Nucleic Acids Research, 2026, **54**, D593–D600
https://doi.org/10.1093/nar/gkaf1259
Advance access publication date: 26 November 2025
Database issue

IID 2025: Physical protein interaction data with detection types, co-purified protein sets, molecular docking, and immune cell networks

Max Kotlyar[1], Chiara Pastrello[1], Mark Abovsky[1], Alexandru Mizeranschi[2,3], Armand Keating[1,4], Luiz-Claudio Cameron[1,5], Vinod Chandran[5,6,7,8], Igor Jurisica[1,9,10,*]

[1]Osteoarthritis Research Program, Division of Orthopedic Surgery, Schroeder Arthritis Institute and Data Science Discovery Centre for Chronic Diseases, Krembil Research Institute, University Health Network, Toronto, ON M5T 0S8, Canada
[2]Research and Development Station for Bovine Arad, Arad 310059, Romania
[3]Institute for Advanced Environmental Research, West University of Timisoara, Timisoara 300086, Romania
[4]Princess Margaret Cancer Centre, Toronto, ON M5G 2M9, Canada
[5]Gladman Krembil Psoriatic Arthritis Program, Schroeder Arthritis Institute, University Health Network, Toronto, ON M5T 0S8, Canada
[6]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada
[7]Division of Rheumatology, Department of Medicine, University of Toronto, Toronto, Canada
[8]Institute of Medical Science, University of Toronto, Toronto, ON, Canada
[9]Departments of Medical Biophysics and Computer Science, and Faculty of Dentistry, University of Toronto, Toronto, ON, Canada
[10]Institute of Neuroimmunology, Slovak Academy of Sciences, Dúbravská cesta 9, Bratislava 845 10, Slovakia

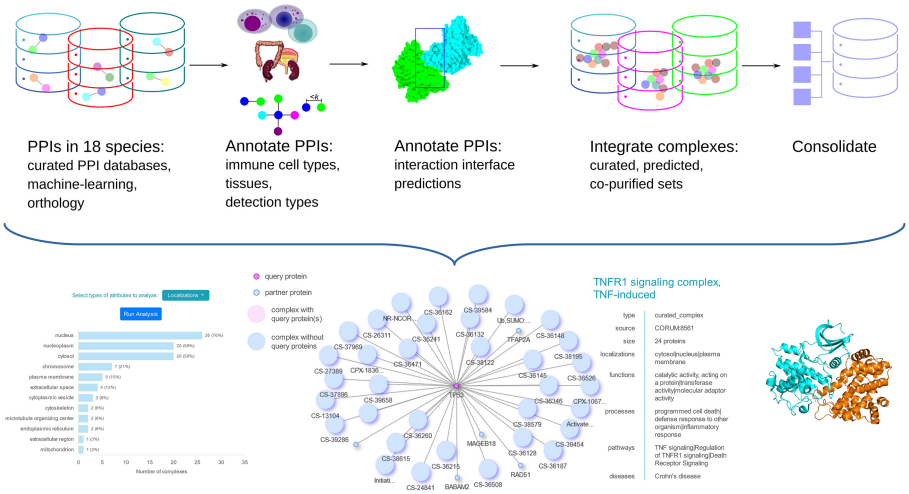
*To whom correspondence should be addressed. Email:juris@ai.utoronto.ca

Abstract

Biomedical research benefits from the rapid growth and diversity of experimentally detected protein–protein interactions (PPIs) by gaining important biological insights. However, increasingly dense PPI networks can be challenging to interpret and apply. The 2025 update of the Integrated Interactions Database (IID) enhances accessibility and utility through several new features. We identify and incorporate network structural components from co-purified protein sets, as well as curated and predicted complexes, enabling users to explore network organization beyond binary interactions. Functional, pathway, and disease associations of these components can be analyzed, enabling interactions to be grouped into higher-order structures with known or provisional biological roles. Users can now filter interactions by five detection types: pairwise, co-purification, colocalization, proximity, and other evidence. To extend the value and information of predicted interactions, we include interaction interface predictions for 53 647 PPIs, generated using the MEGADOCK docking algorithm, adding molecular detail for structural biology and variant impact studies. Finally, we map PPIs to 15 immune cell types and 12 additional normal tissues, offering tissue-specific views of interaction networks increasingly relevant in disease and immunology research. IID 2025 now includes over 1 million experimentally detected human PPIs, representing an 83% increase from the previous release, alongside expanded non-human networks. The portal remains publicly available at https://ophid.utoronto.ca/iid.

Received: September 15, 2025. **Revised:** October 22, 2025. **Accepted:** October 23, 2025

Graphical abstract



Introduction

Protein–protein interaction (PPI) networks are foundational resources in systems biology, offering insights into disease mechanisms [1], drug discovery [2], and the functional relationships among genes identified in experimental studies [3]. These networks have been applied to diverse problems using visualization and computational methods such as random walks [4], network flow [5], and deep learning [6]. However, several persistent challenges limit their full potential, interpretability, and utility.

Two long-term trends in experimentally detected PPI data complicate network analysis. First, the number of reported human PPIs has been doubling approximately every 4 years since 2010, now exceeding 1.2 million (Fig. 1)—far beyond early estimates of the human interactome [7]. Second, a growing proportion of these interactions are derived from methods such as affinity purification mass spectrometry (AP-MS) [8], which identify sets of co-purified proteins rather than direct binary interactions. These sets are typically converted into pairwise interactions using spoke models, resulting in networks dominated by indirect associations.

