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

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

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

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

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

The Prediction and Analysis of Inter- and Intra-Species Protein-Protein Interaction

Theresa Tsun-Hui Tsao[1], Chen-Hsiung Chan[2],
Chi-Ying F. Huang[3] and Sheng-An Lee[4]
[1]Department of Biomedical Electronics and Bioinformatics,
National Taiwan University, Taipei
[2]Department of Medical Informatics, Tzu Chi University, Hualien
[3]Institute of Biomedical Informatics, National Yang Ming University, Taipei
[4]Department of Information Management, Kainan University, Taoyuan, Taiwan
Taiwan

1. Introduction

Protein-protein interactions (PPIs) are essential to cellular processes. Recent developments of high-throughput technologies have uncovered vast numbers of PPIs. However, the experimental evidences are mostly for intra-species interactions of model organisms, especially human. Studies of non-human organisms and inter-species PPIs are few. For organisms such as *Arabidopsis thaliana*, the experimentally detected 5990 PPIs are estimated to be less than 3% of the entire *A. thaliana* interactome (M. Lin et al., 2011). The accuracy of high-throughput PPI experiments is also doubtful (Mrowka et al., 2001; Sprinzak et al., 2003; von Mering et al., 2002). To resolve the above issues, several computational methods have been developed to evaluate and predict PPIs. This chapter focuses on direct PPIs which involve physical interactions of proteins, provides a brief overview of the reliabilities of high-throughput PPI detection technologies, and discusses the weakness and strength of important PPI computational prediction and evaluation methods. The major repositories which store, evaluate, and analyse both detected and predicted PPIs are also introduced.

2. Experimental detection of protein interactions

Not until the past decade, PPIs were identified by time consuming and labour intensive methods, such as low-throughput (small-scale) yeast 2-hybrid (Y2H). The development of high-throughput technologies brought studies of PPIs to an -omics level. Of all the technologies used for PPI detection, the high-throughput Y2H is most mature and commonly used. However, it is also one of the most inaccurate techniques, producing an estimated ~ 50% of false positives (Parrish et al., 2006; von Mering et al., 2002). The error rates of the other high- and medium-throughput technologies are summarized in Table 1 (Mrowka et al., 2001; Parrish et al., 2006; Sprinzak et al., 2003; von Mering et al., 2002). In the past few years, BiFC has became one of the popular *in vivo* technologies as it has a medium throughput and reasonable cost, is technically straightforward, and provides

information on subcellular localizations of proteins. The drawback for BiFC is its occasional false postivies caused by non-specific interactions and background flurorescence. Split-luciferase system has an extremely low background, but does not disclose the subcellular localization of interactions. Protoplast Y2H is similar to Y2H, but technically more challenging as it has to be operated in a nuclei or protoplasts. SUS has high rates of false positives and background signals. The *in vitro* technologies are less favourable as the reactions do not occur in cellular environments and do not examine the cellular localization of proteins.

Technology	Throughput	Accuracy	References
In vivo			
High-throughput Y2H	High	×	(Causier, 2004; Causier & Davies, 2002)
Split-luciferase system	Medium-high	○	(Paulmurugan & Gambhir, 2005,2007)
Protoplast Y2H	Medium	×	(Fujikawa & Kato, 2007)
Split-ubiquitin system (SUS)	Medium	×	(Michnick, 2003; Obrdlik et al., 2004; Reinders et al., 2002; Schulze et al., 2003)
Bimolecular fluorescence complementation (BiFC)	Medium	△	(Citovsky et al., 2008; Hu et al., 2006; Ohad et al., 2007; Ohad & Yalovsky, 2010; Zhou et al., 2011)
In vitro			
Stable-isotope labeling of amino acids in cell culture (SILAC)	High	×	(Gruhler & Kratchmarova, 2008; Mann, 2006)
^{15}N-labeling	High	○	(Huttlin et al., 2007; Nelson et al., 2007)
Chemical crosslinking-MS using protein interaction reporter (PIRs)	High	×	(Anderson et al., 2007; Tang et al., 2005)
Protein microarrays	High	○	(Angenendt et al., 2006; Popescu et al., 2007; Ramachandran et al., 2004)
Single affinity purification-tagging	Medium-high	△	(Berggard et al., 2007)
Tandem affinity purification (TAP) tagging	Medium-high	△	(Rohila et al., 2006; Schoonheim et al., 2007)
Native chromatography or electrophoretic purification	Medium	×	(Liu et al., 2008)

Table 1. The technologies for high- and medium-throughput PPI detection (Morsy et al., 2008) Accuracy of each technology is summarised in this table. The symbol "×" indicates low accuracy, "○" indicates high accuracy, and "△" indicates sound accuracy.

3. Computational prediction and evaluation of protein interactions

As listed in Table 2, the methods for PPI prediction and evaluation can be classified into five categories based on the types of information required for analysis – (1) protein sequences, (2) Gene Ontology (GO), (3) gene expression profiles, (4) topology of the interaction network, and (5) experimental data.

Methods	References
Protein Sequences	
Interologs	(Matthews et al., 2001; Rhodes et al., 2005; von Mering et al., 2007)
Phylogenetic tree similarity	(Jothi et al., 2005)
Gene fusion	(Enright et al., 1999; Marcotte et al., 1999)
Gene neighbouring	(Dandekar et al., 1998; Overbeek et al., 1999; Tamames et al., 1997)
Domain-domain interactions	(Frishman, 2009; Ng et al., 2003)
Gene Expression Profile	
Co-expression correlation coefficience	(Ideker et al., 2002)
Shared Gene Ontology Annotation	
Protein functions	(De Bodt et al., 2009; Jain & Bader, 2010; Wu et al., 2006)
Protein localization	(De Bodt et al., 2009; Jain & Bader, 2010)
Topology Analysis	
Distance between proteins in a PPI network	(Dyer et al., 2007)
Experimental Data	
Cited literatures (text mining)	(Jaeger et al., 2008)
Detected PPI datasets	(von Mering et al., 2002)

Table 2. Computational methods for protein interaction prediction and evaluation

Methods for PPI prediction and evaluation is each developed based on an assumption which states certain criteria are more likely to occur between interacting proteins. These methods are often combined, usually in a Bayesian network (Huttenhower & Troyanskaya, 2006; Jansen et al., 2003; Lee et al., 2006; N. Lin et al., 2004; McDowall et al., 2009; Patil & Nakamura, 2005; Wang et al., 2009; Xu et al., 2011). Some criteria are more relevant to protein interactions than the others. When different methods are combined, the statistical confidence estimated by each method could be weighted according to the confidence level of corresponding assumption.

3.1 Interologs
Homologous proteins often conserve similar functions and PPIs across different organisms, especially in phylogenetically close-related species (Hirsh & Sharan, 2007). These conserved PPIs are designated as interologs.

In the interolog method, it is assumed that if a pair of proteins, A and B, interact and there are two other proteins, A' and B', of which A' is homologous to A and B' is homologous to B, then A' and B' are potentially an interolog to A and B. Interologs can occur among different species or in the same species (Mika & Rost, 2006). The conventional interolog method identifies homologous proteins by comparing the global sequences. For proteins of which only partial sequences are similar, sequence signatures may be compared instead of full sequences (Sprinzak & Margalit, 2001). Structurally similar proteins may also have similar protein interactions, but predicting PPIs by identifying proteins with similar structures is impeded by the limited structural information available (Aytuna et al., 2005; Ogmen et al., 2005). The interologous relationship between the two pairs of proteins, one pair predicted and one pair detected, could be evaluated by functions such as the *s* score. (He et al., 2008). Homologous genes of model organisms can be identified using BLAST or in HomoloGene database that automatically identifies and collects homologs from fully sequenced genomes (Sayers et al., 2011).

The interolog method has been used frequently. The first human PPI network, the Arabidopsis PPI network, and the rice blast fungus PPI network are a few examples constructed by predicted interologs (Geisler-Lee et al., 2007; He et al., 2008; Lehner & Fraser, 2004). Unfortunately, prediction of plant PPIs through a comparative interactome approach is challenged by the unique biology of plants which involves PPIs not commonly found in the other model organisms. Less than 50% of *A. thaliana* proteins have been found to have orthologs in the more extensively studied organisms such as yeast, *Caenorhabditis elegans*, fruit fly, or human (Gollery et al., 2006). Furthermore, the interolog method does not differentiate the functionally significant amino acid residues from the others; neglects the residue-specific requirements for interaction specificity and affinity (Uhrig & Hulskamp, 2006). For the highly homologous members of protein families, the interolog method could be prone to errors.

