

Instructions for WikiPathways Authors

Unleashing Pathway Knowledge

Life is a collection of pathways. And pathway diagrams are our best descriptions of the processes that compose healthy and disease conditions. Unfortunately, the vast majority of pathways are trapped in scientific articles as static figures that are inaccessible to the public and to computational methods of analysis. One of the goals of WikiPathways is to unleash the wealth of pathway knowledge currently trapped in countless pathway figures by transforming them into *pathway models*. By the process described here, you can contribute to this transformation!

Overview

These instructions will guide you through the process of pathway authoring, starting from a pathway figure in a scientific article and ending with an annotated pathway model that can be accessed and used for future learning and research. Some of these steps might not seem important on the surface, but are actually *critical* to building a model that can be integrated into downstream research methods. In brief:

1. Setup your tools and resources.
2. Identify *and assess* target pathway figure.
3. Start a new pathway model.
4. Add *and define* gene products and metabolites.
5. Layout *and connect* reactions and interactions.
6. Grouping and graphical annotations.
7. Add references to scientific articles.
8. Add a description, ontology terms and quality tags.
9. Check, verify and share your work!

1. Tools and Resources

The main editor for WikiPathways is the **PathVisio** tool. To get started, follow these steps:

- [Download and install PathVisio](#)
- Install the WikiPathways plugin for PathVisio by going to “Plugins > Plugin manager” and selecting the “WikiPathways plugin”. Review the [WikiPathways Plugin Tutorial](#).
- Download [relevant mapping databases](#). For example, if you are planning to create a pathway relevant in human, download the human mapping database.

- Also download the Metabolites database from the Metabolite database section.

2. Target Pathway Figures

Searching [PubMed Central \(PMC\)](#) for topics of interest can often lead to pathway figures. You can also add terms like “pathway” to your search. Here are some characteristics of a good target pathway:

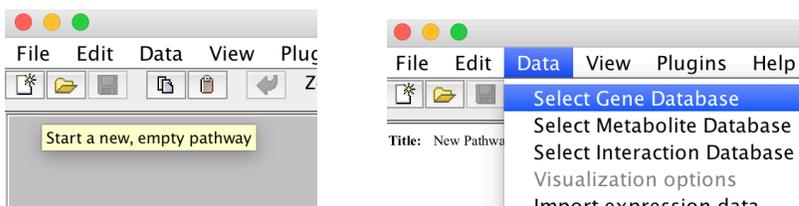
- Represents a biological process of interest to you. You will inevitably learn some biology while authoring the pathway model, so you might as well pick topics you are interested in!
- There are at least 5-10 genes products or metabolites in the pathway. Models with fewer than 5 datanodes will be of limited use.
- The relationships and interactions between the datanodes are clear, e.g., activation, inhibition, binding, catalysis, etc. With the associated caption and text for the figure, you should be able to follow the logic of the interactions.
- Identify any articles referenced in the paper that might be relevant, and even *Google* the gene, metabolites processes described to make sure the biology is clear and to provide additional references for the bibliography of the pathway you are authoring.

Pro tip: Try writing the description of the pathway *in your own words* before you even start. That way you can see right away which parts might not be clear.

3. Start a New Pathway

In the PathVisio Application:

Launch PathVisio on your laptop and click the top-left icon to start a new pathway. Next you will want to select Gene and Metabolite databases from the Data menu. But first you will need to download [the latest databases from pathvisio.org](#). Make sure you select the correct species database. Note that the separate databases for metabolites and interactions work across all species. After selecting both a Gene and Metabolite database from the Data menu, you are ready to go!



At WikiPathways:

After you completed your draft of the pathway in PathVisio, you can either use the [WikiPathways plugin](#) to “Upload New” or go to the [WikiPathways create page](#) and

select “Upload file”. Here is where you will specify the species and title. This is also when you can choose to make the pathway private.



Create new pathway

Select to either use the pathway editor or upload a gpml file:

Use Editor

Upload File

No file selected.

Make pathway private ([how does this work?](#))

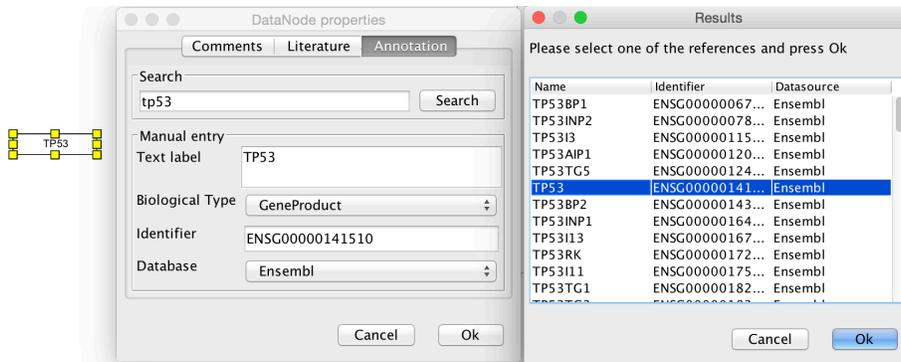
Pro tip: To quickly find a new pathway uploaded via the WikiPathways plugin in PathVisio, use the “New Pathways” link in the left-hand sidebar at WikiPathways.