These trends, combined with other limitations—including lack of biological context (e.g., tissue or developmental stage) and absence of detailed binding information—pose significant challenges for analysis, visualization, and interpretation. Networks often contain thousands of interactions per protein, leading to dense, cluttered graphs that obscure meaningful patterns. Even small subnetworks can be difficult to interpret, especially when they include a mixture of direct and indirect interactions. In cases where spoke models dominate, visualizations may reflect experimental biases more than true biological relationships. These issues also affect network analysis. The high number of PPIs and prevalence of indirect associations compress network distances between proteins, reducing the discriminative power of path-based analyses. Moreover, algorithms may incorrectly assume that most edges represent direct binding and produce misleading results. Lack of biological context further complicates visualization, analysis, and meaningful interpretation. Without information about tissue, cell type, developmental stage, or disease context, networks may conflate interactions from distinct biological settings, reducing interpretability and relevance for translational research. Finally, the absence of detailed binding information—such as interaction interfaces or affinities—limits the utility of PPI networks for studies focused on disease mechanisms or drug discovery, where such data are often critical.

To reduce these limitations, the current update introduces several features that enhance the interpretability and analytical value of PPI networks. Users can now filter interactions by type—including pairwise, co-purified sets, proximity, and colocalization—allowing selection of data that aligns with specific analysis assumptions and improves clarity in visualizations. This filter can be combined with others, such as evidence thresholds (e.g., interactions supported by multiple publications) and biological context (e.g., tissue-specific networks). Retrieved networks now include curated and predicted complexes, and co-purified protein sets from detection methods such as AP-MS. Here, we use the term *complexes* to refer to all three of these protein sets collectively. We annotate complexes with disease, gene ontology, and pathway associations, and determine if they are enriched in retrieved networks. Their inclusion reduces clutter by grouping related nodes and edges while also segmenting networks into functional units that reflect coordinated biological roles. To improve contextual relevance, we add annotations for 12 normal tissues and 15 immune cell types, enabling more precise filtering and interpretation (Supplementary Table 1). Finally, we provide interaction interface predictions for 53 647 protein pairs using the MEGADOCK docking algorithm [9], offering molecular detail that supports structural biology applications and enhances studies of variant impact and drug targeting.

Materials and methods

PPI sources

Experimentally detected PPIs were obtained primarily from seven databases: BioGRID [10] 4.4.244, DIP [11], HPRD [12] Release 9, InnateDB [13] version 5.4, IntAct [14] release 250, MatrixDB [15] version 4.0, and MINT [16]. All databases were downloaded on 16 April 2025. Predicted PPIs were obtained from five studies, as previously described [17]. Orthologous PPIs were generated by mapping experimentally detected

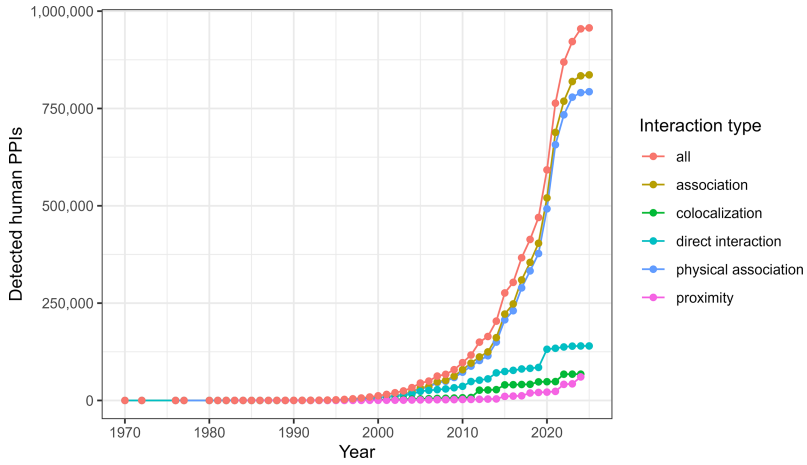


Figure 1. Cumulative number of experimentally detected human PPIs from 1970 to 2025.

PPIs in each of the 18 IID species to orthologous protein pairs in the other 17 species. Mapping was done using 1:1 orthologs downloaded from Ensembl [18] release 114.

Mapping between gene and protein IDs

Mappings between various gene and protein IDs were based on UniProt [19] release 05 February 2025. For a more complete set of mappings between Ensembl and UniProt IDs, mappings from Ensembl [18] release 114 were also used.

Annotation of PPIs with normal tissues and immune cell types

Gene expression in 12 normal human tissues was obtained from the following GEO datasets: GSE3678 [20], GSE4183 [21], GSE8607 [22], GSE8671 [23], GSE9576 [24], GSE11783 [25], GSE14905 [26], GSE24152 [27], GSE53757 [28], GSE63678 [29], GSE65144 [30], and GSE166388 [31]. A PPI was annotated with a tissue if both encoding genes were expressed in that tissue, defined as having MAS5-normalized expression values above 200, following the criteria used by Bossi *et al.* [32].

Gene expression in 15 immune cell types was obtained from the DICE [Database of Immune Cell Expression, Expression quantitative trait loci (eQTLs) and Epigenomics] project [33, 34] build 23 February 2022. A PPI was annotated with a cell type if both encoding genes had a median transcripts per million (TPM) value greater than 1 in that cell type.

