

WikiPathways: Integrating Pathway Knowledge with Clinical Data

73

Denise N. Slenter, Martina Kutmon,
and Egon L. Willighagen

Contents

Introduction	1458
IEMBase and WikiPathways	1458
Machine-Readable Metabolic Pathway Models	1458
Using Pathway Models to Analyze Clinical Data	1461
Limitations	1465
Conclusions	1465
References	1465

Summary

Throughout the chapters in this book, pathways are used to visualize how genetically inheritable metabolic disorders are related. These pathways provide common conceptual models which explain groups of chemical reactions within their biological context. Visual representations of the reactions in biological pathway diagrams provide intuitive ways to study the complex metabolic processes. In order to link (clinical) data to these path-

ways, they have to be understood by computers. Understanding how to move from a regular pathway drawing to its machine-readable counterpart is pertinent for creating proper models. This chapter outlines the various aspects of the digital counterparts of the pathway diagrams in this book, connecting them to databases and using them in data integration and analysis. This is followed by three examples of bioinformatics applications including a pathway enrichment analysis, a biological network extension, and a final example that integrates pathways with clinical biomarker data.

The original version of this chapter was previously published non-open access. A Correction to this chapter is available at https://doi.org/10.1007/978-3-030-67727-5_74

D. N. Slenter (✉) · E. L. Willighagen
Department of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, Maastricht, The Netherlands
e-mail: denise.slenter@maastrichtuniversity.nl;
egon.willighagen@maastrichtuniversity.nl

M. Kutmon
Department of Bioinformatics - BiGCaT, NUTRIM, Maastricht University, Maastricht, The Netherlands

Maastricht Centre for Systems Biology – MaCSBio, Maastricht University, Maastricht, The Netherlands
e-mail: martina.kutmon@maastrichtuniversity.nl

Introduction

The metabolic pathways in this book are common conceptual models which help us understand groups of chemical reactions within their biological context. These conversion reactions are catalyzed by enzymes and triggered by receptors or transporters, causing a wide variety of metabolites to be present in bodily fluids and tissues. Visual representations of the reactions in biological pathway diagrams provide intuitive ways to study complex metabolic processes. If we want to link clinical data to these pathways, we need pathways that can be understood by computers. By creating machine-readable versions of the pathways relevant for (rare) disorders, clinical data can be processed and analyzed in an automated fashion, allowing fast and visual interpretation (Kutmon et al. 2014a; Villaveces et al. 2015). However, modeling these pathways for data analysis comes with some challenges and limitations (Howe et al. 2008; Khare et al. 2016), which are discussed in this chapter in more detail.

Even though there is still a substantial amount of unexplored territory in human molecular biology, data on biological mechanisms, pathways, and related diseases has increasingly become available due to improvements in measurement and computational techniques. A crucial step to enable advanced data analysis is structuring and sharing the knowledge obtained from biological experiments. This chapter provides some examples of what types of data analysis can be performed with machine-readable pathways and related knowledge. More examples are available in literature, e.g., visualization of drug metabolism related expression changes (Jennen et al. 2010), integrating molecular interaction data (Herwig et al. 2016), and fully automatable processing steps of metabolomics data (Stanstrup et al. 2019). We hope to inspire the metabolic rare disease community to aid our quest in transforming, structuring, and collecting knowledge about molecular processes in machine-readable pathway models and databases. Finally, all these data acquisition and modeling steps will lead to a better understanding of the phenotype of a patient.

IEMBase and WikiPathways

The knowledge captured in the chapters of this book holds vital information on genetically inheritable metabolic diseases, genes, and proteins involved, metabolic biomarkers, pathways, relevant literature, diagnosis, and treatment. A great effort has been done digitizing most of this information in the IEMBase (www.iembase.org) (Lee et al. 2018), but one key element missing were machine-readable pathways, which are being added through a collaboration with the WikiPathways pathway database (www.wikipathways.org) (Kelder et al.

2012; Kutmon et al. 2016; Slenter et al. 2018). This database allows researchers to add machine-readable pathways, including literature references, and creates a traceable history of edits. The results are freely accessible and reusable by everyone in the world. Furthermore, adding a pathway to WikiPathways exposes the biological knowledge captured in the model via various other data formats (<http://help.wikipathways.org>) and tools (<http://tools.wikipathways.org>), allowing researchers to directly integrating the new knowledge within their tool of choice.

We aim to provide all pathways in the chapters of this book as machine-readable pathway models in WikiPathways. You can find the currently available pathways on <http://iem.wikipathways.org>.

Machine-Readable Metabolic Pathway Models

Researchers can use different tools to model machine-readable pathways from schematic drawings in publications. WikiPathways stores all pathways in the graphical pathway markup language (GPML), which can be drawn in PathVisio (www.pathvisio.org) (Kutmon et al. 2015). This format is flexible enough to allow modeling of detailed biological phenomena, while relying on a machine-readable structured backbone for automated data analysis. In the following subsections, we highlight relevant topics for creating pathways on genetically inheritable metabolic rare diseases. This information is complemented with an online step-by-step tutorial (academy.wikipathways.org).

