

Protein-Protein Interactions and Disease

Aditya Rao, Gopalakrishnan Bulusu,
Rajgopal Srinivasan and Thomas Joseph
*Life Sciences Division, TCS Innovation Labs, Tata Consultancy Services, Hyderabad
India*

1. Introduction

Protein-protein interactions (PPI), in which, two or more proteins associate with each other by various means, are key to understanding all biological processes that occur within as well as between cells. In effect, biological processes are essentially interactions between multiple proteins (Zhang et al., 2011) with PPI networks controlling the flow of information both within and between biological processes.

Disruptions in PPI networks have been shown to result in diseases. This includes monogenic diseases such as hemophilia where a particular biochemical pathway is disrupted, as well as more complex diseases such as cancer, which involve several signaling pathways (Sam et al., 2007). Conversely, disruption of a set of PPI can lead to a particular disease or, in the case where the set is shared among several networks, to several diseases. While there is a wealth of protein-disease associations in the published literature that have been incorporated in PPI repositories, the challenge is to link such PPI to human disease (Ideker & Sharan, 2008).

In this chapter we discuss several examples of diseases that are caused by disruptions of PPI networks. Our goal is to illustrate through examples how the role of PPI in disease can be studied using a variety of computational tools and data sources. While we discuss tools and data sources that are of general interest, we also discuss methods for studying specific diseases and methods aimed at large scale analysis of PPI data to identify classes of diseases. In each case we provide specific examples from the literature and a brief discussion of the tools used.

2. PPI and disease example

Let us consider cerebral malaria as an example to understand how an analysis of PPI could be used to elucidate the molecular basis of disease. Here, a wide range of experimental and predicted human-*Plasmodium* (host-parasite), human-human (host-host) and *Plasmodium-Plasmodium* (parasite-parasite) PPI are combined and analyzed in the context of key events and processes of cerebral malaria, a dangerous infectious disease (Rao et al., 2010).

Cerebral malaria is a severe form of malarial infection, characterized by cerebral complications, such as neuronal damage and coma (Moxon et al., 2009). The disease is characterized by processes such as sequestration of infected red blood cells to cerebral capillaries and venules, systemic inflammation, hemostasis dysfunction and neuronal

damage (van der Heyde et al., 2006; Wilson et al., 2008). PPI datasets from different sources were first obtained, summarized in Table 1. Since each dataset uses a different nomenclature system for the human and parasite proteins, a crucial step was to normalize all datasets using common gene names. This enabled creation of a unified host-parasite PPI dataset.

No	Source	Reference
1	Davis dataset	Davis et al., 2007
2	Dyer dataset	Dyer et al., 2007
3	Krishnadev dataset	Krishnadev & Srinivasan, 2008
4	Vignali dataset	Vignali et al., 2008
5	Literature PPI data	In-house manual curation

Table 1. Protein-Protein interaction datasets used in the cerebral malaria example.

An automated literature retrieval module was developed using Entrez Programming Utilities (Sayers et al., 2010) to retrieve the list of full-text articles relevant to the malarial parasite. This article set was pruned using the Medical Subject Headings (MeSH) controlled vocabulary for articles relevant to cerebral malaria. The resultant set was augmented by articles retrieved from the Google Scholar database using appropriate disease-specific query terms such as systemic inflammation, hemostasis dysfunction etc. This article corpus had two main uses:

- For extracting biochemical and signaling events of relevance in cerebral malaria.
- Identifying pairs of interacting proteins within the host, within the parasite and between host and parasite.

Gene Ontology (GO) cellular component annotations from PlasmoDB (Aurrecochea et al., 2009), a comprehensive *Plasmodium* resource, were used to prune the unified PPI dataset using the approach of Mahdavi & Lin (2007). In the case of PPI involving parasite proteins, only those proteins that were annotated to be present on the parasite surface or were reported to be released during the relevant stage of the parasite were considered (Lyon et al., 1986). For the human protein annotations, tissue-specific annotations from UniProt (Hubbard et al., 2009) were used in the pruning process.