Annotation of PPIs with detection types

PPIs were classified into five experimental detection types: pairwise, co-purified set, co-localization, proximity, and other. Interactions were annotated as *pairwise* if only two proteins were reported in the detection, though such interactions were not necessarily direct. PPIs were labeled as part of a *co-purified set* when more than two proteins were detected and the interaction type in the source database was neither co-localization nor proximity. *Co-localization* and *proximity* annotations were assigned based on interaction type labels provided by the source database. Interactions that did not meet criteria for any of the four categories were classified as *other*.

Molecular docking

Computational modeling of interacting protein pairs was carried out using molecular docking techniques. Three-dimensional structures of human proteins were sourced from the AlphaFold Protein Structure Database version 2 (https://alphafold.ebi.ac.uk) [35, 36]. Interaction modeling was performed with MEGADOCK version 4.1.1 docking software (https://www.bi.cs.titech.ac.jp/megadock/ppi.html) [9]. Visualization and interpretation of the docking results were conducted using ChimeraX version 1.10.1 (24 July 2025) visualization software (https://www.rbvi.ucsf.edu/chimerax) [37].

MEGADOCK employs a rigid-docking methodology along with a scoring function that relies on shape complementarity, electrostatics, and desolvation free energy. The computational algorithms are tasked with identifying the optimal docking pose by examining 10 800 potential orientations and positions between the two proteins. An example of such analysis is provided in [38].

Protein complexes

Data sources. Curated complexes were downloaded from Complex Portal [39] release 248, Corum [40] release 5.1, and Reactome [41] release 92. Predicted complexes were obtained from hu.MAP3.0 [42]. Co-purified sets were generated from PPIs comprising spoke models in BioGRID [10], DIP [11], InnateDB [13], MINT [16], and IntAct [14]. A co-purified set was the union of proteins in a spoke model.

Filtering complexes. To simplify and characterize PPI networks, complexes were selected based on size and redundancy criteria. Curated complexes were required to contain at least five proteins to exclude very small groups that contribute little to network simplification while retaining the majority of curated complexes, which are valuable for network characterization due to their well-established biological roles. Predicted complexes and co-purified protein sets were filtered to include those with 50–350 proteins. The higher minimum threshold reflects their typically larger size and the need for reliable enrichment analysis, which is more effective for larger groups. The maximum size of 350 was chosen to align with the largest curated complexes (e.g., https://reactome.org/content/detail/R-HSA-975037) and to exclude overly broad groupings that may lack specificity.

To reduce redundancy, complexes were filtered in two steps. First, complexes were prioritized largest-to-smallest in the fol-

lowing order: curated complexes, predicted complexes, and co-purified sets. Second, any complex with >75% similarity to a higher-priority complex—measured as the ratio of intersection to union—was removed. This approach ensures a diverse, non-overlapping set of complexes that can effectively simplify network structure and enhance biological interpretability.

Annotating complexes. Complexes and co-purified protein sets were annotated with disease, function, and pathway information to support biological interpretation. Disease annotations were sourced from the Diseases 2.0 database [43], while Gene Ontology annotations (GOSlim categories for Biological Process, Cellular Component, and Molecular Function) were obtained from the Gene Ontology Consortium [44] and pathway annotations from the PathDIP database [45] version 5. Disease associations were based on experimental and knowledge-channel data in Diseases 2.0 database [43]. A complex was annotated with a disease if at least 10% of its proteins—and a minimum of five—were associated with that disease, and the association was statistically significant ($P < .05$) based on a hypergeometric test. Gene Ontology and pathway annotations followed even stricter criteria: at least 33% of proteins—and a minimum of five—had to be annotated with the category, and the enrichment had to be significant ($P < .05$). These thresholds were chosen to balance sensitivity with confidence, ensuring that annotations reflect meaningful biological associations.

Associating Complexes with PPI Networks. Complexes were associated with PPI networks using two criteria: (a) ensuring reliable association of each complex with the network and (b) limiting redundancy among selected complexes. To establish reliable association, a complex was required to have at least 50% of its proteins present in the network, and the network had to show significant enrichment for the complex members ($P < .05$, hypergeometric test). To reduce overlap among selected complexes, a two-step filtering process was applied. First, complexes were prioritized largest-to-smallest as follows: curated complexes, predicted complexes, and co-purified sets. Second, each complex was evaluated against the current network; if it met the association criteria, its proteins were removed from the network before evaluating the next complex. This iterative approach ensured that selected complexes were distinct and not repeatedly associated with the same subset of network proteins.

Network visualization

To enable real-time visualization of large PPI networks, we implemented a filtering strategy that displays a central subgraph comprising up to 50 nodes. Nodes are prioritized based on the following criteria:

1. Number of neighbors that are query nodes.
2. Preference for displaying query nodes over partner nodes.
3. Total number of neighbors per node.

This approach ensures that the most relevant and densely connected nodes are visualized while maintaining performance and readability. The subgraph is rendered using the vis.js JavaScript library (version 4.21; https://visjs.org).

Results

IID 2025 includes over 1 million experimentally detected human PPIs, representing an 83% increase from the previous release, alongside expanded non-human networks. This dramatic growth underscores the need for extended annotations and improved tools to interpret and apply increasingly complex interaction networks. To address this, the current update introduces four major new features designed to enhance the interpretability and utility of PPI networks: (i) integration of curated complexes, predicted complexes, and co-purified protein sets into retrieved networks; (ii) filtering of PPIs by detection type; (iii) annotation of PPIs with tissues and immune cell types; and (iv) provision of detailed binding interface predictions. These enhancements reduce challenges posed by the rapid expansion of PPI data and the growing prevalence of indirect interactions—particularly those derived from spoke models applied to co-purified protein sets from methods such as AP-MS. While these trends produce networks with much greater information content, they make network interpretation and downstream analysis more complex and challenging.

