APID interactomes: providing proteome-based interactomes with controlled quality for multiple species and derived networks

Diego Alonso-López, Miguel A. Gutiérrez, Katia P. Lopes, Carlos Prieto, Rodrigo Santamaría and Javier De Las Rivas[*]

Cancer Research Center (CiC-IBMCC, CSIC/USAL/IBSAL), Consejo Superior de Investigaciones Científicas (CSIC) and Universidad de Salamanca (USAL), 37007 Salamanca, Spain

ABSTRACT

APID (Agile Protein Interactomes DataServer) is an interactive web server that provides unified generation and delivery of protein interactomes mapped to their respective proteomes. This resource is a new, fully redesigned server that includes a comprehensive collection of protein interactomes for more than 400 organisms (25 of which include more than 500 interactions) produced by the integration of only experimentally validated protein–protein physical interactions. For each protein–protein interaction (PPI) the server includes currently reported information about its experimental validation to allow selection and filtering at different quality levels. As a whole, it provides easy access to the interactomes from specific species and includes a global uniform compendium of 90,379 distinct proteins and 678,441 singular interactions. APID integrates and unifies PPIs from major primary databases of molecular interactions, from other specific repositories and also from experimentally resolved 3D structures of protein complexes where more than two proteins were identified. For this purpose, a collection of 8,388 structures were analyzed to identify specific PPIs. APID also includes a new graph tool (based on Cytoscape.js) for visualization and interactive analyses of PPI networks. The server does not require registration and it is freely available for use at http://apid.dep.usal.es.

INTRODUCTION

Identification of all the specific connections between the elements that comprise a cellular system is crucial to unraveling its molecular architecture and mechanics. In this context, physical molecular interactions between protein pairs (called protein–protein interactions, PPIs) constitute an essential part of the cellular architecture in all living organisms. Genome-wide technologies have provided, over the last two decades, a compendium of the biomolecular entities that configurate many living systems, i.e., all the genes encoded in the genomes of specific organisms and the corresponding derived proteome. Once all these elements became known, the need for comprehensive maps of the molecular physical interactions that occur between such elements was evident, and systematic proteome-scale mapping of specific interactomes began ([1],[2]). Combined global identification of the molecular elements and their physical interactions opened a new avenue for depicting cellular networks and understanding the biomolecular processes that occur in living systems ([3],[4]).

It is clear that over the last decade there has been a great deal of effort to build biological databases and resources providing detailed information about the 'molecular interactions' (MI) determined in thousands of experimental studies in different biological systems, performed either using small-scale or large-scale technologies and reported in thousands of publications. Within these efforts it is worth mentioning the work of international consortiums such as IMEx (http://www.imexconsortium.org) ([5]) which include many primary databases as partners (such as DIP, IntAct, MINT) ([6]–[8]) or observers (such as BioGRID) ([9]), who have made important contributions toward creating well established standards for molecular interactions ([10]), as well as important collaborative efforts for providing integrated access to multiple types of molecular interactions from many resources ([11]–[13]).

As of January 2016, a search in PubMed (www.ncbi.nlm.nih.gov/pubmed) with the term 'protein–protein interaction' revealed 9,687 research articles, most published in the last five years. This indicates that current biomolecular research is highly interested in finding the molecular partners of the proteins or the gene products that are

---

studied in very different biological scenarios. Such interest demands an easy way to provide and visualize interacting proteins in a proteome-wide context. There are many bioinformatics tools and servers that provide information about protein interactions and protein functional associations. An extensive list can be found on Pathguide (http://www.pathguide.org/) (14) which includes more than five hundred biological pathway related resources and molecular interaction related resources. Moreover, there is a group of online resources which provides integration of both experimentally known and computationally predicted interactions, aiming for thorough comprehensiveness and coverage. These include STRING (15), GeneMANIA (16), FunCoup (17), ConsensusPathDB (18), I2D (19) and others. Such resources aim to integrate all types of interactions, as defined in their scopes. The STRING database (Search Tool for the Retrieval of Interacting Genes/Proteins), for example, is dedicated to finding all types of 'functional associations' between proteins, on a global scale (15).

