We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Computational Systems Analysis on Polycystic Ovarian Syndrome (PCOS)

Nor Afiqah-Aleng and Zeti-Azura Mohamed-Hussein

Abstract

Complex diseases are caused by a combination of genetic and environmental factors. Unraveling the molecular pathways from the genetic factors that affect a phenotype is always difficult, but in the case of complex diseases, this is further complicated since genetic factors in affected individuals might be different. Polycystic ovarian syndrome (PCOS) is an example of a complex disease with limited molecular information. Recently, PCOS molecular omics data have increasingly appeared in many publications. We conduct extensive bioinformatics analyses on the data and perform strong integration of experimental and computational biology to understand its complex biological systems in examining multiple interacting genes and their products. PCOS involves networks of genes, and to understand them, those networks must be mapped. This approach has emerged as powerful tools for studying complex diseases and been coined as network biology. Network biology encompasses wide range of network types including those based on physical interactions between and among cellular components and those baised on similarity among patients or diseases. Each of these offers distinct biological clues that may help scientists transform their cellular parts list into insights about complex diseases. This chapter will discuss some computational analysis aspects on the omics studies that have been conducted in PCOS.

Keywords: polycystic ovarian syndrome, PCOS, systems biology, computational systems biology, protein-protein interaction analysis, network biology, pathway analysis

1. Introduction

Findings have shown that most pathological conditions and diseases involve genetic components, in diseases such as cystic fibrosis, hemophilia, and sickle cell disease, are caused by mutations in a single gene [1–3]. However, there are many other common medical problems such as cardiovascular diseases, diabetes mellitus, obesity, and polycystic ovarian syndrome (PCOS), which are not caused by single mutations [4–7]. The etiologies of those problems are much more complex where these disorders are highly associated with multiple genes/proteins in combination with multifactor including genetics, environment, and lifestyle. Many efforts have been done to overcome the complexity of these medical problems.

Studying diseases at the molecular level is one of the efforts in understanding complex diseases. The emergence of biological technology has yielded great advances in deciphering the pathobiology of diseases by generating numerous large omics (genomics, transcriptomics, proteomics, and metabolomics) datasets. These data capture a wide range of disease phenomena including mutations, gene expression, protein expression, metabolite profiling, and genetic and physical interactions between biological molecules, where each dataset offers distinctive of knowledge to understand the diseases. Complex diseases are insufficient by a single level independent omics dataset since those diseases are regulated at multiple systems levels. They can be manifested by integrated omics analysis (integration of multi-omics data).

The multi-omics analysis has brought a new challenge to develop methods or pipelines, statistics, algorithms, and tools for integration, and the assistant of computational systems analysis is in great need. Implementing integrative analysis on these multiple omics data is the best way in deriving systematical and comprehensive views of diseases, achieving a better understanding of disease mechanisms and finding operable personalized health treatments. With the help of computational systems analysis, research in the field of biology and biomedicine has gained tremendous benefits over the past few decades.

Computational systems analysis connects interdisciplinary perspectives with mathematical, algorithms, statistical, modeling and simulations, data repository, and/or network visualizations using computational technique to investigate certain biological phenomenon or condition in a systems view. Currently, there are many studies on the integrated omics data and used network biology, which is one of the main techniques in computational systems analysis to obtain an overview at the systems level in elucidating the pathobiology of human diseases. Network biology could systematically connect all the molecules generated from the omics studies that have been identified to be related to the disease. Other than network biology, there are studies that used simulations approach to have a better understanding of diseases. Database development is another computational systems analysis that serves to provide overall information about the diseases. This chapter encompasses the computational systems analysis, such as network biology, simulations, and data repository, which have been used to understand the pathobiology of human diseases, particularly in PCOS.

2. Network biology in disease

Early biological experiments revealed that proteins, as the main agents of biological function, determine the phenotype of all organisms. In the advent of molecular biology, it is assumed that proteins do not naturally function in isolated forms; instead, they have interactions with one another and also with other molecules (e.g., DNA, RNA, and metabolites) that mediate metabolic, signaling and regulatory pathways, cellular processes, and organismal systems [8]. Most of the biological characteristics or phenotypes arise from the complex interactions between the cell's numerous constituents [9]. Any interruptions to the interactions between those molecules can disturb the normal behavior of the cells and contribute to the medical problems or diseases [10]. Thus, studies on network biology in disease are essential as it can be used to detect interrupter biological events since the network biology plays a role to perceive the biological role within the cells [11].

In network biology, there are two types of analysis that often be performed to understand the pathobiology of diseases, that is, protein-protein interaction and pathway analysis.

2.1 Protein-protein interaction analysis

Protein is a biological molecule that plays an important role in the molecular process in a cell. It acts as an enzyme for metabolic reaction, DNA replication, molecular transporter, antigen defensive system, and cell to cell information transmission [12]. Proteins physically interact with each other to perform a biological function in a cell. Protein-protein interaction (PPI) has become a valuable approach to study the molecular mechanisms of disease [13]. For example, non-metastatic and metastatic breast tumors, as well as the markers of metastasis, have been classified and identified by a network-based method. Based on this study, the said method is more effective because it enables detection of the genes that play a role in metastasis, which is not otherwise picked up during differential expression analysis [14]. Protein networks for type-1 diabetes were constructed by integrating GWAS data with the information from protein-protein interaction databases. Eight new genes were subsequently identified, hence providing better knowledge of the mechanism of type-1 diabetes [15]. Besides, new pathways have been defined from the protein network-based Huntington, giving a deeper understanding of the pathogenesis of Huntington disease [16]. These studies indicate that a network, particularly that of proteins, could be one of the powerful tools in understanding the molecular basis of diseases. Thus, this method could be applied to unveil the molecular basis of PCOS.

There are several approaches including yeast-2-hybrid (Y2H) and mass spectrometry (MS) that have been used to identify the PPI [17–20]. All approaches have generated large interactome and progressively identified PPI network in several organisms such as virus (herpes virus) [21], prokaryote (*Escherichia coli*) [22], eukaryote (yeast) [18–20], nematode [23], fruit fly [24], and human [25, 26]. These PPI datasets have been compiled and stored in PPI databases such as Biological General Repository for Interaction Datasets (BIOGRID) [27], Database of Interacting Proteins (DIP) [28], GeneMANIA [29], Human Integrated Protein-Protein Interaction Reference (HIPPIE) [30], Human Integrated Protein-Protein Human Protein Reference (HPRD) [31], Interologous Interaction Database (I2D) [32], IntAct [33], MIPS Mammalian Protein-Protein Interaction Database (MIPS) [34], Molecular Interaction database (MINT) [35], and STRING [36] (**Table 1**).

Combination of PPI forms a network consists of two main components, i.e. (1) node that represents protein and (2) edge that refers to interaction (**Figure 1**). PPI network has been applied for evolutionary study [45], gene/protein functional prediction [46], and also pathobiology of diseases [47, 48]. There are few analyses that can be applied using the PPI network approach, and PPI network topological analysis is one of the analyses that often are used to study the pathobiology of human diseases. Degree distribution, which is a fraction of a number of the interaction of a node with the number of the interactions in a network, is one of the components in the network topology that have been measured. A node that has a high degree distribution is known as a hub protein. A hub protein is hypothesized to code an essential gene that plays an important role in a cell. Any physical or chemical alterations that occur to this hub protein can interrupt the interaction with other proteins, disturb the normal behavior of the cells and associated to a disease. Previous study by Wachi et al. found that proteins encode for the upregulated genes in the lung squamous cell carcinoma tend to have higher degree distribution [49]. Jonsson and Bates also found