Complexes in Retrieved Networks. The central new feature is the identification and analysis of complexes within retrieved human networks. Identified complexes include curated complexes [39–41] ranging from 5 to 315 proteins, predicted complexes [42], and co-purified sets ranging from 50 to 335 proteins. Complex-related results are presented in three sections: a top panel that visualizes complexes within the network, a table displaying complex properties, and an analysis panel below the table offering summary tools. The table includes key attributes such as source database, complex type (curated, predicted, co-purified), enrichment annotations (disease, Gene Ontology GOSlim categories, and pathways), size, fraction present in the network, and enrichment P-value. Users can sort the table by any column and access detailed information about a complex by clicking on its name or ID. The analysis panel enables summarization of annotation frequencies across complexes, such as the prevalence of disease associations.

To illustrate the utility of complex-level information, we examined PALB2 and IQGAP1 (Q86YC2, P46940)—two proteins recently shown to colocalize in breast cancer, both when PALB2 is wild type and (even more) when it is mutated [46]. As part of the follow-up questions to this discovery, it would be important to know which functions the two proteins exert together. Querying IID for the two proteins returns 1145 interactions. This is a large number and it might be challenging to interpret the functions related to each interaction. Looking at the complexes, though, we can see that the interactions are grouped in 14 complexes, leaving out only 23 proteins. Ten of the proteins and eight of the complexes interact with both proteins, making it easier to focus on specific groups and identify closely related known or novel functions. For example, one of the shared complexes, R-HSA-5683737 (https://reactome.org/content/detail/R-HSA-5683737), is present when a DNA double-strand break occurs and needs to be repaired. This is expected, as PALB2 is a known player in the homologous recombination pathway [46], and IQGAP1 is a known regulator of BRCA1 localization [47].

Filtering of PPIs by detection type. PPIs can be filtered by five detection types: pairwise, co-purified set, colocalization, proximity, and other. These options are available in the *Search Options* page under 'Filter interactions by evidence,' alongside filters for interaction type and evidence quantity. When multiple detection types are selected, users can choose between two filtering modes: the default mode retrieves PPIs that match *any* of the selected types (e.g., selecting 'colocalization' and 'proximity' will return interactions detected by

either method), while the strict mode retrieves only PPIs that match *all* selected types (e.g., selecting both 'colocalization' and 'proximity' will return only interactions supported by both detection methods). This flexible filtering system enables users to tailor network retrieval to match specific analytical assumptions or visualization goals.

Annotation of PPIs with tissue and immune cell context. PPIs are annotated with additional expression data from 12 normal human tissues and 15 immune cell types (Supplementary Table 1). These annotations are available as filtering options in the *Search Options* page under 'Filter interactions by context and other properties' → 'Tissues & cell types.' As with other filters, multiple selected tissues or cell types can be applied in two modes: the default mode retrieves PPIs expressed in *any* of the selected contexts, while the strict mode retrieves only PPIs expressed in *all* selected contexts. Alternatively, tissues and cell types can be specified as annotation criteria—rather than filters—using the 'Specify PPI annotations to retrieve' section of the *Search Options* page.

Tissue-specific analyses can help identify mechanisms linked to only one or multiple cell types. In IID 2025 we include detailed immune cell annotations together with the previous detailed brain and joint information. Considering the immune cell-specific Protein Atlas (https://www.proteinatlas.org/humanproteome/single+cell/immune+cell), several sets of proteins are specifically expressed in only one type of immune cell. For our example, we selected genes specific for NK cells ($n = 65$) and queried them in IID, retrieving immune cells-specific annotations. Of the 9051 retrieved PPIs, 10 are specific for NK cells with the two proteins annotated with NK cells, while 417 are NK specific with only one protein annotated with NK cells. In contrast, 881 interactions were present and had both proteins of the interaction annotated with all the immune cells available in IID. Thus, IID annotation can provide useful insights into the mechanisms affecting the behavior of immune cells, as well as their specific processes. To further interpret the 10 PPIs specific for NK cells, we note that four were self-interactions and three of the remaining involved NCAM1, also known as CD56, a key marker of NK cells [48, 49]. All three such interactions are predicted, leading to newer possible molecular mechanisms to explore. For example, one of the interactors is ITGAD, also known as CD11d, a β2 integrin implicated in leukocyte migration [50] and, more recently, shown to affect the migration of NK cells [51]. This opens the possibility of exploring mechanisms underlying the modulation of NCAM1-mediated NK cell activation [52] and CD11d-driven migration and adhesion to better understand, for example, tissue-specific inflammation.

Binding interfaces. We performed molecular docking and computational modeling to enhance understanding of interactions involving proteins with few experimentally detected partners. Availability of docking data can be selected as a filtering option in the *Search Options* page under 'Filter by interaction context, properties, and docking data' → 'Docking.' If selected, the table of retrieved PPIs includes a column of links to docking results. The top-ranking model is illustrated on the results webpage, with gold color denoting the receptor and the green color signifying the ligand, which is available for download in PDB format. Molecular docking analysis uncovers the interacting residues and atomic contact points. Detailed residue and atom-level contact information can be accessed on a linked web page, identifying the amino acids that are likely to play a role in the interface pertinent to the exploration of complex biological phenomena at a molecular scale, thereby pinpointing potential targets for drug development and the design of PPI inhibitors.