The goal of the web server presented here (APID, Agile Protein Interactomes DataServer) is different because it does not include either 'predicted' protein interactions or 'functional associations' between proteins that do not reveal physical contacts established between two or more proteins based on specific biomolecular forces. In fact, many genetic studies have provided interesting 'functional associations' between individual pairs of genes that are defined as 'genetic interactions' (20) and are reported in several of the resources cited above (7,15,16). However, APID is focused solely on the generation and delivery of unified compendiums of known and experimentaly proven protein–protein physical interactions (PPIs). The protein interactions are provided, including quality levels associated with the number of experiments, methods and publications that report each interaction, and they are organized in interactomes per organism, mapped to their respective proteomes.

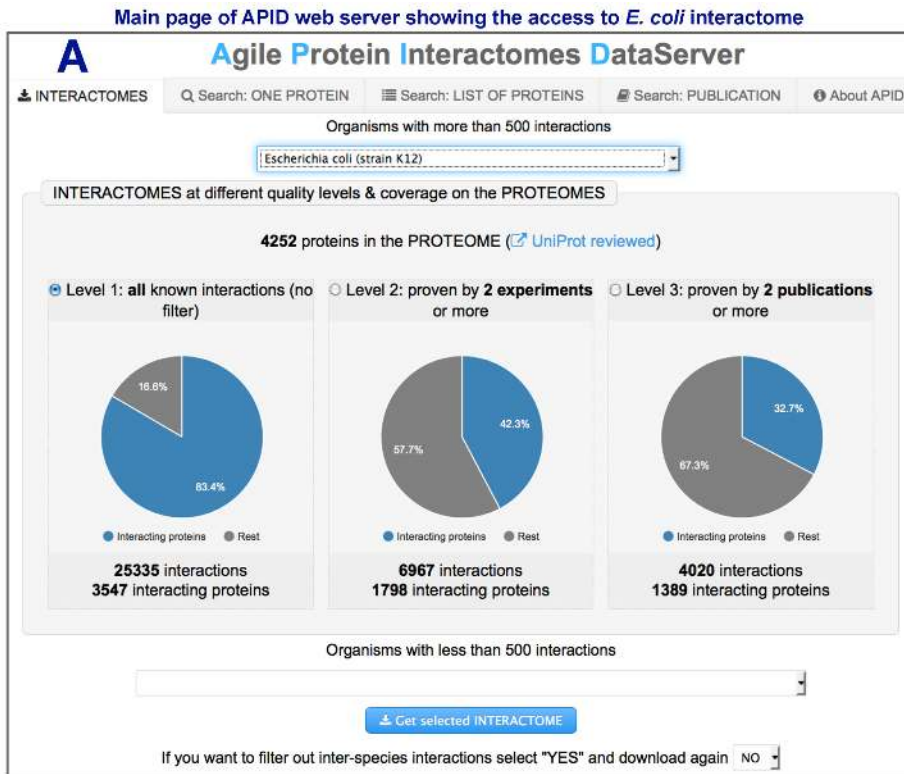## APID: providing proteome-based interactomes at different quality levels

APID (Agile Protein Interactomes DataServer) is a bioinformatics web server developed to provide protein interactomes at different quality levels and allowing their analysis and visualization as networks. This resource is a new, fully redesigned version of the APID web server (21) that provides a comprehensive collection of protein interactions for 448 organisms derived from the integration of known experimentally validated protein–protein physical interactions (22). Construction of the interactomes is done with a methodological approach (detailed below) to report quality levels and coverage over the proteomes for each organism included. Figure 1 presents a view of APID main web page showing an example for the *Escherichia coli (strain K12)* interactome (Figure 1A). In other panels, the figure presents statistics about several interactomes provided (Figure 1B and E), as well as images of the interactions display tool (Figure 1C) and the network tool (Figure 1D) where the colored pie charts on the nodes present user-selected biological functions that are shared by several proteins in the network.

