reviews*, *22*(4), 611–633.
- Miner, L., & Beauchemin, D. (2020). Inductively coupled plasma mass spectrometry coupled to cation exchange chromatography for the determination of trace nickel in alkaline electrolyte†. *Journal of Analytical Atomic Spectrometry*, *35*, 1295–1299.
- Mira, N. P., Teixeira, M. C., & Sá-Correia, I. (2012). Characterization of complex regulatory networks and identification of promoter regulatory elements in yeast: “In Silico” and “wet-Lab” approaches. *Methods in Molecular Biology*, *809*, 27–48.
- Motta, S., & Pappalardo, F. (2013). Mathematical modeling of biological systems. *Briefings in Bioinformatics*, *14*(4), 411–422.
- Nabieva, E., Jim, K., Agarwal, A., Chazelle, B., & Singh, M. (2005). Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps. *Bioinformatics (Oxford, England)*, *21*(Suppl. 1), i302–10.

- Nguyen, C. D., Gardiner, K. J., & Cios, K. J. (2011). Protein annotation from protein interaction networks and Gene Ontology. *Journal of Biomedical Informatics*, 44(5), 824–829. Available from <https://doi.org/10.1016/j.jbi.2011.04.010>.
- Nguyen, H. A., Vu, C. L., Tu, M. P., & Bui, T. L. (2015). Discovery of pathways in protein-protein interaction networks using a genetic algorithm. *Data and Knowledge Engineering*, 96–97. Available from <https://doi.org/10.1016/j.datak.2015.04.002>.
- Nordlund, J. (2020). Advances in whole genome methylomic sequencing. *Epigenetics Methods*, 18, 213–233.
- Onda, Y., Takahagi, K., Shimizu, M., Inoue, K., & Mochida, K. (2018). Multiplex PCR targeted amplicon sequencing (MTA-Seq): Simple, flexible, and versatile SNP genotyping by highly multiplexed PCR amplicon sequencing. *Frontiers in Plant Science*, 9, 201. Available from <https://doi.org/10.3389/fpls.2018.00201>.
- Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E., et al. (2003). The ErgoTM genome analysis and discovery system. *Nucleic Acids Research*, 31(1), 164–171.
- Paital, B. (2015). Mass spectrophotometry: An advanced technique in biomedical sciences. *Advanced Techniques in Biology and Medicine*, 4(3).
- Papin, J. A., Hunter, T., Palsson, B. O., & Subramaniam, S. (2005). Reconstruction of cellular signalling networks and analysis of their properties. *Nature Reviews Molecular Cell Biology*, 6, 99–111.
- Parise, M. T. D., Parise, D., Kato, R. B., Pauling, J. K., Tauch, A., Azevedo, V. A., . . . Baumbach, J. (2020). CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks. *Scientific Data*, 7, 142. Available from <https://doi.org/10.1038/s41597-020-0484-9>.
- Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., et al. (2011). Using graph theory to analyze biological networks. *BioData Mining*, 4, 10. Available from <https://doi.org/10.1186/1756-0381-4-10>.
- Pel, J., Leung, A., Choi, W. W. Y., Despotovic, M., Lloyd Ung, W., Shibahara, G., et al. (2018). Rapid and highly-specific generation of targeted DNA sequencing libraries enabled by linking capture probes with universal primers. *PLoS One*, 13(12), e0208283.
- Peng, W., Wang, J., Wang, W., Liu, Q., Wu, F. X., & Pan, Y. (2012). Iteration method for predicting essential proteins based on orthology and protein-protein interaction networks. *BMC Systems Biology*, 6, 87. Available from <https://doi.org/10.1186/1752-0509-6-87>.
- Peng, X., Wang, J., Peng, W., Wu, F. X., & Pan, Y. (2017). Protein-protein interactions: Detection, reliability assessment and applications. *Briefings in Bioinformatics*, 18(5), 798–819.
- Perez, C., & Germon, R. (2016). *Graph creation and analysis for linking actors: Application to social data. Automating open source intelligence: Algorithms for OSINT*. . . Available from <https://doi.org/10.1016/B978-0-12-802916-9.00007-5>.
- Pilz, G. A., Bottes, S., Betizeau, M., Jörg, D. J., Carta, S., Simons, B. D., et al. (2018). Live imaging of neurogenesis in the adult mouse hippocampus. *Science (New York, N.Y.)*, 359(6376), 658–662.
