Nucleic Acids Research, 2025, gkaf1126
https://doi.org/10.1093/nar/gkaf1126
Database issue

OmniPath: integrated knowledgebase for multi-omics analysis

Dénes Türei [1,*,†], Jonathan Schaul [1,†], Nicolàs Palacio-Escat [1,†], Balázs Bohár [2,†], Yunfan Bai [1], Francesco Ceccarelli [3], Elif Çevrim [4,5], Macabe Daley [1], Melih Darcan [4,5], Daniel Dimitrov [1], Tunca Doğan [4,5,6], Daniel Domingo-Fernández [7], Aurelien Dugourd [8], Attila Gábor [1], Lejla Gul [2], Benjamin A Hall [9], Charles Tapley Hoyt [10], Olga Ivanova [1,11], Michal Klein [12], Toby Lawrence [2], Diego Mañanes [1,13], Dezső Módos [2], Sophia Müller-Dott [1], Márton Ölbei [2], Christina Schmidt [1], Bünyamin Şen [4,5], Fabian J Theis [14,15,16], Atabey Ünlü [4,5], Erva Ulusoy [4,5], Alberto Valdeolivas [1], Tamás Korcsmáros [2,‡], Julio Saez-Rodriguez [1,8,*,‡]

[1]Heidelberg University, Faculty of Medicine, and Heidelberg University Hospital, Institute for Computational Biomedicine, Heidelberg 69120, Germany
[2]Imperial College London, Faculty of Medicine, Department of Metabolism, Digestion and Reproduction, London W12 0NN, United Kingdom
[3]Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, United Kingdom
[4]Biological Data Science Lab, Dept. of Computer Engineering, Hacettepe University, Ankara 06800, Turkey
[5]Department of Bioinformatics, Graduate School of Health Sciences, Hacettepe University, Ankara 06100, Turkey
[6]Department of Health Informatics, Institute of Informatics, Hacettepe University, Ankara 06800,Turkey
[7]Fraunhofer Institute for Algorithms and Scientific ComputingSchloss Birlinghoven 53757,Germany
[8]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton CB10 1SD, United Kingdom
[9]Department of Medical Physics and Biomedical Engineering, University College London, London WC1E 6BT, United Kingdom
[10]RWTH Aachen University, Institute of Inorganic Chemistry, Aachen 52064Germany
[11]Division of Vascular Oncology and Metastasis, German Cancer Research Center (DKFZ), Heidelberg 69120,Germany
[12]Apple Inc. Work done while at Institute of Computational Biology, Helmholtz Center Munich, Neuherberg 85764, Germany
[13]Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid 28029, Spain
[14]Institute of Computational Biology, Helmholtz Center Munich, Neuherberg 85764, Germany
[15]TUM School of Computation, Information and Technology, Technical University of Munich, München 80333, Germany
[16]TUM School of Life Sciences, Technical University of Munich, Freising 85354, Germany

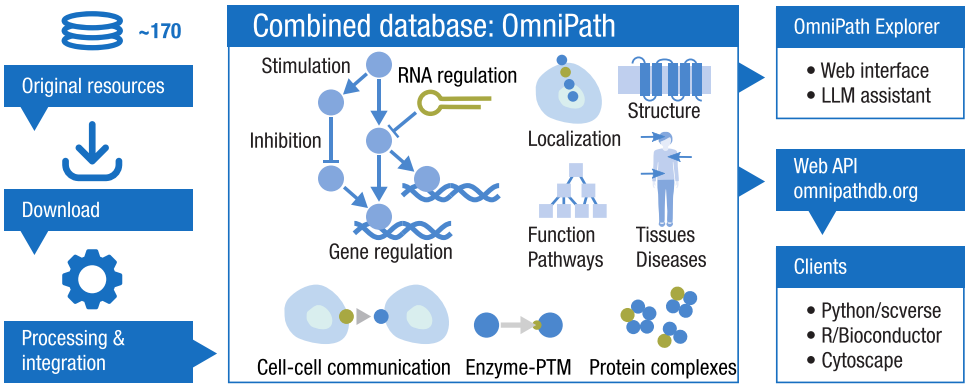
*To whom correspondence should be addressed Email: turei.denes@gmail.com
Correspondrence may also be addressed to Julio Saez-Rodriguez. Email: saezlab@ebi.ac.uk
†The first four authors should be regarded as Joint First Authors.
‡The last two authors should be regarded as Joint Last Authors.

Abstract

Analysis and interpretation of omics data largely benefit from the use of prior knowledge. However, this knowledge is fragmented across resources and often is not directly accessible for analytical methods. We developed OmniPath (https://omnipathdb.org/), a database combining diverse molecular knowledge from 168 resources. It covers causal protein–protein, gene regulatory, microRNA, and enzyme–post-translational modification interactions, cell–cell communication, protein complexes, and information about the function, localization, structure, and many other aspects of biomolecules. It prioritizes literature curated data, and complements it with predictions and large scale databases. To enable interactive browsing of this large corpus of knowledge, we developed OmniPath Explorer, which also includes a large language model agent that has direct access to the database. Python and R/Bioconductor client packages and a Cytoscape plugin create easy access to customized prior knowledge for omics analysis environments, such as scverse. OmniPath can be broadly used for the analysis of bulk, single-cell, and spatial multi-omics data, especially for mechanistic and causal modeling.

Received: August 15, 2025. **Revised:** September 29, 2025. **Accepted:** October 3, 2025

Graphical abstract



Introduction

Bulk, single-cell, and spatial omics technologies provide rich information for understanding biological processes, but interpreting molecular mechanisms and their deregulation in disease remains a challenge. The use of prior knowledge within analytical methods largely expands the extractable information and broadens the scope of testable hypotheses. In particular, by estimating the activity [1] of key processes using prior knowledge of signatures—pathway [2], transcription factor (TF) [2, 3], kinase [5], receptor [6], or ligand–receptor [7] activities—the biological interpretability is greatly enhanced and potential causal drivers can be more easily identified, while the dimensionality of the data is reduced, increasing the statistical power [8]. Furthermore, the integrated outcome of these estimations can be connected by a variety of network inference methods [9, 10] to derive context-specific mechanisms. Databases of prior knowledge have therefore become essential resources for omics data analysis.

