The Planarian Anatomy Ontology (PLANA):

A resource to connect data within and across experimental platforms

Stephanie Nowotarski
Post Doc Sánchez Alvarado Lab
Stowers Institute

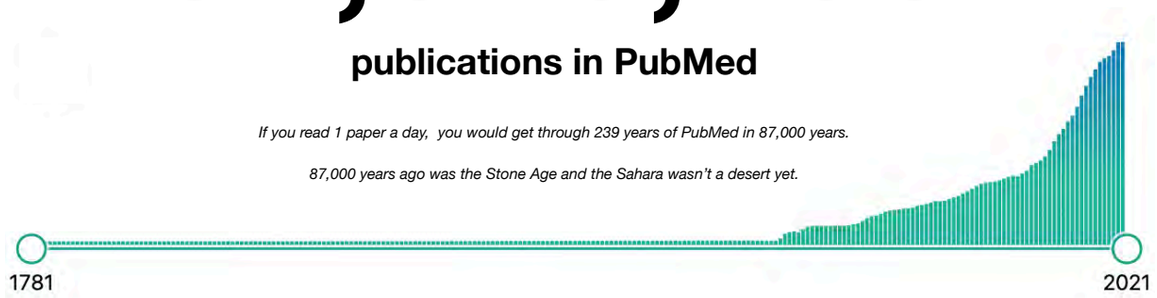
Sofia Robb
Data Scientist
Stowers Institute

31,826,165

publications in PubMed

If you read 1 paper a day, you would get through 239 years of PubMed in 87,000 years.

87,000 years ago was the Stone Age and the Sahara wasn't a desert yet.



pubmed.ncbi.nlm.nih.gov - search "all[sb]"

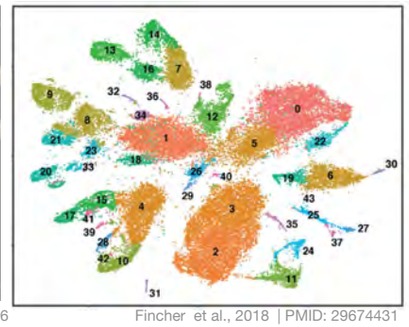
Last updated 12/1/2020 2:05pm central

Knowledge is growing exponentially. Our time is not. Publications area reflection of data…

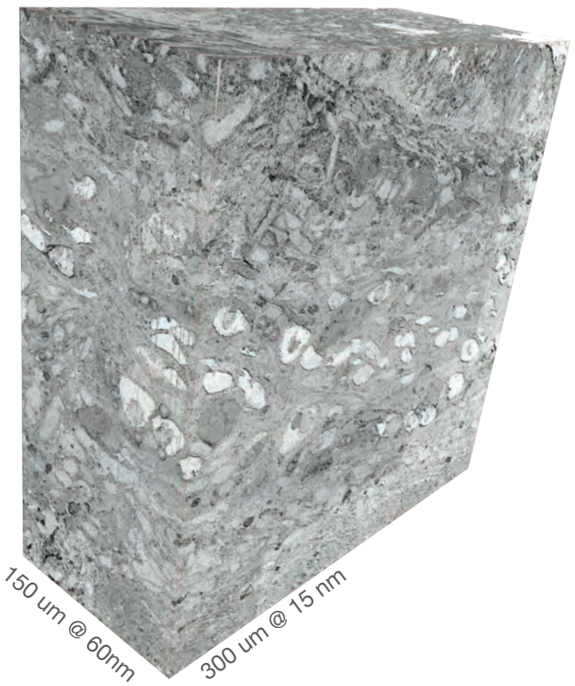
Data is growing too.

How do we annotate high volume anatomical data to use and repurpose?