The resultant PPI subset was then analyzed by mapping the PPI to key events that influence the processes of the disease, as identified from the key review articles. The analysis showed the potential significance of apolipoproteins and heat-shock proteins on efficient *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) presentation, role of the merozoite surface protein (MSP-1) in platelet activation, the role of albumin in astrocyte dysfunction and the effect of parasite proteins in transforming growth factor (TGF)- β regulation. The linking of these PPI to molecular events associated with the disease pathogenesis provides a basis for further experiments to determine the molecular basis of this fatal disease.

3. Tools of the trade

From the example, it is clear that the underpinnings for mapping PPI to disease are: (a) access to various repositories of PPI and (b) ability to filter these PPI in the context of disease

and (c) using different tools for visualizing and analyzing the PPI in the context of diseases. Let us consider each of these in detail.

3.1 PPI repositories

There are a host of repositories that house experimental and predicted PPI data. The cerebral malaria example above considered malaria-specific PPI datasets. However, generic datasets such as BIND, DIP, HPRD, MINT, MIPS and STRING usually have the necessary PPI coverage required for a variety of disease studies.

The Biomolecular Interaction Network Database (BIND), a constituent database of the Biomolecular Object Network Databank, makes available a comprehensive collection of information for specific molecules such as proteins and small molecules (Bader et al., 2003). BIND has been one of the major sources of curated biomolecular interactions, especially PPI. The Database of Interacting Proteins (DIP) contains experimentally determined PPI (Salwinski et al., 2004). It has been created using both manual curation and computational approaches. The Human Protein Reference Database (HPRD) provides a platform to visually depict and integrate information, which are manually curated, pertaining to domain architecture, post-translational modifications, PPI networks and disease association for each protein of the human proteome (Prasad et al., 2009).

The Molecular INTeraction database (MINT) contains experimentally verified PPI that have been manually curated from the scientific literature (Ceol et al., 2010). The Mammalian Protein-Protein Interaction (MIPS) database is a collection of manually curated high-quality PPI data collected from the scientific literature by expert curators (Pagel et al., 2005). STRING is a database of known and predicted protein interactions (Szklarczyk et al., 2011). The interactions include direct (physical) as well as indirect (functional) associations.

Composite PPI resources are also available that integrate PPI data from some of these databases into a single resource. APID (Agile Protein Interaction DataAnalyzer), for instance, is one such resource that integrates experimentally validated PPI from databases such as BIND, DIP, HPRD and MINT, amongst others (Prieto et al., 2006). Protein Interaction Network Analysis (PINA) platform is another example of a composite PPI resource that integrates interactions from MINT, DIP, HPRD and MIPS, amongst others (Wu et al., 2009).

3.2 Integration and filtering of PPI

Databases and tools such as Reactome, GO, MeSH and the Entrez Programming Utilities are crucial for filtering the large number of PPI to obtain a PPI network relevant to a specific disease.

Reactome is a database of biological pathways from various organisms, especially humans (Matthews et al., 2009). This is manually curated by experts. It contains various entities such as proteins, chemicals, localization data, etc. The information in Reactome is cross-referenced to various standard bioinformatics databases such as Entrez Gene, UniProt, Ensembl, etc. The GO project attempts to standardize the description of gene and gene products across species and databases (The Gene Ontology Consortium, 2000). It consists of three ontologies that describe genes and gene products in relation to biological processes, molecular functions and cellular components.

MeSH (<http://www.nlm.nih.gov/mesh/>) is a controlled vocabulary thesaurus maintained by the National Library of Medicine. It is made up of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. It currently has 16 major tree headings including "Diseases" and "Chemicals and Drugs". MeSH terms are used in various methods and tools to filter articles/abstracts and other data. Online Mendelian Inheritance in Man (OMIM) is a database of known Mendelian disorders and their related genes (Hamosh et al., 2002). Currently, there are around 12,000 genes described in this database. OMIM provides information on genotype-phenotype relationships in human Mendelian diseases.

Specific tools are available to access some of these databases, such as the Entrez Programming Utilities (Sayers et al., 2010). These are a set of server-side programs enabling a stable interface to utilize the Entrez query and database system at the National Center for Biotechnology Information. There are currently 38 databases in the Entrez system with a wide variety of information on nucleotide and protein structure and sequences, 3D-molecular structures, disease information and biomedical literature etc.

