Chapter 10

The Integration and Annotation of the Human Interactome in the UniHI Database

Gautam Chaurasia and Matthias Futschik

Abstract

In recent years, remarkable progress has been made toward the systematic charting of human protein interactions. The utilization of the generated interaction data remained however challenging for biomedical researchers due to lack of integration of currently available resources. To facilitate the direct access and analysis of the human interactome, we have developed the Unified Human Interactome (UniHI) database. It provides researchers with a user-friendly Web-interface and integrates interaction data from 12 major resources in its latest version, establishing one of the largest catalogs for human PPIs worldwide. At present, UniHI houses over 250,000 distinct interactions between 22,300 unique proteins and is publically available at http://www.unihi.org.

Key words: Protein–protein interactions, Database, Data integration, Web applications, Bioinformatics

1. Introduction

In a living organism, proteins interact with other proteins to carry out vital cellular functions, such as signal transduction, DNA replication, transcription, protein transport, or metabolic catalysis. Also, many major diseases such as neurological disorders or cancer are characterized by complex interactions of multiple proteins (1–9). The study of human protein interactions may therefore help (i) to improve our general understanding of biological processes, (ii) to decipher the molecular basis of complex diseases, and (iii) to provide new potential therapeutic targets.

For many years, interactions between proteins have been studied in small-scale experiments. This situation has, however, dramatically changed in the last decade. The availability of fully sequenced

Bernhard Suter and Erich E. Wanker (eds.), *Two Hybrid Technologies: Methods and Protocols*, Methods in Molecular Biology, vol. 812, DOI 10.1007/978-1-61779-455-1_10, © Springer Science+Business Media, LLC 2012

genomes (10) and advances in high-throughput approaches (11–14) have led to large-scale studies of protein–protein interactions on a genome-wide scale and to efforts to map the complete protein–protein interaction (PPI) network for an organism, termed also as interactome. Indeed, we have recently witnessed many large-scale protein interaction mapping projects in several model organisms such as *Saccharomyces cerevisiae* (15–18), *Drosophila melanogaster* (19), and *Caenorhabditis elegans* (20). Now, the focus has moved toward a systematic mapping of human protein–protein interactions (21–33). The constructed human PPI maps have been derived from both experimental and computational approaches (11–14, 34, 35) and not only offer a wealth of information but are also expected to a valuable tool for the biomedical research community (6, 7, 36, 37). However, utilization of these interaction maps is impeded by manifold limitations:

A major limitation of current approaches to map the interactome is their limited capability to capture interactions in a comprehensive manner (38, 39). Several studies have shown that the current human PPI maps are incomplete and highly unsaturated (40–42). A further shortcoming of human PPI maps has been a lack of integration. Although many large human PPI maps are publicly accessible, they are distributed at multiple locations. To find comprehensive information on potential interaction partners, researchers have to perform multiple searches at different databases. This, however, can be very time-consuming due to many factors such as various query formats and identifiers used in different interaction databases. Another drawback of many PPI databases is their restriction to the search of interaction partners only for a single query protein. By contrast, comprehensive molecular studies frequently require complex network-oriented searches for interactions of multiple proteins.

Finally, determining the biological function of networks has remained a bottleneck in the analysis of PPI data. In order to understand the complexity of interaction networks, it is necessary to assess not only the function of individual proteins but also their molecular context. Thus, PPI data have to be integrated with other functional data and information to facilitate their interpretation.

UniHI (43–45) was developed to overcome these limitations and to provide a large community of researchers in biomedicine direct access to comprehensive information on human proteins and their potential interaction partners. UniHI currently stores more than a quarter of a million human protein interactions that have been experimentally detected or computationally predicted. Several Web applications are included in UniHI offering a convenient interface for the network analysis, especially for researchers less acquainted with computational investigations. For functional interpretation of interaction networks, PPI data were integrated

with biological pathway data from Encyclopedia of Genes and Genomes (KEGG) (46, 47) and gene expression data from Human Gene Atlas (48). The integrated data can be utilized through advanced tools (i.e., UniHI Express and UniHI Scanner) that enable users to construct tissue-specific interaction networks or to annotate edges with the specific type of interaction.

2. Materials

2.1. PPI Data

Currently, protein interactions in UniHI are derived from 12 large-scale human protein–protein interaction maps (21–33). These maps were generated using yeast-two-hybrid (Y2H) assays, literature review or orthology-based approaches. UniHI includes interaction maps that are based on curated literature (i.e., IntAct (21), BIND (22), BioGrid (23), and REACTOME (27), HPRD (29), DIP (32), on computational text-mining (i.e., COCIT (30)), on large Y2H-based screens (i.e., CCSB-H1 (31) and MDC-Y2H (33)) and on computational prediction (i.e., OPHID (24) and ORTHO (26), HOMOMINT (28)). Further details on these maps are given in the Table 1.

2.2. Gene Expression Data

Gene expression data were collected from Gene Atlas database (48). The data set comprises of expression profiles for 79 human tissues measured in replicates using Affymetrix HG-U133A and custom-designed GNF1H arrays. Expression summaries for the 44,775 transcripts were derived utilizing the MAS5 algorithm by Affymetrix.