Database	Description	URL	Reference
Agile Protein Interactomes DataServer (APID)	Is a collection of integrated known validated protein interactomes from 400 species, including humans	http://apid.dep.usal.es/	[37]
BioGRID	Curates sets of genetic, physical, and chemical interactions in humans and all major organisms	https://thebiogrid.org/	[27]
Database of Interacting Protein (DIP)	Is manually curated by expert curators. This database stores experimental PPI data	https://dip.doe-mbi.ucla.edu	[38]
GeneMANIA	Contains known PPI interactions, which are derived from curated PPI databases and experiments. Predicts PPI interactions in six species, including humans	https://genemania.org/	[29]
Human Integrated Protein-Protein Interaction rEference (HIPPIE)	Contains only human PPI interactions from experiments and curated PPI databases	http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/	[30]
Information Hyperlinked over Proteins (iHOP)	Provides PPIs curated from literature mining	https://bio.tools/ihop	[39]
Human Protein Reference Database (HPRD)	Stores information regarding proteins in humans, including PPIs	https://www.hprd.org/	[31]
Interologous Interaction Database (I2D)	Contains integrated known, experimental, and predicted PPIs for humans and five other species	http://ophid.utoronto.ca/ophidv2.204/	[32]
InnateDB	Stores experimentally verified interactions between genes, proteins, and signaling pathways involved in the innate immune response to microbial infection in humans, mice, and bovines	https://www.innatedb.com/	[40]
IntAct Molecular Interaction Database (IntAct)	All interactions are retrieved from literature and eleven other PPI databases. Is curated by EMBL-EBI and other PPI database teams	https://www.ebi.ac.uk/intact/	[33]
Molecular INTeraction database (MINT)	Compiles experimentally verified PPIs curated from the scientific literature	https://mint.bio.uniroma2.it/	[41]
STRING	Is a PPI database that provides interaction evidence from known interactions (curated databases and experiments), predicted interactions (neighborhood, gene fusion, and co-occurrence), and others (coexpression and text-mining)	https://string-db.org/	[36]
The Extracellular Matrix Interaction Database (MatrixDB)	Stores interactions of extracellular matrix proteins, proteoglycans, and polysaccharides	http://matrixdb.univ-lyon1.fr/	[42]
The Human Protein Interaction Database (HPID)	Compiles proteins from BIND (this PPI database is not open access), DIP, and HPRD and predicts potential PPIs	http://wilab.inha.ac.kr/hpid/	[43]

Database	Description	URL	Reference
The International Molecular Exchange Consortium (IMEx)	Provides nonredundant PPI datasets from major PPI databases like BIND, IntAct, MINT, DIP, and MIPS	https://www.imexconsortium.org/	[44]
The Mammalian Protein-Protein Interaction Database (MIPS)	Compiles high-quality PPIs from experiments	http://mips.helmholtz-muenchen.de/proj/ppi/	[34]

Table 1.
List of PPI databases concerning humans.

346 proteins-related to cancer have two times higher degree connectivity compared to the non-cancer proteins [50]. Number of interaction among proteins that are related to disease in the Online Mendelian Inheritance in Man (OMIM) Morbid Map is higher than the interaction of non-disease proteins [51].

Linkage method is another network analysis that can be used to understand the pathobiology of human diseases [50, 52]. The basic hypothesis in this method is the two proteins (pairwise linkage) that interact with each other tend to be related to the same diseases. Enrichment analysis done by Oti et al. [53] has demonstrated that the proteins that interact with each other are significantly associated with the same diseases. By a pairwise linkage method, they also predicted that Janus kinase 3 (JAK3) as a protein that might be associated with severe combined immunodeficiency syndrome (SCID) as JAK3 directly interacts with proteins of lymphocyte specific protein-tyrosine (LCK), protein-tyrosine phosphatase (PTPRC), and interleukin 2 receptor (IL2RG) [53].

Clustering is also a technique in network analysis in human diseases. A cluster refers to a small group that has similar topological network properties [48, 54]. In this method, it is hypothesized that the proteins in a module tend to be associated with the same diseases. Clusters are identified using algorithms, and there are several clustering algorithms that have been developed to generate the clusters such as CFinder [55], clustering with overlapping neighborhood expansion (ClusterONE) [56], clustering based on maximal cliques (CMC) [57], clique percolation method (CPM) [58],
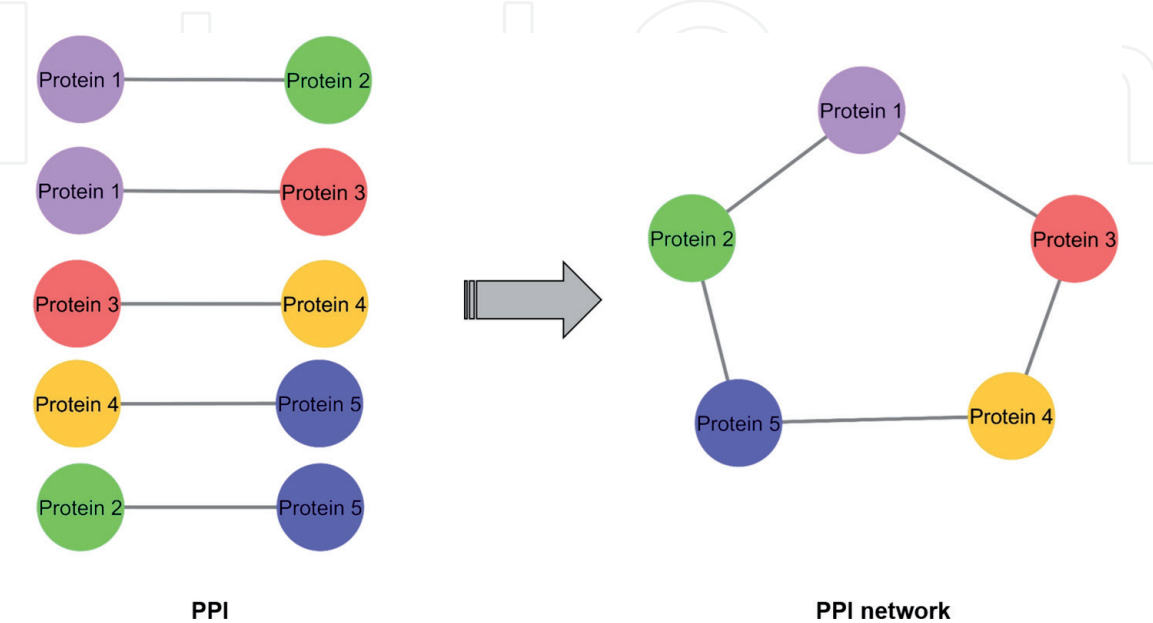

Figure 1.
PPI and PPI network. PPI network is formed from several PPI. The color of the nodes refers to the protein.

density-periphery based clustering (DPClus) [59], density-periphery overlapping-based clustering (DPClusO) [60, 61], identifying protein complex algorithm (IPCA) [62], local clique merging algorithm (LCMA) [63], restricted neighborhood search clustering (RNSC) [64], Markov clustering (MCL) [65], molecular complex detection (MCODE) [66], and so on. Wu et al. developed the clustering algorithm to identify the clusters, and they found that the proteins in the same clusters are associated with the same diseases [67]. Rezaei-Tavirani et al. identified clusters using ClusterONE algorithm to search for potential biomarkers in esophagus adenocarcinoma [68]. Xiao et al. also used clustering methods to identify the candidate proteins for endometriosis biomarkers by their own clustering algorithm and they found the majority of predicted biomarkers in the generated clusters involved in endometriosis pathway [69].

PPI network can be also applied to understand the association between diseases as clinically and there is the occurrence of comorbidity, which is a condition of a patient that is simultaneously affected more than one disease. Disease association network based on PPI analysis can be used as a framework to classify the disease, identify the risk of having other diseases, predict the effect of disease, and search for a more effective therapeutic technique for disease [70, 71]. There are few hypotheses in constructing the disease association network, and one of them is diseases can be associated when those diseases shared the proteins and interactions (**Figure 2**).

The components in a disease association network are similar to the PPI network. It consists of node and edge, where node refers to disease and edge is the interaction of disease. The first human disease network has been constructed among 867 diseases using PPI information by Goh et al. (**Figure 3**) [73]. This network has been used to understand how diseases comorbid to each other by identifying the shared proteins and interactions between the diseases [72]. The disease association network is also useful to predict the disease biomarkers. Ahmed et al. have successfully identified 73 potential biomarkers for neurological diseases, that is, Alzheimer's disease, epilepsy, and dyslexia, by integrating the protein-disease association with the PPI information [74].
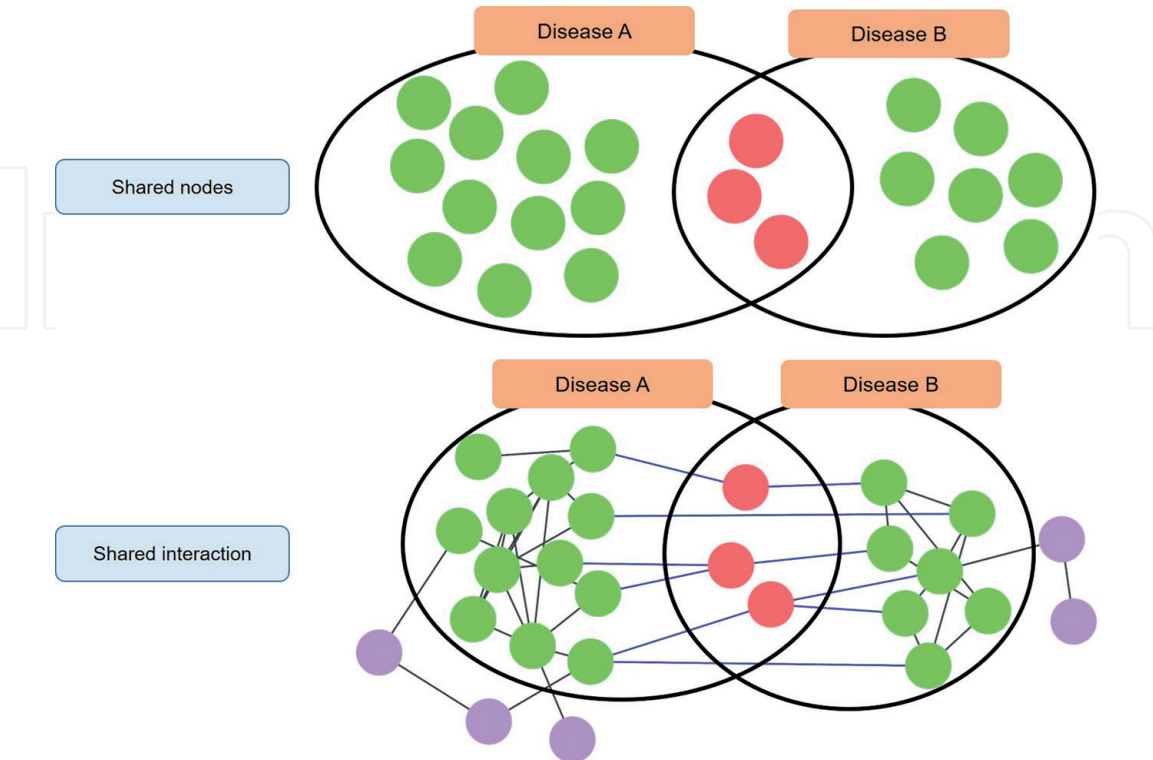

