REVIEW ARTICLE

Protein-protein interaction networks (PPI) and complex diseases

Nahid Safari-Alighiarloo[1], Mohammad Taghizadeh[2], Mostafa Rezaei-Tavirani[1], Bahram Goliaei[2], Ali Asghar Peyvandi[3]

[1] *Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[2] *Bioinformatics Department, Institute of Biochemistry and Biophysics, Tehran University, Tehran, Iran*

[3] *Hearing Disorders Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

ABSTRACT

The physical interaction of proteins which lead to compiling them into large densely connected networks is a noticeable subject to investigation. Protein interaction networks are useful because of making basic scientific abstraction and improving biological and biomedical applications. Based on principle roles of proteins in biological function, their interactions determine molecular and cellular mechanisms, which control healthy and diseased states in organisms. Therefore, such networks facilitate the understanding of pathogenic (and physiologic) mechanisms that trigger the onset and progression of diseases. Consequently, this knowledge can be translated into effective diagnostic and therapeutic strategies. Furthermore, the results of several studies have proved that the structure and dynamics of protein networks are disturbed in complex diseases such as cancer and autoimmune disorders. Based on such relationship, a novel paradigm is suggested in order to confirm that the protein interaction networks can be the target of therapy for treatment of complex multi-genic diseases rather than individual molecules with disrespect the network.

(Please cite as: **Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, Goliaei B, Peyvandi AA. Protein-protein interaction networks (PPI) and complex diseases. Gastroenterol Hepatol Bed Bench 2014;7(1):17-31).**

Introduction

Early biological experiments revealed that proteins, as the main agents of biological function, determine the phenotype of all organisms. By the advent of molecular biology, it has been assumed that proteins are not naturally functional in isolated forms; instead, they have interactions with one another and also with other molecules (e.g. DNA, RNA) that mediate metabolic and signaling pathways, cellular processes, and organismal systems (1). Thus, studies of proteins' interactions are fundamental to perceive their role within the cell. The term 'protein interaction' encompasses a variety of events, such as transient and stable

complexes, as well as physical and functional interactions (2). Protein-protein interaction (PPI) data can be used in a larger scale to map networks of interactions depend on their physical or functional association (3, 4). Protein interaction networks are practical means to abstract basic knowledge and to improve biological and biomedical applications. Although protein interaction networks are incomplete (5) and error-prone (6), systematic studies of them have been confirmed to be especially important for deciphering the relationships between network structure and function (7), discovering novel protein function (8), identifying functionally coherent modules (9, 10), and conserved molecular interaction patterns (11, 12). Since proteins have principle role in biological function, their interactions determine

molecular and cellular mechanisms which control healthy and diseased states in organisms. Diseases are often caused by mutations affecting the binding interface or leading to biochemically dysfunctional allosteric changes in proteins (13). Therefore, the molecular basis of diseases can be enlightened through protein interaction networks, which in turn can appraise methods for prevention, diagnosis, and treatment. Generally, traditional analyses exploit a univariate approach to study gene expression and identify genes with meaningful individual differential expression in the phenotype of interest (14). However, the underlying mechanisms of complex diseases, which arise from the interplay among multiple genetic and environmental factors, cannot be explicated by such univariate approaches Hence, since there are remarkable increase in an availability of human protein interaction data, the focus of bioinformatics development has shifted from understanding networks encoded by model species to understand the networks underlying human disease (15).

Network analysis of complex systems

The concept of network graph theory was first developed in the 18th century by Leonard Euler, but it was not used in real complex networks (such as Biological networks) until the advent of the computer systems (16). There are three main progressions in graph theory at 20th century, random graph theory, small world networks and scale free networks. These developments have framed our understanding of how networks behave as a whole. Random graph theory was developed by Gilbert, Erdosh and Rényi in 1959. This model uses a given probability to locate an edge between two nodes, regardless of the probability of where other edges are placed in the graph (16). Since random graph theory had the weak fit with real world data, two related findings were discovered, regarding the structure of real world networks: small-world networks and scale-free networks.

The two properties (shorter than expected path length and high clustering coefficient) which demonstrate small-world networks have been represented many real world networks including power grids, social networks and telephone call graphs (17). Because the small world network is limited by their transient and spatio-temporal dynamics, they may not be an appropriate model for most complex protein networks. Therefore, small world network without the dimensional constraints may be more suitable for the study of biological systems.

Small world networks can also be characterized by related scale-free networks. Scale-free networks were first formally introduced by Albert and Barabasi (3). Their main feature is that the degree distribution (i.e. the number of connections or edges a node possesses) follows the power-law rule in which the vast majority of nodes have a low degree, while a smaller than expected numbers of nodes, known as hubs, have a very high degree of connectivity. It was later suggested that protein–protein interaction networks obey such power-law distribution (18).

Protein interactions networks

The structure and nature of protein interaction networks as one of the best appreciated in biological networks is a considerable subject in system biology, particularly due to the rich datasets of protein interactions that are available for study. Systematic analysis of physical protein interaction networks initiated in the mid- 1990s with several studies implying to complex relationships between large macromolecular protein complexes such as DNA-polymerase or components of the transcription splicing complexes (19, 20). Furthermore, the growing evidence of mutual interactions between multiple cell-signaling pathways have revealed signal transduction as a network of interconnected pathways rather than a series of insulated linear pathways (21, 22). Therefore, understanding of

protein interactions in the context of large networks is dramatically in attention, and the large datasets of protein interactions are being developed concomitantly. For instance, protein interactions could be gathered from the literature through systematic mining of detailed literature sources (23) and unbiased, high-throughput strategies such a yeast two-hybrid screens which are efficient to map all the interactions of a given organism's proteome (24).