Making prior knowledge available for analysis pipelines is a challenge on its own: it is scattered across many databases and it is not clear which ones are the best suited for the application at hand. Furthermore, each requires different and often significant effort to input into the tools. Excellent original [11–14] and combined [15] databases exist, each with a different focus. Nevertheless these are often limited by various caveats, such as lack of domain knowledge, coverage required by analysis workflows, in particular literature curated, as well as causal interactions, and ease of integration to analysis tools. This challenge prompted us to develop OmniPath, a database combining a growing number of diverse, curated and complementary resources. When first published [16], it only covered causal signaling interactions, and over the years it expanded to include TF–target (regulons) and enzyme–post-translational modification (PTM) relationships, molecular function and localization, intercellular communication, protein complexes, and various other types of molecular knowledge, as described in a subsequent publication [17]. Since then, we included 65 new resources, developed an interactive web page, extended the features of the web API (including provenance details and license support), implemented a higher performance server and added numerous convenient utilities to the Python and R client packages (e.g. identifier and orthology translation). Here we present these novelties and briefly outline our ongoing and future development plans.

Data content

OmniPath consists of five major database domains: *interactions, enzyme–substrate, complexes, annotations*, and *intercellular* (Fig. 1A). These *domains* are integrated databases that we build by combining source databases—what we refer to as *resources* (for a complete list of resources, see Supplementary Table S1); while within the domains we also define application-specific subsets of resources, which we call *datasets*. Licensing conditions are included for each resource, with the majority of them being available for commercial use (Fig. 1B–C and Supplementary Table S1). We also classified resources by their maintenance status, i.e. updates happening frequently, infrequently or never. Strikingly, the majority of *interaction* resources and literature references come from databases which have been published only once without subsequent updates (Fig. 1B–D and Supplementary Table S1). In the integrated database multiple resources supporting the same record suggest a higher confidence, while combining the unique content from each resource increases the coverage. In the *interactions* domain only 22% of the records are unique to a single resource, while in the *complexes* domain >60%, largely due to a lack of agreement between complex prediction methods (Fig. 1E).

The *interactions* domain contains 1419 006 unique molecular interactions from 115 different resources: it covers signaling, gene regulatory (TF–target gene), microRNA (miRNA)–gene, and a limited set of drug–target interactions. Within these interaction types, various datasets are available: the core *omnipath* dataset contains literature-curated, causal protein–protein interactions (PPIs), while further interactions without literature references are provided in separate datasets: *pathwayextra* (causal), *kinaseextra* (kinase–substrate), and *ligrecextra* (ligand–receptor). The largest sources of curated causal signaling interactions are the SIGNOR [11], SignaLink [18], and SPIKE [19] databases (Fig. 1A). Within TF regulons, the *tf_target* dataset contains literature-curated TF–target interactions, while *collectri* and *dorothea* datasets represent comprehensive gene regulatory networks from CollecTRI [3] and DoRothEA [4], respectively, compiled from several curated, high-throughput and predicted sources. Each interaction includes information about its direction and whether it has a stimulatory or inhibitory effect. Hereafter, we refer to this information as "causality", as they are direct, physical interactions with known biochemical background, though we note that it reflects only putative causal effects.