Modeling Biological Entities

The biological entities in GPML-encoded pathways (e.g., gene products, proteins, and metabolites) are captured as *DataNodes* (Fig. 73.1a). These nodes have a textual label and biological type and can contain the following additional information: literature references, free text comments, and a unique database identifier (Fig. 73.1b). The textual label is the visible name of the entity while the type represents which biological entity is modeled: genes are modeled as *GeneProducts*, proteins and enzymes as *Proteins* and chemicals as *Metabolites* (Fig. 73.1a). Additional modeling options include *Complexes*, *Groups*, *RNAs*, (full) *Pathways*, and *States* such as posttranslational modifications (Fig. 73.1c). *DataNodes* can be connected to literature, ideally using PubMed identifiers. Another section allows users to add free text in the comment field, to explain additional details relevant for the pathway. Finally, a database identifier and the accompanying database (Fig. 73.1b) provides the framework required for data integration later (Figs. 73.4

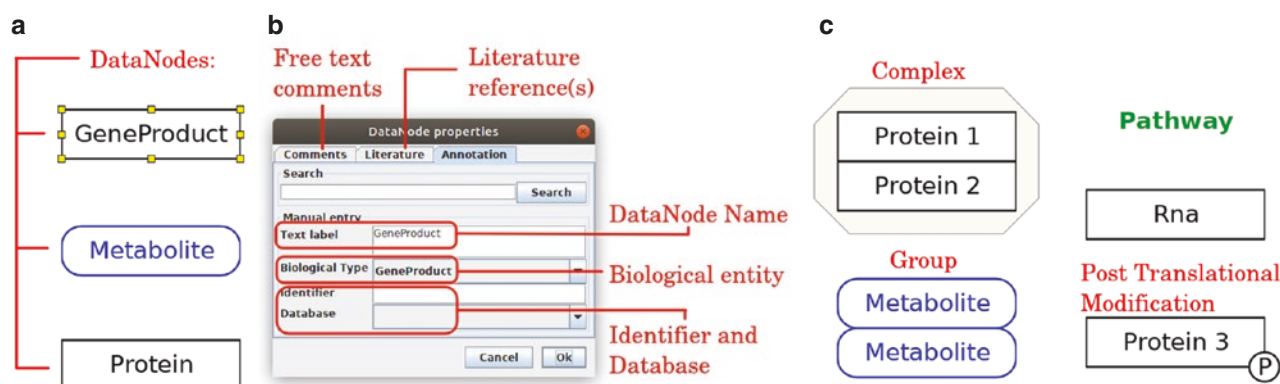


Fig. 73.1 Visualization of modeling properties for biological entities in PathVisio. (a) Main *DataNodes*, where the *GeneProduct* is selected (indicated with small yellow blocks). (b) Pop-up menu allowing connecting a *DataNode* to literature, identifier, and database. (c) Additional

modeling options, where multiple *DataNodes* can be used to form *Complexes* and *Groups*. Nodes can also be used to point to other *Pathways* (including identifier) and *RNA*. *DataNodes* can also be extended with a *State* like posttranslational modifications

and 73.5). An online database identifier allows the retrieval of additional facts about the entity in the pathway. PathVisio makes use of the BridgeDb identifier mapping framework (www.bridgedb.org) (van Iersel et al. 2010) and can therefore support a variety of databases and mappings between them. Since these identifiers are crucial for data analysis, the next two sections discuss them in more detail.

Identifiers for Pathway Entities

There are many online resources storing information about the proteins, molecules, and interactions in pathways. Ideally, each element in the pathway is annotated with a specific identifier from one of the online databases. Because there are many different databases to choose from, we provide some general guidelines for the annotations of *DataNodes* and interactions in pathways.

Genes and proteins can be annotated with over 60 different databases. However, there are some subtle differences between what type of information is modeled in these databases. Ensembl (Cunningham et al. 2019) and NCBI (Entrez) Gene (Agarwala et al. 2018) focus on gene and transcript identifiers, while UniProt (UniProt Consortium 2019) models their data at a protein level. Therefore, one gene identifier in Ensembl could point to multiple UniProt entries. Enzyme commission numbers (EC-codes) (McDonald and Tipton 2014) can be very useful to annotate a group of enzymes which serve a similar biological function or to classify a chemical conversion to a specific reaction mechanism, without knowing the actual protein structure or gene involved. However, since this classification is not specific for one gene or protein, numerous mappings can be created which complicates data analysis. Thus,

regarding identification of biological entities, the more specifically the identifier points to one entity, the more straightforward data can be connected to *GeneProduct* (Ensembl, NCBI Gene) and *Protein* (UniProt) *DataNodes*. For *Complexes* with multiple enzymes, each individual enzyme should receive a unique identifier from UniProt. In addition, distinctive isoforms could also have unique identifiers and be drawn as separate *DataNodes*.

Metabolites can be annotated with over 25 databases, where most of these databases have their own focus (nutrition, toxicology, human metabolism, medication) and possess different levels of chemical detail. These different levels can be quite relevant for the biological implications of these compounds and corresponding interactions, such as stereochemistry (e.g., chirality, isomers), protonation state, (de) phosphorylation, and tautomerization. For several pathways and reactions, the specific level can also be unknown, which is particularly the case for lipid pathways. These are more often considered to behave biologically similar when the head and tail of the lipid are comparable; however a small difference in number of double bonds or location thereof could lead to distinct biological behavior. Identifiers exist to be able to add groups of compounds as *Metabolite DataNodes* in pathways. Nevertheless, this does not solve the issue of straightforward data analysis as discussed for genes and proteins. We would therefore advise to use chemical identifiers which correspond to the known level of chemical detail, e.g., ChEBI (Hastings et al. 2016) for metabolites, DrugBank (Wishart et al. 2018) for drugs, and LIPID MAPS (Fahy et al. 2009) for lipids.

Interactions can currently be annotated with 14 databases, for which some allow easy integration with other resources. A good coverage of metabolic conversions between metabolites is provided by the Rhea database (Lombardot et al. 2019).