3.3 Visualization tools

Important tools that could aid in mapping PPI to disease include:

- Cytoscape (Shannon et al., 2003) for visualizing PPI datasets with nodes representing biological entities and edges representing the relationships between these entities
- Cell Circuits (Mak et al., 2007) for comparison of hand-curated pathway models to hypothetical models derived from large-scale 'omic' data.

Cytoscape plugins such as APID2NET (Hernandez-Toro et al., 2007) and PRINCIPLE (Gottlieb et al., 2011) are also very pertinent. APID2NET retrieves PPI data from the APID server for further analysis within the Cytoscape environment. PRINCIPLE, discussed later in this chapter, is built specifically for exploring PPI-disease associations. Given any disease as a query term, it provides a list of top-ranking genes associated with this disease and a Cytoscape visualization of the sub-networks formed by these genes and their direct interacting neighbors.

IPA (Ingenuity Systems, www.ingenuity.com) is an example of a commercially available platform that enables visualization of dynamically constructed pathway and network models.

4. PPI from literature

What happens when the PPI repositories do not have adequate coverage of the organism or specific protein-set under study? One possibility is that although such repositories do not have these PPI, the PPI have actually been reported in literature. One just needs to go look for them!

Let us consider an example of using text-mining to extract such PPI. In the cerebral malaria study, a basic text-mining approach has been used (Rao et al., 2010). The article corpus was first checked for article-level co-occurrence of pairs of proteins. Full-text articles, wherever available, are automatically downloaded from the respective journal

websites as Portable Document Format (PDF) files and converted to text format using the XPDF conversion utility (The FooLabs, <http://www.foolabs.com/xpdf>). All parasite and host proteins that occur in the full-text of each article were identified using a dictionary lookup approach, with PlasmoDB and UniProt/Ensembl being used to create the parasite and human protein dictionaries respectively. Only those articles that had at least one protein pair (host-parasite, host-host or parasite-parasite) were considered for further analysis.

Özgür et al. (2008) propose a more detailed approach based on integrating automatic text mining and network analysis methods to extract known disease genes and to predict unknown disease genes. They started by collecting an initial set of seed genes known to be related to a disease from curated databases such as OMIM. A disease specific gene network was created using advanced natural language processing techniques that capture both gene names as well as the semantic associations between them.

5. PPI Networks and SNPs

Genome-wide measurement technologies such as microarrays have provided an opportunity to identify genes that are mutated or differentially expressed. In particular, SNP-arrays have been very useful in such studies and have resulted in identification of several genes that are associated with disease-risk or poor prognosis (Karinen et al., 2011). Such genes typically affect cellular functions by altering signaling in regulatory PPI networks.

The mainstay of this approach is the fact that genes related to the same disease are also known to have protein products that physically interact (Navlakha & Kingsford, 2010). However, that by itself is only one crucial component. The other important component is that a genetic disease is associated with a linkage interval on the chromosome if SNPs in the interval are correlated with an increased susceptibility to the disease. These linkage intervals define a potential disease-causing gene set. The computational approach boils down to using both these sources of information—PPI networks and linkage intervals to predict relationships between genes and diseases.

Let us look at a method called CANGES to identify the genetic basis of disease (Karinen et al., 2011). The strength of the method lies in its ability to cohesively integrate many different pieces of information to arrive at testable hypotheses. Genome wide association studies have identified many variations that are possibly linked to one or more diseases. How does one go about prioritizing these variations to get to a set of genes that cause the disease? Clearly, one needs to bring in other known information to help arrive at a decision. The CANGES method combines pathway data, PPI data and genetic variation data with analytical tools to rapidly evaluate the disease causing potential of variations and thus focus attention on one or a few genes. Using this method, a set of 158 SNPs in the p53 gene were identified that plays a central role in cancers. These SNPs are likely to have pathogenic consequences. The same method has also been used, in conjunction with clinical patient data, to identify genes associated with glioblastoma multiforme. It is clear that in the future we will see many more such methods which bring together PPI with several other pieces of information and analytical tools to identify disease genes and gene networks.