As a whole, APID provides easy access to the interactomes from specific species and includes a global uniform compendium of 90,379 distinct proteins and 678,441 singular interactions. APID unifies PPIs from five major primary databases of molecular interactions (BioGRID (9), DIP (6), HPRD (23), IntAct (7), MINT (8)); from some specific repositories not included in the previous ones (BioPlex, http://wren.hms.harvard.edu/bioplex/) (2) and also from experimentally resolved 3D structures of protein complexes (PDB, http://www.rcsb.org/pdb/home/home.do) (24), where more than two proteins had been identified.

To incorporate the 3D structural information, 45,410 interfaces corresponding to 8,388 structures from the PDB were analyzed, searching for specific PPIs involving two different UniProt IDs (i.e. two distinct proteins). Using the criteria defined in PDBsum for protein–protein contacts (25), all of the interfaces between two protein chains were tested for at least one salt bridge, one disulfide bond or one hydrogen bond inferred from the 3D molecular proximity and atomic configuration (25). Interacting protein pairs found in this manner were registered with the corresponding PDB identifiers (PDB IDs), in order to count the specific number of 3D structures that validate each PPI. This process allowed us to assign 8,215 3D structures to 3,220 interactions. Details of the interfaces within these structures are provided on the web server, as they are considered to be one of the most credible proofs of the existence of a protein interaction.

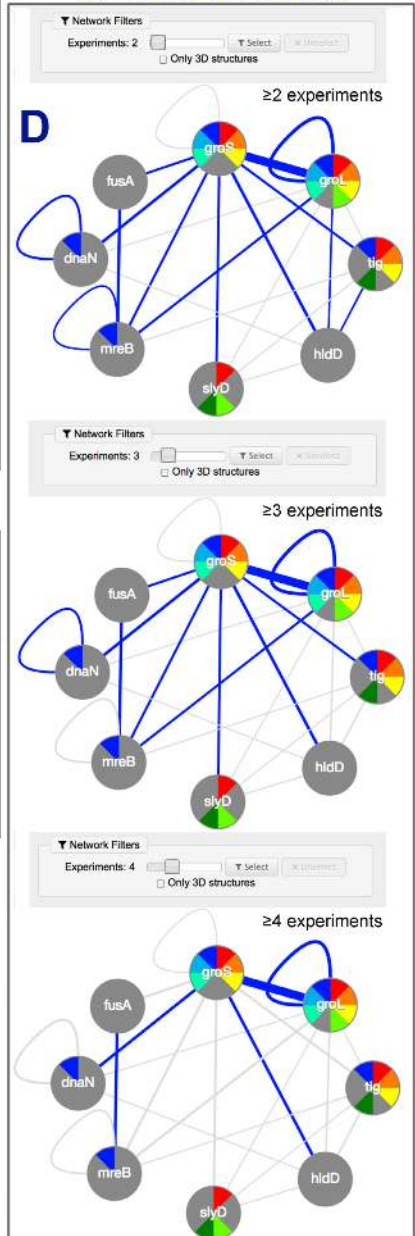## Network viewer to explore and analyze protein interactions

In order to facilitate the construction and interactive exploration of PPI networks, APID includes a new network visualization tool that is a native, web-based app which follows *HTML/CSS/JavaScript* standards and does not require third-party software, such as *Java* or *Flash*, on the client side. The app is based on Cytoscape.js, a *JavaScript* library for programmers (http://js.cytoscape.org/) (26) that was used to build the core of the viewer and which gave us the framework for developing new graphcentric utilities, such as: a drag-and-drop color tagging system for the nodes to highlight multiple functional annotations; realtime filters for interactions based on the number of experiments that demonstrate them or on the presence of 3D structural data; different graph layouts (Figure 1D). In this way, the APID network viewer enables the construction of sub-interactomes using query lists of proteins of interest and visual exploration of the corresponding networks, including an interactive selection of the value of the interactions (i.e. the reliability of the 'edges' in the network) as well as an interactive mapping of the protein functional environments (i.e. the functional annotations of the 'nodes' in the network), that allows identification of nodes with shared functions (Figure 1D). This mapping uses annotation to four biological functional spaces: Gene Ontology (BP, MF and CC) (27); InterPro (28); Pfam (29) and Reactome (30). All of the networks produced with this tool can be exported either as figures (.jpg, .png) or as flat files (.txt) to be read using other software for graph visualization or analysis. In particular, flat files from the networks and all the interactomes provided by APID can be loaded