Interaction Types

As mentioned previously, interactions are separate elements within the pathway model. An interaction clearly connects two or more biological entities to depict a relationship between them. Understanding the biological meaning of a connection between different entities is needed to create accurate machine-readable models. PathVisio supports several types of interaction (standards): basic interactions, molecular interaction map (MIM) interactions (Luna et al. 2011), and the interaction types described in the systems biology graphical notation (SBGN) (van Iersel et al. 2012).

For the pathways in this book, we advise the application of MIM interactions, which include conversion and catalysis for reactions between metabolites and related enzyme(s) (Fig. 73.2a); stimulation and inhibition for signaling functions (Fig. 73.2b); transcription/translation for *GeneProduct* to *Protein* (Fig. 73.2c); modification for posttranslational or other modifications and binding for complex formation (Fig. 73.2d); and translocation for transport of metabolites between different cellular compartments (Fig. 73.2e).

Modeling Diseases and Interactions

Most pathways in this book clearly indicate which step(s) in a pathway are disease causing. This disease information can be added to pathways in several manners. First, diseases can be added to a pathway as text labels and connected to the related gene or protein. Second, a pathway can be tagged with terms from the Human Disease Ontology (Köhler et al. 2019) Pathway Ontology (Petri et al. 2014) and Cell Type Ontology (Diehl et al. 2016) on WikiPathways. The information enables systematic search and browse functionalities with clearly defined child-parent relationships; for example find all pathways, which are linked to the term “inborn error of metabolism pathway” (purl.bioontology.org/ontology/PW/PW:0001589) or all pathways linked to possible child terms. Third, genes and proteins are often linked to OMIM gene entries (Amberger et al. 2015) through BridgeDb, allowing navigation from a specific pathway to disease databases. Finally, including the disease name or class in the title of the pathway or description also helps searching for relevant pathways.

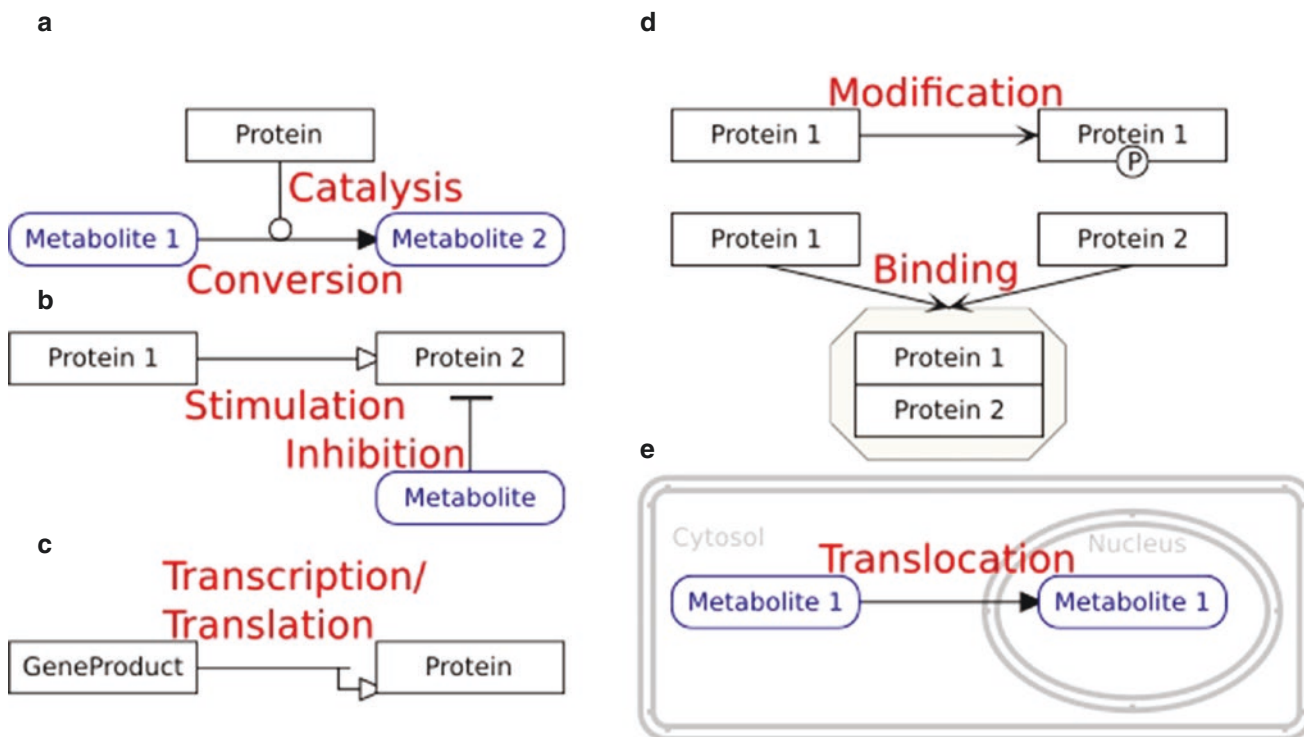


Fig. 73.2 Overview of MIM interaction types and how these can be used to connect biological entities in PathVisio. (a) *Metabolite 1* is enzymatically converted to *Metabolite 2*, which is catalysed by a *Protein*. (b) *Protein 1* stimulates *Protein 2*; *Protein 2* is inhibited by a *Metabolite*. (c) A *GeneProduct* is transcribed and/or translated to a

Protein. (d) top: A *Protein* undergoing a phosphorylation (P) Post-Translational-Modification (PTM); bottom: two *Proteins* binding together to form a complex. (e) A translocation from cytosol to nucleus for a *Metabolite*

Genetically Inheritable Metabolic Disorder Pathways on WikiPathways

A majority of the pathways in the chapters of this book are already available on WikiPathways (iem.wikipathways.org). As an example, Fig. 73.3 visualizes a complete example of a machine-readable version of the purine pathway (Chap. 13, WikiPathways:WP4792, www.wikipathways.org/instance/WP4792), which is linked to over 20 genetically inheritable diseases (available at WikiPathways:WP4224). Even though at first glance, this figure resembles the original drawing quite closely, all the individual metabolites, proteins, and biomarkers are annotated with identifiers and linked to databases, all interactions connect these elements together and have the appropriate MIM types. The following sections provide several examples on how these pathways can now be used for data analysis.

