

UniHI 7: an enhanced database for retrieval and interactive analysis of human molecular interaction networks

Ravi Kiran Reddy Kalathur¹, José Pedro Pinto¹, Miguel A. Hernández-Prieto¹,
Rui S.R. Machado¹, Dulce Almeida¹, Gautam Chaurasia² and Matthias E. Futschik^{1,*}

¹Centre for Molecular and Structural Biomedicine, University of Algarve, Faro, Portugal and ²Institute for Theoretical Biology, Charité, Humboldt-University, Berlin, Germany

Received September 15, 2013; Revised October 17, 2013; Accepted October 18, 2013

ABSTRACT

Unified Human Interactome (UniHI) (<http://www.unihi.org>) is a database for retrieval, analysis and visualization of human molecular interaction networks. Its primary aim is to provide a comprehensive and easy-to-use platform for network-based investigations to a wide community of researchers in biology and medicine. Here, we describe a major update (version 7) of the database previously featured in NAR Database Issue. UniHI 7 currently includes almost 350 000 molecular interactions between genes, proteins and drugs, as well as numerous other types of data such as gene expression and functional annotation. Multiple options for interactive filtering and highlighting of proteins can be employed to obtain more reliable and specific network structures. Expression and other genomic data can be uploaded by the user to examine local network structures. Additional built-in tools enable ready identification of known drug targets, as well as of biological processes, phenotypes and pathways enriched with network proteins. A distinctive feature of UniHI 7 is its user-friendly interface designed to be utilized in an intuitive manner, enabling researchers less acquainted with network analysis to perform state-of-the-art network-based investigations.

INTRODUCTION

The study of molecular systems and networks is now a major field in biology and medicine. The goals of network-based investigations range from prioritization of candidate genes to determination of complex molecular mechanisms underlying a disease or a biological process (1,2). An essential prerequisite for these investigations is the availability of resources for molecular interactions in

model organisms and humans. To address this need, various databases have been established in recent years (3). Especially for protein–protein interactions in humans, many initiatives and research groups have contributed large sets of data derived from the literature, high-throughput methods or computational prediction (4–15). In parallel, a wide range of dedicated programs for network analyses have also been developed (16–18).

However, it is a common experience that current resources and tools pose a considerable challenge to users, especially to researchers less acquainted with concepts in network biology. Frequently, users have to download, map, compile and integrate distinct data types to conduct network-based investigations. These activities require extensive knowledge of data processing and management. Thus, a salient ‘bottleneck’ exists for many interested researchers between the wealth of available molecular interaction data and their utilization.

This observation motivated us to develop a new version of the Unified Human Interactome (UniHI) database for the retrieval, analysis and visualization of human molecular interaction networks: UniHI 7. We provide a platform that enables (i) retrieval of an integrated set of interactions from the major resources, (ii) intuitive use of tools for network-based investigations and (iii) easy utilization of complementary data and information for analysis, evaluation and visualization of retrieved networks.

SCOPE OF UNIHI 7

UniHI 7 integrates ~350 000 molecular interactions for more than 30 000 human proteins. It is based on a complete re-implementation of previous versions of UniHI, with widely extended scope and functionality. Besides protein–protein interactions from 12 different resources [including HPRD (4), BioGrid (5), IntAct (6), DIP (7), BIND (8) and Reactome (9) databases; as well as four interaction maps produced by computational predictions (10–13) and two high-throughput yeast-2-hybrid (Y2H)

*To whom correspondence should be addressed. Tel: +351 289 244 403; Fax: +351 289 818 419; Email: mfutschik@ualg.pt

screens (14,15)], UniHI 7 also comprises curated transcriptional regulatory interactions from three complementary databases TRANSFAC (19), miRTarBase (20) and HTRIdb (21). In addition to these interactions, we also integrated drug target information from DrugBank (22) that can be mapped on the interaction network. Detailed description regarding the incorporated resources can be found on the UniHI 7 web-page and in the Supplementary Materials (Supplementary Table S1). Whereas former UniHI versions can primarily be regarded as integrated databases for human protein–protein interactions (23,24), additional strengths of UniHI 7 lie not only in the integration of regulatory interactions but also in its interactive analysis and visualization tools for molecular networks. Although there are other databases with integrated molecular interaction data (25–34), UniHI provides a distinct and unique set of features ranging from simple filtering options to advanced network analysis tools (Supplementary Materials and Supplementary Table S2). The main application of UniHI 7 is the retrieval and examination of small to medium-sized local networks. It is ideally suited for researchers, who want to explore the molecular context of a single protein or a select set of related proteins using a network-orientated approach.