Pro tip: Choose private pathway when the content is sensitive, e.g., not yet published or public information. Otherwise, it is sufficient to simply tag the pathway as “Under Construction” (see section 7) in order to communicate its state and to automatically exclude it from batch downloads and research workflows, etc.

4. Adding Datanodes

Datanodes are the most important parts of a pathway model. *Every* downstream research method relies on gene products and metabolites being accurately annotated as a datanode. There are three main factors to properly annotating a datanode:

1. **Type:** You should most often use *Gene Product* and *Metabolite* datanode types. ONLY use Protein, RNA or DNA when you specifically intend to model that type of molecule and only that type of molecule and are prepared to use the appropriate datasource (e.g., Proteins should be annotated with UniProt and not with Ensembl or Entrez Gene identifiers). If unsure, simply use *Gene Product* to mean multiple possible forms of the molecule. Note: Pathway is another useful datanode type and you can annotate it with a WikiPathways pathway identifier, like WP1234.
2. **Xref (datasource + identifier):** Rely on the databases underlying WikiPathways and PathVisio to find the correct xref (*external reference*) for a datanode. Search by its name and choose the correct hit. This is the most important annotation of a datanode. Selecting from the search results will *automatically* fill in all the annotation fields accurately, so you don’t have to do it manually!



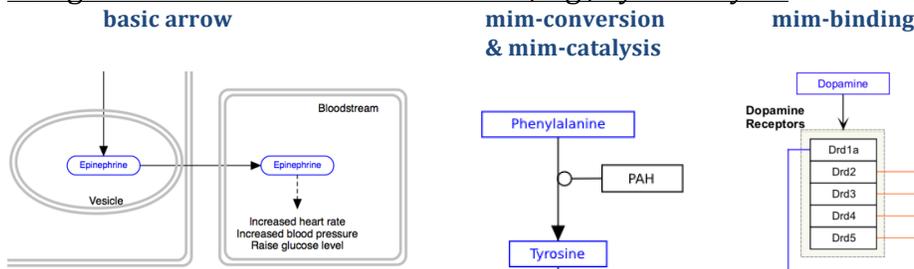
3. **Text Label:** You can override the label provided by the database without affecting the xref. This label is just for display purposes and sometimes in the context of a particular pathway, you might want to show a different name or follow a different convention, while still referring to the same, universally identified molecule.

Pro tip: Sometimes the search results won't be definitive. Don't hesitate to *Google* the gene name. You will often get hits for Wikipedia articles that are actively being curated by a large community of scientists as part of the Gene Wiki project. A project that actively collaborates with WikiPathways!

5. Adding Interactions

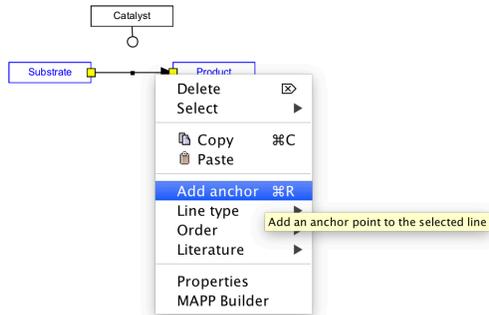
Interactions are easily the second most important part of pathway model. The use of pathways in network analysis absolutely depends on accurate interaction modeling. There are two main factors to consider:

1. **Type:** By default, you should use *Basic Arrow*. Choosing a more specific interaction type should only be done with careful consideration of the meaning of the interaction and type of entities interaction. Commonly used specific interactions include: mim-inhibition, mim-catalysis and mim-binding. Note: mim-conversion should ONLY be used when a metabolite is being converted to another metabolite, e.g., by an enzyme.



2. **anchors:** Then ends of interactions should always be attached to an anchor. Datanodes have anchors indicated by target icons when you drag an interaction end over them. However, you must add anchors to interactions if you intend to have another interaction connect to it. It is NOT enough to just

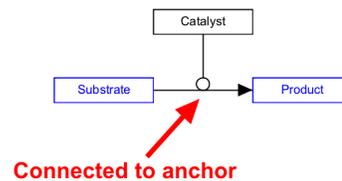
have it be visually close. The act of connecting an interaction end to an anchor, triggers meaningful changes in the underlying GPML model that you are building.



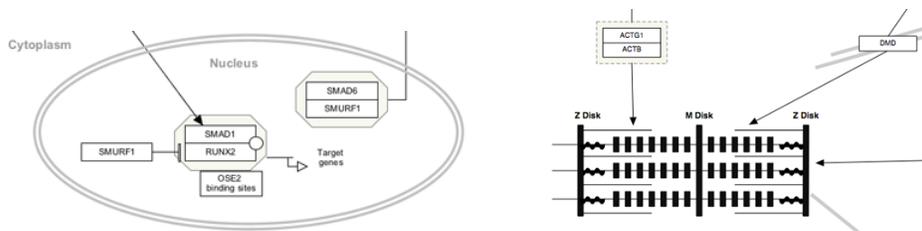
Pro tip: Use the reaction template object when adding catalysis events. With one click you can add 3 nodes, a conversion and an anchored catalysis interaction!