- Pitkänen, E., Rousu, J., & Ukkonen, E. (2010). Computational methods for metabolic reconstruction. *Current Opinion in Biotechnology*, 21(1), 70–77.
- Pornputtpong, N., Nookaew, I., & Nielsen, J. (2015). Human metabolic atlas: An online resource for human metabolism. *Database*, 2015. Available from <https://doi.org/10.1093/database/bav068>, bav068.
- Prieto, C., & De Las Rivas, J. (2006). APID: Agile protein interaction DataAnalyzer. *Nucleic Acids Research*, 34(Web Server issue), W298–W302. Available from <https://doi.org/10.1093/nar/gkl128>.
- Prieto, C., Risueño, A., Fontanillo, C., & De Las Rivas, J. (2008). Human gene coexpression landscape: Confident network derived from tissue transcriptomic profiles. *PLoS One*. Available from <https://doi.org/10.1371/journal.pone.0003911>.
- Querfurth, R., Fischer, A., Schweiger, M. R., Lehrach, H., & Mertens, F. (2012). Creation and application of immortalized bait libraries for targeted enrichment and next-generation sequencing. *Biotechniques*, 52(6). Available from <https://doi.org/10.2144/0000113877>.
- Rao, V. S., Srinivas, K., Sujini, G. N., & Kumar, G. N. S. (2014). Protein-protein interaction detection: Methods and analysis. *International Journal of Proteomics*, 2014, 1–12. Available from <https://doi.org/10.1155/2014/147648>.
- Rauluseviciute, I., Drabløs, F., & Rye, M. B. (2019). DNA methylation data by sequencing: Experimental approaches and recommendations for tools and pipelines for data analysis. *Clinical Epigenetics*, 11, 193. Available from <https://doi.org/10.1186/s13148-019-0795-x>.
- Reid, J. E., & Wernisch, L. (2016). Pseudotime estimation: Deconfounding single cell time series. *Bioinformatics (Oxford, England)*, 32(19), 2973–2980.
- Robinson, J. P. (2013). Wallace H. Coulter: Decades of invention and discovery. *Cytometry Part A*, 83A(5), 424–438. Available from <https://doi.org/10.1002/cyto.a.22296>.
- Rothman, D. H. (2017). Thresholds of catastrophe in the Earth system. *Science Advances*, 3(9), e1700906.
- Roy, S., Bhattacharyya, D. K., & Kalita, J. K. (2014). Reconstruction of gene co-expression network from microarray data using local expression patterns. *BMC Bioinformatics*, 15, S10. Available from <https://doi.org/10.1186/1471-2105-15-S7-S10>.
- Rubakhin, S. S., & Sweedler, J. V. (2010). A mass spectrometry primer for mass spectrometry imaging. *Methods in Molecular Biology*, 656, 21–49. Available from https://doi.org/10.1007/978-1-60761-746-4_2.
- Saint-Antoine, M. M., & Singh, A. (2020). Network inference in systems biology: Recent developments, challenges, and applications. *Current Opinion in Biotechnology*, 63, 89–98.
- Sanchez-Castillo, M., Blanco, D., Tienda-Luna, I. M., Carrion, M. C., & Huang, Y. (2018). A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. *Bioinformatics (Oxford, England)*, 34(6), 964–970.
- Schellenberger, J., Park, J. O., Conrad, T. M., & Palsson, B. T. (2010). BiGG: A biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11, 213. Available from <https://doi.org/10.1186/1471-2105-11-213>.
- Schoeters, F., Munro, C. A., d'Enfert, C., & Van Dijck, P. (2018). A high-throughput candida albicans two-hybrid system. *mSphere*, 3(4). Available from <https://doi.org/10.1128/msphere.00391-18>, e00391-18.

- Schoeters, F., & Van Dijck, P. (2019). Protein-Protein Interactions in *Candida albicans*. *Frontiers in Microbiology*. Available from <https://doi.org/10.3389/fmicb.2019.01792>.
- Seaver, S. M. D., Liu, F., Zhang, Q., Jeffryes, J., Faria, J. P., Edirisinghe, J. N., et al. (2020). The ModelSEED Database for the integration of metabolic annotations and the reconstruction, comparison, and analysis of metabolic models for plants, fungi, and microbes. *BioRxiv*. Available from <https://doi.org/10.1101/2020.03.31.018663>.