Using Pathway Models to Analyze Clinical Data

Pathways can be used to analyze various types of clinical data including whole exome sequencing (WES), transcriptomics and proteomics, genome-wide association studies

(GWAS), metabolomics, or targeted chemical assay data. In this section, we provide several examples on how the created pathway models can be used for data analysis. Scripts and instructions for performing these analyses can be found online (bigcat-um.github.io/IEMPPathwayAnalysis).

Pathway Analysis

The availability of pathways as machine-readable models enables us to visualize molecular data on the elements in the pathways and to perform advanced computational analysis including gene set enrichment and network analysis. While whole-genome sequencing is already well established in clinical practice, transcriptome analysis seems promising for diagnosing rare disease patients (Gonorazky et al. 2019).

Pathway analysis for transcriptomics data is a powerful tool to put the data into a biological context. As an example, we selected a publicly available transcriptomic dataset of patients with Lesch-Nyhan disease (LND) (geo:[GSE24345](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24345), omim:[300322](https://www.ncbi.nlm.nih.gov/omim/300322)) (Kang et al. 2011). The disease is caused by mutations in the *HPRT1* gene (ensembl:[ENSG00000165704](https://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000165704)), producing the hypoxanthine phos-

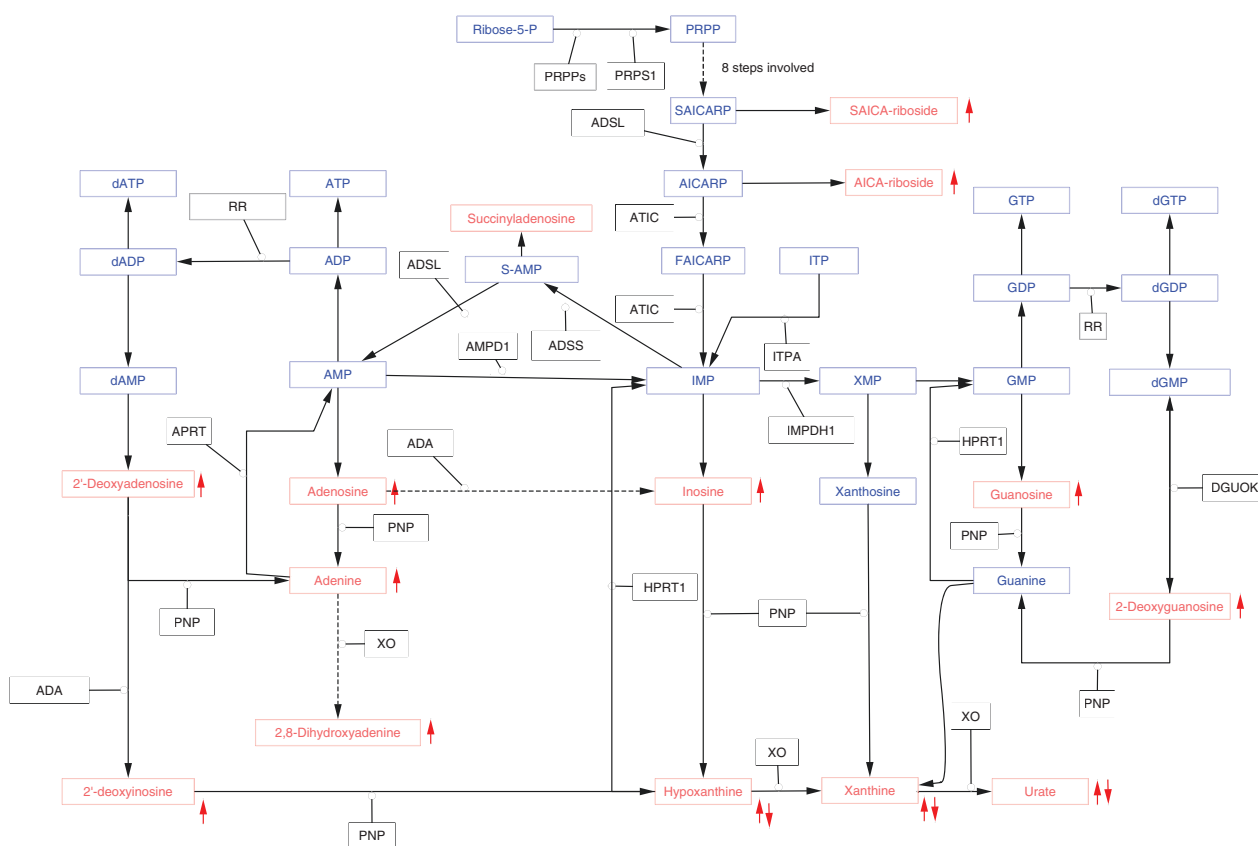


Fig. 73.3 Machine-readable version of the purine pathway, with metabolites in blue, proteins in black, and biomarker molecules in red (Wiki Pathways:WP4792)