3.2 Phylogenetic relationship

Interacting proteins have been observed to have topologically similar phylogenetic trees for the corresponding protein families, presumably due to the co-evolution of cooperating proteins (Fryxell, 1996; Goh et al., 2000; Pages et al., 1997). Based on the above observation, the phylogenetic similarity method was proposed. To compare and construct the phylogenetic trees, firstly, the sequences of two potentially interacting protein families are aligned. Secondly, the evolutionary distance matrixes are calculated from the phylogenetic trees, one for each protein family. Finally, the Pearson's correlation coefficient between the two distance matrixes is calculated as an indication of the likelihood of interactions. Partial protein sequences could be used to construct the phylogenetic trees - for example, poorly conserved sequences have been removed to improve the performance of prediction (Kann et al., 2007).

A similar approach is the phylogenetic profile method. Phylogenetic profile is the profile which records the presence and absence of a protein across all species. Also due to the presumably co-evolution of proteins involved in the same biological process, proteins with similar phylogenetic profiles are more likely to have interactions. The profiles could be compared by Hamming distance (Pellegrini et al., 1999). Although this approach is powerful, it can be applied only to organisms which have been fully sequenced (Frishman, 2009). Additionally, there might be complications with essential proteins which are present

in all organisms (Frishman, 2009). As the second generation sequencing (SGS) technologies exponentially accumulating full genome sequences of non-model organisms, this method is expected to become more favorable.

3.3 Gene fusion and neighboring

Genes which are previously separated in the genome of one organism can be fused into the same gene in another organism. Fused genes almost always encode functional related and physically interacting proteins (Enright et al., 1999; Marcotte et al., 1999). The fusion events might accelerate the formation of protein complexes by increasing the opportunity of correct physical contact between interacting sites.

Similarly, in bacteria, genes which are consistently located in the same operon across many species are likely to express functionally related, and often physically interacting, proteins (Dandekar et al., 1998; Overbeek et al., 1999; Tamames et al., 1997).

3.4 Domain-domain interactions

Just like protein interactions, domain interactions can be predicted by sequence homology among two pairs of interacting domains, by investigating the evolutionary traits of domains, or by identifying conserved neighboring relationship between domains (Frishman, 2009). Interacting proteins are also more likely to contain domains which have been detected or predicted to interact (Ng et al., 2003).

3.5 Co-expression

Interacting proteins are assumed to have similar expression patterns (Dyer et al., 2007). The co-expression correlation coefficients of seven model animals, including human, mouse, chicken, zebra fish, fruit fly, and *Coenorhabditis elegans*, and nematode, are recorded in COXPRESdb (Obayashi et al., 2008; Obayashi & Kinoshita, 2011). The co-expression correlation coefficients of *A. thaliana* and many other flowering plants are recorded in ATTED-II (Obayashi et al., 2011). High-throughput expression data are mostly available on Gene Expression Omnibus (GEO) or TAIR for *A. thaliana* experiments (Garcia-Hernandez et al., 2002; Sayers et al., 2011).

3.6 Gene Ontology (GO)

Interacting proteins are presumably to participate in related biological process and share similar cellular localization (Dyer et al., 2007; Shin et al., 2009). The GO project annotates the cellular components where a protein locates and the biological process in which a protein participates. The annotations are created by structured and controlled vocabularies. The semantic similarities between GO terms assigned to proteins are often used to evaluate the confidence levels of proposed PPIs (De Bodt et al., 2009; Jain & Bader, 2010).

3.7 Topology

As more and more PPIs are revealed, PPI networks can be constructed and analyzed by topology theories. It has been proposed that two proteins which interact with the same protein should have a shorter path between them on the PPI network (Dyer et al., 2007). It has also been proposed that interacting proteins might share more neighboring proteins on a PPI network (J. Chen et al., 2006; Chua et al., 2006).

3.8 Text mining

The protein interactions which have been reported repeatedly in more peer-reviewed literatures might be more trustworthy than the ones which have never or rarely been detected (Jaeger et al., 2008). However, it must be noted that proteins with more valuable functions, such as disease mechanisms, would have been studied more intensively and been documented more frequently.

PubMed and GeneRIF are common sources of text mining materials. The automated data gathering (e.g. text mining *via* natural language processing or biomedical language processing) is not as reliable as manually curated data. It must be noted that manual curation is neither 100% correct due to human errors and inconsistent standards for curation.

3.9 Experimental detections

PPIs detected by low-throughput technologies are generally considered as error free. For the medium- to high-throughput technologies, the reliability of the results varied as listed in Table 1. *In vivo* experiments are usually more accurate than the *in vitro* experiments, as *in vivo* experiments were conducted in cellular environments. Interactions supported by more than one method are generally believed to be more reliable (von Mering et al., 2002). PPI datasets which are more reliable are assumed to have more intersections with the other datasets and higher averaged numbers of documented protein interactions (Shin et al., 2009). Reliable PPI datasets should also contain greater proportion of interactions which have interacting domain pairs (He et al., 2008).

In silica protein docking is another approach which could be used for predicting protein interactions; however, it is impractical for high-throughput predictions due to the extremely large amount of required computation and the lack of detected or predicted structures for most proteins.

4. Protein interaction databases

More than 30 PPI databases have been published and are mostly available online (Fischer et al., 2005). Table 3 listed the frequently referenced databases. The contents of these databases are often overlapped and integrated to create larger non-redundant databases. These collections of PPIs can be used as the foundation for predicting and evaluating the reliability of PPIs.

Database	Validation	Organisms	Reference
MINT	Detected Provides confidence scores for PPIs.	Model organism	(Ceol et al., 2010)
DIP	Detected	Model organisms	(Salwinski et al., 2004)
BIND	Detected	~ 1500 organisms	(Gilbert, 2005; Isserlin et al., 2011; Willis & Hogue, 2006)
BioGRID	Detected	Model organisms	(Breitkreutz et al., 2008; Stark et al., 2011; Stark et al., 2006)

Database	Validation	Organisms	Reference
IntAct	Detected	Model organisms	(Aranda et al., 2010; Brandao et al., 2009)
HiPredict	Data from IntAct, BioGRID, and HPRD Provide confidence scores for PPIs.	Model organisms	(Patil et al., 2011)
MIPS	Detected	Model mammals	(Pagel et al., 2005)
HPRD	Detected	Human	(Goel et al., 2010)
STRING	Detected and predicted data from BIND, BioCarta, BioCyc, BioGRID,DIP, HPRD, IntAct, MINT, REACTOME, textmining, etc Provides confidence scores for PPIs.	~ 1000 organisms	(Szklarczyk et al., 2011; von Mering et al., 2007; von Mering et al., 2005)
HAPPI	Detected and predicted data from HPRD, BIND, MINT, STRING, OPHID, etc Provide confidence scores for PPIs.	Human	(J.Y. Chen et al., 2009)
AtPID	Predicted Provide confidence scores for PPIs.	*A. thaliana*	(Cui et al., 2008; Li et al., 2011)
PAIR	Detected (from IntAct, BioGRID and BIND) and predicted Provide confidence scores for PPIs.	*A. thaliana*	(M. Lin et al., 2011)
AtPIN	Experiemtnal (from BioGRID and IntAct) and predicted (from Geisler-Lee and AtPID) Provide confidence scores for PPIs.	*A. thaliana*	(Cui et al., 2008; Geisler-Lee et al., 2007; Li et al., 2011)
PIG	Data from BIND, IntAct, REACTOME, and MINT	Human-pathogen interacitons	(Driscoll et al., 2009)
HPIDB	Data from BIND, IntAct, REACTOME, MINT, GENERIF and PIG	Host-pathogen interactions, hosts are model organisms	(Kumar & Nanduri, 2010)

Table 3. Major PPI databases

MINT is one of the few repositories which provide confidence scores for experimentally detected PPIs. It uses the number and types of experiments in which a PPI is detected to estimate the confidence of data.

HiPredict is a repository which contains filtered high-confidence PPIs of nine model species from IntAct, BioGRID, and HPRD. While calculating the confidence of PPIs, HiPredict considers (1) the type of experiments which detect the PPIs, (2) the co-expression correlation coefficients of proteins, (3) shared GO terms of proteins, (4) presences of interologs in the same organisms, and (5) domain-domain interactions between proteins. These five criteria are combined in naïve Bayesian networks to give confidence scores.

STRING is one of the largest and most comprehensive PPI repositories. It evaluates PPIs using multiple criteria, including (1) the probability of finding the interacting proteins on the same KEGG pathway, (2) co-mentioning of gene/protein names in PubMed abstracts, (3) co-expression / co-regulation of proteins, (4) presence of interologs, and (5) presence of gene neighboring. Similar to HiPredict, the various criteria are also combined in naïve Bayesian networks.