Figure 2.
Two different approaches to identify the association between diseases. The first approach used shared nodes (red nodes) between diseases. The second approach used shared interactions (blue edges) between disease-related proteins [72].
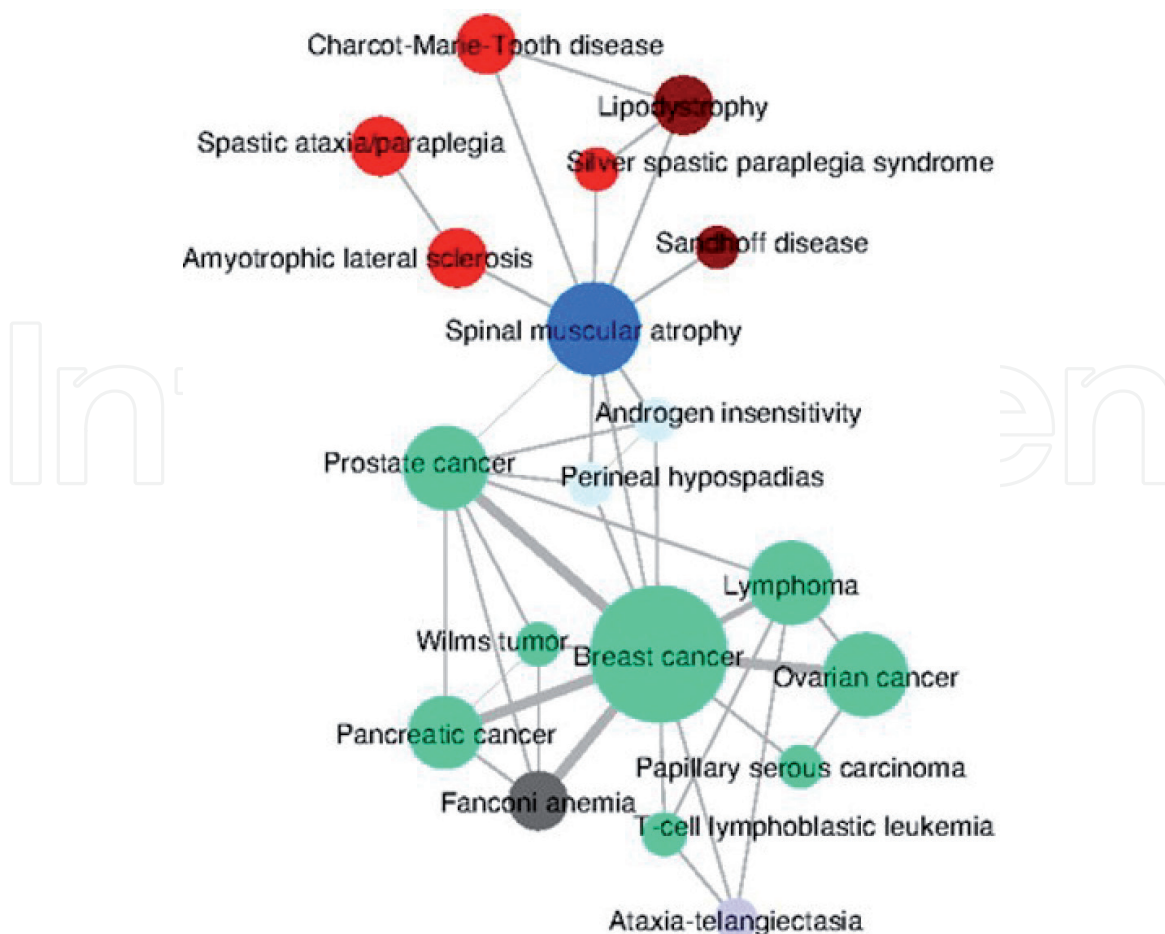
Figure 3.
Human disease association network. This network was constructed by Goh et al. using PPI information [73].
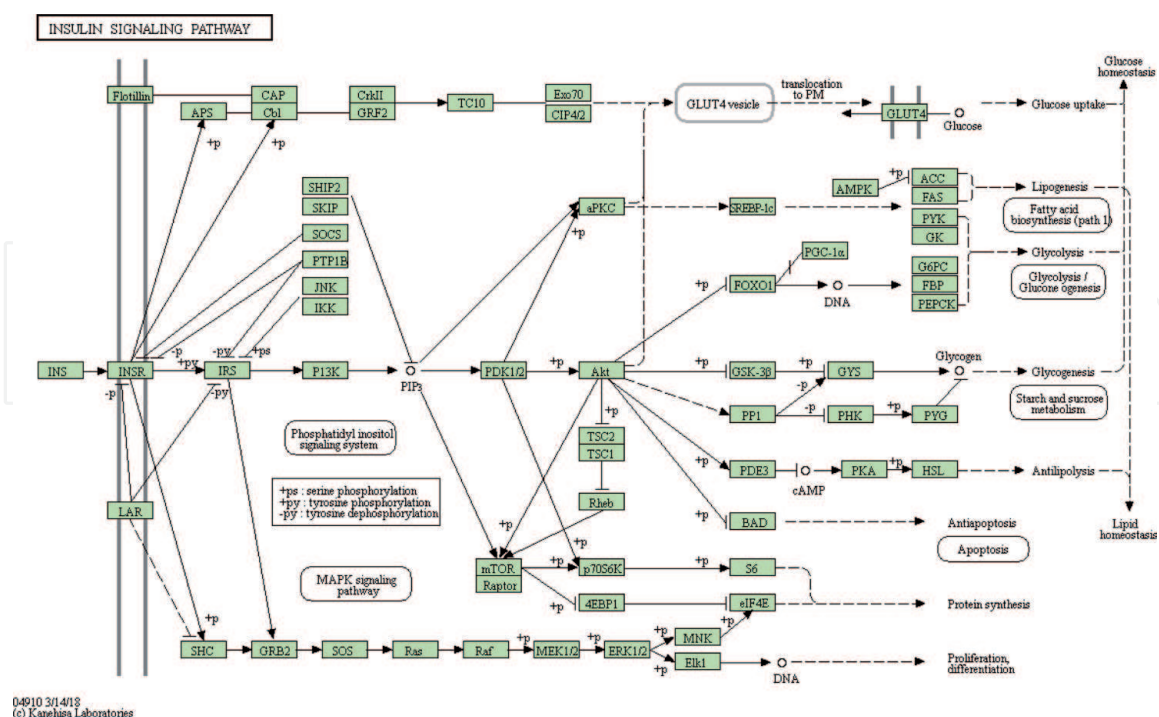

Figure 4.
Example of a biological pathway. This is the insulin signaling pathway that was retrieved from the KEGG database [80].

To summarize, PPI analysis is a powerful approach that can be applied to improve the understanding of the pathobiology of diseases, which in turn can appraise approaches to diagnose, prevent, and treat the diseases. The analysis of

network properties can provide the opportunity to interpret the normal and altered biological behaviors that lead to diseases.