Figure 1. Presentation of APID (Agile Protein Interactomes DataServer) including five panels with different views and information. (**A**) Panel showing the main web page of the APID web server after selecting '*E. coli (strain K12)*' to present its interactomes. (**B**) Table presenting the size that the interactome datasets from *E. coli* produced at three levels of experimental validation and filtering out inter-species interactions. (**C**) Example showing the interactions of protein groS (CH10_ECOLI), including its most reported interaction partner: groL (CH60_ECOLI), that has been validated in 15 experiments and in 11 3D structures. (**D**) View of the closest protein interactors of the groS protein using the APID network tool which allows coloring the nodes according to specific functional terms annotated to the proteins and also allows selection of the edges (in blue) according to the number of experiments that validate each interaction. (**E**) Table with the numbers included in APID about interactome sizes and proteome coverage for eight model organisms.

into Cytoscape (http://www.cytoscape.org/) (31), which is a very useful open source software platform for complex network analysis and visualization that allows the integration of multiple layers of biological information on the networks (31,32).

Comparison to other related tools

As indicated in the introduction, there are multiple bioinformatics tools or platforms that provide information about functional associations and interactions between proteins. A compendium of these can be found in Pathguide (http://www.pathguide.org/) (14). However, as far as we know, there are no servers identical to the new APID described here, focused on the integration of only experimentally validated protein–protein physical interactions. There are multiple applications or servers that took similar approaches to the first version of APID (21) and built integrated compendiums of PPIs for different organisms. Some of the most remarkable and complete ones are: iRefWeb (http://wodaklab.org/iRefWeb/) published in 2010 (33); Hit-Predict (http://hintdb.hgc.jp/htp/) published in 2011 (34); PINA (http://cbg.garvan.unsw.edu.au/pina/) published in 2012 (35); and Mentha (http://mentha.uniroma2.it) published in 2013 (36). These tools are currently accessible, but only two of them have been updated in 2016. We present a comparison of the PPI data corresponding to eight model organisms included in these two servers, the ones currently updated, *versus* APID (see Supplementary Data 1). This comparison indicates that APID provides interactomes with a 48.3% average coverage of the proteomes of these eight species; while iRefWeb shows a coverage of 39.3% and Mentha, 41.9%. These numbers correspond to the versions of these resources downloaded in January 2016. The increase in coverage that APID achieves with respect to the other resources may be due to several reasons: (**i**) the compared datasets may not correspond to updates of the same versions of the primary databases (despite the fact that in all cases the comparisons are of data available in January 2016); (**ii**) APID includes some new sources that are not included in iRefWeb or in Mentha (such as BioPlex and HPRD in the case of Mentha) (2); (**iii**) the different resources may not analyze the same raw files from the primary public databases. In fact, this last reason is probably the most important because, for example, Mentha integrates protein-interaction data curated by experts in compliance with IMEx curation policies, using the PSICQUIC protocol to implement an automatic procedure that, every week and without human intervention, aligns the integrated database with data regularly annotated by the primary databases (36). Therefore, anything that is not in PSICQUIC (11,12) will not be in Mentha. Another important difference is that APID uses the XML files (i.e. PSI-MI XML files) drawn from primary databases, but most of the meta-databases and servers that integrate multiple data from molecular interactions use a simpler format called MITAB (i.e. PSI-MI TAB, which is a common tab delimited format for MI data interchange: https://code.google.com/archive/p/psimi/wikis/PsimiTabFormat.wiki). More details about the procedures and methodology that APID employs to achieve an efficient integration and unification of PPI data are explained below. Finally, other differences observed are that these tools do not offer the same validation procedures with quality levels for the PPIs used in APID and do not integrate any extra information derived from the analyses of interactions in 3D structures of protein complexes.

