

EDITORIAL

Ten simple rules for creating reusable pathway models for computational analysis and visualization

Kristina Hanspers¹, Martina Kutmon^{2,3}, Susan L. Coort², Daniela Digles⁴, Lauren J. Dupuis², Friederike Ehrhart², Finterly Hu^{2,3}, Elisson N. Lopes⁵, Marvin Martens², Nhung Pham^{2,3}, Woosub Shin², Denise N. Slenter², Andra Waagmeester⁶, Egon L. Willighagen², Laurent A. Winckers², Chris T. Evelo^{2,3}, Alexander R. Pico^{1*}

1 Institute of Data Science and Biotechnology, Gladstone Institutes, San Francisco, California, United States of America, **2** Department of Bioinformatics—BiGCaT, NUTRIM, Maastricht University, Maastricht, the Netherlands, **3** Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, the Netherlands, **4** Department of Pharmaceutical Sciences, Division of Pharmaceutical Sciences, University of Vienna, Vienna, Austria, **5** Instituto de Ciencias Biológicas, Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, **6** Micelio, Antwerp, Belgium

☯ These authors contributed equally to this work.

* alex.pico@gladstone.ucsf.edu



OPEN ACCESS

Citation: Hanspers K, Kutmon M, Coort SL, Digles D, Dupuis LJ, Ehrhart F, et al. (2021) Ten simple rules for creating reusable pathway models for computational analysis and visualization. *PLoS Comput Biol* 17(8): e1009226. <https://doi.org/10.1371/journal.pcbi.1009226>

Editor: Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

Published: August 19, 2021

Copyright: © 2021 Hanspers et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the European Union's Horizon 2020 Research and Innovation Programme [OpenRiskNet: 731075 to MM, CTE; EJP-RD: 825575 to FE, WS, CTE; FNS-Cloud: 863059 to SLC, CTE; EU-ToxRisk: 681002 to MM, CTE; NanoCommons: 731032 to ELW, LAW], ZonMw [10430012010015 to MK, SLC, LJD, FE, FH, NP, ELW, CTE] and the National Institute of General Medical Sciences [R01 GM100039 to KH, ARP; R01 GM089820 to AW]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Introduction

Pathway models are an effective way to capture and share our current understanding of biological processes. A pathway model is defined here as a set of interactions among biological entities (e.g., proteins and metabolites) relevant to a particular context, curated and organized to illustrate a particular process. Properly modeled pathways can be used in the analysis and visualization of diverse types of omics and other biomedical data [1,2]. The modeling process involves taking our knowledge about biological pathways—however messy and incomplete—and encoding it in standardized data formats that can be shared, reused, and synthesized with other knowledge in accordance with the Findable, Accessible, Interoperable, and Reusable (FAIR) principles [3]. The rules presented here serve as an introduction and guide to the pathway modeling process, leveraging freely available tools and resources.

Biological pathway information is often conveyed as published figures, and rules for better figures, in general, are also relevant when producing pathway models [4]. However, pathway models are more than just figures. In addition to providing an intuitive depiction of a biological process that is easy to understand for humans, they also provide relevant annotations and metadata to be processed by computers. Similarly, rules for network visualizations can be applied to pathway models [5], but the distinct context, layout, and usage of pathway models necessitate specific guidelines and rules. Pathway models are described with specific languages, such as the Systems Biology Graphical Notation (SBGN) [6], the Systems Biology Markup Language (SBML) [7], the Biological Pathway Exchange (BioPAX) format [8], the Graphical Pathway Markup Language (GPML) [9], and many more.

Here, we describe a set of rules for constructing pathway models to optimize their use both as graphical representations for human consumption and as FAIR resources for computational analysis. The rules range from reusability and dissemination (Rules 1, 9, and 10), intuitive visual concepts (Rules 2, 7, and 8), to enabling computational analysis (Rules 3 to 6). We hope that these rules provide a simple framework for pathway model curators who want to create (re)usable resources for the scientific community.

Competing interests: The authors have declared that no competing interests exist.

Rule 1: If possible, reuse and extend existing models

When constructing a pathway model, first, research existing content from the myriad of online databases containing pathway-related information. Searchable databases for biological pathways include Reactome [10], WikiPathways [11], BioCyc [12], KEGG [13], PharmGKB [14], cBioPortal [15], PANTHER [16], ConsensusPathDB [17], Pathway Figure OCR [18], and Pathway Commons [19]. The pathway models from these databases can be used as a source material to be cited or to help build your pathway modeling project. The WikiPathways database is unique in supporting direct community participation, allowing researchers to make edits, extensions, and new versions of existing content using PathVisio [9]. Each pathway model database provides a different angle or abstraction of the existing knowledge on a given topic, relying on distinct approaches to knowledge collection and curation represented by distinct data formats. For example, the citric acid (tricarboxylic acid (TCA)) cycle pathway content was found to be scattered over 5 different pathway databases [20]. The largest source of pathway information by far is scientific literature, where pathway figures provide critical summaries and conceptual models for primary discoveries, which are often the starting point for pathway curation efforts. The Pathway Figure OCR database project has extracted entities from tens of thousands of pathway figures published over the past 25 years and now enables simple search by standardized gene symbols, diseases, and keywords [18]. Tools and resources are being developed to access this content (<https://gladstone-bioinformatics.shinyapps.io/shiny-25years>) and to facilitate curation by providing starter pathway models with extracted gene sets and by prioritizing pathway figures based on uniqueness against existing pathway databases.

Other valuable sources of pathway information are network and interaction databases such as BioModels [21], NDEX [22], Rhea [23], IntAct [24], Complex Portal [25], and STRING [26]. Some of these databases like NDEX contain context-specific networks, while others are collections of binary interactions (e.g., Rhea and STRING). Most of these databases support searching for more than one entity at a time, as well as by keywords. Ontologies and gene set collections such as Gene Ontology [27] and MSigDB [28] can also be useful in researching the components of a pathway. Additionally, individual entities can be looked up in gene, protein, and chemical databases such as Ensembl [29], UniProt [30], and ChEBI [31], which collect relevant annotations and links to higher-order resources.

By researching these sources, one not only collects supporting evidence for their pathway model but also compiles a more complete and original model. All sources should be cited in the pathway model as part of the modeling process (see Rule 5). There are a variety of free tools available for pathway modeling, including PathVisio [9], CellDesigner [32], Newt [33], SBGN-ED [34], and yED+ySBGN [35]. Many of these are able to import and export existing models in standard formats (SBGN, SBML, GPML, and BioPAX), allowing curators to adapt and extend existing models [36].