HAPPI only collects human PPIs. For the PPIs which have been evaluated, such as data from STRING, the confidence scores are preserved. For the PPIs which have not evaluated, HAPPI calculates the confidence scores based on the type of experimental evidences and the source of data.

PAIR collects 5990 detected protein interactions and 145494 predicted interaction of *A. thaliana* (M. Lin et al., 2011). These PPIs were expected to cover 24% of the entire *A. thaliana* interactome, of which the size was estimated to be 200 000 PPIs (for 20 000 genes) based the size for yeast (18 000 PPIs for 6000 genes). An estimated 44% of the collected PPIs in PAIR are reliable. PAIR predicts PPIs using a machine learning approach with supports the vector machine (SVM) model. In the SVM model, indirect evidences of interactions (i.e. interologs, phylogenetic profile similarity, domain interactions, gene co-expression correlation, shared GO terms, and protein localizations) are combined. The model has been trained using Gold Standard Positives (GSPs), which are reliable PPIs from major repositories. The SVM scores also serve as the confidence scores for the predicted PPIs. The detected PPIs are collected from IntAct, BioGRID, and BIND.

AtPIN integrates (1) the predicted PPIs from Geisler-Lee and AtPID, (2) the curated PPIs from TAIR, and (3) the detected PPIs from BioGRID and IntAct (Brandao et al., 2009). Geisler-Lee (2007) predicts PPIs by identifying interologs. AtPIN calculates confidence scores of PPIs based on (1) the detected or predicted co-localization of interacting proteins and (2) the number of shared neighboring proteins of interacting proteins on the PPI network. It also provides the score calculated by AtPID. **AtPID** combines indirect evidences of interactions, including interologs, phylogenetic profiles, domain interactions, co-expression profile, shared protein functions, protein co-localisation, and gene fusion, in naïve Bayesian networks to predict and evaluate the PPIs of *A. thaliana* (Cui et al., 2008; Li et al., 2011). TAIR is a multi-tasking project which participates in a broad range of *A. thaliana* researches.

The data of protein interactions between hosts and pathogens are scarce. **PIG** integrates the manually curated human-pathogen PPIs from four databases, BIND, IntAct, REACTOME, and MINT, in one platform for searching, visualization, and analysis of PPI networks. The corresponding hyperlinks to UniProt database, Gene Ontology, InterProScan, and PubMed are filed under each protein entry in the user interface for convenient referencing.

Similar to PIG, **HPIDB** integrates several host-pathogen PPI databases, including BIND, IntAct, REACTOME, MINT, GENERIF, and PIG. However, unlike PIG, the PPIs in HPIDB are not limited to human host. Although the majority of data is for human (22386 PPIs), HPIDB also contains host-pathogen PPIs for mouse (147 PPIs), *A. thaliana* (99 PPIs), rat (53 PPIs), cattle (30 PPIs), and chicken (19 PPIs).

A few repositories collect genes which are involved in host-pathogen interactions, but do not contain data on physical protein interactions. **PHIDIAS** is a centralized respiratory for host-pathogen interactions. It collects information for 98 pathogens of two hosts, human, and mouse (Xiang et al., 2007). **PHI-Base** contains information for 405 fungal, oomycete, and bacterial genes which participate in pathogenicity, virulence, and induction of disease resistance (Baldwin et al., 2006; Winnenburg et al., 2006). 176 of these genes are from animal pathogens, 227 from plant pathogens, and 3 from pathogens of fungi. **PathoPlant** contains *A. thaliana* genes which are responsible in the defense against pathogens (Bulow et al., 2007).

5. Identification of drug targets within human-pathogen interactions network

The evolutionary history of human has never been parted with pathogens. Viruses, bacteria, fungi, and nematodes all play critical roles in shaping the human race. Recent advances in metagenomics and human microbiomes suggest that commensal microorganisms have significant influences to the metabolism, immune systems, general wellbeing, and even behaviour patterns of animal hosts.

Despite enormous efforts in preventing, diagnosing, and treating infectious diseases, pathogens still cause insurmountable burden and social-economical impacts to human. The developments of vaccines and drugs have helped to diminish several devastating diseases; however, emerging diseases caused by novel or previously unknown pathogens continuously lead to unexpected outbreaks. To account for current and future threats imposed by pathogens, it is necessary to understand human-pathogen interactions at the molecular level. Viruses require host factors for recognition, entrance, replication, and release. Their gene products form dense interaction networks with host proteins. Most bacteria, fungi, and nematodes, on the other hand, proliferate outside of human cells and interact with host cells with extracellular signals and receptors. The following sections of this chapter review previous works on high-throughput characterization of human-pathogen interactions interactions, mostly between human and viruses. Most works have focused on human-virus interactions.

5.1 Human-virus interactions

High-throughput characterization of intra-species interactions has been the focus of early day PPI studies. Inter-species interactions still constitute a minor part of most interactome databases. Beginning from 2007, several works on high-throughput human-virus interactions and host factor characterization have been published, including the ones for Epstein-Barr virus (EBV), hepatitis C virus (HCV), and influenza virus. Among these sparse inter-species interactions, those between human immunodeficiency virus 1 (HIV-1) and human are most abundant due to the research efforts devoted to this notorious virus. These datasets are summarized in Table 4.

Among these datasets, the HIV-1 human protein interaction database is so far the most comprehensive in terms of recorded interaction number and annotations. The number of human-virus PPIs can be estimated based on the number of human-HIV PPIs, which presumably have not been fully exposed, and the number of human viruses. A severe under-estimated number of human viruses is 200 ~ 1 000 species, which can be deduced to at least 1 ~ 5 million human-virus PPIs yet to be discovered. Despite the small number of human-virus PPIs being detected or predicted, this data is a start point to the research of the viral disease mechanisms and treatments.

Datasets	No. of Interactions	No. of Viral Strains	No. of Viral Proteins	No. of Human Proteins	Sources
Human Immunodeficiency Virus 1 (HIV-1)	5128	1	21	1433	(Pinney et al., 2009)
Epstein-Barr Virus	173	1	42	112	(Calderwood et al., 2007)
Hepatitis C Virus	481	1	11	414	(de Chassey et al., 2008)
Influenza Virus	400	1	10	246	(Konig et al., 2010)
NCBI Interactions	5370	39	86	1530	NCBI FTP Site[1]
IntAct	689	50	124	308	(Aranda et al., 2010)

Table 4. Summary of human-virus interaction datasets

5.2 HIV-1 interactions

The HIV-1, Human Protein Interaction Database (Pinney et al., 2009) was compiled by National Institute of Allergy and Infectious Diseases (NIAID), and hosted by NCBI. Interaction data in this database was collected from published literatures. Unlike other interaction data, entries in this dataset were associated with detailed annotations, including PubMed ID list for references, short phrases describing the interactions, and texts excerpted from the source literature. Interactions in this database are not just revealed by conventional Y2H or immune-co-precipitation, but 70 interactions were annotated with details. For example, the statement "HIV retropepsin cleaves human actin" is supported by four publications and attached with descriptions of the HIV retropepsin and human actin. Occasionally, the texts from the source literatures would provide additional information. In the case of "HIV retropepsin cleaves human alpha-2-macroglobulin precursor", the GeneRIF text states "the cleavage site of alpha 2-Macroglobulin by HIV-1 protease is the Phe684-Tyr685 bond", which depicts the interaction (cleavage) site. Interaction types include cleavage, binding, regulation/modulation, and post-translational modifications.

Analysis of this database found that there were 21 HIV gene products interacting with 1433 human proteins. The top 10 HIV and human proteins which participate in most HIV-human interactions are listed in Table 5.

By simply counting the numbers of PPIs in Table 5, critical host factors in HIV infections could be identified. The C-C chemokine receptor type 5 (CCR5) variants have been implicated in HIV-resistance and immunity (Blanpain et al., 2002). Stem cell-based gene therapy has successfully "cured" HIV with this genetic variant in early phase clinical trials (Symonds et al., 2010). Some host factors were also involved in various types of processes and diseases, such as tumour necrosis factor (TNF), which regulates cell proliferation, apoptosis, and has been implicated in cancer.

---

[1] ftp://ftp.ncbi.nlm.nih.gov/gene/GeneRIF/interactions.gz

Top 10 HIV Proteins	No. of Interactions	Top 10 Human Proteins	No. of Interactions
Tat	1685	T-cell surface glycoprotein CD4 isoform 1 precursor	177
gp120	1252	C-X-C chemokine receptor type 4 isoform b	54
Nef	496	Tumor necrosis factor	47
Vpr	342	Nuclear factor NF-kappa-B p105 subunit isoform 1	41
gp41	253	Nuclear factor NF-kappa-B p105 subunit isoform 1	41
gp160	232	C-C chemokine receptor type 5	40
Rev	139	Interferon gamma precursor	35
Matrix	132	Mitogen-activated protein kinase 1	30
Integrase	122	Major histocompatibility complex, class I, B precursor	30
Retropepsin	98	Cyclin-dependent protein kinase 9	30

Table 5. HIV and human proteins participate in largest numbers of human-HIV interactions.