Utilizing AlphaFold predicted structures as the foundation for docking and ensuring an acceptable level of accuracy with the MEGADOCK scoring function enhances the reliability of the modeling. However, the continued use of a rigid-body approach, constrained by current computational abilities, which does not consider the conformational changes that may occur during protein binding, presents a limitation of the method, and its reliability is contingent upon the specific proteins under investigation. Further modeling with molecular dynamics software can help resolve protein flexibility by simulating how the protein moves, while experimental data showing multiple 'secondary' docking poses can suggest that the scoring function has local energy minima.

Examples of advanced search queries. It is possible to combine all the annotations and the filters to perform simpler or more complex queries. For example, a researcher could query two proteins with immune-related functions (i.e., ICAM1 and DEFA4) and search for all the interactors of such proteins that: (i) are only predicted (to identify novel interactions); (ii) involve at least one protein present in the plasma membrane (to work on signaling mechanisms); (iii) have docking results available (to prioritize further experimental validations); and finally (iv) provide immune cell annotations. This would lead to 52 PPIs that the researcher could use for their investigations. Similarly, a user can query IID for PPIs of TP53 that: (i) have been detected with pairwise experiments (to focus on direct interactions); (ii) are present both in normal lung and lung cancer (to focus on lung-related mechanisms); (iii) are targeted by drugs (to prioritize interactions that could lead to possible treatment); and (iv) have mutations described to decrease or disrupt the interactions (to identify disrupted mechanisms). This would lead to eight interactions that a researcher could prioritize for the follow-up *in vitro* and *in vivo* experiments .

Discussion

Comparison with other PPI resources

The features introduced in this update to IID are largely absent or only partially implemented in other publicly available PPI databases (Table 1). Notably, the integration of protein complexes and predicted interaction interfaces into retrieved PPI results is unique to IID. While filtering by immune cell types is supported in InnateDB [13], the specific cell types available are not listed. Tissue-based filtering is offered by HIPPIE [53], InnateDB [13], MyProteinNet [54], and TissueNet [55], though IID provides more extensive options. Filtering by detection method or interaction type is available in HINT [56], HIPPIE [53], InnateDB [13], and IntAct [14].

Another key differentiating feature of IID is its inclusion of high-confidence computationally predicted PPIs. Although the majority of human PPIs in IID are experimentally detected, the option to incorporate predicted interactions enables users to identify potentially missing links—false negatives—particularly for proteins with few known partners. The sustained rapid growth in experimentally detected PPIs suggests that a substantial fraction of interactions remains undiscovered. Predicted interactions in IID have demonstrated utility in bridging this gap (Fig. 2). Derived from computational studies published between 2005 and 2015, these predictions have been consistently validated by subsequent experimental findings over the past decade.

Table 1. Availability of newly introduced IID features across PPI databases

Database	Last update	Integrated complex information	Filtering by detection types	Filtering by tissues	Filtering by immune cell types	Predicted interaction interfaces
APID [57]	2021	no	no	no	no	no
BioGRID [10]	2025	no	no	no	no	no
HINT [56]	2024	no	yes	no	no	no
HIPPIE [53]	2022	no	yes	yes	no	no
InnateDB [13]	2019	no	yes	yes	yes	no
IntAct [14]	2025	no	yes	no	no	no
MyProteinNet [54]	2015	no	no	yes	no	no
STRING [58]	2025	no	no	no	no	no
TissueNet [55]	2022	no	no	yes	no	no

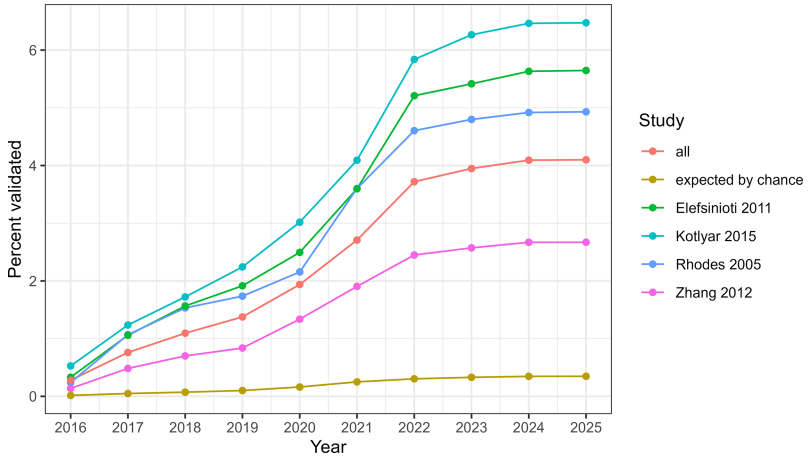


Figure 2. Validation of predicted PPIs over time: percentage of interactions from four prediction studies (2005–2015) confirmed by experimental detection between 2016 and 2025.

Additional distinguishing features of IID include its extensive PPI annotations and support for large-scale, multi-gene queries. Annotations span a wide range of biological contexts, including 158 tissues and cell types, six developmental stages, and 60 disease states. IID also provides information on interaction properties such as evolutionary conservation, mutation impact, and druggability. The web portal is optimized for batch queries, allowing users to input hundreds of gene or protein identifiers, retrieve richly annotated datasets, and apply integrated analysis tools, or download the annotated interactions for additional analysis.

