

UniHI 4: new tools for query, analysis and visualization of the human protein–protein interactome

Gautam Chaurasia^{1,2}, Soniya Malhotra², Jenny Russ¹, Sigrid Schnoegl², Christian Hänig², Erich E. Wanker² and Matthias E. Futschik^{1,3,*}

¹Institute for Theoretical Biology, Charité, Humboldt-University, ²Max-Delbrück-Center for Molecular Medicine, Berlin, Germany and ³Center for Molecular and Structural Biomedicine, University of Algarve, Faro, Portugal

Received September 15, 2008; Revised October 13, 2008; Accepted October 14, 2008

ABSTRACT

Human protein interaction maps have become important tools of biomedical research for the elucidation of molecular mechanisms and the identification of new modulators of disease processes. The Unified Human Interactome database (UniHI, <http://www.unihi.org>) provides researchers with a comprehensive platform to query and access human protein–protein interaction (PPI) data. Since its first release, UniHI has considerably increased in size. The latest update of UniHI includes over 250 000 interactions between ~22 300 unique proteins collected from 14 major PPI sources. However, this wealth of data also poses new challenges for researchers due to the complexity of interaction networks retrieved from the database. We therefore developed several new tools to query, analyze and visualize human PPI networks. Most importantly, UniHI allows now the construction of tissue-specific interaction networks and focused querying of canonical pathways. This will enable researchers to target their analysis and to prioritize candidate proteins for follow-up studies.

INTRODUCTION

Human protein interaction maps play an increasingly important role in biomedical research. They have been shown to be highly valuable in the study of a variety of human diseases and signaling pathways (1–3). The rising popularity of network analyses is reflected in the large number of independently constructed human PPI maps based on experimental and computational approaches. However, these maps generally have limited overlap and frequently lack cross-references (4). To obtain

comprehensive interaction data, researchers were required to perform time-consuming queries of different databases with subsequent error-prone matching of obtained identifiers.

UniHI has been developed to overcome these difficulties (5). It has integrated separate PPI resources to provide a comprehensive platform for querying the human interactome. UniHI is not intended to replace single databases, but to offer a convenient single portal access to human protein interaction data for the biomedical research community. Notably, it allows the identification of network topologies which would not be detectable if PPI resources were examined separately.

The size of UniHI has remarkably increased with more than 250 000 human PPIs currently included. However, this amount of data has its challenges. Even searches with a small number of query proteins can lead to large, highly connected—and often unstructured—networks (frequently referred to as ‘hairballs’). For efficient follow-up analysis, new tools for navigation within the network and prioritization of targets are necessary. Also, flexible visualization is a crucial prerequisite for the display and evaluation of network structures. To meet these challenges, we have implemented several new tools in UniHI for query, analysis and visualization of interaction networks. Beyond its original role as a direct entry gate to the human interactome, UniHI will serve as an integrative platform for the exploration and utilization of human PPI data.

UPDATES AND EXTENSIONS

In our aim to continually provide the most comprehensive human PPI dataset, UniHI has been substantially extended by the inclusion of interactions from two additional major protein interaction databases, i.e. IntAct and BioGRID (6,7). Currently, UniHI includes over 250 000 interactions between more than 22 000 unique proteins

*To whom correspondence should be addressed. Tel: +35 1 914 199 509; Email: mfutschik@ualg.pt