Table 5 also suggests that Tat could be a potential drug target. The crystal structure of Tat which forms complex with cyclin-denpendent protein kinase 9 (CDK9) and cyclin T1 has been solved (Tahirov et al., 2010). The complex structure (PDB ID: 3MI9) reveals that the most part of Tat has physical contact with cyclin T1 and has only a small loop contacting CDK9. The structural information provides valuable insights to the design of Tat inhibitors.

5.3 Epstein-Barr virus interactions

Epstein-Barr virus (EBV) infects human epithelial cells, and is implicated in various types of cancer, such as Burkitt's lymphoma and nasopharyngeal carcinoma. The interactions within EBV proteins, and between EBV and human proteins, have been characterized using Y2H method (Calderwood et al., 2007). Overall, 43 EBV-EBV and 173 human-EBV interactions have been validated with experimental evidences.

Network analysis reveals that most EBV-EBV interactions take place among conserved "core" proteins, thus these interactions may be responsible for the general infection/replication of herpesviruses. On the other hand, most human proteins targeted by EBV are proposed as "hub" proteins, which participate in more human-human interactions and may have crucial roles in the underlying biological processes. The EBV protein targeting most human proteins is BFLF2, with 21 interaction partners. BFLF2 interacts with BFRF1 and changes cellular localization (Gonnella et al., 2005). Deletion of BFLF2 also impairs viral DNA packaging (Granato et al., 2008). The most targeted human proteins are HOMER3 and GRN. HOMER3, which binds to numerous receptors, is involved in diverse biological functions such as neuronal signalling and T-cell activation. Granulin (GRN) is a secreted glycosylated peptide which regulates cell growth and implicates in wound healing and tumorigenesis. BFLF2 interacts with both HOMER3 and GRN; however, the functional implications of interactions were not clear.

5.4 Hepatitis C virus interactions

Hepatitis C virus (HCV) is the pathogen which causes the chronic hepatitis infection. Infection with HCV may lead to cirrhosis and hepatocarcinoma if not properly treated with

antiviral drugs or interferon. Unfortunately, current HCV treatments are expensive and can have severe adverse effects. The human-HCV interaction map would allow us to understand the mechanisms of HCV infection and its chronic nature.

The human-HCV interaction network is constituted by 481 HCV-human interactions (de Chassey et al., 2008). Among these interactions, 314 were determined with Y2H experiments, and others were identified from literature reviews. The most connected HCV proteins include NS3, NS5A, and CORE. Human proteins targeted by most HCV proteins include nuclear receptor subfamily 4, group A, member 1 (NR4A1), homeobox D8 (HOXD8), and SET domain containing 2 (SETD2). NR4A1 is a nuclear transcription factor, which is highly expressed in adrenal cortex, lung, and prostate; however its expression level in liver is low. HOXD8 is important to development; its deletion leads to limb deformation. Expression level of HOXD8 is highest in kidney. SETD2 is a histone methyltransferase and also contains transcription activation domain. Recently, SETD2 has been found as a tumour suppressor gene (Duns et al., 2010). However, the roles of HCV-SETD2 interactions in tumorigenesis remain elusive.

The analyses of EBV-human and HCV-human interaction networks found that viral proteins tend to interact with "hubs" in human protein-protein interaction networks. In human proteins targeted by HCV, three KEGG pathways were significantly enriched, including insulin signalling pathway, TGFβ signalling pathway, and Jak-STAT signalling pathway (de Chassey et al., 2008). Also, "focal adhesion" pathway has been identified as a novel pathway targeted by HCV. In our own analysis using bootstrap to estimate the statistical significance of HCV targeted gene numbers, we have also identified that "focal adhesion" and "ECM-receptor interaction" pathways may be perturbed by HCV infection (Table 6).

Path ID	Title	No. of genes in pathway	# of HCV targets	Random (mean)	Random (SD)	Z-stat	p-value
5160	**Hepatitis C**	134	23	1.82	1.35	15.70	$< 2.2 \times 10^{-16}$
5200	**Pathways in cancer**	328	31	4.60	2.20	12.02	$< 2.2 \times 10^{-16}$
5212	**Pancreatic cancer**	71	12	0.99	0.95	11.54	$< 2.2 \times 10^{-16}$
4510	**Focal adhesion**	202	22	2.80	1.68	11.42	$< 2.2 \times 10^{-16}$
5222	**Small cell lung cancer**	84	13	1.19	1.08	10.98	$< 2.2 \times 10^{-16}$
4520	**Adherents junction**	75	12	1.07	1.03	10.59	$< 2.2 \times 10^{-16}$
4722	**Neurotrophin signaling pathway**	127	15	1.78	1.32	10.04	$< 2.2 \times 10^{-16}$
5215	**Prostate cancer**	89	12	1.32	1.12	9.56	$< 2.2 \times 10^{-16}$
4512	**ECM-receptor interaction**	84	11	1.16	1.05	9.35	$< 2.2 \times 10^{-16}$

Table 6. Top 10 KEGG pathways potentially perturbed by HCV infection.

5.5 Influenza virus host factors

Influenza A virus causes epidemics every now and then. The high transmission rate of influenza virus makes it one of the greatest threats to public health, especially when long diminished strains or emerging strains turned to the surface. The rapidly evolving virus makes it difficult to predict and prepare seasonal vaccines. Drug-resistant strains also challenge our ability to treat and control the disease.

The identification of host factors required by influenza virus may contribute to the prevention and treatment of the virus. Host factors involved in early stage influenza virus replication have been characterized with genome-wide RNA interference (RNAi) screening (Konig et al., 2010). Unlike Y2H experiments, host factors identified with RNAi do not necessary interact with viral proteins directly. Nevertheless, these findings imply that viral diseases may be treated by regulating some of these host factors. One example is the inhibitor for the host factors, CAMK2B, which impedes viral growth and may be developed to new antiviral drugs.

5.6 Human-bacteria interactions

Bacteria cells can reproduce without the cellular machinery of hosts. Studies on human-bacteria interactions thus have been focused on cellular-level interactions. So far, only limited efforts have been devoted to the identification of human-bacteria interactions at the molecular-level. The interactions between three pathogenic bacteria, *Bacillus anthracis*, *Francisella tularensis,* and *Yersinia pestis*, have been characterized using high-throughput Y2H experiments (Dyer et al., 2010). The reported dataset includes 3,073, 1,383, and 4,059 interactions between human or *B. anthracis*, *F. tularensis* or *Y. pestis*, respectively.

The topology of these human-bacteria and human protein-protein interaction networks revealed that many bacteria proteins target "hubs" in human PPI networks. Specifically, several host defence pathways have been identified, including innate immunity and inflammation. Comparative analysis of the three human-pathogen interaction networks also confirmed these findings. Several methods have been used to identify the conserved protein interaction modules (CPIM), and found that these bacteria may have interfered host innate immune responses, including antigen binding and processing, and several immune response pathways.

Analysis of these interactions faces some obstacles. Large proportions of the bacteria proteins (285/943, 66/349, 630/1,218 protiens for *B. anthracis*, *F. tularensis*, and *Y. pestis*, respectively) which interact with human proteins are putative, hypothetical, or uncharacterized. Without sufficient functional annotations, interpretations of these interactions can be superficial. Furthermore, bacterial proteins and human proteins were confined within the membrane of respective cells, and only certain types of proteins can be exported or internalized by host cells. Thus, the annotations or predictions of protein subcellular localization are important tasks for the interpretation and refinements of human-bacteria interaction networks.

5.7 Systematic analysis of host-pathogen interactions

Human-pathogen interactions have been collected and analyzed for their network properties (Dyer et al., 2008). Pooling together human-pathogen interactions should enable identification of common targets and biological processes perturbed by pathogens. A total of

10477 interactions between human and the 190 pathogen strains have been collected from several public databases. Networks for human-bacteria and human-virus interactions have been constructed separately. Most of the interactions were human-virus interactions, notably human-HIV interactions.

Special attention has been paid to human proteins which interact with multiple pathogen groups. Such proteins are believed to be the common targets of these pathogens, and may be the highlights of critical events during pathogen invasion.