SEARCHING FOR INTERACTIONS IN UNIHI 7

The UniHI 7 database can be queried for molecular interactions of single or multiple human proteins. Various identifiers such as gene symbol, Entrez Gene, Uniprot and Ensembl IDs can serve as input. It is also possible to input gene and protein identifiers from the model organisms yeast (*Saccharomyces cerevisiae*), worm (*Caenorhabditis elegans*), fly (*Drosophila melanogaster*) or mouse (*Mus musculus*), which will be automatically mapped to human orthologs. This feature is convenient for researchers, who work with these major model organisms and want to interrogate related human molecular networks. In total, 2977 yeast, 6922 worm, 7998 fly and 15694 mouse genes were mapped to human orthologs included in UniHI 7 using information from the HGNC database (35). As identifiers for model organisms, Entrez Gene IDs can be generally used, as well as systematic names for yeast, WormBase IDs or gene symbols for worm, FlyBase IDs or gene symbols for fly and MGI identifiers or gene symbols for mouse. The list of identifiers and any other data uploaded by the user are only stored during the active session and are accessible only to the user.

PARALLEL PRESENTATION OF PROTEINS, INTERACTIONS AND NETWORK

Whereas in previous UniHI versions, queried proteins, retrieved interactions and resulting networks were presented in sequential order, UniHI 7 displays now on four web-pages in parallel: ‘Proteins’, ‘Physical Interactions’, ‘Regulatory Interactions’ and ‘Network’. This display scheme enables users to readily switch

between the different types of displayed information (Figure 1). The ‘Proteins’ page provides the list of proteins in the UniHI 7, matching the input and the names of the databases or resources in which each protein is included. In addition, hyperlinks to Entrez Gene, Uniprot, RefSeq, OMIM IDs and KEGG (if available) are given. On the ‘Physical Interactions’ and ‘Regulatory Interactions’ pages, the set of detected interactions is shown with various additional information regarding their source, evidence, type and functional annotation. A crucial feature is that interactions can easily be traced back to the original resources and publications, and thus can be critically assessed by users. In addition, all the interactions displayed on these two pages can be downloaded as simple tables, which can used as input for other computational tools.

The retrieved interactions are displayed as a network on the ‘Network’ page. For network visualization, we utilized the recently developed Cytoscape Web, which is a client-side application implemented in Flex/ActionScript and modeled after the popular Cytoscape software (36). To prevent that the visualization tool becomes unresponsive, certain automatic layout and filtering procedures are implemented for larger networks (Supplementary Materials).

Information about proteins and interactions can be interactively explored in the network graphics, avoiding cumbersome comparison with the textual output. For instance, clicking on the network nodes provides information about the corresponding protein and links to other resources such as GeneCards (37) and GeneMania (38) for follow-up study. The displayed network can be exported as simple tab-delimited text, and as image, either as a PNG or PDF file.

INTEGRATED TOOLS FOR NETWORK ANALYSIS

Molecular networks are inherently difficult to analyze and interpret. In fact, the sheer number of retrieved interactions for well-studied proteins (including many kinases and receptors) can be overwhelming for users (Figure 2a). To help users with these challenges, we have implemented several tools for filtering and inspecting networks, as well as for mapping and utilizing complementary data and information. The application of these tools can be customized to different research objectives and can assist in the elucidation of network structure and prioritizing candidate proteins for follow-up studies.

FILTERING OF INTERACTIONS AND UTILIZATION OF GENOMIC DATA

Filtering of interactions can be carried out on resource, published evidence (i.e. number of PubMed references), scale of experiment (i.e. small-scale or large-scale), type of derivation (i.e. literature, computational prediction or Y2H screens), connectivity (i.e. direct or indirect) and interaction (i.e. binary or complex) (Figure 2b). These filtering options can be tailored to produce more reliable and specific networks, e.g. include only interactions reported in multiple publications.