2.2 Pathway analysis

A pathway is a group of molecules that interact to perform the same biological function. PPI network has a type of node that is protein and the undirected network (**Figure 4**). Meanwhile, the pathway consists of a few types of nodes, which are signaling genes, proteins, complex, and metabolites, which are connected by several interactions such as activation, inhibition, binding, and others. A pathway depicts a mechanism in performing a specific biological activity in a cell. As a PPI network, a combination of several pathways forms a pathway network. Pathway information can be retrieved from several pathway databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [75], Reactome [76], WikiPathways [77], BioCyc [78], and BioCarta [79]. **Table 2** shows databases that have pathway information in human.

There are three main types of pathways, that is, signaling, regulatory, and metabolic pathways. Signaling pathway visualizes the cellular response after receiving the extracellular signal. The signal transmission starts when the extracellular gives a signal to activate the receptor that is located in the cell surface. The activated receptor will bind to the signal and alter the intracellular molecules to respond [84]. Any disruption in the signaling pathway can cause disease since the cells cannot be normalized or properly respond when the signals are received [85]. Regulatory pathway displays the gene or protein expression in a cell, either it is upregulated or downregulated. The biological activities such as transcription, translation, and post-translational modification are among the activities that involve the regulatory pathway [86]. Meanwhile, in a metabolic pathway, the primer metabolite will be modified into another metabolite through a series of chemical reactions catalyzed by enzymes [87].

Pathway database such as KEGG also provides pathways that visualize the mechanisms of several complex diseases such as cancer, diabetes mellitus, Alzheimer's disease, Parkinson's disease, and so on [75]. Basically, a complex disease involves several pathways that include all signaling, regulatory, and metabolic pathways. The combination and the integration of several pathways with other types of data such as PPI is one of the valuable approaches that can be used to improve the understanding of complex disease mechanisms [88].

As an analogy, it is essential to have a diagram such as a circuit diagram for an electrician to understand the principle of electricity. A diagram such as a biological network is also important in the medical field to assist the researchers or clinicians to understand the mechanisms of diseases. The biological network can suggest a novel means of developing molecular therapies where the network is the target of therapy rather than individual molecules within the network.

3. Modeling and simulation

Mathematical modeling and computer simulation are another computational systems analysis that has been used to study disease progression and drug development [89, 90]. While the biological network is generally constructed in the static state using annotated genes, proteins, and metabolites and linked these molecules using information from PPI and pathway databases, modeling and simulation are constructed in quasi-steady state, where they require additional data including physicochemical and physiological balances and bounds (mass and energy conversion) [91]. Modeling and simulation have been widely used in several chronic

Database	Description	URL	Reference
BioCarta	A pathway database that provides gene interaction within pathways for human cellular processes	https://cgap.nci.nih.gov/Pathways/BioCarta_Pathways	[79]
HumanCyc	One of the pathway databases in BioCyc that consists of human metabolic pathways	https://humancyc.org/	[81]
Kyoto Encyclopedia of Genes and Genomes (KEGG)	A repository to understand the high-level functions and utilities of the biological system of several organisms including humans that were obtained from genome sequencing and other high-throughput experimental technologies	https://www.kegg.jp/	[80]
PANTHER	Contains several biological information such as pathways of proteins coupled with tools for protein analysis for several organisms including humans	http://pantherdb.org/	[82]
Reactome	A pathway database for several organisms including humans that provides tools for pathway analysis	https://reactome.org/	[83]
Wikipathways	A pathway database that contains pathway information of several organisms such as humans	https://www.wikipathways.org/index.php/WikiPathways	[77]

Table 2.
Lists of human pathway database.

diseases such as diabetes, Alzheimer's disease, coronary heart disease, and infectious diseases such as meningitis and influenza [89, 92–95].

In this approach, there are several types of models that have been applied to understand the human diseases, which are pharmacokinetics (PK) model, pharmacokinetics/pharmacodynamics (PKPD) model, disease progression model, metamodel, and Bayesian model averaging [90]. PK model is widely used in the field of clinical pharmacology as it simulates the rate and extent of drug distribution to different tissues and the rate and impact of drug disposition. It is a very important model as it predicts the impact variability in target patient populations in response to drug administration [96]. PKPD model is another model in the drug development where it integrates PK and PD components. This model establishes and measures the relationships of dose-concentration-response and describes and predicts the effect-time courses in consequence of a drug dose [97]. Meanwhile, the disease progression model is the time course quantitative descriptor of disease status. It was first simulated in 1992 in Alzheimer's disease using the cognitive component of the Alzheimer's disease assessment scale (ADASC) to assess the disease severity [93]. This model characterizes the natural progression of the disease by incorporating biomarkers of disease severity and/or clinical outcomes. Disease progression model is often used to quantify the effects of drug treatment on disease progression by integrating with PK and PKPD models [98]. Metamodel involves model development by combining results from multiple previous studies. In human disease study, this model can be used to compare the effects and safety of new treatments with other treatments, to reevaluate data of mixed or different result situations, and to describe PD or disease progress models [90, 99]. In the meantime, Bayesian model averaging combines models as there is a situation where previous studies show several models for a drug in a certain disease, and it is unclear which model is suitable. The Bayesian model averaging reduces the uncertainty by allowing all existing models to contribute to a simulation with weighing the inputs on the basis of certain criteria such as the quality of data or model [90, 92].

Complex diseases involve many genes, proteins, and metabolites, and these molecules are either activated or deactivated in certain tissues in particular time, depending on the disease status or in the influences of several factors such as drug administration. Hence, modeling and simulation are efficient approaches in the computational systems analysis as these approaches manage to dynamically monitor and understand the progress of diseases in particular situation, which in turn can assist in improving the specific treatment and developing the efficient drugs for complex human diseases.

4. Data repository

Data are the most important resource in computational systems analysis. Most of the analyses require the integration of several data to understand the diseases, particularly complex diseases in a systemic view. For example, several omics data (genomics, transcriptomics, proteomics, and/or metabolomics) were integrated with interactions data (PPI or pathway) to construct network biology. Modeling and simulation also involve omics data integration to capture the complexity of molecular events causing the diseases. In addition, cellular and physiological processes are complex systems [100] that are controlled by signals from the extracellular environment and coordinated by intracellular interaction and transcriptional or gene regulatory networks assembled into functional modules [101]. In order to understand cellular processes as interconnected and interdependent systems and in the context of a biological phenomenon, requires an integrative approach that

draws upon data from as many diverse data sources as possible including data from the literature, public databases, biochemical and kinetic experiments, phenotype studies and high-throughput analyses of the genome, transcriptome, proteome, interactome, and metabolome.

Hence, data repository or database development is one of the main approaches to facilitate the arbitrary querying of the data to perform the computational systems analysis. Besides, recent developments in high-throughput approaches enable the analysis of the transcriptome, proteome, interactome, metabolome, and phenome on a previously unprecedented scale, thus contributing to the deluge of experimental data and scattering in an unorganized way. The data repository is one of the efforts in combining the growing sets of experimental data in a proper way that can be publicly accessed to have further analysis. For example, there are databases such as ArrayExpress [102], Gene Expression Omnibus (GEO) [103], and CIBEX [104] that stores datasets for gene expression studies to be publicly accessed. Other than that, there are also literature databases such as PubMed, Scopus Online, and Google Scholar for the researchers to retrieve published studies, and there are several studies that provide the generated omics datasets in the supplementary section.

There are also databases such as disease databases that have performed several analyses prior to deposit the data into the database. The human disease databases have been developed in order to store information about diseases such as genes, proteins, metabolites, drugs, literature, biological processes, tissues, and others that are related to a particular disease in order to understand the pathobiology, pathogenesis, and pathophysiology of diseases. Currently, databases, such as DisGeNET [105], MalaCards [106], Online Mendelian Inheritance in Man (OMIM) [51], Open Targets [107], GWAS Catalog [108], GWASdb [109], DISEASES [110], and Human Gene Mutation Database (HGMD) [111], have been developed to store several information about human diseases. There are also databases that have been developed that specifically store data or information of a disease such as T2D-Db [112] and T2D@ ZJU [113] for type-2 diabetes, AlzBase [114], AlzGene [115], and NIAGADS [116] for Alzheimer's disease, and The Cancer Genome Atlas (TCGA) [117] and The International Cancer Genome Consortium (ICGC) [118] for human cancers.

Nowadays, the number of databases that hold a growing number of generated data is also increased, which has led to a new challenge in selecting the best and suitable database for further computational systems analysis. Nevertheless, the presence of current available data repository or databases has eased the researchers without having to extensively search the data to integrate the data and visualize the data into a network and/or model in order to harness a comprehensive systems-level understanding of pathophysiological processes of human diseases.

