Lawrence Berkeley National Laboratory
LBL Publications

Title
Biomolecule and Bioentity Interaction Databases in Systems Biology: A Comprehensive Review.

Permalink

Journal

Authors
Baltoumas, Fotis
Zafeiropoulou, Sofia
Karatzas, Evangelos
et al.

Publication Date

DOI

Copyright Information

Peer reviewed

Biomolecule and Bioentity Interaction Databases in Systems Biology: A Comprehensive Review

Fotis A. Baltoumas [1],*, Sofia Zafeiropoulou [1], Evangelos Karatzas [1], Mikaela Koutrouli [1,2], Foteini Thanati [1], Kleanthi Voutsadaki [1], Maria Gkonta [1], Joana Hotova [1], Ioannis Kasionis [1], Pantelis Hatzis [1,3] and Georgios A. Pavlopoulos [1,3],*

[1] Institute for Fundamental Biomedical Research, Biomedical Sciences Research Center "Alexander Fleming", 16672 Vari, Greece; zafeiropoulou@fleming.gr (S.Z.); karatzas@fleming.gr (E.K.); mikaela.koutrouli@cpr.ku.dk (M.K.); fotinithanati@gmail.com (F.T.); voutsadaki@fleming.gr (K.V.); mgkonta97@gmail.com (M.G.); hotova@fleming.gr (J.H.); gkasionis2@gmail.com (I.K.); hatzis@fleming.gr (P.H.)

[2] Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, 2200 Copenhagen, Denmark

[3] Center for New Biotechnologies and Precision Medicine, School of Medicine, National and Kapodistrian University of Athens, 11527 Athens, Greece

* Correspondence: baltoumas@fleming.gr (F.A.B.); pavlopoulos@fleming.gr (G.A.P.); Tel.: +30-210-965-6310 (G.A.P.)

Abstract: Technological advances in high-throughput techniques have resulted in tremendous growth of complex biological datasets providing evidence regarding various biomolecular interactions. To cope with this data flood, computational approaches, web services, and databases have been implemented to deal with issues such as data integration, visualization, exploration, organization, scalability, and complexity. Nevertheless, as the number of such sets increases, it is becoming more and more difficult for an end user to know what the scope and focus of each repository is and how redundant the information between them is. Several repositories have a more general scope, while others focus on specialized aspects, such as specific organisms or biological systems. Unfortunately, many of these databases are self-contained or poorly documented and maintained. For a clearer view, in this article we provide a comprehensive categorization, comparison and evaluation of such repositories for different bioentity interaction types. We discuss most of the publicly available services based on their content, sources of information, data representation methods, user-friendliness, scope and interconnectivity, and we comment on their strengths and weaknesses. We aim for this review to reach a broad readership varying from biomedical beginners to experts and serve as a reference article in the field of Network Biology.

Keywords: biomedical networks; associations; biological interactions; network biology; databases; data integration

1. Introduction

Technological advances in various high-throughput techniques in the last decade have led to an explosion of information about how biomolecules operate and function in living systems. Microarray and RNAseq technologies, for example, provide insights into gene expression levels and changes. scRNAseq technology organizes cells into groups based on their gene expression profiles and mass spectrometry is used for the identification of proteins based on their molecular weights and mass-to-charge ratio. Nuclear magnetic resonance (NMR) and X-ray crystallography are used for the determination of 3D protein structures while genome sequencing can provide insights about genetic variations, polymorphisms, chromatin structure and state, and species identification. Finally, metabolomics are used to study small molecules and metabolites within cells, biofluids, tissues or organisms [1].

Biological systems are composed of a multitude of molecular entities such as genes, proteins, metabolites and other components and essentially all biological processes are regulated by the interactions among these entities. The analysis of these interactions plays an important part in achieving a mechanistic understanding of physiology and pathology of all forms of life, ranging from single-cell microbes to complex, multicellular organisms. This applies not only at the microscopic scale of the cell interior but also at a macroscopic level; studying the relations between different species occupying an ecosystem can help establish the ecological rules that govern a specific environment. As a result, the study of biological interactions has become a staple of systems biology [2].

As current research involves increasing levels of complexity by combining multiple approaches (e.g., genomics, proteomics, transcriptomics, metabolomics, etc.), particularly in the case of biological interactions [3], the necessity for specialized repositories and advanced integration and visualization techniques emerges. One such technique involves the use of biological interaction networks. In network biology, graphs are often used to represent compartments of whole systems and their biomolecular interactions. A node often represents a biomolecule (e.g., gene, protein, chemical, compound, disease, etc.) whereas an edge the relationship between them (e.g., co-expression, co-occurrence, sequence similarity, coevolution, orthology, homology, fusion, common function, etc.) [4].

Biological interaction networks have been used in a wide range of analyses; some of which have been performed at previously unprecedented scales. The most characteristic example is the Human Interactome Network [5], a proteome-scale analysis of protein–protein interactions for the entire human proteome, which has allowed the detection of previously unknown functional relationships. Starting from an initial analysis for a collection of experimental datasets [6], it is currently a reference map for the human proteome and its interactions, containing more than 50,000 binary interactions featuring almost 90% of the protein-coding genome. Similar genome and proteome-wide interaction networks have also been constructed for a number of other model organisms, such as the mouse (*Mus musculus*) [7,8], yeast (*Saccharomyces cerevisiae*) [9,10] and fruit fly (*Drosophila melanogaster*) [11]. In addition, the combination of gene co-expression and protein–protein interaction evidence with information on metabolic pathways and disease associations has led to the creation and analysis of specialized networks dealing with severe pathological conditions, such as cancer [12], AIDS [13], Alzheimer's Disease [14,15], and, most recently, infection with SARS-CoV-2, the virus responsible for the COVID-19 pandemic [16].

The successful generation and analysis of interaction networks, such as those referenced above, requires the presentation of information regarding biological interactions in an organized and concise manner. Although this evidence is, to some extent, available in popular biomedical repositories such as PubMed [17], UniProt [18], GenBank [19], or Ensembl [20], their usefulness for the generation of networks is limited, as these databases have not been designed with the analysis of interactions in mind. The aforementioned issue, coupled with the increasing size and complexity of matrices of interactions, as produced by high throughput methods or generated through computational predictions, has led to the emergence of dedicated biological interaction databases. Nowadays, multiple such interaction databases exist [21], acting as specialized repositories of evidence on gene, protein, and small molecule interactions, as well as associations of these interactions with metabolic pathways, host–pathogen relationships, diseases, and even ecological data. However, while the existence of these repositories has provided more immediate access to interaction evidence, their utilization is not always straightforward, as most of these databases are self-contained systems, each containing their own set of interactions and utilizing their unique organization systems and file formats, which are often incompatible with each other [22]. This makes the retrieval, combination, and manipulation of interaction evidence difficult, particularly for new and inexperienced users.

In this review, we outline, organize, and compare biological interaction databases for a number of interaction types, from protein–protein and protein–small molecule complexes

to disease-association, host–pathogen and environment-organism interactions. Notably, we do not only focus on the major databases but also on more specialized repositories and we thoroughly present, group, and evaluate most of the publicly available services based on their content, sources of information, data representation methods, and scope. Given the ever-rising wealth of available information on biomolecular interactions and biomedical networks, we aim for this review to reach a broad spectrum of readers varying from experts to beginners, and serve as a reference for the biomedical community.

2. File Formats

Before analyzing the various databases, we briefly describe some of the more usual file formats that interaction databases offer. Aiming at a more structured way of storing interaction data, several specific file formats have been introduced. An initial approach was to borrow concepts from graph theory and store a network as an edge list, adjacency matrix, linked-list, or sparse matrix. However, these formats do not allow for metadata storage and, therefore, several other XML-like formats, such as the BioPAX [23], SBML [24], PSI-MI [25], CML [26], and CellML [27] have been introduced. For example, the Systems Biology Markup Language (SBML) is mainly used for biochemical networks and biological processes, the Biological Pathway Exchange format (BioPAX) for biological pathways, the PSI-MI format for data exchange, and the CellML for mathematical models. Notably, both the GraphML [28] format for storing node and edge attributes and the JavaScript Object Notation (JSON) format, which is mainly used for web-based applications, are two of the most widely-used options when building applications. Various databases provide their interaction data in simple text format (tab or comma delimited) or directly in a database schema (e.g., SQLite). To this end, it is worth mentioning that the NDEx [29] is an open-source framework for network sharing. Finally, as each file format comes with certain structural rules, many format-specific parsers (e.g., R/Bioconductor, Biopython) have been implemented and are currently available to facilitate data manipulation.

3. Bioentity Interaction Databases

Interaction databases can be categorized based on three main criteria; (i) their type of interactions, (ii) source of information and (iii) data curation [30]. These criteria are also used to organize and describe the bioentity interaction databases presented in the following subsections of this review. The interaction type essentially defines the identity of each database. For example, protein interaction databases describe the physical, and often functional, interactions of proteins with other proteins or small molecules. Similarly, nucleic acid interaction databases contain the interactions of nucleic acids with various other cellular components, while gene co-expression databases describe interactions based on similar gene expression patterns.

As far as the source of information is concerned, interaction databases can be grouped into three main categories based on their data-acquisition policy. These are (i) primary, (ii) secondary, and (iii) predictive databases [31]. Primary databases directly collect experimental interaction evidence from primary sources, i.e., from scientific publications or from deposited interaction datasets, such as those derived from high-throughput experiments. On the other hand, secondary databases do not collect information from primary sources; instead, they combine and annotate data curated by several primary databases in a single repository. Finally, predictive databases contain not only experimental interaction evidence but also computationally predicted interactions, derived from various methods, such as sequence or structure analysis, or from automatic methods for parsing the literature (e.g., text mining).

The final criterion in classifying interaction databases is their data curation policy; that is, the way the information was collected along with levels of detail and the description, annotation, and classification of this information. Data acquisition can be manual (i.e., handled by curators, or by the scientific community), automated (performed using computational methods), or a combination of the two. As far as the level of curation

is concerned, most interaction databases fall between two extremes, lightly or deeply curated. Lightly curated databases aim to publish the maximum amount of interaction information, without necessarily focusing on details. These interactions are often obtained computationally, through automated methods, such as text mining. As a result, lightly curated databases often contain potentially erroneous interactions, as well as redundant or overlapping information. On the other hand, deeply curated databases offer detailed information on biomolecular complexes and the interacting partners involved. This information is periodically manually annotated, validated through multiple sources, and checked for redundancy; the drawback to this manual, detailed approach is that deeply curated databases often contain significantly fewer interactions [30].

Apart from the above, another database aspect that needs to be taken into account is data availability. Two database features pertaining to this aspect are the existence of programmatic access options and the database's license. Programmatic access refers to the ability to parse a database's contents programmatically, thus allowing the automated retrieval of multiple entries. Its existence in an interaction database can be very important, especially since the analysis of biomolecular interactions in Systems Biology often involves parsing large amounts of data (hundreds of thousands, or even millions of interactions). Programmatic access can typically be performed through an Application Programming Interface (API), dedicated modules in programming languages such as Python or R, or with extensions/plugins written for external applications (e.g., Cytoscape apps) [32]. As far as the licensing model is concerned, it can depend on various factors, from the data sources of each repository to the policies of the foundations hosting the databases. Generally speaking, most of the databases covered in this review offer free access to their data. In some cases, one of the commonly free access licenses is used (e.g., Creative Commons, GNU/GPL, Apache license etc.), while other databases simply offer their data without adopting a license model at all. However, some databases may impose restrictions, by requiring registration with academic credentials, or by offering only paid access to some of their data. Both the licensing model and the existence of programmatic access are evaluated for the databases presented in this review.

3.1. Gene Co-Expression

The key assumption in the construction of co-expression networks is that two genes which are functionally related tend to have similar expression patterns. Hence, poorly characterized genes can be functionally annotated through potentially related genes with similar expression patterns and, as a potential corollary, similar functions [33]. Gene co-expression networks are usually generated by analyzing data from high-throughput gene expression profiling technologies, such as microarrays or RNA–Seq. Normally the co-expression similarity is calculated with the use of metrics such as Pearson or Spearman. In this section, we investigate databases which host such co-expression networks as well as information describing gene-gene relationships across various organisms.

COXPRESdb [34] is a repository that retrieves condition-independent co-expression information from 11 different organisms and focuses on protein-coding RNAs. The major strength of this database is the comparison of multiple co-expression data derived from different transcriptomic technologies (RNA–seq and microarrays) for various species (human, mouse, rat, chicken, zebrafish, fly, nematodes, monkey, dog, budding yeast, and fission yeast). Specifically, the latest update combines gene expression data from 23 different co-expression platforms, of which 123 experiments concern humans, 154 mouse and 154 rat, released by Gene Expression Omnibus (GEO) [35]. In total, COXPRESdb hosts 12 co-expression networks for various species created from ~157,000 microarray and 10,000 RNA–seq samples. Interspecies comparison reveals the evolutionary relationships, whereas the verification of co-expression interactions from multiple platforms minimizes errors [36]. Additional functionalities include: (i) querying of multiple genes simultaneously, (ii) applying topological network analysis and (iii) module detection.

The Search Tool for the Retrieval of Interacting Genes and proteins (STRING) database [37] (described in more detail in Section 3.3.1) primarily hosts protein–protein interactions for more than 14,000 organisms. However, among the several evidence interaction channels (multi-edged networks), one is dedicated to gene co-expression. The majority of the data for this channel is obtained from transcriptomic technologies as well as proteomic expression data (e.g., ProteomeHD database) [38]. In the co-expression network, every pair of genes with similar expression patterns is scored according to how strong the correlation is. The database offers a number of resources for the analysis of interactions, including a versatile REST API and an interface for Cytoscape [39,40], including a specially designed app (stringApp) [41].

GeneMANIA [42] provides gene co-expression networks and comprises functional gene similarities for nine different organisms (*A. thaliana*, *C. elegans*, *D. rerio*, *D. melanogaster*, *E. coli*, *H. sapiens*, *M. musculus*, *R. norvegicus* and *S. cerevisiae*). It integrates genomics and proteomics data from various sources, such as GEO, the Biological General Repository for Interaction Datasets (BioGRID) [43], IRefIndex [44], and Interologous Interaction Database (I2D) [45]. In GeneMANIA, users can query for closely co-expressed genes among 2300 networks, which consist of approximately 600 million interactions involving ~164,000 genes.

GeneFriends [46], Immuno-Navigator [47] and COEXPEDIA [48] are databases dedicated to gene correlation and transcript expression for *H. sapiens* and *M. musculus*. Specifically, GeneFriends is a tool for inferring gene interactions from co-expression networks, while it provides updated gene and transcript networks based on RNA–seq data from 46,475 human and 34,322 mouse samples. The Immuno-Navigator database gathers cell-type specific gene expression and co-expression data derived from the immune system. Currently, it contains data from 4639 human samples, obtained from 19 cell types from 191 studies, as well as 3434 mouse samples, obtained from 24 cell types from 261 studies. On the other hand, COEXPEDIA focuses on co-expression patterns derived from data from individual studies and which are associated with biomedical information related to anatomy, diseases, and chemicals. At the moment, COEXPEDIA contains 8 million interactions inferred from 384 and 248 GEO studies on humans and mice.

Regarding human tissue-specific co-expression networks, HumanBase [49], HumanNet [50], and Brain gene EXPression (BrainEXP) [51] cover the vast majority of known interactions. HumanBase includes the GIANT web server, which provides human tissue-specific networks via multi-gene queries. The gene associations are obtained from projects such as the Encyclopedia of DNA Elements (ENCODE) [52] and The Cancer Genome Atlas (TCGA) [53]. HumanNet (v2) aims to predict gene co-expression interactions and gene-disease associations through a complex combination of a four-level inclusive hierarchy of the human gene networks. The levels consist of protein–protein interactions, co-functional links from genomics data and two extended functional networks by either co-citations or interologs from other species. Finally, BrainEXP provides data about individual co-expressed genes in normal human brains. It currently stores data from 4567 samples from 2863 healthy individuals.

Various databases focus on plants, especially on *A. Thaliana*; ATTED-II [54], CoP [55], PlaNet [56] and PLANEX [57] cover several plant species, while the Arabidopsis Co-expression Tool (ACT) [58] and AraNet [59] are *A. Thaliana*-specific. The latter two provide co-expression patterns involving 21,273 *A. Thaliana* genes from microarrays and genome-scale functional networks. ATTED-II provides co-regulated gene relationships from microarrays and RNA–seq to estimate gene functions. CoP is focused on biological processes, comprising a microarray-based co-expression network for eight different plant species. PlaNet is a platform which integrates several web-tools dedicated to visualization and analysis of co-expression networks for photosynthetic organisms, while PLANEX is a plant co-expression database, based on publicly available GeneChip [60] data obtained from NCBI GEO. Finally, there are two Algae-dedicated co-expression databases based on Next-Generation Sequencing (NGS) data, ALCOdb [61] and AlgaePath [62], whereas

DanioNet [63] is a zebrafish-specific repository. Gene co-expression databases are briefly described in Table 1. Links are summarized in Supplementary Table S1.

Table 1. Gene co-expression databases.

Database Name	Interaction Type	Data Category	Curation Type	Organisms	Data License [1]	Programmatic Access
COXPRESdb [34]	Gene co-expression	Primary	Manual	11 species	Free (CC-BY)	REST API
STRING [37]	Gene co-expression, Protein–protein interactions	Secondary	Automated	>14,000 species	Free (CC-BY)	REST API, R & Python packages, Cytoscape import & app (stringApp) [41]
GeneMANIA [42]	Gene co-expression	Predictive	Automated	9 species	Free	Command line tool, Cytoscape app [64]
GeneFriends [46]	Gene and transcript co-expression	Primary	Manual	2 (*H. sapiens*, *M. musculus*)	Free	Cytoscape import
Immuno-Navigator [47]	Cell-type specific co-expression, related to immune system	Primary	Manual	2 (*H. sapiens*, *M. musculus*)	Free	N/A
COEXPEDIA [48]	Functional co-expression patterns	Primary	Manual	2 (*H. sapiens*, *M. musculus*)	Free	N/A
HumanBase [49]	Tissue-specific co-expression	Primary	Manual	*H. sapiens*	Free	REST API
HumanNet [50]	Gene co-expression	Predictive	Automated	*H. sapiens*	Free (CC-BY-SA)	N/A
BrainEXP [51]	Brain region co-expression	Primary	Manual	*H. sapiens*	Free	N/A
Human Gene Correlation Analysis (HGCA) [65]	Gene co-expression	Predictive	Automated	*H. sapiens*	Free	N/A
ATTED-II [54]	Co-expressed gene networks	Primary	Manual	9 Plant species	Free (CC-BY)	REST API, RDF API (SPARQL endpoint), Cytoscape import
CoP [55]	Co-expressed gene networks	Primary	Manual	8 Plant species	Free	N/A
PlaNet [56]	Co-expressed gene networks	Primary	Manual	11 Plant species	Free for academic users (Max Planck institution license)	Python standalone, Cytoscape import
PLANEX [57]	Co-expressed gene networks	Primary	Manual	8 Plant species	Free (CC-BY)	N/A
ACT [58]	Co-expressed gene networks	Primary	Manual	*A. thaliana*	Free	N/A
AraNet [59]	Co-expressed gene networks	Primary	Manual	*A. thaliana*	Free	N/A
ALCOdb [61]	Gene Co-expression	Primary	Manual	Microalgae	Free (CC0)	N/A
AlgaePath [62]	Gene co-expression	Primary	Manual	Algae	Free	N/A
DanioNet [63]	Gene associations	Predictive	Automated	*D. rerio*	Free	N/A

[1] The license type adopted by each database. In cases where a specific license type is used, it is given in parentheses. License abbreviations: CC-BY, Creative Commons–Attribution (BY); CC-BY-SA, Creative Commons–Attribution–ShareAlike; CC0, Creative Commons public domain.

3.2. RNA and ncRNA Interaction Databases

Non-coding RNAs (ncRNAs) are functionally important due to their interaction with other biomolecules even though they are not translated into proteins. RNA interacting biomolecules may include DNA, other RNAs/ncRNAs, proteins and other chemical compounds, thus influencing various cellular processes. Therefore, in this section, we mainly discuss databases focusing on such RNA interactions.

RNA Bricks2 [66] is a frequently updated database that contains 3D RNA structure motifs and their contact points. It contains more than 4300 RNA structures and RNA–protein complexes originating from the Protein Data Bank (PDB) [67]. RNA network structures are presented as interactive graphs, where nodes depict the basic secondary structure of motifs and edges represent either shared bases or tertiary interactions. RNA Bricks2 contains structure-quality score annotations and offers tools that enable the search of RNA 3D structures and comparisons. It is interconnected with PDB, Rfam [68] and UniProt [18] as the user can browse entries by using identifiers from these databases. Users can query using a FASTA file format, while selected structure data from RNA Bricks2 can be downloaded in PDB format along with a text file that includes a list of interactions. Contact base pairs are annotated through the ClaRNA [69] software and the respective file can be downloaded in CSV format.

As far as ncRNA interaction databases are concerned, NPInter [70] contains functional interactions among various types of ncRNAs (except for tRNAs and rRNAs) and biomolecules such as proteins, RNAs and DNAs. The latest version of NPInter (v4.0) contains a total of 1100,658 interactions, composed by: (i) manually curated literature interactions, (ii) processed high-throughput sequencing data and (iii) interaction data from the RISE [71] database. The interactions concern 35 organisms, while accompanying metadata provide information regarding the interaction class (binding, regulatory, correlation or co-expression) and the tissue/cell line of the experiment, where applicable. NcRNA entries are annotated with identifiers from NONCODE [72], miRBase [73] and circBase [74], while proteins from UniProt [18], Ensembl [20], UniGene (discontinued) and RefSeq [75]. Interactions are downloadable in text format.

Another ncRNA interactions database, snoDB [76], contains manually curated snoRNA interaction data (currently 2089 interactions) from *H. Sapiens*, derived from established databases and literature. SnoDB data are presented in an interactive table view on site and are downloadable in TSV, BED and XLSX file formats. Additional metadata include host genes, species conservation, orthologs and tissue expression where applicable. As snoDB content emanates from various external databases, multiple IDs are used to refer to the same snoRNA with respective links to UCSC [77], RefSeq [75], HUGO Gene Nomenclature Committee (HGNC) [78], Ensembl [20], RNAcentral [79], NCBI, Rfam [68], snoRNAbase (also known as snoRNA–LBME-db) [80], snOPY [81], snoRNA Atlas [82] or RISE [71].

Plant Non-coding RNA Database (PNRD) [83], an updated version of PMRD (plant microRNA database) [84], catalogues plant-related ncRNAs and is currently composed by 25,739 entries, from 11 different ncRNA types across 150 plant species. Nevertheless, its interaction entries regard only miRNAs and their targets, consisting of 178,138 target pairs across 46 plant species. These targets include protein-coding genes, literature ncRNAs and NONCODE lncRNAs and target information has been enriched through psRNATarget [85] and the literature. MiRNA sequence information is mainly derived from miRBase [73] and PMRD [84], while other ncRNAs are mined from NONCODE (v4), Rfam [68], tasiR-NAdb [86], GtRNAdb [87], The Arabidopsis Information Resource (TAIR) [88] and the Rice Genome Annotation Project (RGAP) [89]. All database ncRNA sequences are downloadable in text/FASTA format while miRNA–target information and relative literature, in tabular text format. We also note that PNRD hosts a Cytoscape service for constructing miRNA–gene regulatory networks.