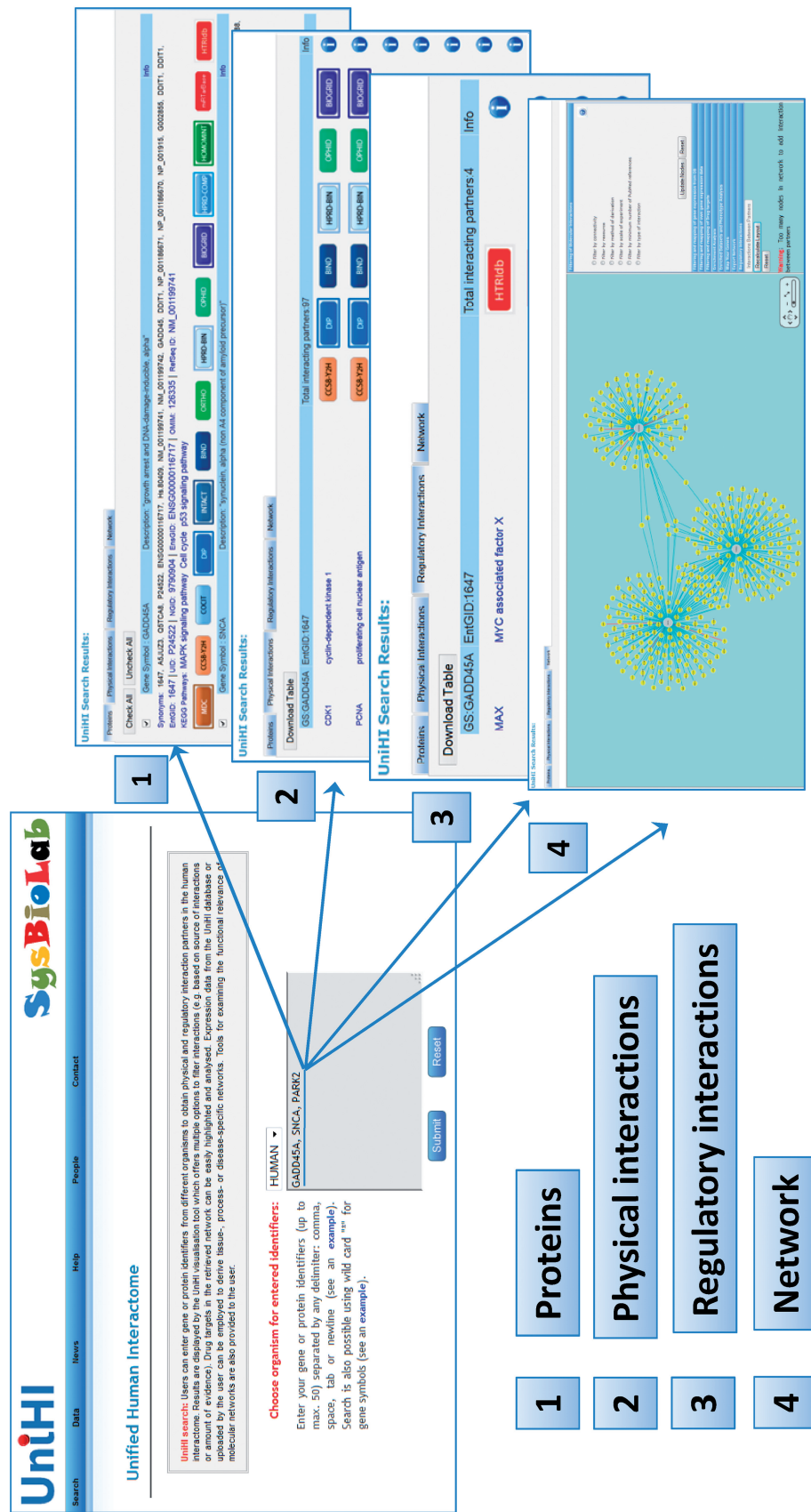


Figure 1. Search and result pages in UniHI 7. After performing initial search on the input page, the user is presented with four result pages providing information about detected proteins, physical and regulatory interactions, as well as a graphic network. The user can easily switch between the four pages by clicking on the relevant tab.