- Selim, K. A., El-Ghwas, D. E., Easa, S. M., & Abdelwahab Hassan, M. I. (2018). Bioethanol a microbial biofuel metabolite; New insights of yeasts metabolic engineering. *Fermentation*, 4(1), 16. Available from <https://doi.org/10.3390/fermentation4010016>.
- Shah, A. (2009). Chromatin immunoprecipitation sequencing (ChIP-Seq) on the SOLiD™ system. *Nature Methods*, 6(4).
- Shameer, S., Logan-Klumpler, F. J., Vinson, F., Cottret, L., Merlet, B., Achcar, F., . . . Jourdan, F. (2015). TrypanoCyc: A community-led biochemical pathways database for *Trypanosoma brucei*. *Nucleic Acids Research*, 43(D1), D637–D644.
- Sharan, R., Suthram, S., Kelley, R. M., Kuhn, T., McCuine, S., Uetz, P., et al. (2005). Conserved patterns of protein interaction in multiple species. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6), 1974–1979.
- Sharma, K. S. (2013). Mass spectrometry—The early years. *International Journal of Mass Spectrometry*, 349–350, 3–8. Available from <https://doi.org/10.1016/j.ijms.2013.05.028>.
- Shen-Orr, S. S., Milo, R., Mangan, S., & Alon, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics*, 31, 64–68.
- Singh, A. J., Ramsey, S. A., Filtz, T. M., & Kioussi, C. (2018). Differential gene regulatory networks in development and disease. *Cellular and Molecular Life Sciences*, 75, 1013–1025. Available from <https://doi.org/10.1007/s00018-017-2679-6>.
- Siuzdak, G. (2004). An introduction to mass spectrometry ionization: An excerpt from the expanding role of mass spectrometry in biotechnology; MCC Press: San Diego, 2005. *JALA: Journal of the Association for Laboratory Automation*, 9(2), 50–63.
- Snider, J., Kotlyar, M., Saraon, P., Yao, Z., Jurisica, I., & Stagljar, I. (2015). Fundamentals of protein interaction network mapping. *Molecular Systems Biology*, 11(12). Available from <https://doi.org/10.15252/msb.20156351>.
- Sobolewski, P., Maślińska, M., Wieczorek, M., Łagun, Z., Malewska, A., Roszkiewicz, M., et al. (2019). Systemic sclerosis—Multidisciplinary disease: Clinical features and treatment. *Reumatologia*, 57(4), 221–233. Available from <https://doi.org/10.5114/reum.2019.87619>.
- Sonawane, A. R., Platig, J., Fagny, M., Chen, C. Y., Paulson, J. N., Lopes-Ramos, C. M., et al. (2017). Understanding tissue-specific gene regulation. *Cell Reports*, 21(4), 1077–1088.
- Sonawane, A. R., Weiss, S. T., Glass, K., & Sharma, A. (2019). Network medicine in the age of biomedical big data. *Frontiers in Genetics*, 10, 294. Available from <https://doi.org/10.3389/fgene.2019.00294>.
- Song, L., Langfelder, P., & Horvath, S. (2012). Comparison of co-expression measures: Mutual information, correlation, and model based indices. *BMC Bioinformatics*, 13, 328. Available from <https://doi.org/10.1186/1471-2105-13-328>.
- Sparks, E. E., Drapek, C., Gaudinier, A., Li, S., Ansariola, M., Shen, N., et al. (2016). Establishment of expression in the SHORTROOT-SCARECROW transcriptional cascade through opposing activities of both activators and repressors. *Developmental Cell*, 39(5), 585–596.
- Spitzer, M. H., & Nolan, G. P. (2016). Mass cytometry: Single cells, many features. *Cell*. Available from <https://doi.org/10.1016/j.cell.2016.04.019>.
- Stevens, A., Cosgrove, K. E., Padidela, R., Skae, M. S., Clayton, P. E., Banerjee, I., & Dunne, M. J. (2013). Can network biology unravel the aetiology of congenital hyperinsulinism? *Orphanet Journal of Rare Diseases*, 8, 21. Available from <https://doi.org/10.1186/1750-1172-8-21>.
- Stoddard, J. L., Niemela, J. E., Fleisher, T. A., & Rosenzweig, S. D. (2014). Targeted NGS: A cost-effective approach to molecular diagnosis of PIDs. *Frontiers in Immunology*, 76(2), 331–382.
