IID 2018 update: context-specific physical protein–protein interactions in human, model organisms and domesticated species

Max Kotlyar[1], Chiara Pastrello[1], Zara Malik[1] and Igor Jurisica[1,2,3,*]

[1]Krembil Research Institute, University Health Network, Toronto, ON M5T 0S8, Canada, [2]Departments of Medical Biophysics and Computer Science, University of Toronto, Toronto, ON M5S 1A4, Canada and [3]Institute of Neuroimmunology, Slovak Academy of Sciences, Bratislava, Slovakia

ABSTRACT

Knowing the set of physical protein–protein interactions (PPIs) that occur in a particular context—a tissue, disease, or other condition—can provide valuable insights into key research questions. However, while the number of identified human PPIs is expanding rapidly, context information remains limited, and for most non-human species context-specific networks are completely unavailable. The Integrated Interactions Database (IID) provides one of the most comprehensive sets of context-specific human PPI networks, including networks for 133 tissues, 91 disease conditions, and many other contexts. Importantly, it also provides context-specific networks for 17 non-human species including model organisms and domesticated animals. These species are vitally important for drug discovery and agriculture. IID integrates interactions from multiple databases and datasets. It comprises over 4.8 million PPIs annotated with several types of context: tissues, subcellular localizations, diseases, and druggability information (the latter three are new annotations not available in the previous version). This update increases the number of species from 6 to 18, the number of PPIs from ~1.5 million to ~4.8 million, and the number of tissues from 30 to 133. IID also now supports topology and enrichment analyses of returned networks. IID is available at http://ophid.utoronto.ca/iid.

INTRODUCTION

Physical protein–protein interaction (PPI) data have become a widely used resource in molecular biology. They are important because most cellular processes, such as growth, metabolism, and repair, occur primarily through PPIs. Consequently, understanding the molecular mechanisms behind diseases and treatments requires knowledge of PPIs. Currently available PPI data, though far from complete, have provided important insights into numerous problems in molecular biology including identification of gene function (1,2), disease genes (3,4), biomarker signatures (5,6), drug targets (7,8), and drug efficacy (9).

While PPI data can help address numerous research problems, effectively using these data can be challenging due to several reasons: false positive and false negative errors, lack of context information (e.g. tissue and disease annotations of PPIs), and difficulty extracting meaningful conclusions from PPI networks. For example, improving a lung cancer signature would require a reliable, comprehensive, lung-specific network involving prognostic signature proteins, and ways of interpreting how this network can improve the signature; unfortunately, meeting these requirements can be difficult. False positive rates have been estimated at over 80% for some PPI detection studies (10), but may be typically lower, and can be reduced by filtering PPIs based on the quantity and reliability of supporting evidence. False negatives (i.e. missing interactions) can often be a bigger problem; about 50% of human proteins have few or no detected interactions (Figure 1)—rendering any PPI-based analysis inapplicable to much of the proteome and affecting data interpretation. The rate of missing interactions is unevenly distributed across proteins; some proteins may have high rates due to technical challenges of detecting their interactions (11), or research bias in favor of other proteins (12). The overall false negative rate for human PPI data may be greater than 50%, based on an estimated human interactome size of 650,000 PPIs (13). The number of detected human PPIs has already exceeded several lower estimates of interactome size (10,14), and the yearly rate of detected PPIs has not plateaued—further implying a large percentage of missing interactions. If PPIs are available, they need to occur in the relevant context, such as the tissue, cell-type, or disease state being studied. However, PPI detection is typically conducted in yeast or cell-lines. The chances of detected PPIs occurring in a relevant context may be low,

---
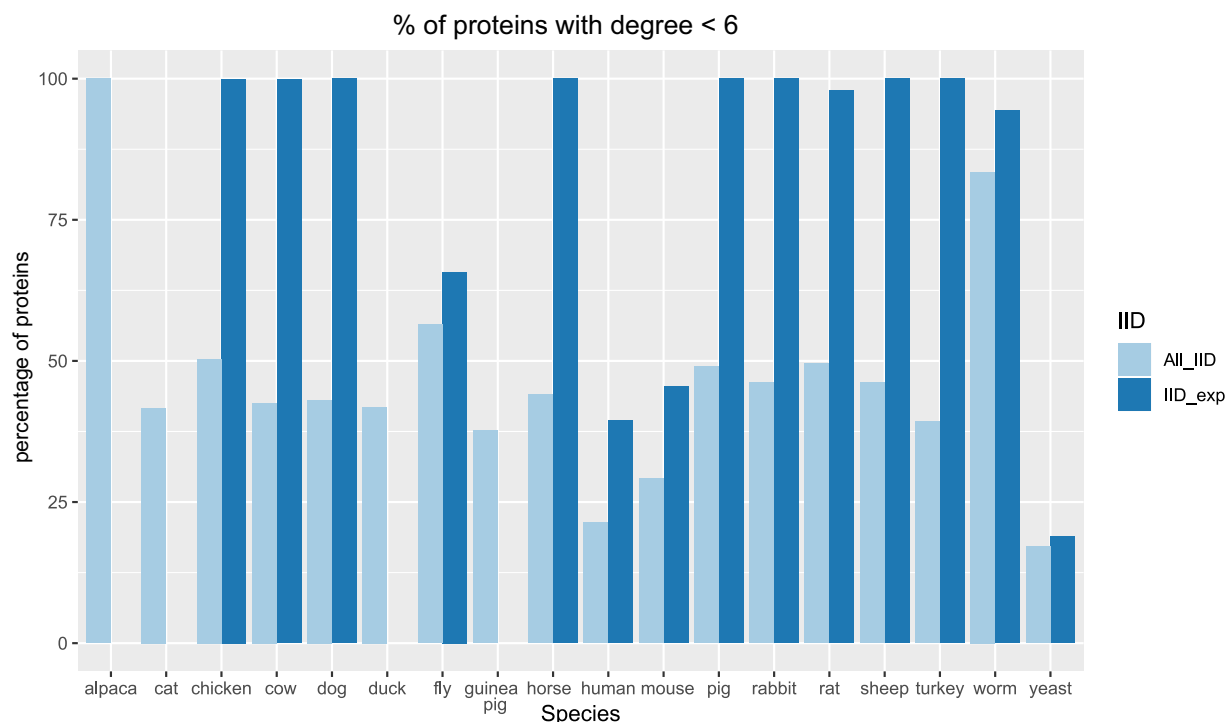
Figure 1. Figure shows the percentage of proteins with degree 5 or lower in each species, taking into consideration the entire set of interactions in IID (light blue) or only the experimental ones (dark blue).