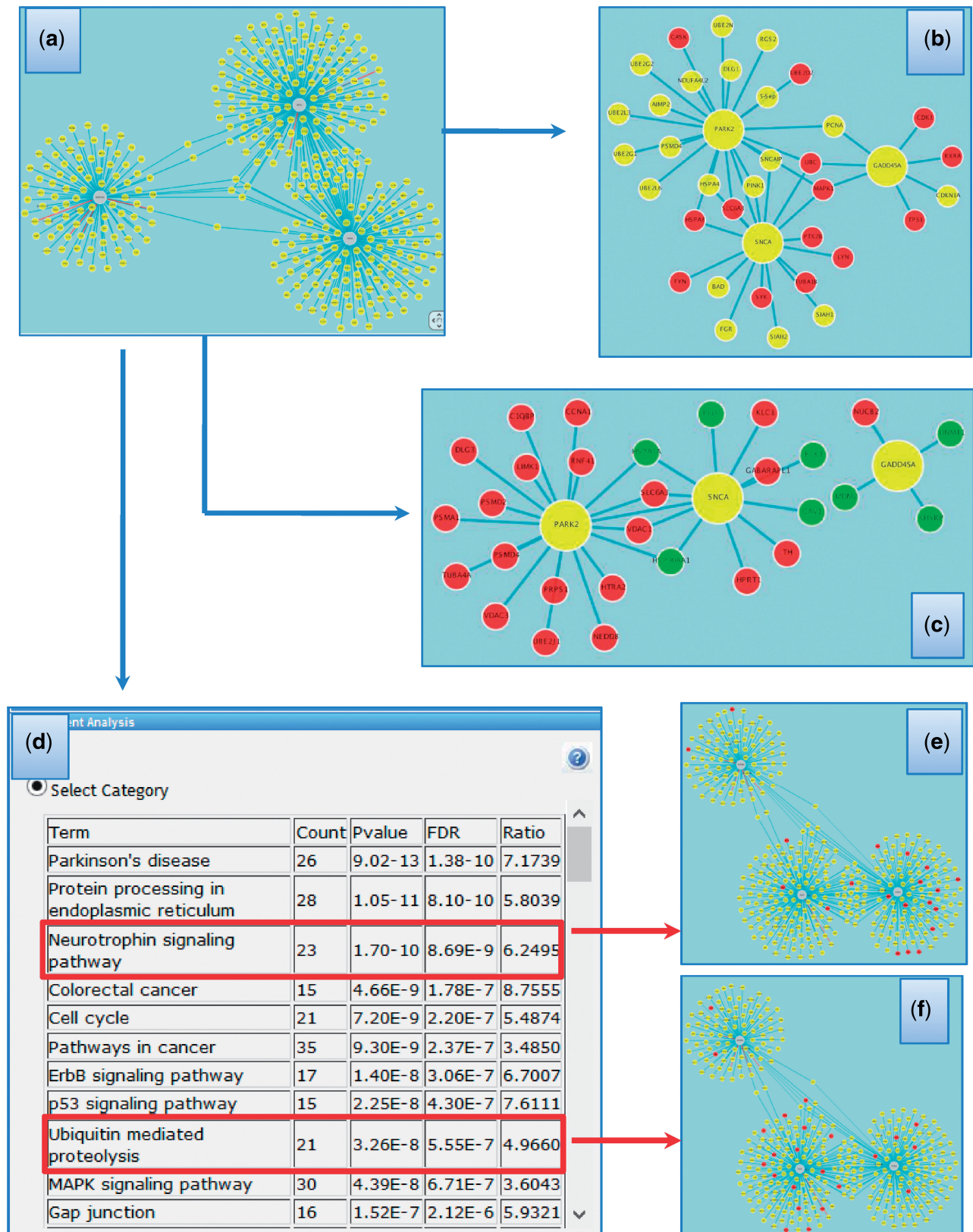


Figure 2. Network analysis and visualization tools: **(a)** original network of the query proteins (GADD45A, SNCA and PARK2); **(b)** network after filtering based on published evidence (i.e. number of PubMed references ≥ 3) and with known drug targets highlighted in red; **(c)** network after filtering with gene expression data (GSE20186) (39). Nodes corresponding to genes with a \log_2 fold change $\geq +0.2$ are displayed in red and those ≤ -0.2 are shown in green; **(d)** significantly enriched KEGG pathways listed in a table; **(e)** filtered network with proteins linked to the selected term 'Neurotrophin signaling pathway' highlighted in red; and **(f)** filtered network with proteins linked to 'Ubiquitin mediated proteolysis' highlighted in red.