Table 1. PPI datasets currently integrated in UniHI

Dataset	Proteins	Interactions	Method	Reference	Database location
MDC-Y2H	1703	3186	Y2H screen	Stelzl <i>et al.</i> (8)	www.mdc-berlin.de/neuroprot
CCSB-Y2H	1549	2754	Y2H screen	Rual <i>et al.</i> (9)	vidal.dfci.harvard.edu (flat file only)
HPRD-BIN	8788	32776	Literature	Mishra <i>et al.</i> (10)	www.hprd.org
HPRD-COMP	1969	8107	Literature	Mishra <i>et al.</i> (10)	www.hprd.org
DIP	1085	1397	Literature	Salwinski <i>et al.</i> (11)	dip.doe-mbi.ucla.edu
BIOGRID	7953	24624	Literature	Breitkreutz <i>et al.</i> (7)	www.thebiogrid.org
INTACT	7273	19404	Literature	Kerrien <i>et al.</i> (6)	www.ebi.ac.uk/intact
BIND	5286	7394	Literature	Bader <i>et al.</i> (12)	www.bind.ca
REACTOME	1554	37332	Literature	Joshi-Tope <i>et al.</i> (13)	www.reactome.org
COCIT	3737	6580	Text mining	Ramani <i>et al.</i> (14)	Bioinformatics.icmb.utexas.edu/idserve/
ORTHO	6225	71466	Orthology	Lehner and Fraser (15)	www.sanger.ac.uk/PostGenomics/signaltransduction/ interactionmap
HOMOMINT	4127	10174	Orthology	Persico <i>et al.</i> (16)	mint.bio.uniroma2.it
OPHID	4785	24991	Orthology	Brown and Jurisica (17)	ophid.utoronto.ca

Number of proteins and interactions in each dataset as well as construction approaches and references.

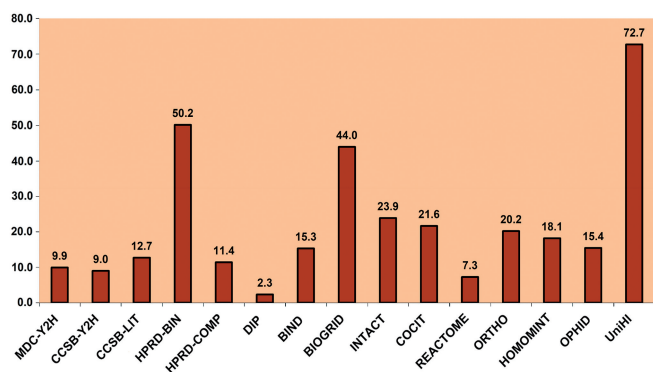


Figure 1. Coverage of the functionally annotated human genome by PPI resources. For annotation, Gene Ontology was utilized. Coverage rates were derived after mapping of proteins to corresponding Entrez Gene IDs. Notably, the coverage of UniHI is considerably larger than of the individual PPI resources.

from 14 distinct sources, establishing it as the largest catalog for human PPIs worldwide (Table 1 and Figure 1). Although the overlap between different PPI resources included in UniHI has increased, they are still strongly divergent. Only a relatively small fraction of $\sim 19\%$ can be found in two or more interaction maps, underlining the continuing need for integrative platforms such as UniHI (Supplementary Tables S1 and S2; Figures S1A and B).

In this study, we optimized the query interface which allows the simultaneous search for interaction partners of several proteins in a network-oriented manner. To facilitate its application, the list of possible protein identifiers has been expanded to include gene symbol, Entrez Gene, Uniprot, NCBI Geneinfo, Ensembl, Biogrid and HPRD IDs. Notably, these identifiers can now also be used for direct hyperlinks to UniHI.

As in previous versions, special care was taken to indicate the origin of the interaction data to the user. Besides links to the original resource, a variety of information regarding the interacting proteins is given. Additionally, updates were carried out for measures of co-annotation and co-expression, which are not only important for the

interpretation of single interactions, but also for higher network structures (18).

NEW INTERACTIVE VISUALIZATION TOOL

Visualization of the retrieved interaction networks remains to be crucial for the evaluation of query results. The complexity of retrieved networks, however, requires highly flexible graphical tools. While the former versions of UniHI only provided non-interactive display, the present update includes interactive graphical tools which offer many attractive features for rapid analysis and adjustment of the extracted information.

For example, nodes (i.e. proteins) can be anchored or hidden allowing filtering of the network and manual adjustment of the layout. Also, information about proteins and interactions can now be accessed directly in the network graphics, thereby avoiding cumbersome comparisons with the textual output. The display can be restricted to direct interactions between query proteins or extended to include bridging proteins.