5. Computational systems analysis progress in PCOS

PCOS is a heterogeneous disorder that may be affected by multiple factors including genetic, lifestyle, and environment. The definition of PCOS is unclear, where it is defined by a combination of different features that lead to its diagnostic criteria remain controversial. PCOS women also experience multi-symptoms, and the diseases that comorbid to PCOS are widely varied [6, 119]. The complexity in PCOS is evident that many genes, proteins, and metabolites involved in the pathobiology of PCOS. All omics platforms have been applied to identifying the molecular basis of PCOS (**Table 3**) [120].

Even though all omics have been performed in PCOS, the pathobiology of PCOS is still far from understood. Since the prevalence of PCOS women is increased and if they are left untreated, PCOS women are at higher risk to develop other

Omics	Description	Examples of previous studies	Reference
Genomics	Identification genetics evidence in PCOS women	Identified 16 loci associated risk of PCOS in Chinese and European subjects	[121–124]
Transcriptomics	Identification of differentially expressed genes (significantly up-regulated and down-regulated genes) between non-PCOS and PCOS women	Identified 243 differential expressed gene in the granulosa cells between non-PCOS and PCOS patients	[125]
Proteomics	Detection of differentially expressed protein between non-PCOS and PCOS women	Identified 186 significantly expressed proteins in the follicular fluid between non-PCOS and PCOS women	[126]
Metabolomics	Detection of altered metabolites between non-PCOS and PCOS women	The altered metabolites in the sera between non-PCOS and PCOS women revealed disruptions in several metabolic pathways such as steroid hormone biosynthesis, amino acids and nucleotides metabolism, and glutathione metabolism, as well as lipids and carbohydrates metabolism	[127]

Table 3.
Omics approaches in PCOS.

chronic diseases (endometrial cancer, type-2 diabetes and cardiovascular diseases), and other approaches such as computational systems analysis need to be done to improve the understanding in PCOS. By far, several studies have integrated the omics platforms using computational systems analysis to provide a systems-level understanding of PCOS.

5.1 PPI and pathway analysis

In PCOS, PPI- and pathway-based analysis is also often used to identify the genes/proteins, ontologies, and pathways that might be involved in this disorder. Among the earliest full-paper study in using PPI analysis in PCOS was published in 2009. In this work, Mohamed-Hussein and Harun combined seven microarray datasets and integrated with PPI information and successfully identified a hypothetical protein, C1ORF123, and several ontologies that might be highly involved in PCOS [128]. Prior to this study, there is an article outline in 2007 by Menke et al. that used a Newman algorithm to identify the small set of modules in the constructed PCOS PPI network that could lead to PCOS phenotypes [129]. Shen et al. [130] have constructed the regulatory network and PPI network by integrating several data such as genome-wide methylated DNA immunoprecipitation (MeDIP), regulatory interactions and PPI to investigate the relationship of insulin resistance (IR) with PCOS. In a regulatory network, the significant methylated genes, CCAAT enhancer binding protein beta (CEBPB) formed a network that regulated other genes that may play a role in both IR and PCOS. Meanwhile, the constructed PPI network showed that the methylated genes in PCOS-IR have a higher number of interactions and might act as key drivers to perform proper cellular functions. Shen et al. [130] also found several enriched pathways such as cancer pathways and MAPK signaling and ontologies including regulation of metabolic process from both constructed networks that might be responsible in both PCOS and IR [130]. Shim et al. used pathway-based analysis on genome-wide

association study (GWAS) dataset of PCOS and successfully identified several PCOS pathways associated with ovulation and insulin secretion [131].

Kori et al. [132] used PPI and pathway analysis by integrating three microarray datasets of PCOS with PPI data, performing the pathway enrichment analysis and comparing the PCOS results with ovarian cancer and endometriosis. These analyses found that PCOS is closely related to endometriosis and ovarian cancer as they shared several molecules and pathways such as MAPK signaling, cell cycle, and apoptosis [132]. The integration of a microarray dataset with PPI information from REACTOME has found several proteins including Rho GTPase activating protein 4 (ARHGAP4), Rho GTPase activating protein 9 (ARHGAP9), ras homolog family member G (RHOG) and LYN proto-oncogene, Src family tyrosine kinase (LYN), and pathways such as RhoA-related pathways, and glycoprotein VI-mediated activation cascade might involve in the PCOS pathogenesis [133].

Other than identifying the molecular basis and the biological functions that might relate to PCOS, PPI and pathway analysis are also applied to decipher the molecular relationship of PCOS with other diseases and improve the knowledge on PCOS treatments. Liu et al. [134] construct a PPI network, which consists of PCOS-related genes and target genes of Erxian decoction (EXD) to understand the pharmacological basis of the EXD action in treating PCOS. EXD is a traditional Chinese medicine composed of six types of herbs that can alleviate several problems such as ovarian failure, which is a problem that commonly experiences by PCOS women. In the constructed network, Liu et al. [134] identified 50 genes that might be key genes that involved in PCOS treatment with EXD since these genes are the EXD targets that are found to be related to PCOS [134]. Ramly et al. [135] also used PPI and pathway analysis to identify protein and pathways to explain the relationships between PCOS and 17 diseases such as migraine, ovarian cancer, and schizophrenia. They used a clustering approach by MCODE [66] to identify shared proteins between PCOS and other diseases and pathway enrichment analysis to identify pathways that might connect PCOS and PCOS-associated diseases [135].

Based on aforementioned studies, it is proved that PPI- and pathway-based can be used to identify genes/proteins, biomarkers, ontologies, and pathways that are related to PCOS, which in turn could improve the diagnosis and treatment in PCOS.

5.2 Data repository in PCOS

As mentioned, there are many datasets that have been generated by the omics platforms to identify the pathobiology of PCOS. The datasets are randomly distributed, and it is very tedious if the researchers intend to retrieve the information about PCOS. Hence, it is essential to have a repository that stores comprehensive information on PCOS.

There are three databases that have been developed by far to deposit the collated molecular information generated by previous studies, which are PCOSBase (www.pcosbase.org) [136], PCOSKB (http://pcoskb.bicnirrh.res.in/) [137], and PCOSDB (http://pcosdb.net/) [138]. Both PCOSKB and PCOSDB contain 241 and 208 genes that related to PCOS, respectively. These databases searched for the PCOS-related genes against scientific literature. Meanwhile, PCOSBase identified 8185 PCOS-related proteins that were obtained from previous disease databases and gene and protein expression studies. All of the PCOS databases provided detailed description for each entry that is related to PCOS and link to the original databases such as UniProt (https://www.uniprot.org/) and NCBI (https://www.ncbi.nlm.nih.gov/) for extensive information. As PCOSBase, biological information such as chromosomal location, gene ontologies, pathways, domains, disease-associated, and tissue localization have been annotated to all PCOS-related proteins. **Figure 5** shows the
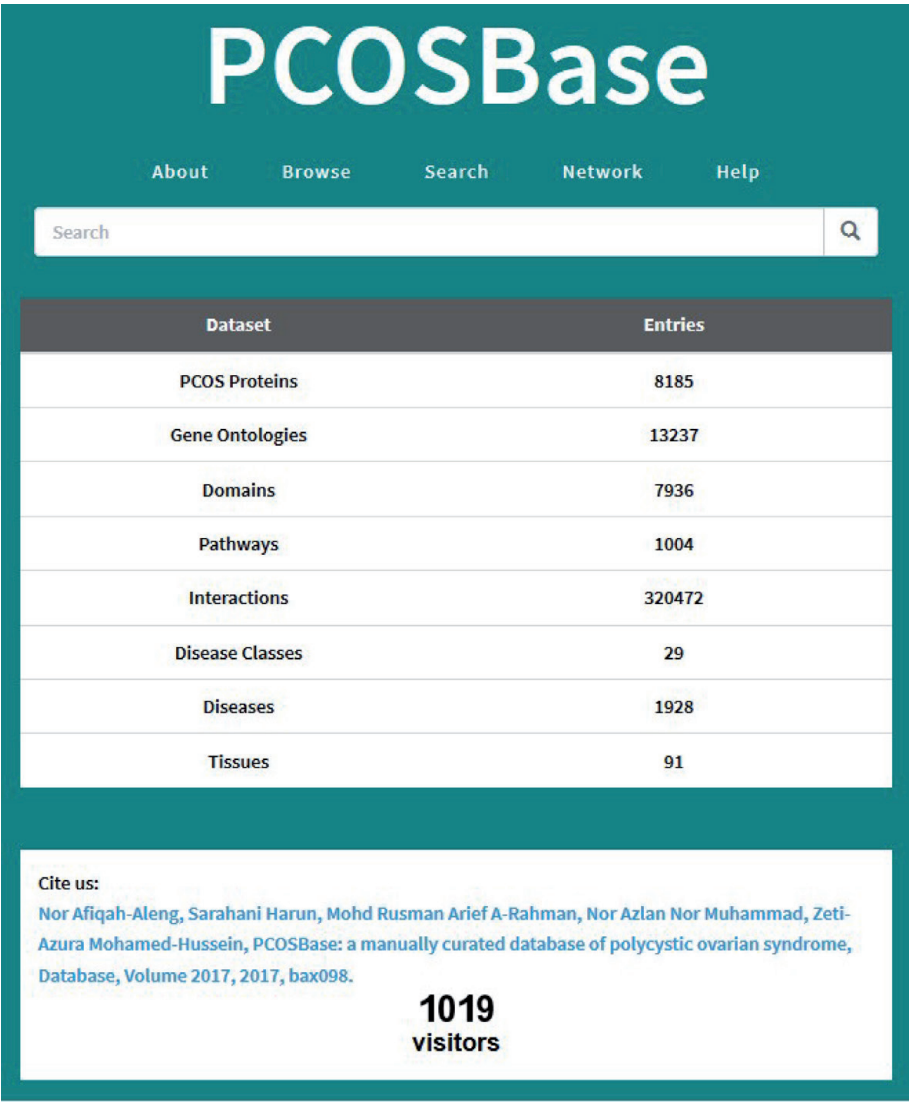
Figure 5.
PCOSBase homepage.