2.3. Gene Annotation

Additional information on proteins (e.g., functional annotations and description of proteins, chromosomal location, and potential disease association of corresponding genes) were imported from National Center for Biotechnology Information (NCBI) (49), Online Mendelian Inheritance in Man (OMIM) (50), and Gene Ontology (GO) (51, 52) databases.

2.4. Pathway Information

Association of genes with pathways were derived from the KEGG pathway database (46, 47), which constitutes a collection of manually drawn pathway maps representing accumulated knowledge of molecular interaction and reaction networks for metabolism, cellular processes, and human diseases, as well as for genetic and environmental information processing. In addition, information about the relations between pathway elements such as regulatory or physical interactions (e.g., phosphorylation, dephosphorylation, activation, inhibition, and binding association) was collected from the KEGG pathways database.

Table 1
Current list of PPI datasets, integrated in UniHI

Dataset	Proteins	Interactions	Method	Database location
MDC-Y2H	1,703	3,186	Y2H SCREEN (33)	http://www.mdc-berlin.de/neuroprot
CCSB-Y2H	1,549	2,754	Y2H SCREEN (31)	http://vidal.dfci.harvard.edu (flat file only)
HPRD-BIN	8,788	32,776	LITERATURE (29)	http://www.hprd.org
DIP	1,085	1,397	LITERATURE (32)	http://dip.doe-mbi.ucla.edu
BioGrid	7,953	24,624	LITERATURE (23)	http://www.thebiogrid.org
IntAct	7,273	19,404	LITERATURE (21)	http://www.ebi.ac.uk/intact
BIND	5,286	7,394	LITERATURE (22)	http://www.bind.ca
COCIT	3,737	6,580	TEXT MINING (30)	http://www.reactome.org
REACTOME	1,554	37,332	LITERATURE (27)	http://Bioinformatics.icmb.utexas.edu/idserve/
ORTHO	6,225	71,466	ORTHOLOGY (26)	http://www.sanger.ac.uk/PostGenomics/signaltransduction/interactionmap
HOMOMINT	4,127	10,174	ORTHOLOGY (28)	http://mint.bio.uniroma2.it
OPHID	4,785	24,991	ORTHOLOGY (24)	http://ophid.utoronto.ca
UniHI	22,307	200,473	INTEGRATION (45)	http://www.unihi.org

Number of proteins and interactions in each dataset as well as construction approach are given

2.5. Gene and Protein Identifiers

Since PPI data were collected from different sources, one of the major problems was to establish common identifiers for their aggregation. For this purpose, lists of different protein identifiers were downloaded from the Web sites of NCBI, HUGO Gene Nomenclature Committee (HGNC) (53), and Ensembl EnsMart (54, 55).

3. Methods

3.1. Architecture of UniHI

The UniHI system should offer users a convenient entry gate to a comprehensive collection of PPI data. With this aim in mind, we have designed its architecture to comply with following requirements: (i) The structure of the platform should be extensible to new data without changing its data structure as the amount of data of protein interactions is growing rapidly; (ii) it should be easily accessible; (iii) queries should be processed in minimal time;

(iv) data should be accurately integrated from different sources; and (v) it should also be consistently updated.

To meet these demands, UniHI system is based on a multitier architecture with different layers, each assigned to a certain task. A data access layer is responsible for writing and retrieving data to and from a locally installed database. A business layer processes the applications and queries by the user. Finally, a presentation layer provides a Web interface and visualization tools for accessing and viewing interactions. The advantage of the implemented UniHI architecture is its modularity and portability. The technical details of the current UniHI system were described in a recent publication (43).

3.2. Mapping of Proteins

As UniHI integrates a large number of different PPI resources, aggregation of heterogeneous interaction maps and building a unique identifier indexing system is a foremost task. For unification of primary data, we first computed complete lists of proteins for each interaction map separately. Subsequently, these lists were compared employing information from NCBI, HGNC (53, 54), and EnsMart (55) to map corresponding identifiers between interaction datasets. After mapping, identical protein identifiers were merged together in a horizontal manner where each protein is a unique entry in a table. A unique identifier was assigned to each protein entry of this table. These unique identifiers were further used for grouping of the redundant interactions from all interaction datasets.

3.3. Search Interface and Filtering Methods

The UniHI system has been designed to facilitate network-oriented functional analyses. It provides several tools to perform different types of network searches and to map the expression and pathway data onto retrieved networks. Several gene and protein identifiers, e.g., gene symbol, Entrez Gene, Uniprot, NCBI Geneinfo, Ensembl, Biogrid, and HPRD IDs, can be used for searching interaction partners. Notably, these identifiers can now also be employed for direct hyperlinks to UniHI. Users can also choose the PPI resource, from which interactions data should be retrieved. This permits, for instance, the exclusion of less validated mapping approaches such as computational prediction. As primary output, a list of interaction partners for the query proteins is presented (Fig. 1). This list can subsequently be filtered based on additional criteria. For instance, interactions can be excluded if they are not associated with a minimum number of PubMed references in which they have been reported (Fig. 2a).