Several computational methods can be used to identify causal genes central to gene-disease relationships from large PPI networks. The methods include network neighbors and neighborhood methods, unsupervised graph partitioning and Markov clustering, semi-supervised graph partitioning, random walks, network flow methods and several of their variants (Navlakha & Kingsford, 2010). Navlakha & Kingsford tested these on two large PPI networks: (a) one derived from the Human Protein Reference Database, consisting of 8776 proteins and 35,820 PPI and (b) the other derived from Online Predicted Human Interaction Database containing 9842 proteins and 73130 PPI. Annotations from OMIM were used to associate diseases with genes and linkage intervals. They observed that the performance of most methods showed a significant correlation with neighborhood homophily. Based on this, they suggest that homophily could be used to assess the quality of network-based predictions of disease-protein relationships. They also observed that the individual methods capture different kinds of structure in the network and these unique abilities can be used together in a consensus method to enhance prediction quality.

6. Structural significance of PPI

One important disease class in which a study of PPI could shed light is cancer. Let us look at a study that analyzed cancer proteins in human PPI networks (Kar et al., 2009). This study is important from a methodology perspective as it uses structural properties of the proteins present in the PPI network. Integrating three-dimensional protein structural information into PPI networks revealed important aspects about cancer-related proteins. Analysis of the structural properties of cancer-related interface proteins showed that the interfaces are, on an average, smaller in size, more planar, less tightly packed and more hydrophilic than those of non-cancer proteins. For instance, in a breast cancer network used in the study, there was significant accuracy in discriminating cancer-protein interfaces from the non-cancer interfaces. Thus, there seems to be a clear distinction between the interfaces.

In addition, they observed that cancer-related proteins tend to interact with their partners via multi-interface hubs, which comprise 56% of cancer-related proteins. Cancer protein networks are therefore more enriched in multi-interface proteins. Cancer proteins, in general, are longer and have larger surface areas. Thus, to participate in many PPI at the same time, these tend to be multi-interface hubs, with distinct interfaces interacting with different proteins.

The processes involved in obtaining relevant PPI with regard to a disease are shown in Figure 1.

7. PPI common across diseases

The hitherto discussed examples link diseases with their possible proteomic underpinnings. Research is also underway that focuses on bridging the gap between PPI and their association to different diseases. The goal is to bring out underlying PPI that are common amongst different sets of diseases. Diseases with overlapping clinical phenotypes are caused by mutations in functionally related genes. Since PPI are the strongest manifestation of a functional relationship between disease genes, applying a network model is an effective approach for revealing the associations among diseases (Zhang et al., 2011).

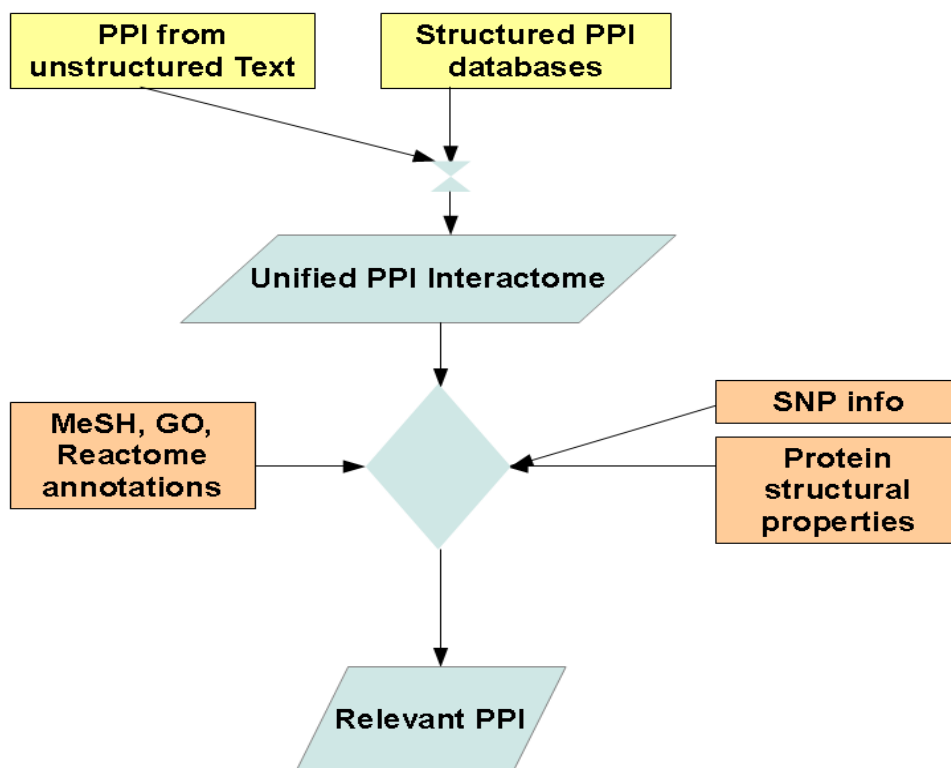


Fig. 1. Processes involved in obtaining relevant PPI with regard to a disease.

