Systems Biology Approaches to Study Disease Comorbidities

6

Tammanna R. Sahrawat

Abstract

Unraveling the relationship between disease phenotypes and disruptions in the underlying cellular functions is an important challenge of contemporary biology and medicine. The traditional approaches to study disease focus primarily on individual genes or proteins related to certain phenotype. Gene-based approach to establish factors that predispose an individual to disease, involved identification of specific genetic defects and single nucleotide polymorphisms and copy number variations. These approaches, though successful, are insufficient since function of various cellular components is exerted through intricate networks of regulatory, metabolic, and protein interactions. On the other hand, the systems biology-based network approaches can facilitate the development of better diagnostic markers and the discovery of core alterations for human complex diseases by system wide analysis on disease diagnosis and the identified disease-responsive genes and modules. Complex diseases result from variations in a large number of correlated genes and their complex interactions rather than by alterations in individual genes. Systems network biology approaches enable investigation of complex interdependencies among a cell's molecular components, to identify significant underlying relationships between apparently distinct disease phenotypes. Such approaches can further facilitate the development of better diagnostic markers and the discovery of core alterations for human complex diseases by system wide analysis on disease diagnosis and the identification of disease-responsive genes and modules.

T. R. Sahrawat (✉)
Centre for Systems Biology and Bioinformatics, UIEAST, Panjab University, Chandigarh, India
e-mail: tammanna@pu.ac.in

R. C. Sobti, A. Sobti (eds.), *Biomedical Translational Research*,
https://doi.org/10.1007/978-981-16-4345-3_6

Keywords

Protein interaction network · Gene interaction network · Biological networks · Computational biology · Systems network biology

6.1 Introduction

The discovery of genes and genetic modules that drive disease has been enabled by study of biological networks. Systems biology aims to investigate multiple biomolecular components and their dynamical behaviors to understand nonlinear interactions that characterize phenotypic alterations leading to complex etiologies of multifactorial diseases such as cancer, diabetes, Alzheimer's disease, etc. The study of complex diseases using traditional approaches such as genetic studies was difficult as multiple factors such as genetic, epigenetic (e.g. environmental factors), cells, and tissue types, and their interactions are involved in the occurrence and development of complex diseases. Furthermore, additional difficulty arises from disruption of normal behaviors of the complex molecular networks of genes and proteins resulting in molecular defects associated with complex diseases. Therefore, for understanding complex diseases, a new paradigm of systems network biology has emerged in view of the availability of large-scale genomic, transcriptomic, proteomic, and metabolomic data that has enabled the identification of risk factors of complex diseases, personalized medicine, and so forth.

Traditional classification of diseases was mainly based on the observed and correlation between pathological analysis and clinical syndromes, which has been widely recognized to have shortcomings on sensitivity and specificity (Loscalzo et al. 2007). In contrast, the network-based systems biology approach can facilitate the development of better diagnostic markers and the discovery of core alterations for human complex diseases by system wide analysis on disease diagnosis and the identified disease-responsive genes and modules (Gijsen et al. 2001).

Network systems biology has facilitated the investigation of complex diseases and their co-occurrence by using all available molecular interaction and ontology data. It is necessary to study disease comorbidities as it is a major problem at the clinical level as it results in the increase in the patient mortality and also complicate the choice of treatment strategies. Furthermore, comorbidity cases are typically associated with polypharmacy (the use of multiple drugs), which can decrease treatment efficacies and can cause unexpected adverse effects, further adding to the disease spectrum in a given patient (Von Lueder and Atar 2014).

6.2 Systems Network Biology Perspective to Study Disease Co-occurrence

Biomolecular networks are integral parts of cellular systems and play a fundamental part in giving rise to life and maintaining the homoeostasis in living organisms. Complex diseases are believed to arise from the dysfunction of these networks rather than single components like genes or proteins (Erler and Linding 2010).