since tissues may express less than half of the genome (15). Estimating the *in vivo* context of interactions requires integrating transcriptomic, proteomic or other data. If PPIs in the relevant context can be detected, the next challenge is to interpret the network and its biological significance.

Our database portal, the Integrated Interactions Database (IID), focuses on addressing the problems of errors, context, and interpretability of PPI data. Given a set of proteins and a context (e.g. tissue, subcellular localization, disease), IID returns a reliable, comprehensive, context-specific interaction network for these proteins, and helps to interpret this network through topological and enrichment analyses. IID provides extensive options for controlling false positive and false negative rates, context, network annotation, and analysis. The content of IID has greatly expanded since the previous release in 2015: the number of species has increased from 6 to 18, the number of tissue contexts has expanded from 30 to 133, three new types of contexts have been added, as well as network analysis.

MATERIALS AND METHODS

PPI sources

Experimentally detected PPIs were obtained primarily from seven curated databases: BioGRID (16) 3.4.158, DIP (17) 2017-02-05, HPRD (18) Release 9, I2D (19) 2.3, InnateDB (20) 5.4, IntAct (21) 4.2.12, and MINT (22) downloaded 2018-05-15. Smaller numbers of PPIs were obtained through targeted curation of literature and from curated PPIs reported in Lefebvre *et al.* (23). Predicted PPIs were obtained from five sources: predictions from Rhodes

et al. (24) with a likelihood ratio cut-off of 381, predictions from Lefebvre *et al.* (23) with probabilities greater than 0.5, predictions from Elefsinioti *et al.* (25) with probabilities greater than 0.7, predictions from Zhang *et al.* (26) with likelihood ratios of at least 600, and FpClass predictions from Kotlyar *et al.* (11) with a false discovery rate less than 0.6. Predicted interactions were available only for human and yeast.

Orthologous PPIs were generated by mapping experimentally detected PPIs in each of the eighteen IID species to orthologous protein pairs in the other 17 species. Mappings were done using 1:1 orthologs from Ensembl (27) release 92.

Mapping between gene and protein IDs

Mappings between various gene and protein IDs were based on UniProt (28) release 2018_06. For a more complete set of mappings between Ensembl and UniProt IDs, mappings from Ensembl release 92 were also used; this enabled more orthologous PPIs and better support for queries using Ensembl IDs.

Assignment of context to PPIs

Tissues. A PPI was assigned to a tissue if its two encoding genes were expressed in the tissue. A gene was considered expressed in a tissue if its mas5 normalized expression was greater than 200, as in Bossi *et al.* (29). Gene expression levels in tissues were determined from 20 gene expression datasets downloaded from NCBI GEO (30): GSE1133, GSE3526, GSE7307, GSE7763, GSE9485, GSE10246, GSE20113, GSE20990, GSE23328,

GSE24207, GSE25138, GSE39796, GSE89347, GSE90449, GSE100083, GSE106641, GSE107494, GSE108033, GSE115799, GSE117834. All datasets were normalized using the mas5 function in the *affy* package (31) in R. In each dataset, disease tissues were removed, replicates were averaged and probeset IDs were mapped to Entrez Gene IDs. If a gene was represented by multiple probesets, the one with the highest variance was selected.

Detailed joint-related tissues. Human PPIs were assigned to joint-related tissues by the same approach as other tissues, described above. Gene expression levels in joint-related tissues were determined from seven gene expression datasets downloaded from NCBI GEO (30): GSE9329, GSE10024, GSE10500, GSE18338, GSE32398, GSE39795, GSE40942.

Detailed brain structures. Human PPIs were assigned to brain structures where both encoding genes were expressed. Normalized microarray gene expression data for brain structures was obtained from the Allen Human Brain Atlas (32) (http://human.brain-map.org/static/download). Probe expression levels were averaged across samples and if a gene was represented by multiple probes, the probe with the highest variance was selected. A gene was considered expressed in a brain structure if its $\log_2$-normalized expression was above 5—a threshold described in the database documentation (http://help.brain-map.org/display/humanbrain/Documentation). A PPI was assigned to a brain structure if its two encoding genes were expressed at or above this level in the structure.