Rule 2: Determine the correct scope and level of detail for the pathway model

To illustrate a particular biological process, pathway models contain entities and their interactions relevant to the particular context. Thus, the scope of the pathway model, i.e., the entities and boundaries we choose, should be based on what we are trying to illustrate. When deciding on the scope of a pathway model, consider which reactions and entities are crucial to understanding the relevant process. As an example, comparing 2 “citric acid cycle” pathway models at WikiPathways, the canonical pathway [37] has a high level of detail describing the enzymatic reactions critical to this pathway, with links to related pathways represented as pathway nodes

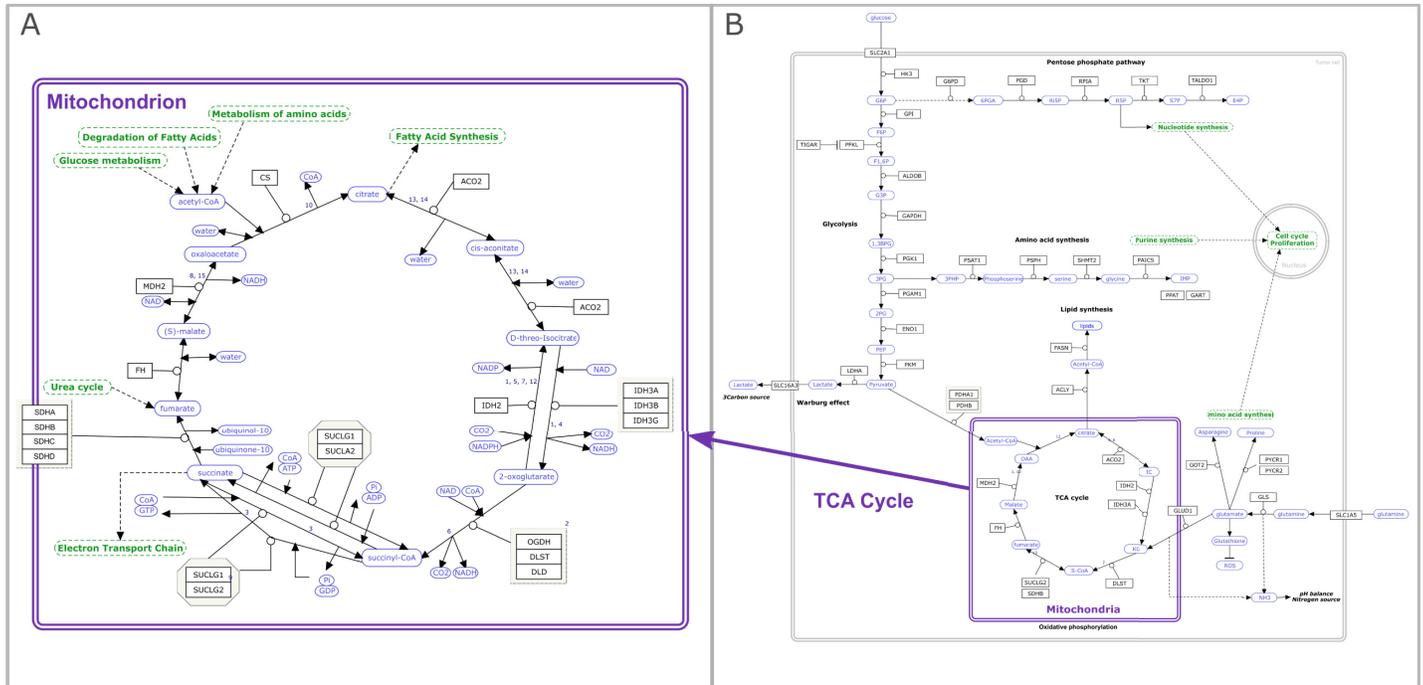


Fig 1. Level of detail/abstraction. The citric acid cycle (TCA cycle or Krebs cycle) represented as the canonical version (panel A, wikipeathways:WP78) [37] and as part of processes involved in metabolic reprogramming in colon cancer (panel B, wikipeathways:WP4290) [38]. In the canonical model, each enzymatic step in the cycle is represented in detail, whereas when the same process is represented in the context of colon cancer, specific steps in the cycle are summarized and abstracted at a higher level. Pathway nodes are depicted in green, metabolites in blue, and genes and proteins in black. TCA, tricarboxylic acid.

<https://doi.org/10.1371/journal.pcbi.1009226.g001>

(Fig 1A). As a comparison, the “metabolic reprogramming in colon cancer” pathway [38] includes a version of the citric acid cycle where individual steps are summarized at a higher level, but where the full pathway includes other relevant metabolic processes (Fig 1B). For metabolic conversions, one option might be to only include the main reaction participants to reduce visual clutter, thereby omitting proton and electron donors or acceptors (usually relatively small molecules) and ignoring stoichiometry. Similarly, in a cancer pathway model where genes in a central signaling cascade are mutated, the central signaling should be illustrated in detail. On the other hand, if ligand binding processes are the most relevant, the downstream signaling events can be condensed and illustrated at a higher level or even using pathway nodes instead of genes and interactions.

Many pathways reference or include (parts of) other pathways. If the inclusion of another pathway is not central to understanding the process, pathways or parts of pathways should be represented by a single pathway node instead of including many peripheral gene products or proteins. If a more detailed model representing part of the pathway is available, we encourage curators to add links to those pathways in their model. The number of nodes included in a model affects its usefulness in downstream analysis and data visualization. For enrichment analysis, pathway size affects the performance and interpretability indicating that meta pathways or other large pathways describing multiple processes should be split into individual smaller pathways for these types of analyses [39]. Smaller models can be merged for topological analysis and other network analysis approaches where larger networks perform well. For visualization, the size depends on the application and purpose of the visualization (providing an overview or showing a specific detail) and other factors impacting interpretability (see Rule 7).

Rule 3: Use standardized naming conventions and identifiers for molecular entities

Understanding biological processes and molecular entities can be a difficult job with the vast amount of synonyms in use. For example, the official gene name of neuroepithelial cell–transforming gene 1 protein (NET1) is also a synonym for the sodium-dependent noradrenaline transporter (SLC6A2). Also, well-known chemicals are commonly referred to by multiple names, for example, paracetamol/acetaminophen, which has well over 500 drug vendor-specific names worldwide [40]. Initiatives such as the HUGO Gene Nomenclature Committee (HGNC) [41] and the Mouse Genome Informatics (MGI) database [42] aim to provide a consistent vocabulary for gene symbols and names in such a way that they are unique and do not use problematic characters, are not autocorrected by tools like Microsoft Excel, are not common words, and are intuitively readable. If existing, species-specific common names and gene symbols should be used. If relevant, the synonym or alternative name can be added as a note or comment to the entity in the pathway model.