Addressing dataset biases and inconsistencies

PPI datasets are subject to multiple sources of bias and inconsistency, including differences in detected relationship types (e.g., direct interaction, colocalization), false positive and false negative rates, bioassay-specific limitations, and research focus biases. As IID integrates data from multiple sources, retrieved interactions inevitably reflect these differences—raising important questions about reliability, biological relevance, and appropriate use.

To address these challenges, IID emphasizes transparency and user control. Rather than assigning quality scores to datasets or interactions—which can be misinterpreted due to trade-offs between false positives and false negatives—IID presents detailed evidence summaries for each interaction. These include detection type, supporting studies and bioassays, and additional support from computational prediction and orthology. Users can filter interactions based on the number of supporting studies or bioassays, evidence or detection type, and biological context, enabling tailored network construction aligned with specific analytical goals. However, as these filters change the type and number of interactions, users need to explicitly describe the choices to ensure reproducibility.

To help reduce false negatives, IID includes high-confidence predicted interactions. These are particularly valuable for proteins with few detected partners—such as the 301 human proteins in IID that currently have only predicted interactions.

Future directions

Future development of IID will focus on expanding biological context and structural insight to further enhance network interpretability. We plan to integrate pathway-level information to clarify the functional organization of interaction networks. Contextual annotation will be significantly broadened using bulk and single-cell RNA sequencing datasets, enabling more precise mapping of interactions across diverse tissues and cell types. Additionally, we aim to increase the coverage of predicted binding interfaces through large-scale molecular docking, providing deeper insight into interaction mechanisms and potential therapeutic targets.

Conclusions

The latest update to IID enhances the interpretability and utility of PPI data by introducing complex-level organization, filtering by detection type, expanded context, and predicted interaction interfaces. These features facilitate functional exploration of large interaction networks and enable more tar-

geted biological insights, as demonstrated through real-world use cases. By combining experimental data with context annotations and computational predictions, IID supports both discovery-driven research and translational applications in systems biology.

Acknowledgements

Author contribution: Max Kotlyar: Conceptualization, Formal analysis, Methodology, Validation, Writing—original draft. Chiara Pastrello: Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing—review & editing. Mark Abovsky: Formal analysis, Methodology, Writing—review & editing. Alexandru Mizeranschi: Methodology, Validation. Armand Keating: Validation, Writing—review & editing. Luiz-Claudio Cameron: Formal analysis, Methodology, Writing—review & editing. Vinod Chandran: Validation, Writing—review & editing. Igor Jurisica: Conceptualization, Formal analysis, Funding, Methodology, Validation, Writing—review & editing.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported in part by funding from Natural Sciences Research Council (NSERC) (RGPIN-2024-04314), CIHR (#519474), Canada Foundation for Innovation (CFI #225404, #30865), and Ontario Research Fund (RDI #34876, RE010-020). Vinod Chandran is supported by the Dr. Dafna D. Gladman Chair in Psoriatic Arthritis Research, a joint Hospital-University Named Chair between the University of Toronto, the University Health Network, and the UHN Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Funding to pay the Open Access publication charges for this article was provided by Natural Sciences Research Council (NSERC) (RGPIN-2024-04314).

Data availability

The IID portal is publicly available at https://ophid.utoronto.ca/iid.