The structure of protein networks

The structure of protein interaction networks have been examined by recent studies in several species. These studies have discovered that regardless of species, the known protein networks are scale-free. It means that some hub proteins have a huge proportion of the interactions while most proteins (are not hub and) only contain a small fraction of ones (25). It is an obvious fact that understanding the structure of a species' protein interaction network only provides one dimension of the biochemical machinery controlling a cell's behavior. Thus, several groups have integrated dynamics of gene expression with protein interaction networks in order to uncover how these networks change in different biological states. For example, the network of proteins involved in the yeast cell cycle was merged with their expression across the cell cycle. The results showed that although most elements of interacting complexes are expressed in a coherent way across the stages of the cell cycle, only a single or a small number of key proteins interacting with these complexes are expressed in a single phase (26). A "just in time" model was suggested by these data describing dynamic protein complexes where most of the proteins involved in a dynamic process were co-expressed regardless of stage, while they are not active. It is because of missing key elements of the given complex. Thus, the complexes are dynamically activated by expressing key elements at a specific period; thereby completing the

complex for its stage is a particular purpose. The dynamic modular structure is another component of the protein network that it has also been observed in the human protein interaction network (27). This phenomenon represents that modular structure is not species specific or an artifact of the analysis of expression and interactome of yeast (28).

Network topology is also introduced to characterize a network structure. There are four higher-level topological indices including average degree (K), clustering coefficient(C), average path length (L), and diameter (D). It is possible to calculate four topological distributions such as degree distribution P (k), degree distribution of cluster coefficients C (k), shortest path distribution SP (i), and topological coefficient distribution TC (k), which take more attentions (29–31) and are comprehensively used in cellular networks, such as PPI networks (32, 7), MNs (Metabolic Networks) (33), gene co-expression networks (GCEN) (34), and domain interaction networks (35). To get more, some basic definitions of topological terms are described in table1.The topological features of cellular networks are efficiently explained by these criteria which also provide vast insights into cellular evolution, molecular function, network stability, and dynamic responses (31, 33–35).

Approaches for discovering protein-protein interactions networks

To know more about protein interactions in systemic screen, two basics methods can be utilized; Experimental technologies tend to identify of PPI, and Computational methods are an excellent candidate for prediction of protein interactions.

A-Experimental identification of PPI

1- Biophysically Methods

The main source of knowledge about protein interactions has resulted from biophysical

Table 1. Basic definitions in PPI network

Term	Definition
Node or (Vertices)	Each protein in network
Edge or (link)	Physical or functional interactions between proteins
Hub	Each "high-degree" node of a network
Modules	Group of sub networks in which each sub network includes a high number of inside-sub network links and a low number of between- sub network links
Degree (k)	The number of adjacent links
Average degree (<k>)	The mean of all degree values of nodes in a network.
Clustering Coefficient (<C>)	The proportion of links between the nodes within the i-neighborhood divided by the number of links that could possibly exist between them
Shortest Path Length	The steps (number of links) needed to connect every pair of nodes through their shortest path.
Diameter	The minimum number of links that separate the two most distant nodes in a network.
Betweenness centrality	Measures how often nodes occur on the shortest paths between other nodes
Heterogeneity of a network	the coefficient of variation of the degree distribution

methods, particularly from those based on structural information (e.g. X-ray crystallography, NMR spectroscopy, fluorescence, atomic force microscopy). Interacting partners are identified by biophysical methods and they also provide detailed information about the biochemical features of the interactions (e.g. binding mechanism, allosteric changes involved) (1).

2- High-Throughput methods
2-1- Direct high-throughput methods
Yeast two-hybrid (Y2H) is one of the prevalent straight high-throughput methods. The Y2H system examines the interaction of two given proteins by fusing each of them to a transcription binding domain. If the transcription complex is activated, it means that the proteins interact. In this situation a reporter gene is transcribed that its product can be detected (36).

2-2- Indirect high-throughput methods
Protein interactions have been deduced by several high-throughput methods via looking at characteristics of the genes encoding the putative interacting partners. For instance, gene co-expression is based on the assumption that the genes of interacting proteins must be co-expressed

to provide the products for protein interaction. Synthetic lethality, on the other hand, introduces mutations on two separate genes, which are viable alone but lethal when combined, as a way to deduce physically interacting proteins (37).

B-Computational predictions of PPIs

Although experimental biophysics approaches can provide specific interaction details, they have some deficiencies; they are expensive, extremely laborious, and can only be implemented for a few complexes at a time. The prediction of PPIs can be done by computational approaches as a fast and an inexpensive alternative to complete experimental efforts. Furthermore, computational interaction studies are useful to confirm experimental data and are effective to select potential targets for further experimental screening (38-40). More importantly, such invaluable methods prepare great chance to follow proteins within the context of their interaction networks at different functional levels (i.e. at the complex, pathway, cell, or organismal level). Therefore, they give us a great chance to convert lists of pair-wise relationships

into complete network maps. Since computational techniques are based on different principles, they can also reveal functional relationships. Besides, such approaches even provide information about interaction details (e.g. domain interactions), which may avoid some experimental methods. Computational interaction prediction methods can be classified into two types: methods predicting protein domain interactions from existing empirical data about PPI, and methods relying entirely on theoretical information to predict protein-protein or domain-domain interactions (41).

1- Empirical predictions

The computational techniques based on experimental data exploit the relative frequency of interacting domains (42), maximum likelihood estimation of domain interaction probability (43,44), co-expression (45), or network properties (46–48) to predict protein and domain interactions. Since empirical computations rely on an existing protein network to infer new nodes, they disseminate the inaccuracies of the experimental methods, which is the main disadvantage of such computational predictions.

2- Theoretical predictions

Theoretical techniques regard a great range of biological considerations; they use an accepted assumption that interacting proteins coevolve to preserve their function (e.g. mirror tree, phylogenetic profiling (49–51), happen in the same organisms (e.g. (52, 53)), conserve gene order (e.g. gene neighbors method (54,55)) or are fused in some organisms (e.g. the Rosetta Stone method (56,57)).