The availability of high throughput genomic, transcriptomic, proteomic, and metabolomic data provides the opportunity to investigate the essential mechanisms by biomolecular networks, in which each node is a biological molecule, and each edge represents the dysfunction interaction or association of a pair of molecules related to disease complex diseases. Hub proteins in PPI networks correspond to essential proteins having a higher probability of involvement in the disease (He and Zhang 2006). Therefore, the omics datasets can be integrated into higher level understandings of a disease process. There exist various biomolecular networks, namely, protein–protein interaction network [PPIN], gene regulatory network [GBN], metabolic network [MBN], and signal transduction network [STN]; these networks interacts with each other to build complex network of interactions (Hu et al. 2016). The human genetic interactions are one of the most important predictors of a link between diseases. Any malfunctions observed in these networks can indicate a presence of disease phenotype.

6.3 Analysis of Disease Co-occurrence Using Biological Networks

The study of disease similarities can provide information of etiology, common pathophysiology, and/or suggest treatment that can be appropriated from one disease to another (Butte and Kohane 2006). Different diseases in the same disorder class may exhibit concordance in protein networks and biological processes, which can be viewed in terms of a "diseasome." To investigate comorbidity between diseases, various types of information such as genetic, symptomatic, and phenotypic in association with penetrance models and medicare data have been used (Rzhetsky et al. 2007; Suthram et al. 2010; Mathur and Dinakarpandian 2012).

Two important computer languages, viz., R (Team RC 2013) and python (Van Rossum and Drake Jr 1995), provide a programming environment for detailed, data-driven biological network analysis and visualization. Many packages are easily downloadable from bioconductor using R platform such as "neat" (efficient network enrichment analysis test) (Signorelli et al. 2016) Version 1.2.1 that integrates gene enrichment analysis (GEA) tests with information on known relations between genes, represented by means of a gene network. Clustering biological functions using gene ontology and semantic similarity can be performed using ViSEAGO (Brionne et al. 2019). Online open source software cytoscape (Shannon et al. 2003) can be used for integrating biomolecular interaction networks with high-throughput expression data to decipher the associations among protein-protein, protein-DNA,

and genetic interactions. RCytoscape (Shannon et al. 2013) merges the powers of R and Cytoscape, thereby resulting in a synergistic tool that combines statistical and data handling facilities with the powerful network visualization and analysis capabilities of cytoscape software.

Another common observation is that a complex disease may lead to the appearance of another phenotype through physiological changes referred to as pleiotropy. Example it has been reported that patients with type 1 diabetes mellitus having hyperglycemia can develop diabetic retinopathy. Thereby treatment of DM (Type I) can prevent the occurrence of the complications associated with it. However, this is not the case when two or more diseases are co-occurring owing to shared genetics or shared diseased pathways. Therefore, it is important to identify the underlying pathways using phenome-wide association studies (PheWAS) that detect genotype-to-phenotype associations by combining and linking genotype data with detailed clinical data from health records (Denny et al. 2013; Bush et al. 2016). Genome-wide association studies (GWAS) can be used to detect pleiotropic risk variants, while studying comorbidity of diseases, which may not otherwise be found in single-disease studies (Ellinghaus et al. 2016). BUHMBOX (Bellou et al. 2020; Gao 2020) is one such online tool that can distinguish between pleiotropy and heterogeneity where the latter refers to misclassification in which a subgroup of cases in one disease is genetically identical to another disease (Ellinghaus et al. 2016).

To understand complex relationships between multifactorial diseases in a population, a systems network biology approach involving computational biology helps to elucidate their underlying common factors. Some of the approaches are as follows:

6.3.1 Using Protein-Protein Interaction to Study Diseases

Proteins interact with one another to form physical or functional interactions, and any perturbations may alter the interactions causing disturbance in the biological functions and leading to one or more diseases. Therefore, studying protein-protein interactions (PPIs) can help us interpret the underlying mechanism of a disease and its relationship with other diseases (Baudot et al. 2009).

Studying the co-occurrence between two or more diseases involves construction and analysis of the PPIs. The PPIs can be retrieved from two types of databases:

1. Primary PPI databases such as MINT and InAct that contain data from the experimental evidence in the literature and use a manual curation process.
2. Metadatabase or a predictive database that include STRING and UniHI, which use the interactions defined in primary databases to computationally predict and refine the interactions.