The analysis of human proteins targeted by multiple viral pathogens have revealed that viruses perturb host cells mainly through controlling cell cycle, regulating apoptosis, and transporting viral particles across membrane (Dyer et al., 2008). Human-bacteria interactions, on the other hand, perturb Gene Ontology processes like "immune system process" and "immune response". It is notable that much of these perturbed pathways were linked to inflammatory and cancer, suggesting multiple roles of pathogens in various diseases.

Another analysis on human-viral interaction network also highlighted the mechanisms of non-infectious diseases (Navratil et al., 2011). Totally 2,099 manually curated interactions of 416 viral proteins from 110 species have been collected. This human-virus interaction network has been integrated with human PPI network. Disease gene annotations from OMIM have been evaluated for their associations with viral proteins. Links between virus and auto-immune diseases have been found, including type 1 diabetes. A comparison between human-virus interaction network and human type I interferon network also revealed that viruses attack host at multiple levels, from receptors to transcription factors (Navratil et al., 2010).

We have also performed similar analysis with human-virus interactions collected from NCBI interactions[2], IntAct (Aranda et al., 2009), and other sources. The association of KEGG (Kanehisa et al., 2010) disease pathways and human-virus interactions have been analyzed. Several KEGG pathways have been identified with high significance, including "systemic lupus erythematosus", "pathways in cancer", "chemokine signalling pathway", "focal adhesion", and "T cell receptor signalling pathway". These findings are *in par* with studies described in previous sections; all pointing to pathogens gain their foothold in host cells through modulating host defence mechanisms. In the meantime, inflammation, autoimmune diseases and cancers may arise as results of these modulations.

6. Conclusion

At the present, the number of confident PPI data is scarce, especially for non-human organisms and inter-species interactions. The prediction of PPIs, as well as the evaluation of accuracy of detected and predicted PPIs, are important topics which require further advances in methodology, tools and data generation. It is believed that, in recent years, as the second generation sequencing (SGS) rapidly discloses full genome sequences and exponentially accumulates high-throughput expression data, more and more inter- and intra-species networks PPI will be constructed for, not only model organisms, but also crops, biofuel producing algae and bacteria, between host-pathogens, and between symbiotic organisms.

---

[2] ftp://ftp.ncbi.nlm.nih.gov/gene/GeneRIF/interactions.gz

7. References

Anderson, G. A.; Tolic, N.; Tang, X.; Zheng, C. & Bruce, J. E. (2007). Informatics strategies for large-scale novel cross-linking analysis. *J Proteome Res*, Vol. 6, No. 9, pp. 3412-21

Angenendt, P.; Kreutzberger, J.; Glokler, J. & Hoheisel, J. D. (2006). Generation of high density protein microarrays by cell-free in situ expression of unpurified pcr products. *Mol Cell Proteomics*, Vol. 5, No. 9, pp. 1658-66

Aranda, B.Achuthan, P.Alam-Faruque, Y.Armean, I.Bridge, A.Derow, C.Feuermann, M.Ghanbarian, A. T.Kerrien, S.Khadake, J., et al. (2009). The intact molecular interaction database in 2010. *Nucleic Acids Res*, Vol. 38, No. Database issue, pp. D525-31

Aranda, B.Achuthan, P.Alam-Faruque, Y.Armean, I.Bridge, A.Derow, C.Feuermann, M.Ghanbarian, A. T.Kerrien, S.Khadake, J., et al. (2010). The intact molecular interaction database in 2010. *Nucleic Acids Res*, Vol. 38, No. Database issue, pp. D525-31

Aytuna, A. S.; Gursoy, A. & Keskin, O. (2005). Prediction of protein-protein interactions by combining structure and sequence conservation in protein interfaces. *Bioinformatics*, Vol. 21, No. 12, pp. 2850-5

Baldwin, T. K.; Winnenburg, R.; Urban, M.; Rawlings, C.; Koehler, J. & Hammond-Kosack, K. E. (2006). The pathogen-host interactions database (phi-base) provides insights into generic and novel themes of pathogenicity. *Mol Plant Microbe Interact*, Vol. 19, No. 12, pp. 1451-62

Berggard, T.; Linse, S. & James, P. (2007). Methods for the detection and analysis of protein-protein interactions. *Proteomics*, Vol. 7, No. 16, pp. 2833-42

Blanpain, C.; Libert, F.; Vassart, G. & Parmentier, M. (2002). Ccr5 and hiv infection. *Receptors Channels*, Vol. 8, No. 1, pp. 19-31

Brandao, M. M.; Dantas, L. L. & Silva-Filho, M. C. (2009). Atpin: Arabidopsis thaliana protein interaction network. *BMC Bioinformatics*, Vol. 10, p 454

Breitkreutz, B. J.Stark, C.Reguly, T.Boucher, L.Breitkreutz, A.Livstone, M.Oughtred, R.Lackner, D. H.Bahler, J.Wood, V., et al. (2008). The biogrid interaction database: 2008 update. *Nucleic Acids Res*, Vol. 36, No. Database issue, pp. D637-40

Bulow, L.; Schindler, M. & Hehl, R. (2007). Pathoplant: A platform for microarray expression data to analyze co-regulated genes involved in plant defense responses. *Nucleic Acids Res*, Vol. 35, No. Database issue, pp. D841-5

Calderwood, M. A.Venkatesan, K.Xing, L.Chase, M. R.Vazquez, A.Holthaus, A. M.Ewence, A. E.Li, N.Hirozane-Kishikawa, T.Hill, D. E., et al. (2007). Epstein-barr virus and virus human protein interaction maps. *Proc Natl Acad Sci U S A*, Vol. 104, No. 18, pp. 7606-11

Causier, B. & Davies, B. (2002). Analysing protein-protein interactions with the yeast two-hybrid system. *Plant Mol Biol*, Vol. 50, No. 6, pp. 855-70

Causier, B. (2004). Studying the interactome with the yeast two-hybrid system and mass spectrometry. *Mass Spectrom Rev*, Vol. 23, No. 5, pp. 350-67

Ceol, A.; Chatr Aryamontri, A.; Licata, L.; Peluso, D.; Briganti, L.; Perfetto, L.; Castagnoli, L. & Cesareni, G. (2010). Mint, the molecular interaction database: 2009 update. *Nucleic Acids Res*, Vol. 38, No. Database issue, pp. D532-9

Chen, J.; Hsu, W.; Lee, M. L. & Ng, S. K. (2006). Increasing confidence of protein interactomes using network topological metrics. *Bioinformatics*, Vol. 22, No. 16, pp. 1998-2004

Chen, J. Y.; Mamidipalli, S. & Huan, T. (2009). Happi: An online database of comprehensive human annotated and predicted protein interactions. *BMC Genomics*, Vol. 10 Suppl 1, p S16

Chua, H. N.; Sung, W. K. & Wong, L. (2006). Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. *Bioinformatics*, Vol. 22, No. 13, pp. 1623-30

Citovsky, V.; Gafni, Y. & Tzfira, T. (2008). Localizing protein-protein interactions by bimolecular fluorescence complementation in planta. *Methods*, Vol. 45, No. 3, pp. 196-206

Cui, J.; Li, P.; Li, G.; Xu, F.; Zhao, C.; Li, Y.; Yang, Z.; Wang, G.; Yu, Q. & Shi, T. (2008). Atpid: Arabidopsis thaliana protein interactome database--an integrative platform for plant systems biology. *Nucleic Acids Res*, Vol. 36, No. Database issue, pp. D999-1008

Dandekar, T.; Snel, B.; Huynen, M. & Bork, P. (1998). Conservation of gene order: A fingerprint of proteins that physically interact. *Trends Biochem Sci*, Vol. 23, No. 9, pp. 324-8

De Bodt, S.; Proost, S.; Vandepoele, K.; Rouze, P. & Van de Peer, Y. (2009). Predicting protein-protein interactions in arabidopsis thaliana through integration of orthology, gene ontology and co-expression. *BMC Genomics*, Vol. 10, p 288

de Chassey, B.Navratil, V.Tafforeau, L.Hiet, M. S.Aublin-Gex, A.Agaugue, S.Meiffren, G.Pradezynski, F.Faria, B. F.Chantier, T., et al. (2008). Hepatitis c virus infection protein network. *Mol Syst Biol*, Vol. 4, p 230

Driscoll, T.; Dyer, M. D.; Murali, T. M. & Sobral, B. W. (2009). Pig--the pathogen interaction gateway. *Nucleic Acids Res*, Vol. 37, No. Database issue, pp. D647-50

Duns, G.; van den Berg, E.; van Duivenbode, I.; Osinga, J.; Hollema, H.; Hofstra, R. M. & Kok, K. (2010). Histone methyltransferase gene setd2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. *Cancer Res*, Vol. 70, No. 11, pp. 4287-91