Experimentally proven protein–protein physical interactions, unified and weighted

The APID server presents a way to evaluate and qualify PPIs based on identification of the distinct 'experiments' from the literature (i.e. from specific scientific articles reported in PubMed) that prove a given protein pair interaction. In other words, APID counts the number of 'experiments' as the number of times that the interaction between two proteins has been tested and demonstrated in a research lab with one specific method and reported in a published article. This is a different approach to the procedure followed by other PPI resources that count 'evidences' defined as the 'aggregated experimental evidences retrieved from the different databases' (8). Moreover, often these PPI resources build and provide a 'score' calculated for each interaction that is based on such counts of 'evidences' (8,11).

In APID, 'evidences' correspond to 'curation events' and they provide larger numbers than the 'experiments' because several primary databases can curate the same published articles and, when they do, it does not mean that a new experiment was done to test and validate the interaction. In fact, we performed an analysis to show that counting 'curation events' produces a clear overestimation of the interactions and, therefore, an overestimation of the size of the interactomes. Supplementary Data 2 presents a graphic comparison of the number of interactions included in the human interactome considering several numbers of 'experiments' or 'curation events'. This analysis shows that an interactome validated with 3 or more 'curation events' per interaction will be 48.5% larger than an interactome validated with 3 or more 'experiments'; thus demonstrating that producing scores based on curation events may not be very accurate.

Another fact is that in APID, counting the experiments is a simple and transparent process, since it does not attempt to calculate a 'score' derived from a rational combination of factors. The procedures to calculate such scores need to reach a compromise between every variable that describes an interaction and, therefore, are usually quite arbitrary and can sometimes be difficult to understand or confusing for the users. In fact, to illustrate the problems associated with the definition of an integrated score, we compared the results in APID for two well-known interactions that are validated by very different experimental approaches: HRAS (P01112) interaction with RAF1 (P04049) was validated in 36 singular experiments and HRAS interaction with SOS1 (Q07889) was validated in only seven singular experiments; by contrast we found 18 distinct PDB 3D structures that validate HRAS interaction with SOS1 but only three PDB structures that validate HRAS interaction with RAF1. It is very difficult to make a fair decision to rank and give a higher score to RASH–RAF1 interaction or to RASH–SOS1 interaction based on these numbers: which is better, 36 singular reported experiments or 18 distinct PDB structures?. For these reasons, we prefer to leave this discussion

open, providing all the experimental results for each singular interaction in APID and allowing the users to employ their own criteria to sort or rank the interactions. This ranking may even follow different approaches appropriate to different types of interactomic studies.

According to the strategy described, for each PPI pair the APID server provides a combination of four counts that measure the level of experimental validation: (**i**) the number of 'experiments' (calculated as described above); (**ii**) the number of 'methods' that validate such interaction (following PSI-MI ontology for the identification of different 'interaction detection methods') (37,38); (**iii**) the number of 'publications' that have reported such interaction (including specific PMIDs from PubMed); (**iv**) the number of '3D structures' from the PDB that include two proteins interacting in a specific way at molecular level (i.e. with H-bonds or other types of specific bonding inferred from the PDB) (24,25).

Architecture of the web server and procedures for integrating and unifying PPI data

The APID server was built with a protein and proteome-centered strategy, using the UniProt database (http://www.uniprot.org) as the main guide to identify and handle all of the proteins and map them into the reference proteomes of each species (based on the new proteome identifiers that UniProt recently developed: http://www.uniprot.org/proteomes/) (39). In this way UniProt, including both SwissProt and TrEMBL, was used as the main reference database and we used protein or gene identifier recursive mapping to UniProtKB AC/ID as the key way to integrate and unify data, thus avoiding duplications or incorrect identifications.

To provide a global view of the methodology and procedures followed to build APID, a graphic scheme presents the main workflow with the pipelines and steps applied to integrate the PPI data. This scheme is included as Supplementary Data 3 and also as a figure on the APID website.