phoribosyltransferase 1 enzyme (uniprot:P00492) which enables cells to recycle purines. After statistical analysis with GEO2R (Davis and Meltzer 2007), the differential gene expression data was visualized on the previously described purine pathway (Fig. 73.3) using the WikiPathways app in Cytoscape (apps.cytoscape.org/apps/wikipathways) (Kutmon et al. 2014b); see Fig. 73.4. Cytoscape (www.cytoscape.org) (Shannon et al. 2003) is a widely adopted network analysis and visualization software tool, and the WikiPathways app provides direct access to the WikiPathways pathway content. The log2 fold changes (difference in gene expression between patients and control group) are visualized as a gradient from blue (less expressed) over white (not changed) to red (more expressed). A strong downregulation of the *HPRT1* gene is immediately visible by the corresponding blue gene boxes. Based on this dataset, it seems that several of the other enzymes in the pathway (*PNP*, *APRT*, *PRPS1*) attempt to compensate for the lower transcript availability of *HPRT1* and are upregulated in LND patients. Other enzymes, which are up- or downstream from the affected purine pathway, could also show changes relevant for the phenotype of the

patient. Therefore, enrichment analysis on other pathway models can be used to assess the relevance of these pathways and find interesting affected processes in a certain dataset. Enrichment analysis showed several immune-related processes including the complement system as well as the Wnt signaling pathway affected in LND patients. Reimand et al. published a protocol that provides a description of pathway enrichment analysis and a practical step-by-step guide (Reimand et al. 2019).

Tutorial videos and R scripts for the data visualization (1A, Fig. 73.4) and pathway enrichment analysis (1B) can be found in the PathwayAnalysis folder on GitHub (bigcat-um.github.io/IEMPathwayAnalysis).

Pathways as a Source for Network Biology

Well-annotated pathway models are also an immensely useful source for network biology approaches. They can be used to extend biological pathways and networks (combinations of pathways) with additional knowledge, such as drug targets.

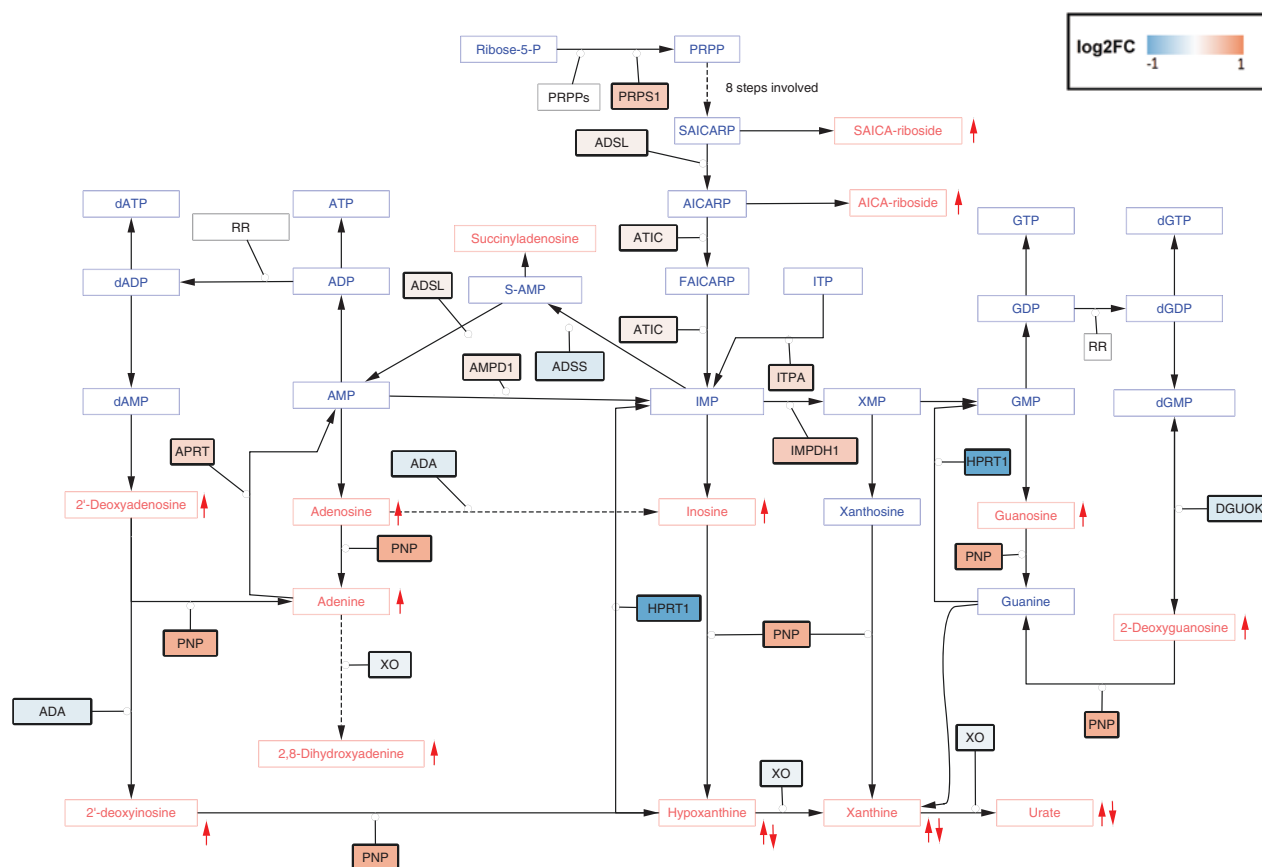


Fig. 73.4 Visualization of changes in gene expression in patients with Lesch-Nyhan disease in the purine metabolism pathway (WikiPathways: WP4792)