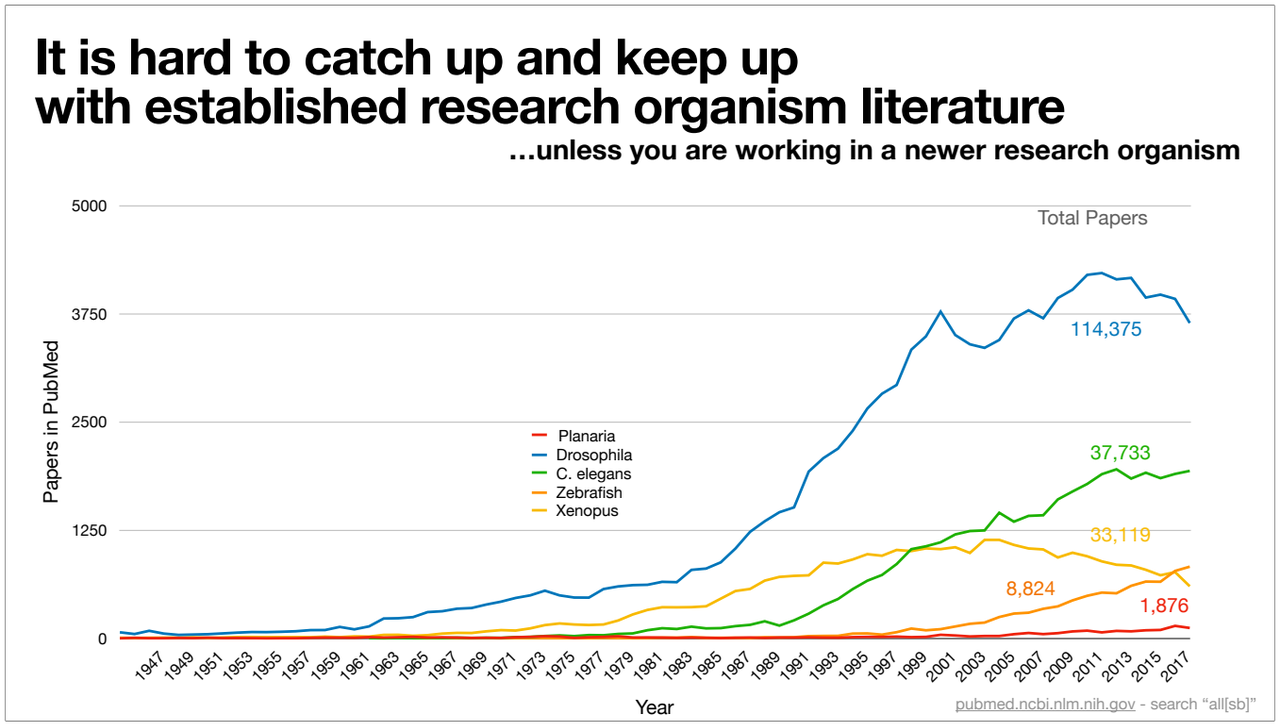
Reddien et al., 2005 |PMID:15866156



Fincher et al., 2018 | PMID: 29674431



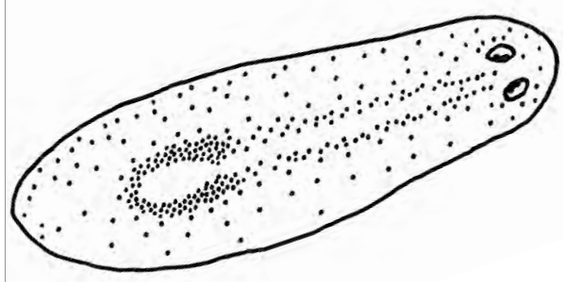
…and data is expanding too. From large scale screens, to single cell data, to high resolution 3D imaging techniques, it’s everywhere. How to we handle this, and do so in a way where data can cross talk more easily between papers and organization structures?