- Stynen, B., van Dijck, P., & Tournu, H. (2010). A CUG codon adapted two-hybrid system for the pathogenic fungus *Candida albicans*. *Nucleic Acids Research*, 38(19), e184.
- Stynen, B., Tournu, H., Tavernier, J., & Van Dijck, P. (2012). Diversity in genetic in vivo methods for protein-protein interaction studies: From the yeast two-hybrid system to the mammalian split-luciferase system. *Microbiology and Molecular Biology Reviews*, 76(2), 331–382.
- Su, G., Morris, J. H., Demchak, B., & Bader, G. D. (2014). Biological network exploration with Cytoscape 3. *Current Protocols in Bioinformatics*, 2014, 8.13.1–8.13.24. Available from <https://doi.org/10.1002/0471250953.bi0813s47>.
- Sueur, J., Krause, B., & Farina, A. (2019). Climate change is breaking earth's beat. *Trends in Ecology and Evolution*, 34(11). Available from <https://doi.org/10.1016/j.tree.2019.07.014>.
- Sulonen, A. M., Ellonen, P., Almusa, H., Lepistö, M., Eldfors, S., Hannula, S., et al. (2011). Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biology*, 12, R94. Available from <https://doi.org/10.1186/gb-2011-12-9-r94>.
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, 49(D1), D605–D612.
- Tang, X., Wang, J., Zhong, J., & Pan, Y. (2014). Predicting essential proteins based on weighted degree centrality. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 11(2). Available from <https://doi.org/10.1109/TCBB.2013.2295318>.
- Taroni, J. N., Greene, C. S., Martyanov, V., Wood, T. A., Christmann, R. B., Farber, H. W., . . . Whitfield, M. L. (2017). A novel multi-network approach reveals tissue-specific cellular modulators of fibrosis in systemic sclerosis. *Genome Medicine*, 9, 27. Available from <https://doi.org/10.1186/s13073-017-0417-1>.
- Telford, W. G., Hawley, T. S., & Hawley, R. G. (2003). Analysis of violet-excited fluorochromes by flow cytometry using a violet laser diode. *Cytometry Part A*. Available from <https://doi.org/10.1002/cyto.a.10046>.

- Thibon, C., Pons, A., Mouakka, N., Redon, P., Méreau, R., & Darriet, P. (2015). Comparison of electron and chemical ionization modes for the quantification of thiols and oxidative compounds in white wines by gas chromatography-tandem mass spectrometry. *Journal of Chromatography A*, *1415*(123–133). Available from <https://doi.org/10.1016/j.chroma.2015.08.027>.
- Tørresen, O. K., Star, B., Mier, P., Andrade-Navarro, M. A., Bateman, A., Jarnot, P., et al. (2019). Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Research*, *47*(21), 10994–11006.
- Urban, P. L. (2016). Quantitative mass spectrometry: An overview. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2079). Available from <https://doi.org/10.1098/rsta.2015.0382>.
- Valet, G. (2003). Past and present concepts in flow cytometry: A European perspective. *Journal of Biological Regulators and Homeostatic Agents*, *17*(3), 213–222.
- Waiho, K., Afiqah-Aleng, N., Iryani, M. T. M., & Fazhan, H. (2021). Protein–protein interaction network: An emerging tool for understanding fish disease in aquaculture. *Reviews in Aquaculture*, *13*(1), 156–177.
- Walt, D. R. (2000). Bead-based fiber-optic arrays. *Science (New York, N.Y.)*, *287*(5452), 451–452.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63.
- Wang, Z., & Zhang, J. (2007). In search of the biological significance of modular structures in protein networks. *PLoS Computational Biology*. Available from <https://doi.org/10.1371/journal.pcbi.0030107>.
- Whitaker, J. W., Letunic, I., McConkey, G. A., & Westhead, D. R. (2009). metaTIGER: A metabolic evolution resource. *Nucleic Acids Research*, *37* (Database issue), D531–D538.
- Wilschefska, S., & Baxter, M. (2019). Inductively coupled plasma mass spectrometry: Introduction to analytical aspects. *Clinical Biochemist Reviews*, *40*(3), 115–133.
- Winter, D. R., Song, L., Mukherjee, S., Furey, T. S., & Crawford, G. E. (2013). DNase-seq predicts regions of rotational nucleosome stability across diverse human cell types. *Genome Research*, *23*(7), 1118–1129.