Using the WikiPathways app in Cytoscape, pathways can be visualized in two different views: the pathway view and the network view. The network view retains the existing biological interactions in the pathway only and applies an automatic layout allowing the pathways to be used for network analysis. Pathways can then be automatically enriched with additional information, e.g., regulatory mechanisms, known drugs, or disease annotations. The CyTargetLinker app for Cytoscape was specifically designed for this purpose (cytargetlinker.github.io) (Kutmon et al. 2019). In the following example, we will extend the purine metabolism pathway with approved drugs from DrugBank; see Fig. 73.5. The pathway contains fifteen enzymes, shown as rounded rectangles. In

DrugBank, 21 drugs (shown as green diamonds) are reported to target one or more of these enzymes. Nine of the 15 enzymes are targeted by at least one drug, as shown in orange. The LND-causing gene *HPRT1* (highlighted with the red rectangle) is targeted by three drugs, azathioprine, mercaptopurine, and α -phosphoribosyl pyrophosphoric acid. The first two are immunosuppressants and inhibit *HPRT1*. The last molecule is a key substance in the biosynthesis of histidine, tryptophan, purine, and pyrimidine nucleotides and has unknown pharmacological action on *HPRT1*.

Tutorial videos and R scripts for the following example can be found in the Network Analysis folder on GitHub ([big-cat-um.github.io/IEMPathwayAnalysis](https://github.com/big-cat-um/IEMPathwayAnalysis)).

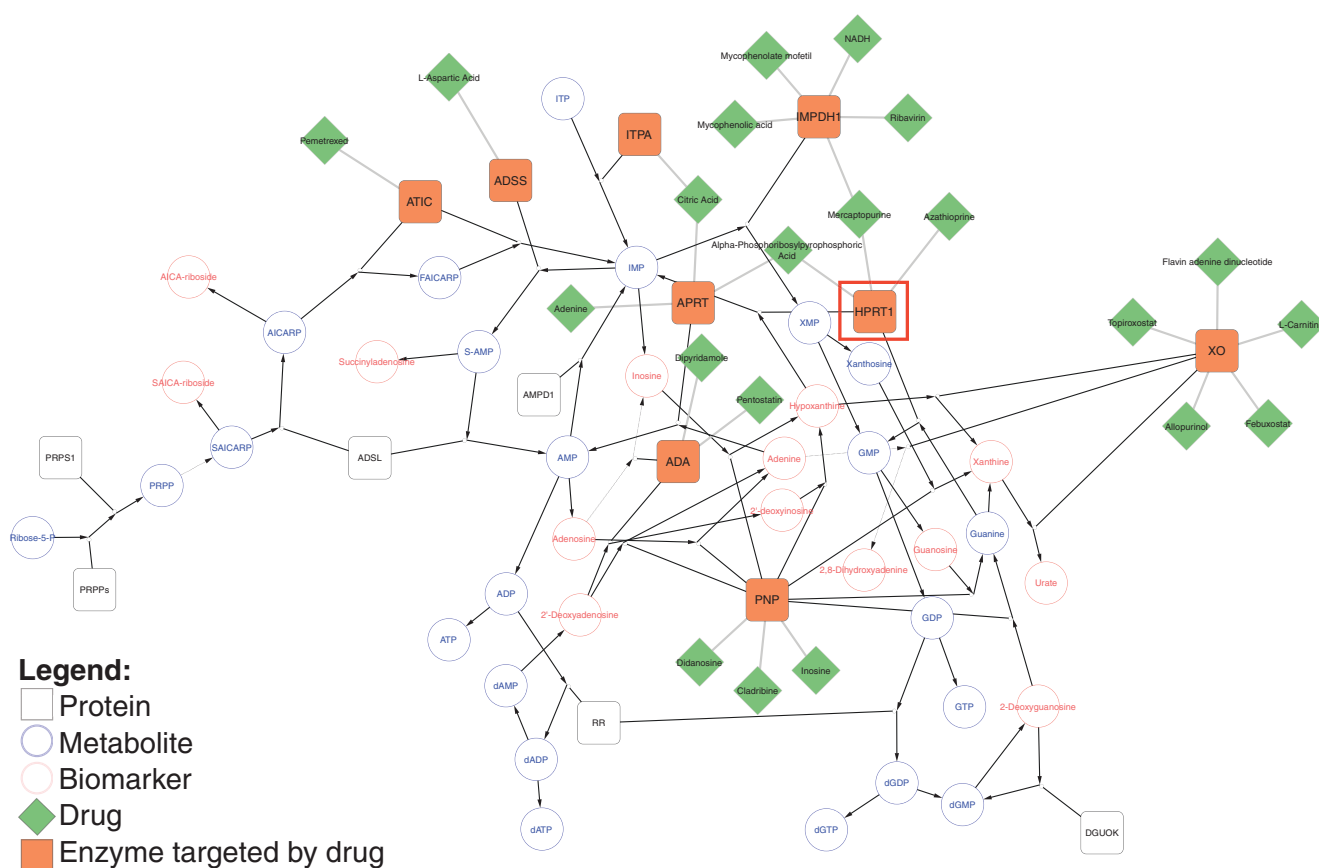


Fig. 73.5 Known drug-target interactions (from DrugBank) for the purine metabolism pathway