Protein networks and diseases

Protein networks are useful resources to identify novel pathways to gain basic knowledge of diseases. Protein interaction sub networks are group of the proteins that are interacting with each other's in functional complexes and pathways (58).

Now, new methods are being developed to accurately extract interaction sub networks to yield pathway hypotheses that can be used to understand different aspects of disease progression (59, 60).

Some of the findings that have been revealed by combining PPI and pathway analysis are here: (a) over 39,000 protein interactions have been recognized in the human cell (61), (b) although, in a few diseases like cancer, disease genes tend to encode highly-connected proteins (hubs), disease genes are generally nonessential and occupy peripheral positions in the human interactome (62-64), (c) disease genes tend to cluster together and co-occur in central network locations (65). (d) Proteins involved in similar phenotypes (e.g. all cancer proteins) are highly interconnected (62). (e) Viral networks differ significantly from cellular networks, which raise the hypothesis that other intracellular pathogens might also have distinguishing topologies (66). (f) Etiologically unrelated diseases often present similar symptoms because separate biological processes often use common molecular pathways (67).

It is noticeable that PPI networks can be used to explore the differences between healthy and diseased states (68, 69). Since the identification of disease-associated interacting proteins can give us the ability of recognizing potentially interesting disease-associated gene candidates (i.e. the genes coding for the interacting proteins are putative disease causing genes), Protein interaction studies play a major role in the prediction of genotype-phenotype associations. Therefore, it is suggested that one of the best ways to know more about novel disease genes is to study the interaction partners of known disease associated proteins (70). Gandhi et al. (71) found that mutations on the genes of interacting proteins lead to similar disease phenotypes, presumably because of their functional relationship. Therefore, protein interactions can be used to prioritize gene

candidates in studies investigating the genetic basis of disease (72).

It is possible to introduce markers to create new prognostic tools by identifying disease sub networks, and determining activated pathways in diseased states. For instance, Chuang et al. (59) identified a set of sub network markers via using a protein-network- based approach by which they were able to classify metastatic vs. non metastatic tumors in individual patients accurately.

Disease networks can improve drug design by determining key nodes as potential drug targets. If, for example, the target is a hub (a highly connected protein), its inhibition may affect many activities that are critical for the suitable function of the cell and might thus be unsuitable as a drug target. On the other hand, less connected nodes (e.g. nodes affecting a single disease pathway) could constitute sensitive points of the disease related network, which are more proper candidates for drug targets (73, 74).

Understanding of complex diseases through protein interactions network

Many complex diseases are resulted from a complex interplay of multiple genes, and heterogeneity. It means that such diseases do not comply the standard Mendelian patterns of inheritance. Besides, environmental factors connected to the risk allele immensely determine the development of diseases phenotype (75). Since several factors play critical role in complex diseases phenotype, Genotype-Phenotype correlation is complicated; thus building suitable protein network with the genetically associated genes in complex diseases would provide excellent hypotheses for further experimentations to follow the molecular pathways of developing diseases phenotypes. Analysis of protein-protein interaction (PPI) networks is being increasingly

recognized as an momentous mean, which help us characterize the underlying biology of genes associated to complex diseases, in particular immune-mediated ones (76,77).

It is logical to hypothesize that those genes which are truly associated with the same trait will be involved in similar biological processes. Identification of candidate genes by which the pathogenesis of complex diseases will be further elucidated is a great challenge of biomedical research. By recent assembling of dependable molecular interaction data, progress has been expanding in the discovery of novel susceptibility genes. Concomitantly, expectations are increasing about opportunities of computational approaches for distinguishing disease-related genes from non-disease ones. The results of the latest studies on the prediction of candidate genes dependent upon PPI networks alone or in addition to gene expression profiles (78-80) could reflect potential candidate genes. This approach also promotes a better understanding of the role of PPI topological features in the prediction of susceptible genes. There are some invaluable points that have been resulted in previous studies (81-83); for instance, direct interacting partners of a protein likely tend to share similar functions with it, and causative genes of some complex disease tends to reside in the same network groups such as biological modules, protein complexes, pathways or sub networks of a given biological network. Some further graph-theoretical analyses of molecular interaction networks (84-86) have succeeded in identifying biological network modules and deciphering the association between genes and diseases. Consequently, a unified basic assumption mentions that genes sharing similar network topological features with known disease genes may result in the same phenotype. These relationships suggest a novel paradigm for treatment of complex mutagenic diseases where the protein interaction network is the target of therapy more than individual molecules within the network.