UniHI 7 also stores gene expression in 19 different human tissue types derived from the SymAtlas (40,41). Users can apply this data to highlight or exclude proteins (based on a chosen threshold level) to derive tissue-specific networks. In addition to using gene expression data stored in UniHI 7, users can upload their own expression data to filter and examine human molecular interaction networks. This feature can be applied to detect network proteins, which have distinct expression patterns related to physiological processes or diseases. Two different types of expression data can be used: absolute gene expression, i.e. positive values for transcript levels such as detected by Affymetrix GeneChips or RNA-Seq and differential gene expression, i.e. changes in expression derived from two-color arrays or by subtraction of absolute expression measurements. Thresholds for differential expression data can be set as maximum *P*-values or minimum fold changes (Figure 2c).

In addition, gene lists, for example, derived from RNAi screens or high-throughput assays, can be uploaded and utilized for annotation and filtering of interaction networks. Together with its capacity to overlay expression data, this option makes UniHI 7 an efficient platform for network-based analyses in the ‘post-genomic’ era.

EXPLORING THE NETWORK: DRUG TARGETS, GENE ANNOTATION, PHENOTYPES AND DISEASES

Small molecules (drugs) can influence activity of single proteins, alter pathogenic mechanisms and are of crucial importance in numerous therapeutic interventions. To facilitate identification of known drug targets in the retrieved networks, UniHI 7 provides relevant highlighting and filtering options (Figure 2b) as well as information about the drugs and their mechanisms of action. From the DrugBank database, information for 4203 drugs targeting 2139 annotated proteins were altogether, imported into UniHI 7 (22) (Supplementary Materials).

The functional relevance of networks is inherently difficult to assess. Hence, we have implemented a user-friendly integrated tool, which carries out enrichment analyses for molecular functions, biological processes, cellular location [as defined by Gene Ontology (42)], protein families [as defined by Pfam (43)] and pathways [as defined by KEGG (44)] of network proteins (Figure 2d). The significance of overrepresentation of network proteins in a Gene Ontology category, Pfam protein family or KEGG pathway is calculated using the hypergeometric test (which is equivalent to the one-tailed Fisher’s exact test) with the terms from human genome as background distribution. For significant terms (i.e. GO categories, Pfam families or KEGG pathways), the number of included network proteins, the *P*-value and the false discovery rate for enrichment are displayed in a table. The associated proteins can easily be highlighted in the network graphics by clicking on the corresponding term (Figure 2e and f).

Finally, phenotype information can be assessed for network proteins in the new version of UniHI. For

this purpose, we have integrated gene–phenotype associations, curated in Mouse Genome Database (45) and mapped to their human orthologs, for several major phenotypes such as cardiovascular system or embryogenesis. In addition, we have collected genes linked to aging in humans from the GenAge database (46), genes associated with cancer from the Cancer Gene Census catalogue (47) and genes linked to human diseases from the OMIM database (48) (Supplementary Materials). Similarly to the type of analysis describe earlier, the UniHI user can assess whether phenotypic associations are overrepresented among network proteins, and highlight the associated proteins within the network. A help page with detailed description of the different tools and sample outputs for typical analyses is available on the UniHI 7 webpage.

IMPLEMENTATION

The architecture of UniHI 7 comprises a database and an application layer. The database layer is implemented using MySQL, an open source SQL relational database management system. The application layer is implemented using a J2EE architecture including, e.g. JDBC to connect to the back-end database, DAO for interacting with the database and accessing data and JavaServerPages to generate web pages. Data retrieval from the database is performed using the Hibernate library. The communication between client and the application layer is through a Tomcat server. To perform enrichment analyses, UniHI 7 connects via Rserve (<http://www.rforge.net/Rserve/>) to the R/Bioconductor software (17). Matching of gene and protein identifiers was carried out using information from HGNC (35) and applying the g:Convert web tool (49). UniHI 7 performs best with small- to medium-sized networks (with up to several hundred interactions). For larger networks, the visualization and analysis becomes increasingly time consuming.