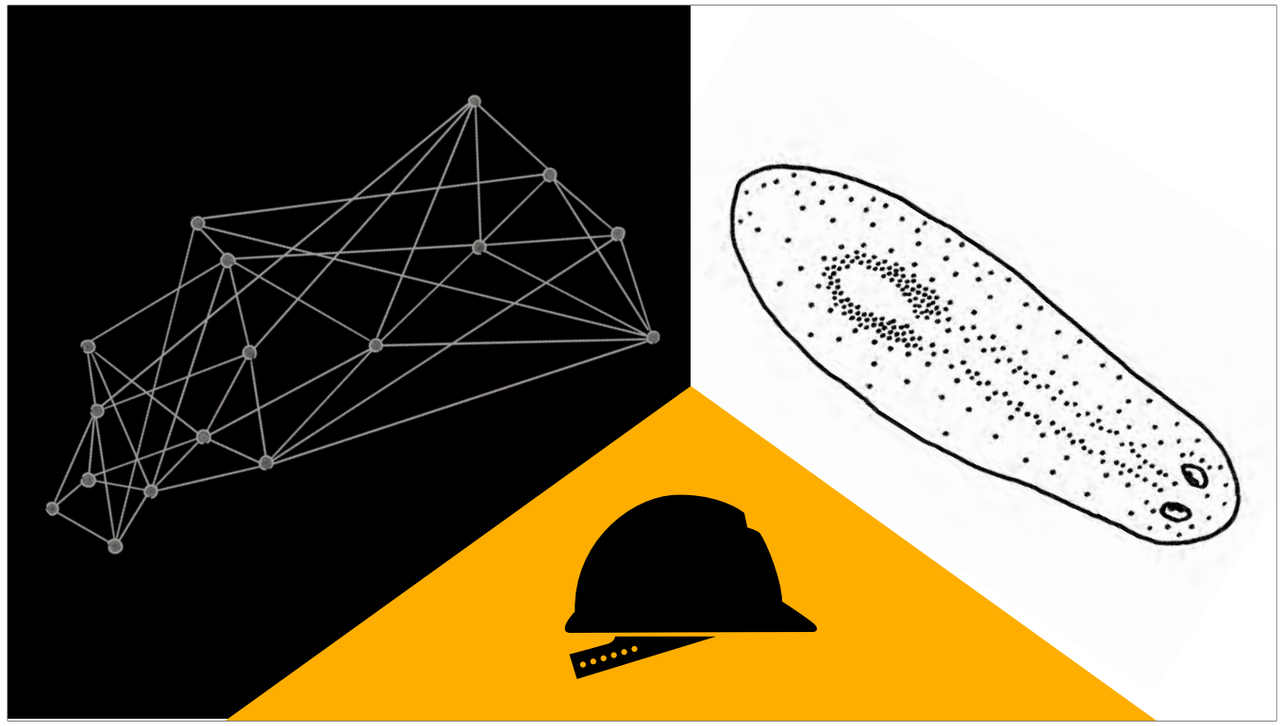
If we head back to the papers data, it's immediately evident as someone who was a grad student working on drosophila, the amount of information I had to sift through was HUGE. It's hard to catch up and keep up. The advantage of working on a newer research organism is that the paper pool is much smaller, which means you are poised to organize from the get go.

But a down side of working on a new model organism is that all the big tools the “model” organisms have… you don’t. So they need to be built by folks trying to get a lot of tools up and running quickly while answering important biological questions. Plus these tools need to be developed open sourced (cause budgets!) And transparently (cause participation and buy-in).