The application of PPIs in surveying of a couple serious complex diseases

Cancer

Cancer is a complex disease, and many genes have been reported to involve in the development of cancers. A systematic investigation of cancer proteins in the human protein-protein interaction network may provide important biological information for uncovering the molecular mechanisms of cancer and, potentially, other complex diseases while traditional approach are not so promising because of their focusing on studying individual genes or loci. In this set of materials, Ergun et al., 2007, integrated known genetic modifiers of prostate cancer with expression dynamics and protein interaction networks. Their efforts were lead to the development of methods which were suitable enough to reveal molecular network differences between aggressive and non–aggressive prostate cancers (87). Equally, Chuang et al 2007 used protein networks as well as ~8000 gene expression panels in order to discover an effective approach for the classification of metastatic versus non-metastatic tumors (59). Another study was investigated by Taylor et al 2009, for finding the indicator of breast cancer prognosis. By merging expression data from 250 breast cancer tumors with the human protein interaction network, they discovered that there is a significant change in the modular structure of the network in patients with good outcome breast cancer (disease free survival greater than 5 years) versus those with poor outcome (27). There are some invaluable findings which come from exploring of topological features of PPIs. For instance, in 2006, Pall and et al published a paper in which they showed that topological features of human proteins translated from known cancer genes isn't the samefor proteins which not documented as being mutated in cancer. Cancer proteins particularly tend to interact with more number of proteins, and they also prefer to participate in central hubs rather than peripheral ones; therefore, it is reflecting their greater centrality and participation in networks as backbone of the proteome. Moreover, their results indicated that cancer proteins contain a high proportion of structural domains which comprised a high propensity to mediate protein interactions. An underlying evolutionary distinction between the two groups of proteins was uncovered by such observations in which the central roles of proteins, whose mutations lead to cancer, were reflected (64). Moreover, in 2009 Jingchun et al., explored global and local network characteristics of the proteins encoded by cancer genes (cancer proteins) in the human interactome. This study confirmed earlier results in which they implied that the network topology of the cancer proteins was much different from non-cancer ones. By the investigation of topological features of the proteins encoded by essential genes (essential proteins) or control genes (control proteins) versus cancer proteins, they concluded that cancer proteins tended to have higher degrees, higher betweenness, shorter shortest-path distance, and weaker clustering coefficient in the human interactome than two other proteins. Finally, they achieved to this fact that cancer proteins have non-randomly distribution in the human interactome and their strongly connected with each other (88). Concurrently, Li and his colleagues in 2009 also discovered that Topological features of PPI networks, protein domain compositions and GO annotations promote the identification of cancer genes. They introduced the SVM classifier which was able to merge multiple characteristics, and it was useful for prioritizing candidate cancer genes for experimental validations (89).With the intense investigation, Tijana et al. in 2010 also provided clear evidence to substantiate that PPI network has different structure around cancer genes than from the structure around non-cancer genes. It seems

that by following of such underlying principles of this phenomenon, they achieve quite promising results, which increase the understanding of complex diseases (90). As an important epigenetic modification; DNA methylation has a central role in the development of mammals and in the event of complex diseases. Genes that interact directly or indirectly may have the same or similar functions in the biological processes in which they are involved and together contribute to the related disease phenotypes. Network theory appears to have a potential for uncovering the complex relations between genes. A protein-protein interaction (PPI) network represents a platform by which we have this chance to systematically identify disease-related genes from the relations between genes with similar functions. To prove this hypothesis, the network theory was combined with epigenetic characteristics by Hui and et al. in 2011. 154 potential cancer-related genes with abnormal methylation were prioritized that might contribute to the further understanding of cancers (91). After a while in 2012, based on combined network topological features, Zhang et al., introduced a novel computational method that enable them to construct a combined classifier. They used this method to predict candidate genes for coronary artery diseases (CAD). As a result, 276 novel candidate genes were predicted by such machine, and were shown to share similar functions to known disease genes (92). Along the same vein, Wu et al. in 2012 proved this fact that combining of gene expression and network data is a promising approach to prioritize disease-associated genes. In this paper, they developed a method, Networked Gene Prioritizer (NGP). Several breast cancer and lung cancer datasets were used to demonstrate that NGP performs better than the existing methods. The top-ranked genes by NGP-PLK1, MCM2, MCM3, MCM7, MCM10 and SKP2 might arrange to promote cell cycle related processes in cancer but not normal cells (93). Furthermore, recently, Network

topological characteristics (including degree, betweenness, clustering coefficient and shortest-path distance) of cancer proteins of the human nuclear and tyrosine kinases receptors network (NR-RTK) were also explored by Choura which constructed in their earlier work. Their results had so similarity to earlier ones in this area as they delineated the network topology of cancer proteins in this network. They discovered that relative to the non-cancer proteins, the cancer proteins have likely higher degree, higher betweenness, similar clustering coefficient and similar shortest-path distance. Finally, they found that the cancer proteins were occupied mainly in signaling pathways which their dysfunction is directly related to cancer arrival. These findings are useful for cancer candidate protein prioritization and confirmation, and identification of key pathways involved in cancer diseases (94).

Autoimmune diseases
Autoimmunity is the breakdown of an organism to recognize its own parts as itself, which results in an immune response against its own cells and tissues. Any disease that results from such an abnormal immune response is named as an autoimmune disease. There are some points about these diseases: 1) Today, more than 80 clinically separated diseases are classified as autoimmune diseases.2) They happen in 3–5% of the population, usually as a result of a numerous of genetic and environmental parameters, which lead to an alteration in immune reactivity (95,96). The genetic architecture of autoimmune diseases as complex ones may be repercussions of heterogeneity, incomplete penetrance, polygenic inheritance, and environmental factors. Thus, it seems that constructing meaningful biological pathways via protein-protein interaction network can provide invaluable information about molecular mechanism of autoimmune diseases. For this purpose, Tuller et al. in 2013 sought to identify intracellular regulatory mechanisms in

peripheral blood mononuclear cells (PBMCs), which are either common in numerous autoimmune diseases or just in some of them. They integrated large-scale data such as protein–protein interactions, gene expression and demographical information of hundreds of patients and healthy subjects, related to six autoimmune diseases with available large-scale gene expression measurements: multiple sclerosis (MS), systemic lupus erythematosus (SLE), juvenile rheumatoid arthritis (JRA), Crohn's disease (CD), ulcerative colitis (UC) and type1 diabetes (T1D).This study cleared that the most of the cellular processes have been shared in most of the analyzed autoimmune diseases. These processes especially were involved cell proliferation (epidermal growth factor, platelet-derived growth factor, nuclear factor-kB, Wnt/b-catenin signaling, stress-activated protein kinase c-Jun NH2-terminal kinase), inflammatory response (for example, interleukins IL-2 and IL-6, the cytokine granulocyte–macrophage colony-stimulating factor and the B-cell receptor), general signaling cascades (such as mitogen-activated protein kinase, extracellular signal-regulated kinase, TRK and p38) and apoptosis. However, apoptosis and chemotaxis are two ones which activated via different sub signaling pathways in each of aforementioned diseases (97). At that time, Amit et al., also identified some new genes/SNPs leading to share some autoimmune diseases phenotype through genome-wide association studies (GWAS) and protein interaction network. Interaction of some autoimmune diseases associated genes with numerous environmental and endogenous factors indicates their crucial role in autoimmunity. Furthermore, interaction of newly associated genes has been reported with existing drugs which have been used long before the reorganization of these associated genes. Thus, progressive therapeutic strategies could be designed with grouping patients according to their risk allele(s) in specific genes that directly or