7.1 Background

Traditionally, diseases are defined as 'similar' mainly by their clinical appearance, with no correlation to underlying molecular processes. Conceptually, each monogenic disease has a collection of specific phenotypic features. This is true for about 2000 human single gene diseases with a defined genetic phenotype. Syndromes are defined in medicine as a set of phenotypes which, occurring together, serve to define a trait or disease. However, phenotypes very often overlap in the case of many syndromes. Recognition of this overlap brought about the concept of 'syndrome families' taking into account the common features shared between diseases (Sam et al., 2007).

The clustering of syndromes into these families in combination with genetic insights has led to the discovery that what were often thought as two different disorders were really variable expressions of the same disorder. On the other hand, it has long been known that mutations at different loci can lead to the same genetic disease. It has also been hypothesized that this genetic heterogeneity has its roots at the PPI level, suggesting that other genes associated with the phenotype also have some functional role. Therefore, it is plausible that functional

properties of shared molecular networks reflect phenotypic overlap of diseases. Thus, PPI networks provide unique opportunities for exploring disease pathways (Sam et al., 2007).

Let us continue with cancer as a disease theme. Sam et al. (2007) highlight an example that links Fanconi's Anemia and cancer. Fanconi's Anemia is a hereditary DNA-repair deficiency disease characterized by defects in a set of DNA repair proteins, leading to, among others, hypersensitivity to DNA damaging agents. This disorder is caused by a mutation in any one of the genes in Fanconi's Anemia complementation group. Symptoms of the disease include anemia, several congenital malformations, etc. Importantly, patients suffering from it exhibit a strong predisposition to different cancers. In the study, this link was substantiated with 14 potential PPI common between Fanconi's Anemia and colorectal neoplasms.

7.2 PPI and common phenotypes

Let us consider another example where a PPI network has been systematically combined with disease-protein relationship data derived from mining GO annotations with phenotypic context (Sam et al., 2007). PPI associated with pairs of diseases were identified and the statistical significance of the occurrence of interactions in the protein interaction knowledgebase calculated. This study demonstrates that the associations between diseases are directly correlated to their underlying PPI networks. A subset of PhenoGO (Lussier et al., 2006; Sam et al., 2009) restricted to human diseases was examined to study the relationships between diseases according to the following criteria. Two basic types of relationships were considered, which determine whether two diseases share PPI networks: a) an identity relationship where common proteins are shared by two diseases, and b) direct interactions between protein A of one disease and protein B of another. A total of 10 pairs of diseases were identified that are significantly correlated due to their shared proteins and PPI. These pairs were analyzed based on mentions in literature, and their correlations were confirmed.

Xeroderma pigmentosum and Cockayne syndrome provide an example of how two diseases are correlated through their PPI networks. Xeroderma pigmentosum is a disorder causing susceptibility of the skin to ultraviolet radiation as a result of deficiencies in one of the XPA-XPG complementation group genes involved in nucleotide excision repair (NER). Cockayne syndrome results from deficiencies in transcription-coupled repair genes, like ERCC6 and ERCC8, leading to a number of conditions including abnormal sensitivity to sunlight (Sam et al., 2007; Spivak et al., 2004). There were 27 direct PPI and 5 common proteins shared amongst these two diseases. Majority of the proteins in the common networks between the two diseases are related to DNA repair processes - Global Genomic NER and Transcription-coupled NER. While the Global Genomic NER repairs lesions from non-transcribed regions of genome independent to transcription, the Transcription-coupled NER repairs UV induced damage in the transcribed strands of active genes. Both the diseases are seen to be associated with these processes, suggesting defects in the DNA damage repair processes are the cause of the diseases.