6. Grouping and Graphical Annotations

It's not all simple datanodes and interactions, unfortunately. Very often you will need to represent a collection of datanodes as a grouped entity, for example, a protein complex. Select two or more datanodes, right-click and choose *Create Complex*. You can also arbitrarily group items to help with layout management.



WikiPathways also has support for pre-made graphics to help represent cellular compartments, like the mitochondria and nucleus, as well as arbitrary graphics to draw whatever the pathway author needs to in order to provide sufficient context for human interpretation. See [this described in the guidelines](#).



7. Adding References

A bibliography of references is important for a couple reasons. First, it gives appropriate credit to the effort of those whose work you are building upon. And, second, it is the foundation of trust upon which future authors rely when considering whether to use and cite your work. *Would you trust a pathway or article*

with zero references? WikiPathways tries to make it as easy as possible to add references. You can click anywhere on the background and add *Literature for pathway*. Or you can click on individual data nodes or interactions and add references for specific parts of pathways. This helps researchers examine the references for the parts of the pathway they care about. Enter a Pubmed ID and the reference tool will do the rest of the work.

8. Describe Your Pathway

There are pencil “edit” icons on each pathway page to allow you to edit the title, description and any quality tags already added.

Pathway Titles

Try to be descriptive and specific, but also concise. It is often helpful to [browse current pathway titles](#) to make sure you distinguish your pathway appropriately.

Pathway Descriptions

It is important to avoid plagiarism here. You must use your own words to describe your newly authored pathway model. The best descriptions *walk the reader through the key paths* in the model. Also describe the overall point of pathway, e.g., what type of research it is based on or most relevant to. Feel free to include links here to web pages that are relevant as either source material or that provide more in depth descriptions of their own.

Quality tags

These are unique to WikiPathways and help communicate to other what the status and goals of a pathway are. E.g., you should always add an “Under construction” tag when you first create or upload a pathway. Some tags will be added by WikiPathways bots and administrators, as well as by other curators. Your goal is to get your pathway tagged with “Approved version” (which means it’s ready for researchers to use) and “Featured version” (which means it represents high-quality content and will be periodically featured on the front page). And, don’t forget to remove the “Under construction” tag when you are ready to share your pathway.

9. Share Your Pathway

Remember: you are not alone! The whole point of WikiPathways is to *collaboratively* curate the pathways of biology as a *community*. No individual, nor single team, can be expected to accurately model *all of biology*. So, double-check your work, then invite others to review, correct, and augment your work. Finally, be sure to share your work more broadly so that it can be most effectively used in future research. Here are some ideas:

Checklist to review your own work:

1. Did you indicate the correct species?

2. Expand list of Datanodes at bottom of page. Is it as accurate and complete as you can manage? Click on the linkouts to verify any identities that are you unsure of. Are there any gaps you can still try to fill in?
3. Did you use the correct interaction types? Are they properly anchored to datanodes or other interactions?
4. Did you add all the references you used in making the pathway model? There has to be at least one reference in the bibliography. Ideally, more than one.
5. Did you provide a descriptive and unique title? An adequate description in your own words? Some ontology terms? Proper quality tags?

Inviting others to review your work:

- Other pathway curators you know.
- Post a message on [wikipathways-discuss](#) asking for curation help.
- Contact the authors of the articles you cite.

Share your work:

- Post a message on [wikipathways-discuss](#) announcing the finished version of the pathway. If it passes [our curation criteria](#), we will tag it for release. If it is especially high quality, we will tag it as *featured* and share it on the WikiPathways [twitter](#) account and [blog](#).
- Use social media. You can use a [URL shortener](#) to include links directly to your pathway.

Glossary

datanode – a node or entity on a pathway that is annotated by a database entry. E.g., a Gene Product annotated with an Ensembl identifier.

datasource – third-party sources for the databases underlying WikiPathways. E.g., Ensembl, Entrez Gene and UniProt.

gene product – a datanode type commonly used to represent the protein or transcript products of a gene.

gpml – Graphical Pathway Meta Language, the native XML file format for WikiPathways.

identifier – the universally recognized label for a gene product, metabolite or interaction provided and curated by a datasource.

interaction – the relationship between two entities. E.g., activation or inhibition of a datanode by another datanode.

metabolite – chemicals, drugs and small molecules that are active in a biological process.

mim – Molecular Interaction Maps, a project that defined the arrowhead glyphs for a set of interactions.

model – here used in the sense of a *pathway model* to refer to a pathway diagram that follows a set of database, syntactic and semantic conventions to allow consistent and programmatic parsing and integration with other data types.

xref - External Reference, a datasource and identifier pair that together uniquely define an entity in any context, universally.