3.4. Visualization of Protein Interactions

Visualization of the retrieved interaction networks remains crucial for the evaluation of query results. Thus, we have implemented a visualization tool for interaction data that offer many attractive features for rapid analysis and adjustment of the extracted information.

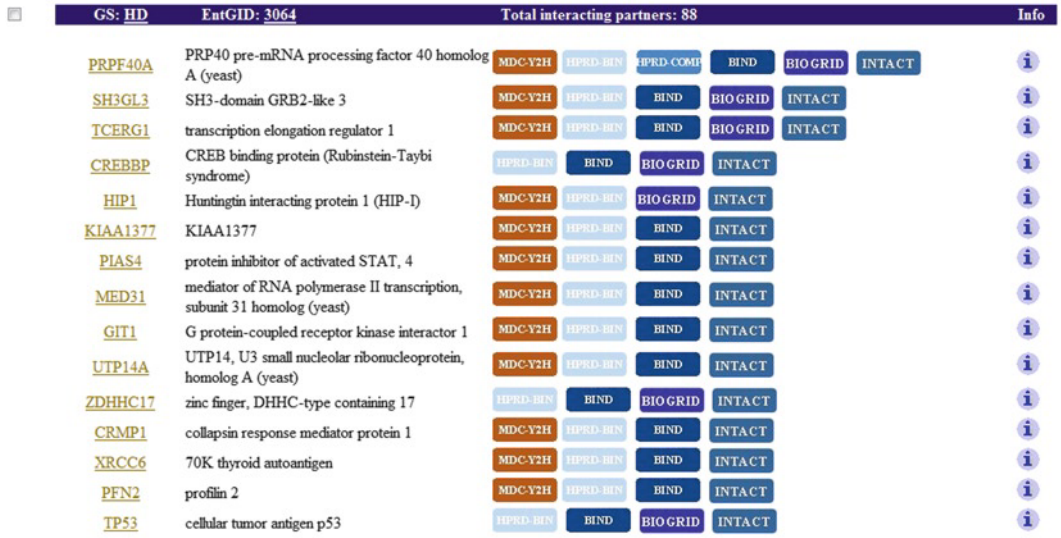

Fig. 1. Textual representation of a query result for protein interactions in UniHI. Multiple links indicate identification of the interaction by different methods. For easy discrimination between maps, specific colors have been assigned. *Shades of blue* have been used for datasets derived by literature search, *shades of green* for orthology-based maps, *shades of red* for maps derived from Y2H screens.

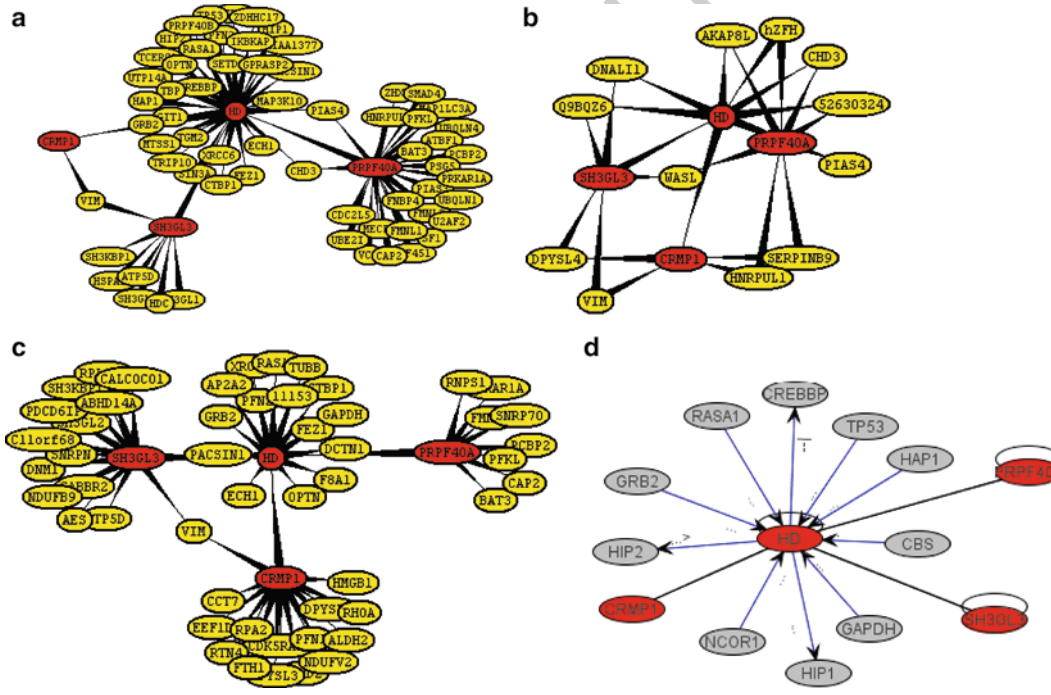

Fig. 2. Graphical representation and analysis of PPI networks using UniHI Search, UniHI Express and UniHI Scanner tools. Display of the interaction partners (*yellow or gray*) of the query proteins (*red*) HD, CRMP1, PRPF40A, and SH3GL3. (**a**) UniHI Search: Using the number of PubMed as a criterion, only those interaction partners are shown, which have been reported in more than two publications. (**b**) Only brain-specific networks with a medium expression level. (**d**) UniHI Scanner: Annotation of found network using KEGG pathway. *Gray nodes* represent proteins included in KEGG "Huntington Disease" pathway. UniHI Scanner allows a display of the intersection between the retrieved PPI network and the pathway. Additional information is given regarding the type of interaction (e.g., phosphorylation (+P), dephosphorylation (–P), activation (–>), or inhibition (–I)) facilitating the assessment of the retrieved interactions.