example homepage of PCOSBase, where it provides search box to facilitate the users to search with keywords and shows the number of entries for each functional details that are deposited in the database.

All of these databases are developed as an effort for other researchers in identifying PCOS biomarkers. Besides, the information from the databases has been used to integrate with other information such as PPI and pathway to have a systems-level view of PCOS. PCOSBase provided a menu ("Network") that contained a biological network of PCOS as examples of analysis on the PCOS-related proteins from this database. The network provided in the database can give an insight into improving the knowledge, particularly in PCOS.

6. Conclusion and future perspective

PCOS is an endocrine disorder that linked many clinical symptoms and the diversity of diseases. The PCOS complexity requires the development of novel analysis methods such as the simultaneous analysis of omics data using computational systems analysis. In addition, the availability of multi-omics datasets has opened the avenue to gain new insights into related molecular pathophysiological changes in PCOS. Thus, the previously generated data should be fully utilized as a whole to have a systems-view of PCOS. As mentioned in this chapter, the

computation systems analysis such as PPI and pathway analysis has been performed, and several examples of studies using this approach have been provided. The specific data repository of PCOS has also been developed, which could be used for further analysis by PCOS researchers. However, there is a lack of studies that integrate the omics datasets using modeling and simulation to investigate PCOS in a systems-level. This approach should be put into consideration in the future as this approach can dynamically elucidate the PCOS progression and improve the PCOS diagnosis and treatment. Although there is a limitation particularly the state of the incompleteness of biological information such as human interactome and pathway annotation, the analysis on current data by computational systems analysis should be continuously performed as these efforts could constantly enhance the knowledge of a complex syndrome, which is PCOS.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

Author details

Nor Afiqah-Aleng[1,2] and Zeti-Azura Mohamed-Hussein[1,3]*

1 Centre for Bioinformatics Research, Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor, Malaysia

2 Institute of Marine Biotechnology, Universiti Malaysia Terengganu (UMT), Terengganu, Malaysia

3 Centre for Frontier Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor, Malaysia

*Address all correspondence to: zeti.hussein@ukm.edu.my

IntechOpen

References

[1] Chakravorty S, Williams TN. Sickle cell disease: A neglected chronic disease of increasing global health importance. Archives of Disease in Childhood. 2015;**100**:48-53

[2] Mannucci P, Tuddenham E. The hemophilias—From royal genes to gene therapy. The New England Journal of Medicine. 2001;**344**:1773-1779

[3] Davidson DJ, Porteous DJ. The genetics of cystic fibrosis lung disease. Thorax. 1998;**53**:389-397

[4] Kharroubi AT, Darwish HM. Diabetes mellitus: The epidemic of the century. World Journal of Diabetes. 2015;**6**:850-867

[5] Lara-Pezzi E, Dopazo A, Manzanares M. Understanding cardiovascular disease: A journey through the genome (and what we found there). Disease Models & Mechanisms. 2012;**5**:434-443

[6] Escobar-Morreale HF. Polycystic ovary syndrome: Definition, aetiology, diagnosis and treatment. Nature Reviews. Endocrinology. 2018;**14**:270-284

[7] Rutter H. The complex systems challenge of obesity. Clinical Chemistry. 2018;**64**:44-46

[8] Gonzalez MW, Kann MG. Protein interactions and disease. PLoS Computational Biology. 2012;**8**:e1002819

[9] Barabási AL, Oltvai ZN. Network biology: Understanding the cell's functional organization. Nature Reviews Genetics. 2004;**5**:101-113

[10] Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: A network-based approach to human disease. Nature Reviews Genetics. 2011;**12**:56-68

[11] Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, et al. Protein-protein interaction networks (PPI) and complex diseases. Gastroenterology and Hepatology From Bed to Bench. 2014;7:17-31

[12] Milo R. What is the total number of protein molecules per cell volume? A call to rethink some published values. BioEssays. 2013;**35**:1050-1055

[13] Sevimoglu T, Arga KY. The role of protein interaction networks in systems biomedicine. Computational and Structural Biotechnology Journal. 2014;**11**:22-27

[14] Chuang H-Y, Lee E, Liu Y-T, et al. Network-based classification of breast cancer metastasis. Molecular Systems Biology. 2007;**3**:140

[15] Bergholdt R, Brorsson C, Palleja A, et al. Identification of novel type 1 diabetes candidate genes by integrating genome-wide association data, protein-protein interactions, and human pancreatic islet gene expression. Diabetes. 2012;**61**:954-962

[16] Tourette C, Li B, Bell R, et al. A large scale huntingtin protein interaction network implicates RHO GTPase signaling pathways in Huntington disease. The Journal of Biological Chemistry. 2014;**289**:6709-6726

[17] Ito T, Chiba T, Ozawa R, et al. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences. 2001;**98**:4569-4574

[18] Uetz P, Giot L, Cagney G, et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature. 2000;**403**:623-627

[19] Gavin AC, Aloy P, Grandi P, et al. Proteome survey reveals modularity

of the yeast cell machinery. Nature. 2006;**440**:631-636

[20] Krogan NJ, Cagney G, Yu H, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature. 2006;**440**:637-643

[21] Fossum E, Friedel CC, Rajagopala SV, et al. Evolutionarily conserved herpesviral protein interaction networks. PLoS Pathogens. 2009;**5**:e1000570

[22] Peregrín-Alvarez JM, Xiong X, Su C, et al. The modular organization of protein interactions in *Escherichia coli*. PLoS Computational Biology. 2009;**5**:e1000523

[23] Li S, Armstrong CM, Bertin N, et al. A map of the interactome network of the metazoan *C. elegans*. Science. 2004;**303**:540-543

[24] Giot L, Bader JS, Brouwer C, et al. A protein interaction map of *Drosophila melanogaster*. Science. 2003;**302**:1727-1736

[25] Rual JF, Venkatesan K, Hao T, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005;**437**:1173-1178

[26] Stelzl U, Worm U, Lalowski M, et al. A human protein-protein interaction network: A resource for annotating the proteome. Cell. 2005;**122**:957-968

[27] Stark C, Breitkreutz B-J, Reguly T, et al. BioGRID: A general repository for interaction datasets. Nucleic Acids Research. 2006;**34**:D535-D539

[28] Salwinski L, Miller CS, Smith AJ, et al. The database of interacting proteins: 2004 update. Nucleic Acids Research. 2004;**32**:D449-D451

[29] Franz M, Rodriguez H, Lopes C, et al. GeneMANIA update 2018. Nucleic Acids Research. 2018;**46**:W60-W64

[30] Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: Enhancing meaningfulness and reliability of protein-protein interaction networks. Nucleic Acids Research. 2017;**45**:D408-D414

[31] Prasad KS, Goel R, Kandasamy K, et al. Human protein reference database—2009 update. Nucleic Acids Research. 2009;**37**:D767-D772

[32] Brown KR, Jurisica I. Unequal evolutionary conservation of human protein interactions in interologous networks. Genome Biology. 2007;**8**:R95

[33] Orchard S, Ammari M, Aranda B, et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Research. 2014;**42**:D358-D363

[34] Pagel P, Kovac S, Oesterheld M, et al. The MIPS mammalian protein-protein interaction database. Bioinformatics. 2005;**21**:832-834

[35] Chatr-aryamontri A, Ceol A, Palazzi LM, et al. The molecular interaction database. Nucleic Acids Research. 2007;**35**:D572-D574

[36] Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Research. 2017;**45**:D362-D368