The PPI data retrieved from the databases can be visualized and analyzed using various bioinformatics tools like Cytoscape, NAViGaTOR, POINeT, Gephi, and Pajek (Leong et al. 2013). These network analysis tools have the ability to take data

from varied sources. They perform various analysis using different algorithms to identify various critical hub nodes that may be responsible for various diseases along with gene ontology (GO) annotations by various plugins such as BiNGO and ClueGO, which use gene ontology (GO), Reactome, and KEGG annotations (Maere et al. 2005; Bindea et al. 2009; Kanehisa et al. 2017; Fabregat et al. 2018).

6.3.2 Using Gene-Gene Interaction Networks to Study Diseases

Gene products, i.e., the proteins, work together to perform a particular task. One of the important genetic interactions is that between the genetic variants, which can be lethal or have a suppressor effect. It is important to study the complex networks of these genes to understand the complexity of the diseases and their co-occurrence (Bebek 2012). Co-expression networks of the genes can be constructed to study the correlations between the expressions of the different genes (Cho et al. 2012). Like the PPI data, analysis of gene-gene interactions also involves the retrieval of genes from the databases such as gene database (NCBI), KEGG or WikiPathways, BioGRID, IRefIndex, and I2D (Zuberi et al. 2013). This is followed by building an analysis of a gene-gene interaction network using various bioinformatics tools like Cytoscape application GeneMANIA and CytoCluster (Zuberi et al. 2013; Li et al. 2017).

6.3.3 Pathway Enrichment Analysis or Functional Enrichment Analysis

Pathway enrichment or functional enrichment aids in achieving deep biological insights of the functional roles of the proteins, genes, and metabolites from multi-omics datasets by comparing the activity of the pathways or biological processes of interest in two or more states or cohorts to be compared. Three approaches for pathway enrichment are as follows (Paczkowska et al. 2020):

1. Over-representation analysis approach gives information about the number of genes in each pathway and whether the pathway is upregulated or downregulated. MetPA is a software based on this approach (Khatri et al. 2012).
2. Functional Class Scoring Approach uses all available molecular measurements for the analysis of the genes. An example is gene set enrichment analysis (GSEA) software that analyzes set of the genes that are differentially expressed, which can later be visualized using EnrichmentMap plugin in Cytoscape (Khatri et al. 2012; Reimand et al. 2019; Subramanian et al. 2005).

6.3.4 Disease-Gene Interaction

Disease-gene networks can be used to identify the dysregulated genes of a particular disease that are responsible for causing other diseases too. DisGeNET is a repository that contains information about genes, genetic variants, and their interaction with the complex human diseases including Mendelian, environmental, and rare diseases (Bauer-Mehren et al. 2010). DisGeNET integrates data from expert curated repositories, GWAS catalogues, animal models, and the scientific literature. The information provided in it is accessible through a web interface, a Cytoscape App, an RDF SPARQL endpoint, scripts in several programming languages, and R package (Piñero et al. 2016). The information retrieved can be used to decipher information of a gene and its cross-talk with the other diseases.

6.4 Identifying Critical Hub Proteins and Probable Therapeutic Targets

In the omics era, the study of interconnections amongst genome, transcriptome, proteome and metabolome using systems biology approaches has led to increasing awareness of complex diseases and their comorbidities. Studies based on the systems network biology approaches revealed genes that frequent between breast and bone cancer were found to be enriched in various signaling pathways such as cell cycle, transcription misregulation in cancer, p53 signaling, breast cancer, integrated breast cancer, FOXO signaling pathway, and DNA damage pathways, thereby indicating that they may be responsible for common molecular origin of these cancers (Sahrawat and Kaur 2017). Another study identified common molecular signatures and pathways that interact between inflammatory bowel disease and colorectal cancer, and the indispensable pathological mechanisms revealed 177 common differentially expressed genes (DEGs) between the two diseases (Al Mustanjid et al. 2020). Using a systems biology network approach involving PPI network analysis, critical protein was identified, viz., HSPA8/HSC70, which could be targeted to affect the regulation of CFTR (cystic fibrosis transmembrane conductance regulator), which is mutated in cystic fibrosis (Sahrawat and Bhalla 2013). The genes and their protein products may act as potential biomarkers for early detection of predisposition to co-occurrence of diseases and potential therapeutic targets based on the common molecular underpinnings of comorbid diseases such as diabetes, depression, and cardiovascular disease (Sahrawat and Talwar 2020) and age-related diseases such as Alzheimer's disease and diabetes for better understanding of pathophysiology (Sahrawat and Dwivedi 2020).Thus, systems biology approach may hold great promise to identify probable therapeutic targets for diseases that are comorbid.