This procedure was used to assign human PPIs to 38 brain structures, each represented by at least 20 samples. PPIs were also assigned to 64 higher level brain structures that subsume these 38 structures according to the Human Brain Atlas ontology (http://help.brain-map.org/display/api/Atlas+Drawings+and+Ontologies#AtlasDrawingsandOntologies-StructuresAndOntologies). A PPI assigned to a given low-level structure, was also assigned to all ancestors of this structure in the ontology.

Subcellular localizations. PPIs were assigned to 13 high-level subcellular localizations, based on Gene Ontology (GO) (33,34) compartment annotations of the interacting proteins. A PPI was assigned to a localization if both proteins were annotated with the localization or with its descendent terms in the GO compartment ontology. GO compartment annotations for proteins were obtained from UniProt (28) release 2018_06.

Diseases. PPIs were assigned to 37 diseases and 54 disease categories from Disease Ontology (35), based on gene-disease associations from DisGeNET (36) v5.0. A PPI was assigned to a disease if its two encoding genes were associated with the disease in DisGeNET. To increase the reliability of gene-disease associations, only associations supported by at least two publications were used.

DisGeNET disease names were mapped to Disease Ontology names by using UMLS (37) concept IDs. PPIs were annotated with these diseases and also with categories from Disease Ontology that encompassed these diseases; a PPI

assigned to a disease was also assigned to all ancestors of the disease in the ontology. PPIs were annotated with 91 diseases and higher level disease categories. Non-human PPIs were assigned to diseases based on disease associations of orthologous human protein pairs.

Drug target categories. PPIs were assigned to four major classes of drug targets (38): enzymes, ion channels, receptors, and transporters. A PPI was assigned to a class if one or both proteins were annotated with the GO category of this class according to UniProt (28) or with a descendent of the category in the GO ontology.

Drug targets. PPIs were annotated with drugs that target either of the interacting proteins according to DrugBank (39) v5.0. PPIs were also annotated with drugs that target orthologs of the interacting proteins.

Topology analysis

Topology analysis calculates degree, clustering coefficient, and normalized betweenness centrality of proteins in returned networks. Degree and clustering coefficient are calculated by custom javascript code and normalized betweenness centrality is calculated by cytoscape.js (40).

Enrichment analysis

Enrichment *P*-values are calculated using a hypergeometric cumulative distribution (hcd) function implemented in javascript. To calculate the enrichment of a given PPI annotation, PPI_a (e.g. presence in plasma membrane), in the returned network, the following parameters are used with the hcd function: N = number of PPIs matching the user-selected evidence and species (e.g. number of experimentally detected PPIs in mouse); M = number of PPIs matching the selected species and evidence type, and having annotation PPI_a; n = number of PPIs in the returned network; m = number of PPIs in the returned network, with annotation PPI_a. Enrichment is available for the following annotations: tissues (not detailed structures), subcellular localizations, diseases, and drug target categories.

WEBSITE DESCRIPTION

IID provides access to detected and predicted PPIs in 18 species (Table 1). PPIs are annotated with tissue, subcellular localization, disease and druggability information. These annotations can be used for filtering PPIs or helping to interpret the resulting network. Returned networks can be analyzed by topology or enrichment for PPI annotations.

Inputs

Required inputs to IID comprise gene or protein IDs and their species. IDs may include gene symbols, Entrez, Ensembl, and UniProt. Optional inputs control how IID searches for PPIs (e.g. retrieves interactions between pairs of query proteins, or between query proteins and any others), the required evidence for PPIs, the context for filtering PPIs, and PPI annotations included in output.

Table 1. Number of proteins and interactions per type of evidence per species

Species			PPIs			
Common name	Latin name	Proteins	Experimental	Orthologous	Predicted	Total
alpaca[*]	*Vicugna pacos*	13	0	13	0	13
cat	*Felis silvestris catus*	14 491	0	296 308	0	296 308
chicken	*Gallus gallus domesticus*	11 744	399	223 386	0	223 701
cow	*Bos taurus*	14 812	561	301 684	0	302 123
dog	*Canis lupus familiaris*	14 568	59	292 826	0	292 857
duck	*Anas platyrhynchos*	11 569	0	221 125	0	221 125
fly	*Drosophila melanogaster*	10 275	62 249	51 916	0	111 975
guinea pig	*Cavia porcellus*	14 252	0	294 510	0	294 510
horse	*Equus caballus*	14 572	5	303 500	0	303 504
human	*Homo sapiens*	19 250	334 315	50 866	667 804	975 877
mouse	*Mus musculus*	16 297	37 683	287 031	0	316 402
pig	*Sus scrofa*	14 733	76	300 884	0	300 945
rabbit	*Oryctolagus cuniculus*	13 444	135	257 965	0	258 056
rat	*Rattus norvegicus*	15 468	6 929	276 002	0	281 909
sheep	*Ovis aries*	14 476	3	289 985	0	289 986
turkey	*Meleagris gallopavo*	10 960	2	201 945	0	201 947
worm	*Caenorhabditis elegans*	6 898	13 723	46 595	0	59 463
yeast	*Saccharomyces cerevisiae*	6 318	161 851	9 736	61 720	197 041
Totals		**224 140**	**617 990**	**3 706 277**	**729 524**	**4 927 742**

[*]IID contains few alpaca proteins and PPIs because most alpaca proteins have not been identified: UniProt contains 164 alpaca protein IDs, corresponding to 28 unique Ensembl genes.

Controlling error rates

IID provides ways of controlling false positive and false negative rates of retrieved PPIs. The false positive rate can be controlled by setting a minimum number of publications or bioassays supporting each PPI. PPIs supported by a single publication and bioassay have been considered less reliable (12), but increasing these thresholds may remove true PPIs detected only by specialized assays or in specific contexts (41), and thus may substantially increase false negative rates.