- Wong, E., Baur, B., Quader, S., & Huang, C. H. (2012). Biological network motif detection: Principles and practice. *Briefings in Bioinformatics*, *13*(2), 202–215.
- Wrzodek, C., Dräger, A., & Zell, A. (2011). KEGGtranslator: Visualizing and converting the KEGG PATHWAY database to various formats. *Bioinformatics (Oxford, England)*, *27*(16), 2314–2315.
- Wu, Z., Liao, Q., & Liu, B. (2020). A comprehensive review and evaluation of computational methods for identifying protein complexes from protein-protein interaction networks. *Briefings in Bioinformatics*, *21*(5), 1531–1548.
- Xing, S., Wallmeroth, N., Berendzen, K. W., & Grefen, C. (2016). Techniques for the analysis of protein-protein interactions in vivo. *Plant Physiology*. Available from <https://doi.org/10.1104/pp.16.00470>.
- Xu, A., Cheng, J., Sheng, H., Wen, Z., Lin, Y., Zhou, Z., et al. (2019). Clinical management and gene mutation analysis of children with congenital hyperinsulinism in south China. *Ocrinology*, *11*(4), 400–409.
- Xu, D. J., & Noyes, M. B. (2015). Understanding DNA-binding specificity by bacteria hybrid selection. *Briefings in Functional Genomics*, *14*(1), 3–16.
- Xu, J., & Lan, Y. (2015). Hierarchical feedback modules and reaction hubs in cell signaling networks. *PLoS One*, *18*(13), 1031–1043.
- Yan, W., Zhang, D., Shen, C., Liang, Z., & Hu, G. (2018). Recent advances on the network models in target-based drug discovery. *Current Topics in Medicinal Chemistry*, *18*(13), 1031–1043.
- Yu, D., Kim, M., Xiao, G., & Hwang, T. H. (2013). Review of biological network data and its applications. *Genomics and Informatics*, *11*(4), 200–210.
- Yuan, L., Guo, L. H., Yuan, C. A., Zhang, Y., Han, K., Nandi, A. K., et al. (2019). Integration of multi-omics data for gene regulatory network inference and application to breast cancer. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *16*(3), 782–791.
- Zaki, N., Efimov, D., & Berengueres, J. (2013). Protein complex detection using interaction reliability assessment and weighted clustering coefficient. *BMC Bioinformatics*, *14*, 163. Available from <https://doi.org/10.1186/1471-2105-14-163>.
- Zamal, F. Al, & Ruths, D. (2012). On the contributions of topological features to transcriptional regulatory network robustness. *BMC Bioinformatics*, *13*, 318. Available from <https://doi.org/10.1186/1471-2105-13-318>.
- Zhang, P., & Itan, Y. (2019). Biological network approaches and applications in rare disease studies. *Genes*, *10*(10), 797. Available from <https://doi.org/10.3390/genes10100797>.
- Zhang, P., Tao, L., Zeng, X., Qin, C., Chen, S., Zhu, F., et al. (2017). A protein network descriptor server and its use in studying protein, disease, metabolic and drug targeted networks. *Briefings in Bioinformatics*, *18*(6), 1057–1070.
- Zhao, J., & Lei, X. (2019). Predicting essential proteins based on second-order neighborhood information and information entropy. *IEEE Access*, *7*, 136012–136022. Available from <https://doi.org/10.1109/ACCESS.2019.2942843>.
- Zhou, G., & Xia, J. (2018). OmicsNet: A web-based tool for creation and visual analysis of biological networks in 3D space. *Nucleic Acids Research*, *46*(W1), W514–W522.
- Zhou, Y., Liu, Z., Welch, J. D., Gao, X., Wang, L., Garbutt, T., et al. (2019). Single-cell transcriptomic analyses of cell fate transitions during human cardiac reprogramming. *Cell Stem Cell*, *25*(1), 149–164. Available from <https://doi.org/10.1016/j.stem.2019.05.020>, e9.
- Zhu, C., Kushwaha, A., Berman, K., & Jegga, A. G. (2012). A vertex similarity-based framework to discover and rank orphan disease-related genes. *BMC Systems Biology*, *6*, S8. Available from <https://doi.org/10.1186/1752-0509-6-S3-S8>.