Tarbase [90] also focuses on miRNA interactions. It contains manually curated, experimentally supported, miRNA–gene interactions from the literature as well as from raw libraries like GEO and the DNA Data Bank of Japan (DDBJ) [91]. It contains more than 1 million entries that correspond to 670,000 unique, experimentally supported miRNA–target pairs. The interactions within Tarbase are derived from more than 33 high-throughput techniques, applied to 516 cell types and 85 tissues, under 451 experimental conditions, across 18 species. This information is provided as query metadata along with the positive/negative miRNA–target regulation per species and binding locations. Tarbase also incorporates data from miRTarBase [92] and miRecords [93], and supports Ensembl and

miRBase identifier queries. The database is interconnected with the Ensembl genome browser and other DIANA-tools, including microT-CDS [94] for in silico identification of miRNA targets, LncBase v2.0 [95] for miRNA–lncRNA interactions identification (see Section 3.2.2) and DIANA-miRPath v3.0 [96] for miRNA functional characterization. Data is available in text format, after filling a request form on the site. The core information of all aforementioned RNA interaction databases is appended on Table 2.

Table 2. RNA interaction databases.

Database Name	Interaction Type	Data Category	Curation Type	Organisms	Data License	Programmatic Access
RNA Bricks2 [66]	RNA–(RNA, proteins, metal ions, water molecules, small molecule ligands)	Secondary	Manual	All organisms with available RNA structures in PDB	Free	N/A
NPInter [70]	lncRNA–miRNA–ncRNA–circRNA–vtRNAs-snoRNA–snRNA–sRNA–piRNAs-mRNA–protein-pseudogene-DNA	Primary, Secondary	Manual	35 species	Free	N/A
snoDB [76]	snoRNA–(rRNA, snRNA, non-canonical)	Primary, Secondary	Manual	*H. Sapiens*	Free	N/A
PNRD [83]	miRNA–(protein-coding genes, ncRNAs, lncRNAs)	Primary, Secondary	Automated	46 plant species	Free	Cytoscape service
Tarbase [90]	miRNA–gene	Primary, Secondary	Manual	18 species	Free	N/A

3.2.1. RNA–Protein Interactions

The inherent instability of RNA molecules coupled with the diversity and versatility of their functions are partly responsible for their constant chaperoning by a plethora of different protein complexes. Besides the regulatory binding of proteins to RNA molecules, RNAs also interact with specific proteins to perform specialized functions [97]. Notably, despite the significant contribution of recently developed transcriptome-wide methods and integrative analyses, deciphering the intricate principles of RNA–protein networks is undoubtedly challenging.

In order to facilitate the understanding of these complex, yet vital, interactions, RNA–protein interaction databases integrate experimentally validated and computationally predicted data from published literature and high-throughput technologies, visualizing RNA interactomes [98]. Regarding the contents provided by each resource, RNA–protein interaction databases may be characterized either as comprehensive, incorporating data from multiple sources, specialized, focusing on interactions validated by various experimental methods or predictive, utilizing computational methods, apart from experimental data, to predict possible interactions.

Protein–RNA interaction database (PRD) [99] is a comprehensive database which integrates literature-based physical RNA–protein interactions at the gene level. The current version of PRD contains 10,817 interactions between proteins and protein-coding RNAs, tRNAs, rRNAs, miRNAs, and viral RNAs in 22 organisms, corresponding to 1539 unique gene pairs. Each interaction is enriched with further information curated from multiple other resources, concerning RNA and protein binding sites/motifs, Gene Ontology (GO) [100] terms, detection methods, and biological functions.

The RNA Interactome Database (RNAInter) [101], previously named RAID, is another comprehensive and manually curated database of RNA-associated interactions (RNA–Protein/RNA–RNA), integrating experimentally validated and computationally predicted data from the published literature and 35 other resources. Apart from the fuzzy/batch search, interaction networks, and RNA dynamic expression data that are included in RNAInter, four RNA interactome tools are also embedded, namely, RIscoper [102], IntaRNA [103], PRIdictor [104], and DeepBind [105]. Currently, RNAInter contains 41,322,577 RNA-associated interactions of 22 different RNA types in 154 species, including 34,106,998 RPIs. Identifiers for external databases, such as miRBase, NCBI, HGNC,

Ensembl, Online Mendelian Inheritance in Man (OMIM) [106,107], Human Protein Reference Database (HPRD) [108], and UniProt are also provided. Data can be browsed by interaction type, detection method or species and are downloadable in text format, as well as obtainable through an API.

Furthermore, POSTAR3 [109] and doRiNA [110] constitute more specialized repositories, concerning post-translational regulatory RNA–Protein interactions. Both databases provide functional association prediction and contain structural information about binding sites of RNA–binding proteins and RNAs originating from cutting-edge high-throughput sequencing techniques. In particular, POSTAR2 provides the largest collection of RNA–binding protein (RBP) binding sites and functional annotations in 6 species, namely human, mouse, fly, worm, *A. thaliana* and yeast. Three modules (RBP, RNA, and translatome modules) and RBP–RNA interaction network in *H. sapiens* are supported, offering both functional and structural insights into translational and post-translational regulation. On the other hand, doRiNA integrates experimentally validated RBPs and miRNA target site data for *H. sapiens*, *M. musculus*, and *C. elegans*, while computational methods for all species are also used for miRNA target site prediction.

As far as predictive databases are concerned, Protein–RNA Interface Database (PRIDB) [111] contains a total of 30,056 RNA–Protein interactions (5694 unique RNA chains and 1702 unique protein chains) and incorporates structural information facilitating the analysis of RNA–protein complexes and their interface, by providing a user-friendly format. The RNA–Binding Protein DataBase (RBPDB) [112] is a manually curated resource of experimentally observed RNA–binding data for 1171 RBPs in humans, mice, flies, and worms. Finally, RNA binding site DataBase (RsiteDB) [113] is another predictive database aiming to describe, classify, and predict interactions between protein binding sites and single-stranded RNA bases. Table 3 contains information regarding all aforementioned RNA–protein interaction databases.

Table 3. RNA–Protein interactions databases.

Database Name	Interaction Type	Data Category	Curation Type	Organisms	Data License	Programmatic Access
PRD [99]	RNA–protein	Primary	Manual	22 species	Free	N/A
RNAInter (RAID) [101]	RNA–protein, RNA–RNA, RNA–DNA, RNA–compound, RNA–histone modification	Primary, Secondary, Predictive	Manual	154 species	Free	REST API
POSTAR3 [109]	RNA–protein	Primary	Automated	6 (*H. sapiens, M. musculus, C. elegans, D. melanogaster, A. thaliana, S. cerevisiae*)	Free	N/A
doRinA [110]	miRNA–protein	Primary	Manual	4 (*H. sapiens, M. musculus, C. elegans, D. melanogaster*)	Free	REST API, Python API
PRIDB [111]	RNA–protein	Secondary, Predictive	Automated	5 (*H. sapiens, M. musculus, C. elegans, D. melanogaster, E. coli*)	Free	N/A
RBPDB [112]	RNA–protein	Primary, Predictive	Manual	4 (*H. sapiens, M. musculus, C. elegans, D. melanogaster*)	Free	N/A
RsiteDB [113]	RNA–protein	Secondary, Predictive	Manual	All organisms with available 3D protein-RNA structures in PDB	Free	N/A

3.2.2. LncRNA–Target Interactions

Long non-coding RNAs (lncRNAs) are transcripts defined as greater than 200 nucleotides in size, which lack protein coding capacity. LncRNAs play a crucial role in biological processes such as cell cycle regulation, immune responses, and embryonic stem cell pluripotency. Studying lncRNAs is also important in order to understand the underlying mechanisms related to the pathogenesis of various diseases, such as cancer. Here,

we evaluate relevant databases that compile and integrate information about lncRNA–target interactions.

LncRNA2Target [114] contains a comprehensive repository of lncRNAs and their target genes regarding *H. Sapiens* and *M. Musculus*, hosting 152,137 associations from 1047 manuscripts (manual literature extraction) and 224 datasets. High-throughput microarray or RNA–seq datasets were used to identify all differentially expressed genes by checking expression before and after knockdown of lncRNAs. All lncRNAs were annotated by NCBI Genbank, Ensembl, GENCODE [115], and Entrez ID/symbols and gene targets by Entrez ID/symbols [116]. Furthermore, each interaction provides a link to the relative publication through a PubMed identifier (PMID). Users can browse and download all lncRNA–target interaction data in text and XLSX format.

EVLncRNAs [117] contains lncRNA interactions validated by low-throughput experiments, such as qRT-PCR, knock-down, western blot, northern blot, and luciferase reporter assays. These interactions are mainly curated from the literature and consist of lncRNA interactions with biomolecules such as DNA, RNA, proteins, and transcription factors (TFs), similar to the RNAInter database, which has already been discussed in the section "*RNA–protein interactions*". EVLncRNAs also incorporates lncRNA interaction entries from other databases, such as lncRNAdb [118] (discontinued), LncRNADisease [119], and Lnc2Cancer [120] (both discussed below), along with enhanced, manually curated metadata. Its current version (v2.0, July 2020) covers a total of 4010 lncRNAs and 6244 biomolecular interactions across 124 species, and 11,257 lncRNA–disease associations across 1082 diseases. Additional metadata are offered for each entry, such as chromosome position, assembly version, type of interaction (binding, regulation or co-expression), lncRNA class, and validation method. Accession numbers to NCBI and Ensembl, as well as PMID links are provided. EVLncRNAs allows data downloading in XLSX format. In addition, EVLncRNAs provides network visualization of all available interactions on site, as well as links to tools for lncRNA prediction. However, predicted interactions are not included in the database itself.

DIANA-LncBase [95] accommodates experimentally verified tissue and cell type specific miRNA–lncRNA interactions in *H. sapiens* and *M. musculus*. MiRNA–lncRNA interactions are derived from the manual curation of published literature and the analysis of high-throughput datasets. The current version of DIANA-LncBase (v3.0) catalogues ~240,000 interactions regarding ~500,000 entries. Interactions can be retrieved by querying with miRNA or gene names/identifiers (for lncRNAs) from Ensembl, RefSeq, miRBase, and the publication of Cabili et al. [121]. Additional filtering criteria, such as species, cell types/tissues, and methodologies can be applied. A second module in DIANA-LncBase contains information about lncRNAs in different cell types, as well as their subcellular localization, in the nucleus and/or cytoplasm. Queried miRNA–lncRNA interactions and lncRNA expression profiles are downloadable in CSV and JSON format, even though there is no option to download all database interactions.

ChIPBase [122] contains interactions of trans-acting factors, such as TFs, transcription cofactors (TCFs), chromatin-remodeling factors (CRFs), DNA-binding proteins, and histone modifications with various types of RNAs such as miRNA, lncRNAs, and other ncRNAs, from ChIP-seq data across 10 species. ChIP-seq peak datasets of these trans-acting factors are retrieved from GEO, ENCODE, the modENCODE project [123], and the NIH Roadmap Epigenomics Project [124]. All experiments contain metadata regarding cell line/tissue, dataset IDs, and Ensembl IDs for the studied genes. Experiments within ChIPBase can be queried and downloaded in text format (one experiment at a time).

As far as lncRNA–disease association databases are concerned, LncRNADisease [119] is a collection of experimentally and/or computationally validated lncRNA–disease and circular RNA–disease associations. The current version (v2.0) contains more than 200,000 lncRNA–disease and circRNA–disease associations in total, across 4 species. All experimentally supported data are manually retrieved from the literature and the computationally supported data were predicted by four algorithms, LRLSLDA [125], LDAP [126], RWRlncD [127], and

LncDisease [128]. Each ncRNA–disease association entry contains detailed information, including gene symbol, gene category, disease information, and regulatory relationship, along with a confidence score. Each disease name is mapped to Disease Ontology (DO) [129] and Medical Subject Headings (MeSH) [130]. All database interactions are downloadable in XLSX format.

Lnc2Cancer [120] is another lncRNA–disease interaction database that focuses on cancer subtypes. The database provides lncRNA–cancer and circRNA–cancer associations, along with their mode of regulation (up or down), supported by experiments. The current version (v3.0) contains 10,303 entries for 2659 human lncRNAs, 743 circRNAs, and 216 cancer subtypes. For every lncRNA or circRNA interaction, flags are provided as additional metadata, relative to their involvement in regulatory mechanisms (miRNA, TF, genetic variant, methylation, and enhancer), biological functions (cell growth, apoptosis, autophagy, EMT, immunity, and coding ability) and clinical applications (metastasis, recurrence, circulation, drug-resistance, and prognosis). lncRNA names are coherent with names from HGNC, Ensembl, GENCODE, Genbank or Refseq, whereas circRNA names are derived from circBase or Circbank [131]. Online data can be browsed by lncRNA/circRNA or cancer names and all interaction data are downloadable in XLSX format.

NONCODE [72] catalogues a variety of ncRNAs, focusing mainly on lncRNAs. NON-CODE entries are derived from the literature and the latest versions of several public databases (Ensembl, RefSeq, lncRNAdb and LNCipedia [132]). The bioentity interaction data within NONCODE concern lncRNA–disease associations (32,226) obtained from four lncRNA–disease databases (LncRNADisease, Lnc2Cancer, Mammalian ncRNA–Disease Repository (MNDR—discussed in Section 3.5) [133] and LncRNAWiki [134]) and lncRNA–SNP associations obtained from LincSNP [135] (724,579 total SNPs), which is further discussed below. All entries are accompanied by the respective PMID and each SNP provides a link to dbSNP [136]. NONCODE contains detailed information regarding the sequence, structure, expression, function, conservation, and disease relevance of lncRNAs. All NONCODE sequences are downloadable in FASTA format and all lncRNAs and their respective genes in BED format. However, there is no dedicated download page for the bioentity interaction datasets.

Another lncRNA–disease related database, lncRNASNP2 [137] provides information of SNPs in human and mouse lncRNAs, as well as their impact on lncRNA structure and function. lncRNASNP2 current version (v2) contains 10,205,295 SNPs on 141,353 *H. sapiens* lncRNA transcripts and 5,104,701 SNPs on 117,405 *M. musculus* lncRNA transcripts. lncR-NASNP2 transcripts are obtained from 170,002 NONCODE lncRNA genes. lncRNASNP2 also contains predicted lncRNA–miRNA interactions and lncRNA–disease associations. MiRNA sequences were collected from miRBase and disease-associated miRNAs from the Human microRNA Disease Database (HMDD) [138]. Moreover, lncRNASNP2 contains noncoding variants from COSMIC [139,140] cancer data as well as TCGA cancer mutations. All interaction data are downloadable in text format. Online search and prediction tools are also available, enabling the analysis of user-uploaded lncRNAs.

Finally, a similar SNP-centric database, LincSNP [135], stores and annotates disease or phenotype-associated variants, including SNPs, linkage disequilibrium SNPs (LD SNPs), somatic mutations, and RNA editing in human lncRNAs and circRNAs or their regulatory elements. The latter consist of transcription factor binding sites (TFBSs), enhancers, DNase I hypersensitive sites (DHSs), topologically associated domains (TADs), footprints, and open chromatin regions. LincSNP contains entries of experimentally supported variant–lncRNA/circRNA–disease/trait associations retrieved from the literature. LincSNp also incorporates lncRNA information from five databases (Ensembl, LncRBase [135], NON-CODE, LNCipedia, and GENCODE). Moreover, disease-associated SNPs were obtained from nine different sources (dbGaP [141], Genetic Association Database (GAD) [142], Gene-wide association study (GWAS) Central [143], Johnson and O'Donnell [144], the National Human Genome Research Institute (NHGRI) GWAS Catalog [145], PharmGKB [146], GWASdb [147], GRASP [148], and LnCeVar [149]). LD-SNPs were collected after analysis

by VCFtools [150] and somatic mutations from the COSMIC database [140]. In addition, associations between functional variants, lncRNAs, circRNAs, and their regulatory elements were constructed using BEDTools [151]. Queried interactions are downloadable in spreadsheet, CSV, and PDF formats, while all bioentity site interactions are also downloadable in BED format. All aforementioned lncRNA-related database information is summarized in Table 4.

Table 4. LncRNA–target interaction databases.

Database Name	Interaction Type	Data Category	Curation Type	Organisms	Data License [1]	Programmatic Access
LncRNA2Target [114]	lncRNA–gene	Primary	Manual	2 (*H. sapiens*, *M. musculus*)	Free	N/A
EVLncRNAs [117]	lncRNA–DNA, lncRNA–RNA, lncRNA–protein, lncRNA–TF, DNA-TF, peptide–protein, lncRNA–disease	Primary, Secondary	Manual	124 species	Free	N/A
RNAInter (RAID) [101]	lncRNA–DNA, lncRNA–RNA, lncRNA–protein, lncRNA–histone modification, lncRNA–compound	Primary, Secondary, Predictive	Manual	15 species	Free	REST API
DIANA-LncBase [95]	lncRNA–miRNA	Primary, Predictive	Manual, Automated	2 (*H. sapiens*, *M. musculus*)	Free	N/A
ChIPBase [122]	(TF, TCF, CRF, DNA-binding protein, histone modification)–(ncRNA, protein, lncRNA, miRNA)	Primary, Predictive	Automated	10 species	Free	N/A
LncRNADisease [119]	lncRNA–disease, circRNA–disease	Primary, Predictive	Manual	4 (*H. sapiens*, *G. gallus*, *R. norvegicus*, *M. musculus*)	Free	N/A
Lnc2Cancer [120]	lncRNA–cancer, circRNA–cancer	Primary	Manual	*H. sapiens*	Free	N/A
NONCODE [72]	lncRNA–disease, lncRNA–SNP-disease	Secondary	Automated	39 (16 animals, 23 plants)	Free (CC-BY-NC)	N/A
lncRNASNP2 [137]	lncRNA–SNP, lncRNA–disease, lncRNA–miRNA	Secondary, Predictive	Automated	2 (*H. sapiens*, *M. musculus*)	Free	N/A
LincSNP [135]	SNP-lncRNA, SNP-circRNA, LD SNP-lncRNA, LD SNP-circRNA, Somatic mutation-lncRNA, Somatic mutation-circRNA, RNA editing-lncRNA, RNA editing-circRNA	Primary, Secondary, Predictive	Manual	1 (*H. sapiens*)	Free (Open Source)	N/A

[1] The license type adopted by each database. In cases where a specific license type is used, it is given in parentheses. License abbreviations: CC-BY-NC, Creative Commons–Attribution–Non-commercial.

3.3. Protein Interaction Databases

In the following section we discuss databases containing protein interactions. As proteins are responsible for nearly every cell function, the investigation of their interactions is critical to the study of every biological process, as well as the study of diseases and the design of novel pharmaceuticals. As a result, the vast majority of currently available biomolecular interaction databases currently focus on proteins and their interactions, either with other proteins (protein–protein interactions), or with chemical compounds, such as ligands, drugs, and other substances (protein–small molecule interactions).

3.3.1. Protein–Protein Interactions (PPIs)

Proteins rarely act alone inside the cell. Instead, the vast majority of cell functions, from gene expression and metabolic pathways to structural support, cell growth, and cell death, are conducted by multiple proteins, frequently coordinating their action through the formation of protein complexes. Protein–protein interactions are of paramount importance in biological research. Studying the interactions behind a protein–protein complex that conducts a biological process is critical for elucidating the mechanisms that govern that

process, as well as for designing better treatments for the diseases that are caused when these interactions are disrupted. For this reason, a significant number of protein–protein interaction databases have emerged in the literature. In this subsection we present a subset of these databases, focusing mainly on repositories that can be of use in Systems Biology, and particularly in the creation and analysis of biological networks.

IntAct [152,153] is a large, open-source, manually curated molecular interaction database hosted by the European Bioinformatics Institute (EBI). All interactions contained in the database are derived from experimental results, obtained from the literature by the database's curators, or from interaction datasets submitted by the scientific community. IntAct is the largest biomolecular interaction database, as it currently holds more than 11 million binary interactions, the vast majority of which involve protein–protein complexes. In addition to its own data, the database also integrates experimental interaction evidence deposited in MINT [154,155], another major protein–protein interaction database (described in more detail below), as well as interactions derived from UniProtKB/Swiss-Prot and PDB [153]. Each interaction is annotated with details about the experimental procedures followed, as well as accompanied by relevant publications. This annotation evidence is also used to evaluate the confidence of each interaction, by applying a numerical score (Mi-score). Interactions are available for download in the PSI-MI format, both for the entire database and for manually selected datasets, dedicated to specific proteomes or diseases. In addition, database offers a number of resources for the analysis of interactions, including a REST API based on PSICQUIC (Proteomics Standard Initiative Common QUery InterfaCe), an import interface for Cytoscape [39] and a dedicated Cytoscape app (IntAct app) [156] and an embedded network viewer based on Cytoscape-web (a preliminary implementation of Cytoscape.js [157]). IntAct is a major participant in the International Molecular Exchange (IMEx) Consortium, a combined effort to provide an integrative, non-redundant dataset of biomolecular interactions [158].

Similar to IntAct, the MINT (Molecular INTeraction) database [154,155] focuses on experimental evidence derived from peer-reviewed publications. Its data consist of direct (physical) and indirect (functionally inferred) interaction evidence, with each binary interaction entry also containing information on promoter regions, mRNA transcripts, and the functional annotation of its protein partners. Starting from 2014, all interactions deposited in MINT are also integrated into the IntAct database [153]. In addition, MINT has adopted the database organization scheme and infrastructure of IntAct, including the use of the IntAct Mi-score to evaluate data confidence. In contrast to MINT, which exclusively relies on manual curation, the Database of Interacting Proteins (DIP) [159] catalogs experimentally determined interactions that are curated, both manually by expert curators and automatically, using computational approaches. DIP combines information from a variety of sources to create a single, consistent set of protein–protein interactions, each of which is annotated with cross-references to major biological repositories, such as UniProt, RefSeq, and GO. The Integrated Interactions Database (IID) [160] is a database of experimentally detected and predicted protein–protein interactions in 18 species, including human, 5 model organisms, and 12 domesticated species. IID collects experimental evidence from nine PPI databases and combines them with computational predictions using a number of different approaches. Each interaction is annotated with information on the experimental or computational procedure followed, as well as with cell type and tissue expression evidence, where available. In addition, IID offers a number of tools for the creation and topological analysis of PPI networks. MINT, DIP, and IID are all active participants in the IMEx Consortium and utilize the same REST API for programmatic access [158].

BioGRID [43] is a curated biological interaction database, comprising primarily protein–protein interactions, as well as genetic and chemical interactions and post-translational modifications. It strives to provide a comprehensive curated resource for all major model organism species while attempting to remove redundancy to create a single data map. BioGRID is currently one of the largest repositories of biomolecular interactions, containing

over 1,740,000 protein–protein interactions curated from both high-throughput datasets and individual focused studies, derived from over 70,000+ publications. Although BioGRID is not an active participant in the IMEx Consortium, it complies with the latter's guidelines and data format and has been classified as an IMEx Observer. The database provides programmatic access through a REST API, as well as the PSICQUIC API of IMEx members, in addition to integration with the Cytoscape network analysis program.