7.3 Of PRINCE and PRINCIPLES

PRINCIPLE is very relevant tool specifically built for finding out common diseases based on PPI. It is a Cytoscape plugin implementation of the PRINCE algorithm (Vanunu, et al.,

2010). Given a query disease, it provides a list of top ranking genes associated with it and an additional visualization of the sub-networks formed by these top ranking genes and their direct interacting neighbors. The underlying logic is that genes causing similar diseases often lie close to one another in a PPI network (Oti & Brunner, 2007; Oti, et al., 2006). Given a disease as the query term, PRINCE (a) identifies a set of phenotypically similar diseases, (b) retrieves the known causal genes of these diseases based on their similarity to the query and (c) propagates the scores of the prior set of genes over a human PPI network to provide association scores for all genes. It uses a comprehensive set of weighted PPI compiled from disparate sources (Vanunu, et al., 2010), disease-disease similarity measures (van Driel, et al., 2006), and on the disease-gene associations present in OMIM.

7.4 Human disease network – The holy grail!

Zhang et al. (2011) constructed an expanded Human Disease Network by combining disease-gene information with PPI information. Work such as this is very important, since a network model to represent relationships between diseases is very useful in looking at relationships amongst diseases on a large scale. Analysis of the network's topological features and functional properties showed that the network was hierarchical. Most diseases in the network were connected to only a few diseases, while a small set of diseases were linked to many different diseases. Also, diseases in a specific disease category tended to cluster together, and genes associated with the same disease were functionally related. While this might intuitively sound obvious, it establishes a molecular basis for disease-disease associations.

The limitation of the network is that only known and available disease phenotypic data has been incorporated. However, as more data is made available in databases and in literature, this network provides an ideal template to analyze relationships amongst diseases from a PPI perspective.

8. The road ahead

This is a new field and there are many more approaches than what has been brought out in this chapter. For instance, Bandyopadhyay et al. (2006) use a network analysis of gene expression and PPI data to identify active pathways related to HIV pathogenesis. A functional analysis of the detected sub-networks provides useful insights into various stages of the HIV replication cycle. Chen et al. (2006) developed a framework to mine disease-related proteins from OMIM and PPI data. They demonstrate the power of their method by applying it to Alzheimer's disease. The key to their method is a scoring function that ranks proteins according to their relevance to a particular disease pathway.

Methods to arrive at high-precision predictions that are translatable to effective steps in disease prevention, diagnosis and prognosis should be the goal of PPI studies. The generated leads should be tested experimentally to determine their relevance.