6.5 Concluding Remarks

In systems biology, the individual molecules of the biological systems do not exist in isolation. On the contrary, they work in conjunction with one another to produce a specific phenotypic pattern specific to a given cell type. Study of bimolecular networks using systems biology approaches can be used to identify biomarkers and critical control proteins that can play an important role in the understanding disease etiologies and design treatment regimens for diseases that are comorbid due to the common genes/proteins in the pathways of complex human diseases. The elucidation of such common genes/proteins in comorbid diseases can go a long way for better quality treatment for the patient in a timely manner.

References

Al Mustanjid MA, Mahumud SMH, Royel MRI, Rahman MH, Islam T, Rahman MR, Moni MA (2020) Detection of molecular signatures and pathways shared in inflammatory bowel disease and colorectal cancer: a bioinformatics and systems biology approach. Genomics 112(5): 3416–3426. https://doi.org/10.1016/j.ygeno.2020.06.001

Baudot A, Gomez-Lopez G, Valencia A (2009) Translational disease interpretation with molecular networks. Genome Biol 10(6):221. https://doi.org/10.1186/gb-2009-10-6-221

Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI (2010) DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene–disease networks. Bioinformatics 26(22): 2924–2926. https://doi.org/10.1093/bioinformatics/btq538

Bebek G (2012) Identifying gene interaction networks. In: Statistical human genetics. Humana Press, pp 483–494

Bellou E, Stevenson-Hoare J, Escott-Price V (2020) Polygenic risk and pleiotropy in neurodegenerative diseases. Neurobiol Dis 142:104953

Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics (Oxford, England) 25(8): 1091–1093. https://doi.org/10.1093/bioinformatics/btp101

Brionne A, Juanchich A, Hennequet-Antier C (2019) ViSEAGO: a bioconductor package for clustering biological functions using gene ontology and semantic similarity. BMC BioData Mining 12(16). https://doi.org/10.1186/s13040-019-0204-1

Bush WS, Oetjens M, Crawford DC (2016) Unravelling the human genome–phenome relationship using phenome-wide association studies. Nat Rev Genet 17(3):129. https://doi.org/10.1038/nrg.2015.36

Butte AJ, Kohane IS (2006) Creation and implications of a phenome-genome network. Nat Biotechnol 24(1):55–62. https://doi.org/10.1038/nbt1150

Cho DY, Kim YA, Przytycka TM (2012) Network biology approach to complex diseases. PLoS Comput Biol 8(12). https://doi.org/10.1371/journal.pcbi.1002820

Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Basford MA (2013) Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol 31(12):1102. https://doi.org/10.1038/nbt.2749

Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, Folseraas T (2016) Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet 48(5):510. https://doi.org/10.1038/ng.3528

Erler JT, Linding R (2010) Network-based drugs and biomarkers. J Pathol 220(2):290–296. https://doi.org/10.1002/path.2646

Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, Milacic M, Roca CD, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Viteri G, Weiser J, Wu G, D'Eustachio P (2018) The reactome pathway knowledgebase. Nucleic Acids Res 46(D1):D649–D655. https://doi.org/10.1093/nar/gkx1132

Gao XR (2020) Pleiotropy in eye disease and related traits. In: Genetics and genomics of eye disease. Academic Press, pp 315–336. https://doi.org/10.1016/B978-0-12-816222-4.00020-4

Gijsen R, Hoeymans N, Schellevis FG, Ruwaard D, Satariano WA, van den Bos GA (2001) Causes and consequences of comorbidity: a review. J Clin Epidemiol 54(7):661–674. https://doi.org/10.1016/S0895-4356(00)00363-2

He X, Zhang J (2006) Why do hubs tend to be essential in protein networks? PLoS Genet 2(6):e88