Names are often more descriptive and understandable to researchers with knowledge of the context; however, this is not trivial for computers. In pathway models, this problem is in part overcome by annotating molecular entities with both a human-readable label and a computer-readable identifier. These resolvable identifiers connect to online knowledge bases containing information about the individual molecular entities. Many of the pathway curation tools have integrated identifier resolution, and curators only need to define the identifier and database for the entity. Given the large number of online databases, choosing the proper identifier for a specific entity can be difficult. It is important to use the most precise identifier with the biological entity. For example, use UniProt for a specific protein, instead of the gene identifier, or RefSeq [43] if a specific protein sequence is meant. For genes, use Ensembl or NCBI gene identifiers [44], which are regularly updated by genome assembly and annotation projects. For species not covered by these 2 sources or where full genomes are not (completely) known yet, gene identifiers from other assembly databases can be used. For resolvability purposes, we advise to register these databases in identifiers.org [45]. For specific gene products, like microRNAs, identifiers from more targeted databases can be used (e.g., miRBase) [46]. Similarly, use ChEBI or Wikidata [47] for chemical compounds, and for lipids, LIPID MAPS [48] or SwissLipids [49]. Annotating groups of proteins acting together or in parallel can be increasingly difficult when the number of individual proteins is large. The Complex Portal [25] provides identifiers for protein complexes, which also enables the possibility to (automatically) check for the participants in the complex. In contrast, gene products acting in parallel (e.g., isoenzymes and assembly factors) that are not directly part of a complex can often be combined into groups. For protein families, InterPro [50] or Enzyme Commission (EC) codes [51] can be used, but only if no other details can be found in the literature. By explicitly stating if proteins belong to either a complex (indicating each subunit is needed to perform an action) or group (meaning each protein alone can perform the same action), the understanding of a pathway model and the possibility for analysis will improve.

In general, consistency in labeling and annotation with identifiers is essential for users and analysis tools to interpret molecular pathway models. This is in line with the FAIR principles [3], which were established to improve the usefulness of data sources for humans and systems.

Rule 4: Use standardized interaction types

Pathway models need detailed interactions to describe how different biological entities influence each other. Therefore, specifying the meaning of interactions or the relationships between entities constitutes a core task in generating a pathway model. Pathway models use various

line styles (e.g., arrows and T bars) to show which biological entities are interacting and how. Although the machine readability of pathway models is made possible by connecting entities, simply connecting entities is not enough. The lines and arrows in diagrams can have several biological meanings, e.g., metabolic conversion, stimulation, and modification, which should be easily distinguishable from one another. This warrants the necessity of a standardized set of biological interaction types used within a pathway model.

Several established standards describe different types of biological interactions, all with a specific focus, for example, SBGN or Molecular Interaction Maps (MIMs) [52]. We advise curators to select and stick to one of the available drawing standards within a pathway model. Additionally, ontologies like the Systems Biology Ontology (SBO) [53], BioPAX, Biological Expression Language (BEL) [54], or Molecular Interactions Controlled Vocabulary (PSI-MI) [55] can be used to further specify and annotate the interactions. Distinguishing between visual interaction types (different arrowheads) and annotation details is important for users to understand the individual biological meanings at first glance. As an example, drawing standards usually have different styles for stimulation, catalysis, and necessary stimulation (MIM and SBGN), while ontologies like SBO could provide further details (e.g., necessary versus absolute stimulation).

By using a standardized set of interactions, pathway authors can specify the correct biological semantics for the interactions in the pathway model. Detailed annotation (type, evidence, provenance, and identifier) of interactions facilitate their use to integrate pathway models with other resources [56]. Mappings between different standards and vocabularies will then allow harmonization between pathway models from different resources or with different drawing styles to efficiently combine and integrate pathway models in computational analysis such as drug discovery [57]. To facilitate the use of such standardized interaction types, pathway tools often include them predefined in drawing panels. Whenever possible, adding identifiers to interactions further improves the identification for analysis purposes. As an example, the Rhea database provides stoichiometric and balanced reactions for metabolic conversion interactions, which is key for pharmacodynamic and kinetic data modelers. Adding Rhea identifiers to interactions also moves the pathway models a step closer to kinetic modeling.

Finally, if you decide to use nonstandard interaction types, always define their meaning in the pathway model in a visual legend.

Rule 5: Provide literature references, provenance, and evidence for pathway content

Information added to a pathway model, such as proteins or interactions, should be based on scientific evidence or at least a documented hypothesis. The origin of that information, also called provenance, and an evaluation of the evidence should be referenced such that people exploring the pathway model can easily access and verify the source of the information. Curators often add focused references for specific biological entities (e.g., publication about the role of a mutated protein in a pathway) or interactions (e.g., evidence for the inhibition of protein A by protein B). Additionally, many tools allow adding general literature references important for the understanding of the pathway model as a whole. In addition to the reference itself, by using the Evidence and Conclusion Ontology (ECO) [58], one can provide additional information about the certainty of the information, for example, when a pathway or interaction has been inferred from another species.

When provenance is provided in a machine-readable manner, this information can be used to automatically filter pathway content by the presence and type of supporting evidence. Furthermore, the coverage of a pathway model and pathway database can be assessed with more

ease. Finally, linking the pathway model content to other databases and retrieving additional information thereof can be done effortlessly.

Rule 6: Annotate pathway models with a title, a description, and ontology terms

To improve readability and reusability, accurately describing the pathway content on different levels is crucial. First, a descriptive and clear title is important to capture the overarching concept of the pathway model. The title should be concise and limited to a single line when possible. In general, abbreviations should be avoided and only included when the process refers to a specific gene or protein. Second, think of the pathway description as a textual representation of the pathway model, describing each main step as well as the overall process and outcome. While it might seem counterintuitive, writing the description before creating the model helps to make sure that the description is a proper representation of the model in the end. The description should be unique; do not simply copy the description from another source (such as the original caption of a figure). Include literature references within your description, whether they refer to specific components in the model or provide the provenance from which the model was adapted (see also Rule 5). Finally, machine-readable metadata about the pathway diagram improves the findability and organization of the pathway models. Using existing ontologies like the Pathway Ontology [59], the Gene Ontology, the Cell Ontology [60], or the Disease ontology [61] provides the specific context to a pathway model.

Rule 7: Increase readability with graphical annotations and intuitive layout

Illustrating a biological process in a graphical format clearly and legibly is a challenge. The layout of reactions and entities is critical to readability. For most pathway models, a layout with top-to-bottom and left-to-right orientation is recommended. Cellular compartments and other graphical annotations are helpful to further organize the layout and should be labeled for clear identification. In general, labels on nodes describing proteins, gene products, and metabolites should have one consistent font, size, and placement throughout the pathway model. However, using a larger font size helps to highlight and draw attention to nodes of particular importance. To increase readability, the placement of nodes should avoid creating overlapping interactions as much as possible. Additionally, the layout should be optimized to avoid using multiple copies of the same node, unless there is a biological meaning. For example, when illustrating the transport of a protein into a compartment, 2 copies of the same node are necessary to describe the process properly. Similarly, illustrating 2 different complexes composed of an overlapping set of proteins requires multiple copies of the same node, as does illustrating multiple reactions involving the same common reactants, for example, ATP/ADP.

[Fig 2](#), illustrating nicotinamide adenine dinucleotide (NAD⁺) metabolism as described at Wiki-Pathways [62], exemplifies these recommendations. The pathway is organized from top to bottom, starting with the entry of precursors (tryptophan, nicotinic acid, nicotinamide, nicotinamide riboside, and nicotinamide ribotide) at the top, using 2 copies of the relevant nodes to illustrate transport. Within the cell, processes are localized to either the cytoplasm, mitochondria, or nucleus.