References

1. Navlakha S, Kingsford C. The power of protein interaction networks for associating genes with diseases. *Bioinformatics* 2010;26:1057–63. https://doi.org/10.1093/bioinformatics/btq076
2. Isik Z, Baldow C, Cannistraci CV *et al.* Drug target prioritization by perturbed gene expression and network information. *Sci Rep* 2015;5:17417. https://doi.org/10.1038/srep17417
3. Mostafavi S, Morris Q. Combining many interaction networks to predict gene function and analyze gene lists. *Proteomics* 2012;12:1687–96. https://doi.org/10.1002/pmic.201100607
4. Smedley D, Köhler S, Czeschik JC *et al.* Walking the interactome for candidate prioritization in exome sequencing studies of Mendelian diseases. *Bioinformatics* 2014;30:3215–22. https://doi.org/10.1093/bioinformatics/btu508
5. Jeong H, Qian X, Yoon B-J. Effective comparative analysis of protein–protein interaction networks by measuring the steady-state network flow using a Markov model. *BMC Bioinformatics* 2016;17:395. https://doi.org/10.1186/s12859-016-1215-2
6. Cui J, Yang S, Yi L *et al.* Recent advances in deep learning for protein–protein interaction: a review. *BioData Mining* 2025;18:43. https://doi.org/10.1186/s13040-025-00457-6
7. Stumpf MPH, Thorne T, de Silva E *et al.* Estimating the size of the human interactome. *Proc Natl Acad Sci USA* 2008;105:6959–64. https://doi.org/10.1073/pnas.0708078105
8. Meyer K, Selbach M. Quantitative affinity purification mass spectrometry: a versatile technology to study protein-protein interactions. *Front Genet* 2015;6:237. https://doi.org/10.3389/fgene.2015.00237
9. Hayashi T, Matsuzaki Y, Yanagisawa K *et al.* MEGADOCK-Web: an integrated database of high-throughput structure-based protein–protein interaction predictions. *BMC Bioinformatics* 2018;19:62. https://doi.org/10.1186/s12859-018-2073-x
10. Oughtred R, Rust J, Chang C *et al.* The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci* 2021;30:187–200. https://doi.org/10.1002/pro.3978
11. Salwinski L, Miller CS, Smith AJ *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* 2004;32:D449–51. https://doi.org/10.1093/nar/gkh086
12. Keshava Prasad TS, Goel R, Kandasamy K *et al.* Human Protein Reference Database—2009 update. *Nucleic Acids Res* 2009;37:D767–72. https://doi.org/10.1093/nar/gkn892
13. Breuer K, Foroushani AK, Laird MR *et al.* InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res* 2013;41:D1228–33. https://doi.org/10.1093/nar/gks1147
14. Orchard S, Ammari M, Aranda B *et al.* The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucl Acids Res* 2014;42:D358–63. https://doi.org/10.1093/nar/gkt1115
15. Samarasinghe KW, Kotlyar M, Vallet SD *et al.* MatrixDB 2024: an increased coverage of extracellular matrix interactions, a new Network Explorer and a new web interface. *Nucleic Acids Res* 2025;53:D1677–82. https://doi.org/10.1093/nar/gkae1088
16. Licata L, Briganti L, Peluso D *et al.* MINT, the molecular interaction database: 2012 Update. *Nucleic Acids Res* 2012;40:D857–61. https://doi.org/10.1093/nar/gkr930
17. Kotlyar M, Pastrello C, Malik Z *et al.* IID 2018 update: context-specific physical protein–protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res* 2019;47:D581–9. https://doi.org/10.1093/nar/gky1037
18. Dyer SC, Austine-Orimoloye O, Azov AG *et al.* Ensembl 2025. *Nucleic Acids Res* 2025;53:D948–57. https://doi.org/10.1093/nar/gkae1071
19. Consortium The UniProt. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res* 2025;53:D609–17. https://doi.org/10.1093/nar/gkae1010
20. Reyes I, Reyes N, Suriano R *et al.* Gene expression profiling identifies potential molecular markers of papillary thyroid carcinoma. *CBM* 2019;24:71–83. https://doi.org/10.3233/CBM-181758
21. Gyorffy B, Molnar B, Lage H *et al.* Evaluation of microarray preprocessing algorithms based on concordance with RT-PCR in clinical samples. *PLoS One* 2009;4:e5645. https://doi.org/10.1371/journal.pone.0005645
22. Gashaw I, Grümmer R, Klein-Hitpass L *et al.* Gene signatures of testicular seminoma with emphasis on expression of ets variant gene 4. *Cell Mol Life Sci* 2005;62:2359–68. https://doi.org/10.1007/s00018-005-5250-9
23. Sabates-Bellver J, Van der Flier LG, de Palo M *et al.* Transcriptome profile of human colorectal adenomas. *Mol Cancer Res*