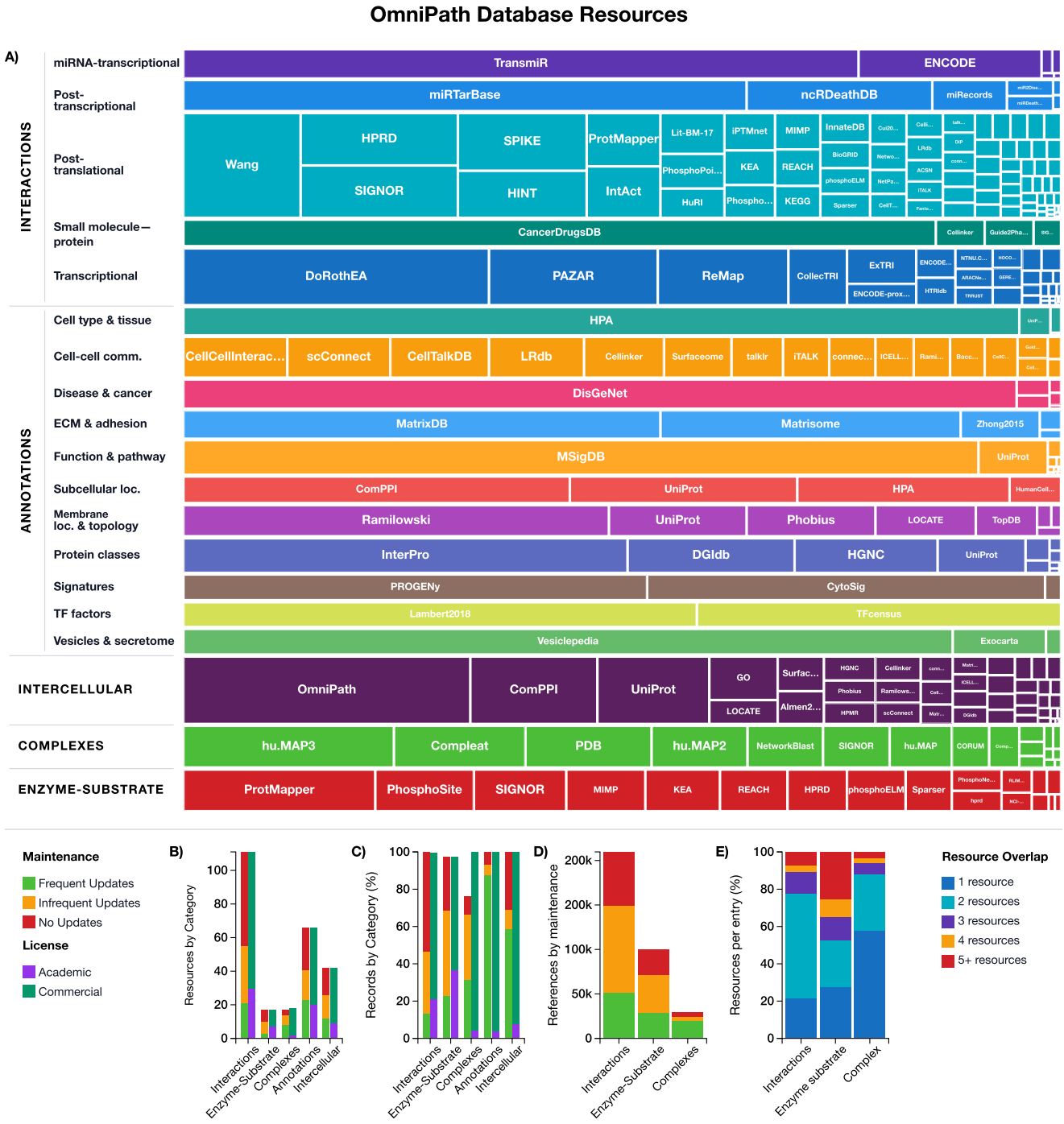


Figure 1. Data content of OmniPath. (**A**) Resource distribution across database categories—Treemap showing relative sizes of individual resources within their database domain and subcategory. The size of each cell is proportional to the number of records. (**B**) Number of resources in each database domain, by maintenance and license categories. (**C**) Percentage of records by database domain and by maintenance and license category. (**D**) Literature reference count by database and maintenance status. (**E**) Resource overlap within the databases: percentage of entries appearing in 1, 2, 3, 4, or 5 + resources. Interactive versions of these visualizations are available at https://explore.omnipathdb.org/.

The *enzyme–substrate* domain is a collection of 115 215 enzyme–PTM interactions from 18 resources. Each record describes the residue, site, and type of modification, most predominantly phosphorylation, followed by dephosphorylation and acetylation. PhosphoSitePlus [20] contributes the majority of curated evidence, complemented by further curated and prediction-based resources.

The *complexes* database enumerates 52 086 human protein complexes, integrating 18 resources, annotated with stoichiometry and literature references. CORUM [21] and Complex Portal [22] are the primary sources of curated complexes, accompanied by several smaller or prediction-based resources.

The largest database domain in OmniPath is the *annotations* with its 5895 462 entries, providing a broad variety

of protein and gene function, localization, structure, and expression information. This includes pathway memberships, roles in biological processes and diseases, for example, the functional gene sets from MSigDB [23]; various classifications, such as the protein families from HGNC [24]; protein localizations—for example, CSPA [25] to annotate cell surface proteins; and weighted functional signatures, such as the pathway response scores from PROGENy [2] or cytokine responses from CytoSig [26]. The data extracted from the 67 resources (Supplementary Table S1) is provided as it is, without integration across resources.

The *intercell* domain integrates annotation resources into a curated atlas of cell–cell communication. Using the function, localization and structure annotations described above, it classifies proteins into categories such as ligand, receptor, transporter, matrix protein, or secreted enzyme, and tags them with membrane associations and subcellular localization [17]. In the R and Python clients, these annotations can be merged with molecular interactions, enabling application-specific customization, for example, by establishing tissue-specific ligand–receptor networks [27].

The *interactions* and *enzyme–substrate* database domains, built with human data, are translated to mouse and rat by orthologous gene pairs. For this translation, we used NCBI HomoloGene [28], Ensembl [29], and the Orthologous Matrix (OMA; 9). Our translation utilities are available in the *pypath* and *OmnipathR* packages, allowing translation to other organisms.

Web page

With the current update, we introduce *OmniPath Explorer*, an interactive web application (https://explore.omnipathdb.org/) to access the OmniPath resource. *OmniPath Explorer* consists of browsable pages of database content and a chat assistant (Fig. 2A–C).

The first allows users to search for protein and gene names and explore the data through an easy-to-use but comprehensive graphical interface. Contents of the five database domains are presented in five different views. The left side control panel enables filtering by a broad range of variables, for example, type, causality, amount of evidence for *interactions* (Fig. 2A), or location, scope and causality for *intercell*. In the *interactions* part, all partners of a given molecule appear as a list, with the interaction type and causality encoded by colors and symbols. The provenances are also included for each interaction, with links to the original databases and articles in PubMed. The *annotations* part groups the resources by topic, and presents data from the selected ones in tabular format. Alternatively, free text search in annotation records is also available (Fig. 2B). In all views, proteins of interest are presented at the top, with a basic overview from UniProt [31]: the organism, molecular weight, polypeptide length, and literature statements about the functions, localization, function, classification, Gene Ontology, PTM information, and links to PubMed and further resources (Fig. 2D).

The *chat* part integrates an LLM assistant that accepts natural language questions, writes and executes SQL queries, and interprets the results to provide answers. The generated SQL queries and their output are shown alongside the answers, these can be checked for correctness, edited, and exported as tables (Fig. 2C). The LLM is provided with an expanding set of query templates which guide it to formulate queries. This enables noncomputational users to explore the database content and flexibly access and integrate data across multiple tables or resources.

Web API

The web service serves data in tabular or JSON format, and consists of five main endpoints (query types), corresponding to the five database domains of OmniPath. It supports the filtering of records by practically any of the variables: by resources, molecules, organisms, and other variables specific for the database domains, for example, filter TF regulons according to specific confidence levels [4], or PTM residues at the enzyme–substrate domain, or cell–cell communication roles at the *intercell* domain. Optional columns can be selected by the *fields* parameter. *Annotations* are returned as a long-format data frame and require pivoting into wide format, which is supported by the client packages.

We recently added two new columns to the *interaction* records, both containing JSON blobs. The *extra_attrs* column includes resource-specific interaction attributes, such as the mechanism or detection method of the interaction. The client packages are able to extract specific variables from the JSON blob to data frame columns. The *evidences* column contains the provenance information in full detail. This enables the client packages to do precise filtering, for example, discarding all information that is not licensed for commercial use. License-based filtering is also supported by the web API's *license* parameter.