The STRING database is a large collection of experimentally derived and computationally inferred interactions [37]. STRING is a secondary database (or meta-database), compiling evidence from various sources, including experimental evidence from several primary PPI databases and computationally inferred interactions from literature text mining of scientific texts, de novo prediction of genomic features, and inference based on orthology with model organisms. A major aim of this database is the widest possible coverage of interactions in as many different organisms as possible. As such, STRING currently contains interaction evidence, either experimental or computational, for more than 14,000 species. Each interaction in STRING is annotated as direct/physical or indirect/functional, based on its data sources, and is ranked using a confidence score. STRING provides users with a versatile network visualization platform for the generation and analysis of PPI networks [161], including the analysis of topological features, as well as functional enrichment with terms from GO, KEGG (Kyoto Encyclopedia of Genes and Genomes) [162], Reactome [163], DO, Pfam, InterPro, and the Simple Modular Architecture Research Tool (SMART) [164]. In addition, the database offers programmatic access through a REST API, packages for the R and Python languages, direct integration with Cytoscape and a specially designed Cytoscape (stringApp), capable of building PPI networks with the characteristic STRING visualization style [41].

I2D, formerly known as Online Predicted Human Interaction Database (OPHID), contains protein–protein interactions for a number of mammals and other eukaryotic species [45]. It contains experimental evidence, obtained from high-throughput experiments as well as other databases, such as IntAct or BioGRID, and predicted interactions, inferred by mapping experimental results between different species. In addition, the database implements NAViGaTOR [165], a web-based network analysis platform for the visualization and analysis of PPI networks derived from its data. Although a significant portion of its content has migrated to IID [160], I2D remains one of the most comprehensive sources of known and predicted eukaryotic PPIs for model organisms, such as *S. cerevisiae*, *C. elegans*, *D. melonogaster*, *R. norvegicus*, *M. musculus*, and *H. sapiens*.

The Protein Interactions Network Analysis (PINA) database [166] is an integrated platform for the visualization and analysis of protein–protein interactions through the use of PPI networks. PINA consists of a non-redundant dataset of protein–protein interactions from seven model organisms (*H. sapiens*, *M. musculus*, *R. norvegicus*, *D. melanogaster*, *C. elegans*, *S. cerevisiae*, and *A. thaliana*), obtained from integrating data from five manually curated databases (IntAct, MINT, BioGRID, HPRD, and DIP). The database offers a large number of tools for the construction, visualization, and analysis of PPI networks. In addition, PINA has implemented search and visualization schemes for the analysis of interactions associated with various types of cancer, integrating PPI evidence with RNA–seq transcriptomes and mass spectrometry-based proteomes.

The Compartmentalized Protein–Protein Interactions (ComPPI) database [167] is a large collection of protein–protein interactions from four model organisms (*H. sapiens*, *D. melanogaster*, *S. cerevisiae*, and *C. elegans*). The database currently contains 791,059 interactions, obtained from several other PPI databases and manually curated for redundancy. These interactions are combined with evidence on protein subcellular localization, tissue and cell type expression evidence, and can be utilized to produce tissue-specific, cell-specific, and even subcellular location-specific interaction networks.

CORUM (Comprehensive Resource of Mammalian protein complexes) [168] is a collection of manually annotated protein complexes from mammalian organisms. Its annotation includes protein complex function, localization, subunit composition, literature references,

functional enrichment with GO terms, and associations with diseases. All information is obtained from individual experiments published in scientific articles, while data from high-throughput experiments is excluded. For this reason, the total number of interactions in CORUM is relatively small compared to other repositories; however, its data curation for each entry is significantly more detailed. Similar to CORUM, ComplexPortal [169] is a manually annotated and curated resource on macromolecular complexes, with emphasis on protein–protein and, to a lesser extent, protein–nucleic acid, and protein–small molecule complexes. Its interactions are derived from physical molecular interaction evidence extracted and cross-referenced from the literature and deposited in IntAct, by curator inference from information on homologs in closely related species. A key characteristic of ComplexPortal is its strict definition of the term "macromolecular complex" as an assembly of any two or more bioentities that are stable enough in vitro to be reconstituted and have been demonstrated to have a specific molecular function. This means that only constant protein–protein complexes are included in the database, while transient interactions, like those formed in processes such as signal transduction, are discarded. Another key feature of the database is its rich annotation, as each macromolecular complex is accompanied by detailed description of its stoichiometry, function, and relation to biological processes and diseases.

In addition to the major PPI databases described above, a number of web services that focus on specific systems also exist. These include databases which cover the interactions of specialized groups of proteins (or often a single protein class or family) with biomedical or pharmacological interest. Characteristic examples include major drug targets, such as G-protein coupled receptors (GPCRs), Receptor-Tyrosine Kinases (RTKs), and ion channels. A number of databases exist that specialize in describing the features of these proteins, including their interactions. For example, GPCRdb [170] contains both structural and functional evidence on the interactions of GPCRs with ligands and heterotrimeric G-proteins. In the same vein, hGPCRnet (Human GPCR network) provides a network visualization and associated database for PPIs in human GPCR signaling pathways, accompanied with annotation regarding cell and tissue expression [171]. As far as RTKs are concerned, one detailed resource is PrimesDB (Protein interaction machines in oncogenic EGF receptor signalling) [172], which focuses exclusively on PPIs related to the signaling mechanisms of EGFR and ERBB. EGFR and ERBB are two major biomarkers and drug targets in several diseases, including various forms of cancer. PrimesDB also offers tools for the visualization of PPI networks and is a participant in the IMEx Consortium. Finally, Channelpedia [173] is a community-driven database on the features of ion channels, including the interactions between their subunits. All of the aforementioned protein classes and their interactions are also collected and presented in the IUPHAR/BPS Guide to Pharmacology [174], a manually curated dataset of biomolecular interactions implicated in the signaling pathways of human, mouse and rat GPCRs, ion channels, RTKs and other drug targets. Apart from specialization into protein classes/families, databases also exist that provide information on PPIs observed in specific subcellular locations, such as organelles, vesicles, the cell membrane or the extracellular matrix. MitoProteome [175] is a database describing proteins present in mitochondria and their interactions. PerMemDB [176] collects experimental and computationally predicted information on peripheral membrane proteins, including their interactions with transmembrane proteins. Finally, the protein–protein interactions of the extracellular matrix (ECM) are covered by MatrixDB [177], a manually curated database on the PPIs of ECM proteins and proteoglycans. Table 5 presents a collection of available PPI databases.

Table 5. A collection of protein–protein interaction databases.

Database Name	Interaction Type	Data Category	Curation Type	Organisms	Data License [1]	Programmatic Access
IntAct [152]	protein–protein, protein–chemical	Primary & Secondary	Manual	1452 species	Free (Apache 2.0)	REST API (PSICQUIC) Cytoscape import & app [156]
MINT [155]	protein–protein	Primary	Manual	669 species	Free	REST API (PSICQUIC) Cytoscape import
DIP [159]	protein–protein	Primary	Manual (primarily) & Automated	834 species	Free (CC-BY-ND)	REST API (PSICQUIC) Cytoscape import
IID [160]	protein–protein	Predictive	Manual	18 (*H. sapiens*, 5 model organisms, 12 domesticated species)	Free	REST API (PSICQUIC) Cytoscape import
BioGRID [43]	protein–protein, protein–chemical, gene co-expression, PTMs	Predictive	Automated (primarily) & Manual	86 species	Free (GNU/GPL)	REST API, Cytoscape import
STRING [37]	protein–protein, gene co-expression	Secondary	Automated	>14,000 taxa	Free (CC-BY)	REST API (dedicated and PSICQUIC), R & Python packages, Cytoscape import & app (stringApp) [41]
I2D [45]	protein–protein	Predictive	Manual	8 species	Free	Cytoscape import
PINA [166]	protein–protein	Secondary	Manual	7 species	FREE (NIH GDS)	RESTful API, Cytoscape app [178]
ComPPI [167]	protein–protein	Secondary	Manual	4 (*H. sapiens, D. melanogaster, S. cerevisiae, and C. elegans*)	Free (CC-BY-SA)	N/A
CORUM [168]	protein–protein	Primary	Manual	3 (*H. sapiens M. musculus, R. norvegicus*)	Free (CC-BY-NC)	N/A
ComplexPortal [169]	protein–protein, protein–small molecule, protein–nucleic acid	Secondary	Manual	26 species	Free	N/A
GPCRdb [170]	protein–protein, specializes in GPCR structural complexes	Secondary	Manual	All metazoa with known GPCRs	Free (Apache 2.0)	REST API
hGPCRnet [171]	protein–protein, specializes in GPCR signaling pathways	Secondary	Manual	*H. sapiens*	Free	N/A
PrimesDB [172]	protein–protein, specializes in RTKs	Secondary	Manual	*H. sapiens*	Free (requires registration)	REST API (PSICQUIC)
Channelpedia [173]	protein–protein, specializes in ion channels	Primary	Manual	Mammals	Free (CC-BY-NC)	N/A
IUPHAR/BPS Guide to Pharmacology [174]	protein–protein, protein–chemical	Secondary	Manual	3 (*H. sapiens M. musculus, R. norvegicus*)	Free (CC-BY-SA)	REST API
MitoProteome [175]	protein–protein, specializes in mitochondria	Secondary	Manual	*H. sapiens*	Free	N/A
PerMemDB [176]	protein–protein, specializes in peripheral membrane proteins	Predictive	Manual	1009 species	Free	N/A
MatrixDB [177]	protein–protein, specializes in proteins of the extracellular matrix	Primary	Manual	12 model organisms, primary focus on *H. sapiens*	Free (CC-BY)	REST API (PSICQUIC)
ConsensusPath DB [179]	protein–protein, protein–chemical	Secondary	Manual	*H. sapiens*	Free for academic users	SOAP/WSDL API, Cytoscape app [180]

[1] The license type adopted by each database. In cases where a specific license type is used, it is given in parentheses. License abbreviations: CC-BY, Creative Commons–Attribution; CC-BY-ND, Creative Commons–Attribution–No derivatives; CC-BY-SA, Creative Commons–Attribution–Share Alike; CC-BY-NC, Creative Commons–Attribution–Non Commercial; GNU/GPL, GNU General Public License; NIH GDS, NIH General Data Sharing policy.

3.3.2. Protein–Small Molecule Interactions

The interactions of proteins with small molecules are vital for a wide range of biological functions. Inside a cell, small molecules play a twofold role as substrates, cofactors, and products in various biochemical reactions and as ligands or hormones which regulate protein functions [181]. Additionally, bioactive small molecules are often used as probes to identify therapeutic protein targets in drug discovery. Information on the structures, calculated properties, and bioactivities for a large number of chemicals and drug-like compounds is integrated in specialized databases, including PubChem [182], ChEMBL [183], and SIDER [184], with the aim of deciphering their properties and facilitating the drug discovery process. Another essential data resource involves databases focused on protein-chemical interactions, which gather information on the existence, stoichiometry, and biological or biomedical relevance of protein–small molecule complexes [185]. In Table 6, we have collected the relevant information on protein–small molecule interactions databases.

The primary, and most often used source of information in protein-small molecule interactions comes from databases focusing on experimentally studied protein-chemical complexes. DrugBank [186] is currently one of the most popular databases in this category. It is a manually curated and publicly available resource that provides primarily experimental information about small molecules (i.e., chemical, pharmacological, and pharmaceutical) and their protein targets (i.e., sequence, structure, metabolic pathways). In addition to drug-drug interactions, the database incorporates information for physical drug-target interactions and interactions with proteins known to metabolize a compound. Despite its name, however, the database does not focus solely on drugs, but also provides information on other compound types, such as metabolites. DrugBank is a frequently updated resource and its latest release (April 2021) integrates 14,524 drug entries, including 2684 approved small molecule drugs, 1464 approved biologics (proteins, peptides, vaccines, and allergenics), 131 nutraceuticals, and over 6654 experimental (discovery-phase) drugs. Finally, 5249 non-redundant protein (i.e., drug target/enzyme/transporter/carrier) sequences are associated with the aforementioned drug entries.

Another important, experimentally focused protein-small molecule interaction database is BindingDB [187]. BindingDB is a specialized repository of experimentally validated and measured binding affinities between drug-like compounds and therapeutically relevant protein targets. In particular, the latest version of BindingDB incorporates 41,328 Entries, each with a DOI, containing 2,259,122 binding data for 977,487 small molecules, which are mapped to 8516 protein targets. The database is continuously curated, deriving data mainly from scientific articles as well as from US patents. The search interface is well-designed and enables combined query criteria, including target name, sequence, molecular weight, source organism, compound name, SMILES string, binding potency, and article or patent information, while restricted searches by data source (e.g., BindingDB, ChEMBL, PubChem, and patents) is also allowed.

Apart from the primary databases described above, several secondary repositories also exist, combining information from multiple sources. STITCH (Search Tool for Interactions of Chemicals) [188], the "sister" database of STRING, is a manually curated resource to explore both known and predicted interactions between 9,600,000 proteins from 2031 eukaryotic and prokaryotic genomes and over 430,000 chemicals. Known interaction evidence is mainly derived from experimentally validated data as well as from manually curated datasets, including KEGG and Reactome. Protein–small molecule interactions are also accompanied by protein–protein interaction evidence, derived from STRING, to help illustrate the effect of chemicals on supramolecular assemblies. Text mining-based associations are compiled after parsing articles from PubMed Central (PMC) and PubMed. Like STRING, STITCH offers a REST API for programmatic access, as well as integration with Cytoscape.

Similar to STITCH, ConsensusPathDB [179] contains human interaction data referring to biochemical reactions and protein, genetic, metabolic, signaling, or drug-target

interactions as well as gene regulatory interactions involving different types of physical entities. SuperTarget [189] is another secondary database which hosts information from various databases. It contains 332,828 drug-target interactions along with pathways, protein–protein interactions, and drug-target-related ontologies, based on information retrieved from DrugBank, BindingDB, SuperCYP [190], ConsensusPathDB and CORUM. Metrabase (Metabolism and Transport Database) [191] is another comprehensive cheminformatics and bioinformatics database providing manually curated data extracted from published literature and other resources (TP-Search [192], ChEMBL, Human Protein Atlas [193], DrugBank, and UniProt) related to human metabolism and transport of chemical compounds across biological membranes and their interactions with proteins. Apart from transporter/enzyme-ligand associations, Metrabase incorporates experimentally validated information on non-substrate, non-inhibitor, and non-inducer compounds, aiming to assist the prediction of models based on the characteristics of both the positive and the negative class. Another example is Transformer [194], a database that focuses on the metabolism and transport of chemical compounds in the human body and, more specifically, xenobiotics. It contains integrated data on transformation, transportation, conjugation, and excretion of drugs, prodrugs, alimentary and Traditional Chinese Medicine compounds as well as their effect on enzymes and proteins, also providing the ability of interactive visualization.

A major field of interest in the study of protein–small molecule interactions involves the structural analysis of protein–ligand complexes. A number of specialized databases exist for this purpose. Some of these repositories are, essentially, subsets of PDB, containing analysis on the stoichiometry of protein–heteroatom interactions often found in the PDB entries of experimental 3D structures. PLI (Protein–Ligand Interaction) [195] and PLIC (Protein–Ligand Interaction Clusters) [196] are two such databases, which, as their names indicate, focus on protein-ligand associations. PLI database incorporates all the interactions between proteins and small molecules identified in the PDB with a Het_id code, while PLIC, by analyzing similarities in binding sites and employing computational tools, provides clusters of similar binding sites from PDB. Notably, PLIC, unlike other protein-ligand specific databases, not only reports similarities in interactions but also hosts data on attributes, such as binding site shape, protein–ligand contacts, and energetics among similar protein–ligand interactions.

In addition to the above, a number of structural databases also exist that complement crystallographic evidence with computational predictions derived from energy calculations, protein-ligand docking predictions or ab initio simulations. NLDB (Natural Ligand Database) [197] is a predictive database focusing on 3D protein-ligand interactions specifically in enzymatic reactions of metabolic pathways registered in KEGG. Based on the latest update, NLDB offers data about known human genome polymorphisms on protein structures, as well as 87,400 experimentally validated protein–ligand complex structures in PDB, defined as natural complexes, while 31,672 analog complexes and 70,570 ab initio complexes were predicted based on known protein structures in a complex with a similar ligand and by docking simulations accordingly. In cases of unknown complex structures, 3D interactions are predicted by implementing state-of-the-art software programs and subsequently generating a database of the 3D protein–ligand interactions in various enzymatic reactions. NLDB also provides an enrichment analysis function based on a set of KEGG compound IDs. PoSSuM (Pocket Similarity Search using Multi-Sketches) [198] is another predictive database that aims to retrieve similar small-molecule binding pockets on proteins with both different and similar global folds, contributing to structure-based drug discovery. It employs the *SketchSort* [199] algorithm for all-pair similarity searches, resulting in more than 163 million similar pairs of binding sites with annotations. Finally, PDID (Protein-Drug Interaction Database) [200] is a database of predicted protein–ligand interactions in the structural human proteome. PDID incorporates 9652 structures from 3746 proteins and provides a comprehensive set of 16,800 putative protein–drug interactions between 51 popular, FDA-approved drugs and over 10,000 protein structures, which were generated from approximately 1.1 million all-atom structure-based predictions.

The databases described above offer generalized information on the existence and properties of protein–small molecule complexes. However, specialized repositories also exist, focusing on the protein–chemical interactions associated with specific systems, phenotypes or diseases. One characteristic example involves cancer-specific databases, such as CancerDR [201], CAncerREsource 2 [202], and canSAR [203]. As their names indicate, these databases focus on protein–drug interactions related particularly to cancer. CancerDR incorporates 148 anticancer drugs which are mapped to 116 drug targets in 1000 cancer cell lines, also offering information about the function, structure, and gene sequences of each of these targets. In addition, CancerREsource 2 contains not only comprehensive data on 90,744 interactions between drugs and cancer-relevant protein targets, but also mRNA expression and non-synonymous mutation data from large-scale cancer genomics experiments. Similarly to the previously mentioned databases, canSAR is a comprehensive database which integrates protein–drug interactions between 564,407 proteins from all species and 3,312,866 compounds with unique chemical structures, as well as genomic and structural data.

Finally, one important category of specialized protein–small molecule databases focuses on the interactions of kinases, a large group of enzymes that participates in a multitude of cell processes and which, as such, has been implicated in a wide range of diseases. Kinase-specific databases include KIDFamMap (Kinase-inhibitor-disease family map) [204] and KLIFS (Kinase-Ligand Interaction Fingerprints and Structures database) [205], which contain protein–chemical information oriented to the kinase superfamily. In particular, KIDFamMap includes 189,987 kinase-inhibitor interactions derived from BindingDB and grouped into 1210 kinase-inhibitor families according to their pharma-interfaces, providing associations between 399 human protein kinases, 35,788 kinase inhibitors, and 339 diseases. KLIFS is another comprehensive kinase database that focuses on the interactions between 3499 kinase inhibitors and 312 kinases, based on the chemical structure of their catalytic domains. Finally, kinase-substrate interactions are also included in the IUPHAR/BPS Guide to Pharmacology [174], a database which, among other drug targets, includes a special section dedicated to the functionality and pharmacology of kinases.

Table 6. Protein–small molecule interactions databases.

Database Name	Interaction Type	Data Category	Curation Type	Organisms	Data License [1]	Programmatic Access
DrugBank [186]	protein–chemical	Primary	Manual	*H. sapiens*	Free for academic users (CC-BY-NC)	REST & SQL API
BindingDB [187]	protein–chemical	Primary	Manual, Automated	*H. sapiens*	Free (CC-BY-SA)	RESTful API
STITCH [188]	protein–chemical	Secondary, Predictive	Automated	2031 eukaryotic and prokaryotic genomes	Free for academic users (EMBL License)	REST API, Cytoscape app [41]
ConsensusPathDB [179]	protein–protein, protein–chemical	Secondary	Manual	*H. sapiens*	Free for academic users	SOAP/WSDL API, Cytoscape app [180]
SuperTarget [189]	protein–protein, protein–chemical	Secondary	Manual	*H. sapiens*	Free (CC-BY-NC-SA)	N/A
Metrabase [191]	protein–chemical	Primary	Manual	*H. sapiens*	Free (CC-BY-SA)	N/A
Transformer [194]	protein–chemical	Primary	Manual	*H. sapiens*	Free	N/A
PLI [195]	protein–chemical	Secondary	Automated	100 species	Free	N/A
PLIC [196]	protein–chemical	Secondary	Automated	All organisms with available protein–ligand structures in PDB	Free	REST API
NLDB [197]	protein–chemical	Secondary & Predictive	Automated	All organisms with available protein–ligand structures in PDB	Free	N/A
PoSSuM [198]	protein–small molecule	Secondary	Manual	*H. sapiens*	Free	N/A

Table 6. *Cont.*

Database Name	Interaction Type	Data Category	Curation Type	Organisms	Data License [1]	Programmatic Access
PDID [200]	protein–chemical	Secondary	Manual	*H. sapiens*	Free (Open Source)	N/A
CancerDR [201]	protein–chemical	Primary	Manual	*H. sapiens*	Free	N/A
CAncerREsource [202]	protein–chemical	Primary	Manual	*H. sapiens*	Free	N/A
canSAR [203]	protein–chemical	Primary	Manual	*H. sapiens*	Free (Public Domain)	N/A
KIDFamMap [204]	protein–protein, protein–chemical	Primary	Manual	*H. sapiens*	Free	N/A
KLIFS [205]	protein–chemical	Primary	Manual, Automated	*H. sapiens,* *M. musculus*	Free (Open Source)	REST API
TDR Targets [206]	protein–chemical	Primary	Manual	*11 host organisms, 35 pathogens*	Free	N/A
T3DB [207]	protein–chemical	Primary	Manual	*H. sapiens*	Free	N/A
BioLiP [208]	protein–chemical	Secondary	Semi-manual	~100 species	Free	N/A
Binding MOAD [209]	protein–chemical	Primary	Manual	>100 species	Free	N/A
ASDCD [210]	protein–chemical	Primary	Manual	*H. sapiens*	Free	N/A
PRRDB 2.0 [211]	protein–small molecule	Primary	Manual	7 species	Free	N/A

[1] The license type adopted by each database. In cases where a specific license type is used, it is given in parentheses. License abbreviations: CC-BY-SA, Creative Commons–Attribution–Share Alike; CC-BY-NC, Creative Commons–Attribution–Non Commercial; CC-BY-NC-SA, Creative Commons–Attribution–Non-Commercial–Share Alike.