For quality control, users can specify the PPI resource, from which interactions should be retrieved. This allows, e.g. the exclusion of less validated mapping approaches such as computational prediction. As additional criteria, interactions can be filtered based on a minimum number of PubMed references in which they have been reported.

UNIHI SCANNER

Pathway-focused interaction networks

Pathway information can provide highly useful clues about the functions and dynamics of interactions. Especially for the elucidation of local network structures, knowledge of interrelated pathways can be of crucial importance. We therefore constructed a new tool called UniHI Pathway Scanner (Figure 2A and B). It provides the possibility to examine the intersection of canonical pathways from KEGG with the extracted networks (19). In this way, it enables researchers to detect possible

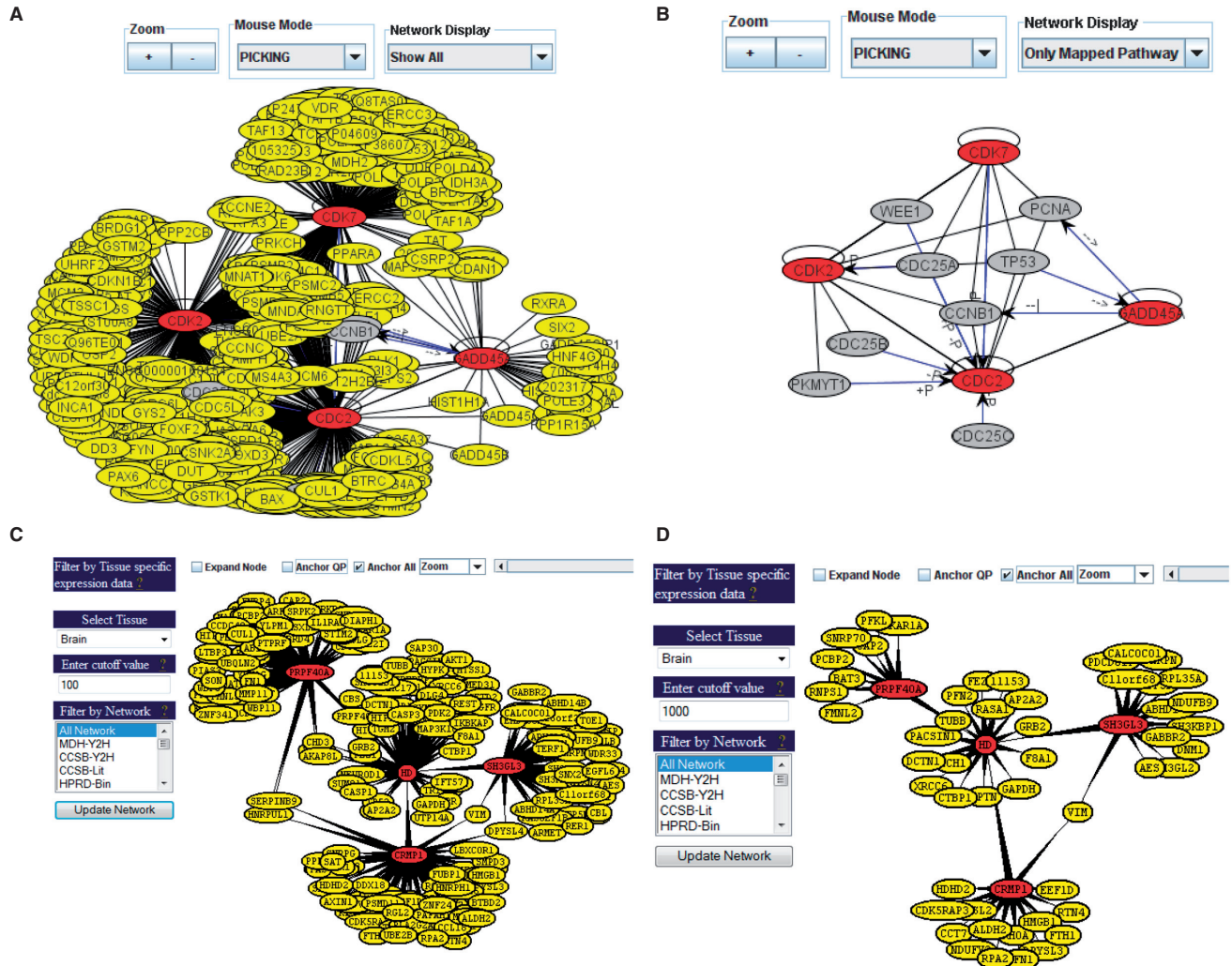


Figure 2. Graphical representation and analysis of PPI networks using UniHI Scanner and UniHI Express. (A) Display of the interaction partners (yellow or gray) of the query proteins (red) GADD45, CDK1, CDK2 and CDK7. Gray nodes represent proteins included in the KEGG ‘cell cycle’ pathway. UniHI Scanner allows a focused display of the intersection between the retrieved PPI network and the pathway (B). Additional information is given regarding the type of interaction (e.g. phosphorylation (+P), dephosphorylation (–P), activation (–>+) or inhibition (–|)), facilitating the assessment of the retrieved interactions. (C) Construction of tissue-specific networks by UniHI Express: Interaction partners (yellow) of HD, CRMP1, SH3GL3 and PRPF40A (red) which have minimal expression values in brain tissue are displayed. The selection of larger expression thresholds can lead to a considerable reduction of the network, allowing the prioritization of proteins and interactions for follow-up studies (D).

modifiers of pathways as well as proteins involved in the cross-talk between different pathways. Users can switch between the graphical display of the complete network and the intersection with selected pathways. UniHI Scanner does not only show the proteins included in the pathway but also the KEGG annotation of the interactions (e.g. phosphorylation, activation or inhibition) between nodes (see also Supplementary Materials). We expect that this will be a highly attractive feature for the large community of researchers working in cell signaling.

UNIHI EXPRESS

Tissue-specific PPI networks