narrowly have interaction with the specified drugs. Hence, it is not only the further efficient molecular basis against these diseases which will be recognized by this drug-susceptible gene network but also it will be determined which drug could be more promising for those patients carrying risk allele(s) in that gene(98). Earlier results in this field are also noticeable. To begin with, Berghold et al. in 2007 developed an integrative analysis method for analysis, combining genetic interactions, and a high-confidence human protein interaction network to discover novel genes in type 1 diabetes as one of the uncured autoimmune diseases. Consequential networks were ranked by the outstanding of the enrichment of proteins from interacting regions. They identified a number of new protein network modules and novel candidate genes/proteins for type 1 diabetes (99). Along the same lines, Gao and Wang in 2009 investigated the proteins interactions efficacy in type 1 diabetes (T1D).Their study illustrates the potential of the PPI information in prioritizing positional candidate genes for T1D. Based on their results, it was brought out that the use of protein–protein interactions can immensely increase the possibility of finding positional candidate disease genes when applied on a large scale. Such invaluable systemic approaches can lead to novel candidate gene predictions (100).

Three years later, Bergholdt et al. also integrated type 1 diabetes GWAS (genome-wide association studies) data with protein-protein interactions to construct biological networks for type 1 diabetes. In this study, they were successful to identify 17 protein interaction networks which help them elucidate the mechanisms behind type 1diabetes pathogenesis and, thus, may provide the basis for the design of novel treatment strategies (101).

Multiple sclerosis (MS) is an idiopathic autoimmune neurodegenerative disease. To survey Multiple Sclerosis (MS) molecular mechanism through PPIs, Baranzini et al. in 2009 recognized several pathways according on network-based

analysis of genome-wide association studies in multiple sclerosis. Identification of sub-networks of genes from several immunological pathways including cell adhesion, communication and signaling was the main results of their studies. Remarkably, neural pathways, namely axon-guidance and synaptic potentiation, were also over-represented in MS. The potential involvement of neural pathways in MS susceptibility was the novelty of their woks as they discovered it for the first time (102). Later, Ragnedda et al. in 2012 conducted a protein-protein interaction (PPI) analysis of gene products coded in loci recently reported to be MS associated at the genome-wide significance level and in loci suggestive of MS association. They showed that the halves of genes identified by network analysis located in loci currently are suggestive in MS association. These are included some genes such as SYK, IL-6, CSF2RB, FCLR3, EIF4EBP2 and CHST12 which have immune-related functions. They conclude the fact that more common variants remain to be found as MS associated (103).Recently, in 2013, International Multiple Sclerosis Genetics Consortium identified several high-confidence candidates' genes by using a protein-interaction-network-based pathway analysis (PINBPA) on two large genetic MS studies comprising a total of15,317 cases and 29,529 controls. They believed that PINBPA is a powerful approach to gaining further insights into the biology of associated genes and to prioritizing candidates for subsequent genetic studies of complex traits (104).

Rheumatoid arthritis (RA) is another chronic autoimmune disease that primarily attacks synovial joints. Despite the advances in diagnosis and treatment of RA, novel molecular targets are still needed to improve the accuracy of diagnosis and the therapeutic outcomes. By integrating Rheumatoid Arthritis's gene expression data and protein interactions, You et al. in 2012 was able to reconstruct associated sub networks which delineate key RA associated cellular processes and transcriptional regulation. They concluded that exploiting such network models are capable of recognizing potential targets that will serve for some momentous clinical goals. For instance, they can be useful resources for the discovery of therapeutic targets and diagnostic markers, as well as providing novel insights into RA pathogenesis (105).Systemic lupus erythematosus (SLE) is a prototype systemic autoimmune disease characterized by flares of high morbidity. According to former studies, systems approach will enable us to discover myriad of facts about this lethal diseases. To achieve this purpose, Genome-wide pathway analysis of genome-wide association studies on Systemic Lupus Erythematosus (SLE) and Rheumatoid Arthritis (RA) was done by Young Ho Lee et al in 2012. They identified five candidate SNPs and thirteen pathways, involving bystander cytokine network, B cell activation and collagen metabolic processing, which may contribute to SLE susceptibility. Additionally, they revealed candidate causal non-HLA SNPs, genes, and pathways of RA (106).

One of the important autoimmune diseases in GI track is celiac disease. Celiac disease (CD) is autoimmune disorder which differentiated by an intestinal inflammation triggered by gluten (107). Similar to other autoimmune diseases mentioned before, CD is the result of an immune response to self-antigens leading to tissue damage and creation of autoantibodies such as tissue transglutaminase. The results of previous studies indicated that, gene's expression in the small intestine of patients with celiac disease is different from control patients. For this purpose, recently, Genome Wide Association Studies (GWAS) have been successful in finding genetic risk variants behind CD (108). These genes and GWAS pathway, together expose a new potential biological mechanism that could influence the genesis of celiac disease. On the other hands, the

development of GWAS technologies has lead to the discovery of more than 100 IBD loci which some genes are shared between Crohn's disease (CD) and ulcerative colitis (UC), and other are IBD subtype-specific   like autophagy genes, epithelial barrier gene (109).