The false negative rate can be reduced by allowing more types of interaction evidence: experimental (i.e., detection by bioassays), orthology based, or predicted. Experimental evidence is typically considered most reliable, but is largely unavailable for most non-human species, and even in human, less than 50% of PPIs may have been detected by bioassays. Using orthology-based PPIs may dramatically decrease the false negative rate in most non-human species, but the false positive rates of these PPIs have not been extensively benchmarked. Computationally predicted PPIs may also substantially decrease the false negative rate, but are currently available in IID for human and yeast networks only. Predicted PPIs comprise high-confidence predictions from five computational studies (11,23–26), which conducted extensive assessments of false positive rates, in most cases with experimental validation. These predictions decrease the number of low-degree proteins and PPI 'orphans' (11), making PPI-based analysis methods (e.g. for improving disease signatures) applicable to a larger portion of the proteome and less biased.
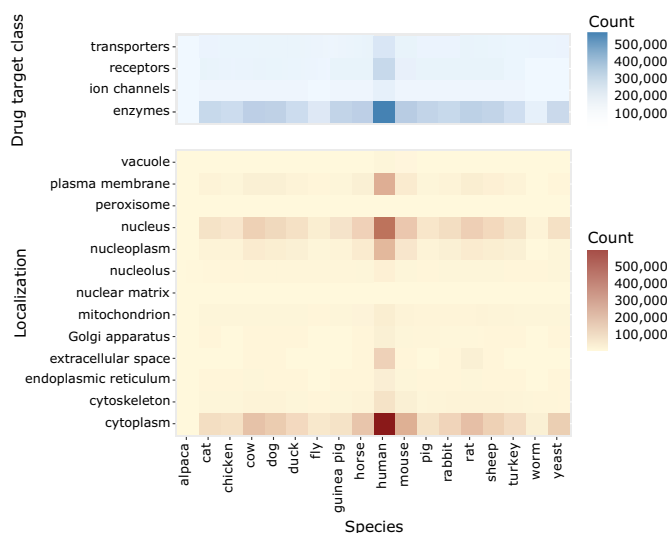
Specifying context

IID enables filtering PPIs by tissue, subcellular localization, disease and druggability. Tissue options include 26 high-level categories (e.g. adipose tissue, brain, Figure 2A), and

comprehensive options for joint-related tissues (five categories, Figure 2B) and human brain structures (102 categories, Figure 2C). As visible in Figure 2A, options for non-human species are more limited. IID uses gene expression data from GEO (30) and Allen Brain Atlas (32) to assign tissues—a PPI is annotated with tissues where the two encoding genes are expressed above background noise. This annotation approach has been used previously (29,42–44), and resulting networks have been shown to outperform unfiltered networks for applications such as prioritization of disease genes (45–47). As an example, we queried IID for interactions of SLC22A6, a protein involved in renal sodium-dependent transport and excretion of organic anions (https://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC22A6). A researcher who would be interested in knowing the molecular basis of SLC22A6's role in kidney and who would collect all interactions of SLC22A6 would use a misleading network: as highlighted in Figure 2D, only two-thirds of SLC22A6 PPIs are predicted to be in kidney. The output of IID is a tab-separated file that can be used for network visualization and analysis—in our example we used NAViGaTOR 3.08 (http://ophid.utoronto.ca/navigator) (48).

Subcellular localizations comprise 13 high-level GO cellular compartment categories (e.g. Golgi apparatus, cytoplasm) (Figure 3). A PPI is annotated with a localization if the two proteins are annotated with the localization or its Gene Ontology descendants. Similarly, a PPI is annotated with a disease if the two encoding genes are associated with the disease according to DisGeNET (36). PPIs are also annotated with higher level disease categories, based on Disease Ontology (35). Figure 4 shows the distribution of human PPIs per disease. The last context type, druggability, helps identify PPIs that may be amenable to modulation by drugs (Figure 3). There are two ways to filter by druggability: using drug target classes or drug targets. Filtering

Figure 2. Tissue distributions of PPIs in each IID species (**A**). Distribution in human of detailed joint (**B**) and brain tissues (**C**). Network of SLC22A6, a protein involved in renal sodium-dependent transport and excretion of organic anions. Blue edges indicate PPIs in kidney, yellow edges indicate PPIs in synovial macrophages, green edges indicate PPIs in both tissues, and black edges indicate PPIs without tissue annotations (**D**). Data from IID, network layout generated using NAViGaTOR 3.08 (48).

Figure 3. Drug target class (top) and localization (bottom) distributions of PPIs in each IID species.

by target classes returns PPIs where one or both interacting proteins are members of protein classes (enzymes, ion channels, receptors, transporters) that are commonly targeted by drugs. Filtering by drug targets returns PPIs where one or both interacting proteins are targeted by drugs or have orthologs that are targeted.

IID enables users to select any number of contexts and combine these contexts in different ways. Within each context type (e.g. tissue), users can specify whether returned PPIs can be in any of the selected contexts (e.g. present in either kidney or liver) or must be in all selected contexts (e.g. present in kidney and liver). If multiple context types are selected (e.g. tissues and subcellular localizations), the context types will be combined as conjunctions.

Output and downloads

Results are returned in a tabular format with one PPI per row. Users can choose to include interaction evidence (PubMed IDs, detection methods) in the results, as well as any context annotations. Full networks for each species, including context annotations, can be downloaded in tab-delimited format.