With coverage as one of the main objectives, the procedure begins with an exhaustive parsing of the complete raw PSI-MI XML files from the five major public databases of molecular interactions: BioGRID (9), DIP (6), HPRD (23), IntAct (7), MINT (8). A TSV file with the data from BioPlex project (2) is also downloaded, parsed and integrated. For this part of the workflow, we designed a protocol based on JAMI (Java Framework for Molecular Interactions) (40) that processed all of the XML entries contained in the downloaded files. This approach allowed us to acquire all of the information contained in the source databases, and design a pipeline to discard any dataset that was incomplete or not appropriate, such as: (**i**) any participant of an interaction that is not a protein; (**ii**) any apparent participant with an ID that could not be matched to a UniProt ID; (**iii**) any Uniprot ID that was obsolete and deprecated and could not be replaced by a current UniProt ID. This procedure guaranteed that every participant in an interaction was registered as a protein and mapped to the UniProt database (SwissProt or TrEMBL). Gene names (i.e. official gene symbols such as KRAS for RASK_HUMAN or Tp53 for P53_MOUSE) were added as an annotation after the ID mapping to facilitate identification of participants in each

interaction and the use of the PPI data in other resources that employ gene identifiers. At the beginning of the workflow (Supplementary Data 3), for the records reporting protein interactions that include more than two proteins (i.e. records with multiple proteins) we applied the *spoke model* to expand the data and generate binary interactions from these co-complex data (22).

Once the ID mapping was completed, a unification pipeline was followed to merge data. For example: (**i**) curation events from different sources that reported the same interactions after protein ID matching, and (**ii**) isoforms of the same protein reported as different interactors. This unification allowed the identification of singular interaction pairs and eliminated many duplications. Unification of the interaction protein pairs was always performed following HUPO Proteomics Standards (PSI-MI) (37,38) including the ontology of terms with its hierarchy (as shown in http://www.ebi.ac.uk/ols/beta/ontologies/mi).

APID: download protein interactomes and visualize networks of specific PPI sets

The APID web server is fully functional, free and open to all users at URL: http://apid.dep.usal.es. The server's first page allows downloading of protein interactions for more than 400 organisms at three different quality levels: level **1**) all known interactions; level **2**) interactions proven by 2 or more experiments; level **3**) interactions reported in two or more research publications. Data for organisms with more than 500 known interactions are presented in an alphabetically ordered drop-down list to allow rapid access. The rest of the organisms are included in a second similar drop-down list. For each organism the interactomes can be downloaded, including interactions with proteins from other species (inter-species interactions) or by simply filtering out such interactions. The server also includes pages with search engines for single proteins ('Search: ONE PROTEIN') or lists of proteins ('Search: LIST OF PROTEINS'), using either UniProt AC/ID identifiers or standard gene/protein Symbols. On another page the server includes a search tool ('Search: PUBLICATION') to query by published articles (i.e. a PubMed ID number, PMID) in order to find all of the PPIs that have been reported in a given publication, including all of the information about such interactions that is currently integrated in the server. APID includes examples in all of these search pages.

Search results deliver PPIs in a tabular format, showing all the interactor pairs with protein names (UniProt IDs) and taxonomy IDs (http://www.ncbi.nlm.nih.gov/taxonomy) to identify the species, plus all experimental evidences counted and presented in five different columns: (**i**) number of experiments, (**ii**) number of methods, (**iii**) number of publications, (**iv**) number of 3D structures (PDBs) and (**v**) number of curation events (including source databases) (see Figure 1C). The data can be sorted by any of the columns and filtered to select a minimum number of experiments, methods, publications or curation events. Once a set or subset of interactions is displayed, the web allows one to build a network with the network viewer app for the proteins and interactions selected.

This website also contains a section called 'About_APID' with useful information including a 'HELP' page with a brief tutorial presenting some simple cases that illustrate how to use the server. It also includes a page named 'METHODOLOGY' that provides a global view of the procedures followed to build APID with a figure presenting the main workflow with the pipelines and steps applied to integrate the PPI data (this figure is also included here as Supplementary Data 3). Another page named 'DOWNLOADS' allows downloading (in MITAB format) of all of the raw curation events from PPIs that are integrated in APID resulting from unification of the primary public databases, grouped into single files by organism. Two other pages ('STATISTICS' and 'ACKNOWLEDGEMENTS') provide more information about source databases, versions, updates, references and technologies. The site also includes a 'Show_HELP' button on all pages, which presents captions with brief descriptions of each one of the elements viewed on a given page. Throughout the site, server links to the corresponding source databases are included, such as: UniProt for proteins; UniProt-Proteomes for proteomes; PubMed for publications; PDB for 3D structures; and the corresponding primary molecular interaction databases for all the singular curation events reported.