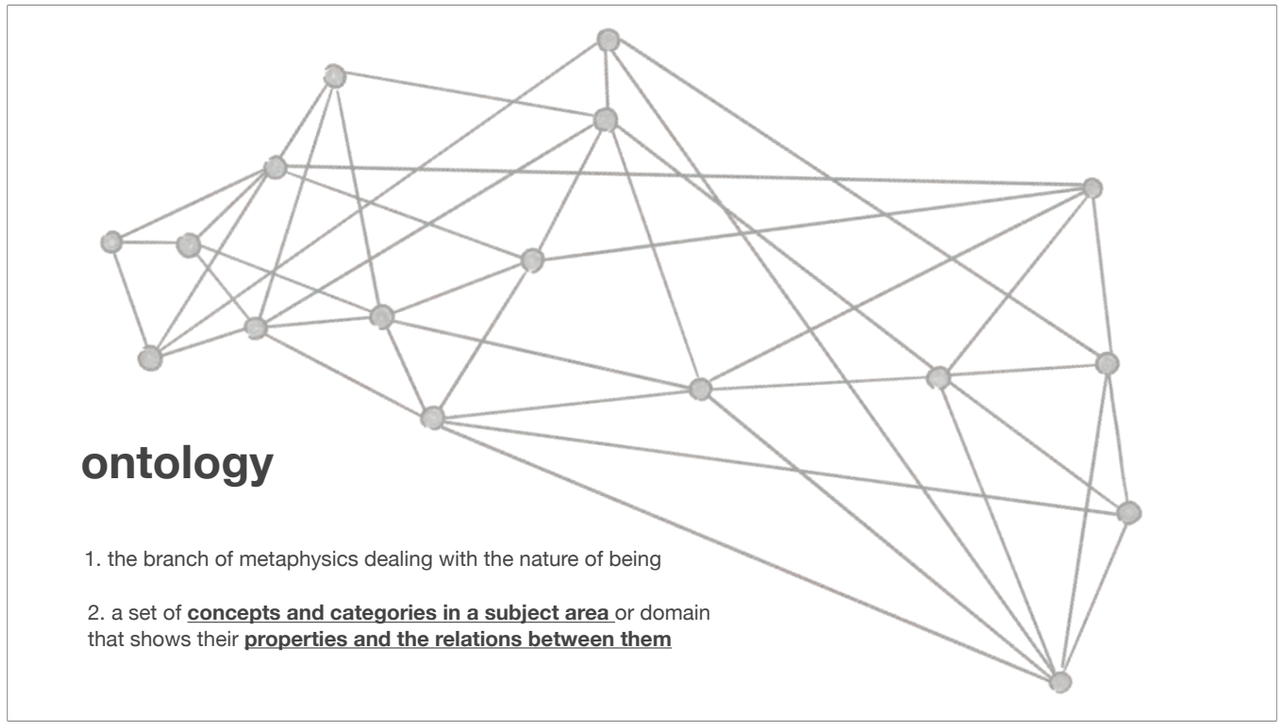
Planarian Anatomy



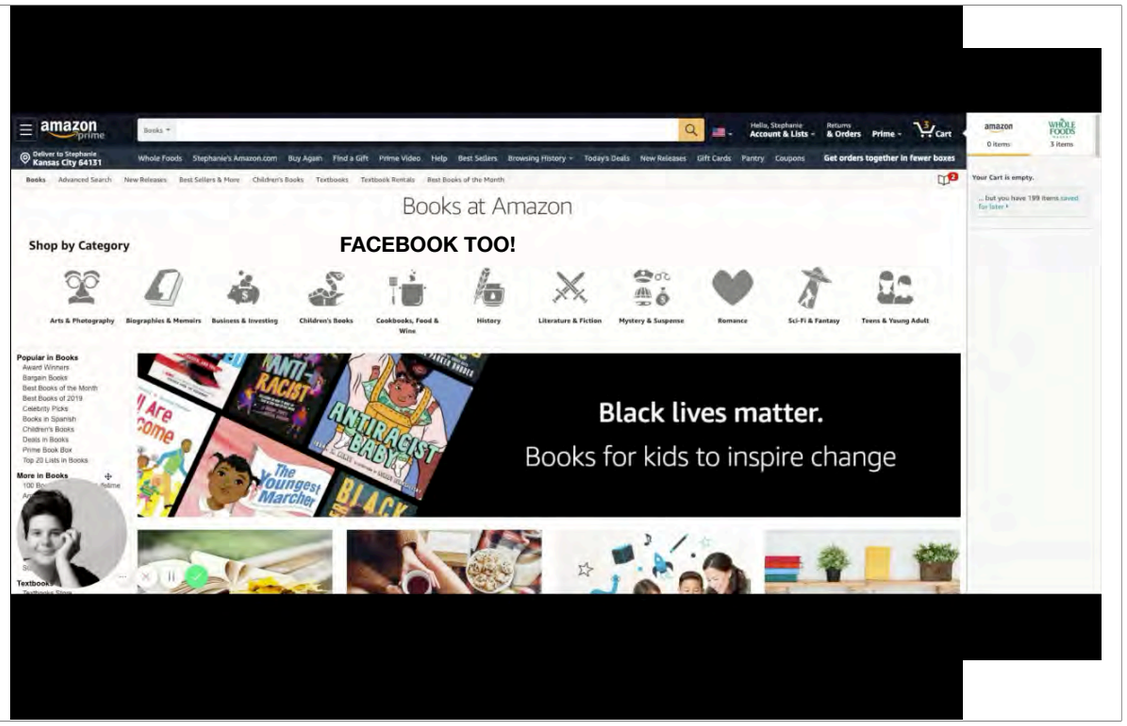
I’m going to talk about how we built one of these data tools for S.mediterranea, an anatomy ontology.