Analysis

IID provides topology and enrichment analysis for returned networks. Topology analysis can identify important proteins in the network based on degree and betweenness. Proteins of high degree (hubs) tend to be conserved across species and frequently have a large impact on phenotype (49), though high degree may also be due to research bias (50). Such proteins may be the best candidates for further investigating pathways, disease signatures, or drug side-effects. Topology analysis can also help identify protein complexes comprising more than two proteins, by calculating clustering coefficients. Proteins with high clustering coefficients may form complexes involving most of their interaction partners. Proteins in the same complex typically have

similar properties. Consequently, a complex can be helpful for predicting the properties of its members, such as function, subcellular localization and disease.

IID enrichment analysis can help identify conditions where the network is physiologically important. Typically, enrichment analysis determines whether a set of proteins (genes) is enriched for certain annotations, relative to a background population such as all proteins in the known interactome or the proteome. However, IID determines if retrieved PPIs (rather than proteins) are enriched for annotations, relative to all PPIs in the same species, and with the same interaction evidence that was selected in the query. For example, if a user searched for mouse PPIs supported by experimental evidence, then enrichment will be calculated relative to all mouse PPIs with experimental evidence. Enrichment analysis can be done on tissue, subcellular localization, disease, or drug annotations.

Novel features in IID 2018

This update substantially expands both the content and functionality of IID 2015-09. The number of species has increased from 6 to 18 (Table 1). While the first 6 species were human and common model organisms, the 12 new species are meant to support veterinary and agricultural research. The total number of PPIs has increased from ~1.5 million to ~4.8 million. Available context annotations for PPIs have substantially expanded as well. The number of tissues increased from 30 to 133 with the addition of detailed human brain structures and joint-related tissues. Three new context types have been added: subcellular localizations, diseases, and druggability information. The functionality of IID now includes two types of network analysis: topology analysis to identify important parts of the network and enrichment analysis of tissues, localizations, diseases, and druggability.

The addition of comprehensive options for brain and joint-related tissues supports the use of PPI networks in neurological and arthritis research. Brain disorders are increasing in incidence worldwide, but there is no cure for diseases like neurodegenerative disorders, autism, or schizophrenia. Unfortunately, failure rates in drug development for neurologic and psychiatric diseases are quite high, due to the complexity of the human brain—linked to difficulties developing appropriate animal models, and resulting in pharmaceutical companies losing interest in the field (51). Similarly, the degenerative disease osteoarthritis affects a large part of the population globally, yet remains without curative treatment (52). We previously demonstrated that many drug targets and evolutionarily recent proteins (like the ones present in brain) are understudied. With the current IID update we aim to provide the tools to fill this research gap, and enable molecular and pharmacological researchers to improve the success of drug development strategies (11).

IID displays available brain tissues as an ontology tree, and joint-related and high-level tissues as lists; users can select any number of these tissues. Moreover, IID provides annotations for druggability of PPIs (calculated as described in methods). Figure 3 shows the number of PPIs per species, annotated with different classes of targets.
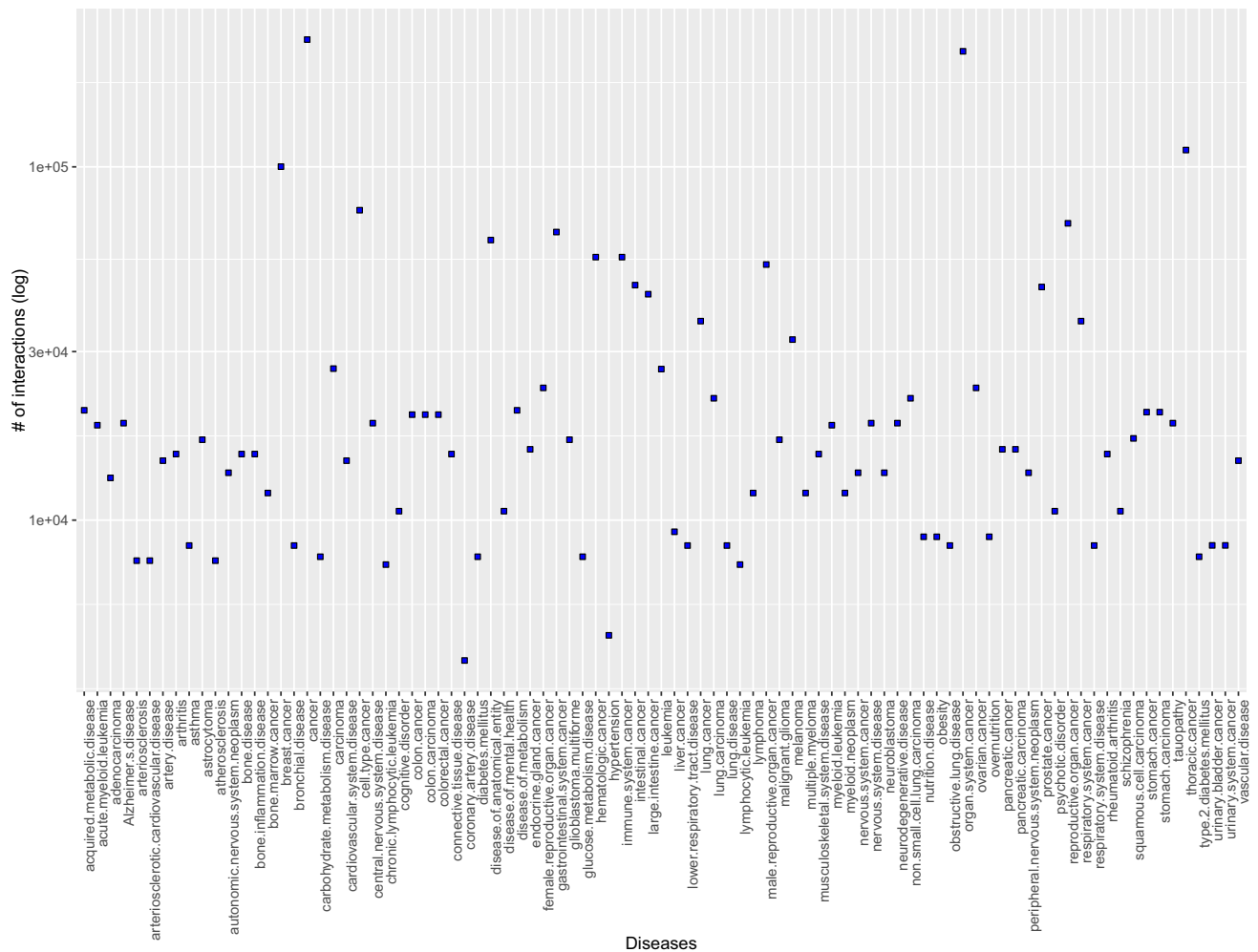
Figure 4. Disease distributions of human PPIs. PPIs are annotated with a disease if both interactors are annotated with the disease in DisGeNET.