CONCLUSIONS

UniHI 7 is intended to serve as a bridge between resources for interaction data and more advanced software. It provides a user-friendly web-based platform to study networks underlying molecular mechanisms in human health and disease. Customization allows users to: (i) adjust the set of included interactions, (ii) overlay retrieved sub-networks with other types of data, (iii) inspect networks for relevant genes, (iv) determine potential network functions and (v) associations with phenotypes. We hope that UniHI 7 will considerably facilitate network-orientated investigations for many researchers, especially for those who are new to this field.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Christian Lopes and Gary Bader for their help with the integration of Cytoscape Web into UniHI 7, Paulo Martel for his support in the configuration of the server, Trudi Semeniuk for critical reading of the manuscript and both referees for their constructive comments.

FUNDING

Funding for the presented work was provided by Portuguese *Fundação para a Ciência e a Tecnologia (FCT)* [PTDC/BIA-GEN/116519/2010 and IBB/CBME, LA] and *CHDI* Foundation [A-2666]. R.K. is recipient of a FCT scholarship [SFRH/BPD/70718/2010]. D.A. was supported by the FCT [PTDC/BIA-BCM/117975/2010]. Funding for open access charge: Portuguese *Fundação para a Ciência e a Tecnologia* [PTDC/BIA-GEN/116519/2010].

Conflict of interest statement. None declared.

REFERENCES

- Magger, O., Waldman, Y.Y., Ruppin, E. and Sharan, R. (2012) Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS Comput. Biol.*, **8**, e1002690.
- Zhang, B., Gaiteri, C., Bodea, L.G., Wang, Z., McElwee, J., Podtelezchnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R. *et al.* (2013) Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*, **153**, 707–720.
- Bader, G.D., Cary, M.P. and Sander, C. (2006) Pathguide: a pathway resource list. *Nucleic Acids Res.*, **34**, D504–D506.
- Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A. *et al.* (2009) Human Protein Reference Database—2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
- Chatr-Aryamontri, A., Breitkreutz, B.J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., O'Donnell, L. *et al.* (2013) The BioGRID interaction database: 2013 update. *Nucleic Acids Res.*, **41**, D816–D823.
- Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U. *et al.* (2012) The IntAct molecular interaction database in 2012. *Nucleic Acids Res.*, **40**, D841–D846.
- 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–D451.
- Isserlin, R., El-Badrawi, R.A. and Bader, G.D. (2011) The Biomolecular Interaction Network Database in PSI-MI 2.5. *Database (Oxford)*, **2011**, baq037.
- Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **39**, D691–D697.
- 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.
- Lehner, B. and Fraser, A.G. (2004) A first-draft human protein-interaction map. *Genome Biol.*, **5**, R63.
- Brown, K.R. and Jurisica, I. (2005) Online predicted human interaction database. *Bioinformatics*, **21**, 2076–2082.
- 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.
- Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppe, 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.
- Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L. and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
- Hu, Z., Chang, Y.C., Wang, Y., Huang, C.L., Liu, Y., Tian, F., Granger, B. and Delisi, C. (2013) VisANT 4.0: integrative network platform to connect genes, drugs, diseases and therapies. *Nucleic Acids Res.*, **41**, W225–W231.
- Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K. *et al.* (2006) TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
- Hsu, S.D., Lin, F.M., Wu, W.Y., Liang, C., Huang, W.C., Chan, W.L., Tsai, W.T., Chen, G.Z., Lee, C.J., Chiu, C.M. *et al.* (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **39**, D163–D169.
- Bovolenta, L.A., Acencio, M.L. and Lemke, N. (2012) HTRIDb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics*, **13**, 405.
- Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V. *et al.* (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.*, **39**, D1035–D1041.
- 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–D594.
- 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–D660.
- Kamburov, A., Stelzl, U., Lehrach, H. and Herwig, R. (2013) The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.*, **41**, D793–D800.
- Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G.D. and Sander, C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.
- Turner, B., Razick, S., Turinsky, A.L., Vlasblom, J., Crowdy, E.K., Cho, E., Morrison, K., Donaldson, I.M. and Wodak, S.J. (2010) iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database (Oxford)*, **2010**, baq023.
- del-Toro, N., Dumousseau, M., Orchard, S., Jimenez, R.C., Galeota, E., Launay, G., Goll, J., Breuer, K., Ono, K., Salwinski, L. *et al.* (2013) A new reference implementation of the PSICQUIC web service. *Nucleic Acids Res.*, **41**, W601–W606.
- Tarcea, V.G., Weymouth, T., Ade, A., Bookvich, A., Gao, J., Mahavisno, V., Wright, Z., Chapman, A., Jayapandian, M., Ozgur, A. *et al.* (2009) Michigan molecular interactions r2: from interacting proteins to pathways. *Nucleic Acids Res.*, **37**, D642–D646.
- Schaefer, M.H., Fontaine, J.F., Vinayagam, A., Porras, P., Wanker, E.E. and Andrade-Navarro, M.A. (2012) HIPPIE: Integrating protein interaction networks with experiment based quality scores. *PLoS One*, **7**, e31826.
- Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C. *et al.* (2013) STRING v9.1: protein-protein interaction