Conclusion

It has been described in this review that protein interaction networks can elucidate the molecular basis of diseases, which in turn can appraise methods for prevention, diagnosis, and treatment. When the properties of these protein networks have been analyzed, novel higher order structures have been revealed. Therefore, it can provide an opportunity to interpret complex biological behaviors and alterations (in network dynamics) associated with complex diseases such as cancer and autoimmune diseases. These network relationships suggest a novel means of developing molecular therapies where the network is the target of therapy rather than individual molecules within the network. Hence, we expect that such systemic vantage of view should be applicable to complex diseases such as cancers and autoimmune diseases which are needed new efficient diagnosis and therapies, and offers new opportunities for enhancing our understanding of complex diseases.

References

1- Mileidy WG, Maricel GK. Chapter 4: protein interactions and disease. PLoS Comput Biol 2012; 8:e1002819.

2- Samira J, Patrick A. From protein interaction networks to novel therapeutic strategies. IUBMB Life 2012; 64: 529–37.

3- Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004;5: 101–13.

4- Grindrod P, Kibble. M Review of uses of network and graph theory concepts within proteomics. Expert Rev Proteomics 2004;11: 229–38.

5- Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, et al. High-quality binary protein interaction map of the yeast interactome network. Science. 2008;322:104-10.

6- De Las Rivas J, Fontanillo C. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. PLoS Comput Biol 2010; 6: e1000807.

7- Yook S H, Oltvai Z N, Barabasi A L. Functional and topological characterization of protein interaction networks. Proteomics 2004;4: 928–942.

8- Sharan R, Ulitsky I, Shamir R. Network-based prediction of protein function. Mol Syst Biol 2007; 3, 88.

9- Dittrich MT, Klau GW, Rosenwald A, Dandekar T, Muller T. Identifying functional modules in protein-protein interaction networks: an integrated exact approach. Bioinformatics 2008; 24: i223–31.

10- Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci USA 2003;100: 12123–12.

11- Jaeger S, Sers CT, Leser U. Combining modularity, conservation, and interactions of proteins significantly increases precision and coverage of protein function prediction. BMC Genomics 2010; 11:717.

12- Sharan R, Suthram S, Kelley RM, Kuhn T, McCuine S. Conserved patterns of protein interaction in multiple species. Proc Natl Acad Sci USA 2005; 1974–79.

13- Kann MG. Protein interactions and disease:computational approaches to uncover the etiology of diseases. Brief Bioinform 2007; 8: 333–46.

14- Erten S, Chowdhury SA, Guan X, Nibbe RK, Barnholtz-Sloan JS, Chance MR, et al Identifying stage-specific protein subnetworks for colorectal cancer. BMC Proceedings 2012; 6:S1.

15- Kann MG. Protein interactions and disease: computational approaches to uncover the etiology of diseases. Brief Bioinform. 2007;8:333-46.

16- Bollobás B, Editor. Random graphs. Cambridge, UK: Cambridge University Press; 2001.

17- Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 1998; 393, 440-42.

18- Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. Nature 2000; 406: 378-382.

19- Bartel PL, Roecklein JA, SenGupta D, Fields S. A protein linkage map of Escherichia coli bacteriophage T7. Nat Genet 1996; 12: 72-77.

20- Fromont-Racine M, Rain JC, Legrain P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nat Genet 1997; 16: 277-282.

21- Letamendia A, Labbe E, Attisano L. Transcriptional regulation by Smads: crosstalk between the TGF-beta and Wnt pathways. J Bone Joint Surg Am 2001; 83: S31-39.

22- Shuai K. Modulation of STAT signaling by STAT-interacting proteins. Oncogene 2000; 19: 2638-44.

23- Roberts PM. Mining literature for systems biology. Briefings in bioinformatics 2006;7: 399-406.

24- Walhout AJ, Vidal M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. Methods (San Diego, Calif) 2001; 24: 297-306.

25- Albert R. Scale-free networks in cell biology. J Cell Sci 2005; 118: 4947-57.

26- Batada NN, Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, Hurst LD, et al. Stratus not altocumulus: a new view of the yeast protein interaction network. PLoS Biol 2006; 4: e317.

27- Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 2009; 27: 199-204.

28- Bertin N, Simonis N, Dupuy D, Cusick ME, Han JD, Fraser HB, et al. Confirmation of organized modularity in the yeast interactome. PLoS Biol 2007; 5: e153.

29- Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, et al. A human protein-protein nteraction network: a resource for annotating the proteome. Cell 2005;122: 957–968.

30- Yu H, Zhu X, Greenbaum D, Karro J, Gerstein M. TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics Nucleic. Acids Res 2004; 32: 328–37.

31- Boccalettia S, Latora V, Moreno Y, Chavezf M, Hwanga Yook DU. Complex networks: Structure and dynamics. Physics Reports 2006; 424:4–5.

32- Li D, Li J, Ouyang S, Wang J, Wu S, Wan P, et al. Protein interaction networks of *Saccharomyces cerevisiae*, caenorhabditiselegans and drosophila melanogaster: Large-scale organization and robustness. Proteomics 2006;6:456–61.

33- Zhu D, Qin ZS. Structural comparison of metabolic networks in selected single cell organisms. BMC Bioinformatics 2005;6: 8.

34- Bergmann S, Ihmels J, Barkai N. Similarities and differences in genome-wide expression data of six organisms. PLoS Biol 2004; 2: E9.

35- Wuchty S, Almaas E. Stable evolutionary signal in a Yeast protein interaction network. BMC Evol Biol 2005; 5:24.

36- Shoemaker BA, Panchenko AR. Deciphering protein–protein interactions. Part I. Experimental techniques and databases. PLoS Comput Biol 2007; 3: e42.

37- Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED. The genetic landscape of a cell. Science 2010; 327: 425–31.

38- Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A 2001; 98: 4569–74.

39- Mrowka R, Patzak A, Herzel H. Is there a bias in proteome research? Genome Res 2001; 11:1971–73.