3.4. Signaling and Metabolic Pathway Interactions

The interactions between all aforementioned molecules (DNA, RNA, proteins, etc.) cause cascading effects that may consequently affect biological mechanisms and processes through signaling and metabolic pathways. Analysis, processing, and interpretation of the vast and ever-growing amounts of -omics- data has made the implementation of pathway-oriented approaches necessary in most fields in Biology. The complexity of biological processes and their innumerable underlying interactions is most effectively and efficiently conceptualized with the representation and visualization of biological pathways [199]. Herein, we summarize a variety of databases dedicated to signaling and metabolic pathway interactions. Table 7 contains information on the discussed signaling and metabolic pathway interaction databases.

WikiPathways [212] is a manually curated database, launched in 2007 that is continuously updated on an almost daily basis. It is a collaborative platform based on the MediaWiki software, which incorporates customized graphical tools for editing and facilitating the representation of biological pathways and processes. The community has consistently been involved in the construction and revision of the pathway models comprising the database. Wikipathways also incorporates content from a large selection of databases, providing users the ability to query pathways from a variety of fields, such as Renal Genomics, the Reactome database, Diseases, Lipids and Micronutrients, through dedicated thematic sections (portals). The WikiPathways database includes a total of 2958 pathways (April 2021) consisting of proteins, genes, metabolites, and drugs, covering *H. sapiens* along with 29 other species and comprises 46,105 interactions between the represented bioentities. A designated wiki page is ascribed to each pathway, including features such as a pathway diagrams, short analysis, list of references as well as a list of all pathway components. The database content is freely accessible through a browser, an API or a specially designed Cytoscape app [213], and is downloadable in multiple formats, such as: (i) image formats (PNG, SVG, PDF), (ii) gene lists (GMT, Eu.Gene format), and (iii) machine-readable formats (GPML, RDF, BioPAX, XGMML, SBGN, SBML) for further pathway analysis by various tools, such as PathVisio [214] and Cytoscape [39]. Links to

other databases are provided for pathway components via the BridgeDb web service [215], such as NCBI, GO, Ensembl, UCSC Genome Browser, UniProt-TrEMBL, WIKIGENES [216], PDB, and IUPHAR/BPS Guide to Pharmacology.

Reactome [163] contains manually curated information derived from 33,453 literature references and in principle constitutes an extended metabolic map of *H. sapiens*. It includes detailed information of cellular processes on a molecular level, visualizing them in coherent data models. Such processes range from transport and DNA replication to signal transduction and intricate metabolic functions. Orthologous molecular reactions are also included for various other species, where applicable. The database (version 76) contains 10,867 human genes, 415 drugs, 1856 small molecules which serve as natural substrates, catalysts or regulators, 11,073 discrete proteins and 13,732 reactions incorporated into 2516 human pathways grouped in 26 superpathways (i.e., immune system, metabolism, diseases). The entities are linked to various databases of the relevant type, such as NCBI, Ensembl, UniProt, KEGG (Gene and Compound), ChEBI [217], PubMed, and GO. Reactome data is downloadable in various formats (DOC, PDF, SBML, SBGN, BioPAX 2, BioPAX 3, OWL, PNG, SVG, JPEG, GIF) and can be queried via an API, as well as through a Cytoscape app (ReactomeFIViz) [218].

KEGG [162], rather than constituting a single database, is an integrated database framework comprising 15 databases which are manually curated and an additional computationally generated one. Among them, KEGG PATHWAY [219] contains biological pathways represented graphically by manually drawn pathway maps, similar to Reactome. Listed entities include molecules, genes, proteins, and pathways, as well as disease genes and drug targets. Within the pathway maps, sequenced genes are linked to higher order functions in the context of individual cells or entire organisms. Such functions are depicted by a web of interactions and chemical reactions, drawn in the format of KEGG pathway maps, BRITE hierarchies, and KEGG modules. KEGG contains 34,042,792 genes, 781,759 pathways and 11,505 reactions pertaining to 545 eukaryotes, 6234 bacteria, and 343 Archaea (April 2021). Links are provided to other databases for bioentities included in the various pathways, such as GO, UniProt, other KEGG Databases, Rhea [220], NCBI, PubChem, CheMBL, KNApSAcK [221], PDB (Chemical Components) while PubMed references are also incorporated. The database provides an API, while the content can be downloaded in multiple formats, such as PNG, RDF and KGML. In addition, multiple Cytoscape apps have been developed, both from the database's curators and from third-party users, that integrate KEGG data visualization and analysis with Cytoscape [222–224].

Similar to the aforementioned databases, CBN (Causal Biological Network) [225] provides over 120 manually curated network models using Biological Expression Language (BEL) [226] integrating over 80,000 literature-based information pieces in order to describe signaling pathways and their biomolecular interactions. More specifically, it showcases the relationships in pathways across a wide spectrum of biological fields in 3 species (*H. sapiens*, *M. musculus*, *R. norvegicus*) using interactive network visualizations. These fields include cell fate, cell stress, cell proliferation, inflammation, tissue repair, and angiogenesis in the framework of the pulmonary and cardiovascular systems. Furthermore, the visualizations incorporate interacting entities, including proteins, DNA variants, coding and non-coding RNAs, chemicals, lipids, and processes (e.g., phosphorylation). Pathway compartments are annotated with metadata regarding species, tissue, and cell type and are also accompanied by their original references in PubMed. The networks can be downloaded in several formats, such as JSON GRAPH, SIF and SVG for further analysis.

Finally, the INDRA (Integrated Network and Dynamical Reasoning Assembler) database [227] is an automated system for the retrieval of interaction information on bioentities. Based on the INDRA model assembly system, the database aggregates knowledge extracted by multiple machine-reading systems from all available abstracts and open-access full text articles, and combines this with mechanisms from pathway databases. Queries allow searching for genes, chemicals, biological processes and other concepts of interest, and returns a ranked list of relevant interactions and molecular pathways. INDRA

sources include the PubMed and PubMed Central literature repositories, as well as a large number of other biological information databases, including DrugBank, BioGRID, CBN and many others. The database can be queried through the IDNRA REST API, as well as through a standalone application implemented in Python.

Table 7. Signaling and metabolic pathway interaction databases.

Database Name	Interaction Type	Data Category	Curation Type	Organisms	Data License [1]	Programmatic Access
WikiPathways [212]	pathway-pathway, intra-pathway biomolecular interactions (proteins, genes, drugs, metabolites)	Primary, Secondary	Manual	30 species	Free (CC0)	REST API, RDF API (SPARQL endpoint), PathwayWidget (embedded iframe), API libraries (R, Java, Perl, PHP, Python), Cytoscape app [213]
Reactome [163]	pathway-pathway, intra-pathway biomolecular interactions (proteins, genes, drugs)	Primary	Manual	16 species	Free (CC0)	REST API, ReactomeFIViz app (Cytoscape app) [218], reactome2py(Python), ReactomePA (R), DiagramJs (JavaScript)
KEGG Pathway [219]	pathway-pathway, intra-pathway biomolecular interactions (proteins, genes, drugs)	Primary	Manual	545 eukaryotes, 6234 bacteria, 343 Archaea	Free for standard & API access, paid license for FTP data access	REST API, KEGGscape (Cytoscape app) [222]
CBN [225]	protein-DNA-variant-RNA–ncRNA–chemical-lipid-process	Primary	Manual	3 species (*H. sapiens*, *M. musculus*, *R. norvegicus*)	Free	N/A
INDRA [227]	pathway-pathway, intra-pathway biomolecular interactions (proteins, genes, drugs)	Predictive	Automated	Any	Free (BSD license)	REST API & Python package

[1] The license type adopted by each database. In cases where a specific license type is used, it is given in parentheses. License abbreviations: CC0; Creative Commons–Public Domain.

3.5. Disease-Related Interactions

Perturbations in signaling and metabolic pathway interactions are often the cause of disease. Various databases contain such biomolecular interactions that are implicated in diseases. In this section, we discuss some of these disease-related databases covering biomolecule-biomolecule, biomolecule-disease and bioentity-disease interactions.

Regarding biomolecule-biomolecule interactions, the CIDeR database [228] contains interactions between disease-related biomolecules (and other bioentities, such as environment and phenotype) mainly for metabolic and neurological disorders. There are currently 109,779 interactions between 12,406 biological entries, derived from 11,341 parsed articles. The information is manually curated and each interaction entry is accompanied by its source PubMed ID and the related disease. CIDeR contains a variety of interaction types, such as expression increase/decrease, co-occurrence, co-localization, processing, phosphorylation, transport, and folding. It also holds information about interacting biomolecules such as genes, proteins, complexes, SNPs, mutations, variants, chemical compounds, ncRNAs, and miRNAs. Finally, CIDeR contains interacting bioentities, such as biological processes, pathways, and phenotypes. Each entry is also accompanied by additional metadata (where applicable) regarding the affected organism, tissue/cell line, and gender. CIDeR provides interconnectivity with the Entrez Gene, KEGG, OMIM, miRBase, GO, CORUM, Mammalian Phenotype Ontology (MPO) [229], and BRENDA Tissue Ontology (BTO) [230]. Interactions can be visualized as an interactive 2D network and downloaded in a CSV or SBML format.

MiRNA SNP Disease Database (MSDD) [231] is another database which comprises human disease-related biomolecular interactions. Similar to CIDeR, its data are derived from the literature and are manually curated. MSDD focuses on disease miRNA–SNP interactions, with accompanying metadata, such as the relevant gene and tissue, SNP position

relative to the associated miRNA, its allele, and the dysfunction pattern (increase/decrease). Specifically, MSDD provides 525 associations between 182 human miRNAs and 197 SNPs, regarding 153 genes and 164 human diseases. Information was mined in 2387 articles (last update: June 2017). The site allows the user to download MSDD data in text format, while also offering the choice to limit entries to a selected organ. Annotation information regarding miRNAs is derived from miRBase and SNPs from dbSNP.

Other databases provide direct links of biomolecules to diseases, without specifying direct inter-biomolecular interactions. DisGeNET [232] contains both curated and non-curated automatically mined information regarding disease-gene, disease-variant, and disease-disease associations. DisGeNET receives regular updates and its current version (v7.0) covers 1,134,942 gene-disease and 369,554 variant-disease associations, regarding 30,170 disease entries (UMLS [215]), 21,671 genes (NCBI), and 194,515 variants (dbSNP). Curated gene-disease associations are derived from UniProt, ClinGen [233], Genomics England PanelApp [234], PsyGeNET [235], Orphanet [236], the Human Phenotype Ontology (HPO) [237], and Comparative Toxicogenomics Database (CTD) [238], while curated variant-disease associations from UniProt, ClinVar [239], GWASdb, and the GWAS Catalog. DisGeNET data is downloadable in tab-delimited and SQLite database formats. All interaction data are also accessible programmatically through a REST API, an RDF API, the disgenet2r R package and the Cytoscape application. These programmatic endpoints enable downloading data in JSON, XML and TSV formats, as well as allowing disease ID mapping in UMLS, MeSH, OMIM, HPO, DO, Monarch Disease Ontology (MONDO) [240], NCI [241], and ICD-9 [242] databases.

EnDisease [243] is a manually curated database of enhancer-disease associations. The EnDisease database contains 535 total associations between 133 diseases and 454 enhancers, extracted from 199 published articles in 11 species. The data are downloadable in text format and represent the chromosomal position of the enhancer, the targeted gene and its UCSC identifier, and the related disease, with a respective entry link to the OMIM database. Additional metadata describe the cell type or mutation (where applicable), as well as the PubMed ID of the extracted association.

MNDR [133] is a frequently updated database that provides curated associations between ncRNAs and diseases along with a confidence score. MNDR data are derived from the literature, established databases as well as from predictive algorithms. Specifically, the current version (v3.1) includes 393,651 miRNA–disease, 295,834 lncRNA–disease, 300,630 circRNA–disease, 13,624 piRNA–disease, and 1573 snoRNA–disease associations, for a total of 1,005,312 associations regarding 1614 disease and 11 mammal species. The database entries are downloadable in text format. MNDR also provides an API to programmatically query associations, searching by ncRNA symbol or ID, disease name or DO/MeSH ID. As far as interconnectivity is concerned, MNDR entries contain an official gene symbol or miRBase ID, as well as DO and MeSH identifiers.

The Nervous System Disease NcRNAome Atlas (NSDNA) [244] is another ncRNA–disease association database that specializes in nervous system diseases. Its current version documents 26,128 associations between 144 nervous system diseases and 8736 ncRNAs, regarding 11 species, where information has been manually curated from 1410 articles. The data can be downloaded in text or spreadsheet format. Accompanying metadata describe the organism, tissue, expression pattern, detection method, target, and potential treatment of the association. Regarding database interoperability, NSDNA miRNA symbols were taken from miRBase, lncRNA from NONCODE and lncRNAdb, siRNA from siRecords [228], snoRNA from snoRNA–LBME-db, and piRNA from piRNABank [229]. The relative PubMed ID is also assigned to each ncRNA–disease association.

Several peptides and proteins have been found to possess an inherent tendency to misfold from their native functional state into intractable aggregates. These aggregates, known as "amyloid fibrils", have been associated with a diverse group of diseases known as "amyloidoses"; examples include the Alzheimer's and Parkinson's diseases, Type 2 diabetes, Creutzfeldt-Jakob Syndrome and many others. AmyCo (the Amyloidoses

Collection) [245] is a freely available collection of amyloidoses and other clinical disorders related to amyloid deposition. AmyCo classifies 75 diseases into 2 distinct categories, amyloidoses and other clinical conditions associated with amyloidoses. Each disease is associated with its precursor proteins (causative proteins), co-deposited proteins of amyloid deposits and affected tissues or organs. Database entries are also supplemented with detailed annotation and are linked to MeSH, OMIM, PubMed and UniProt databases.

Finally, there are databases linking bioentities, such as phenotypes, to diseases. The Human Phenotype Ontology (HPO) [237] provides human phenotype-disease associations, along with the implicated genes (where applicable) of each phenotype. HPO data are manually curated entries from the OMIM database. OMIM is a regularly updated, major gene-phenotype association database. The HPO is downloadable in OBO and OWL ontology formats. HPO also allows downloading text files with gene-phenotype and phenotype-gene associations, as found in the OMIM, Orphanet, and DECIPHER [246] databases. Gene entries are accompanied by Entrez Gene IDs. Since 2019, HPO provides a REST API to programmatically query HPO entries based on phenotype terms, diseases or genes.

Lastly, NeuroDNet [247] provides manually curated associations of diseases with genetic risk factors and with network models. These models are graphs containing parsed literature information regarding interactions of genes, proteins, and signaling pathways for a neurodegenerative disease. The database contains genetic risk factors regarding 12 neurodegenerative diseases and 16 total disease models for 8 diseases. Disease model networks are visualized through the Celldesigner [232] software and can be downloaded in SBML format. Disease entries are linked to the OMIM database, genes to NCBI, and proteins to the UniProt database, while association links are provided for each PubMed article reference. Information regarding the discussed disease-related databases is appended in Table 8.

Table 8. Disease-related databases.

Database Name	Interaction Type	Data Category	Curation Type	Organism	Data License [1]	Programmatic Access
CIDeR [228]	cellular component–protein complex–disease–drug–environment–gene–protein–mutant–mRNA–variant–ncRNA–miRNA–organism–phenotype-process–protein modification-SNP-tissue-cell line-CpG site	Primary	Manual	22 species	Free	N/A
MSDD [231]	miRNA–SNP	Primary	Manual	1 (*H. sapiens*)	Free	N/A
DisGeNET [232]	disease–gene, disease–variant, disease–disease	Primary	Manual, Automated	1 (*H. sapiens*)	Free (CC-BY-NC-SA)	REST API, RDF API (SPARQL endpoint), R (disgenet2r), Cytoscape
EnDisease [243]	enhancer–disease	Primary	Manual	11 species	Free	N/A
MNDR [133]	ncRNA–disease	Primary, Secondary, Predictive	Manual, Automated	11 species	Free	REST API
NSDNA [244]	ncRNA–disease	Primary	Manual	11 species	Free	N/A
AmyCo [245]	Protein–disease	Secondary	Manual	1 (*H. sapiens*)	Free	N/A
HPO [237]	phenotype–disease	Secondary	Manual	1 (*H. sapiens*)	Free	REST API
NeuroDNet [247]	genetic risk factor–disease, network model–disease	Primary	Manual	1 (*H. sapiens*)	Free (Open Source)	N/A

[1] The license type adopted by each database. In cases where a specific license type is used, it is given in parentheses. License abbreviations: CC-BY-NC-SA Creative Commons–Attribution–Non-Commercial–Share Alike.

Host–Pathogen Interactions

A discrete category of interactions that may lead to a disease concerns host–pathogen interactions. Here, we present bioenity interaction databases focusing on such host–pathogen interactions.

Viruses.STRING [248], an extension of STRING, is a database that contains intra-virus and virus–host PPIs. These annotated PPIs are either physical or functional. Interaction data are derived through text-mining, experimental data from BioGrid, Mint/IntAct [153], DIP, HPIDB [249], and VirusMentha [250], and orthologous relationships from eggNOG 4.5 [251]. As of 2021, Viruses.STRING covers 1,380,838,440 interactions between 2031 organisms and more than 9.5 million viral proteins. The site generates interactive networks of the queried interactions and all node entries are linked to Uniprot. Furthermore, the protein entries are also linked to Ensembl, KEGG, GeneCards [252], and neXtProt [253] databases. The data can be fully accessed and analyzed through a REST API and the Cytoscape STRING app. The generated interaction networks can be downloaded in SVG, TSV, XML, and MFA (multi-fasta) formats. All interaction data files are downloadable in text format and the whole database schema in SQL format.

ViRBase [254] is another viral-host interactions database that, apart from just proteins, mainly focuses on ncRNA interactions. More specifically, it includes manually curated associations between viral ncRNAs (especially lncRNAs and miRNAs) and host ncRNAs or proteins. The database (v2.1) currently consists of 781,476 ncRNA interactions between 93 viruses and 27 hosts, derived from 491 articles. microRNA entries were collected from miRBase, lncRNAs from lncRNAdb and the functional lncRNA database [118], snoRNAs from sno/scaRNAbase [255] and snoRNA–LBME-db [80], whereas ICTVdb (International Committee on Taxonomy of Viruses) [256] records provided virus names and abbreviations. Detailed views of the interaction entries consist of confidence scores, detection methods, tissue/cell line of origin and expression changes, where applicable. Furthermore, data can be queried through an API and are also downloadable in XLSX and text formats.

Another host–pathogen interactions database is TDR Targets [206], a repository on protein–chemical interactions involved in neglected disease pathogens, such as those implicated in tropical diseases like African trypanosomiasis (sleeping sickness) or dengue fever. In its latest version, TDR Targets incorporates experimentally determined and computationally predicted annotations on the chemical compounds and metabolites of pathogens associated with diseases and the drugs utilized in the treatment of these conditions, as well as on the sequence and structure features of the proteins targeted.

MVP (Microbe Versus Phage) [257] database focuses on interactions between phages and prokaryotes (bacteria/archaea). The database incorporates known viral sequences from NCBI, putative prophage regions in bacterial sequences from NCBI and EMBL, as well as viral and prophage sequences from ICTV published datasets, and metagenomic datasets from EBI. For the detection of putative prophage sequences in bacterial/archaeal genomes, the Phage_Finder tool was used [258]. All the viral sequences (50,782) were clustered based on their sequence similarity, resulting in 33,097 viral groups. Interactions and associations between the prophage sequences and microbes are based on 30,321 published sources, including projects such as Uncovering Earth's virome [259] and ICTV. All phage clusters and prokaryotes in MVP are provided with NCBI taxonomic IDs and all associations are downloadable in text format, whereas visualized networks can be downloaded in SVG and PNG formats.

Finally, HoPaCI-DB [260] further zooms in on two bacteria, *P. aeruginosa* and *C. burnetii*, and their host interactions. All listed interacting entries are manually curated and consist of either biomolecules such as proteins, nucleic acids or chemical compounds, or bioentities such as cellular processes, phenotypes or environmental factors. Its current version contains 4443 interactions, regarding 371 entries, mined from 290 articles. Database interactions are presented on site either in tabular format or as graph structures. Entries in HoPaCI-DB are mapped to Entrez Gene, KEGG, OMIM, miRBase, GO or CORUM identifiers, depending on their type, and all interactions are accompanied by a relative PubMed ID.

Additional metadata describe the type of interaction (e.g., localization, expression change, phosphorylation, etc.) as well as cell type, cell line, and tissue, where applicable. Database interactions are downloadable in CSV and SBML formats. Table 9 incorporates information regarding the aforementioned host–pathogen interaction databases.

Table 9. Host–pathogen interactions databases.

Database Name	Interaction Type	Data Category	Curation Type	Organisms	Data License	Programmatic Access
Viruses.STRING [248]	viral protein–viral protein, viral protein–host protein	Primary, Secondary, Predictive	Automated	2031 (1678 bacteria, 238 eukaryotes, 115 archaea)	Free for academic users (EMBL License)	REST API, Cytoscape app (stringApp) [41]
ViRBase [254]	host/viral ncRNA/protein–host/viral ncRNA/protein (ncRNA = mRNA, miRNA, pseudo-snoRNA, snRNA, lncRNA, rRNA, circRNA, shRNA)	Primary, Predictive	Manual	93 viruses, 27 host organisms	Free	REST API
TDR Targets [206]	protein–chemical	Primary	Manual	11 host organisms, 35 pathogens	Free	N/A
MVP [257]	bacteria–viral clusters, archaea-viral clusters	Secondary, Predictive	Automated	9122 bacteria, 123 archaea, 33,097 viral clusters (phages)	Free	N/A
HoPaCI-DB [260]	gene/protein-process-disease-organism-tissue/cell line-cellular component-phenotype-complex/PPI-drug-ncRNA/miRNA-organism-gene/protein mutant-environment-mRNA/protein variant	Primary	Manual	15 species	Free	N/A

3.6. Ecological Interactions

Finally, in a more macroscopic view, interactions can be captured between the different species and their relations (prey, pollinate, parasite, etc.). Data banks that include information about such ecosystem interactions aim to capture biodiversity, as well as key biotic and abiotic factors in environmental processes. The following databases include species interactions and trophic webs.

Global Biotic Interactions (GloBI) [261] is an open source database that contains interactions between living organisms and environmental factors. GloBI interaction data are retrieved both from web resources (data journals and APIs) and from directly contacting authors/data managers and are manually curated. The most recent data (May 2021) include 7,824,407 interaction records between approximately 240,000 species. These interactions comprise species' relationships, such as predator–prey, pollinator–plant, pathogen–host, parasite–host, and describe 33 different interaction types, such as "*eats*", "*kills*", "*interacts with*", "*parasite of*". The web interface represents interactions in the form of search widgets, interactive maps, hairballs, and bundle diagrams. The records that contain known taxa are cross-referenced with entries in NCBI, World Register of Marine Species (WoRMS) [262], Integrated Taxonomic Information System (ITIS) and Global Biodiversity Information Facility (GBIF), and the site entries are accompanied by links to Wikidata. Dataset collections with interactions are available in TSV, CSV, RDF formats as well as in sqlite, Darwin Core Archive [263], and Neo4j database formats. Data can also be accessed programmatically through a REST API, as well as through R (rglobi) and JavaScript (eol-globi-data-js) libraries or SPARQL and Cypher queries. GloBI is also integrated in the Encyclopedia of Life (EOL) [264] and Gulf of Mexico Species Interactions (GoMexSI) [265] projects.