Rule 8: Consider data visualization when using colors in your pathway model

Various types of data, like transcriptomics, proteomics, metabolomics, and fluxomics data, can be visualized on pathway models. Whereas data linked to molecular entities

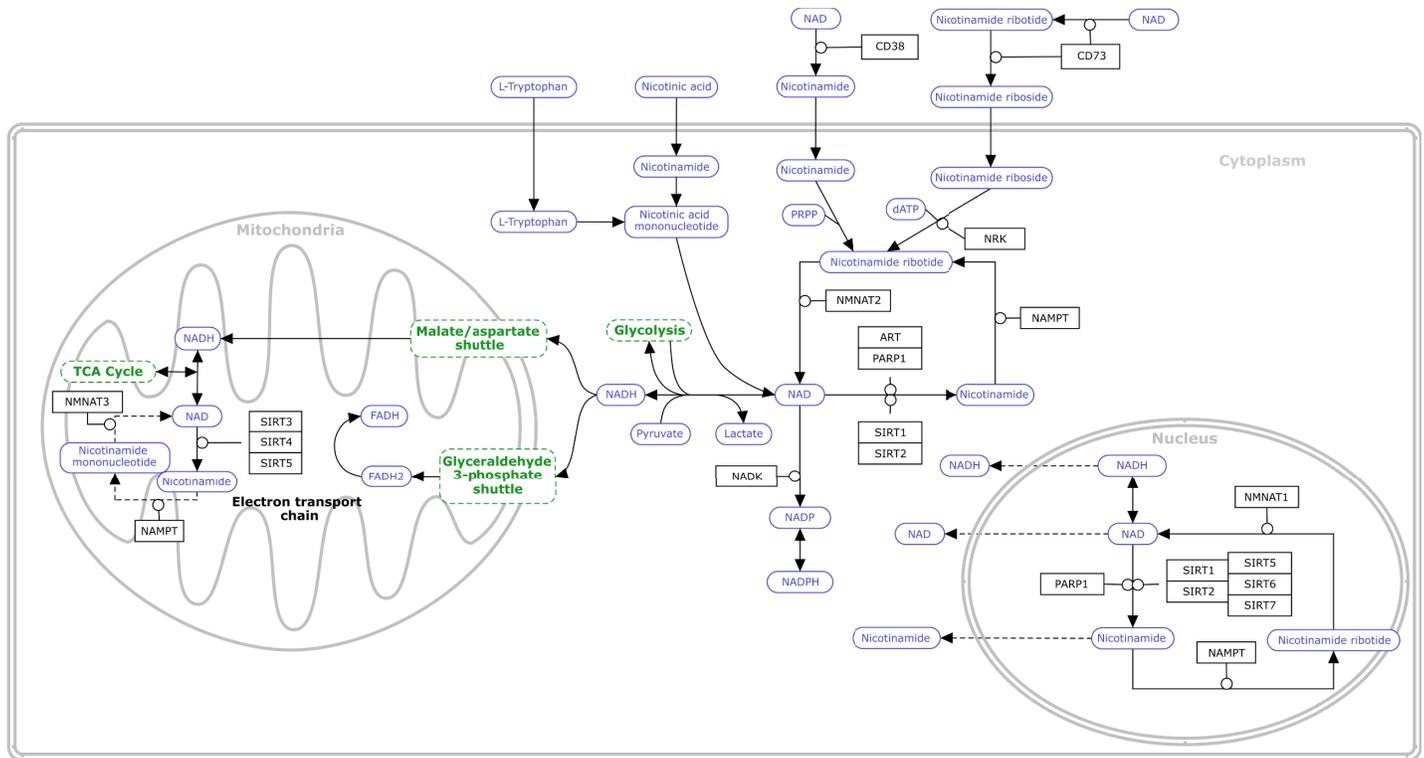


Fig 2. Use of graphical annotations in the NAD⁺ metabolism pathway at WikiPathways (wikipathways:WP3644) [62]. Since one of the most common uses of pathway models is data visualization on nodes, nodes must be adequately sized to allow for efficient data visualization. Accounting for this, node size can then be optimized to the overall size and complexity of the pathway model, utilizing a smaller node size for models with many nodes and a larger node size for sparse models. A consistent representation of different molecular types in terms of size and shape is important, for example, using a specific shape for gene products/proteins and another for metabolites. Any nonstandard nodes or entities should be defined in a legend in the model, as well as in the textual description. NAD⁺, nicotinamide adenine dinucleotide.

<https://doi.org/10.1371/journal.pcbi.1009226.g002>

(nodes) are often depicted by using a color scheme to fill the nodes, data related to interaction types are often depicted by changing the width or color of the interaction linking the data nodes.

To avoid interfering with such data visualization, the fill color for nodes and edges should be limited when designing your pathway, as they can interfere with data visualization and lead to misinterpretation at the process level. We recommended differentiating molecular entities using node shapes to designate state information (e.g., posttranslational modifications) using dedicated glyphs and to depict different interaction types using standard arrowheads (see Rule 4), instead of using colors. If you decide to use colors in your pathway, consider readability for all users by choosing a visual accessible color palette.

In Fig 3A, node fill color was used to describe gene mutations of interest to a disease. When data are visualized as node color (Fig 3B), the fill color related to mutations is overwritten, and the information is lost. Fig 3C illustrates a better solution to illustrating genes with mutations by adding a small graphic to relevant nodes, which is preserved when data are visualized on the node, as shown in Fig 3D.

Finally, when using different colors and line width is unavoidable, include a legend to define the color and line annotations. A legend can assist in differentiating between the data and the pathway entities themselves when visualizing data on the pathway model.