Protein interactions are known to be highly dynamic and to strongly depend on many biological factors. Current

protein interaction maps, however, only give a static view of the human interactome. Experimentally validated protein interactions are generally identified under a variety of conditions in numerous cell and tissue types. Current interaction maps do not fully reflect physiological states, because only a selection of proteins is present in a cell at a certain point in time. Biomedical research, however, is usually focused on specific tissues involved in pathogenesis. Addressing this need, we developed and implemented UniHI Express as a new tool in our database (Figure 2C and D). It allows the filtering of PPIs based on gene expression in selected tissues and thus enables the construction of tissue-specific networks (see also Supplementary Materials). First preliminary studies show that the use of UniHI Express can be highly efficient to prioritize interactions. The expression data were derived from Gene Expression Atlas and merged to a number of main tissue

types to facilitate their utilization (20). UniHI Express represents a first step towards a dynamic representation of the human interactome.

CONCLUSIONS AND FUTURE DIRECTIONS

Human interaction maps are rapidly increasing in size and have proven to be highly valuable for the study of human health and disease. UniHI will continue to extend its scope by the incorporation of newly available PPI resources and to consolidate the frequently divergent data. In this context, we like to invite other data providers and researchers to participate in the UniHI project. Clearly, the wealth of interaction data poses new challenges in follow-up analysis. The new tools included in UniHI allow researchers a more rapid inspection and prioritization of extracted interactions. Tissue-specific networks can help to focus on biologically relevant interactions, whereas use of pathway information can give important hints about functional modules of interacting proteins. We hope that these and further extensions of UniHI will support scientists in the exploration and utilization of the human interactome.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank both referees for their constructive suggestions and Kris Gunsalus for insightful discussions regarding the implementation of UniHI.

FUNDING

Deutsche Forschungsgemeinschaft (grant SFB 618-sub-project A5). Funding for open access charge: SFB 618 grant of the Deutsche Forschungsgemeinschaft (DFG).