PPIs are not static but rather occur in specific environments or conditions and change with time (53). We focused on two types of annotations that can change with time—localization and disease conditions. Localization, for example, is important because even if a PPI is reported in a database, if the two binding proteins do not share the same localization, the interaction is unlikely to happen *in vivo* (54). We added 13 localization annotations in this update, and Figure 3 shows the distribution of PPIs per species annotated with each localization. Finally, we annotated PPIs with 91 diseases based on DisGeNET (36). Available diseases are displayed as an ontology tree, and users can retrieve PPIs present in at least one or in multiple diseases of interest.

Comparison with other PPI resources

Compared to other PPI resources, IID is one of the broadest and largest physical interaction databases, and provides more options for reducing false negatives, specifying context, and analyzing networks (especially in non-human species). Several resources, including APID (55), HIPPIE v2.0 (44), HINT (56), iRefWeb (57), MyProteinNet (43), STRING (58) and TissueNet v.2 (42) provide some of the same functionality, but have important differences in their options for error-reduction, filtering by context, and network analysis.

Control of false positive rate is quite similar among these resources—all provide PPI scores, calculated in various ways, to indicate the reliability of PPIs. Reduction of the false negative rate is achieved by integration of PPIs from multiple databases that conduct literature curation. IID is the only PPI resource that also offers high-confidence predicted physically binding PPIs, which further reduce the false negative rate (e.g. for human, about two-thirds of available PPIs are predicted). Several databases, including STRING (58) and FunCoup (59), provide predictions for functional rather than physical interactions.

Filtering PPIs by context is supported by HIPPIE v2.0, MyProteinNet, and TissueNet v.2. All three provide filtering by tissue, HIPPIE v2.0 and MyProteinNet also provide filtering by Gene Ontology, and HIPPIE v2.0 provides filtering by disease as well. IID supports filtering by these contexts as well as by druggability, detailed brain structures and joint-related tissues. Users can specify whether PPIs can be in any of the selected contexts or should be present in all of them. Also, IID provides context filtering for the largest

number (17) of non-human species; HIPPIE v2.0 and TissueNet v.2 are available only for human, and MyProteinNet is available for 11 species.

Network analysis is supported by HIPPIE v2.0 and STRING. HIPPIE v2.0 analyses enrichment of disease and GO annotations of network proteins. STRING provides summary topology statistics for networks, and enrichment analysis of pathways and functions. IID provides both topology and enrichment analysis; it identifies important network nodes, and calculates enrichment of tissues, localizations, diseases, and druggability for network interactions, rather than network proteins.

CONCLUSION

IID helps address key challenges of using PPI data: high error rates, lack of context, and networks that are difficult to interpret. IID provides unique functionality for reducing false negatives by integrating multiple curated and high-confidence computationally-predicted interaction sources. It specifies context by using ontologies and multiple tissue, localization, disease, and drug-related data resources. It helps interpret returned networks by providing topological and enrichment analyses. Importantly, IID supports non-human species, many of which are vitally important in biomedical research but lack comprehensive, context-specific PPI networks. Future IID updates will focus on including more species, reliably transferring interaction information between species, and further expanding interaction annotations from ontologies and relevant data sets.