40- von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, et al. Comparative assessment of large-scale data sets of protein-protein interactions. Nature 2002; 417: 399–403.

41- Shoemaker BA, Panchenko AR, Bryant SH. Finding biologically relevant protein domain interactions: conserved binding mode analysis. Protein Sci 2006; 15: 352–61.

42- Sprinzak E, Margalit H. Correlated sequence-signatures as markers of protein-protein interaction. J Mol Biol 2001;311: 681–92.

43- Deng M, Mehta S, Sun F, Chen T. Inferring domain-domain interactions from protein- protein interactions. Genome Res 2002;12: 1540– 48.

44- Nye TM, Berzuini C, Gilks WR, Babu MM, Teichmann SA. Statistical analysis of domains in interacting protein pairs. Bioinformatics 2005;21: 993–1001.

45- Fraser HB, Hirsh AE, Wall DP, Eisen MB. Coevolution of gene expression among interacting

proteins. Proc Natl Acad Sci U S A2004;101: 9033–38.

46- Kanaan SP, Huang C, Wuchty S, Chen DZ, Izaguirre JA. Inferring protein-protein interactions from multiple protein domain combinations. Methods Mol Biol 2009;541: 43–59.

47- Guimaraes KS, Przytycka TM. Interrogating domain-domain interactions with parsimony based approaches. BMC Bioinformatics 2008; 9: 171.

48- Guimaraes KS, Jothi R, Zotenko E, PrzytyckaTM .Predicting domain-domain interactions using a parsimony approach. Genome Biol 2006;7: R104.

49- Gertz J, Elfond G, Shustrova A, Weisinger M, Pellegrini M. Inferring protein interactions from phylogenetic distance matrices. Bioinformatics 2003;19: 2039–45.

50- Jothi R, Cherukuri PF, Tasneem A, Przytycka TM .Co-evolutionary analysis of domains in interacting proteins reveals insights into domain-domain interactions mediating protein- protein interactions. J Mol Biol 2006; 362: 861–75.

51- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci U S A 1999;96: 4285–88.

52- Huynen MA, Bork P. Measuring genome evolution. Proc Natl Acad Sci U S A1998; 95: 5849– 5856.

53- Dandekar T, Snel B, Huynen M, Bork P. Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci 1998;2: 324–28.

54- Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. Use of contiguity on the chromosome to predict functional coupling. In Silico Biol 1999;1: 93–108.

55- Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO. Detecting protein function and protein-protein interactions from genome sequences. Science1999; 285: 751–53.

56- Enright AJ, IliopoulosI, Kyrpides NC, Ouzounis CA. Protein interaction maps for complete genomes based on gene fusion events. Nature 1999;402: 86–90.

57- Juan D, Pazos F, Valencia. A Co-evolution and co-adaptation in protein networks. FEBS Letter 2008 9;582:1225-30.

58- Barabási A, Gulbahce N, and LoscalzoJ .Network Medicine: A Network-based approach to human disease. Nat Rev Genet 2011; 12: 56–68.

59- Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol2007; 3: 140.

60- Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signaling circuits in molecular interaction networks. Bioinformatics 2002;18: S233–40.

61- Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S. Human protein reference database–2009 update. Nucleic Acids Res 2009;37: 767–72.

62- Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. Proc Natl Acad Sci U S A 2007;104: 8685–90.

63- Wachi S, Yoneda K, Wu R. Interactome transcript to me analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. Bioinformatics 2005;21: 4205–208.

64- Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. Bioinformatics 2006; 22: 2291–97.

65- Ideker T, Sharan R. Protein networks in disease. Genome Res 2008;18: 644–52.

66- Uetz P, Dong YA, Zeretzke C, Atzler C, Baiker A, Berger B, et al. Herpesviral protein networks and their interaction with the human proteome. Science 2006;311: 239–42.

67- Lim J, Hao T, Shaw C, Patel AJ, Szabó G, Rual JF, et al. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell 2006;125: 801–14.

68- Charlesworth JC, Curran JE, Johnson MP, Goring HH, Dyer TD. Transcriptomic epidemiology of smoking: the effect of smoking on gene expression in lymphocytes. BMC Med Genomics 2010;3: 29.

69- Sarajlić A, Janjić V, Stojković N, Radak D, Pržulj N. Network Topology Reveals Key Cardiovascular Disease Genes. PLoS ONE 2013;8: e71537.

70- Oti M, Snel B, Huynen MA, Brunner HG. Predicting disease genes using protein-protein interactions. J Med Genet 2006;43: 691–98.

71- Gandhi TK, Zhong J, Mathivanan S, Karthick L, Chandrika KN, Mohan SS, et al. Analysis of the human protein interactome and comparison with

yeast, worm and fly interaction datasets. Nat Genet 2006;38: 285–93.

72- Chen L, Tai J, Zhang L, Shang Y, Li X, Qu X, et al. Global risk transformative prioritization for prostate cancer candidate genes in molecular networks. Mol Biosyst 2011;7: 2547–53.

73- Jiang Z, Zhou Y .Using bioinformatics for drug target identification from the genome. Am J Pharmacogenomics 2005;5: 387–96.

74- Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M. Drug-target network. Nat Biotechnol 2007; 25: 1119–26.

75- Bauer-Mehren A, Bundschus M, Rautschka M, Mayer MA, Sanz F, Furlong LI. Gene-disease network analysis reveals functional modules in Mendelian, complex and environmental diseases PLoS ONE 2011; 6: e20284.

76- Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JD, Hao T, et al. Predictive models of molecular machines involved in Caenorhabditiselegans early embryogenesis. Nature 2005;436:861-65.

77- Goñi J, Esteban FJ, de Mendizábal NV, Sepulcre J, Ardanza-Trevijano S, Agirrezabal I, et al. A computational analysis of protein-protein interaction networks in neurodegenerative diseases BMC Systems Biology 2008; 2:52.