Causality of *interactions* is represented by three columns *(is_directed, is_stimulation, is_inhibition)*; in addition, *"consensus"* alternatives of these columns provide a majority vote across all resources.

The web service also features a few auxiliary queries that provide meta-information about the contents. The *databases* and *datasets* queries return the list of resources and datasets in the *interactions* domain, the *queries* query returns the list of valid parameters for each query, while the *resources* query maps the use of resources within the databases, and also includes license information. The *annotations_summary* and *intercell_summary* queries, for each resource, list all variables and all possible values.

Python, R, and Cytoscape clients

The *omnipath* Python package is available in the PyPI repository (https://pypi.org/project/omnipath/), while the *OmnipathR* R package (https://bioconductor.org/packages/OmnipathR) is part of Bioconductor [33]. Query types, interaction types and interaction datasets are represented in the Python package by classes, in the R package by functions. All web service query parameters can be provided as arguments to the *"get"* method of the Python classes and similarly to the functions in the R package. The results are returned as data frames. Both packages provide utilities to pivot the annotation data frames from long to wide format, combine networks and annotations, translate data to other organisms by orthologous gene pairs, and update the causality of interactions based on detailed provenance data in the *evidences* column.

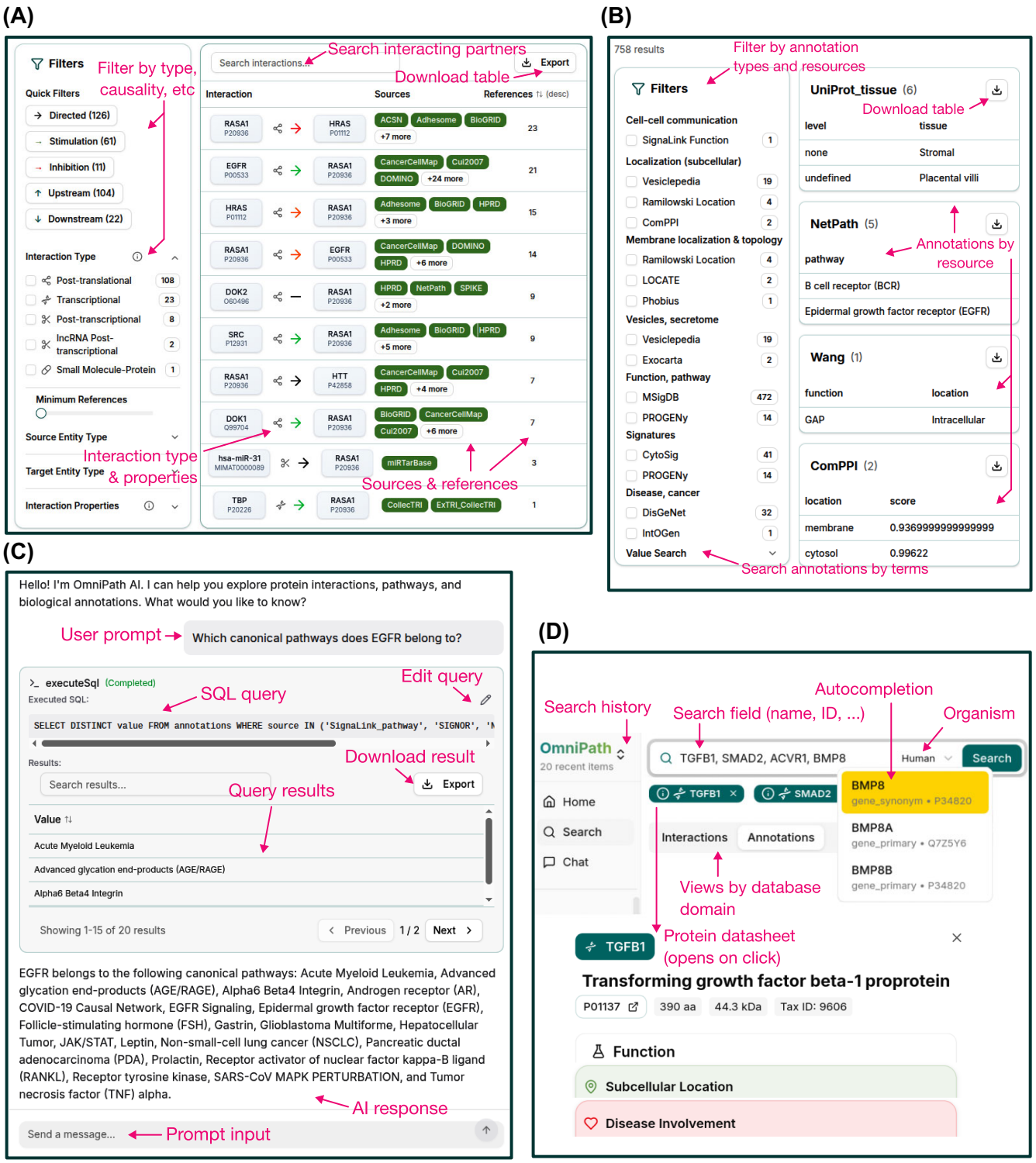


Figure 2. The OmniPath Explorer web application. (**A**) Interaction browser. (**B**) Annotation browser. (**C**) Chat interface to the OmniPath large language model (LLM) agent. (**D**) Main page search and menu; protein datasheet.

Besides the OmniPath client functionalities, *OmnipathR* provides direct access to 26 resources (Supplementary Table S1), among them several metabolomics related ones, such as MetalinksDB [34], a network of annotated metabolite-protein interactions combining 14 resources, or RaMP-DB [35], a comprehensive resource of metabolite identifiers and structures. *OmnipathR* also comes with a prior knowledge processing toolkit, including translation utilities for 102 identifier types, orthologous gene pair translation (by HomoloGene, Ensembl [29], and OMA [30]), translation ambiguity analysis, handling of organism names, traversal of the Gene Ontology [36] tree, full featured clients to the Ensembl BioMart [29], KEGG [37], and UniProt [32] APIs. It is able to convert *interaction* data frames to igraph [38] networks or deliver them to Bio Model Analyzer [39].