FUNDING

REFERENCES

1. Tian,W., Zhang,L.V., Taşan,M., Gibbons,F.D., King,O.D., Park,J., Wunderlich,Z., Cherry,J.M. and Roth,F.P. (2008) Combining guilt-by-association and guilt-by-profiling to predict Saccharomyces cerevisiae gene function. *Genome Biol. BioMed. Central*, **9**, S7.
2. Mostafavi,S. and Morris,Q. (2012) Combining many interaction networks to predict gene function and analyze gene lists. *Proteomics*, **12**, 1687–1696.
3. Navlakha,S. and Kingsford,C. (2010) The power of protein interaction networks for associating genes with diseases. *Bioinformatics*, **26**, 1057–1063.
4. Lee,I., Blom,U.M., Wang,P.I., Shim,J.E. and Marcotte,E.M. (2011) Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.*, **21**, 1109–1121.
5. Wang,Y.-C., Chen,B.-S., Parkin,D., Bray,F., Ferlay,J., Pisani,P. *et al.* (2011) A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. *BMC Med. Genomics. BioMed. Central*, **4**, 2.
6. Cun,Y. and Fröhlich,H. (2013) Network and data integration for biomarker signature discovery via network smoothed t-statistics. *PLoS One*, **8**, e73074.
7. Yeh,S.-H., Yeh,H.-Y. and Soo,V.-W. (2012) A network flow approach to predict drug targets from microarray data, disease genes and interactome network - case study on prostate cancer. *J. Clin. Bioinforma*, **2**, 1.
8. Isik,Z., Baldow,C., Cannistraci,C.V. and Schroeder,M. (2015) Drug target prioritization by perturbed gene expression and network information. *Sci. Rep.*, **5**, 17417.
9. Guney,E., Menche,J., Vidal,M. and Barábasi,A.-L. (2016) Network-based in silico drug efficacy screening. *Nat. Commun.*, **7**, 10331.
10. 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.
11. Kotlyar,M., Pastrello,C., Pivetta,F., Lo Sardo,A., Cumbaa,C., Li,H., Naranian,T., Niu,Y., Ding,Z., Vafaee,F. *et al.* (2015) In silico prediction of physical protein interactions and characterization of interactome orphans. *Nat Methods*, **12**, 79–84.
12. Vidal,M. (2016) How much of the human protein interactome remains to be mapped? *Sci. Signal.*, **9**, eg7.
13. Stumpf,M.P., Thorne,T., de Silva,E., Stewart,R., An,H.J., Lappe,M. and Wiuf,C. (2008) Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 6959–6964.
14. Venkatesan,K., Rual,J.-F., Vazquez,A., Stelzl,U., Lemmens,I., Hirozane-Kishikawa,T., Hao,T., Zenkner,M., Xin,X., Goh,K.I. *et al.* (2009) An empirical framework for binary interactome mapping. *Nat. Methods*, **6**, 83–90.
15. Emig,D., Kacprowski,T. and Albrecht,M. (2011) Measuring and analyzing tissue specificity of human genes and protein complexes. *EURASIP J. Bioinform. Syst. Biol.*, **2011**, 5.
16. Chatr-aryamontri,A., Oughtred,R., Boucher,L., Rust,J., Chang,C., Kolas,N.K., O'Donnell,L., Oster,S., Theesfeld,C., Sellam,A. *et al.* (2017) The BioGRID interaction database: 2017 update. *Nucleic Acids Res.*, **45**, D369–D379.
17. 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.
18. 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 AcidsRes.*, **37**, D767–D772.
19. Brown,K.R. and Jurisica,I. (2007) Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol.*, **8**, R95.
20. Breuer,K., Foroushani,A.K., Laird,M.R., Chen,C., Sribnaia,A., Lo,R., Winsor,G.L., Hancock,R.E., Brinkman,F.S. and Lynn,D.J. (2013) InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. *Nucleic Acids Res.*, **41**, D1228–D1233.
21. 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.
22. 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.
23. Lefebvre,C., Lim,W.K., Basso,K., Favera,R.D. and Califano,A. (2007) A Context-Specific network of Protein-DNA and Protein-Protein interactions reveals new regulatory motifs in Human B cells. In: *Systems Biology and Computational Proteomics*. Springer, Berlin, pp. 42–56.
24. Rhodes,D.R., Tomlins,S.A., Varambally,S., Mahavisno,V., Barrette,T., Kalyana-Sundaram,S., Ghosh,D., Pandey,A. and Chinnaiyan,A.M. (2005) Probabilistic model of the human protein–protein interaction network. *Nat. Biotechnol.*, **23**, 951–959.
25. Elefsinioti,A., ÖS,S., Hegele,A., Plake,C., Hubner,N.C., Poser,I., Sarov,M., Hyman,A., Mann,M., Schroeder,M., Stelzl,U. *et al.* (2011) Large-scale de novo prediction of physical protein–protein association. *Mol. Cell Proteomics*, **10**, doi:10.1074/mcp.M111.010629.

26. Zhang,Q.C., Petrey,D., Deng,L., Qiang,L., Shi,Y., Thu,C.A., Bisikirska,B., Lefebvre,C., Accili,D., Hunter,T. *et al.* (2012) Structure-based prediction of protein–protein interactions on a genome-wide scale. *Nature*, **490**, 556–560.

27. Zerbino,D.R., Achuthan,P., Akanni,W., Amode,M.R., Barrell,D., Bhai,J., Billis,K., Cummins,C., Gall,A., Girón,C.G. *et al.* (2018) Ensembl 2018. *Nucleic Acids Res.*, **46**, D754–D761.

28. Consortium,U. (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.

29. Bossi,A. and Lehner,B. (2009) Tissue specificity and the human protein interaction network. *Mol. Syst. Biol.*, **5**, 260.

30. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.

31. Gautier,L., Cope,L., Bolstad,B.M. and Irizarry,R.A. (2004) affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, **20**, 307–315.

32. Hawrylycz,M.J., Lein,E.S., Guillozet-Bongaarts,A.L., Shen,E.H., Ng,L., Miller,J.A., van de Lagemaat,L.N., Smith,K.A., Ebbert,A., Riley,Z.L. *et al.* (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, **489**, 391–399.

33. Ashburner,M., Ball,C., Blake,J., Botstein,D., Butler,H., Cherry,J., 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.