For filtering of the network and manual adjustment of the layout, nodes (representing proteins) can be anchored or hidden. Also, information about proteins and interactions can be accessed directly via the visualization interface, thereby avoiding cumbersome review of textual output. Moreover, the display can be restricted to direct interactions between query proteins or extended to include bridging proteins (Fig. 2b).

3.5. Advanced Tools for Network Analysis

Addressing the need of a more dynamic interactome and its functional interpretation, we developed UniHI Express and UniHI Pathway Scanner as two new tools in our database (Fig. 2c, d). UniHI Express allows the filtering of PPI based on the expression in a selected tissue and thus enables the construction of tissue-specific networks. First preliminary studies show that the usage of UniHI express can be highly efficient to prioritize interactions. Furthermore, UniHI Pathway Scanner provides the possibility to examine the intersection of canonical pathways from KEGG with the extracted networks. Thus, it enables researchers to detect possible modifiers of pathways as well as proteins involved in the cross talk between different pathways.

3.5.1. UniHI Express

Biomedical researchers frequently study processes that occur in specific tissues or cell types, e.g., degeneration of neuronal tissue. By contrast, current collections of protein interactions are derived from experiments using various cell and tissue types. Thus, PPI networks retrieved from these resources represent rather a gross summary of possible interactions, thereby neglecting the actual conditions in a specific tissue of interest. In fact, since only a small percentage of proteins correspond to ubiquitously expressed housekeeping genes, the probability might be high, that many proteins included in retrieved networks are not expressed in a chosen tissue. Hence, researchers are required to examine carefully the presence of proteins in their model system of interest. This is a considerable task considering the large number of interaction partners that queries with even a small number of proteins can produce. We, therefore, integrated gene expression data to allow researchers the construction of tissue-specific networks. As tissue expression data set, the Human Gene Expression Atlas was utilized (48).

To enable the integration with PPI data, microarray probes were mapped to their corresponding Entrez Gene IDs using the annotation generated by the curators of the Gene Expression atlas. Expression values were averaged over probes which correspond to the same nonredundant Entrez Gene IDs. To facilitate the use of UniHI Express, the samples were assigned to 19 main tissue classes. These main classes comprise adrenal gland, brain, heart, kidney, liver, lung, prostate, pancreas, placenta, muscle, salivary gland, thymus, thyroid, tonsil, lymph node, testis, trachea, uterus, and uterine corpus.

To obtain a unique tissue expression profile, we averaged expression values of tissues samples belonging to the same class.

Employing UniHI Express, users can filter the interacting proteins by requiring a minimum expression in the selected tissue class. Note that the cutoff value is not applied to the query protein. By adjusting the expression threshold, the PPI network can be reduced to include only highly expressed or can be extended to include also lowly expressed proteins (Fig. 2c). Additionally, the PPI resources to be queried can be specified.

3.5.2. UniHI Scanner

PPI maps represent the chain of molecular events connected with each other. Functional analyses of these events can help us to understand the roles of individual proteins and the interplay between them (56–58). Indeed, PPI networks need to be integrated with other functional data to gain new insights into molecular mechanisms. Therefore, we have integrated the UniHI protein interaction data with pathway data from KEGG (46, 47).

As platform for integration of PPI with pathway data, we have implemented a new tool termed UniHI Pathway Scanner, which offers a search Web interface. Here, users can provide a list of query proteins and choose a list of pathways to be scanned against identified network. Additionally, the source of interaction data can be selected. Subsequently, UniHI Scanner performs three functions. First, interaction partners of search proteins are identified and a PPI network is created. Second, a pathway network is created from the KEGG pathway IDs provided by user. Edges of this pathway network contain the explicit information about the mode of interactions (such as phosphorylation, activation, or inhibition). Finally, the PPI and pathway networks are intersected. Nodes and edges in the PPI network are annotated if they are found in the pathway network. Thus, directed edges in an annotated network represent the molecular relations found in KEGG pathways between proteins (Fig. 10.2d). For viewing both types of networks, an interactive visualization tool is provided. Users can choose between the display of the full PPI network or the annotated intersection. Additionally, researchers can rearrange the layout of PPI network. For instance, the whole network or individuals nodes can be moved and their location adjusted. Details of networks can be magnified by the zoom function. With these features, UniHI Pathway Scanner is anticipated to be a valuable tool to identify modifiers of pathways or to detect proteins involved in cross talk between pathways.

4. Notes

In the postgenomic era, one of the daunting tasks of proteomics is to chart the complete protein–protein interaction networks that occur within organisms (59, 60). Although, the availability of many [AU1]

large genome sequences and advances in high-throughput methods provided us a platform to construct PPI maps, the hitherto constructed interactome are far from complete (61, 62).