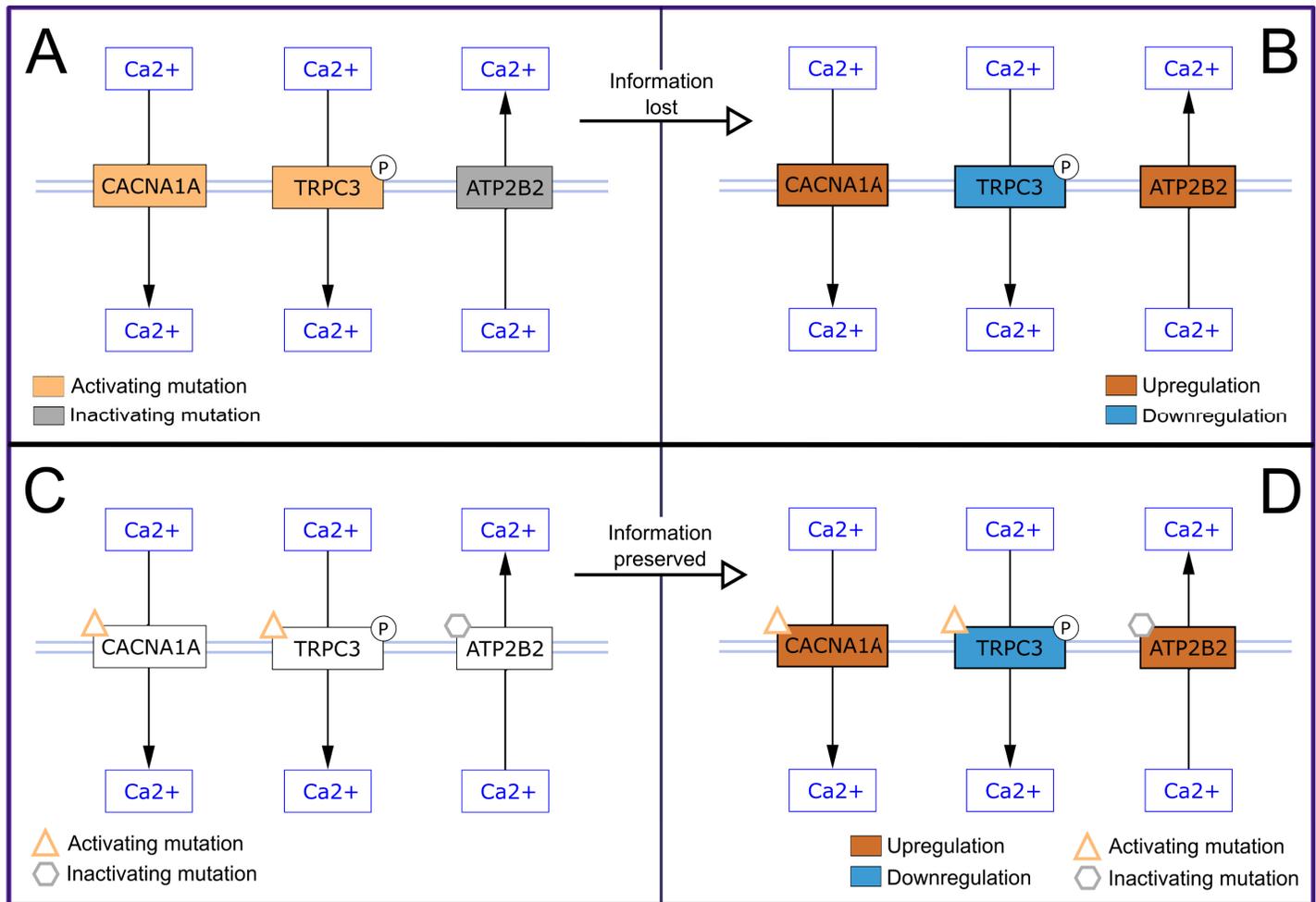


Fig 3. Three visualizations of part of the “PKC-gamma calcium signaling pathway in ataxia” at WikiPathways (wikipathways:WP4760) [63]. Visualizations of part of the “PKC-gamma calcium signaling pathway in ataxia.” (A) Node fill color is used to describe genes with mutations relevant to ataxia; orange indicates an activating mutation, and gray indicates an inactivating mutation. (B) When experimental data are visualized as node fill color, the mutation information is lost. (C) Mutations are shown as an added graphic to nodes; an orange triangle indicates an activating mutation, and a gray hexagon indicates an inactivating mutation. (D) Data visualization does not interfere with the mutation information, and both data types are visualized. Pathways were visualized in Cytoscape [64]. PKC, protein kinase C.

<https://doi.org/10.1371/journal.pcbi.1009226.g003>

Rule 9: Communicate and disseminate your pathway model widely

Once your model is ready, you want people to reuse and cite it. With the above rules, you already ensured the easy reuse of the model. But full dissemination still starts with communicating to others that you have created a new pathway model. When your model is being developed as part of a research article, particularly if featured as a figure therein, include at least the machine-readable model as a supporting information. This simplifies the reuse of the knowledge by the readers of your article. You can even take this a step further and make your model available in an open pathway knowledge base, for example, WikiPathways or NDEX, or more general archives like figshare [65] or Zenodo [66]. This also provides a URL, allowing other researchers to find and cite your model easier.

Once you are confident about the content, you can further communicate your work via social media (e.g., LinkedIn [67] and Twitter [68,69]), or, if the model is of wider interest, also in Wikipedia (www.wikipedia.org). You may also be able to include it on your website as an image linking to the machine-readable version or as a compact identifier in text, such as

wikipathways:WP78 for WikiPathways WP78 [70]. For pathway models at WikiPathways, an interactive view of the model can be embedded in a web page [71].

As an extension to the World Wide Web, the semantic web is a viable platform to share novel pathway content on. Various pathway resources, for example, WikiPathways, disseminate content in the Resource Description Framework (RDF), which allows rapid reuse of the content in other research workflows that use the semantic web [72]. Network representations of the pathways can be deposited into NDEX, making it accessible directly from network analysis tools such as Cytoscape [64]. The pathway content (gene lists) can also be distributed to multiple enrichment analysis tools, for example, MSigDb. Adding your model to a community-curated resource like WikiPathways automatically ensures broad dissemination of your pathway model in multiple formats and also allows for early review from peers and automated version control during the development of the model.

Rule 10: Maintain your pathway model as an evolving resource

Pathway models are never finished and should be considered living models, both in terms of content and graphical aspects, as seen in Fig 4. When research reveals new findings or biological insights, these should be added to the pathway model if it fits into the scope of the pathway model as originally intended. Just like research, the development, curation, and maintenance of