34. The Gene Ontology Consortium. (2017) Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.*, **45**, D331–D338.

35. Kibbe,W.A., Arze,C., Felix,V., Mitraka,E., Bolton,E., Fu,G., Mungall,C.J., Binder,J.X., Malone,J., Vasant,D. *et al.* (2015) Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.*, **43**, D1071–D1078.

36. Piñero,J., À,B., Queralt-Rosinach,N., Gutiérrez-Sacristán,A., Deu-Pons,J., Centeno,E., García-García,J., Sanz,F. and Furlong,L.I. (2017) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic. Acids. Res.*, **45**, D833–D839.

37. Bodenreider,O. (2004) The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.*, **32**, D267–D270.

38. Imming,P., Sinning,C. and Meyer,A. (2006) Drugs, their targets and the nature and number of drug targets. *Nat. Rev. Drug Discov.*, **5**, 821–834.

39. Wishart,D.S., Feunang,Y.D., Guo,A.C., Lo,E.J., Marcu,A., Grant,J.R., Sajed,T., Johnson,D., Li,C., Sayeeda,Z. *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.

40. Franz,M., Lopes,C.T., Huck,G., Dong,Y., Sumer,O. and Bader,G.D. (2015) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, **32**, btv557.

41. Snider,J., Kotlyar,M., Saraon,P., Yao,Z., Jurisica,I. and Stagljar,I. (2015) Fundamentals of protein interaction network mapping. *Mol. Syst. Biol.*, **11**, 848.

42. Basha,O., Barshir,R., Sharon,M., Lerman,E., Kirson,B.F., Hekselman,I. and Yeger-Lotem,E. (2017) The TissueNet v.2 database: A quantitative view of protein–protein interactions across human tissues. *Nucleic Acids Res.*, **45**, D427–D431.

43. Basha,O., Flom,D., Barshir,R., Smoly,I., Tirman,S. and Yeger-Lotem,E. (2015) MyProteinNet: build up-to-date protein interaction networks for organisms, tissues and user-defined contexts. *Nucleic Acids Res.*, **43**, W258–W263.

44. Alanis-Lobato,G., Andrade-Navarro,M.A. and Schaefer,M.H. (2017) HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res.*, **45**, D408–D414.

45. Guan,Y., Gorenshteyn,D., Burmeister,M., Wong,A.K., Schimenti,J.C., Handel,M.A., Bult,C.J., Hibbs,M.A. and Troyanskaya,O.G. (2012) Tissue-Specific functional networks for prioritizing phenotype and disease genes. *PLoS Comput. Biol.*, **8**, e1002694.

46. 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. Public Library Sci.*, **8**, e1002690.

47. Greene,C.S., Krishnan,A., Wong,A.K., Ricciotti,E., Zelaya,R.A., Himmelstein,D.S., Zhang,R., Hartmann,B.M., Zaslavsky,E., Sealfon,S.C. *et al.* (2015) Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.*, **47**, 569–576.

48. Djebbari,A., Ali,M., Otasek,D., Kotlyar,M., Fortney,K., Wong,S., Hrvojic,A. and Jurisica,I. (2011) NAViGaTOR: Large scalable and interactive navigation and analysis of large graphs. *Internet Math.*, **7**, 314–347.

49. He,X. and Zhang,J. (2006) Why do hubs tend to be essential in protein networks? *PLoS Genet.*, **2**, e88.

50. Schaefer,M.H., Serrano,L. and Andrade-Navarro,M.A. (2015) Correcting for the study bias associated with protein–protein interaction measurements reveals differences between protein degree distributions from different cancer types. *Front Genet.*, **6**, 260.

51. Pankevich,D.E., Altevogt,B.M., Dunlop,J., Gage,F.H. and Hyman,S.E. (2014) Improving and accelerating drug development for nervous system disorders. *Neuron*, **84**, 546–553.

52. Anandacoomarasamy,A. and March,L. (2010) Current evidence for osteoarthritis treatments. *Ther. Adv. Musculoskelet. Dis.*, **2**, 17–28.

53. Zhang,Y., Lin,H., Yang,Z., Wang,J., Liu,Y. and Sang,S. (2016) A method for predicting protein complex in dynamic PPI networks. *BMC Bioinformatics. BioMed. Central*, **17**, 229.

54. Veres D,V., Gyurkó,D.M., Thaler,B., Szalay,K.Z., Fazekas,D., Korcsmáros,T. and Csermely,P. (2015) ComPPI: a cellular compartment-specific database for protein–protein interaction network analysis. *Nucleic Acids Res.*, **43**, D485–D493.

55. Alonso-López,D., Gutiérrez,M.A., Lopes,K.P., Prieto,C., Santamaría,R. and De Las Rivas,J. (2016) APID interactomes: providing proteome-based interactomes with controlled quality for multiple species and derived networks. *Nucleic Acids Res.*, **44**, W529–W535.

56. Das,J. and Yu,H. (2012) HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst. Biol.*, **6**, 92.

57. 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.

58. Szklarczyk,D., Morris,J.H., Cook,H., Kuhn,M., Wyder,S., Simonovic,M., Santos,A., Doncheva,N.T., Roth,A., Bork,P. *et al.* (2017) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.*, **45**, D362–D368.

59. Ogris,C., Guala,D., Kaduk,M. and Sonnhammer,E.L.L. (2018) FunCoup 4: new species, data, and visualization. *Nucleic Acids Res.*, **46**, D601–D607.