Conflict of interest statement. None declared.

REFERENCES

- Goehler,H., Lalowski,M., Stelzl,U., Waelter,S., Stroedicke,M., Worm,U., Droege,A., Lindenberg,K.S., Knoblich,M., Haenig,C. *et al.* (2004) A protein interaction network links GIT1, an enhancer of Huntingtin aggregation, to Huntington's disease. *Mol. Cell*, **15**, 853–865.
- Liu,M., Liberzon,A., Kong,S.W., Lai,W.R., Park,P.J., Kohane,I.S. and Kasif,S. (2007) Network-based analysis of affected biological processes in type 2 diabetes models. *PLoS Genetics*, **3**, e96.
- Lage,K., Karlberg,E.O., Störing,Z.M., Ólason,P., Pedersen,A.G., Rigina,O., Hinsby,A.M., Tümer,Z., Pociot,F., Tommerup,N. *et al.* (2007) A human phenome–interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.*, **25**, 309–316.
- Futschik,M.E., Chaurasia,G. and Herzel,H. (2007) Comparison of human protein–protein interaction maps. *Bioinformatics*, **23**, 605–611.
- Chaurasia,G., Iqbal,Y., Hänig,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–D594.
- Kerrien,S., Alam-Faruque,Y., Aranda,B., Bancarz,I., Bridge,A., Derow,C., Dimmer,E., Feuermann,M., Friedrichsen,A., Huntley,R. *et al.* (2006) IntAct – open source resource for molecular interaction data. *Nucleic Acids Res.*, **34**, D561–D565.
- Breitkreutz,B.J., Stark,C., Reguly,T., Boucher,L., Breitkreutz,A., Livstone,M., Oughtred,R., Lackner,D.H., Bähler,J., Wood,V. *et al.* (2008) The BioGRID interaction database: 2008 update. *Nucleic Acids Res.*, **36**, D637–D640.
- Stelzl,U., Worm,U., Lalowski,M., Haenig,C., Brembeck,F.H., Goehler,H., Stroedicke,M., Zenkner,M., Schoenherr,A., Koeppen,S. *et al.* (2005) A human protein–protein interaction network: a resource for annotating the proteome. *Cell*, **122**, 957–968.
- 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. *et al.* (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, **437**, 1173–1178.
- Mishra,G.R., Suresh,M., Kumaran,K., Kannabiran,N., Suresh,S., Bala,P., Shivakumar,K., Anuradha,N., Reddy,R., Raghavan,M.T. *et al.* (2006) Human protein reference database—2006 update. *Nucleic Acids Res.*, **34**, D411–D414.
- 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 Acid Res.*, **32**, D449–D451.
- Bader,G.D., Donaldson,I., Wolting,C., Ouellette,B.F., Pawson,T. and Hogue,C.W. (2001) BIND—the biomolecular interaction network database. *Nucleic Acids Res.*, **29**, 242–245.
- Joshi-Tope,G., Gillespie,M., Vastrik,I., D'Eustachio,P., Schmidt,E., de Bono,B., Jassal,B., Gopinath,G.R., Wu,G.R., Matthews,L. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D443.
- Ramani,A.K., Bunesco,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.
- Lehner,B. and Fraser,A.G. (2004) A first-draft human protein–interaction map. *Genome Biol.*, **5**, R63.
- 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.
- Brown,K.R. and Jurisica,I. (2005) Online predicted human interaction database. *Bioinformatics*, **21**, 2076–2082.
- Futschik,ME., Chaurasia,G., Tschaut,A., Russ,J. and Babu,MM. (2007) Functional and transcriptional coherency of modules in the human protein interaction network. *J. Integr. Bioinform.*, **4**, 76.
- Kanehisa,M., Araki,M., Goto,S., Hattori,M., Hirakawa,M., Itoh,M., Katayama,T., Kawashima,S., Okuda,S., Tokimatsu,T. and Yamanishi,Y. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, D480–D484.
- Su,AI., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.