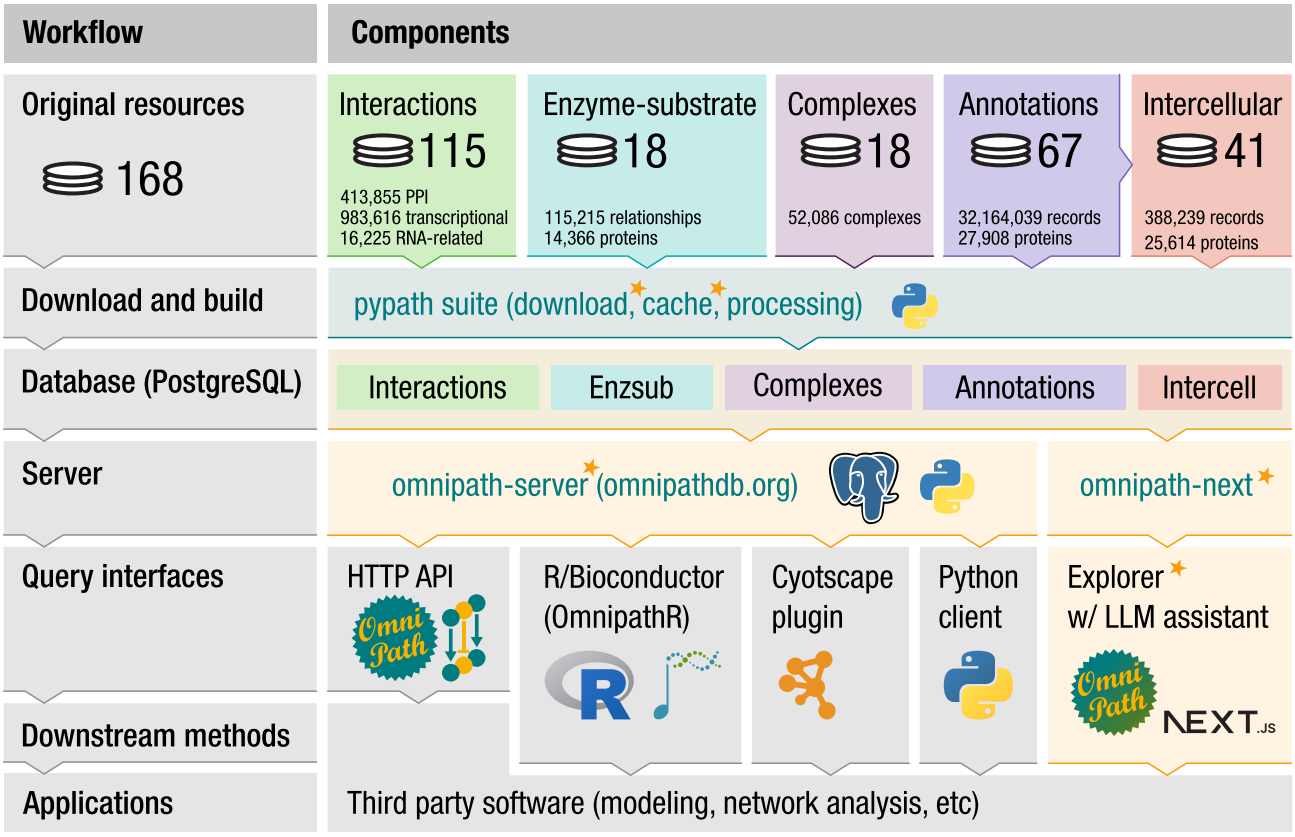


Figure 3. Technical architecture of OmniPath. The combined database, built from 168 different resources by the pypath Python software suite, is publicly available through the HTTP API at https://omnipathdb.org/. The OmniPath Explorer web app allows interactive browsing with LLM assistance, while client packages for R, Python, and Cytoscape provide convenient access and utilities for seamless integration into analysis workflows. The components new or largely renewed in the current update are highlighted in orange boxes and marked with orange stars.

The *OmniPath* Cytoscape app [40] supports the *interactions*, *enzyme–substrate*, and a limited set of the *annotations* database domains. It imports the data directly into Cytoscape from the datasets and resources specified by the user.

Implementation

The database build and the web API of OmniPath are implemented in a suite of Python packages (Fig. 3). The *pypath* package is responsible for resource-specific parsing and compiling the combined databases. It features a number of processing utilities, most importantly, the identifier and orthologous gene pair translation. The *pypath.inputs* module is a collection of clients for 200 original resources. These clients use the *download-manager* and *cache-manager* packages for robust network transactions and local caching. The *pypath.core* module builds the OmniPath databases. All clients and the complete database build are tested daily by an automated pipeline and a status report is published at https://status.omnipathdb.org/. For original resources that became temporarily or permanently inaccessible, we host these on our own server at https://rescued.omnipathdb.org/. Another Python package, the *omnipath-server* loads the databases into PostgreSQL and operates the web service. OmniPath Explorer is a TypeScript application built with the Next.js framework, and uses the same PostgreSQL database as the web API (Fig. 3). The database is updated periodically, with the old versions archived at https://archive.omnipathdb.org/. By default, the LLM agent uses the openly-accessible Google Gemini Flash 2.5 model.

Discussion

OmniPath is an integrated database combining 168 molecular resources into a single, continuously updated framework, including signed and directed PPIs, enzyme–PTM relationships, ligand–receptor pairs, protein complexes, and extensive functional annotations. By harmonizing data from all these diverse resources, OmniPath facilitates access to prior-knowledge for a broad range of use cases. The OmniPath Explorer presents all evidence for molecular interactions and comprehensive annotations in one place. This allows users to explore interactively the complete knowledgebase and quickly look up specific information and find the most suitable resources for their analysis. The web API and its clients for popular bioinformatics environments (Python, R, Cytoscape) enable effortless creation of customized prior knowledge for diverse applications. With its coverage of small and curated resources, OmniPath also fills a critical gap among other large meta-resources, such as STRING [13] or PathwayCommons [15], and major interaction or pathway databases with original curation effort, like IntAct [12] or Reactome [14]. The heterogeneous curation protocols and quality of constituting resources can be seen as a potential weakness of OmniPath as an integrated database; however, alternatives—literature mined networks [41], high-throughput screens [12], correlation-based

approaches [13]—come with their own limitations, while an integrated database creates new opportunities to search for potentially erroneous records, and also to compile on demand any custom combination of the resources.