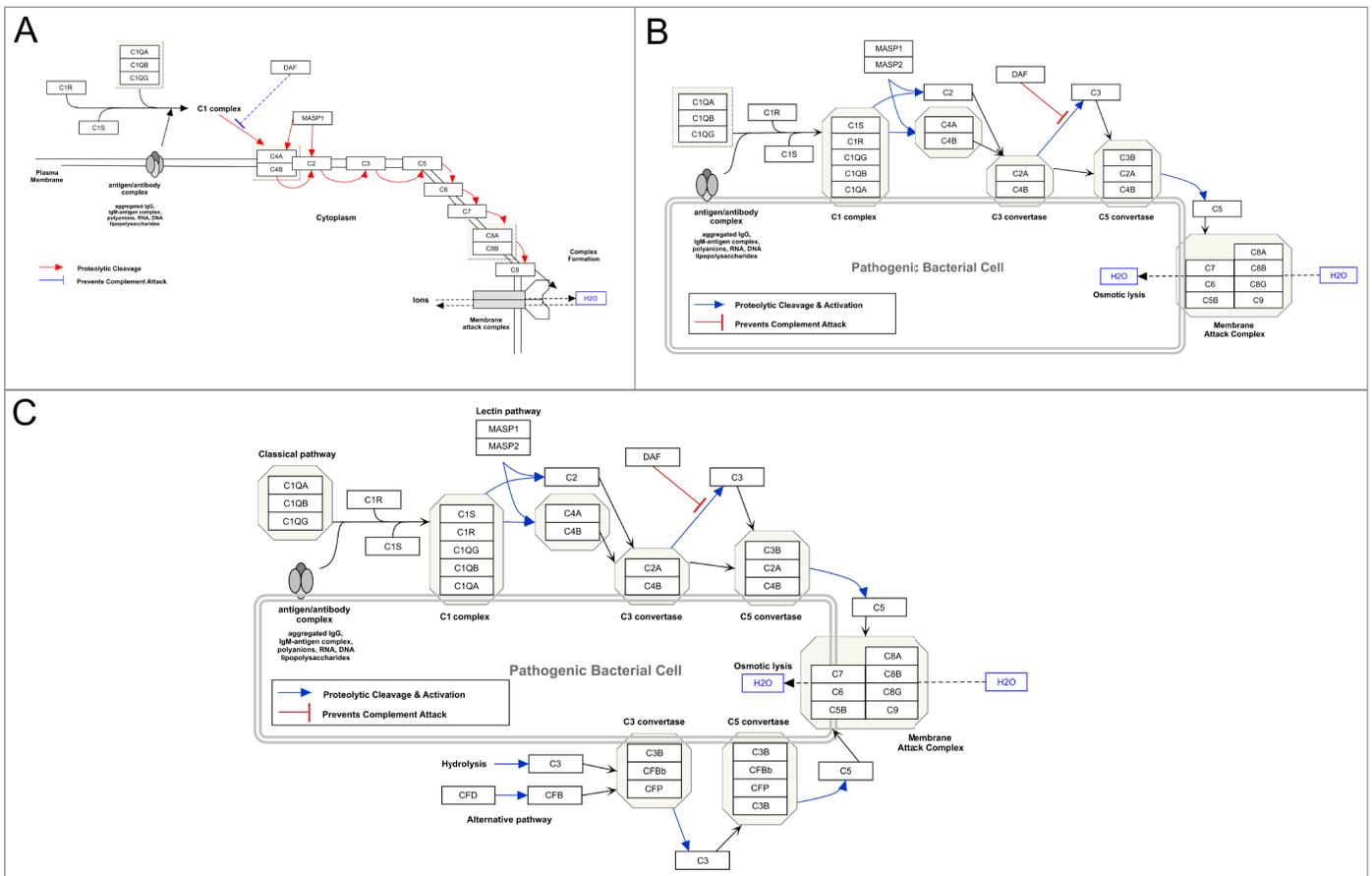


Fig 4. The development of the complement activation pathway model at 3 time points. (A) Revision number 63169 as of May 8, 2013 [73]. **(B)** Revision number 82136 as of September 9, 2015 [74]. **(C)** Revision number 106816 as of September 17, 2019 [75].

<https://doi.org/10.1371/journal.pcbi.1009226.g004>

pathway models are community efforts that bring various perspectives and information sources together.

Updates are aimed at improving the graphical layout and content and should be described by the editor in accompanying comments. An example of a pathway timeline is illustrated in Fig 4, showing 3 versions of the complement activation pathway [75] on WikiPathways: (A) The first revision from 2013 is simple, limited to the core molecular entities and interactions and a basic layout. (B) In 2015, the pathway model has been extended in terms of molecular content and their interactions (addition of MASP1/2, details of the complement proteins, etc.), and the layout was updated with the addition of a cell shape, increasing readability. (C) After continuous improvement, the latest revision includes the addition of the alternate pathway as well as the definition of the 2 other subpathways, classical and lectin.

Another reason to keep your pathway updated is the identifier annotation for genes, proteins, metabolites, and other pathway elements (see Rule 3). Databases like Ensembl, ChEBI, and even WikiPathways phase out identifiers in data curation processes, and like broken website links, replacing phased out identifiers in your pathway ensures that links to external databases and general interoperability are preserved over time.

Conclusions

Pathway diagrams are intuitive tools for collecting and communicating molecular details of biological processes. Every year, thousands of pathway diagrams are published in the literature often as static images, which slow down the distribution of the knowledge, limits reusability, and prevents computational analysis [18].

Pathway diagrams, in general, should be made available, ideally in an online pathway database, as models created with appropriate pathway modeling tools and standards. By following the 10 simple rules described in this paper, pathway models maximize potential for computational analysis and reuse.

Acknowledgments

We would like to thank all WikiPathways contributors for their work creating pathway models, which has helped greatly in the development of these rules.

References

1. Kutmon M, Evelo CT, Coort SL. A network biology workflow to study transcriptomics data of the diabetic liver. *BMC Genomics*. 2014; 15(1):971. <https://doi.org/10.1186/1471-2164-15-971> PMID: 25399255
2. Benson HE, Watterson S, Sharman JL, Mpamhanga CP, Parton A, Southan C, et al. Is systems pharmacology ready to impact upon therapy development? A study on the cholesterol biosynthesis pathway. *Br J Pharmacol*. 2017; 174(23):4362–82. <https://doi.org/10.1111/bph.14037> PMID: 28910500
3. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016; 3(1). <https://doi.org/10.1038/sdata.2016.18> PMID: 26978244
4. Rougier NP, Droettboom M, Bourne PE. Ten Simple Rules for Better Figures. *PLoS Comput Biol*. 2014; 10(9):e1003833. <https://doi.org/10.1371/journal.pcbi.1003833> PMID: 25210732
5. Marai GE, Pinaud B, Bühler K, Lex A, Morris JH. Ten simple rules to create biological network figures for communication. *PLoS Comput Biol*. 2019; 15(9):e1007244. <https://doi.org/10.1371/journal.pcbi.1007244> PMID: 31557157
6. Touré V, Novère NL, Waltemath D, Wolkenhauer O. Quick tips for creating effective and impactful biological pathways using the Systems Biology Graphical Notation. *PLoS Comput Biol*. 2018; 14(2): e1005740. <https://doi.org/10.1371/journal.pcbi.1005740> PMID: 29447151