I’m going to spend the first part of the talk talking about what the heck is an ontology. Give y’all some background as a cell and developmental biologist who in the past really only thought about microscopes and images, with things you need to know to sound half way competent. This will mean some jargon :). Then I’ll move on to how we are using our ontology to structure data for S.med. And then Sofia is going to talk about the nuts and bolts about how we built ours/ how we are advising others to build theirs.



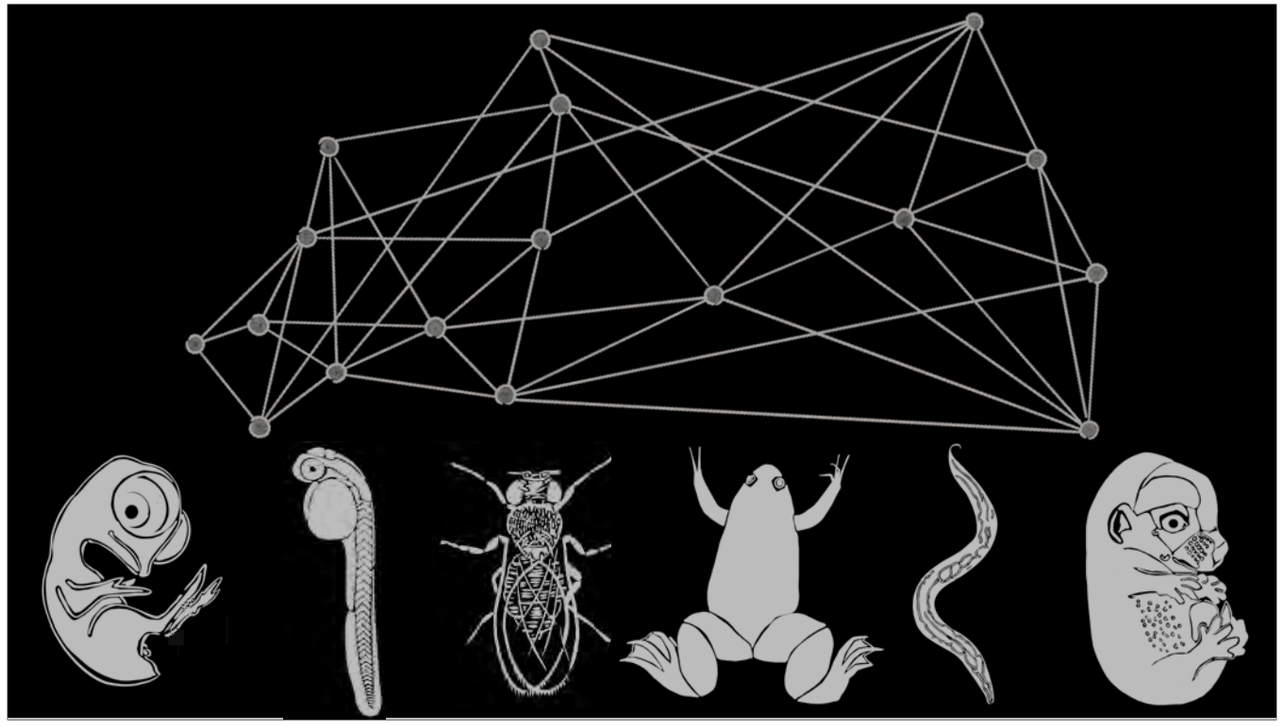
When I first started talking to Sofia about organizing imaging anatomical imaging data as a new post doc she emphatically said “Oh, we need to build and ontology!” I nodded my head in agreement and immediately went to look up what the heck one was. When I saw the first definition I thought I do not have the right docotorate of philosophy for this. Luckily there was a second definition and that’s the kind we are talking about today. It’s just a set of concepts and a relationship between them.