Dyer, M. D.; Murali, T. M. & Sobral, B. W. (2007). Computational prediction of host-pathogen protein-protein interactions. *Bioinformatics*, Vol. 23, No. 13, pp. i159-66

Dyer, M. D.; Murali, T. M. & Sobral, B. W. (2008). The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog*, Vol. 4, No. 2, p e32

Dyer, M. D.; Neff, C.; Dufford, M.; Rivera, C. G.; Shattuck, D.; Bassaganya-Riera, J.; Murali, T. M. & Sobral, B. W. (2010). The human-bacterial pathogen protein interaction networks of bacillus anthracis, francisella tularensis, and yersinia pestis. *PLoS One*, Vol. 5, No. 8, p e12089

Enright, A. J.; Iliopoulos, I.; Kyrpides, N. C. & Ouzounis, C. A. (1999). Protein interaction maps for complete genomes based on gene fusion events. *Nature*, Vol. 402, No. 6757, pp. 86-90

Fischer, T. B.; Paczkowski, M.; Zettel, M. F. & Tsai, J. (2005). A guide to protein interaction databases, In: *The proteomics protocols handbook,* ed. J. M. Walker, pp (753-799), Humana Press, 978-1-59259-890-8,

Frishman, D., Albrecht, M., Blankenburg, H., Bork, P., Harrington, E. D., Hermjakob, H., Jensen, L. J., Juan, D. A., Lengauer, T., Pagel, P., Schachter, V. and Valencia, A.

(2009). Protein-protein interactions: Analysis and prediction, In: *Modern genome annotation: The biosapiens network.* , ed. D. a. V. Frishman, A., pp (353-410), Springer, Wien, Austria

Fryxell, K. J. (1996). The coevolution of gene family trees. *Trends Genet*, Vol. 12, No. 9, pp. 364-9

Fujikawa, Y. & Kato, N. (2007). Split luciferase complementation assay to study protein-protein interactions in arabidopsis protoplasts. *Plant J*, Vol. 52, No. 1, pp. 185-95

Garcia-Hernandez, M.Berardini, T. Z.Chen, G.Crist, D.Doyle, A.Huala, E.Knee, E.Lambrecht, M.Miller, N.Mueller, L. A., et al. (2002). Tair: A resource for integrated arabidopsis data. *Funct Integr Genomics*, Vol. 2, No. 6, pp. 239-53

Geisler-Lee, J.; O'Toole, N.; Ammar, R.; Provart, N. J.; Millar, A. H. & Geisler, M. (2007). A predicted interactome for arabidopsis. *Plant Physiol*, Vol. 145, No. 2, pp. 317-29

Gilbert, D. (2005). Biomolecular interaction network database. *Brief Bioinform*, Vol. 6, No. 2, pp. 194-8

Goel, R.; Muthusamy, B.; Pandey, A. & Prasad, T. S. (2010). Human protein reference database and human proteinpedia as discovery resources for molecular biotechnology. *Mol Biotechnol*,

Goh, C. S.; Bogan, A. A.; Joachimiak, M.; Walther, D. & Cohen, F. E. (2000). Co-evolution of proteins with their interaction partners. *J Mol Biol*, Vol. 299, No. 2, pp. 283-93

Gollery, M.; Harper, J.; Cushman, J.; Mittler, T.; Girke, T.; Zhu, J. K.; Bailey-Serres, J. & Mittler, R. (2006). What makes species unique? The contribution of proteins with obscure features. *Genome Biol*, Vol. 7, No. 7, p R57

Gonnella, R.Farina, A.Santarelli, R.Raffa, S.Feederle, R.Bei, R.Granato, M.Modesti, A.Frati, L.Delecluse, H. J., et al. (2005). Characterization and intracellular localization of the epstein-barr virus protein bflf2: Interactions with bfrf1 and with the nuclear lamina. *J Virol*, Vol. 79, No. 6, pp. 3713-27

Granato, M.; Feederle, R.; Farina, A.; Gonnella, R.; Santarelli, R.; Hub, B.; Faggioni, A. & Delecluse, H. J. (2008). Deletion of epstein-barr virus bflf2 leads to impaired viral DNA packaging and primary egress as well as to the production of defective viral particles. *J Virol*, Vol. 82, No. 8, pp. 4042-51

Gruhler, S. & Kratchmarova, I. (2008). Stable isotope labeling by amino acids in cell culture (silac). *Methods Mol Biol*, Vol. 424, pp. 101-11

He, F.; Zhang, Y.; Chen, H.; Zhang, Z. & Peng, Y. L. (2008). The prediction of protein-protein interaction networks in rice blast fungus. *BMC Genomics*, Vol. 9, p 519

Hirsh, E. & Sharan, R. (2007). Identification of conserved protein complexes based on a model of protein network evolution. *Bioinformatics*, Vol. 23, No. 2, pp. e170-6

Hu, C. D.; Grinberg, A. V. & Kerppola, T. K. (2006). Visualization of protein interactions in living cells using bimolecular fluorescence complementation (bifc) analysis. *Curr Protoc Cell Biol*, Vol. Chapter 21, p Unit 21 3

Huttenhower, C. & Troyanskaya, O. G. (2006). Bayesian data integration: A functional perspective. *Comput Syst Bioinformatics Conf*, pp. 341-51

Huttlin, E. L.; Hegeman, A. D.; Harms, A. C. & Sussman, M. R. (2007). Comparison of full versus partial metabolic labeling for quantitative proteomics analysis in arabidopsis thaliana. *Mol Cell Proteomics*, Vol. 6, No. 5, pp. 860-81

Ideker, T.; Ozier, O.; Schwikowski, B. & Siegel, A. F. (2002). Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, Vol. 18 Suppl 1, pp. S233-40

Isserlin, R.; El-Badrawi, R. A. & Bader, G. D. (2011). The biomolecular interaction network database in psi-mi 2.5. *Database (Oxford)*, Vol. 2011, p baq037

Jaeger, S.; Gaudan, S.; Leser, U. & Rebholz-Schuhmann, D. (2008). Integrating protein-protein interactions and text mining for protein function prediction. *BMC Bioinformatics*, Vol. 9 Suppl 8, p S2

Jain, S. & Bader, G. D. (2010). An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology. *BMC Bioinformatics*, Vol. 11, p 562

Jansen, R.; Yu, H.; Greenbaum, D.; Kluger, Y.; Krogan, N. J.; Chung, S.; Emili, A.; Snyder, M.; Greenblatt, J. F. & Gerstein, M. (2003). A bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*, Vol. 302, No. 5644, pp. 449-53

Jothi, R.; Kann, M. G. & Przytycka, T. M. (2005). Predicting protein-protein interaction by searching evolutionary tree automorphism space. *Bioinformatics*, Vol. 21 Suppl 1, pp. i241-50

Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M. & Hirakawa, M. (2010). Kegg for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*, Vol. 38, No. Database issue, pp. D355-60

Kann, M. G.; Jothi, R.; Cherukuri, P. F. & Przytycka, T. M. (2007). Predicting protein domain interactions from coevolution of conserved regions. *Proteins*, Vol. 67, No. 4, pp. 811-20

Konig, R.Stertz, S.Zhou, Y.Inoue, A.Hoffmann, H. H.Bhattacharyya, S.Alamares, J. G.Tscherne, D. M.Ortigoza, M. B.Liang, Y., et al. (2010). Human host factors required for influenza virus replication. *Nature*, Vol. 463, No. 7282, pp. 813-7

Kumar, R. & Nanduri, B. (2010). Hpidb--a unified resource for host-pathogen interactions. *BMC Bioinformatics*, Vol. 11 Suppl 6, p S16

Lee, H.; Deng, M.; Sun, F. & Chen, T. (2006). An integrated approach to the prediction of domain-domain interactions. *BMC Bioinformatics*, Vol. 7, p 269

Lehner, B. & Fraser, A. G. (2004). A first-draft human protein-interaction map. *Genome Biol*, Vol. 5, No. 9, p R63

Li, P.; Zang, W.; Li, Y.; Xu, F.; Wang, J. & Shi, T. (2011). Atpid: The overall hierarchical functional protein interaction network interface and analytic platform for arabidopsis. *Nucleic Acids Res*, Vol. 39, No. Database issue, pp. D1130-3

Lin, M.; Shen, X. & Chen, X. (2011). Pair: The predicted arabidopsis interactome resource. *Nucleic Acids Res*, Vol. 39, No. Database issue, pp. D1134-40

Lin, N.; Wu, B.; Jansen, R.; Gerstein, M. & Zhao, H. (2004). Information assessment on predicting protein-protein interactions. *BMC Bioinformatics*, Vol. 5, p 154