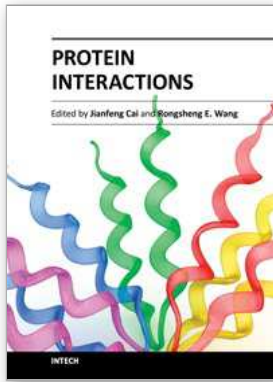
9. References

Aurrecochea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W,

- Miller JA, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoeckert CJ Jr, Treatman C & Wang H. (2009). PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Research* D539-543.
- Bader GD, Betel D & Hogue CW. (2003). BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Research* 31:248-250.
- Bandyopadhyay S, Kelley R & Ideker T. (2006). Discovering regulated networks during IV-1 latency and reactivation. *Pacific Symposium on Biocomputing* 354-366.
- Ceol A, ChatrAryamontri A, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L & Cesareni G. (2010). MINT, the molecular interaction database: 2009 update. *Nucleic Acids Research* (Database issue):D532-D539.
- Chen JY, Shen C & Sivachenko AY. (2006). Mining Alzheimer disease relevant proteins from integrated protein interactome data. *Pacific Symposium on Biocomputing* 367-378.
- Davis FP, Barkan DT, Eswar N, McKerrow JH & Sali A. (2007). Host pathogen protein interactions predicted by comparative modeling. *Protein Science* 16:2585-2596.
- Dyer MD, Murali TM & Sobral BW. (2007). Computational prediction of host pathogen protein-protein interactions. *Bioinformatics* 23:59-66.
- Gottlieb A., Magger O., Berman I., Ruppin E. & Sharan R. (2011). PRINCIPLE: A tool for associating genes with diseases via network propagation. *Bioinformatics*.doi: 10.1093/bioinformatics/btr584.
- Hernandez-Toro J, Prieto C & De las Rivas J. (2007). APID2NET: unified interactome graphic analyzer. *Bioinformatics* 23:2495-2497.
- Hamosh A, Scott AF, Amberger J, Bocchini C, Valle D & McKusick VA. (2002). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research* 30:52-55.
- Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, Braghi E, Brent S, Chen Y, Clapham P, Clarke L, Coates G, Fairley S, Fitzgerald S, Fernandez-Banet J, Gordon L, Graf S, Haider S, Hammond M, Holland R, Howe K, Jenkinson A, Johnson N, Kahari A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Kulesha E, Lawson D, Longden I, Megy K, Meidl P, Overduin B, Parker A, Pritchard B, Rios D, Schuster M, Slater G, Smedley D, Spooner W, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wilder S, Zadissa A, Birney E, Cunningham F, Curwen V, Durbin R, Fernandez-Suarez XM, Herrero J, Kasprzyk A, Proctor G, Smith J, Searle S & Flicek P. (2009). Ensembl 2009. *Nucleic Acids Research* 37:D690-697.
- Ideker T & Sharan R. (2008). Protein networks in disease. *Genome Research* 18: 644-652.
- Kann M, Ofran Y, Punta M & Radivojac P. (2006). Protein Interactions and disease. *Pacific Symposium on Biocomputing* 11:351-353.
- Kar G, Gursoy A & Keskin O. (2009). Human cancer protein-protein interaction network: a structural perspective. *PLoS Computational Biology* 5:e1000601.
- Karinen S, Heikkinen T, Nevanlinna H & Hautaniemi S. (2011). Data integration workflow for search of disease driving genes and genetic variants. *PLoS One* 6:e18636.
- Krishnadev O & Srinivasan N. (2008). A data integration approach to predict host- pathogen protein-protein interactions: application to recognize protein interactions between human and a malarial parasite. *In Silico Biology* 8:235-250.
- Lussier Y, Borlawsky T, Rappaport D, Liu Y, Friedman C. (2006). PhenoGO: assigning phenotypic context to gene ontology annotations with natural language processing. *Pacific Symposium on Biocomputing* 64-75.

- Lyon JA, Haynes JD, Diggs CL, Chulay JD & Pratt-Rossiter JM. (1986). *Plasmodium falciparum* antigens synthesized by schizonts and stabilized at the merozoite surface by antibodies when schizonts mature in the presence of growth inhibitory immune serum. *Journal of Immunology* 136:2252-2258.
- Mahdavi MA & Lin YH. (2007). False positive reduction in protein-protein interaction predictions using gene ontology annotations. *BMC Bioinformatics* 8:262.
- Mak HC, Daly M, Gruebel B & Ideker T. (2007). CellCircuits: a database of protein network models. *Nucleic Acids Research* 35:D538-545.
- Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, Mahajan S, May B, Schmidt E, Vastrik I, Wu G, Birney E, Stein L & D'Eustachio P. (2009). Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Research (Database issue)*: D619-622.
- Moxon CA, Heyderman RS & Wassmer SC. (2009). Dysregulation of coagulation in cerebral malaria. *Molecular and Biochemical Parasitology* 166:99-108.
- Navlakha S & Kingsford C. (2010). The power of protein interaction networks for associating genes with diseases. *Bioinformatics* 26:1057-1063.
- Oti M & Brunner HG. (2007). The modular nature of genetic diseases. *Clinical Genetics* 71(1):1-11.
- Oti M, Snel B, Huynen MA & Brunner HG. (2006). Predicting disease genes using protein-protein interactions. *Journal of Medical Genetics* 43(8):691-698.
- Özgür A, Vu T, Erkan G & Radev DR. (2008). Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics* 24, i277-i1285.
- Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stümpflen V, Mewes HW, Ruepp A, Frishman D (2005). The MIPS mammalian protein-protein interaction database. *Bioinformatics* 21(6):832-834.
- Prasad TSK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R & Pandey A. (2009). Human Protein Reference Database - 2009 Update. *Nucleic Acids Research* 37: D767-72.
- Prieto C. & De Las Rivas J. (2006). APID: Agile Protein Interaction DataAnalyzer. *Nucl. Acids Research* 34:W298-W302.
- Rao A, Kumar MK, Joseph T & Bulusu G. (2010). Cerebral malaria: Insights from host-parasite protein protein interactions. *Malaria Journal* 9:155.
- Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU & Eisenberg D. (2004). The Database of Interacting Proteins: 2004 update. *Nucleic Acids Research* 32:D449-451.
- Sam L, Liu Y, Li J, Friedman C & Lussier YA. (2007). Discovery of protein interaction networks shared by diseases. *Pacific Symposium on Biocomputing* 76-87.
- Sam LT, Mendonça EA, Li J, Blake J, Friedman C & Lussier YA. (2009). PhenoGO: an integrated resource for the multiscale mining of clinical and biological data. *BMC Bioinformatics* 10:S8.
- Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V,