- networks, with increased coverage and integration. *Nucleic Acids Res.*, **41**, D808–D815.
32. Kozhenkov,S., Dubinina,Y., Sedova,M., Gupta,A., Ponomarenko,J. and Baitaluk,M. (2010) BiologicalNetworks 2.0—an integrative view of genome biology data. *BMC Bioinformatics*, **11**, 610.
 33. Chen,J.Y., Mamidipalli,S. and Huan,T. (2009) HAPPI: an online database of comprehensive human annotated and predicted protein interactions. *BMC Genomics*, **10**(Suppl. 1), S16.
 34. Prieto,C. and De Las Rivas,J. (2006) APID: Agile Protein Interaction DataAnalyzer. *Nucleic Acids Res.*, **34**, W298–W302.
 35. Gray,K.A., Daugherty,L.C., Gordon,S.M., Seal,R.L., Wright,M.W. and Bruford,E.A. (2013) Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res.*, **41**, D545–D552.
 36. Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
 37. Safran,M., Dalah,I., Alexander,J., Rosen,N., Iny Stein,T., Shmoish,M., Nativ,N., Bahir,I., Doniger,T., Krug,H. *et al.* (2010) GeneCards Version 3: the human gene integrator. *Database (Oxford)*, **2010**, baq020.
 38. Warde-Farley,D., Donaldson,S.L., Comes,O., Zuberi,K., Badrawi,R., Chao,P., Franz,M., Grouios,C., Kazi,F., Lopes,C.T. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, **38**, W214–W220.
 39. Zheng,B., Liao,Z., Locascio,J.J., Lesniak,K.A., Roderick,S.S., Watt,M.L., Eklund,A.C., Zhang-James,Y., Kim,P.D., Hauser,M.A. *et al.* (2010) PGC-1alpha, a potential therapeutic target for early intervention in Parkinson’s disease. *Sci. Transl. Med.*, **2**, 52ra73.
 40. Russ,J. and Futschik,M.E. (2010) Comparison and consolidation of microarray data sets of human tissue expression. *BMC Genomics*, **11**, 305.
 41. Su,A.I., 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. U.S.A.*, **101**, 6062–6067.
 42. 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. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
 43. Punta,M., Coggill,P.C., Eberhardt,R.Y., Mistry,J., Tate,J., Boursnell,C., Pang,N., Forslund,K., Ceric,G., Clements,J. *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res.*, **40**, D290–D301.
 44. Kanehisa,M., Araki,M., Goto,S., Hattori,M., Hirakawa,M., Itoh,M., Katayama,T., Kawashima,S., Okuda,S., Tokimatsu,T. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, D480–D484.
 45. Bult,C.J., Eppig,J.T., Blake,J.A., Kadin,J.A. and Richardson,J.E. (2013) The mouse genome database: genotypes, phenotypes, and models of human disease. *Nucleic Acids Res.*, **41**, D885–D891.
 46. Tacutu,R., Craig,T., Budovsky,A., Wuttke,D., Lehmann,G., Taranukha,D., Costa,J., Fraifeld,V.E. and de Magalhaes,J.P. (2013) Human Ageing Genomic Resources: integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Res.*, **41**, D1027–D1033.
 47. Futreal,P.A., Coin,L., Marshall,M., Down,T., Hubbard,T., Wooster,R., Rahman,N. and Stratton,M.R. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.
 48. Amberger,J., Bocchini,C.A., Scott,A.F. and Hamosh,A. (2009) McKusick’s Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res.*, **37**, D793–D796.
 49. Reimand,J., Arak,T. and Vilo,J. (2011) g:Profiler—a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res.*, **39**, W307–W315.