The Web of Life [266] is a database similar to GloBI, which contains interactions between animals–plants, plants–plants, and host–plants, and visualizes ecological networks on the web in a coordinate-based system. A key difference with GloBI is that Web of Life only provides an "*interacts with*" type of association. At this moment, Web of Life contains

186 interaction networks, regarding 13,244 animal and plant species, which have been assembled by data from both published and unpublished projects. Other than the name of the species and the respective publications, there are no identifiers linking terms to other databases. All networks are provided as adjacency lists and are downloadable in CSV, XLS, JSON, and Pajek formats.

Another database that contains trophic interactions between ~7000 animals and plants in adjacency matrices, similarly to the Web of Life, is the Food Web (GlobalWeb) [267]. By representing the trophic interactions in a network, it is easier to detect the endangered and invasive species that might result from anthropogenic activities, such as fisheries. Currently, Food Web contains 358 food web graphs (adjacency matrix CSV format) that contain information manually mined from 123 reference papers. Again, no identifiers from other databases are provided.

Finally, a more specialized ecological database, focusing on interactions between bats and plants or other organisms, is Bat Eco-Interactions [268]. It currently (May 2021) contains 13,383 interactions that occur between 479 bat species and 2135 other organisms, mined from 622 peer-reviewed articles. Interaction data are available in CSV format after registration. The database receives regular updates with bat–parasite and ba–mammal interactions, which include taxonomic and location metadata. Table 10 summarizes the interaction data of the four aforementioned databases.

Table 10. Ecological interactions databases.

Database Name	Interaction Type	Data Category	Curation Type	Organisms	Data License	Programmatic Access
GloBI [261]	predator–prey, pollinator–plant, pathogen–host, parasite–host	Secondary	Manual, Automated	644,512 known taxa, 77,245 unknown taxa	Free (Open Source)	REST API, R (rglobi), JavaScript (eol-globi-data-js), SPARQL and Cypher endpoints
The Web of Life [266]	host–parasite, plant–herbivore, food webs, anemone–fish, plant–ant, pollination, seed dispersal	Primary	Manual	13,244 animals/plants	Free	N/A
Food web [267]	fish–animal type–detritus–plankton	Primary	Manual	~7000 animals/plants	Free	N/A
Bat Eco-Interactions [268]	cohabitation–consumption–host–pollination–prey–predation–roost–seed dispersal–transport–visitation bat/living organisms	Primary	Manual	484 bat species, 2146 other species	Free	N/A

4. Data Visualization

Network visualization plays a key role in understanding, communicating, exploring and identifying patterns (e.g., important edges, highly connected nodes or communities) in an interactome. For this purpose, several interactive applications have been implemented and a plethora of review papers have been written [269–273]. Briefly, Cytoscape [39,40], Cytoscape.js [157], Gephi [274], Pajek [275], Ondex [276], Proviz [277], VisANT [278], Osprey [279], Arena3D [280,281], Arena3D^web [282], Graphia (Kajeka) [283], NORMA [284], and BioLayout Express 3D [285] are state-of-the-art tools worth mentioning. Similarly, pathway-specific applications include the Pathview [286], BioTapestry [287], PathVisio [214], Interactive Pathways Explorer (iPath) [288], MapMan [289], KEGG [162], Reactome [163], WikiPathways [212] and Pathway Commons [290].

In addition to the aforementioned visualizers, several complementary tools have been implemented for network topological analysis and cluster set comparisons. Typical examples are the Network Analyzer [291], ZoomOut [292], Network Analysis Toolkit (NEAT) [293], NAP [294,295], VICTOR [296] and the Stanford Network Analysis Project (SNAP) [297]. Back-end libraries for network storage and analysis which are worth mentioning are igraph [298], NetworkX [299], GraphViz and Graph-tool [300]. Finally, widely used

force-directed layout algorithms [301] for efficient graph drawing include the Fruchterman-Reingold [302], Yifan-Hu [303], Large Graph Layout (LGL) [304] and Kamada-Kawai [305] algorithms.

5. Functional Enrichment

In order to annotate clusters within a network and gain insights into the biology of the bioentities which tend to form distinct communities, a functional enrichment analysis is necessary. Briefly, functional enrichment analysis is an approach to identify classes of bioentities (e.g., Gene Ontologies, pathways etc.) in which genes or proteins were found to be over-represented. Among several applications which have been proposed [306,307], tools which are worthy mentioning include: g:Profiler [308], Panther [309], DAVID [310], WebGestalt [311], EnrichR [312], AmiGO [313], GeneSCF [314], AllEnricher [315], aGO-tool [316], ClueGo [317], GSEA [318], GOrilla [319], Flame [320], clusterProfiler [321] and NASQAR [322].

6. Conclusions

While great efforts have been made in the fields of network biology and biomedical data integration, and despite the numerous databases and repositories for organizing data in a more structured way, a number of challenges remain to be addressed. Scalability is one of the major future challenges. The overgrowth trend to biomedical data has been clear at least since 2015 [323], when it was reported that Twitter was producing 1–17 petabytes of information per year, Astronomy with 1000 petabytes per year, YouTube with 1000–2000 petabytes per year, and Genomics with 2000–40,000 petabytes per year. Due to these orders of magnitude in data accumulation, biomedical repositories need to adjust to the new big-data era and adopt new technologies which can cope with today's complexity and exponential information growth.

Efficient indexing, compression algorithms for massive volumes of information, and usage of cloud computing and distributed systems would constitute significant enhancements. In addition, despite the plethora of biomedical databases, users (especially less experienced ones) still prefer non-biomedical search engines, such as Google, Bing or Yandex to query biomedical terms. This can mainly be attributed to the poor integration between databases and their inefficient search engines, which often do not allow for any user-friendly flexibility. Some progress has been made in this area with the development of systems that integrate multiple resources in a common framework. Perhaps the most characteristic example is the IMEx Consortium [158], which integrates information from multiple interaction databases (IntAct, MINT, DIP etc.) with additional annotation from other sources (e.g., Mechanobiology [324]), and provides a common API (IMEx PSICQUIC) to retrieve and combine information from all its participants. Another example is the Network Data Exchange (NDEx) [29], which also integrates multiple sources in a unified format and access point. However, these systems support only a limited number of the currently available biomolecule and bioentity interaction databases; instead, the vast majority of the databases presented in this review are isolated systems, with poor APIs, documentation, and data accessibility.

In terms of designs, many of the currently available databases often come with unfriendly or complicated GUIs, thus being unattractive, overwhelming and difficult-to-use. Another important issue is the limited cross-talk between the various repositories (web services) along with the lack of ID conversion tools, which rarely cover a broad enough spectrum of common database identifiers. Among the issues that still remain to be addressed are symbol disambiguation, redundant information across repositories, better literature mining tools (e.g., OnTheFly [325] or INDRA [227]), richer metadata, more accurate name entity recognition techniques to link free text with database records [326,327], utilization of semantics, interoperability, and more frequent/automated updating and maintenance. These tasks will undoubtedly keep bioinformaticians busy in the next few years and their

successful tackling promises to offer scientists from all ranks and expertise the necessary tools to successfully navigate the ever-increasing complexity of biological data.

References

1. Lightbody, G.; Haberland, V.; Browne, F.; Taggart, L.; Zheng, H.; Parkes, E.; Blayney, J.K. Review of applications of high-throughput sequencing in personalized medicine: Barriers and facilitators of future progress in research and clinical application. *Brief. Bioinform.* **2019**, *20*, 1795–1811. [CrossRef]
2. Sonawane, A.R.; Weiss, S.T.; Glass, K.; Sharma, A. Network Medicine in the Age of Biomedical Big Data. *Front. Genet.* **2019**, *10*, 294. [CrossRef] [PubMed]
3. Pavlopoulos, G.A.; Iacucci, E.; Iliopoulos, I.; Bagos, P. Interpreting the Omics 'era' Data. In *Multimedia Services in Intelligent Environments*; Tsihrintzis, G.A., Virvou, M., Jain, L.C., Eds.; Springer International Publishing: Heidelberg, Germany, 2013; Volume 25, pp. 79–100, ISBN 978-3-319-00374-0.
4. Pavlopoulos, G.A.; Soldatos, T.G.; Barbosa-Silva, A.; Schneider, R. A reference guide for tree analysis and visualization. *BioData Min.* **2010**, *3*, 1. [CrossRef]
5. Luck, K.; Kim, D.-K.; Lambourne, L.; Spirohn, K.; Begg, B.E.; Bian, W.; Brignall, R.; Cafarelli, T.; Campos-Laborie, F.J.; Charloteaux, B.; et al. A reference map of the human binary protein interactome. *Nature* **2020**, *580*, 402–408. [CrossRef] [PubMed]
6. Rolland, T.; Taşan, M.; Charloteaux, B.; Pevzner, S.J.; Zhong, Q.; Sahni, N.; Yi, S.; Lemmens, I.; Fontanillo, C.; Mosca, R.; et al. A Proteome-Scale Map of the Human Interactome Network. *Cell* **2014**, *159*, 1212–1226. [CrossRef] [PubMed]
7. Kim, E.; Hwang, S.; Kim, H.; Shim, H.; Kang, B.; Yang, S.; Shim, J.H.; Shin, S.Y.; Marcotte, E.M.; Lee, I. MouseNet v2: A database of gene networks for studying the laboratory mouse and eight other model vertebrates. *Nucleic Acids Res.* **2016**, *44*, D848–D854. [CrossRef]
8. Alanis-Lobato, G.; Möllmann, J.S.; Schaefer, M.H.; Andrade-Navarro, M.A. MIPPIE: The mouse integrated protein–protein interaction reference. *Database* **2020**, *2020*, baaa035. [CrossRef]
9. Schwikowski, B.; Uetz, P.; Fields, S. A network of protein–protein interactions in yeast. *Nat. Biotechnol.* **2000**, *18*, 1257–1261. [CrossRef] [PubMed]
10. Tong, A.H.Y. Global Mapping of the Yeast Genetic Interaction Network. *Science* **2004**, *303*, 808–813. [CrossRef]
11. Guruharsha, K.G.; Rual, J.-F.; Zhai, B.; Mintseris, J.; Vaidya, P.; Vaidya, N.; Beekman, C.; Wong, C.; Rhee, D.Y.; Cenaj, O.; et al. A Protein Complex Network of Drosophila melanogaster. *Cell* **2011**, *147*, 690–703. [CrossRef]
12. Li, Z.; Ivanov, A.A.; Su, R.; Gonzalez-Pecchi, V.; Qi, Q.; Liu, S.; Webber, P.; McMillan, E.; Rusnak, L.; Pham, C.; et al. The OncoPPi network of cancer-focused protein–protein interactions to inform biological insights and therapeutic strategies. *Nat. Commun.* **2017**, *8*, 14356. [CrossRef]
13. Ivanov, S.; Lagunin, A.; Filimonov, D.; Tarasova, O. Network-Based Analysis of OMICs Data to Understand the HIV–Host Interaction. *Front. Microbiol.* **2020**, *11*, 1314. [CrossRef]
14. Karbalaei, R.; Allahyari, M.; Rezaei-Tavirani, M.; Asadzadeh-Aghdaei, H.; Zali, M.R. Protein-protein interaction analysis of Alzheimer's disease and NAFLD based on systems biology methods unhide common ancestor pathways. *Gastroenterol. Hepatol. Bed Bench* **2018**, *11*, 27–33.

15. Apostolakou, A.E.; Sula, X.K.; Nastou, K.C.; Nasi, G.I.; Iconomidou, V.A. Exploring the conservation of Alzheimer-related pathways between H. sapiens and C. elegans: A network alignment approach. *Sci. Rep.* **2021**, *11*, 4572. [CrossRef] [PubMed]

16. Gordon, D.E.; Jang, G.M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K.M.; O'Meara, M.J.; Rezelj, V.V.; Guo, J.Z.; Swaney, D.L.; et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**, *583*, 459–468. [CrossRef]

17. Lindberg, D.A. Internet access to the National Library of Medicine. *Eff. Clin. Pract.* **2000**, *3*, 256–260. [PubMed]

18. The UniProt Consortium UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [CrossRef] [PubMed]

19. Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2016**, *44*, D67–D72. [CrossRef]

20. Howe, K.L.; Achuthan, P.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; Bhai, J.; et al. Ensembl 2021. *Nucleic Acids Res.* **2021**, *49*, D884–D891. [CrossRef]

21. Miryala, S.K.; Anbarasu, A.; Ramaiah, S. Discerning molecular interactions: A comprehensive review on biomolecular interaction databases and network analysis tools. *Gene* **2018**, *642*, 84–94. [CrossRef]

22. Bajpai, A.K.; Davuluri, S.; Tiwary, K.; Narayanan, S.; Oguru, S.; Basavaraju, K.; Dayalan, D.; Thirumurugan, K.; Acharya, K.K. Systematic comparison of the protein-protein interaction databases from a user's perspective. *J. Biomed. Inform.* **2020**, *103*, 103380. [CrossRef] [PubMed]

23. Demir, E.; Cary, M.P.; Paley, S.; Fukuda, K.; Lemer, C.; Vastrik, I.; Wu, G.; D'Eustachio, P.; Schaefer, C.; Luciano, J.; et al. The BioPAX community standard for pathway data sharing. *Nat. Biotechnol.* **2010**, *28*, 935–942. [CrossRef] [PubMed]

24. Hucka, M.; Finney, A.; Sauro, H.M.; Bolouri, H.; Doyle, J.C.; Kitano, H.; Arkin, A.P.; Bornstein, B.J.; Bray, D.; Cornish-Bowden, A.; et al. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics* **2003**, *19*, 524–531. [CrossRef] [PubMed]

25. Hermjakob, H.; Montecchi-Palazzi, L.; Bader, G.; Wojcik, J.; Salwinski, L.; Ceol, A.; Moore, S.; Orchard, S.; Sarkans, U.; von Mering, C.; et al. The HUPO PSI's Molecular Interaction format—a community standard for the representation of protein interaction data. *Nat. Biotechnol.* **2004**, *22*, 177–183. [CrossRef] [PubMed]

26. Murray-Rust, P.; Rzepa, H.S.; Wright, M. Development of chemical markup language (CML) as a system for handling complex chemical content. *New J. Chem.* **2001**, *25*, 618–634. [CrossRef]

27. Lloyd, C.M.; Halstead, M.D.B.; Nielsen, P.F. CellML: Its future, present and past. *Prog. Biophys. Mol. Biol.* **2004**, *85*, 433–450. [CrossRef]

28. Tamassia, R. (Ed.) *Handbook of Graph Drawing and Visualization*; Discrete mathematics and its applications; First issued in paperback; CRC Press: Boca Raton, FL, USA, 2016; ISBN 978-1-138-03424-2.

29. Pratt, D.; Chen, J.; Pillich, R.; Rynkov, V.; Gary, A.; Demchak, B.; Ideker, T. NDEx 2.0: A Clearinghouse for Research on Cancer Pathways. *Cancer Res.* **2017**, *77*, e58–e61. [CrossRef]

30. Koh, G.C.K.W.; Porras, P.; Aranda, B.; Hermjakob, H.; Orchard, S.E. Analyzing Protein–Protein Interaction Networks. *J. Proteome Res.* **2012**, *11*, 2014–2031. [CrossRef]

31. De Las Rivas, J.; Fontanillo, C. Protein-protein interactions essentials: Key concepts to building and analyzing interactome networks. *PLoS Comput. Biol.* **2010**, *6*, e1000807. [CrossRef]

32. Lotia, S.; Montojo, J.; Dong, Y.; Bader, G.D.; Pico, A.R. Cytoscape app store. *Bioinformatics* **2013**, *29*, 1350–1351. [CrossRef]

33. van Dam, S.; Võsa, U.; van der Graaf, A.; Franke, L.; de Magalhães, J.P. Gene co-expression analysis for functional classification and gene–disease predictions. *Brief. Bioinform.* **2018**, *19*, 575–592. [CrossRef]

34. Obayashi, T.; Kagaya, Y.; Aoki, Y.; Tadaka, S.; Kinoshita, K. COXPRESdb v7: A gene coexpression database for 11 animal species supported by 23 coexpression platforms for technical evaluation and evolutionary inference. *Nucleic Acids Res.* **2019**, *47*, D55–D62. [CrossRef] [PubMed]

35. 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. NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **2013**, *41*, D991–D995. [CrossRef]

36. Pembroke, W.G.; Hartl, C.L.; Geschwind, D.H. Evolutionary conservation and divergence of the human brain transcriptome. *Genome Biol.* **2021**, *22*, 52. [CrossRef]

37. Szklarczyk, D.; Gable, A.L.; Nastou, K.C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N.T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **2021**, *49*, D605–D612. [CrossRef]

38. Kustatscher, G.; Grabowski, P.; Schrader, T.A.; Passmore, J.B.; Schrader, M.; Rappsilber, J. Co-regulation map of the human proteome enables identification of protein functions. *Nat. Biotechnol.* **2019**, *37*, 1361–1371. [CrossRef]

39. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]

40. Shannon, P.T.; Grimes, M.; Kutlu, B.; Bot, J.J.; Galas, D.J. RCytoscape: Tools for exploratory network analysis. *BMC Bioinform.* **2013**, *14*, 217. [CrossRef]

41. Doncheva, N.T.; Morris, J.; Gorodkin, J.; Jensen, L.J. Cytoscape stringApp: Network analysis and visualization of proteomics data. *J. Proteome Res.* **2018**, *18*, 623–632. [CrossRef] [PubMed]

42. Franz, M.; Rodriguez, H.; Lopes, C.; Zuberi, K.; Montojo, J.; Bader, G.D.; Morris, Q. GeneMANIA update 2018. *Nucleic Acids Res.* **2018**, *46*, W60–W64. [CrossRef]
43. Oughtred, R.; Rust, J.; Chang, C.; Breitkreutz, B.-J.; Stark, C.; Willems, A.; Boucher, L.; Leung, G.; Kolas, N.; Zhang, F.; et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* **2021**, *30*, 187–200. [CrossRef]
44. Razick, S.; Magklaras, G.; Donaldson, I.M. iRefIndex: A consolidated protein interaction database with provenance. *BMC Bioinform.* **2008**, *9*, 405. [CrossRef] [PubMed]
45. Brown, K.R.; Jurisica, I. Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol.* **2007**, *8*, R95. [CrossRef] [PubMed]
46. Raina, P.; Lopes, I.; Chatsirisupachai, K.; Farooq, Z.; de Magalhães, J.P. GeneFriends 2021: Updated co-expression databases and tools for human and mouse genes and transcripts. *bioRxiv* **2021**. [CrossRef]
47. Vandenbon, A.; Dinh, V.H.; Mikami, N.; Kitagawa, Y.; Teraguchi, S.; Ohkura, N.; Sakaguchi, S. Immuno-Navigator, a batch-corrected coexpression database, reveals cell type-specific gene networks in the immune system. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E2393–E2402. [CrossRef] [PubMed]
48. Yang, S.; Kim, C.Y.; Hwang, S.; Kim, E.; Kim, H.; Shim, H.; Lee, I. COEXPEDIA: Exploring biomedical hypotheses via co-expressions associated with medical subject headings (MeSH). *Nucleic Acids Res.* **2017**, *45*, D389–D396. [CrossRef]
49. 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. Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* **2015**, *47*, 569–576. [CrossRef] [PubMed]
50. Hwang, S.; Kim, C.Y.; Yang, S.; Kim, E.; Hart, T.; Marcotte, E.M.; Lee, I. HumanNet v2: Human gene networks for disease research. *Nucleic Acids Res.* **2019**, *47*, D573–D580. [CrossRef]
51. Jiao, C.; Yan, P.; Xia, C.; Shen, Z.; Tan, Z.; Tan, Y.; Wang, K.; Jiang, Y.; Huang, L.; Dai, R.; et al. BrainEXP: A database featuring with spatiotemporal expression variations and co-expression organizations in human brains. *Bioinformatics* **2019**, *35*, 172–174. [CrossRef]
52. The ENCODE Project Consortium An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [CrossRef]
53. The Cancer Genome Atlas Research Network Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **2008**, *455*, 1061–1068. [CrossRef] [PubMed]
54. Obayashi, T.; Aoki, Y.; Tadaka, S.; Kagaya, Y.; Kinoshita, K. ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. *Plant Cell Physiol.* **2018**, *59*, e3. [CrossRef] [PubMed]
55. Ogata, Y.; Suzuki, H.; Sakurai, N.; Shibata, D. CoP: A database for characterizing co-expressed gene modules with biological information in plants. *Bioinformatics* **2010**, *26*, 1267–1268. [CrossRef]
56. van Dijk, A.D.J. (Ed.) *Plant Genomics Databases: Methods and Protocols*; Methods in molecular biology; Humana Press: New York, NY, USA, 2017; ISBN 978-1-4939-6656-1.
57. Yim, W.; Yu, Y.; Song, K.; Jang, C.; Lee, B.-M. PLANEX: The plant co-expression database. *BMC Plant Biol.* **2013**, *13*, 83. [CrossRef]
58. Manfield, I.W.; Jen, C.-H.; Pinney, J.W.; Michalopoulos, I.; Bradford, J.R.; Gilmartin, P.M.; Westhead, D.R. Arabidopsis Co-expression Tool (ACT): Web server tools for microarray-based gene expression analysis. *Nucleic Acids Res.* **2006**, *34*, W504–W509. [CrossRef]
59. Lee, I.; Ambaru, B.; Thakkar, P.; Marcotte, E.M.; Rhee, S.Y. Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana. *Nat. Biotechnol.* **2010**, *28*, 149–156. [CrossRef] [PubMed]
60. Dalma-Weiszhausz, D.D.; Warrington, J.; Tanimoto, E.Y.; Miyada, C.G. The Affymetrix GeneChip®Platform: An Overview. In *Methods in Enzymology*; Elsevier: Amsterdam, The Netherlands, 2006; Volume 410, pp. 3–28, ISBN 978-0-12-182815-8.
61. Aoki, Y.; Okamura, Y.; Ohta, H.; Kinoshita, K.; Obayashi, T. ALCOdb: Gene Coexpression Database for Microalgae. *Plant Cell Physiol.* **2016**, *57*, e3. [CrossRef]
62. Zheng, H.-Q.; Chiang-Hsieh, Y.-F.; Chien, C.-H.; Hsu, B.-K.; Liu, T.-L.; Chen, C.-N.; Chang, W.-C. AlgaePath: Comprehensive analysis of metabolic pathways using transcript abundance data from next-generation sequencing in green algae. *BMC Genom.* **2014**, *15*, 196. [CrossRef]
63. Shim, H.; Kim, J.H.; Kim, C.Y.; Hwang, S.; Kim, H.; Yang, S.; Lee, J.E.; Lee, I. Function-driven discovery of disease genes in zebrafish using an integrated genomics big data resource. *Nucleic Acids Res.* **2016**, *44*, 9611–9623. [CrossRef]
64. Montojo, J.; Zuberi, K.; Rodriguez, H.; Kazi, F.; Wright, G.; Donaldson, S.L.; Morris, Q.; Bader, G.D. GeneMANIA Cytoscape plugin: Fast gene function predictions on the desktop. *Bioinformatics* **2010**, *26*, 2927–2928. [CrossRef]
65. Michalopoulos, I.; Pavlopoulos, G.A.; Malatras, A.; Karelas, A.; Kostadima, M.-A.; Schneider, R.; Kossida, S. Human gene correlation analysis (HGCA): A tool for the identification of transcriptionally co-expressed genes. *BMC Res. Notes* **2012**, *5*, 265. [CrossRef]
66. Chojnowski, G.; Waleń, T.; Bujnicki, J.M. RNA Bricks—A database of RNA 3D motifs and their interactions. *Nucleic Acids Res.* **2014**, *42*, D123–D131. [CrossRef]
67. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef] [PubMed]