Not surprisingly, this is also the case for the human interactome. Additionally, currently available human PPI data had been stored in various locations. In a recent survey, we observed that interaction maps derived from different resources are remarkably distinct (45). In fact, less than 19% of all interactions occurred in multiple maps indicating a low degree of saturation of single maps. The small number of shared interactions is striking considering that more than 50% of all proteins were included in two or more maps. Thus, current PPI datasets are highly complementary sharing few interactions between many common proteins. For researchers, it is clearly tempting to use as many resources as possible for the in silico construction of their networks of interest. However, it is important to keep in mind that the single human PPI maps are generated using different approaches and that they might contain characteristic biases, i.e., over- or underrepresentation of proteins from certain categories, due to the sensitivity and specificity of applied methods toward specific classes of proteins. In a comparative analysis, we indeed detected a strong sampling and detection biases linked to the method used to generate a PPI map (41). For example, RNA binding proteins were overrepresented in orthology-based maps (e.g., ORTHO), whereas signal transducers were overproportionally sampled in literature-based maps. A significant depletion of membrane proteins was observed in all networks and not only in Y2H-based maps as previously expected. Moreover, different maps of different origin showed considerable divergence with respect to their internal structure (40). These findings are necessary to be considered to avoid pitfalls in the application of PPI maps. We therefore advise researchers to carefully check whether the results of their network analyses might be affected by the choice of resource. Note that this kind of examination is facilitated by the option implemented in UniHI to select or deselect particular primary PPI resources.

The quality of PPI data also remains a crucial issue. It has been frequently stated that especially data produced by high-throughput methods contain a high rate of false positive or negative interactions (40, 41, 61, 62). However, a recent study questioned this view and rather suggested that PPI data sets based on small scale studies include a considerable rate of unreliable interactions (63). A major reason for this controversy is the lack of an unbiased high confidence human interaction data set which would allow a stringent assessment of the quality of different mapping approaches. Although the existence of such "gold standard" is obviously desirable, the possibility of its generation in the nearby future remains doubtful. Moreover, as protein interactions frequently require specific protein modifications, the structure of

interaction networks depends critically on the cellular conditions and is highly dynamic.

Since there is no unique quality index for interaction data, we have provided in UniHI additional information that enables researchers to assess the reliability of the retrieved interactions. In particular, we have computed two different measures: coannotation and coexpression. Coannotation assumes that two proteins are likely to interact if they share same functions, involved in same biological process or are allocated to the same cellular compartment. For computing coannotation, we assessed the similarity of GO categories assigned to interacting proteins. The similarity of GO categories was approximated by calculating the length of the shared path from the root category. Large shared path lengths indicate that the GO categories are in proximity to each other within the GO graph and thus can be considered similar. In case of multiple GO assignments for proteins, the largest shared path length is counted.

Similarly, coexpression provides an indication whether two proteins are coregulated on transcriptional level, and has been shown to correlate with the probability of interaction (64, 65). To measure coexpression for interacting proteins, Spearman rank correlation of expression levels was calculated. Additionally, the corresponding quantiles for correlation coefficients were derived allowing users to assess the significance of coexpression. For instance, a quantile of 0.05 signifies that the corresponding correlation coefficient is within the top 5% of the total distribution of observed coefficients.

Evidently, the usage of gene expression as proxy for protein abundance has its limitation. However, it can give researchers valuable first indications to prioritize interactions in possible follow-up studies. For future releases, we expect that the inclusion of quantitative measurement of protein abundance in tissues – once such data become available on a proteome-wide scale – will lead to considerably more accurate tissue-specific PPI networks. Nevertheless, the current release of UniHI constitutes a first important step toward a dynamic representation of the human interactome.

Acknowledgments

The development and implementation of UniHI was funded by the SFB 618 grant of the *Deutsche Forschungsgesellschaft* (DFG). We would also like to acknowledge the support of M.F by the Portuguese *Fundação para a Ciência e Tecnologia* (IBB/CBME, LA, FEDER/POCI 2010).