Integrating a multitude of resources in a uniform format opens the way towards their benchmarking, as it is done, for example, in NetworkCommons [10] to evaluate the performance of network inference methods. Including various interaction types—signaling pathways, transcriptional, and miRNA regulation, ligand–receptor, etc—in a uniform network, together with the annotations and the above mentioned inference methods, enables the generation and investigation of complex, multilayered hypotheses that would be limitedly possible with other resources.

The integrative design of OmniPath translates into practical impact across diverse tasks in the analysis of omics data. The client packages feature integrations with several downstream analysis tools. For example, integration with the enrichment package Decoupler [1] enables seamless use of signature-based TF, pathway, kinase, and cytokine activity estimations in bulk and single-cell workflows, including within scverse [42], as well as automated cell type annotations using PanglaoDB [43]. Similarly, OmniPath delivers highly customizable ligand–receptor networks directly into the LIANA+ [7] cell–cell communication inference framework, also part of the scverse ecosystem. Networks from OmniPath are readily available in the CORNETO [44] network optimization framework, its multi-omics causal variant COSMOS [45], and further network inference methods in NetworkCommons [10], to derive context-specific mechanisms from omics data. For metabolism, OmniPath not only delivers prior knowledge in the MetaProViz R package [46], but also enables connecting knowledge and metabolomics features with its extensive identifier translation utilities. OmniPath is also connected to other prior-knowledge processing systems: the curated knowledge from OmniPath is presented in the INDRA natural language processing system for molecular interactions [41]; a script is available to write OmniPath interactions into neo4j importable CSV using the BioCypher library [47]; while OmniPath's prior-knowledge processing toolkit is used in building the CROssBARv2 database [48].

The server-client architecture isolates resource-specific download and processing from analysis workflows, for a reduced complexity and enhanced robustness. Built-in support for resource license constraints facilitates applications in commercial settings by keeping compliance straightforward. OmniPath Explorer's LLM agent turns text-based questions into SQL queries. This user-friendly mode of operation assists both new and advanced users to design database queries, while directly answering natural language questions for noncomputational users. We are integrating OmniPath into BioContextAI, a collection of biomedical LLM agents [49], to facilitate integration with LLMs via the model context protocol. We consider the current implementation to be only a first step towards LLM assistance. Through an improved integration with the knowledge contained in OmniPath, and access to additional computational tools, our LLM agent may evolve to answer more complex questions and contribute to larger tasks together with other LLMs. We aim to iterate on our implementation by continuously learning about its current utility and limitations from user feedback.

In our future work, we also plan covering more metabolite, drug, and microbiome-related knowledge, with special focus on causal relationships and functional annotations. We intend to further develop OmniPath Explorer with more views, richer interactivity, and information presented in each view. In the Python client side, we have plans to further improve the integration with network inference methods in NetworkCommons and single-cell workflows of the scverse ecosystem.

We remain committed to developing OmniPath as a free open-source resource for the community, particularly for researchers analyzing omics data. We welcome feedback, content suggestions, feature requests, and bug reports, which can be submitted via GitHub (https://github.com/saezlab) or email (omnipathdb@gmail.com).

Funding

D.T. was supported by the Landesinstitut für Bioinformatik-infrastruktur in Baden-Württemberg. J.S. was supported by the European Union's Horizon 2020 Programme [grant no. 965193 (DECIDER)]. N.P.E. acknowledges the support of the German Research Foundation (DFG) through the CRC 1550 "Molecular Circuits of Heart Disease" [grant no. DFG-508152189]. The work of L.G., B.B., and T.K. was supported by the UKRI BBSRC Institute Strategic Programme Food Microbiome and Health [BB/X011054/1 and its constituent project BBS/E/F/000PR13631]. T.D., E.Ç., M.D., B.Ş., A.Ü., and EU were supported by TUBITAK ARDEB 3501 Career Development Program [project no: 120E531]. S.M.D. was funded by the LiSyM-cancer network supported by the German Federal Ministry of Research, Technology and Space (BMFTR) [031L0257B]. D.M. acknowledges financial support from Imperial College London through an Imperial College Research Fellowship grant award. C.S. was funded by SmartCare [03LW0233K]. Funding to pay the Open Access publication charges for this article was provided by the European Bioinformatics Institute (EMBL-EBI).

Acknowledgements

We are grateful to Ömer Kaan Vural, Tennur Kılıç, Ahmet Rifaioglu, and Forrest Hyde for their contributions in the *pypath.inputs* module and to Daniele Bottazzi, Jannik Franken, Edwin Carreño, and Youssef Zerta for their ongoing contributions to related software packages.

Author contributions: Dénes Türei (Conceptualization [lead], Data curation [lead], Software [lead], Supervision [lead], Writing—original draft [lead], Writing—review and editing [lead]), Jonathan Schaul (Software [lead], Visualization [lead], Writing—review & editing [equal]), Nicolàs Palacio-Escat (Software [lead], Visualization [lead], Writing—review & editing [equal]), Balázs Bohár (Software [lead]), Yunfan Bai (Data curation [equal], Software [equal]), Francesco Ceccarelli (Software [equal]), Elif Çevrim (Software [equal]), Macabe Daley (Software [equal]), Melih Darcan (Software [equal]), Daniel Dimitrov (Software [equal]), Tunca Doğan (Software [equal], Supervision [equal]), Daniel Domingo-Fernández (Software [equal]), Aurelien Dugourd (Software [equal], Supervision [equal]), Attila Gábor (Project administration [equal], Software [equal]), Lejla Gul (Project administration [equal], Software [equal], Supervision [equal], Writing—review & editing [equal]), Benjamin A Hall (Software [equal]), Charles Hoyt (Software [equal]), Olga Ivanova (Software [equal]), Michal Klein (Software [equal]), Toby Lawrence (Software [equal]), Diego Mañanes (Software

[equal]), Dezső Módos (Software [equal]), Sophia Müller-Dott (Software [equal]), Márton Ölbei (Software [equal]), Christina Schmidt (Supervision [equal]), Bünyamin Şen (Software [equal]), Fabian Theis (Software [equal], Supervision [equal]), Atabey Ünlü (Software [equal]), Erva Ulusoy (Software [equal]), Alberto Valdeolivas (Software [equal]), Tamás Korcsmáros (Conceptualization [lead], Supervision [equal], Writing—review and editing [lead]), Julio Saez-Rodriguez (Conceptualization [lead], Supervision [equal], Writing—review & editing [lead]).