Liu, X.; Yang, W. C.; Gao, Q. & Regnier, F. (2008). Toward chromatographic analysis of interacting protein networks. *J Chromatogr A*, Vol. 1178, No. 1-2, pp. 24-32

Mann, M. (2006). Functional and quantitative proteomics using silac. *Nat Rev Mol Cell Biol*, Vol. 7, No. 12, pp. 952-8

Marcotte, E. M.; Pellegrini, M.; Thompson, M. J.; Yeates, T. O. & Eisenberg, D. (1999). A combined algorithm for genome-wide prediction of protein function. *Nature*, Vol. 402, No. 6757, pp. 83-6

Matthews, L. R.; Vaglio, P.; Reboul, J.; Ge, H.; Davis, B. P.; Garrels, J.; Vincent, S. & Vidal, M. (2001). Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "Interologs". *Genome Res*, Vol. 11, No. 12, pp. 2120-6

McDowall, M. D.; Scott, M. S. & Barton, G. J. (2009). Pips: Human protein-protein interaction prediction database. *Nucleic Acids Res*, Vol. 37, No. Database issue, pp. D651-6

Michnick, S. W. (2003). Protein fragment complementation strategies for biochemical network mapping. *Curr Opin Biotechnol*, Vol. 14, No. 6, pp. 610-7

Mika, S. & Rost, B. (2006). Protein-protein interactions more conserved within species than across species. *PLoS Comput Biol*, Vol. 2, No. 7, p e79

Morsy, M.; Gouthu, S.; Orchard, S.; Thorneycroft, D.; Harper, J. F.; Mittler, R. & Cushman, J. C. (2008). Charting plant interactomes: Possibilities and challenges. *Trends Plant Sci*, Vol. 13, No. 4, pp. 183-91

Mrowka, R.; Patzak, A. & Herzel, H. (2001). Is there a bias in proteome research? *Genome Res*, Vol. 11, No. 12, pp. 1971-3

Navratil, V.; de Chassey, B.; Meyniel, L.; Pradezynski, F.; Andre, P.; Rabourdin-Combe, C. & Lotteau, V. (2010). System-level comparison of protein-protein interactions between viruses and the human type i interferon system network. *J Proteome Res*, Vol. 9, No. 7, pp. 3527-36

Navratil, V.; de Chassey, B.; Combe, C. R. & Lotteau, V. (2011). When the human viral infectome and diseasome networks collide: Towards a systems biology platform for the aetiology of human diseases. *BMC Syst Biol*, Vol. 5, p 13

Nelson, C. J.; Huttlin, E. L.; Hegeman, A. D.; Harms, A. C. & Sussman, M. R. (2007). Implications of 15n-metabolic labeling for automated peptide identification in arabidopsis thaliana. *Proteomics*, Vol. 7, No. 8, pp. 1279-92

Ng, S. K.; Zhang, Z. & Tan, S. H. (2003). Integrative approach for computationally inferring protein domain interactions. *Bioinformatics*, Vol. 19, No. 8, pp. 923-9

Obayashi, T.; Hayashi, S.; Shibaoka, M.; Saeki, M.; Ohta, H. & Kinoshita, K. (2008). Coxpresdb: A database of coexpressed gene networks in mammals. *Nucleic Acids Res*, Vol. 36, No. Database issue, pp. D77-82

Obayashi, T. & Kinoshita, K. (2011). Coxpresdb: A database to compare gene coexpression in seven model animals. *Nucleic Acids Res*, Vol. 39, No. Database issue, pp. D1016-22

Obayashi, T.; Nishida, K.; Kasahara, K. & Kinoshita, K. (2011). Atted-ii updates: Condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. *Plant Cell Physiol*, Vol. 52, No. 2, pp. 213-9

Obrdlik, P.El-Bakkoury, M.Hamacher, T.Cappellaro, C.Vilarino, C.Fleischer, C.Ellerbrok, H.Kamuzinzi, R.Ledent, V.Blaudez, D., et al. (2004). K+ channel interactions detected by a genetic system optimized for systematic studies of membrane protein interactions. *Proc Natl Acad Sci U S A*, Vol. 101, No. 33, pp. 12242-7

Ogmen, U.; Keskin, O.; Aytuna, A. S.; Nussinov, R. & Gursoy, A. (2005). Prism: Protein interactions by structural matching. *Nucleic Acids Res*, Vol. 33, No. Web Server issue, pp. W331-6

Ohad, N.; Shichrur, K. & Yalovsky, S. (2007). The analysis of protein-protein interactions in plants by bimolecular fluorescence complementation. *Plant Physiol*, Vol. 145, No. 4, pp. 1090-9

Ohad, N. & Yalovsky, S. (2010). Utilizing bimolecular fluorescence complementation (bifc) to assay protein-protein interaction in plants. *Methods Mol Biol*, Vol. 655, pp. 347-58

Overbeek, R.; Fonstein, M.; D'Souza, M.; Pusch, G. D. & Maltsev, N. (1999). Use of contiguity on the chromosome to predict functional coupling. *In Silico Biol*, Vol. 1, No. 2, pp. 93-108

Pagel, P.Kovac, S.Oesterheld, M.Brauner, B.Dunger-Kaltenbach, I.Frishman, G.Montrone, C.Mark, P.Stumpflen, V.Mewes, H. W., et al. (2005). The mips mammalian protein-protein interaction database. *Bioinformatics*, Vol. 21, No. 6, pp. 832-4

Pages, S.; Belaich, A.; Belaich, J. P.; Morag, E.; Lamed, R.; Shoham, Y. & Bayer, E. A. (1997). Species-specificity of the cohesin-dockerin interaction between clostridium thermocellum and clostridium cellulolyticum: Prediction of specificity determinants of the dockerin domain. *Proteins*, Vol. 29, No. 4, pp. 517-27

Parrish, J. R.; Gulyas, K. D. & Finley, R. L., Jr. (2006). Yeast two-hybrid contributions to interactome mapping. *Curr Opin Biotechnol*, Vol. 17, No. 4, pp. 387-93

Patil, A. & Nakamura, H. (2005). Filtering high-throughput protein-protein interaction data using a combination of genomic features. *BMC Bioinformatics*, Vol. 6, p 100

Patil, A.; Nakai, K. & Nakamura, H. (2011). Hitpredict: A database of quality assessed protein-protein interactions in nine species. *Nucleic Acids Res*, Vol. 39, No. Database issue, pp. D744-9

Paulmurugan, R. & Gambhir, S. S. (2005). Firefly luciferase enzyme fragment complementation for imaging in cells and living animals. *Anal Chem*, Vol. 77, No. 5, pp. 1295-302

Paulmurugan, R. & Gambhir, S. S. (2007). Combinatorial library screening for developing an improved split-firefly luciferase fragment-assisted complementation system for studying protein-protein interactions. *Anal Chem*, Vol. 79, No. 6, pp. 2346-53

Pellegrini, M.; Marcotte, E. M.; Thompson, M. J.; Eisenberg, D. & Yeates, T. O. (1999). Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc Natl Acad Sci U S A*, Vol. 96, No. 8, pp. 4285-8

Pinney, J. W.; Dickerson, J. E.; Fu, W.; Sanders-Beer, B. E.; Ptak, R. G. & Robertson, D. L. (2009). Hiv-host interactions: A map of viral perturbation of the host system. *AIDS*, Vol. 23, No. 5, pp. 549-54

Popescu, S. C.; Popescu, G. V.; Bachan, S.; Zhang, Z.; Seay, M.; Gerstein, M.; Snyder, M. & Dinesh-Kumar, S. P. (2007). Differential binding of calmodulin-related proteins to their targets revealed through high-density arabidopsis protein microarrays. *Proc Natl Acad Sci U S A*, Vol. 104, No. 11, pp. 4730-5

Ramachandran, N.; Hainsworth, E.; Bhullar, B.; Eisenstein, S.; Rosen, B.; Lau, A. Y.; Walter, J. C. & LaBaer, J. (2004). Self-assembling protein microarrays. *Science*, Vol. 305, No. 5680, pp. 86-90

Reinders, A.; Schulze, W.; Kuhn, C.; Barker, L.; Schulz, A.; Ward, J. M. & Frommer, W. B. (2002). Protein-protein interactions between sucrose transporters of different affinities colocalized in the same enucleate sieve element. *Plant Cell*, Vol. 14, No. 7, pp. 1567-77

Rhodes, D. R.; Tomlins, S. A.; Varambally, S.; Mahavisno, V.; Barrette, T.; Kalyana-Sundaram, S.; Ghosh, D.; Pandey, A. & Chinnaiyan, A. M. (2005). Probabilistic model of the human protein-protein interaction network. *Nat Biotechnol*, Vol. 23, No. 8, pp. 951-9