Linking Chemical (Biomarker) Data with RDF

The two examples in the previous sections work well for transcriptomics data. However, connecting metabolomics or chemical biomarker data to pathways can be challenging with the methods presented above, since the amount of data is significantly lower. Furthermore, as previously mentioned in Sect. 73.3.2, various levels of chemical detail exist within pathway models. To overcome these problems, we want to highlight another method to connect biomarkers data with pathways using an automated approach. The workflow for this example is visualized in Fig. 73.6a. One starts with chemical data from targeted (clinically validated) assays or with (un)targeted metabolomics data from mass spectrometry (MS) and/or nuclear magnetic resonance (NMR). After preprocessing of this data, for example, in R (Stanstrup et al. 2019), and annotation of relevant peaks with the corresponding chemical structure, one can also add an identifier from a (supported) database to the data. For subsequent data analysis steps, we used BridgeDb to link these identifiers to their corresponding InChIKey (a shortened version of the InChI) (Heller et al. 2015), which links the chemical structure to the original compound identifier in a machine-readable manner. The InChIKey consists of three parts (separated by a bar “-”); the first describes the general structure of a molecule, the second its stereochemistry, and the third the charge of the molecule. For example, “CKLJMWZTIZZHCS-REOHCLBHSA-M” is the InChIKey for L-aspartic acid monoanion, where the “M” stands for a -1 charge. In order to link these compounds to (metabolic) pathways, we will use a SPARQL query (Galgonek

et al. 2016). This query is a structured method to ask questions to a database such as WikiPathways, which has been converted to the Resource Description Framework (RDF) format (Waagmeester et al. 2016). This RDF format unifies and harmonizes pathway data over all models, with several filtering and search options; data can be queried for specific species, *DataNodes*, literature, and more. Figure 73.6b provides an example of a SPARQL query where we investigated the occurrence of five compounds within the pathway models of WikiPathways. More details on the structure of the WikiPathways RDF and example queries are available on the Semantic Web portal: rdf.wikipathways.org. We also provide a beginners’ tutorial on how to write SPARQL queries in the SPARQL folder on GitHub ([bigcat-um.github.io/IEMPPathwayAnalysis](https://github.com/bigcat-um/IEMPPathwayAnalysis)).

To visualize the concepts of SPARQL and data mapping, we chose the five naturally charged proteinogenic amino acids (positive charge: aspartic acid (Asp, D) and glutamic acid (Glu, E); negative charge: arginine (Arg, R), histidine (His, H), and lysine (Lys, K)). Lines 6–10 in Fig. 73.6b provide the InChIKeys for these compounds, which we will link to pathway data in line 15 (`wp:bdbInChIKey ?inchikey`). The results of the example query reveal in which pathways these compounds can be found, which are listed from lowest to highest (line 22) occurrence counts (line 4) in pathways, to find the most relevant pathways for further analysis. This results in 15 pathways, which all have one of these compounds present (RDF data release 2021-03-10). However, since pathways can be annotated with the uncharged form of these amino acids as well, we can also query all compounds



Fig. 73.6 Proposed workflow for data analysis of metabolic biomarkers with semantic web technologies. **(a)** Moving from chemical assay data to data interpretation using BridgeDb, SPARQL and Cytoscape. **(b)** Example of a SPARQL query on metabolic data in the WikiPathways RDF

with the same stereochemistry independent of charge status, obtaining a more complete result. By changing the end of each compounds InChIKey to their neutral counterpart (“CKLJMWTZIZZHCS-REOHCLBHSA-M” becomes “CKLJMWTZIZZHCS-REOHCLBHSA-N”) and changing line 15 to “`rdfs:seeAlso ?inchikey`,” we obtain a total of 38 pathways, with 15 pathways containing 2 of the 5 amino acids (either in their neutral or charged form). This example is explained in more detail in the SPARQL folder on GitHub (bigcat-um.github.io/IEMPathwayAnalysis).

The next step in this workflow downloads the relevant pathways from WikiPathways in Cytoscape; see Sect. 73.4.1 and 73.4.2 for further details. When moving to a network tool, more advanced analysis approaches can be used. The chemical data can be visualized on the nodes in the network, while the edges (interactions between nodes) can be used to visualize fluxomic data (if available). After the data has been added, interpretation can be facilitated by additional features in Cytoscape, such as visualizing the chemical structure of the compounds, connecting several pathways (as networks) to obtain a more complete overview, adding other experimental data (e.g., transcriptomics; see Sect. 73.4.1), or expanding the network with biological knowledge from other databases (see Sect. 73.4.2).

Limitations

Every type of data analysis comes with its own set of limitations. For pathway analysis, the results depend on the coverage of the selected database(s), the mappings between identifiers from the dataset to the pathway knowledge, the statistics, and cut-off value used for the fold change to discover significant findings. Especially regarding this first issue, WikiPathways is a useful resource, since users can add missing information on pathways and interaction themselves, allowing them and others to use them directly in data analysis. Furthermore, the machine-readable model behind this database is flexible enough to accommodate the needs of researchers from the genetic inheritable metabolic disease research area. We hope that this chapter, as well as the created examples of machine-readable pathways from the figures in this book, provides other users inspiration to add more biological knowledge to databases. For more information on how to model these pathways, please visit help.wikipathways.org.

Conclusions

While regular pathway drawings are a great resource to visualize relevant biological interactions, these figures are not interactive and not directly reusable for data analysis. Understanding how to move from a regular pathway drawing to its machine-readable counterpart is pertinent for creating

proper models. As shown in this chapter, having a digital pathway can link to reference databases and allows us to perform data integration and analysis. This will require some time and effort from the side of the user; however once these skills are mastered, adding new information is relatively easy, will decrease other research time needed for data analysis, and aid the research community as a whole.