78- George RA, Liu JY, Feng LL, Bryson-Richardson RJ, Fatkin D, Wouters MA. Analysis of protein sequence and interaction data for candidate disease gene prediction. Nucleic Acids Res 2006;34: e130.

79- Liu B, Jiang T, Ma S, Zhao H, Li J, Jiang X, et al. Exploring candidate genes for human brain diseases from a brain-specific gene network. Biochem Biophys Res Commun 2006;349: 1308–14.

80- Emre G, Baldo O. Exploiting protein-protein interaction networks for genome-wide disease-gene prioritization. PLoS One 2012;7:e43557

81- Hishigaki H, Nakai K, Ono T, Tanigami A, Takagi T. Assessment of prediction accuracy of protein function from protein–protein interaction data. Yeast 2001;18: 523–31.

82- Chua HN, Sung WK, Wong L. Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. Bioinformatics 2006;22: 1623–30.

83- Wu X, Jiang R, Zhang MQ, Li S. Network-based global inference of human disease genes. Mol Syst Biol 2008;4: 189.

84- Sam L, Liu Y, Li J, Friedman C, Lussier YA. Discovery of protein interaction networks shared by diseases. Pac Symp Biocomput 2007:76-87.

85- Ozgur A, Vu T, Erkan G, Radev DR. Identifying gene-disease associations using centrality on a literature mined gene-interaction network. Bioinformatics 2008;24: i277-85.

86- Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci U S A 2003;100: 12123–28.

87- Ergun A, Lawrence CA, Kohanski MA, Brennan TA, Collins J. Network biology approach to prostate cancer. Mol Syst Biol 2007;3: 82.

88- Jingchun S, Zhongming Zh. A comparative study of cancer proteins in the human protein-protein interaction network BMC Genomics 2010; 11:S5.

89- Li L, Kangyu Zh. Discovering cancer genes by integrating network and functional properties BMC Med Genomics 2009; 2:61.

90- Milenkovic T, Memisevic V, Ganesan AK, Przulj N. Systems-level cancer gene identification from protein interaction network topology applied to melanogenesis-related functional genomics data. J R Soc Interface 2010; 7: 423–37

91- Liu H, Su J, Li J, Liu H, Lv J, Li B, et al. Prioritizing cancer-related genes with aberrant methylation based on a weighted protein-protein interaction network. BMC Syst Biol 2011;5:158.

92- Zhang L, Li X, Tai J, Li W, Chen L. Predicting candidate genes based on combined network topological features: a case study in coronary artery disease. PLoS One 2012;7: e39542.

93- Chao W, Jun Zh, Xuegong Zh. Integrating gene expression and protein-protein interaction network to prioritize cancer-associated genes BMC Bioinformatics 2012; 13:182.

94- houra M, Rebaï A. Topological features of cancer proteins in the human NR-RTK interaction network. J Recept Signal Transduct Res 2012; 32:257-62.

95- Noel R, Rose IRM. The Autoimmune Diseases. Academic Press: New York, 2009.

96- Karopka T, Fluck J, Mevissen HT, Glass A. The autoimmune disease database: dynamically compiled literature-derived database. BMC Bioinform 2006; 7: 325.

97- Tuller T, Atar S, Ruppin E, Gurevich M, Achiron A. Common and specific signatures of gene

expression and protein-protein interactions in autoimmune diseases. Genes Immun 201; 14:67-82.

98- Maiti K, Nath K. Gene network analysis of small molecules with autoimmune disease associated genes predicts a novel strategy for drug efficacy. Autoimm Rev 2013;12: 510–22.

99- Bergholdt R, Størling ZM, Lage K, Karlberg EO, Olason PI, Aalund M, et al. Integrative analysis for finding genes and networks involved in diabetes and other complex diseases. Genome Biol 2007; 8:R253.

100- Shouguo G, Wang X. Predicting type 1 diabetes candidate genes using human protein-protein interaction networks. J Comput Sci Syst Biol 2009; 2: 133.

101- Bergholdt R, Brorsson C, Palleja A, Berchtold LA, Fløyel T, Bang-Berthelsen CH, 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;61:954-62.

102- Baranzini SE. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. Human Molecular Genetics 2009; 18: 33-43.

103- Ragnedda G, Disanto G, Giovannoni G, Ebers GC, Sotgiu S, Ramagopalan SV. protein-protein interaction analysis highlights additional loci of interest for Multiple Sclerosis PLoS One 2012;7: e46730.

104- International Multiple Sclerosis Genetics Consortium. Network-based multiple sclerosis pathway analysis with gwas data from 15,000 cases and 30,000 controls. Am J Hum Genet. 2013; 92: 1–12.

105- You S, Cho CS, Lee I, Hood L, Hwang D, Kim WU. A systems approach to rheumatoid arthritis. PLoS One 2012; 7: e51508.

106- Lee YH, Bae SC, Choi SJ, Ji JD, Song GG. Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. Mol Biol Rep. 2012;39:10627-35.

107- Rostami Nejad M, Ehsani-Ardakani MJ, Dabiri R, Aldulaimi D, Rostami K. Geographic trends and Risk of Gastrointestinal Cancer among Patients with Celiac Disease in Asian pacific regions. Gastroenterol Hepatol Bed Bench 2013;6:170-177

108- Östensson M, Montén C, Bacelis J, Gudjonsdottir AH, Adamovic S, Ek J, et al. A possible mechanism behind autoimmune disorders discovered by genome-wide linkage and association analysis in celiac disease. PLoS One 2013; 8:e70174.

109- Atella G, Fiocchi C, Caprili R. News from the "5th International Meeting on Inflammatory Bowel Diseases" CAPRI 2010. J Crohns Colitis. 2010;4:690-702.