7. Byrnes RW, Cotter D, Maer A, Li J, Nadeau D, Subramaniam S. An editor for pathway drawing and data visualization in the Biopathways Workbench. *BMC Syst Biol.* 2009; 3(1):99. <https://doi.org/10.1186/1752-0509-3-99> PMID: 19799790
8. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, et al. The BioPAX community standard for pathway data sharing. *Nat Biotechnol.* 2010; 28(9):935–42. <https://doi.org/10.1038/nbt.1666> PMID: 20829833
9. Kutmon M, van Iersel MP, Bohler A, Kelder T, Nunes N, Pico AR, et al. PathVisio 3: An Extendable Pathway Analysis Toolbox. *PLoS Comput Biol.* 2015; 11(2):e1004085. <https://doi.org/10.1371/journal.pcbi.1004085> PMID: 25706687
10. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* 2019. <https://doi.org/10.1093/nar/gkz1031> PMID: 31691815
11. Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, et al. WikiPathways: connecting communities. *Nucleic Acids Res.* 2020. <https://doi.org/10.1093/nar/gkaa1024> PMID: 33211851
12. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief Bioinform.* 2017; 20(4):1085–93. <https://doi.org/10.1093/bib/bbx085> PMID: 29447345
13. Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000; 28(1):27–30. <https://doi.org/10.1093/nar/28.1.27> PMID: 10592173
14. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics Knowledge for Personalized Medicine. *Clin Pharmacol Ther.* 2012; 92(4):414–7. <https://doi.org/10.1038/clpt.2012.96> PMID: 22992668
15. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* 2012; 2(5):401–4. <https://doi.org/10.1158/2159-8290.CD-12-0095> PMID: 22588877
16. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 2018; 47(D1):D419–26. <https://doi.org/10.1093/nar/gky1038> PMID: 30407594
17. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.* 2012; 41(D1):D793–800. <https://doi.org/10.1093/nar/gks1055> PMID: 23143270
18. Hanspers K, Riutta A, Summer-Kutmon M, Pico AR. Pathway information extracted from 25 years of pathway figures. *Genome Biol.* 2020; 21(1). <https://doi.org/10.1186/s13059-020-02181-2> PMID: 33168034
19. Rodchenkov I, Babur O, Luna A, Aksoy BA, Wong JV, Fong D, et al. Pathway Commons 2019 Update: integration, analysis and exploration of pathway data. *Nucleic Acids Res.* 2019. <https://doi.org/10.1093/nar/gkz946> PMID: 31647099
20. Stobbe MD, Houten SM, Jansen GA, van Kampen AH, Moerland PD. Critical assessment of human metabolic pathway databases: a stepping stone for future integration. *BMC Syst Biol.* 2011; 5(1):165. <https://doi.org/10.1186/1752-0509-5-165> PMID: 21999653
21. Malik-Sheriff RS, Glont M, Nguyen TVN, Tiwari K, Roberts MG, Xavier A, et al. BioModels—15 years of sharing computational models in life science. *Nucleic Acids Res.* 2019. <https://doi.org/10.1093/nar/gkz1055> PMID: 31701150
22. Pratt D, Chen J, Welker D, Rivas R, Pillich R, Rynkov V, et al. NDEX, the Network Data Exchange. *Cell Syst.* 2015; 1(4):302–5. <https://doi.org/10.1016/j.cels.2015.10.001> PMID: 26594663
23. Morgat A, Axelsen KB, Lombardot T, Alcántara R, Aimo L, Zerara M, et al. Updates in Rhea—a manually curated resource of biochemical reactions. *Nucleic Acids Res.* 2014; 43(D1):D459–64. <https://doi.org/10.1093/nar/gku961> PMID: 25332395
24. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 2013; 42(D1):D358–63. <https://doi.org/10.1093/nar/gkt1115> PMID: 24234451
25. Meldal BHM, Bye-A-Jee H, Gajdoš L, Hammerová Z, Horáčková A, Melicher F, et al. Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes. *Nucleic Acids Res.* 2018; 47(D1):D550–8. <https://doi.org/10.1093/nar/gky1001> PMID: 30357405
26. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2018; 47(D1):D607–13. <https://doi.org/10.1093/nar/gky1131> PMID: 30476243
27. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 2021; 49(D1):D325–34. <https://doi.org/10.1093/nar/gkaa1113> PMID: 33290552

28. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102(43):15545–50. <https://doi.org/10.1073/pnas.0506580102> PMID: 16199517
29. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. *Nucleic Acids Res*. 2019. <https://doi.org/10.1093/nar/gkz966> PMID: 31691826
30. Consortium UniProt. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2018; 47(D1): D506–15. <https://doi.org/10.1093/nar/gky1049> PMID: 30395287
31. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res*. 2015; 44(D1):D1214–9. <https://doi.org/10.1093/nar/gkv1031> PMID: 26467479
32. Kitano H, Funahashi A, Matsuoka Y, Oda K. Using process diagrams for the graphical representation of biological networks. *Nat Biotechnol*. 2005; 23(8):961–6. <https://doi.org/10.1038/nbt1111> PMID: 16082367
33. Sari M, Bahceci I, Dogrusoz U, Sumer SO, Aksoy BA, Babur Ö, et al. SBGNviz: A Tool for Visualization and Complexity Management of SBGN Process Description Maps. *PLoS ONE*. 2015; 10(6):e0128985. <https://doi.org/10.1371/journal.pone.0128985> PMID: 26030594
34. Czauderna T, Klukas C, Schreiber F. Editing, validating and translating of SBGN maps. *Bioinformatics*. 2010; 26(18):2340–1. <https://doi.org/10.1093/bioinformatics/btq407> PMID: 20628075
35. ySBGN. Github. Available from: <https://github.com/sbgn/ySBGN>
36. Hoksza D, Gawron P, Ostaszewski M, Hasenauer J, Schneider R. Closing the gap between formats for storing layout information in systems biology. *Brief Bioinform*. 2019; 21(4):1249–60. <https://doi.org/10.1093/bib/bbz067> PMID: 31273380
37. Dahlquist K, Stobbe M, Pico A, Hanspers K, van Iersel M, Kelder T, et al. TCA Cycle (aka Krebs or citric acid cycle) (Homo sapiens). Available from: https://www.wikipathways.org/instance/WP78_r113981
38. Hanspers K, Willighagen E, Pico A. Metabolic reprogramming in colon cancer (Homo sapiens). Available from: https://www.wikipathways.org/instance/WP4290_r113958
39. Karp PD, Midford PE, Caspi R, Khodursky A. Pathway size matters: the influence of pathway granularity on over-representation (enrichment analysis) statistics. *BMC Genomics*. 2021; 22(1). <https://doi.org/10.1186/s12864-021-07502-8> PMID: 33726670
40. acetaminophen. [cited 2020 Oct 7]. Available from: <https://www.wikidata.org/wiki/Q57055>
41. Bruford EA, Braschi B, Denny P, Jones TEM, Seal RL, Tweedie S. Guidelines for human gene nomenclature. *Nat Genet*. 2020; 52(8):754–8. <https://doi.org/10.1038/s41588-020-0669-3> PMID: 32747822
42. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, Anagnostopoulos A, et al. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*. 2018; 47(D1):D801–6. <https://doi.org/10.1093/nar/gky1056> PMID: 30407599
43. O’Leary NA, Wright MW, Brister JR, Ciufu S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2015; 44(D1):D733–45. <https://doi.org/10.1093/nar/gkv1189> PMID: 26553804
44. Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, Ermolaeva O, et al. Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res*. 2014; 43(D1):D36–42. <https://doi.org/10.1093/nar/gku1055> PMID: 25355515
45. Bernal-Llinares M, Ferrer-Gómez J, Juty N, Goble C, Wimalaratne SM, Hermjakob H. Identifiers.org: Compact Identifier services in the cloud. *Bioinformatics*. 2020. <https://doi.org/10.1093/bioinformatics/btaa864> PMID: 33031499
46. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res*. 2018; 47(D1):D155–62. <https://doi.org/10.1093/nar/gky1141> PMID: 30423142
47. Waagmeester A, Stupp G, Burgstaller-Muehlbacher S, Good BM, Griffith M, Griffith OL, et al. Wikidata as a knowledge graph for the life sciences. *Elife*. 2020; 9:e52614. <https://doi.org/10.7554/eLife.52614> PMID: 32180547
48. Fahy E, Sud M, Cotter D, Subramaniam S. LIPID MAPS online tools for lipid research. *Nucleic Acids Res*. 2007; 35(Web Server):W606–12. <https://doi.org/10.1093/nar/gkm324> PMID: 17584797
49. Aimo L, Liechti R, Hyka-Nouspikel N, Niknejad A, Gleizes A, Götz L, et al. The SwissLipids knowledge-base for lipid biology. *Bioinformatics*. 2015; 31(17):2860–6. <https://doi.org/10.1093/bioinformatics/btv285> PMID: 25943471
50. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*. 2018; 47(D1): D351–60. <https://doi.org/10.1093/nar/gky1100> PMID: 30398656