- Mizrachi I, Ostell J, Panchenko A, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, John Wilbur W, Yaschenko E & Ye J. (2010). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 38:D5-16.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B & Ideker T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13:2498-2504.
- Spivak G. (2004). The many faces of Cockayne syndrome. *Proceedings of the ational Academy of Sciences of the United States of America* 101(43):15273-15274.
- Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguéz P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research* 39(Database issue):D561-568.
- The Gene Ontology Consortium. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics* 25(1):25-29.
- The UniProt Consortium. (2009). The Universal Protein Resource (UniProt). *Nucleic Acids Research* 37:D169-D174.
- van der Heyde HC, Nolan J, Combes V, Gramaglia I & Grau GE. (2006). A unified hypothesis for the genesis of cerebral malaria: sequestration, inflammation and hemostasis leading to microcirculatory dysfunction. *Trends in Parasitology* 22:503-508.
- van Driel MA, Bruggeman J, Vriend G, Brunner HG & Leunissen JA. (2006). A text-mining analysis of the human phenome. *European Journal of Human Genetics* 14(5):535-542.
- Vanunu, O., Magger, O., Ruppín, E., Shlomi, T. & Sharan R. Associating genes and protein complexes with disease via network propagation. (2010). *PLoS Computational Biology* 6, e1000641.
- Vignali M, McKinlay A, LaCount DJ, Chettier R, Bell R, Sahasrabudhe S, Hughes RE & Fields S. (2008). Interaction of an atypical *Plasmodium falciparum* ETRAMP with human apolipoproteins. *Malaria Journal* 7:211.
- Wilson NO, Huang MB, Anderson W, Bond V, Powell M, Thompson WE, Armah HB, Adjei AA, Gyasi R, Tettey Y & Stiles JK. (2008). Soluble factors from *Plasmodium falciparum*-infected erythrocytes induce apoptosis in human brain vascular endothelial and neuroglia cells. *Molecular and Biochemical Parasitology* 162:172-176.
- Wu J, Vallenius T, Ovaska K, Westermarck J, Mäkelä TP, Hautaniemi S. (2009). Integrated network analysis platform for protein-protein interactions *Nature Methods* 6(1):75-77.
- Zhang X, Zhang R, Jiang Y, Sun P, Tang G, Wang X, Lv H & Li X. (2011). The expanded human disease network combining protein-protein interaction information. *European Journal of Human Genetics* 19:783-788.



Protein Interactions

Edited by Dr. Jianfeng Cai

ISBN 978-953-51-0244-1

Hard cover, 464 pages

Publisher InTech

Published online 16, March, 2012

Published in print edition March, 2012

Protein interactions, which include interactions between proteins and other biomolecules, are essential to all aspects of biological processes, such as cell growth, differentiation, and apoptosis. Therefore, investigation and modulation of protein interactions are of significance as it not only reveals the mechanism governing cellular activity, but also leads to potential agents for the treatment of various diseases. The objective of this book is to highlight some of the latest approaches in the study of protein interactions, including modulation of protein interactions, development of analytical techniques, etc. Collectively they demonstrate the importance and the possibility for the further investigation and modulation of protein interactions as technology is evolving.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Aditya Rao, Gopalakrishnan Bulusu, Rajgopal Srinivasan and Thomas Joseph (2012). Protein-Protein Interactions and Disease, Protein Interactions, Dr. Jianfeng Cai (Ed.), ISBN: 978-953-51-0244-1, InTech, Available from: <http://www.intechopen.com/books/protein-interactions/protein-protein-interactions-and-disease>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.