[37] Alonso-López D, Campos-Laborie FJ, Gutiérrez MA, et al. APID database: redefining protein-protein interaction experimental evidences and binary interactomes. Database. 2019;**2019**:baz005

[38] Xenarios I. DIP: The database of interacting proteins: 2001 update. Nucleic Acids Research. 2001;**29**:289-291

[39] Hoffmann R, Valencia A. A gene network for navigating the literature. Nature Genetics. 2004;**36**:664

[40] Breuer K, Foroushani AK, Laird MR, et al. InnateDB: Systems biology of innate immunity and beyond—Recent updates and continuing curation. Nucleic Acids Research. 2013;**41**:D1228-D1233

[41] Licata L, Briganti L, Peluso D, et al. The molecular interaction database: 2012 update. Nucleic Acids Research. 2012;**40**:D857-D861

[42] Launay G, Salza R, Multedo D, et al. MatrixDB, the extracellular matrix interaction database: Updated content, a new navigator and expanded functionalities. Nucleic Acids Research. 2015;**43**:D321-D327

[43] Han K, Park B, Kim H, et al. HPID: The human protein interaction. Bioinformatics. 2004;**20**:2466-2470

[44] Orchard S, Kerrien S, Abbani S, et al. Protein interaction data curation—The international molecular exchange consortium (IMEx). Nature Methods. 2012;**9**:345-350

[45] Kuchaiev O, Pržulj N. Integrative network alignment reveals large regions of global network similarity in yeast and human. Bioinformatics. 2011;**27**:1390-1396

[46] Memisevic V, Milenkovic T, Przulj N. Complementarity of network and sequence information in homologous proteins. Journal of Integrative Bioinformatics. 2010;**7**:135

[47] Ideker T, Sharan R. Protein networks in disease. Genome Research. 2008;**18**:644-652

[48] Liu W, Wu A, Pellegrini M, et al. Integrative analysis of human protein, function and disease networks. Scientific Reports. 2015;**5**:14344

[49] Wachi S, Yoneda K, Wu R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. Bioinformatics. 2005;**21**:4205-4208

[50] Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. Bioinformatics. 2006;**22**:2291-2297

[51] Amberger JS, Bocchini CA, Schiettecatte F, et al. OMIM.org: Online mendelian inheritance in man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Research. 2015;**43**:D789-D798

[52] Krauthammer M, Kaufmann CA, Gilliam TC, et al. Molecular triangulation: Bridging linkage and molecular-network information for identifying candidate genes in Alzheimer's disease. Proceedings of the National Academy of Sciences. 2004;**101**:15148-15153

[53] Oti M, Snel B, Huynen MA, et al. Predicting disease genes using protein-protein interactions. Journal of Medical Genetics. 2006;**43**:691-698

[54] Navlakha S, Kingsford C. The power of protein interaction networks for associating genes with diseases. Bioinformatics. 2010;**26**:1057-1063

[55] Adamcsek B, Palla G, Farkas IJ, et al. CFinder: Locating cliques and overlapping modules in biological networks. Bioinformatics. 2006;**22**:1021-1023

[56] Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. Nature Methods. 2012;**9**:471-472

[57] Liu G, Wong L, Chua HN. Complex discovery from weighted PPI networks. Bioinformatics. 2009;**25**:1891-1897

[58] Palla G, Derényi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society. Nature. 2005;**435**:814-818

[59] Altaf-Ul-Amin M, Shinbo Y, Mihara K, et al. Development and implementation of an algorithm for detection of protein complexes in large interaction networks. BMC Bioinformatics. 2006;**7**:207

[60] Altaf-Ul-Amin M, Wada M, Kanaya S. Partitioning a PPI network into overlapping modules constrained by high-density and periphery tracking. ISRN Biomathematics. 2012;**2012**:726429

[61] Mohammad BK, Wakamatsu N, Md A-U-A. DPClusOST: A software tool for general purpose graph clustering. Journal of Computer Aided Chemistry. 2017;**18**:76-93

[62] Li M, Chen JE, Wang JX, et al. Modifying the DPClus algorithm for identifying protein complexes based on new topological structures. BMC Bioinformatics. 2008;**9**:398

[63] Li X-L, Tan S-H, Foo C-S, et al. Interaction graph mining for protein complexes using local clique merging. Genome Informatics: International Conference on Genome Informatics. 2005;**16**:260-269

[64] King AD, Pržulj N, Jurisica I. Protein complex prediction via cost-based clustering. Bioinformatics. 2004;**20**:3013-3020

[65] Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Research. 2002;**30**:1575-1584

[66] Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003;**4**:2

[67] Wu X, Jiang R, Zhang MQ, et al. Network-based global inference of human disease genes. Molecular Systems Biology. 2008;**4**:189

[68] Rezaei-Tavirani M, Rezaei-Tavirani S, Mansouri V, et al. Protein-protein interaction network analysis for a biomarker panel related to human esophageal adenocarcinoma. Asian Pacific Journal of Cancer Prevention. 2017;**18**:3357-3363

[69] Xiao H, Yang L, Liu J, et al. Protein-protein interaction analysis to identify biomarker networks for endometriosis. Experimental and Therapeutic Medicine. 2017;**14**:4647-4654

[70] Hidalgo CA, Blumm N, Barabási AL, et al. A dynamic network approach for the study of human phenotypes. PLoS Computational Biology. 2009;**5**:e1000353

[71] Sun K, Gonçalves JP, Larminie C, et al. Predicting disease associations via biological network analysis. BMC Bioinformatics. 2014;**15**:304

[72] Ko Y, Cho M, Lee JS, et al. Identification of disease comorbidity through hidden molecular mechanisms. Scientific Reports. 2016;**6**:39433

[73] Goh K-I, Cusick ME, Valle D, et al. The human disease network. Proceedings of the National Academy of Sciences. 2007;**104**:8685-8690

[74] Ahmed SS, Ahameethunisa AR, Santosh W, et al. Systems biological approach on neurological disorders: A novel molecular connectivity to aging and psychiatric diseases. BMC Systems Biology. 2011;**5**:6

[75] Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research. 2016;**44**:D457-D462

[76] Croft D, O'Kelly G, Wu G, et al. Reactome: A database of reactions,

pathways and biological processes. Nucleic Acids Research. 2011;**39**:D691-D697

[77] Pico AR, Kelder T, Van Iersel MP, et al. WikiPathways: Pathway editing for the people. PLoS Biology. 2008;**6**:e184

[78] Karp PD, Billington R, Caspi R, et al. The BioCyc collection of microbial genomes and metabolic pathways. Briefings in Bioinformatics. 2017;**2017**:bbx085

[79] Nishimura D. A view from the web. BioCarta. Biotech Software & Internet Report. 2001;**2**:117-120

[80] Kanehisa M, Furumichi M, Tanabe M, et al. KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research. 2016;**45**:353-361

[81] Trupp M, Altman T, Fulcher CA, et al. Beyond the genome (BTG) is a (PGDB) pathway genome database: HumanCyc. Genome Biology. 2010;**11**:O12

[82] Mi H, Thomas P. PANTHER pathway: An ontology-based pathway database coupled with data analysis tools. Methods in Molecular Biology. 2009;**563**:123-140

[83] Fabregat A, Jupe S, Matthews L, et al. The reactome pathway knowledgebase. Nucleic Acids Research. 2018;**46**:D649-D655

[84] Torres-Ayuso P, Sahoo S, Ashton G, et al. Signaling pathway screening platforms are an efficient approach to identify therapeutic targets in cancers that lack known driver mutations: A case report for a cancer of unknown primary origin. NPJ Genomic Medicine. 2018;**3**:15

[85] Sebastian-Leon P, Vidal E, Minguez P, et al. Understanding disease mechanisms with models of signaling

pathway activities. BMC Systems Biology. 2014;**8**:121

[86] Varala K, Marshall-Colón A, Cirrone J, et al. Temporal transcriptional logic of dynamic regulatory networks underlying nitrogen signaling and use in plants. Proceedings of the National Academy of Sciences of the United States of America. 2018;**115**:6494-6499

[87] Doong SJ, Gupta A, Prather KLJ. Layered dynamic regulation for improving metabolic pathway productivity in *Escherichia coli*. Proceedings of the National Academy of Sciences. 2018;**115**:2964-2969

[88] Alaimo S, Marceca G, Ferro A, et al. Detecting disease specific pathway substructures through an integrated systems biology approach. Non-Coding RNA. 2017;**3**:20

[89] Barhak J, Isaman DJ, Ye W, et al. Chronic disease modeling and simulation software. Journal of Biomedical Informatics. 2010;**43**:791-799

[90] Mould D, Upton R. Basic concepts in population modeling, simulation, and model-based drug development. CPT: Pharmacometrics & Systems Pharmacology. 2012;**1**:e6