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

JSR reports in the last 3 years funding from GSK and Pfizer, and fees and honoraria from Travere Therapeutics, Stadapharm, Astex, Owkin, Pfizer, Grunenthal, Tempus, and Moderna. Work of T.K. and B.B. was partially supported by Roche. F.J.T. consults for Immunai Inc., CytoReason Ltd, Cellarity, BioTuring Inc., Valinor Discovery, and Genbio.AI Inc., and has an ownership interest in RN.AI Therapeutics, Dermagnostix GmbH and Cellarity. A.V. is currently an employee and shareholder of F. Hoffmann-La Roche Ltd.

Data availability

The pypath Python package is available at https://github.com/saezlab/pypath, the auxiliary packages are linked from this repository. The web API implementation is available at https://github.com/saezlab/omnipath-server, while the OmniPath Explorer app at https://github.com/saezlab/omnipath-next. OmnipathR is provided at https://github.com/saezlab/OmnipathR, the omnipath Python client at https://github.com/saezlab/omnipath and the Cytoscape app at https://github.com/saezlab/Omnipath_Cytoscape. All these components are free software, the database build, web app, server, and Cytoscape app are distributed under GPLv3 license, while OmnipathR and omnipath client packages under MIT license.

OmniPath data is available at https://omnipathdb.org/ and by the web app at https://explore.omnipathdb.org/, under the licenses of the constituting resources. Old releases of OmniPath are accessible at https://archive.omnipathdb.org/, and original resources hosted by OmniPath at https://rescued.omnipathdb.org/.

The code has been archived in the following Zenodo repositories:

pypath: https://doi.org/10.5281/zenodo.17294424
omnipath-next: https://doi.org/10.5281/zenodo.17294409
OmnipathR: https://doi.org/10.5281/zenodo.17294392
omnipath (Python client): https://doi.org/10.5281/zenodo.17294394
omnipath-server: https://doi.org/10.5281/zenodo.17294418

References

1. Badia-I-Mompel P, Vélez Santiago J, Braunger J *et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinforma Adv* 2022;2:vbac016. https://doi.org/10.1093/bioadv/vbac016
2. Schubert M, Klinger B, Klünemann M *et al.* Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun* 2018;9:20. https://doi.org/10.1038/s41467-017-02391-6
3. Müller-Dott S, Tsirvouli E, Vazquez M *et al.* Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic Acids Res* 2023;51:10934–49. https://doi.org/10.1093/nar/gkad841
4. Garcia-Alonso L, Holland C, Ibrahim M *et al.* Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res* 2019; 29:1363–75. https://doi.org/10.1101/gr.240663.118
5. Wirbel J, Cutillas P, Saez-Rodriguez J. Phosphoproteomics-based profiling of kinase activities in cancer cells. In: Von Stechow L (ed.), *Cancer Systems Biology*. New York: Springer, 2018, 1711, pp. 103–32.
6. Barsi S, Varga E, Dimitrov D *et al.* RIDDEN: data-driven inference of receptor activity from transcriptomic data. *PLoS Comput Biol* 2025;21:e1013188. https://doi.org/10.1371/journal.pcbi.1013188
7. Dimitrov D, Schäfer P, Farr E *et al.* LIANA+ provides an all-in-one framework for cell–cell communication inference. *Nat Cell Biol* 2024;26:1613–22. https://doi.org/10.1038/s41556-024-01469-w
8. Schäfer PSL, Dimitrov D, Villablanca EJ *et al.* Integrating single-cell multi-omics and prior biological knowledge for a functional characterization of the immune system. *Nat Immunol* 2024;25:405–17. https://doi.org/10.1038/s41590-024-01768-2
9. Garrido-Rodriguez M, Zirngibl K, Ivanova O *et al.* Integrating knowledge and omics to decipher mechanisms via large-scale models of signaling networks. *Mol Syst Biol* 2022;18:e11036. https://doi.org/10.15252/msb.202211036
10. Paton V, Türei D, Ivanova O *et al.* NetworkCommons: bridging data, knowledge and methods to build and evaluate context-specific biological networks. *Bioinformatics* 2025;41:btaf048. https://doi.org/10.1093/bioinformatics/btaf048
11. Lo Surdo P, Iannuccelli M, Contino S *et al.* SIGNOR 3.0, the SIGnaling network open resource 3.0: 2022 update. *Nucleic Acids Res* 2023; 51:D631–7. https://doi.org/10.1093/nar/gkac883
12. del Toro N, Shrivastava A, Ragueneau E *et al.* The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res* 2022; 50:D648–53. https://doi.org/10.1093/nar/gkab1006
13. Szklarczyk D, Nastou K, Koutrouli M *et al.* The STRING database in 2025: protein networks with directionality of regulation. *Nucleic Acids Res* 2025; 53:D730–7. https://doi.org/10.1093/nar/gkae1113
14. 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
15. Rodchenkov I, Babur O, Luna A *et al.* Pathway Commons 2019 Update: integration, analysis and exploration of pathway data. *Nucleic Acids Res* 2019;48:D489–97. https://doi.org/10.1093/nar/gkz946
16. Türei D, Korcsmáros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat Methods* 2016;13:966–7.
17. Türei D, Valdeolivas A, Gul L *et al.* Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol Syst Biol* 2021;17:e9923.
18. Csabai L, Fazekas D, Kadlecsik T *et al.* SignaLink3: a multi-layered resource to uncover tissue-specific signaling networks. *Nucleic Acids Res* 2022; 50:D701–9. https://doi.org/10.1093/nar/gkab909
19. Paz A, Brownstein Z, Ber Y *et al.* SPIKE: a database of highly curated human signaling pathways. *Nucleic Acids Res* 2011; 39:D793–9. https://doi.org/10.1093/nar/gkq1167
20. Hornbeck PV, Zhang B, Murray B *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* 2015; 43:D512–20. https://doi.org/10.1093/nar/gku1267
21. 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