Luckily sorting through massive amounts of data by different people with different goals has been worked out quite well by online retail. Let's start where amazon started out in the very beginning: books. When you are on amazon you are looking at some serious categorial logistics that are fairly straightforward off the mark.

Types. You can have a book be in sci-fi and mystery and also be a bargain book. The semantic web uses ontologies to organize information

Remember IS A (that's going to come up later!)

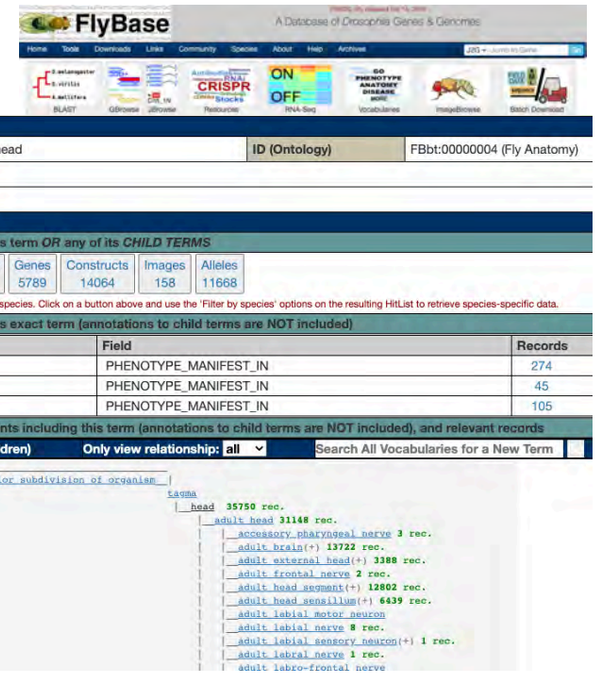


Okay so ontologies organize information for online really well. What are ontologies used for for the “model” organisms?