References

1. Baranzini, S. E., Galwey, N. W., Wang, J., Khankhanian, P., Lindberg, R., Pelletier, D., Wu, W., Uitdehaag, B. M., Kappos, L., Polman, C. H., Matthews, P. M., Hauser, S. L., Gibson, R. A., Oksenberg, J. R., and Barnes, M. R. (2009) Pathway and network-based analysis of genome-wide association studies in multiple sclerosis, *Hum Mol Genet 18*, 2078–2090.
2. Calvano, S. E., Xiao, W., Richards, D. R., Felciano, R. M., Baker, H. V., Cho, R. J., Chen, R. O., Brownstein, B. H., Cobb, J. P., Tschoeke, S. K., Miller-Graziano, C., Moldawer, L. L., Mindrinos, M. N., Davis, R. W., Tompkins, R. G., and Lowry, S. F. (2005) A network-based analysis of systemic inflammation in humans, *Nature 437*, 1032–1037.
3. Chuang, H. Y., Lee, E., Liu, Y. T., Lee, D., and Ideker, T. (2007) Network-based classification of breast cancer metastasis, *Mol Syst Biol 3*, 140.
4. Ergun, A., Lawrence, C. A., Kohanski, M. A., Brennan, T. A., and Collins, J. J. (2007) A network biology approach to prostate cancer, *Mol Syst Biol 3*, 82.
5. Goehler, H., Lalowski, M., Stelzl, U., Waelter, S., Stroedicke, M., Worm, U., Droege, A., Lindenberg, K. S., Knoblich, M., Haenig, C., Herbst, M., Suopanki, J., Scherzinger, E., Abraham, C., Bauer, B., Hasenbank, R., Fritzsche, A., Ludewig, A. H., Bussow, K., Coleman, S. H., Gutekunst, C. A., Landwehrmeyer, B. G., Lehrach, H., and Wanker, E. E. (2004) A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease, *Mol Cell 15*, 853–865.
6. Goh, K. I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barabasi, A. L. (2007) The human disease network, *Proc Natl Acad Sci U S A 104*, 8685–8690.
7. Ideker, T., and Sharan, R. (2008) Protein networks in disease, *Genome Res 18*, 644–652.
8. Lage, K., Karlberg EO., Størling, Z., Ólason, P., Pedersen, A., Rigina, O., Hinsby, A., Tümer, Z., Pociot, F., Tommerup, N., Moreau, Y., and Brunak, S. (2007) A human phenome-interactome network of protein complexes implicated in genetic disorders, *Nature Biotechnology*, 309–316.
9. Lim, J., Hao, T., Shaw, C., Patel, A. J., Szabo, G., Rual, J. F., Fisk, C. J., Li, N., Smolyar, A., Hill, D. E., Barabasi, A. L., Vidal, M., and Zoghbi, H. Y. (2006) A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration, *Cell 125*, 801–814.
10. Ruder, K., and Winstead, E. A Quick Guide to Sequenced Genomes, http://www.genomenewsnetwork.org/resources/sequenced_genomes/.
11. Fields, S., and Song, O. (1989) A novel genetic system to detect protein-protein interactions, *Nature 340*, 245–246.
12. Figeys, D., McBroom, L. D., and Moran, M. F. (2001) Mass spectrometry for the study of protein-protein interactions, *Methods 24*, 230–239.
13. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification, *Methods 24*, 218–229.
14. Suter, B., Kittanakom, S., and Stagljar, I. (2008) Two-hybrid technologies in proteomics research, *Curr Opin Biotechnol 19*, 316–323.
15. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figeys, D., and Tyers, M. (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry, *Nature 415*, 180–183.
16. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome, *Proc Natl Acad Sci U S A 98*, 4569–4574.
17. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae, *Nature 403*, 623–627.
18. Yu, H., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., Hao, T., Rual, J. F., Dricot, A., Vazquez, A., Murray, R. R., Simon, C., Tardivo, L., Tam, S., Svrzikapa, N., Fan, C., de Smet, A. S., Motyl,

A., Hudson, M. E., Park, J., Xin, X., Cusick, M. E., Moore, T., Boone, C., Snyder, M., Roth, F. P., Barabasi, A. L., Tavernier, J., Hill, D. E., and Vidal, M. (2008) High-quality binary protein interaction map of the yeast interactome network, *Science 322*, 104–110.

19. Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E., Vijayadamodar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioime, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carrolla, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C. A., Finley, R. L., Jr., White, K. P., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R. A., McKenna, M. P., Chant, J., and Rothberg, J. M. (2003) A protein interaction map of Drosophila melanogaster, *Science 302*, 1727–1736.
20. Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T., Goldberg, D. S., Li, N., Martinez, M., Rual, J. F., Lamesch, P., Xu, L., Tewari, M., Wong, S. L., Zhang, L. V., Berriz, G. F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H. W., Elewa, A., Baumgartner, B., Rose, D. J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S. E., Saxton, W. M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K. C., Harper, J. W., Cusick, M. E., Roth, F. P., Hill, D. E., and Vidal, M. (2004) A map of the interactome network of the metazoan C. elegans, *Science 303*, 540–543.
21. Aranda, B., Achuthan, P., Alam-Faruque, Y., Armean, I., Bridge, A., Derow, C., Feuermann, M., Ghanbarian, A. T., Kerrien, S., Khadake, J., Kerssemakers, J., Leroy, C., Menden, M., Michaut, M., Montecchi-Palazzi, L., Neuhauser, S. N., Orchard, S., Perreau, V., Roechert, B., van Eijk, K., and Hermjakob, H. The IntAct molecular interaction database in 2010, *Nucleic Acids Res 38*, D525-531.
22. Bader, G. D., Betel, D., and Hogue, C. W. (2003) BIND: the Biomolecular Interaction Network Database, *Nucleic Acids Res 31*, 248–250.
23. Breitkreutz, B. J., Stark, C., Reguly, T., Boucher, L., Breitkreutz, A., Livstone, M., Oughtred, R., Lackner, D. H., Bahler, J., Wood, V., Dolinski, K., and Tyers, M. (2008) The BioGRID Interaction Database: 2008 update, *Nucleic Acids Res 36*, D637–640.
24. Brown, K. R., and Jurisica, I. (2005) Online predicted human interaction database, *Bioinformatics 21*, 2076–2082.
25. Ewing, R. M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M. D., O'Connor, L., Li, M., Taylor, R., Dharsee, M., Ho, Y., Heilbut, A., Moore, L., Zhang, S., Ornatsky, O., Bukhman, Y. V., Ethier, M., Sheng, Y., Vasilescu, J., Abu-Farha, M., Lambert, J. P., Duewel, H. S., Stewart, II, Kuehl, B., Hogue, K., Colwill, K., Gladwish, K., Muskat, B., Kinach, R., Adams, S. L., Moran, M. F., Morin, G. B., Topaloglou, T., and Figeys, D. (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry, *Mol Syst Biol 3*, 89.
26. Lehner, B., and Fraser, A. G. (2004) A first-draft human protein-interaction map, *Genome Biol 5*, R63.
27. 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., and D'Eustachio, P. (2009) Reactome knowledgebase of human biological pathways and processes, *Nucleic Acids Res 37*, D619–622.
28. Persico, M., Ceol, A., Gavrila, C., Hoffmann, R., Florio, A., and Cesareni, G. (2005) HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms, *BMC Bioinformatics 6 Suppl 4*, S21.
29. Prasad, T. S., Kandasamy, K., and Pandey, A. (2009) Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology, *Methods Mol Biol 577*, 67–79.
30. Ramani, A. K., Bunescu, R. C., Mooney, R. J., and Marcotte, E. M. (2005) Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome, *Genome Biol 6*, R40.
31. Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P., and Vidal, M. (2005) Towards a proteome-scale map of the human protein-protein interaction network, *Nature 437*, 1173–1178.

32. Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D. (2004) The Database of Interacting Proteins: 2004 update, *Nucleic Acids Res 32*, D449–451.
33. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksoz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E. E. (2005) A human protein-protein interaction network: a resource for annotating the proteome, *Cell 122*, 957–968.
34. Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S., Emili, A., Snyder, M., Greenblatt, J. F., and Gerstein, M. (2003) A Bayesian networks approach for predicting protein-protein interactions from genomic data, *Science 302*, 449–453.
35. Matthews, L. R., Vaglio, P., Reboul, J., Ge, H., Davis, B. P., Garrels, J., Vincent, S., and Vidal, M. (2001) Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs", *Genome Res 11*, 2120–2126.
36. Braun, P., Rietman, E., and Vidal, M. (2008) Networking metabolites and diseases, *Proc Natl Acad Sci U S A 105*, 9849–9850.
37. Yildirim, M. A., Goh, K. I., Cusick, M. E., Barabasi, A. L., and Vidal, M. (2007) Drug-target network, *Nat Biotechnol 25*, 1119–1126.
38. Koegl, M., and Uetz, P. (2007) Improving yeast two-hybrid screening systems, *Brief Funct Genomic Proteomic 6*, 302–312.
39. Venkatesan, K., Rual, J. F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K. I., Yildirim, M. A., Simonis, N., Heinzmann, K., Gebreab, F., Sahalie, J. M., Cevik, S., Simon, C., de Smet, A. S., Dann, E., Smolyar, A., Vinayagam, A., Yu, H., Szeto, D., Borick, H., Dricot, A., Klitgord, N., Murray, R. R., Lin, C., Lalowski, M., Timm, J., Rau, K., Boone, C., Braun, P., Cusick, M. E., Roth, F. P., Hill, D. E., Tavernier, J., Wanker, E. E., Barabasi, A. L., and Vidal, M. (2009) An empirical framework for binary interactome mapping, *Nat Methods 6*, 83–90.
40. Chaurasia, G., Herzel, H., Wanker, E. E., and Futschik, M. E. (2006) Systematic functional assessment of human protein-protein interaction maps, *Genome Inform 17*, 36–45.
41. Futschik, M. E., Chaurasia, G., and Herzel, H. (2007) Comparison of human protein-protein interaction maps, *Bioinformatics 23*, 605–611.
42. Futschik, M. E., Tschaut, A., Chaurasia, G., and Herzel, H. (2007) Graph-theoretical comparison reveals structural divergence of human protein interaction networks, *Genome Inform 18*, 141–151.
43. Chaurasia, G., Iqbal, Y., Hänig, C., Herzel, H., Wanker, E., and Futschik, M. Flexible web-based integration of distributed large-scale human protein interaction maps, *Journal of Integrative Bioinformatics 4*, 51.
44. Chaurasia, G., Iqbal, Y., Hanig, C., Herzel, H., Wanker, E. E., and Futschik, M. E. (2007) UniHI: an entry gate to the human protein interactome, *Nucleic Acids Res 35*, D590-594.
45. Chaurasia, G., Malhotra, S., Russ, J., Schnoegl, S., Hanig, C., Wanker, E. E., and Futschik, M. E. (2009) UniHI 4: new tools for query, analysis and visualization of the human protein-protein interactome, *Nucleic Acids Res 37*, D657–660.
46. Kanehisa, M., and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res 28*, 27–30.
47. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs, *Nucleic Acids Res 38*, D355–360.
48. Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M. P., Walker, J. R., and Hogenesch, J. B. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes, *Proc Natl Acad Sci U S A 101*, 6062–6067.
49. Sayers, E. W., Barrett, T., Benson, D. A., Bolton, E., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., Dicuccio, M., Federhen, S., Feolo, M., Geer, L. Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D. J., Lu, Z., Madden, T. L., Madej, T., Maglott, D. R., Marchler-Bauer, A., Miller, V., Mizrachi, I., Ostell, J., Panchenko, A., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T. A., Wagner, L., Wang, Y., John Wilbur, W., Yaschenko, E., and Ye, J. Database resources of the National Center for Biotechnology Information, *Nucleic Acids Res 38*, D5–16.
50. Lenffer, J., Nicholas, F. W., Castle, K., Rao, A., Gregory, S., Poidinger, M., Mailman, M. D., and Ranganathan, S. (2006) OMIA (Online Mendelian Inheritance in Animals): an enhanced platform and integration into the Entrez search interface at NCBI, *Nucleic Acids Res 34*, D599–601.

51. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat Genet 25*, 25–29.

52. Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G. M., Blake, J. A., Bult, C., Dolan, M., Drabkin, H., Eppig, J. T., Hill, D. P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J. M., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S., Fisk, D. G., Hirschman, J. E., Hong, E. L., Nash, R. S., Sethuraman, A., Theesfeld, C. L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S. Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E. M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., and White, R. (2004) The Gene Ontology (GO) database and informatics resource, *Nucleic Acids Res 32*, D258–261.

53. Bruford, E. A., Lush, M. J., Wright, M. W., Sneddon, T. P., Povey, S., and Birney, E. (2008) The HGNC Database in 2008: a resource for the human genome, *Nucleic Acids Res 36*, D445–448.

54. Birney, E., Andrews, D., Bevan, P., Caccamo, M., Cameron, G., Chen, Y., Clarke, L., Coates, G., Cox, T., Cuff, J., Curwen, V., Cutts, T., Down, T., Durbin, R., Eyras, E., Fernandez-Suarez, X. M., Gane, P., Gibbins, B., Gilbert, J., Hammond, M., Hotz, H., Iyer, V., Kahari, A., Jekosch, K., Kasprzyk, A., Keefe, D., Keenan, S., Lehvaslaiho, H., McVicker, G., Melsopp, C., Meidl, P., Mongin, E., Pettett, R., Potter, S., Proctor, G., Rae, M., Searle, S., Slater, G., Smedley, D., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Storey, R., Ureta-Vidal, A., Woodwark, C., Clamp, M., and Hubbard, T. (2004) Ensembl 2004, *Nucleic Acids Res 32*, D468–470.

55. Kasprzyk, A., Keefe, D., Smedley, D., London, D., Spooner, W., Melsopp, C., Hammond, M., Rocca-Serra, P., Cox, T., and Birney, E. (2004) EnsMart: a generic system for fast and flexible access to biological data, *Genome Res 14*, 160–169.

56. Barabasi, A. L., and Oltvai, Z. N. (2004) Network biology: understanding the cell's functional organization, *Nat Rev Genet 5*, 101–113.

57. Spirin, V., and Mirny, L. A. (2003) Protein complexes and functional modules in molecular networks, *Proc Natl Acad Sci U S A 100*, 12123–12128.

58. Yook, S. H., Oltvai, Z. N., and Barabasi, A. L. (2004) Functional and topological characterization of protein interaction networks, *Proteomics 4*, 928–942.

59. Bork, P., Jensen, L. J., von Mering, C., Ramani, A. K., Lee, I., and Marcotte, E. M. (2004) Protein interaction networks from yeast to human, *Curr Opin Struct Biol 14*, 292–299.

60. Hart, G. T., Ramani, A. K., and Marcotte, E. M. (2006) How complete are current yeast and human protein-interaction networks?, *Genome Biol 7*, 120.

61. Bader, G. D., and Hogue, C. W. (2002) Analyzing yeast protein-protein interaction data obtained from different sources, *Nat Biotechnol 20*, 991–997.

62. von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., and Bork, P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions, *Nature 417*, 399–403.

63. Cusick, M. E., Yu, H., Smolyar, A., Venkatesan, K., Carvunis, A. R., Simonis, N., Rual, J. F., Borick, H., Braun, P., Dreze, M., Vandenhaute, J., Galli, M., Yazaki, J., Hill, D. E., Ecker, J. R., Roth, F. P., and Vidal, M. (2009) Literature-curated protein interaction datasets, *Nat Methods 6*, 39–46.

64. Ge, H., Liu, Z., Church, G. M., and Vidal, M. (2001) Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae, *Nat Genet 29*, 482–486.

65. Mrowka, R., Patzak, A., and Herzel, H. (2001) Is there a bias in proteome research?, *Genome Res 11*, 1971–1973.

Author Queries

Chapter No.: 10 0001340572

Queries	Details Required	Author's Response
AU1	As per Walker series, notes should be in the numbered list form. Also, all notes should be cited in text in sequential order (e.g., see Note 1). Please check.	
AU2	Please provide volume number for reference 8.	