Rohila, J. S.Chen, M.Chen, S.Chen, J.Cerny, R.Dardick, C.Canlas, P.Xu, X.Gribskov, M.Kanrar, S., et al. (2006). Protein-protein interactions of tandem affinity purification-tagged protein kinases in rice. *Plant J*, Vol. 46, No. 1, pp. 1-13

Salwinski, L.; Miller, C. S.; Smith, A. J.; Pettit, F. K.; Bowie, J. U. & Eisenberg, D. (2004). The database of interacting proteins: 2004 update. *Nucleic Acids Res*, Vol. 32, No. Database issue, pp. D449-51

Sayers, E. W.Barrett, T.Benson, D. A.Bolton, E.Bryant, S. H.Canese, K.Chetvernin, V.Church, D. M.DiCuccio, M.Federhen, S., et al. (2011). Database resources of the national center for biotechnology information. *Nucleic Acids Res*, Vol. 39, No. Database issue, pp. D38-51

Schoonheim, P. J.; Veiga, H.; Pereira Dda, C.; Friso, G.; van Wijk, K. J. & de Boer, A. H. (2007). A comprehensive analysis of the 14-3-3 interactome in barley leaves using a complementary proteomics and two-hybrid approach. *Plant Physiol*, Vol. 143, No. 2, pp. 670-83

Schulze, W. X.; Reinders, A.; Ward, J.; Lalonde, S. & Frommer, W. B. (2003). Interactions between co-expressed arabidopsis sucrose transporters in the split-ubiquitin system. *BMC Biochem*, Vol. 4, p 3

Shin, C. J.; Wong, S.; Davis, M. J. & Ragan, M. A. (2009). Protein-protein interaction as a predictor of subcellular location. *BMC Syst Biol*, Vol. 3, p 28

Sprinzak, E. & Margalit, H. (2001). Correlated sequence-signatures as markers of protein-protein interaction. *J Mol Biol*, Vol. 311, No. 4, pp. 681-92

Sprinzak, E.; Sattath, S. & Margalit, H. (2003). How reliable are experimental protein-protein interaction data? *J Mol Biol*, Vol. 327, No. 5, pp. 919-23

Stark, C.; Breitkreutz, B. J.; Reguly, T.; Boucher, L.; Breitkreutz, A. & Tyers, M. (2006). Biogrid: A general repository for interaction datasets. *Nucleic Acids Res*, Vol. 34, No. Database issue, pp. D535-9

Stark, C.Breitkreutz, B. J.Chatr-Aryamontri, A.Boucher, L.Oughtred, R.Livstone, M. S.Nixon, J.Van Auken, K.Wang, X.Shi, X., et al. (2011). The biogrid interaction database: 2011 update. *Nucleic Acids Res*, Vol. 39, No. Database issue, pp. D698-704

Symonds, G. P.; Johnstone, H. A.; Millington, M. L.; Boyd, M. P.; Burke, B. P. & Breton, L. R. (2010). The use of cell-delivered gene therapy for the treatment of hiv/aids. *Immunol Res*, Vol. 48, No. 1-3, pp. 84-98

Szklarczyk, D.Franceschini, A.Kuhn, M.Simonovic, M.Roth, A.Minguez, P.Doerks, T.Stark, M.Muller, J.Bork, P., et al. (2011). The string database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*, Vol. 39, No. Database issue, pp. D561-8

Tahirov, T. H.; Babayeva, N. D.; Varzavand, K.; Cooper, J. J.; Sedore, S. C. & Price, D. H. (2010). Crystal structure of hiv-1 tat complexed with human p-tefb. *Nature*, Vol. 465, No. 7299, pp. 747-51

Tamames, J.; Casari, G.; Ouzounis, C. & Valencia, A. (1997). Conserved clusters of functionally related genes in two bacterial genomes. *J Mol Evol*, Vol. 44, No. 1, pp. 66-73

Tang, X.; Munske, G. R.; Siems, W. F. & Bruce, J. E. (2005). Mass spectrometry identifiable cross-linking strategy for studying protein-protein interactions. *Anal Chem*, Vol. 77, No. 1, pp. 311-8

Uhrig, J. F. & Hulskamp, M. (2006). Plant gtpases: Regulation of morphogenesis by rops and ros. *Curr Biol*, Vol. 16, No. 6, pp. R211-3

von Mering, C.; Krause, R.; Snel, B.; Cornell, M.; Oliver, S. G.; Fields, S. & Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, Vol. 417, No. 6887, pp. 399-403

von Mering, C.; Jensen, L. J.; Snel, B.; Hooper, S. D.; Krupp, M.; Foglierini, M.; Jouffre, N.; Huynen, M. A. & Bork, P. (2005). String: Known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res*, Vol. 33, No. Database issue, pp. D433-7

von Mering, C.; Jensen, L. J.; Kuhn, M.; Chaffron, S.; Doerks, T.; Kruger, B.; Snel, B. & Bork, P. (2007). String 7--recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res*, Vol. 35, No. Database issue, pp. D358-62

Wang, C.; Cheng, J. & Su, S. (2009). Prediction of interacting protein pairs from sequence using a bayesian method. *Protein J*, Vol. 28, No. 2, pp. 111-5

Willis, R. C. & Hogue, C. W. (2006). Searching, viewing, and visualizing data in the biomolecular interaction network database (bind). *Curr Protoc Bioinformatics*, Vol. Chapter 8, p Unit 8 9

Winnenburg, R.; Baldwin, T. K.; Urban, M.; Rawlings, C.; Kohler, J. & Hammond-Kosack, K. E. (2006). Phi-base: A new database for pathogen host interactions. *Nucleic Acids Res*, Vol. 34, No. Database issue, pp. D459-64

Wu, X.; Zhu, L.; Guo, J.; Zhang, D. Y. & Lin, K. (2006). Prediction of yeast protein-protein interaction network: Insights from the gene ontology and annotations. *Nucleic Acids Res*, Vol. 34, No. 7, pp. 2137-50

Xiang, Z.; Tian, Y. & He, Y. (2007). Phidias: A pathogen-host interaction data integration and analysis system. *Genome Biol*, Vol. 8, No. 7, p R150

Xu, Y.; Hu, W.; Chang, Z.; Duanmu, H.; Zhang, S.; Li, Z.; Yu, L. & Li, X. (2011). Prediction of human protein-protein interaction by a mixed bayesian model and its application to exploring underlying cancer-related pathway crosstalk. *J R Soc Interface*, Vol. 8, No. 57, pp. 555-67

Zhou, J.; Lin, J.; Zhou, C.; Deng, X. & Xia, B. (2011). An improved bimolecular fluorescence complementation tool based on superfolder green fluorescent protein. *Acta Biochim Biophys Sin (Shanghai)*, Vol. 43, No. 3, pp. 239-44
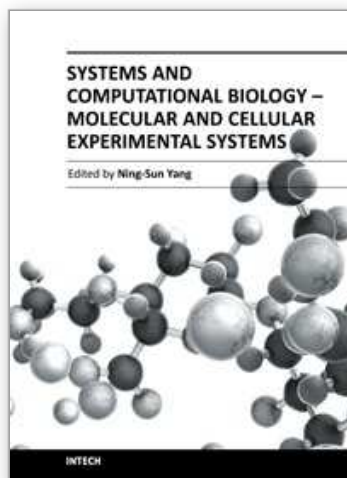
Systems and Computational Biology - Molecular and Cellular Experimental Systems

Edited by Prof. Ning-Sun Yang

Whereas some "microarray" or "bioinformatics" scientists among us may have been criticized as doing "cataloging research", the majority of us believe that we are sincerely exploring new scientific and technological systems to benefit human health, human food and animal feed production, and environmental protections. Indeed, we are humbled by the complexity, extent and beauty of cross-talks in various biological systems; on the other hand, we are becoming more educated and are able to start addressing honestly and skillfully the various important issues concerning translational medicine, global agriculture, and the environment. The two volumes of this book presents a series of high-quality research or review articles in a timely fashion to this emerging research field of our scientific community.

How to reference

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

Theresa Tsun-Hui Tsao, Chen-Hsiung Chan, Chi-Ying F. Huang and Sheng-An Lee (2011). The Prediction and Analysis of Inter- and Intra-Species Protein-Protein Interaction, Systems and Computational Biology - Molecular and Cellular Experimental Systems, Prof. Ning-Sun Yang (Ed.), ISBN: 978-953-307-280-7, InTech, Available from: http://www.intechopen.com/books/systems-and-computational-biology-molecular-and-cellular-experimental-systems/the-prediction-and-analysis-of-inter-and-intra-species-protein-protein-interaction

INTECH
open science | open minds