[91] Edwards LM, Thiele I. Appling systems biology methods to the study of human physiology in extreme environment. Extreme Physiology & Medicine. 2013;**2**:8

[92] Wang D, Lertsithichai P, Nanchahal K, et al. Risk factors of coronary heart disease: A Bayesian model averaging approach. Journal of Applied Statistics. 2003;**30**:813-826

[93] Holford NH, Peace KE. Results and validation of a population pharmacodynamic model for cognitive effects in Alzheimer patients treated with tacrine. Proceedings of the

National Academy of Sciences. 1992;**89**:11471-11475

[94] Gambhir M, Bozio C, O'Hagan JJ, et al. Infectious disease modeling methods as tools for informing response to novel influenza viruses of unknown pandemic potential. Clinical Infectious Diseases. 2015;**60**:S11-S19

[95] Jackson ML, Diallo AO, Médah I, et al. Initial validation of a simulation model for estimating the impact of serogroup A Neisseria meningitidis vaccination in the African meningitis belt. PLoS One. 2018;**13**:e0206117

[96] Kiang TKL, Sherwin CMT, Spigarelli MG, et al. Fundamentals of population pharmacokinetic modelling. Clinical Pharmacokinetics. 2015;**51**:515-525

[97] Meibohm B, Derendorf H. Basic concepts of pharmacokinetic/pharmacodynamic (PK/PD) modelling. International Journal of Clinical Pharmacology and Therapeutics. 1997;**35**:401-413

[98] Cook SF, Bies RR. Disease progression modeling: Key concepts and recent developments. Current Pharmacology Reports. 2016;**2**:221-230

[99] Burke DS, Grefenstette JJ. Toward an integrated meta-model of public health dynamics for preparedness decision support. Journal of Public Health Management and Practice. 2013;**19**:S12-S15

[100] Kitano H. Computational systems biology. Nature. 2002;**420**:206-210

[101] Hartwell LH, Hopfield JJ, Leibler S, et al. From molecular to modular cell biology. Nature. 1999;**402**:C47-C52

[102] Kolesnikov N, Hastings E, Keays M, et al. Array express update-simplifying data submissions. Nucleic Acids Research. 2015;**43**:D1113-D1116

[103] Clough E, Barrett T. The gene expression omnibus database. Methods in Molecular Biology. 2016;**1418**:93-110

[104] Ikeo K, Ishi-i J, Tamura T, et al. CIBEX: Center for information biology gene expression database. Comptes Rendus Biologies. 2003;**326**:1079-1082

[105] Pinero J, Bravo A, Queralt-Rosinach N, et al. DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Research. 2017;**45**:D833-D839

[106] Rappaport N, Twik M, Plaschkes I, et al. MalaCards: An amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. Nucleic Acids Research. 2017;**45**:D877-D887

[107] Carvalho-Silva D, Pierleoni A, Pignatelli M, et al. Open targets platform: New developments and updates two years on. Nucleic Acids Research. 2019;**47**:D1056-D1065

[108] Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. Nucleic Acids Research. 2014;**42**:D1001-D1006

[109] Li MJ, Liu Z, Wang P, et al. GWASdb v2: An update database for human genetic variants identified by genome-wide association studies. Nucleic Acids Research. 2016;**44**:D869-D876

[110] Pletscher-Frankild S, Pallejà A, Tsafou K, et al. DISEASES: Text mining and data integration of disease-gene associations. Methods. 2015;**74**:83-89

[111] Stenson PD, Mort M, Ball EV, et al. The human gene mutation database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and

next-generation sequencing studies. Human Genetics. 2017;**136**:665-677

[112] Agrawal S, Dimitrova N, Nathan P, et al. T2D-Db: An integrated platform to study the molecular basis of type 2 diabetes. BMC Genomics. 2008;**9**:320

[113] Yang Z, Yang J, Liu W, et al. T2D@ ZJU: A knowledgebase integrating heterogeneous connections associated with type 2 diabetes mellitus. Database. 2013;**2013**:bat052

[114] Bai Z, Han G, Xie B, et al. AlzBase: An integrative database for Ggene dysregulation in Alzheimer's disease. Molecular Neurobiology. 2016;**53**:310-319

[115] Nia BV, Kang C, Tran MG, et al. Meta analysis of human AlzGene database: Benefits and limitations of using C. elegans for the study of Alzheimer's disease and co-morbid conditions. Frontiers in Genetics. 2017;**8**:55

[116] Kuzma A, Valladares O, Cweibel R, et al. NIAGADS: The NIA genetics of Alzheimer's disease data storage site. Alzheimer's & Dementia. 2016;**12**:1200-1203

[117] Hutter C, Zenklusen JC. The cancer genome atlas: Creating lasting value beyond its data. Cell. 2018;**173**:283-285

[118] Zhang J, Baran J, Cros A, et al. International cancer genome consortium data portal-a one-stop shop for cancer genomics data. Database. 2011;**2011**:bar026

[119] Azziz R, Carmina E, Chen Z, et al. Polycystic ovary syndrome. Nature Reviews Disease Primers. 2016;**2**:16057

[120] Afiqah-Aleng N, Mohamed-Hussein ZA. Computational systems biology approach on polycystic ovarian syndrome (PCOS). Journal of Molecular and Genetic Medicine. 2019;**13**:1000392

[121] Chen Z-J, Zhao H, He L, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. Nature Genetics. 2011;**43**:55-59

[122] Shi Y, Zhao H, Shi Y, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. Nature Genetics. 2012;**44**:1020-1025

[123] Hayes MG, Urbanek M, Ehrmann DA, et al. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. Nature Communications. 2015;**6**:1-12

[124] Day FR, Hinds DA, Tung JY, et al. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. Nature Communications. 2015;**6**:8464

[125] Lan C-W, Chen M-J, Tai K-Y, et al. Functional microarray analysis of differentially expressed genes in granulosa cells from women with polycystic ovary syndrome related to MAPK/ERK signaling. Scientific Reports. 2015;**5**:14994

[126] Ambekar AS, Kelkar DS, Pinto SM, et al. Proteomics of follicular fluid from women with polycystic ovary syndrome suggests molecular defects in follicular development. The Journal of Clinical Endocrinology and Metabolism. 2015;**100**:744-753

[127] Dong F, Deng D, Chen H, et al. Serum metabolomics study of polycystic ovary syndrome based on UPLC-QTOF-MS coupled with a pattern recognition approach. Analytical and Bioanalytical Chemistry. 2015;**407**:4683-4695

[128] Mohamed-Hussein ZA, Harun S. Construction of a polycystic ovarian

syndrome (PCOS) pathway based on the interactions of PCOS-related proteins retrieved from bibliomic data. Theoretical Biology and Medical Modelling. 2009;**6**:18

[129] Menke NB, Bonchev DG, Witten TM, et al. A novel computational approach to the genetics of polycystic ovarian syndrome (PCOS). Fertility and Sterility. 2007;**88**:S73

[130] Shen H, Qiu L, Zhang Z, et al. Genome-wide methylated DNA immunoprecipitation analysis of patients with polycystic ovary syndrome. PLoS One. 2013;**8**:e64801

[131] Shim U, Kim HN, Lee H, et al. Pathway analysis based on a genome-wide association study of polycystic ovary syndrome. PLoS One. 2015;**10**:e0136609

[132] Kori M, Gov E, Arga KY. Molecular signatures of ovarian diseases: Insights from network medicine perspective. Systems Biology in Reproductive Medicine. 2016;**62**:266-282

[133] Shen H, Liang Z, Zheng S, et al. Pathway and network-based analysis of genome-wide association studies and RT-PCR validation in polycystic ovary syndrome. International Journal of Molecular Medicine. 2017;**40**:1385-1396

[134] Liu L, Du B, Zhang H, et al. A network pharmacology approach to explore the mechanisms of Erxian decoction in polycystic ovary syndrome. Chinese Medicine. 2018;**13**:46

[135] Ramly B, Afiqah-Aleng N, Mohamed-Hussein Z-A. Protein–protein interaction network analysis reveals several diseases highly associated with polycystic ovarian syndrome. International Journal of Molecular Sciences. 2019;**20**:2959

[136] Afiqah-Aleng N, Harun S, A-Rahman MR, et al. PCOSBase: A manually curated database of polycystic ovarian syndrome. Database. 2017;**2017**:bax098

[137] Joseph S, Barai RS, Bhujbalrao R, et al. PCOSKB: A knowledgebase on genes, diseases, ontology terms and biochemical pathways associated with polycystic ovary syndrome. Nucleic Acids Research. 2016;**44**:D1032-D1035

[138] Maniraja JM, Vetrivel U, Munuswamy D, et al. PCOSDB: PolyCystic ovary syndrome database for manually curated genes associated with the disease. Bioinformation. 2016;**12**:4-8