Hu JX, Thomas CE, Brunak S (2016) Network biology concepts in complex disease comorbidities. Nat Rev Genet 17(10):615

Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45(D1):D353–D361. https://doi.org/10.1093/nar/gkw1092

Khatri P, Sirota M, Butte AJ (2012) Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 8(2):e1002375

Leong GW, Lee SC, Lau CC, Klappa P, Omar MSS (2013) Comparison of computational tools for protein-protein interaction (PPI) mapping and analysis. J Teknol 63(1):1–8

Li M, Li D, Tang Y, Wu F, Wang J (2017) CytoCluster: a Cytoscape plugin for cluster analysis and visualization of biological networks. Int J Mol Sci 18(9):1880. https://doi.org/10.3390/ijms18091880

Loscalzo J, Kohane I, Barabasi AL (2007) Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. Mol Syst Biol 3(1):124. https://doi.org/10.1038/msb4100163

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics (Oxford, England) 21(16):3448–3449. https://doi.org/10.1093/bioinformatics/bti551

Mathur S, Dinakarpandian D (2012) Finding disease similarity based on implicit semantic similarity. J Biomed Inform 45(2):363–371. https://doi.org/10.1016/j.jbi.2011.11.017

Paczkowska M, Barenboim J, Sintupisut N, Fox NS, Zhu H, Abd-Rabbo D, Mee MW, Boutros PC, Drivers PCAWG, Functional Interpretation Working Group, Reimand J, PCAWG Consortium (2020) Integrative pathway enrichment analysis of multivariate omics data. Nat Commun 11(1):735. https://doi.org/10.1038/s41467-019-13983-9

Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, Furlong LI (2016) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res 45(D1):D833–D839. https://doi.org/10.1093/nar/gkw943

Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Merico D (2019) Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc 14(2):482–517. https://doi.org/10.1038/s41596-018-0103-9

Rzhetsky A, Wajngurt D, Park N, Zheng T (2007) Probing genetic overlap among complex human phenotypes. Proc Natl Acad Sci 104(28):11694–11699. https://doi.org/10.1073/pnas.0704820104

Sahrawat TR, Bhalla S (2013) Identification of critical target protein for cystic fibrosis using systems biology network approach. Int J Bioautomation 17(4):227–240

Sahrawat TR, Dwivedi J (2020) Investigating the relationship between diabetes and Alzheimer's disease: a network systems biology approach. J Sci Sustain Dev 7(1):1–11. https://doi.org/10.4314/jssd.v7i1.1

Sahrawat TR, Kaur H (2017) Unraveling the relationship between breast and bone cancer: a systems biology approach. Biosci Discov 8(4):783–790

Sahrawat TR, Talwar D (2020) Network-centric identification of disease co-occurrences: a systems biology approach. Sumerianz J Med Healthc 3(11):103–110. https://doi.org/10.1371/journal.pone.0003178.g001

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504. http://www.genome.org/cgi/doi/10.1101/gr.1239303

Shannon PT, Grimes M, Kutlu B, Bot JJ, Galas DJ (2013) RCytoscape: tools for exploratory network analysis. BMC Bioinformatics 14(1):217

Signorelli M, Vinciotti V, Wit EC (2016) NEAT: an efficient network enrichment analysis test. BMC Bioinformatics 17:352. https://doi.org/10.1186/s12859-016-1203-6

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102(43):15545–15550. https://doi.org/10.1073/pnas.0506580102

Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, Butte AJ (2010) Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. PLoS Comput Biol 6(2). https://doi.org/10.1371/journal.pcbi.1000662.g001

Team RC (2013) R: a language and environment for statistical computing

Van Rossum G, Drake FL Jr (1995) Python reference manual. Centrum voor Wiskunde en Informatica, Amsterdam

Von Lueder TG, Atar D (2014) Comorbidities and polypharmacy. Heart Fail Clin 10(2):367–372. https://doi.org/10.1016/j.hfc.2013.12.003

Zuberi K, Franz M, Rodriguez H, Montojo J, Lopes CT, Bader GD, Morris Q (2013) GeneMANIA prediction server 2013 update. Nucleic Acids Res 41(W1):W115–W122. https://doi.org/10.1093/nar/gkt533