2007;5:1263–75. https://doi.org/10.1158/1541-7786.MCR-07-0267
24. Leja J, Essaghir A, Essand M *et al.* Novel markers for enterochromaffin cells and gastrointestinal neuroendocrine carcinomas. *Mod Pathol* 2009;22:261–72. https://doi.org/10.1038/modpathol.2008.174
25. Gamper M, Viereck V, Geissbühler V *et al.* Gene expression profile of bladder tissue of patients with ulcerative interstitial cystitis. *Bmc Genomics* 2009;10:199. https://doi.org/10.1186/1471-2164-10-199
26. Yao Y, Richman L, Morehouse C *et al.* Type I interferon: potential therapeutic target for psoriasis? *PLoS One* 2008;3:e2737. https://doi.org/10.1371/journal.pone.0002737
27. Zhang Z, Furge KA, Yang XJ *et al.* Comparative gene expression profiling analysis of urothelial carcinoma of the renal pelvis and bladder. *BMC Med Genomics* 2010;3:58. https://doi.org/10.1186/1755-8794-3-58
28. von Roemeling CA, Radisky DC, Marlow LA *et al.* Neuronal pentraxin 2 supports clear cell renal cell carcinoma by activating the AMPA-selective glutamate receptor-4. *Cancer Res* 2014;74:4796–810. https://doi.org/10.1158/0008-5472.CAN-14-0210
29. Pappa KI, Polyzos A, Jacob-Hirsch J *et al.* Profiling of discrete gynecological cancers reveals novel transcriptional modules and common features shared by other cancer types and embryonic stem cells. *PLoS One* 2015;10:e0142229. https://doi.org/10.1371/journal.pone.0142229
30. von Roemeling CA, Marlow LA, Pinkerton AB *et al.* Aberrant lipid metabolism in anaplastic thyroid carcinoma reveals stearoyl CoA desaturase 1 as a novel therapeutic target. *J Clin Endocrinol Metab* 2015;100:E697–709. https://doi.org/10.1210/jc.2014-2764
31. Qiu X, Zheng L, Liu X *et al.* ULK1 inhibition as a targeted therapeutic strategy for psoriasis by regulating keratinocytes and their crosstalk with neutrophils. *Front Immunol* 2021;12:714274. https://doi.org/10.3389/fimmu.2021.714274
32. Bossi A, Lehner B. Tissue specificity and the human protein interaction network. *Mol Syst Biol* 2009;5:260. https://doi.org/10.1038/msb.2009.17
33. Schmiedel BJ, Singh D, Madrigal A *et al.* Impact of genetic polymorphisms on human immune cell gene expression. *Cell* 2018;175:1701–15. https://doi.org/10.1016/j.cell.2018.10.022
34. Chandra V, Bhattacharyya S, Schmiedel BJ *et al.* Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. *Nat Genet* 2021;53:110–9. https://doi.org/10.1038/s41588-020-00745-3
35. Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–9. https://doi.org/10.1038/s41586-021-03819-2
36. Varadi M, Bertoni D, Magana P *et al.* AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* 2024;52:D368–75. https://doi.org/10.1093/nar/gkad1011
37. Meng EC, Goddard TD, Pettersen EF *et al.* UCSF ChimeraX: tools for structure building and analysis. *Protein Sci* 2023;32:e4792. https://doi.org/10.1002/pro.4792
38. Yuan Z, Ostrowska-Podhorodecka Z, Cox T *et al.* Annexin A2 contributes to release of extracellular vimentin in response to inflammation. *FASEB J* 2025;39:e70621. https://doi.org/10.1096/fj.202500793R
39. Balu S, Huget S, Medina Reyes JJ *et al.* Complex portal 2025: predicted human complexes and enhanced visualisation tools for the comparison of orthologous and paralogous complexes. *Nucleic Acids Res* 2025;53:D644–50. https://doi.org/10.1093/nar/gkae1085
40. Steinkamp R, Tsitsiridis G, Brauner B *et al.* CORUM in 2024: protein complexes as drug targets. *Nucleic Acids Res* 2025;53:D651–7. https://doi.org/10.1093/nar/gkae1033
41. Milacic M, Beavers D, Conley P *et al.* The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res* 2024;52:D672–8. https://doi.org/10.1093/nar/gkad1025
42. Fischer SN, Claussen ER, Kourtis S *et al.* hu.MAP3.0: atlas of human protein complexes by integration of >25,000 proteomic experiments. *Mol Syst Biol* 2025;21:911–43.
43. Grissa D, Junge A, Oprea TI *et al.* Diseases 2.0: a weekly updated database of disease–gene associations from text mining and data integration. *Database (Oxford)* 2022;2022:baac019. https://doi.org/10.1093/database/baac019
44. Ontology Consortium Gene, Aleksander SA, Balhoff J *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* 2023;224:iyad031. https://doi.org/10.1093/genetics/iyad031
45. Pastrello C, Kotlyar M, Abovsky M *et al.* PathDIP 5: improving coverage and making enrichment analysis more biologically meaningful. *Nucleic Acids Res* 2024;52:D663–71. https://doi.org/10.1093/nar/gkad1027
46. Pérez-Rodríguez N-D, Martín-Ramírez R, González-Fernández R *et al.* Functional disruption of IQGAP1 by truncated PALB2 in two cases of breast cancer: implications for proliferation and invasion. *Biomedicines* 2025;13:1804.
47. Osman MA, Antonisamy WJ, Yakirevich E. IQGAP1 control of centrosome function defines distinct variants of triple negative breast cancer. *Oncotarget* 2020;11:2493–511. https://doi.org/10.18632/oncotarget.27623
48. Van Acker HH, Capsomidis A, Smits EL *et al.* CD56 in the immune system: more than a marker for cytotoxicity? *Front Immunol* 2017;8:892. https://doi.org/10.3389/fimmu.2017.00892
49. Martinez AL, Shannon MJ, Sloan T *et al.* CD56/NCAM mediates cell migration of human NK cells by promoting integrin-mediated adhesion turnover. *MBoC* 2024;35:ar64. https://doi.org/10.1091/mbc.E23-12-0463
50. Blythe EN, Weaver LC, Brown A *et al.* β2 integrin CD11d/CD18: from expression to an emerging role in staged leukocyte migration. *Front Immunol* 2021;12:775447. https://doi.org/10.3389/fimmu.2021.775447
51. Bourel C, Al Khaldi M, Dubuissez M *et al.* CD11d, an NK cell ally in their fight against tumours. *J Immunol* 2024;212:1436_4363. https://doi.org/10.4049/jimmunol.212.supp.1436.4363
52. Picard LK, Claus M, Fasbender F *et al.* Human NK cells responses are enhanced by CD56 engagement. *Eur J Immunol* 2022;52:1441–51. https://doi.org/10.1002/eji.202249868
53. Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res* 2017;45:D408–14. https://doi.org/10.1093/nar/gkw985
54. Basha O, Flom D, Barshir R *et al.* MyProteinNet: build up-to-date protein interaction networks for organisms, tissues and user-defined contexts. *Nucleic Acids Res* 2015;43:W258–63. https://doi.org/10.1093/nar/gkv515
55. Ziv M, Gruber G, Sharon M *et al.* The TissueNet v.3 database: protein–protein interactions in adult and embryonic human tissue contexts. *J Mol Biol* 2022;434:167532. https://doi.org/10.1016/j.jmb.2022.167532
56. Das J, Yu H. HINT: high-quality protein interactomes and their applications in understanding human disease. *BMC Syst Biol* 2012;6:92. https://doi.org/10.1186/1752-0509-6-92
57. Alonso-López Di, Campos-Laborie FJ, Gutiérrez MA *et al.* APID database: redefining protein–protein interaction experimental evidences and binary interactomes. *Database* 2019;2019:baz005.
58. Szklarczyk D, Kirsch R, Koutrouli M *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 2023;51:D638–46. https://doi.org/10.1093/nar/gkac1000

Received: September 15, 2025. **Revised:** October 22, 2025. **Accepted:** October 23, 2025