Finally, the web server presented here is a fully redesigned PPI resource providing agile access to protein interactomes, but it maintains the value and credit of the first APID version (published in 2006 in *Nucleic Acids Research, Web Server* issue) (15) keeping the same acronym for its name. We feel that this will allow it to be of better service to the research community and facilitate a broader use.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

REFERENCES

1. Rolland,T., Tasan,M., Charloteaux,B., Pevzner,S.J., Zhong,Q., Sahni,N., Yi,S., Lemmens,I., Fontanillo,C., Mosca,R. *et al.* (2014) A proteome-scale map of the human interactome network. *Cell*, **159**, 1212–1226.
2. Huttlin,E.L., Ting,L., Bruckner,R.J., Gebreab,F., Gygi,M.P., Szpyt,J., Tam,S., Zarraga,G., Colby,G., Baltier,K. *et al.* (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell*, **162**, 425–440.
3. Barabasi,A.L. and Oltvai,Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.*, **5**, 101–113.
4. Vidal,M., Cusick,M.E. and Barabasi,A.L. (2011) Interactome networks and human disease. *Cell*, **144**, 986–998.
5. Orchard,S., Kerrien,S., Abbani,S., Aranda,B., Bhate,J., Bidwell,S., Bridge,A., Briganti,L., Brinkman,F.S., Cesareni,G. *et al.* (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat. Methods*, **9**, 345–350.
6. 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.
7. 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.
8. Licata,L., Briganti,L., Peluso,D., Perfetto,L., Iannuccelli,M., Galeota,E., Sacco,F., Palma,A., Nardozza,A.P., Santonico,E. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.*, **40**, D857–D861.
9. Chatr-Aryamontri,A., Breitkreutz,B.J., Oughtred,R., Boucher,L., Heinicke,S., Chen,D., Stark,C., Breitkreutz,A., Kolas,N., O'Donnell,L. *et al.* (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Res.*, **43**, D470–D478.
10. Orchard,S., Salwinski,L., Kerrien,S., Montecchi-Palazzi,L., Oesterheld,M., Stumpflen,V., Ceol,A., Chatr-aryamontri,A., Armstrong,J., Woollard,P. *et al.* (2007) The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat. Biotechnol.*, **25**, 894–898.
11. Aranda,B., Blankenburg,H., Kerrien,S., Brinkman,F.S., Ceol,A., Chautard,E., Dana,J.M., De Las Rivas,J., Dumousseau,M., Galeota,E. *et al.* (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat. Methods*, **8**, 528–529.
12. 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.
13. Orchard,S., Ammari,M., Aranda,B., Breuza,L., Briganti,L., Broackes-Carter,F., Campbell,N.H., Chavali,G., Chen,C., del-Toro,N. *et al.* (2014) The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.
14. Bader,G.D., Cary,M.P. and Sander,C. (2006) Pathguide: a pathway resource list. *Nucleic Acids Res.*, **34**, D504–D506.
15. Szklarczyk,D., Franceschini,A., Wyder,S., Forslund,K., Heller,D., Huerta-Cepas,J., Simonovic,M., Roth,A., Santos,A., Tsafou,K.P. *et al.* (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, **43**, D447–D452.
16. Zuberi,K., Franz,M., Rodriguez,H., Montojo,J., Lopes,C.T., Bader,G.D. and Morris,Q. (2013) GeneMANIA prediction server 2013 update. *Nucleic Acids Res.*, **41**, W115–W122.
17. Schmitt,T., Ogris,C. and Sonnhammer,E.L. (2014) FunCoup 3.0: database of genome-wide functional coupling networks. *Nucleic Acids Res.*, **42**, D380–D388.
18. Kamburov,A., Stelzl,U., Lehrach,H. and Herwig,R. (2013) The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.*, **41**, D793–D800.
19. Niu,Y., Otasek,D. and Jurisica,I. (2010) Evaluation of linguistic features useful in extraction of interactions from PubMed; application to annotating known, high-throughput and predicted interactions in I2D. *Bioinformatics*, **26**, 111–119.
20. Mani,R., St Onge,R.P., Hartman,J.L. 4th, Giaever,G. and Roth,F.P. (2008) Defining genetic interaction. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 3461–3466.
21. Prieto,C. and De Las Rivas,J. (2006) APID: Agile Protein Interaction DataAnalyzer. *Nucleic Acids Res.*, **34**, W298–W302.
22. De Las Rivas,J. and Fontanillo,C. (2010) Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS Comput. Biol.*, **6**, e1000807.
23. 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.
24. Rose,P.W., Prlic,A., Bi,C., Bluhm,W.F., Christie,C.H., Dutta,S., Green,R.K., Goodsell,D.S., Westbrook,J.D., Woo,J. *et al.* (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.*, **43**, D345–D356.
25. de Beer,T.A., Berka,K., Thornton,J.M. and Laskowski,R.A. (2014) PDBsum additions. *Nucleic Acids Res.*, **42**, D292–D296.
26. Franz,M., Lopes,C.T., Huck,G., Dong,Y., Sumer,O. and Bader,G.D. (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, **32**, 309–311.