68. Burge, S.W.; Daub, J.; Eberhardt, R.; Tate, J.; Barquist, L.; Nawrocki, E.P.; Eddy, S.R.; Gardner, P.P.; Bateman, A. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **2013**, *41*, D226–D232. [CrossRef]

69. Waleń, T.; Chojnowski, G.; Gierski, P.; Bujnicki, J.M. ClaRNA: A classifier of contacts in RNA 3D structures based on a comparative analysis of various classification schemes. *Nucleic Acids Res.* **2014**, *42*, e151. [CrossRef]

70. Teng, X.; Chen, X.; Xue, H.; Tang, Y.; Zhang, P.; Kang, Q.; Hao, Y.; Chen, R.; Zhao, Y.; He, S. NPInter v4.0: An integrated database of ncRNA interactions. *Nucleic Acids Res.* **2019**, *48*, D160–D165. [CrossRef] [PubMed]

71. Gong, J.; Shao, D.; Xu, K.; Lu, Z.; Lu, Z.J.; Yang, Y.T.; Zhang, Q.C. RISE: A database of RNA interactome from sequencing experiments. *Nucleic Acids Res.* **2018**, *46*, D194–D201. [CrossRef] [PubMed]

72. Fang, S.; Zhang, L.; Guo, J.; Niu, Y.; Wu, Y.; Li, H.; Zhao, L.; Li, X.; Teng, X.; Sun, X.; et al. NONCODEV5: A comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* **2018**, *46*, D308–D314. [CrossRef] [PubMed]

73. Kozomara, A.; Griffiths-Jones, S. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **2014**, *42*, D68–D73. [CrossRef]

74. Glažar, P.; Papavasileiou, P.; Rajewsky, N. circBase: A database for circular RNAs. *RNA* **2014**, *20*, 1666–1670. [CrossRef] [PubMed]

75. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [CrossRef]

76. Bouchard-Bourelle, P.; Desjardins-Henri, C.; Mathurin-St-Pierre, D.; Deschamps-Francoeur, G.; Fafard-Couture, É.; Garant, J.-M.; Elela, S.A.; Scott, M.S. snoDB: An interactive database of human snoRNA sequences, abundance and interactions. *Nucleic Acids Res.* **2020**, *48*, D220–D225. [CrossRef]

77. Haeussler, M.; Zweig, A.S.; Tyner, C.; Speir, M.L.; Rosenbloom, K.R.; Raney, B.J.; Lee, C.M.; Lee, B.T.; Hinrichs, A.S.; Gonzalez, J.N.; et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **2019**, *47*, D853–D858. [CrossRef]

78. Braschi, B.; Denny, P.; Gray, K.; Jones, T.; Seal, R.; Tweedie, S.; Yates, B.; Bruford, E. Genenames.org: The HGNC and VGNC resources in 2019. *Nucleic Acids Res.* **2019**, *47*, D786–D792. [CrossRef] [PubMed]

79. The RNAcentral Consortium; Sweeney, B.A.; Petrov, A.I.; Burkov, B.; Finn, R.D.; Bateman, A.; Szymanski, M.; Karlowski, W.M.; Gorodkin, J.; Seemann, S.E.; et al. RNAcentral: A hub of information for non-coding RNA sequences. *Nucleic Acids Res.* **2019**, *47*, D221–D229. [CrossRef]

80. Lestrade, L. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* **2006**, *34*, D158–D162. [CrossRef] [PubMed]

81. Yoshihama, M.; Nakao, A.; Kenmochi, N. snOPY: A small nucleolar RNA orthological gene database. *BMC Res. Notes* **2013**, *6*, 426. [CrossRef] [PubMed]

82. Jorjani, H.; Kehr, S.; Jedlinski, D.J.; Gumienny, R.; Hertel, J.; Stadler, P.F.; Zavolan, M.; Gruber, A.R. An updated human snoRNAome. *Nucleic Acids Res.* **2016**, *44*, 5068–5082. [CrossRef]

83. Yi, X.; Zhang, Z.; Ling, Y.; Xu, W.; Su, Z. PNRD: A plant non-coding RNA database. *Nucleic Acids Res.* **2015**, *43*, D982–D989. [CrossRef]

84. Zhang, Z.; Yu, J.; Li, D.; Zhang, Z.; Liu, F.; Zhou, X.; Wang, T.; Ling, Y.; Su, Z. PMRD: Plant microRNA database. *Nucleic Acids Res.* **2010**, *38*, D806–D813. [CrossRef]

85. Dai, X.; Zhuang, Z.; Zhao, P.X. psRNATarget: A plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **2018**, *46*, W49–W54. [CrossRef]

86. Zhang, C.; Li, G.; Zhu, S.; Zhang, S.; Fang, J. tasiRNAdb: A database of ta-siRNA regulatory pathways. *Bioinformatics* **2014**, *30*, 1045–1046. [CrossRef]

87. Chan, P.P.; Lowe, T.M. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **2009**, *37*, D93–D97. [CrossRef] [PubMed]

88. Lamesch, P.; Berardini, T.Z.; Li, D.; Swarbreck, D.; Wilks, C.; Sasidharan, R.; Muller, R.; Dreher, K.; Alexander, D.L.; Garcia-Hernandez, M.; et al. The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* **2012**, *40*, D1202–D1210. [CrossRef]

89. Kawahara, Y.; de la Bastide, M.; Hamilton, J.P.; Kanamori, H.; McCombie, W.R.; Ouyang, S.; Schwartz, D.C.; Tanaka, T.; Wu, J.; Zhou, S.; et al. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice* **2013**, *6*, 4. [CrossRef] [PubMed]

90. Karagkouni, D.; Paraskevopoulou, M.D.; Chatzopoulos, S.; Vlachos, I.S.; Tastsoglou, S.; Kanellos, I.; Papadimitriou, D.; Kavakiotis, I.; Maniou, S.; Skoufos, G.; et al. DIANA-TarBase v8: A decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Res.* **2018**, *46*, D239–D245. [CrossRef]

91. Kodama, Y.; Mashima, J.; Kosuge, T.; Kaminuma, E.; Ogasawara, O.; Okubo, K.; Nakamura, Y.; Takagi, T. DNA Data Bank of Japan: 30th anniversary. *Nucleic Acids Res.* **2018**, *46*, D30–D35. [CrossRef] [PubMed]

92. Chou, C.-H.; Chang, N.-W.; Shrestha, S.; Hsu, S.-D.; Lin, Y.-L.; Lee, W.-H.; Yang, C.-D.; Hong, H.-C.; Wei, T.-Y.; Tu, S.-J.; et al. miRTarBase 2016: Updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.* **2016**, *44*, D239–D247. [CrossRef]

93. Xiao, F.; Zuo, Z.; Cai, G.; Kang, S.; Gao, X.; Li, T. miRecords: An integrated resource for microRNA-target interactions. *Nucleic Acids Res.* **2009**, *37*, D105–D110. [CrossRef]

94. Paraskevopoulou, M.D.; Georgakilas, G.; Kostoulas, N.; Vlachos, I.S.; Vergoulis, T.; Reczko, M.; Filippidis, C.; Dalamagas, T.; Hatzigeorgiou, A.G. DIANA-microT web server v5.0: Service integration into miRNA functional analysis workflows. *Nucleic Acids Res.* **2013**, *41*, W169–W173. [CrossRef]

95. Karagkouni, D.; Paraskevopoulou, M.D.; Tastsoglou, S.; Skoufos, G.; Karavangeli, A.; Pierros, V.; Zacharopoulou, E.; Hatzigeorgiou, A.G. DIANA-LncBase v3: Indexing experimentally supported miRNA targets on non-coding transcripts. *Nucleic Acids Res.* **2019**, *38*, D101–D110. [CrossRef]

96. Vlachos, I.S.; Zagganas, K.; Paraskevopoulou, M.D.; Georgakilas, G.; Karagkouni, D.; Vergoulis, T.; Dalamagas, T.; Hatzigeorgiou, A.G. DIANA-miRPath v3.0: Deciphering microRNA function with experimental support. *Nucleic Acids Res.* **2015**, *43*, W460–W466. [CrossRef]

97. Ramanathan, M.; Porter, D.F.; Khavari, P.A. Methods to study RNA–protein interactions. *Nat. Methods* **2019**, *16*, 225–234. [CrossRef] [PubMed]

98. Yi, Y.; Zhao, Y.; Huang, Y.; Wang, D. A Brief Review of RNA-Protein Interaction Database Resources. *Non-Coding RNA* **2017**, *3*, 6. [CrossRef] [PubMed]

99. Fujimori, S.; Hino, K.; Saito, A.; Miyano, S.; Miyamoto-Sato, E. PRD: A protein–RNA interaction database. *Bioinformation* **2012**, *8*, 729–730. [CrossRef] [PubMed]

100. Gene Ontology Consortium. The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res.* **2021**, *49*, D325–D334. [CrossRef]

101. Lin, Y.; Liu, T.; Cui, T.; Wang, Z.; Zhang, Y.; Tan, P.; Huang, Y.; Yu, J.; Wang, D. RNAInter in 2020: RNA interactome repository with increased coverage and annotation. *Nucleic Acids Res.* **2020**, *48*, D189–D197. [CrossRef]

102. Zhang, Y.; Liu, T.; Chen, L.; Yang, J.; Yin, J.; Zhang, Y.; Yun, Z.; Xu, H.; Ning, L.; Guo, F.; et al. RIscoper: A tool for RNA–RNA interaction extraction from the literature. *Bioinformatics* **2019**, *35*, 3199–3202. [CrossRef]

103. Mann, M.; Wright, P.R.; Backofen, R. IntaRNA 2.0: Enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res.* **2017**, *45*, W435–W439. [CrossRef]

104. Tuvshinjargal, N.; Lee, W.; Park, B.; Han, K. PRIdictor: Protein–RNA Interaction predictor. *Biosystems* **2016**, *139*, 17–22. [CrossRef]

105. Alipanahi, B.; Delong, A.; Weirauch, M.T.; Frey, B.J. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* **2015**, *33*, 831–838. [CrossRef]

106. Amberger, J.S.; Bocchini, C.A.; Schiettecatte, F.; Scott, A.F.; Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **2015**, *43*, D789–D798. [CrossRef]

107. Amberger, J.S.; Hamosh, A. Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes. *Curr. Protoc. Bioinforma.* **2017**, *58*. [CrossRef]

108. Keshava Prasad, T.S.; Goel, R.; Kandasamy, K.; Keerthikumar, S.; Kumar, S.; Mathivanan, S.; Telikicherla, D.; Raju, R.; Shafreen, B.; Venugopal, A.; et al. Human Protein Reference Database—2009 update. *Nucleic Acids Res.* **2009**, *37*, D767–D772. [CrossRef]

109. Zhu, Y.; Xu, G.; Yang, Y.T.; Xu, Z.; Chen, X.; Shi, B.; Xie, D.; Lu, Z.J.; Wang, P. POSTAR2: Deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res.* **2019**, *47*, D203–D211. [CrossRef] [PubMed]

110. Blin, K.; Dieterich, C.; Wurmus, R.; Rajewsky, N.; Landthaler, M.; Akalin, A. DoRiNA 2.0—Upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.* **2015**, *43*, D160–D167. [CrossRef]

111. Lewis, B.A.; Walia, R.R.; Terribilini, M.; Ferguson, J.; Zheng, C.; Honavar, V.; Dobbs, D. PRIDB: A protein-RNA interface database. *Nucleic Acids Res.* **2011**, *39*, D277–D282. [CrossRef] [PubMed]

112. Cook, K.B.; Kazan, H.; Zuberi, K.; Morris, Q.; Hughes, T.R. RBPDB: A database of RNA-binding specificities. *Nucleic Acids Res.* **2011**, *39*, D301–D308. [CrossRef]

113. Shulman-Peleg, A.; Nussinov, R.; Wolfson, H.J. RsiteDB: A database of protein binding pockets that interact with RNA nucleotide bases. *Nucleic Acids Res.* **2009**, *37*, D369–D373. [CrossRef]

114. Cheng, L.; Wang, P.; Tian, R.; Wang, S.; Guo, Q.; Luo, M.; Zhou, W.; Liu, G.; Jiang, H.; Jiang, Q. LncRNA2Target v2.0: A comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.* **2019**, *47*, D140–D144. [CrossRef] [PubMed]

115. Harrow, J.; Frankish, A.; Gonzalez, J.M.; Tapanari, E.; Diekhans, M.; Kokocinski, F.; Aken, B.L.; Barrell, D.; Zadissa, A.; Searle, S.; et al. GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.* **2012**, *22*, 1760–1774. [CrossRef] [PubMed]

116. Maglott, D.; Ostell, J.; Pruitt, K.D.; Tatusova, T. Entrez Gene: Gene-centered information at NCBI. *Nucleic Acids Res.* **2011**, *39*, D52–D57. [CrossRef] [PubMed]

117. Zhou, B.; Zhao, H.; Yu, J.; Guo, C.; Dou, X.; Song, F.; Hu, G.; Cao, Z.; Qu, Y.; Yang, Y.; et al. EVLncRNAs: A manually curated database for long non-coding RNAs validated by low-throughput experiments. *Nucleic Acids Res.* **2018**, *46*, D100–D105. [CrossRef]

118. Quek, X.C.; Thomson, D.W.; Maag, J.L.V.; Bartonicek, N.; Signal, B.; Clark, M.B.; Gloss, B.S.; Dinger, M.E. lncRNAdb v2.0: Expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* **2015**, *43*, D168–D173. [CrossRef]

119. Bao, Z.; Yang, Z.; Huang, Z.; Zhou, Y.; Cui, Q.; Dong, D. LncRNADisease 2.0: An updated database of long non-coding RNA–associated diseases. *Nucleic Acids Res.* **2019**, *47*, D1034–D1037. [CrossRef]

120. Gao, Y.; Shang, S.; Guo, S.; Li, X.; Zhou, H.; Liu, H.; Sun, Y.; Wang, J.; Wang, P.; Zhi, H.; et al. Lnc2Cancer 3.0: An updated resource for experimentally supported lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA–seq data. *Nucleic Acids Res.* **2021**, *49*, D1251–D1258. [CrossRef]

121. Cabili, M.N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **2011**, *25*, 1915–1927. [CrossRef]

122. Zhou, K.-R.; Liu, S.; Sun, W.-J.; Zheng, L.-L.; Zhou, H.; Yang, J.-H.; Qu, L.-H. ChIPBase v2.0: Decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res.* **2017**, *45*, D43–D50. [CrossRef]

123. Gerstein, M.B.; Lu, Z.J.; Van Nostrand, E.L.; Cheng, C.; Arshinoff, B.I.; Liu, T.; Yip, K.Y.; Robilotto, R.; Rechtsteiner, A.; Ikegami, K.; et al. Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project. *Science* **2010**, *330*, 1775–1787. [CrossRef]

124. Bernstein, B.E.; Stamatoyannopoulos, J.A.; Costello, J.F.; Ren, B.; Milosavljevic, A.; Meissner, A.; Kellis, M.; Marra, M.A.; Beaudet, A.L.; Ecker, J.R.; et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **2010**, *28*, 1045–1048. [CrossRef]

125. Chen, X.; Yan, G.-Y. Novel human lncRNA–disease association inference based on lncRNA expression profiles. *Bioinformatics* **2013**, *29*, 2617–2624. [CrossRef]

126. Lan, W.; Li, M.; Zhao, K.; Liu, J.; Wu, F.-X.; Pan, Y.; Wang, J. LDAP: A web server for lncRNA-disease association prediction. *Bioinformatics* **2017**, *33*, 458–460. [CrossRef] [PubMed]

127. Sun, J.; Shi, H.; Wang, Z.; Zhang, C.; Liu, L.; Wang, L.; He, W.; Hao, D.; Liu, S.; Zhou, M. Inferring novel lncRNA–disease associations based on a random walk model of a lncRNA functional similarity network. *Mol. BioSyst.* **2014**, *10*, 2074–2081. [CrossRef]

128. Wang, J.; Ma, R.; Ma, W.; Chen, J.; Yang, J.; Xi, Y.; Cui, Q. LncDisease: A sequence based bioinformatics tool for predicting lncRNA–disease associations. *Nucleic Acids Res.* **2016**, *44*, e90. [CrossRef]

129. Schriml, L.M.; Mitraka, E.; Munro, J.; Tauber, B.; Schor, M.; Nickle, L.; Felix, V.; Jeng, L.; Bearer, C.; Lichenstein, R.; et al. Human Disease Ontology 2018 update: Classification, content and workflow expansion. *Nucleic Acids Res.* **2019**, *47*, D955–D962. [CrossRef] [PubMed]

130. Lipscomb, C.E. Medical Subject Headings (MeSH). *Bull. Med. Libr. Assoc.* **2000**, *88*, 265–266. [PubMed]

131. Liu, M.; Wang, Q.; Shen, J.; Yang, B.B.; Ding, X. Circbank: A comprehensive database for circRNA with standard nomenclature. *RNA Biol.* **2019**, *16*, 899–905. [CrossRef]

132. Volders, P.-J.; Anckaert, J.; Verheggen, K.; Nuytens, J.; Martens, L.; Mestdagh, P.; Vandesompele, J. LNCipedia 5: Towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.* **2019**, *47*, D135–D139. [CrossRef] [PubMed]

133. Ning, L.; Cui, T.; Zheng, B.; Wang, N.; Luo, J.; Yang, B.; Du, M.; Cheng, J.; Dou, Y.; Wang, D. MNDR v3.0: Mammal ncRNA–disease repository with increased coverage and annotation. *Nucleic Acids Res.* **2021**, *49*, D160–D164. [CrossRef]

134. Ma, L.; Li, A.; Zou, D.; Xu, X.; Xia, L.; Yu, J.; Bajic, V.B.; Zhang, Z. LncRNAWiki: Harnessing community knowledge in collaborative curation of human long non-coding RNAs. *Nucleic Acids Res.* **2015**, *43*, D187–D192. [CrossRef] [PubMed]

135. Gao, Y.; Li, X.; Shang, S.; Guo, S.; Wang, P.; Sun, D.; Gan, J.; Sun, J.; Zhang, Y.; Wang, J.; et al. LincSNP 3.0: An updated database for linking functional variants to human long non-coding RNAs, circular RNAs and their regulatory elements. *Nucleic Acids Res.* **2021**, *49*, D1244–D1250. [CrossRef] [PubMed]

136. Day, I.N.M. dbSNP in the detail and copy number complexities. *Hum. Mutat.* **2010**, *31*, 2–4. [CrossRef]

137. Miao, Y.-R.; Liu, W.; Zhang, Q.; Guo, A.-Y. lncRNASNP2: An updated database of functional SNPs and mutations in human and mouse lncRNAs. *Nucleic Acids Res.* **2018**, *46*, D276–D280. [CrossRef] [PubMed]

138. Huang, Z.; Shi, J.; Gao, Y.; Cui, C.; Zhang, S.; Li, J.; Zhou, Y.; Cui, Q. HMDD v3.0: A database for experimentally supported human microRNA–disease associations. *Nucleic Acids Res.* **2019**, *47*, D1013–D1017. [CrossRef] [PubMed]

139. Forbes, S.A.; Bindal, N.; Bamford, S.; Cole, C.; Kok, C.Y.; Beare, D.; Jia, M.; Shepherd, R.; Leung, K.; Menzies, A.; et al. COSMIC: Mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **2011**, *39*, D945–D950. [CrossRef]

140. Tate, J.G.; Bamford, S.; Jubb, H.C.; Sondka, Z.; Beare, D.M.; Bindal, N.; Boutselakis, H.; Cole, C.G.; Creatore, C.; Dawson, E.; et al. COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **2019**, *47*, D941–D947. [CrossRef]

141. Mailman, M.D.; Feolo, M.; Jin, Y.; Kimura, M.; Tryka, K.; Bagoutdinov, R.; Hao, L.; Kiang, A.; Paschall, J.; Phan, L.; et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **2007**, *39*, 1181–1186. [CrossRef] [PubMed]

142. Becker, K.G.; Barnes, K.C.; Bright, T.J.; Wang, S.A. The Genetic Association Database. *Nat. Genet.* **2004**, *36*, 431–432. [CrossRef]

143. Beck, T.; Hastings, R.K.; Gollapudi, S.; Free, R.C.; Brookes, A.J. GWAS Central: A comprehensive resource for the comparison and interrogation of genome-wide association studies. *Eur. J. Hum. Genet.* **2014**, *22*, 949–952. [CrossRef] [PubMed]

144. Johnson, A.D.; O'Donnell, C.J. An Open Access Database of Genome-wide Association Results. *BMC Med. Genet.* **2009**, *10*, 6. [CrossRef]

145. Welter, D.; MacArthur, J.; Morales, J.; Burdett, T.; Hall, P.; Junkins, H.; Klemm, A.; Flicek, P.; Manolio, T.; Hindorff, L.; et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **2014**, *42*, D1001–D1006. [CrossRef]

146. Altman, R.B. PharmGKB: A logical home for knowledge relating genotype to drug response phenotype. *Nat. Genet.* **2007**, *39*, 426. [CrossRef]

147. Li, M.J.; Liu, Z.; Wang, P.; Wong, M.P.; Nelson, M.R.; Kocher, J.-P.A.; Yeager, M.; Sham, P.C.; Chanock, S.J.; Xia, Z.; et al. GWASdb v2: An update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.* **2016**, mboxemph44, D869–D876. [CrossRef]

148. Eicher, J.D.; Landowski, C.; Stackhouse, B.; Sloan, A.; Chen, W.; Jensen, N.; Lien, J.-P.; Leslie, R.; Johnson, A.D. GRASP v2.0: An update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res.* **2015**, *43*, D799–D804. [CrossRef]

149. Wang, P.; Li, X.; Gao, Y.; Guo, Q.; Ning, S.; Zhang, Y.; Shang, S.; Wang, J.; Wang, Y.; Zhi, H.; et al. LnCeVar: A comprehensive database of genomic variations that disturb ceRNA network regulation. *Nucleic Acids Res.* **2020**, *48*, D111–D117. [CrossRef]

150. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]

151. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [CrossRef] [PubMed]

152. Hermjakob, H.; Montecchi-Palazzi, L.; Lewington, C.; Mudali, S.; Kerrien, S.; Orchard, S.; Vingron, M.; Roechert, B.; Roepstorff, P.; Valencia, A.; et al. IntAct: An open source molecular interaction database. *Nucleic Acids Res.* **2004**, *32*, D452–D455. [CrossRef] [PubMed]

153. 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. The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **2014**, *42*, D358–D363. [CrossRef] [PubMed]

154. Chatr-aryamontri, A.; Ceol, A.; Palazzi, L.M.; Nardelli, G.; Schneider, M.V.; Castagnoli, L.; Cesareni, G. MINT: The Molecular INTeraction database. *Nucleic Acids Res.* **2007**, *35*, D572–D574. [CrossRef]

155. Licata, L.; Briganti, L.; Peluso, D.; Perfetto, L.; Iannuccelli, M.; Galeota, E.; Sacco, F.; Palma, A.; Nardozza, A.P.; Santonico, E.; et al. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* **2012**, *40*, D857–D861. [CrossRef] [PubMed]

156. Ragueneau, E.; Shrivastava, A.; Morris, J.H.; Del-Toro, N.; Hermjakob, H.; Porras, P. IntAct App: A Cytoscape application for molecular interaction network visualisation and analysis. *Bioinformatics* **2021**, btab319. [CrossRef] [PubMed]

157. Franz, M.; Lopes, C.T.; Huck, G.; Dong, Y.; Sumer, O.; Bader, G.D. Cytoscape.js: A graph theory library for visualisation and analysis. *Bioinformatics* **2016**, *32*, 309–311. [CrossRef]

158. Orchard, S.; Kerrien, S.; Abbani, S.; Aranda, B.; Bhate, J.; Bidwell, S.; Bridge, A.; Briganti, L.; Brinkman, F.S.L.; Brinkman, F.; et al. Protein interaction data curation: The International Molecular Exchange (IMEx) consortium. *Nat. Methods* **2012**, *9*, 345–350. [CrossRef]

159. Xenarios, I.; Rice, D.W.; Salwinski, L.; Baron, M.K.; Marcotte, E.M.; Eisenberg, D. DIP: The database of interacting proteins. *Nucleic Acids Res.* **2000**, *28*, 289–291. [CrossRef]

160. Kotlyar, M.; Pastrello, C.; Malik, Z.; Jurisica, I. IID 2018 update: Context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res.* **2019**, *47*, D581–D589. [CrossRef] [PubMed]

161. Koutrouli, M.; Hatzis, P.; Pavlopoulos, G.A. Exploring Networks in the STRING and Reactome Database. In *Systems Medicine*; Wolkenhauer, O., Ed.; Academic Press: Oxford, UK, 2021; pp. 507–520, ISBN 978-0-12-816078-7.

162. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef]

163. Jassal, B.; Matthews, L.; Viteri, G.; Gong, C.; Lorente, P.; Fabregat, A.; Sidiropoulos, K.; Cook, J.; Gillespie, M.; Haw, R.; et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **2020**, *48*, D498–D503. [CrossRef] [PubMed]

164. Letunic, I.; Khedkar, S.; Bork, P. SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Res.* **2021**, *49*, D458–D460. [CrossRef] [PubMed]

165. Brown, K.R.; Otasek, D.; Ali, M.; McGuffin, M.J.; Xie, W.; Devani, B.; van Toch, I.L.; Jurisica, I. NAViGaTOR: Network Analysis, Visualization and Graphing Toronto. *Bioinformatics* **2009**, *25*, 3327–3329. [CrossRef]

166. Du, Y.; Cai, M.; Xing, X.; Ji, J.; Yang, E.; Wu, J. PINA 3.0: Mining cancer interactome. *Nucleic Acids Res.* **2021**, *49*, D1351–D1357. [CrossRef] [PubMed]

167. Veres, D.V.; Gyurkó, D.M.; Thaler, B.; Szalay, K.Z.; Fazekas, D.; Korcsmáros, T.; Csermely, P. ComPPI: A cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res.* **2015**, *43*, D485–D493. [CrossRef]

168. Giurgiu, M.; Reinhard, J.; Brauner, B.; Dunger-Kaltenbach, I.; Fobo, G.; Frishman, G.; Montrone, C.; Ruepp, A. CORUM: The comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res.* **2019**, *47*, D559–D563. [CrossRef] [PubMed]

169. Meldal, B.H.M.; Bye-A-Jee, H.; Gajdoš, L.; Hammerová, Z.; Horácková, A.; Melicher, F.; Perfetto, L.; Pokorný, D.; Lopez, M.R.; Türková, A.; et al. Complex Portal 2018: Extended content and enhanced visualization tools for macromolecular complexes. *Nucleic Acids Res.* **2019**, *47*, D550–D558. [CrossRef]

170. Kooistra, A.J.; Mordalski, S.; Pándy-Szekeres, G.; Esguerra, M.; Mamyrbekov, A.; Munk, C.; Keserű, G.M.; Gloriam, D.E. GPCRdb in 2021: Integrating GPCR sequence, structure and function. *Nucleic Acids Res.* **2021**, *49*, D335–D343. [CrossRef]

171. Apostolakou, A.E.; Baltoumas, F.A.; Stravopodis, D.J.; Iconomidou, V.A. Extended Human G-Protein Coupled Receptor Network: Cell-Type-Specific Analysis of G-Protein Coupled Receptor Signaling Pathways. *J. Proteome Res.* **2020**, *19*, 511–524. [CrossRef]

172. PRIMES: Protein Interaction Machines in Oncogenic EGF Receptor Signalling. Available online: https://www.research.ed.ac.uk/en/projects/primes-protein-interaction-machines-in-oncogenic-egf-receptor-sig (accessed on 14 August 2021).

173. Ranjan, R.; Khazen, G.; Gambazzi, L.; Ramaswamy, S.; Hill, S.L.; Schürmann, F.; Markram, H. Channelpedia: An integrative and interactive database for ion channels. *Front. Neuroinformatics* **2011**, *5*, 36. [CrossRef]

174. Armstrong, J.F.; Faccenda, E.; Harding, S.D.; Pawson, A.J.; Southan, C.; Sharman, J.L.; Campo, B.; Cavanagh, D.R.; Alexander, S.P.H.; Davenport, A.P.; et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: Extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. *Nucleic Acids Res.* **2020**, *48*, D1006–D1021. [CrossRef] [PubMed]

175. Cotter, D.; Guda, P.; Fahy, E.; Subramaniam, S. MitoProteome: Mitochondrial protein sequence database and annotation system. *Nucleic Acids Res.* **2004**, *32*, D463–467. [CrossRef]

176. Nastou, K.C.; Tsaousis, G.N.; Iconomidou, V.A. PerMemDB: A database for eukaryotic peripheral membrane proteins. *Biochim. Biophys. Acta Biomembr.* **2020**, *1862*, 183076. [CrossRef] [PubMed]

177. Clerc, O.; Deniaud, M.; Vallet, S.D.; Naba, A.; Rivet, A.; Perez, S.; Thierry-Mieg, N.; Ricard-Blum, S. MatrixDB: Integration of new data with a focus on glycosaminoglycan interactions. *Nucleic Acids Res.* **2019**, *47*, D376–D381. [CrossRef]

178. Cowley, M.J.; Pinese, M.; Kassahn, K.S.; Waddell, N.; Pearson, J.V.; Grimmond, S.M.; Biankin, A.V.; Hautaniemi, S.; Wu, J. PINA v2.0: Mining interactome modules. *Nucleic Acids Res.* **2012**, *40*, D862–D865. [CrossRef] [PubMed]

179. Herwig, R.; Hardt, C.; Lienhard, M.; Kamburov, A. Analyzing and interpreting genome data at the network level with Consensus-PathDB. *Nat. Protoc.* **2016**, *11*, 1889–1907. [CrossRef] [PubMed]

180. Pentchev, K.; Ono, K.; Herwig, R.; Ideker, T.; Kamburov, A. Evidence mining and novelty assessment of protein–protein interactions with the ConsensusPathDB plugin for Cytoscape. *Bioinformatics* **2010**, *26*, 2796–2797. [CrossRef]

181. Li, X.; Wang, X.; Snyder, M. Systematic investigation of protein-small molecule interactions. *IUBMB Life* **2013**, *65*, 2–8. [CrossRef] [PubMed]

182. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395. [CrossRef]

183. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107. [CrossRef]

184. Kuhn, M.; Letunic, I.; Jensen, L.J.; Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **2016**, *44*, D1075–D1079. [CrossRef]

185. McFedries, A.; Schwaid, A.; Saghatelian, A. Methods for the Elucidation of Protein-Small Molecule Interactions. *Chem. Biol.* **2013**, *20*, 667–673. [CrossRef]

186. 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. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [CrossRef]

187. Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045–D1053. [CrossRef]

188. Szklarczyk, D.; Santos, A.; von Mering, C.; Jensen, L.J.; Bork, P.; Kuhn, M. STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* **2016**, *44*, D380–D384. [CrossRef]

189. Hecker, N.; Ahmed, J.; von Eichborn, J.; Dunkel, M.; Macha, K.; Eckert, A.; Gilson, M.K.; Bourne, P.E.; Preissner, R. SuperTarget goes quantitative: Update on drug-target interactions. *Nucleic Acids Res.* **2012**, *40*, D1113–D1117. [CrossRef] [PubMed]

190. Preissner, S.; Kroll, K.; Dunkel, M.; Senger, C.; Goldsobel, G.; Kuzman, D.; Guenther, S.; Winnenburg, R.; Schroeder, M.; Preissner, R. SuperCYP: A comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions. *Nucleic Acids Res.* **2010**, *38*, D237–D243. [CrossRef] [PubMed]

191. Mak, L.; Marcus, D.; Howlett, A.; Yarova, G.; Duchateau, G.; Klaffke, W.; Bender, A.; Glen, R.C. Metrabase: A cheminformatics and bioinformatics database for small molecule transporter data analysis and (Q)SAR modeling. *J. Cheminform.* **2015**, *7*, 31. [CrossRef] [PubMed]

192. Ozawa, N.; Shimizu, T.; Morita, R.; Yokono, Y.; Ochiai, T.; Munesada, K.; Ohashi, A.; Aida, Y.; Hama, Y.; Taki, K.; et al. Transporter Database, TP-Search: A Web-Accessible Comprehensive Database for Research in Pharmacokinetics of Drugs. *Pharm. Res.* **2004**, *21*, 2133–2134. [CrossRef]

193. Pontén, F.; Jirström, K.; Uhlen, M. The Human Protein Atlas—a tool for pathology. *J. Pathol.* **2008**, *216*, 387–393. [CrossRef]

194. Hoffmann, M.F.; Preissner, S.C.; Nickel, J.; Dunkel, M.; Preissner, R.; Preissner, S. The Transformer database: Biotransformation of xenobiotics. *Nucleic Acids Res.* **2014**, *42*, D1113–D1117. [CrossRef] [PubMed]

195. Gallina, A.M.; Bisignano, P.; Bergamino, M.; Bordo, D. PLI: A web-based tool for the comparison of protein-ligand interactions observed on PDB structures. *Bioinformatics* **2013**, *29*, 395–397. [CrossRef]

196. Anand, P.; Nagarajan, D.; Mukherjee, S.; Chandra, N. PLIC: Protein–ligand interaction clusters. *Database* **2014**, *2014*, bau029. [CrossRef]

197. Murakami, Y.; Omori, S.; Kinoshita, K. NLDB: A database for 3D protein–ligand interactions in enzymatic reactions. *J. Struct. Funct. Genom.* **2016**, *17*, 101–110. [CrossRef] [PubMed]

198. Ito, J.; Ikeda, K.; Yamada, K.; Mizuguchi, K.; Tomii, K. PoSSuM v.2.0: Data update and a new function for investigating ligand analogs and target proteins of small-molecule drugs. *Nucleic Acids Res.* **2015**, *43*, D392–D398. [CrossRef]

199. Tabei, Y.; Tsuda, K. SketchSort: Fast All Pairs Similarity Search for Large Databases of Molecular Fingerprints. *Mol. Inform.* **2011**, *30*, 801–807. [CrossRef]

200. Wang, C.; Hu, G.; Wang, K.; Brylinski, M.; Xie, L.; Kurgan, L. PDID: Database of molecular-level putative protein–drug interactions in the structural human proteome. *Bioinformatics* **2016**, *32*, 579–586. [CrossRef] [PubMed]

201. Kumar, R.; Chaudhary, K.; Gupta, S.; Singh, H.; Kumar, S.; Gautam, A.; Kapoor, P.; Raghava, G.P.S. CancerDR: Cancer Drug Resistance Database. *Sci. Rep.* **2013**, *3*, 1445. [CrossRef] [PubMed]

202. Gohlke, B.-O.; Nickel, J.; Otto, R.; Dunkel, M.; Preissner, R. CancerResource—Updated database of cancer-relevant proteins, mutations and interacting drugs. *Nucleic Acids Res.* **2016**, *44*, D932–D937. [CrossRef] [PubMed]

203. Coker, E.A.; Mitsopoulos, C.; Tym, J.E.; Komianou, A.; Kannas, C.; Di Micco, P.; Villasclaras Fernandez, E.; Ozer, B.; Antolin, A.A.; Workman, P.; et al. canSAR: Update to the cancer translational research and drug discovery knowledgebase. *Nucleic Acids Res.* **2019**, *47*, D917–D922. [CrossRef] [PubMed]

204. Chiu, Y.-Y.; Lin, C.-T.; Huang, J.-W.; Hsu, K.-C.; Tseng, J.-H.; You, S.-R.; Yang, J.-M. KIDFamMap: A database of kinase-inhibitor-disease family maps for kinase inhibitor selectivity and binding mechanisms. *Nucleic Acids Res.* **2013**, *41*, D430–D440. [CrossRef]

205. Kanev, G.K.; de Graaf, C.; Westerman, B.A.; de Esch, I.J.P.; Kooistra, A.J. KLIFS: An overhaul after the first 5 years of supporting kinase research. *Nucleic Acids Res.* **2021**, *49*, D562–D569. [CrossRef]

206. Urán Landaburu, L.; Berenstein, A.J.; Videla, S.; Maru, P.; Shanmugam, D.; Chernomoretz, A.; Agüero, F. TDR Targets 6: Driving drug discovery for human pathogens through intensive chemogenomic data integration. *Nucleic Acids Res.* **2020**, *48*, D992–D1005. [CrossRef]

207. Wishart, D.; Arndt, D.; Pon, A.; Sajed, T.; Guo, A.C.; Djoumbou, Y.; Knox, C.; Wilson, M.; Liang, Y.; Grant, J.; et al. T3DB: The toxic exposome database. *Nucleic Acids Res.* **2015**, *43*, D928–D934. [CrossRef] [PubMed]

208. Yang, J.; Roy, A.; Zhang, Y. BioLiP: A semi-manually curated database for biologically relevant ligand–protein interactions. *Nucleic Acids Res.* **2012**, *41*, D1096–D1103. [CrossRef] [PubMed]

209. Smith, R.D.; Clark, J.J.; Ahmed, A.; Orban, Z.J.; Dunbar, J.B.; Carlson, H.A. Updates to Binding MOAD (Mother of All Databases): Polypharmacology Tools and Their Utility in Drug Repurposing. *J. Mol. Biol.* **2019**, *431*, 2423–2433. [CrossRef] [PubMed]

210. Chen, X.; Ren, B.; Chen, M.; Liu, M.-X.; Ren, W.; Wang, Q.-X.; Zhang, L.-X.; Yan, G.-Y. ASDCD: Antifungal Synergistic Drug Combination Database. *PLoS ONE* **2014**, *9*, e86499. [CrossRef] [PubMed]

211. Kaur, D.; Patiyal, S.; Sharma, N.; Usmani, S.S.; Raghava, G.P.S. PRRDB 2.0: A comprehensive database of pattern-recognition receptors and their ligands. *Database* **2019**, *2019*, baz076. [CrossRef] [PubMed]

212. Martens, M.; Ammar, A.; Riutta, A.; Waagmeester, A.; Slenter, D.N.; Hanspers, K.; Miller, R.A.; Digles, D.; Lopes, E.N.; Ehrhart, F.; et al. WikiPathways: Connecting communities. *Nucleic Acids Res.* **2021**, *49*, D613–D621. [CrossRef]

213. Kutmon, M.; Lotia, S.; Evelo, C.T.; Pico, A.R. WikiPathways App for Cytoscape: Making biological pathways amenable to network analysis and visualization. *F1000Research* **2014**, *3*, 152. [CrossRef]

214. Kutmon, M.; van Iersel, M.P.; Bohler, A.; Kelder, T.; Nunes, N.; Pico, A.R.; Evelo, C.T. PathVisio 3: An extendable pathway analysis toolbox. *PLoS Comput. Biol.* **2015**, *11*, e1004085. [CrossRef]

215. van Iersel, M.P.; Pico, A.R.; Kelder, T.; Gao, J.; Ho, I.; Hanspers, K.; Conklin, B.R.; Evelo, C.T. The BridgeDb framework: Standardized access to gene, protein and metabolite identifier mapping services. *BMC Bioinform.* **2010**, *11*, 5. [CrossRef]

216. Hoffmann, R. A wiki for the life sciences where authorship matters. *Nat. Genet.* **2008**, *40*, 1047–1051. [CrossRef]

217. Hastings, J.; Owen, G.; Dekker, A.; Ennis, M.; Kale, N.; Muthukrishnan, V.; Turner, S.; Swainston, N.; Mendes, P.; Steinbeck, C. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **2016**, *44*, D1214–D1219. [CrossRef] [PubMed]

218. Wu, G.; Dawson, E.; Duong, A.; Haw, R.; Stein, L. ReactomeFIViz: A Cytoscape app for pathway and network-based data analysis. *F1000Research* **2014**, *3*, 146. [CrossRef]

219. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [CrossRef]

220. Alcántara, R.; Axelsen, K.B.; Morgat, A.; Belda, E.; Coudert, E.; Bridge, A.; Cao, H.; de Matos, P.; Ennis, M.; Turner, S.; et al. Rhea—A manually curated resource of biochemical reactions. *Nucleic Acids Res.* **2012**, *40*, D754–D760. [CrossRef]

221. Saitō, K.; Dixon, R.A.; Willmitzer, L. (Eds.) *Plant Metabolomics*; Biotechnology in Agriculture and Forestry; Springer: Berlin/Heidelberg, Germany, 2006; ISBN 978-3-540-29781-9.

222. Nishida, K.; Ono, K.; Kanaya, S.; Takahashi, K. KEGGscape: A Cytoscape app for pathway data integration. *F1000Research* **2014**, *3*, 144. [CrossRef]

223. Nersisyan, L.; Samsonyan, R.; Arakelyan, A. CyKEGGParser: Tailoring KEGG pathways to fit into systems biology analysis workflows. *F1000Research* **2014**, *3*, 145. [CrossRef] [PubMed]

224. Cytoscape App Store—CytoKegg. Available online: https://apps.cytoscape.org/apps/cytokegg (accessed on 15 August 2021).

225. Boué, S.; Talikka, M.; Westra, J.W.; Hayes, W.; Di Fabio, A.; Park, J.; Schlage, W.K.; Sewer, A.; Fields, B.; Ansari, S.; et al. Causal biological network database: A comprehensive platform of causal biological network models focused on the pulmonary and vascular systems. *Database* **2015**, *2015*, bav030. [CrossRef]

226. Slater, T. Recent advances in modeling languages for pathway maps and computable biological networks. *Drug Discov. Today* **2014**, *19*, 193–198. [CrossRef]

227. Gyori, B.M.; Bachman, J.A.; Subramanian, K.; Muhlich, J.L.; Galescu, L.; Sorger, P.K. From word models to executable models of signaling networks using automated assembly. *Mol. Syst. Biol.* **2017**, *13*, 954. [CrossRef]

228. Lechner, M.; Höhn, V.; Brauner, B.; Dunger, I.; Fobo, G.; Frishman, G.; Montrone, C.; Kastenmüller, G.; Waegele, B.; Ruepp, A. CIDeR: Multifactorial interaction networks in human diseases. *Genome Biol.* **2012**, *13*, R62. [CrossRef] [PubMed]

229. Smith, C.L.; Eppig, J.T. The mammalian phenotype ontology: Enabling robust annotation and comparative analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2009**, *1*, 390–399. [CrossRef] [PubMed]

230. Gremse, M.; Chang, A.; Schomburg, I.; Grote, A.; Scheer, M.; Ebeling, C.; Schomburg, D. The BRENDA Tissue Ontology (BTO): The first all-integrating ontology of all organisms for enzyme sources. *Nucleic Acids Res.* **2011**, *39*, D507–D513. [CrossRef]

231. Yue, M.; Zhou, D.; Zhi, H.; Wang, P.; Zhang, Y.; Gao, Y.; Guo, M.; Li, X.; Wang, Y.; Zhang, Y.; et al. MSDD: A manually curated database of experimentally supported associations among miRNAs, SNPs and human diseases. *Nucleic Acids Res.* **2018**, *46*, D181–D185. [CrossRef]

232. Piñero, J.; Ramírez-Anguita, J.M.; Saüch-Pitarch, J.; Ronzano, F.; Centeno, E.; Sanz, F.; Furlong, L.I. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **2020**, *48*, D845–D855. [CrossRef]

233. Rehm, H.L.; Berg, J.S.; Brooks, L.D.; Bustamante, C.D.; Evans, J.P.; Landrum, M.J.; Ledbetter, D.H.; Maglott, D.R.; Martin, C.L.; Nussbaum, R.L.; et al. ClinGen—The Clinical Genome Resource. *N. Engl. J. Med.* **2015**, *372*, 2235–2242. [CrossRef] [PubMed]

234. Martin, A.R.; Williams, E.; Foulger, R.E.; Leigh, S.; Daugherty, L.C.; Niblock, O.; Leong, I.U.S.; Smith, K.R.; Gerasimenko, O.; Haraldsdottir, E.; et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat. Genet.* **2019**, *51*, 1560–1565. [CrossRef]

235. Gutiérrez-Sacristán, A.; Grosdidier, S.; Valverde, O.; Torrens, M.; Bravo, À.; Piñero, J.; Sanz, F.; Furlong, L.I. PsyGeNET: A knowledge platform on psychiatric disorders and their genes: Table 1. *Bioinformatics* **2015**, *31*, 3075–3077. [CrossRef] [PubMed]

236. Weinreich, S.S.; Mangon, R.; Sikkens, J.J.; en Teeuw, M.E.; Cornel, M.C. Orphanet: A European database for rare diseases. *Ned. Tijdschr. Geneeskd.* **2008**, *152*, 518–519.