22. 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
23. Liberzon A, Subramanian A, Pinchback R *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27:1739–40. https://doi.org/10.1093/bioinformatics/btr260
24. Seal RL, Braschi B, Gray K *et al.* Genenames.org: the HGNC resources in 2023. *Nucleic Acids Res* 2023;51:D1003–9. https://doi.org/10.1093/nar/gkac888
25. Bausch-Fluck D, Hofmann A, Bock T *et al.* A mass spectrometric-derived cell surface protein atlas. *PLoS One* 2015;10:e0121314. https://doi.org/10.1371/journal.pone.0121314
26. Jiang P, Zhang Y, Ru B *et al.* Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. *Nat Methods* 2021;18:1181–91. https://doi.org/10.1038/s41592-021-01274-5
27. Poletti M, Treveil A, Csabai L *et al.* Mapping the epithelial–immune cell interactome upon infection in the gut and the upper airways. *Npj Syst Biol Appl* 2022;8:15. https://doi.org/10.1038/s41540-022-00224-x
28. Sayers EW, Beck J, Bolton EE *et al.* Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res* 2025;53:D20–9. https://doi.org/10.1093/nar/gkae979
29. 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
30. Altenhoff AM, Train C-M, Gilbert KJ *et al.* OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Res* 2021; 49:D373–9. https://doi.org/10.1093/nar/gkaa1007
31. UniProt Consortium T, Bateman A, Martin M-J *et al.* UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res* 2025; 53:D609–17. https://doi.org/10.1093/nar/gkae1010
32. UniProt Consortium T, Bateman A, Martin M-J *et al.* UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res* 2025; 53:D609–17. https://doi.org/10.1093/nar/gkae1010
33. Huber W, Carey VJ, Gentleman R *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 2015;12:115–21. https://doi.org/10.1038/nmeth.3252
34. Farr E, Dimitrov D, Schmidt C *et al.* MetalinksDB: a flexible and contextualizable resource of metabolite-protein interactions. *Brief Bioinform* 2024;25:bbae347. https://doi.org/10.1093/bib/bbae347
35. Braisted J, Patt A, Tindall C *et al.* RaMP-DB 2.0: a renovated knowledgebase for deriving biological and chemical insight from metabolites, proteins, and genes. *Bioinformatics* 2023;39:btac726. https://doi.org/10.1093/bioinformatics/btac726
36. The Gene Ontology Consortium, Aleksander SA, Balhoff J *et al.*, The Gene Ontology knowledgebase in 2023. *Genetics* 2023;224:iyad031.
37. Kanehisa M, Furumichi M, Sato Y *et al.* KEGG: biological systems database as a model of the real world. *Nucleic Acids Res* 2025; 53:D672–7. https://doi.org/10.1093/nar/gkae909
38. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems* 2006;1695:1–9.
39. Benque D, Bourton S, Cockerton C *et al.* Bio Model Analyzer: visual tool for modeling and analysis of biological networks. In: *Computer Aided Verification (CAV) 2012. LNCS 7358*. Berlin, Heidelberg: Springer Verlag, 2012, 686–92.
40. Ceccarelli F, Turei D, Gabor A *et al.* Bringing data from curated pathway resources to Cytoscape with OmniPath. *Bioinformatics* 2020;36:2632–3. https://doi.org/10.1093/bioinformatics/btz968
41. Bachman JA, Gyori BM, Sorger PK. Automated assembly of molecular mechanisms at scale from text mining and curated databases. *Mol Syst Biol* 2023;19:e11325. https://doi.org/10.15252/msb.202211325
42. Virshup I, Bredikhin D, Heumos L *et al.* The scverse project provides a computational ecosystem for single-cell omics data analysis. *Nat Biotechnol* 2023;41:604–6. https://doi.org/10.1038/s41587-023-01733-8
43. Franzén O, Gan L, Björkegren J. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database J Biol Databases Curation* 2019;2019:baz046.
44. Rodriguez-Mier P, Garrido-Rodriguez M, Gabor A *et al.* Unifying multi-sample network inference from prior knowledge and omics data with CORNETO. *Nat Mach Intell* 2025;7:1168–86. https://doi.org/10.1038/s42256-025-01069-9
45. Dugourd A, Lafrenz P, Mañanes D *et al.* Modeling causal signal propagation in multi-omic factor space with COSMOS. bioRxiv, https://doi.org/10.1101/2024.07.15.603538, 19 Jult 2024, preprint: not peer reviewed.
46. Schmidt C, Turei D, Prymidis D *et al.* Integrated metabolomics data analysis to generate mechanistic hypotheses with MetaProViz. *bioRxiv*, https://doi.org/10.1101/2025.08.18.670781, 22 August 2025, preprint: not peer reviewed.
47. Lobentanzer S, Aloy P, Baumbach J *et al.* Democratizing knowledge representation with BioCypher. *Nat Biotechnol* 2023;41:1056–9. https://doi.org/10.1038/s41587-023-01848-y
48. Doğan T, Atas H, Joshi V *et al.* CROssBAR: comprehensive resource of biomedical relations with knowledge graph representations. *Nucleic Acids Res* 2021; 49:e96.
49. Kuehl M, Schaub DP, Carli F *et al.* Community-based biomedical context to unlock agentic systems. *bioRxiv*, https://doi.org/10.1101/2025.07.21.665729, 25 July 2025, preprint: not peer reviewed.

Received: August 15, 2025. **Revised:** September 29, 2025. **Accepted:** October 3, 2025