27. Huntley,R.P., Sawford,T., Mutowo-Meullenet,P., Shypitsyna,A., Bonilla,C., Martin,M.J. and O'Donovan,C. (2015) The GOA database: gene Ontology annotation updates for 2015. *Nucleic Acids Res.*, **43**, D1057–D1063.

28. Mitchell,A., Chang,H.Y., Daugherty,L., Fraser,M., Hunter,S., Lopez,R., McAnulla,C., McMenamin,C., Nuka,G., Pesseat,S. *et al.* (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.*, **43**, D213–D221.

29. Finn,R.D., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Mistry,J., Mitchell,A.L., Potter,S.C., Punta,M., Qureshi,M., Sangrador-Vegas,A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.

30. Croft,D., Mundo,A.F., Haw,R., Milacic,M., Weiser,J., Wu,G., Caudy,M., Garapati,P., Gillespie,M., Kamdar,M.R. *et al.* (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res.*, **42**, D472–D477.

31. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

32. Cline,M.S., Smoot,M., Cerami,E., Kuchinsky,A., Landys,N., Workman,C., Christmas,R., Avila-Campilo,I., Creech,M., Gross,B. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protocols*, **2**, 2366–2382.

33. 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*, **2010**, baq023.

34. Patil,A., Nakai,K. and Nakamura,H. (2011) HitPredict: a database of quality assessed protein-protein interactions in nine species. *Nucleic Acids Res.*, **39**, D744–D749.

35. Cowley,M.J., Pinese,M., Kassahn,K.S., Waddell,N., Pearson,J.V., Grimmond,S.M., Biankin,A.V., Hautaniemi,S. and Wu,J. (2012) PINA v2.0: mining interactome modules. *Nucleic Acids Res.*, **40**, D862–D865.

36. Calderone,A., Castagnoli,L. and Cesareni,G. (2013) mentha: a resource for browsing integrated protein-interaction networks. *Nat. Methods*, **10**, 690–691.

37. Hermjakob,H., Montecchi-Palazzi,L., Bader,G., Wojcik,J., Salwinski,L., Ceol,A., Moore,S., Orchard,S., Sarkans,U., von Mering,C. *et al.* (2004) The HUPO PSI's molecular interaction format–a community standard for the representation of protein interaction data. *Nat. Biotechnol.*, **22**, 177–183.

38. Kerrien,S., Orchard,S., Montecchi-Palazzi,L., Aranda,B., Quinn,A.F., Vinod,N., Bader,G.D., Xenarios,I., Wojcik,J., Sherman,D. *et al.* (2007) Broadening the horizon–level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.*, **5**, 44.

39. UniProt_Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.

40. Orchard,S., Albar,J.P., Binz,P.A., Kettner,C., Jones,A.R., Salek,R.M., Vizcaino,J.A., Deutsch,E.W. and Hermjakob,H. (2014) Meeting new challenges: The 2014 HUPO-PSI/COSMOS Workshop: 13-15 April 2014, Frankfurt, Germany. *Proteomics*, **14**, 2363–2368.