51. Boyce S, Tipton KF. Enzyme Classification and Nomenclature. In eLS, editor. Chichester, UK: John Wiley & Sons, Ltd; 2001. p. 1–11.
52. Luna A, Karac EI, Sunshine M, Chang L, Nussinov R, Aladjem MI, et al. A formal MIM specification and tools for the common exchange of MIM diagrams: an XML-Based format, an API, and a validation method. *BMC Bioinformatics*. 2011; 12(1):167. <https://doi.org/10.1186/1471-2105-12-167> PMID: 21586134
53. Courtot M, Juty N, Knüpfer C, Waltemath D, Zhukova A, Dräger A, et al. Controlled vocabularies and semantics in systems biology. *Mol Syst Biol*. 2011; 7(1):543. <https://doi.org/10.1038/msb.2011.77> PMID: 22027554
54. Madan S, Szostak J, Elayavilli RK, Tsai RTH, Ali M, Qian L, et al. The extraction of complex relationships and their conversion to biological expression language (BEL) overview of the BioCreative VI (2017) BEL track. *Database (Oxford)*. 2019; 2019. <https://doi.org/10.1093/database/baz084> PMID: 31603193
55. Sivade M, Alonso-López D, Ammari M, Bradley G, Campbell NH, Ceol A, et al. Encompassing new use cases—level 3.0 of the HUPO-PSI format for molecular interactions. *BMC Bioinformatics*. 2018; 19(1). <https://doi.org/10.1186/s12859-018-2118-1> PMID: 29642841
56. Touré V, Verduyck S, Acencio ML, Lovering RC, Orchard S, Bradley G, et al. The Minimum Information about a Molecular Interaction CAusal STatement (MI2CAST). *Bioinformatics*. 2020. <https://doi.org/10.1093/bioinformatics/btaa622> PMID: 32637990
57. Miller RA, Woollard P, Willighagen EL, Digles D, Kutmon M, Loizou A, et al. Explicit interaction information from WikiPathways in RDF facilitates drug discovery in the Open PHACTS Discovery Platform. *F1000Res*. 2018; 7:75. <https://doi.org/10.12688/f1000research.13197.2> PMID: 30416713
58. Giglio M, Tauber R, Nadendla S, Munro J, Olley D, Ball S, et al. ECO, the Evidence & Conclusion Ontology: community standard for evidence information. *Nucleic Acids Res*. 2019; 47(D1):D1186–94. <https://doi.org/10.1093/nar/gky1036> PMID: 30407590
59. Petri V, Jayaraman P, Tutaj M, Hayman G, Smith JR, Pons JD, et al. The pathway ontology—updates and applications. *J Biomed Semantics*. 2014; 5(1):7. <https://doi.org/10.1186/2041-1480-5-7> PMID: 24499703
60. Meehan TF, Masci AM, Abdulla A, Cowell LG, Blake JA, Mungall CJ, et al. Logical Development of the Cell Ontology. *BMC Bioinformatics*. 2011; 12(6). <https://doi.org/10.1186/1471-2105-12-6> PMID: 21208450
61. Schriml LM, Arze C, Nadendla S, Chang YWW, Mazaitis M, Felix V, et al. Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res*. 2011; 40(D1):D940–6. <https://doi.org/10.1093/nar/gkr972> PMID: 22080554
62. Hanspers K, Pico A, Mélius J, Kutmon M. NAD+ metabolism (Homo sapiens). Available from: https://www.wikipathways.org/instance/WP3644_r113960
63. Hanspers K, Willighagen E. PKC-gamma calcium signaling pathway in ataxia (Homo sapiens). Available from: https://www.wikipathways.org/instance/WP4760_r108400
64. Shannon P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*. 2003; 13(11):2498–504. <https://doi.org/10.1101/gr.1239303> PMID: 14597658
65. figshare. Available from: <https://figshare.com/>
66. Zenodo. Available from: <https://zenodo.org/>
67. LinkedIn. [cited 2020 Oct 23]. Available from: <https://www.linkedin.com/>
68. Ekins S, Perlstein EO. Ten Simple Rules of Live Tweeting at Scientific Conferences. *PLoS Comput Biol*. 2014; 10(8):e1003789. <https://doi.org/10.1371/journal.pcbi.1003789> PMID: 25144683
69. Cheplygina V, Hermans F, Albers C, Bielczyk N, Smeets I. Ten simple rules for getting started on Twitter as a scientist. *PLoS Comput Biol*. 2020; 16(2):e1007513. <https://doi.org/10.1371/journal.pcbi.1007513> PMID: 32040507
70. Wimalaratne SM, Juty N, Kunze J, Janée G, McMurry JA, Beard N, et al. Uniform resolution of compact identifiers for biomedical data. *Sci Data*. 2018; 5(1):1–8. <https://doi.org/10.1038/s41597-018-0002-5> PMID: 30482902
71. PathwayWidget. [cited 2020 Oct 7]. Available from: <https://www.wikipathways.org/index.php/PathwayWidget>
72. Waagmeester A, Kutmon M, Riutta A, Miller R, Willighagen EL, Evelo CT, et al. Using the Semantic Web for Rapid Integration of WikiPathways with Other Biological Online Data Resources. *PLoS Comput Biol*. 2016; 12(6):e1004989. <https://doi.org/10.1371/journal.pcbi.1004989> PMID: 27336457
73. Salomonis N, Hanspers K, Adriaens M. Complement Activation (Homo sapiens). Available from: https://www.wikipathways.org/instance/WP545_r63169

74. Salomonis N, Pico A, Hanspers K, Kutmon M, Adriaens M, Chichester C. Complement Activation (Homo sapiens). Available from: https://www.wikipathways.org/instance/WP545_r82136
75. Salomonis N, Pico A, Hanspers K, Kutmon M, Adriaens M, Chichester C, Willighagen E. Complement Activation (Homo sapiens). Available from: https://www.wikipathways.org/instance/WP545_r106816