237. Köhler, S.; Gargano, M.; Matentzoglu, N.; Carmody, L.C.; Lewis-Smith, D.; Vasilevsky, N.A.; Danis, D.; Balagura, G.; Baynam, G.; Brower, A.M.; et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* **2021**, *49*, D1207–D1217. [CrossRef] [PubMed]

238. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; Wiegers, J.; Wiegers, T.C.; Mattingly, C.J. Comparative Toxicogenomics Database (CTD): Update 2021. *Nucleic Acids Res.* **2021**, *49*, D1138–D1143. [CrossRef]

239. Landrum, M.J.; Chitipiralla, S.; Brown, G.R.; Chen, C.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; Kaur, K.; Liu, C.; et al. ClinVar: Improvements to accessing data. *Nucleic Acids Res.* **2020**, *48*, D835–D844. [CrossRef]

240. Shefchek, K.A.; Harris, N.L.; Gargano, M.; Matentzoglu, N.; Unni, D.; Brush, M.; Keith, D.; Conlin, T.; Vasilevsky, N.; Zhang, X.A.; et al. The Monarch Initiative in 2019: An integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.* **2020**, *48*, D704–D715. [CrossRef]

241. Grever, M.R.; Schepartz, S.A.; Chabner, B.A. The National Cancer Institute: Cancer drug discovery and development program. *Semin. Oncol.* **1992**, *19*, 622–638.

242. Fung, K.W.; Xu, J.; Ameye, F.; Gutiérrez, A.R.; Busquets, A. Re-purposing the ICD-9-CM Procedures Index for Coding in ICD-10-PCS and SNOMED CT. *AMIA Annu. Symp. Proc.* **2018**, *2018*, 450–459. [PubMed]

243. Zeng, W.; Min, X.; Jiang, R. EnDisease: A manually curated database for enhancer-disease associations. *Database* **2019**, *2019*, baz020. [CrossRef] [PubMed]

244. Wang, J.; Cao, Y.; Zhang, H.; Wang, T.; Tian, Q.; Lu, X.; Lu, X.; Kong, X.; Liu, Z.; Wang, N.; et al. NSDNA: A manually curated database of experimentally supported ncRNAs associated with nervous system diseases. *Nucleic Acids Res.* **2017**, *45*, D902–D907. [CrossRef] [PubMed]

245. Nastou, K.C.; Nasi, G.I.; Tsiolaki, P.L.; Litou, Z.I.; Iconomidou, V.A. AmyCo: The amyloidoses collection. *Amyloid* **2019**, *26*, 112–117. [CrossRef]

246. Bragin, E.; Chatzimichali, E.A.; Wright, C.F.; Hurles, M.E.; Firth, H.V.; Bevan, A.P.; Swaminathan, G.J. DECIPHER: Database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res.* **2014**, *42*, D993–D1000. [CrossRef]

247. Vasaikar, S.V.; Padhi, A.K.; Jayaram, B.; Gomes, J. NeuroDNet—An open source platform for constructing and analyzing neurodegenerative disease networks. *BMC Neurosci.* **2013**, *14*, 3. [CrossRef]

248. Cook, H.; Doncheva, N.; Szklarczyk, D.; von Mering, C.; Jensen, L. Viruses.STRING: A Virus-Host Protein-Protein Interaction Database. *Viruses* **2018**, *10*, 519. [CrossRef]

249. Ammari, M.G.; Gresham, C.R.; McCarthy, F.M.; Nanduri, B. HPIDB 2.0: A curated database for host–pathogen interactions. *Database* **2016**, *2016*, baw103. [CrossRef]

250. Calderone, A.; Licata, L.; Cesareni, G. VirusMentha: A new resource for virus-host protein interactions. *Nucleic Acids Res.* **2015**, *43*, D588–D592. [CrossRef]

251. Huerta-Cepas, J.; Szklarczyk, D.; Forslund, K.; Cook, H.; Heller, D.; Walter, M.C.; Rattei, T.; Mende, D.R.; Sunagawa, S.; Kuhn, M.; et al. eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **2016**, *44*, D286–D293. [CrossRef]

252. Stelzer, G.; Rosen, N.; Plaschkes, I.; Zimmerman, S.; Twik, M.; Fishilevich, S.; Stein, T.I.; Nudel, R.; Lieder, I.; Mazor, Y.; et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses: The GeneCards Suite. In *Current Protocols in Bioinformatics*; Bateman, A., Pearson, W.R., Stein, L.D., Stormo, G.D., Yates, J.R., Eds.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2016; pp. 1–30, ISBN 978-0-471-25095-1.

253. Lane, L.; Argoud-Puy, G.; Britan, A.; Cusin, I.; Duek, P.D.; Evalet, O.; Gateau, A.; Gaudet, P.; Gleizes, A.; Masselot, A.; et al. neXtProt: A knowledge platform for human proteins. *Nucleic Acids Res.* **2012**, *40*, D76–D83. [CrossRef]

254. Li, Y.; Wang, C.; Miao, Z.; Bi, X.; Wu, D.; Jin, N.; Wang, L.; Wu, H.; Qian, K.; Li, C.; et al. ViRBase: A resource for virus–host ncRNA-associated interactions. *Nucleic Acids Res.* **2015**, *43*, D578–D582. [CrossRef]

255. Xie, J.; Zhang, M.; Zhou, T.; Hua, X.; Tang, L.; Wu, W. Sno/scaRNAbase: A curated database for small nucleolar RNAs and cajal body-specific RNAs. *Nucleic Acids Res.* **2007**, *35*, D183–D187. [CrossRef] [PubMed]

256. Fauquet, C.; Fargette, D. International Committee on Taxonomy of Viruses and the 3,142 unassigned species. *Virol. J.* **2005**, *2*, 64. [CrossRef] [PubMed]

257. Gao, N.L.; Zhang, C.; Zhang, Z.; Hu, S.; Lercher, M.J.; Zhao, X.-M.; Bork, P.; Liu, Z.; Chen, W.-H. MVP: A microbe–phage interaction database. *Nucleic Acids Res.* **2018**, *46*, D700–D707. [CrossRef]

258. Fouts, D.E. Phage_Finder: Automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Res.* **2006**, *34*, 5839–5851. [CrossRef]

259. Paez-Espino, D.; Eloe-Fadrosh, E.A.; Pavlopoulos, G.A.; Thomas, A.D.; Huntemann, M.; Mikhailova, N.; Rubin, E.; Ivanova, N.N.; Kyrpides, N.C. Uncovering Earth's virome. *Nature* **2016**, *536*, 425–430. [CrossRef]

260. Bleves, S.; Dunger, I.; Walter, M.C.; Frangoulidis, D.; Kastenmüller, G.; Voulhoux, R.; Ruepp, A. HoPaCI-DB: Host- *Pseudomonas* and *Coxiella* interaction database. *Nucleic Acids Res.* **2014**, *42*, D671–D676. [CrossRef]

261. Poelen, J.H.; Simons, J.D.; Mungall, C.J. Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecol. Inform.* **2014**, *24*, 148–159. [CrossRef]

262. Vandepitte, L.; Vanhoorne, B.; Decock, W.; Vranken, S.; Lanssens, T.; Dekeyzer, S.; Verfaille, K.; Horton, T.; Kroh, A.; Hernandez, F.; et al. A decade of the World Register of Marine Species—General insights and experiences from the Data Management Team: Where are we, what have we learned and how can we continue? *PLoS ONE* **2018**, *13*, e0194599. [CrossRef]

263. Wieczorek, J.; Bloom, D.; Guralnick, R.; Blum, S.; Döring, M.; Giovanni, R.; Robertson, T.; Vieglais, D. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* **2012**, *7*, e29715. [CrossRef]

264. Parr, C.S.; Wilson, N.; Leary, P.; Schulz, K.; Lans, K.; Walley, L.; Hammock, J.; Goddard, A.; Rice, J.; Studer, M.; et al. The Encyclopedia of Life v2: Providing Global Access to Knowledge About Life on Earth. *Biodivers. Data J.* **2014**, *2*, e1079. [CrossRef] [PubMed]

265. Lehtonen, J.; Heiska, S.; Pajari, M.; Tegelberg, R.; Saarenmaa, H.; Jones, M.B.; Gries, C. The process of digitizing natural history collection specimens at Digitarium. In Proceedings of the Environmental Information Management Conference 2011, Santa Barbara, CA, USA, 28–29 September 2011. [CrossRef]

266. Fortuna, M.A.; Ortega, R.; Bascompte, J. The Web of Life. *arXiv* **2014**, arXiv:1403.2575.

267. Thompson, R.M.; Brose, U.; Dunne, J.A.; Hall, R.O.; Hladyz, S.; Kitching, R.L.; Martinez, N.D.; Rantala, H.; Romanuk, T.N.; Stouffer, D.B.; et al. Food webs: Reconciling the structure and function of biodiversity. *Trends Ecol. Evol.* **2012**, *27*, 689–697. [CrossRef] [PubMed]

268. Bat Eco-Interactions. Available online: https://www.batbase.org/ (accessed on 14 August 2021).

269. Pavlopoulos, G.A.; Wegener, A.-L.; Schneider, R. A survey of visualization tools for biological network analysis. *BioData Min.* **2008**, *1*, 12. [CrossRef] [PubMed]

270. Pavlopoulos, G.A.; Secrier, M.; Moschopoulos, C.N.; Soldatos, T.G.; Kossida, S.; Aerts, J.; Schneider, R.; Bagos, P.G. Using graph theory to analyze biological networks. *BioData Min.* **2011**, *4*, 10. [CrossRef] [PubMed]

271. Pavlopoulos, G.A.; Malliarakis, D.; Papanikolaou, N.; Theodosiou, T.; Enright, A.J.; Iliopoulos, I. Visualizing genome and systems biology: Technologies, tools, implementation techniques and trends, past, present and future. *GigaScience* **2015**, *4*, 38. [CrossRef] [PubMed]

272. O'Donoghue, S.I.; Gavin, A.-C.; Gehlenborg, N.; Goodsell, D.S.; Hériché, J.-K.; Nielsen, C.B.; North, C.; Olson, A.J.; Procter, J.B.; Shattuck, D.W.; et al. Visualizing biological data—Now and in the future. *Nat. Methods* **2010**, *7*, S2–S4. [CrossRef]

273. Gehlenborg, N.; O'Donoghue, S.I.; Baliga, N.S.; Goesmann, A.; Hibbs, M.A.; Kitano, H.; Kohlbacher, O.; Neuweger, H.; Schneider, R.; Tenenbaum, D.; et al. Visualization of omics data for systems biology. *Nat. Methods* **2010**, *7*, S56–S68. [CrossRef] [PubMed]

274. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proc. Int. AAAI Conf. Web Soc. Media* **2009**, *3*, 361–362.

275. Mrvar, A.; Batagelj, V. Analysis and visualization of large networks with program package Pajek. *Complex Adapt. Syst. Model.* **2016**, *4*, 1–8. [CrossRef]

276. Köhler, J.; Baumbach, J.; Taubert, J.; Specht, M.; Skusa, A.; Rüegg, A.; Rawlings, C.; Verrier, P.; Philippi, S. Graph-based analysis and visualization of experimental results with ONDEX. *Bioinformatics* **2006**, *22*, 1383–1390. [CrossRef] [PubMed]

277. Iragne, F.; Nikolski, M.; Mathieu, B.; Auber, D.; Sherman, D. ProViz: Protein interaction visualization and exploration. *Bioinformatics* **2005**, *21*, 272–274. [CrossRef]

278. Hu, Z.; Hung, J.-H.; Wang, Y.; Chang, Y.-C.; Huang, C.-L.; Huyck, M.; DeLisi, C. VisANT 3.5: Multi-scale network visualization, analysis and inference based on the gene ontology. *Nucleic Acids Res.* **2009**, *37*, W115–W121. [CrossRef]

279. Breitkreutz, B.-J.; Stark, C.; Tyers, M. Osprey: A network visualization system. *Genome Biol.* **2003**, *4*, R22. [CrossRef]

280. Pavlopoulos, G.A.; O'Donoghue, S.I.; Satagopam, V.P.; Soldatos, T.G.; Pafilis, E.; Schneider, R. Arena3D: Visualization of biological networks in 3D. *BMC Syst. Biol.* **2008**, *2*, 104. [CrossRef] [PubMed]

281. Secrier, M.; Pavlopoulos, G.A.; Aerts, J.; Schneider, R. Arena3D: Visualizing time-driven phenotypic differences in biological systems. *BMC Bioinform.* **2012**, *13*, 45. [CrossRef]

282. Karatzas, E.; Baltoumas, F.A.; Panayiotou, N.A.; Schneider, R.; Pavlopoulos, G.A. Arena3Dweb: Interactive 3D visualization of multilayered networks. *Nucleic Acids Res.* **2021**, *49*, W36–W45. [CrossRef]

283. Freeman, T.C.; Horsewell, S.; Patir, A.; Harling-Lee, J.; Regan, T.; Shih, B.B.; Prendergast, J.; Hume, D.A.; Angus, T. Graphia: A Platform for the Graph-Based Visualisation and Analysis of Complex Data. *bioRxiv* **2020**. [CrossRef]

284. Koutrouli, M.; Karatzas, E.; Papanikolopoulou, K.; Pavlopoulos, G.A. NORMA: The Network Makeup Artist—A Web Tool for Network Annotation Visualization. *Genom. Proteom. Bioinform.* **2021**. [CrossRef]

285. Theocharidis, A.; van Dongen, S.; Enright, A.J.; Freeman, T.C. Network visualization and analysis of gene expression data using BioLayout Express(3D). *Nat. Protoc.* **2009**, *4*, 1535–1550. [CrossRef] [PubMed]

286. Luo, W.; Brouwer, C. Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **2013**, *29*, 1830–1831. [CrossRef]

287. Longabaugh, W.J.R. BioTapestry: A Tool to Visualize the Dynamic Properties of Gene Regulatory Networks. *Methods Mol. Biol.* **2012**, *786*, 359–394. [CrossRef]

288. Darzi, Y.; Letunic, I.; Bork, P.; Yamada, T. iPath3.0: Interactive pathways explorer v3. *Nucleic Acids Res.* **2018**, *46*, W510–W513. [CrossRef]

289. Thimm, O.; Bläsing, O.; Gibon, Y.; Nagel, A.; Meyer, S.; Krüger, P.; Selbig, J.; Müller, L.A.; Rhee, S.Y.; Stitt, M. MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **2004**, *37*, 914–939. [CrossRef]

290. Rodchenkov, I.; Babur, O.; Luna, A.; Aksoy, B.A.; Wong, J.V.; Fong, D.; Franz, M.; Siper, M.C.; Cheung, M.; Wrana, M.; et al. Pathway Commons 2019 Update: Integration, analysis and exploration of pathway data. *Nucleic Acids Res.* **2020**, *48*, D489–D497. [CrossRef]

291. Doncheva, N.T.; Assenov, Y.; Domingues, F.S.; Albrecht, M. Topological analysis and interactive visualization of biological networks and protein structures. *Nat. Protoc.* **2012**, *7*, 670–685. [CrossRef]

292. Athanasiadis, E.I.; Bourdakou, M.M.; Spyrou, G.M. ZoomOut: Analyzing Multiple Networks as Single Nodes. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2015**, *12*, 1213–1216. [CrossRef]

293. Brohée, S.; Faust, K.; Lima-Mendez, G.; Sand, O.; Janky, R.; Vanderstocken, G.; Deville, Y.; van Helden, J. NeAT: A toolbox for the analysis of biological networks, clusters, classes and pathways. *Nucleic Acids Res.* **2008**, *36*, W444–W451. [CrossRef]

294. Theodosiou, T.; Efstathiou, G.; Papanikolaou, N.; Kyrpides, N.C.; Bagos, P.G.; Iliopoulos, I.; Pavlopoulos, G.A. NAP: The Network Analysis Profiler, a web tool for easier topological analysis and comparison of medium-scale biological networks. *BMC Res. Notes* **2017**, *10*, 278. [CrossRef]

295. Koutrouli, M.; Theodosiou, T.; Iliopoulos, I.; Pavlopoulos, G.A. The Network Analysis Profiler (NAP v2.0): A web tool for visual topological comparison between multiple networks. *EMBnet.J.* **2021**, *26*, e943. [CrossRef]

296. Karatzas, E.; Gkonta, M.; Hotova, J.; Baltoumas, F.A.; Kontou, P.I.; Bobotsis, C.J.; Bagos, P.G.; Pavlopoulos, G.A. VICTOR: A visual analytics web application for comparing cluster sets. *Comput. Biol. Med.* **2021**, *135*, 104557. [CrossRef]

297. Leskovec, J.; Sosič, R. SNAP: A General-Purpose Network Analysis and Graph-Mining Library. *ACM Trans. Intell. Syst. Technol.* **2016**, *8*, 1–20. [CrossRef] [PubMed]

298. Csardi, G.; Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Syst.* **2006**, *1695*, 1–9.

299. Hagberg, A.; Swart, P.; S Chult, D. Exploring Network Structure, Dynamics, and Function Using Networkx. United States. 2008. Available online: https://www.osti.gov/servlets/purl/960616 (accessed on 15 August 2021).

300. Peixoto, T.P. The Graph-Tool Python Library. Figshare. Dataset. 2015. Available online: https://figshare.com/articles/dataset/graph_tool/1164194 (accessed on 14 August 2021).

301. Pavlopoulos, G.A.; Paez-Espino, D.; Kyrpides, N.C.; Iliopoulos, I. Empirical Comparison of Visualization Tools for Larger-Scale Network Analysis. *Adv. Bioinforma.* **2017**, *2017*, 1278932. [CrossRef]

302. Fruchterman, T.M.J.; Reingold, E.M. Graph drawing by force-directed placement. *Softw. Pract. Exp.* **1991**, *21*, 1129–1164. [CrossRef]

303. Yifan, H. Efficient, high-quality force-directed graph drawing. *Math. J.* **2005**, *10*, 37–71.

304. Adai, A.T.; Date, S.V.; Wieland, S.; Marcotte, E.M. LGL: Creating a map of protein function with an algorithm for visualizing very large biological networks. *J. Mol. Biol.* **2004**, *340*, 179–190. [CrossRef]

305. Kamada, T.; Kawai, S. An algorithm for drawing general undirected graphs. *Inf. Process. Lett.* **1989**, *31*, 7–15. [CrossRef]

306. Maleki, F.; Ovens, K.; Hogan, D.J.; Kusalik, A.J. Gene Set Analysis: Challenges, Opportunities, and Future Research. *Front. Genet.* **2020**, *11*, 654. [CrossRef] [PubMed]

307. Shi Jing, L.; Fathiah Muzaffar Shah, F.; Saberi Mohamad, M.; Moorthy, K.; Deris, S.; Zakaria, Z.; Napis, S. A Review on Bioinformatics Enrichment Analysis Tools Towards Functional Analysis of High Throughput Gene Set Data. *Curr. Proteom.* **2015**, *12*, 14–27. [CrossRef]

308. Raudvere, U.; Kolberg, L.; Kuzmin, I.; Arak, T.; Adler, P.; Peterson, H.; Vilo, J. g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **2019**, *47*, W191–W198. [CrossRef] [PubMed]

309. Mi, H.; Ebert, D.; Muruganujan, A.; Mills, C.; Albou, L.-P.; Mushayamaha, T.; Thomas, P.D. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **2021**, *49*, D394–D403. [CrossRef]

310. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [CrossRef] [PubMed]

311. Liao, Y.; Wang, J.; Jaehnig, E.J.; Shi, Z.; Zhang, B. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **2019**, *47*, W199–W205. [CrossRef]

312. Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G.V.; Clark, N.R.; Ma'ayan, A. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **2013**, *14*, 128. [CrossRef]

313. Carbon, S.; Ireland, A.; Mungall, C.J.; Shu, S.; Marshall, B.; Lewis, S.; AmiGO Hub; Web Presence Working Group. AmiGO: Online access to ontology and annotation data. *Bioinformatics* **2009**, *25*, 288–289. [CrossRef] [PubMed]

314. Subhash, S.; Kanduri, C. GeneSCF: A real-time based functional enrichment tool with support for multiple organisms. *BMC Bioinform.* **2016**, *17*, 365. [CrossRef]

315. Zhang, D.; Hu, Q.; Liu, X.; Zou, K.; Sarkodie, E.K.; Liu, X.; Gao, F. AllEnricher: A comprehensive gene set function enrichment tool for both model and non-model species. *BMC Bioinform.* **2020**, *21*, 106. [CrossRef] [PubMed]

316. Schölz, C.; Lyon, D.; Refsgaard, J.C.; Jensen, L.J.; Choudhary, C.; Weinert, B.T. Avoiding abundance bias in the functional annotation of post-translationally modified proteins. *Nat. Methods* **2015**, *12*, 1003–1004. [CrossRef] [PubMed]

317. Bindea, G.; Mlecnik, B.; Hackl, H.; Charoentong, P.; Tosolini, M.; Kirilovsky, A.; Fridman, W.-H.; Pagès, F.; Trajanoski, Z.; Galon, J. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **2009**, *25*, 1091–1093. [CrossRef]

318. Han, H.; Lee, S.; Lee, I. NGSEA: Network-Based Gene Set Enrichment Analysis for Interpreting Gene Expression Phenotypes with Functional Gene Sets. *Mol. Cells* **2019**, *42*, 579–588. [CrossRef]

319. Eden, E.; Navon, R.; Steinfeld, I.; Lipson, D.; Yakhini, Z. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinform.* **2009**, *10*, 48. [CrossRef]

320. Thanati, F.; Karatzas, E.; Baltoumas, F.A.; Stravopodis, D.J.; Eliopoulos, A.G.; Pavlopoulos, G.A. FLAME: A Web Tool for Functional and Literature Enrichment Analysis of Multiple Gene Lists. *Biology* **2021**, *10*, 665. [CrossRef]

321. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **2012**, *16*, 284–287. [CrossRef] [PubMed]

322. Yousif, A.; Drou, N.; Rowe, J.; Khalfan, M.; Gunsalus, K.C. NASQAR: A web-based platform for high-throughput sequencing data analysis and visualization. *BMC Bioinform.* **2020**, *21*, 267. [CrossRef] [PubMed]

323. Stephens, Z.D.; Lee, S.Y.; Faghri, F.; Campbell, R.H.; Zhai, C.; Efron, M.J.; Iyer, R.; Schatz, M.C.; Sinha, S.; Robinson, G.E. Big Data: Astronomical or Genomical? *PLoS Biol.* **2015**, *13*, e1002195. [CrossRef]

324. MBInfo | Defining Mechanobiology. Available online: https://www.mechanobio.info/ (accessed on 16 August 2021).

325. Baltoumas, F.A.; Zafeiropoulou, S.; Karatzas, E.; Paragkamian, S.; Thanati, F.; Iliopoulos, I.; Eliopoulos, A.G.; Schneider, R.; Jensen, L.J.; Pafilis, E.; et al. OnTheFly[2.0]: A Text-Mining Web Application for Automated Biomedical Entity Recognition, Document Annotation, Network and Functional Enrichment Analysis. *bioRxiv* **2021**. [CrossRef]

326. Pafilis, E.; Buttigieg, P.L.; Ferrell, B.; Pereira, E.; Schnetzer, J.; Arvanitidis, C.; Jensen, L.J. EXTRACT: Interactive extraction of environment metadata and term suggestion for metagenomic sample annotation. *Database* **2016**, *2016*, baw005. [CrossRef] [PubMed]

327. Pavlopoulos, G.A.; Promponas, V.J.; Ouzounis, C.A.; Iliopoulos, I. Biological information extraction and co-occurrence analysis. *Methods Mol. Biol.* **2014**, *1159*, 77–92. [CrossRef] [PubMed]