References

- Agarwala R, Barrett T, Beck J, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2018;46:D8–D13.
- Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 2015;43:D789–98.
- Cunningham F, Achuthan P, Akanni W, et al. Ensembl 2019. *Nucleic Acids Res.* 2019;47:D745–51.
- Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics.* 2007;23:1846–7.
- Diehl AD, Meehan TF, Bradford YM, Brush MH, Dahdul WM, Dougall DS, et al. The Cell Ontology 2016: Enhanced content, modularization, and ontology interoperability. *J Biomed Semantics.* 2016;7(1):44.
- Fahy E, Subramaniam S, Murphy RC, Nishijima M, Raetz CRH, Shimizu T, Spener F, van Meer G, Wakelam MJO, Dennis EA. Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res.* 2009;50:S9–S14.
- Galgonek J, Hurt T, Michlíková V, Onderka P, Schwarz J, Vondrášek J. Advanced SPARQL querying in small molecule databases. *J Cheminform.* 2016;8:31.
- Gonorazky HD, Naumenko S, Ramani AK, et al. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *Am J Hum Genet.* 2019;104:466–83.
- 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:D1214–9.
- Heller SR, McNaught A, Pletnev I, Stein S, Tchekhovskoi D. InChI, the IUPAC international chemical identifier. *J Cheminform.* 2015;7:23.
- Herwig R, Hardt C, Lienhard M, Kamburov A. Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nat Protoc.* 2016;11:1889–907.
- Howe D, Costanzo M, Fey P, et al. The future of biocuration. *Nature.* 2008;455:47–50.
- Jennen DGJ, Gaj S, Giesbertz PJ, van Delft JHM, Evelo CT, Kleinjans JCS. Biotransformation pathway maps in WikiPathways enable direct visualization of drug metabolism related expression changes. *Drug Discov Today.* 2010;15:851–8.
- Kang TH, Guibinga G-H, Friedmann T, Friedmann T. HPRT deficiency coordinately dysregulates canonical Wnt and Presenilin-1 Signaling: a neuro-developmental regulatory role for a housekeeping gene? *PLoS One.* 2011;6:e16572.
- Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, Pico AR. WikiPathways: building research communities on biological pathways. *Nucleic Acids Res.* 2012;40:D1301–7.
- Khare R, Good BM, Leaman R, Su AI, Lu Z. Crowdsourcing in biomedicine: challenges and opportunities. *Brief Bioinform.* 2016;17:23–32.

- Köhler S, Carmody L, Vasilevsky N, et al. Expansion of the human phenotype ontology (HPO) knowledge base and resources. *Nucleic Acids Res.* 2019;47:D1018–27.
- Kutmon M, Ehrhart F, Willighagen EL, Evelo CT, Coort SL. CyTargetLinker app update: a flexible solution for network extension in Cytoscape. *F1000Res.* 2019;7:743.
- Kutmon M, Evelo CT, Coort SL. A network biology workflow to study transcriptomics data of the diabetic liver. *BMC Genomics.* 2014a;15:971.
- Kutmon M, Lotia S, Evelo CT, Pico AR. WikiPathways app for Cytoscape: making biological pathways amenable to network analysis and visualization. *F1000Res.* 2014b;3:152.
- Kutmon M, Riutta A, Nunes N, et al. WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic Acids Res.* 2016;44:D488–94.
- Kutmon M, van Iersel MP, Bohler A, Kelder T, Nunes N, Pico AR, Evelo CT. PathVisio 3: an extendable pathway analysis toolbox. *PLoS Comput Biol.* 2015;11:e1004085.
- Lee JJY, Wasserman WW, Hoffmann GF, van Karnebeek CDM, Blau N. Knowledge base and mini-expert platform for the diagnosis of inborn errors of metabolism. *Genet Med.* 2018;20:151–8.
- Lombardot T, Morgat A, Axelsen KB, et al. Updates in Rhea: SPARQLing biochemical reaction data. *Nucleic Acids Res.* 2019;47:D596–600.
- Luna A, Sunshine ML, van Iersel MP, Aladjem MI, Kohn KW. PathVisio-MIM: PathVisio plugin for creating and editing molecular interaction maps (MIMs). *Bioinformatics.* 2011;27:2165–6.
- McDonald AG, Tipton KF. Fifty-five years of enzyme classification: advances and difficulties. *FEBS J.* 2014;281:583–92.
- Petri V, Jayaraman P, Tutaj M, et al. The pathway ontology—updates and applications. *J Biomed Semantics.* 2014;5:7.
- Reimand J, Isserlin R, Voisin V, et al. Pathway enrichment analysis and visualization of omics data using g:profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc.* 2019;14:482–517.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504.
- Slenter DN, Kutmon M, Hanspers K, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* 2018;46:D661–7.
- Stanstrup J, Broeckling C, Helmus R, et al. The metaRbolomics toolbox in Bioconductor and beyond. *Metabolites.* 2019;9:200.
- UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019;47:D506–15.
- van Iersel MP, Pico AR, Kelder T, Gao J, Ho I, Hanspers K, Conklin BR, Evelo CT. The BridgeDb framework: standardized access to gene, protein and metabolite identifier mapping services. *BMC Bioinform.* 2010;11:5.
- van Iersel MP, Villeger AC, Czauderna T, et al. Software support for SBGN maps: SBGN-ML and LibSBGN. *Bioinformatics.* 2012;28:2016–21.
- Villaveces J, Koti P, Habermann B. Tools for visualization and analysis of molecular networks, pathways, and -omics data. *Adv Appl Bioinforma Chem.* 2015;11.
- Waagmeester A, Kutmon M, Riutta A, Miller R, Willighagen EL, Evelo CT, Pico AR. Using the semantic web for rapid integration of WikiPathways with other biological online data resources. *PLoS Comput Biol.* 2016;12:e1004989.
- Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46:D1074–82.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