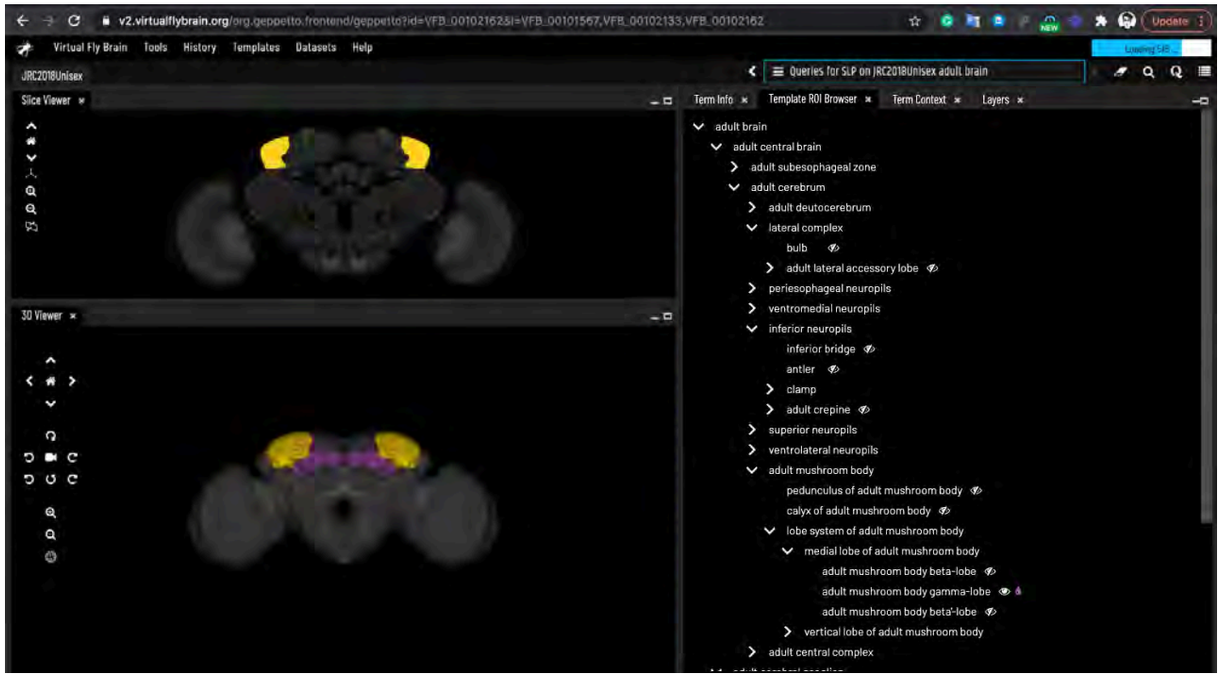
What are canonical model organisms using their anatomy ontologies for?

Or browse the following hierarchy structures:

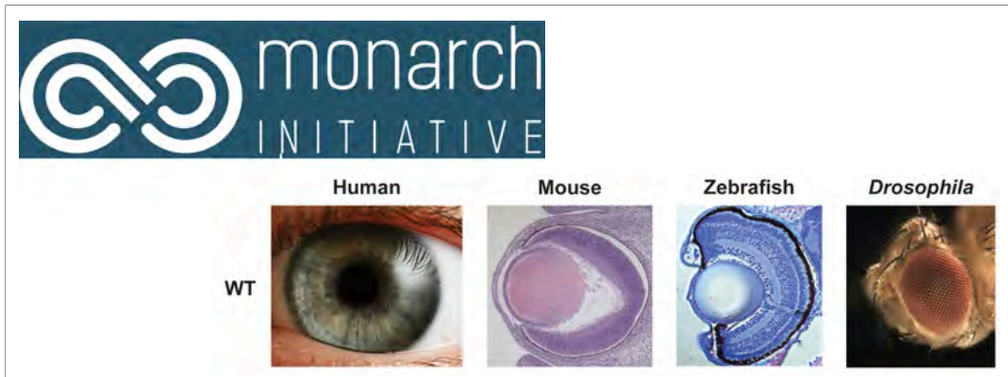
- (●) Anatomy (FBbt)
- ( ) Biological Process (GO)
- ( ) Cellular Component (GO)
- ( ) Molecular Function (GO)
- ( ) Human Disease (DO)
- ( ) Molecular Interaction (MI)
- ( ) Development (FBdv)
- ( ) Allele Class (FBcv)
- ( ) Phenotypic Class (FBcv)
- ( ) Origin of mutation (FBcv)
- ( ) Stock descriptor (FBsv)
- ( ) Publication descriptor (FBcv)
- ( ) Imaging method (FBbi)
- ( ) Sequence ontology (SO)
- ( ) Gene Class (SO)
- ( ) Chromosome Structure Variation (SO)
- ( ) Sequence Variant (SO)



Drosophila’s fly base uses multiple ontologies (they call it vocabularies). The most prevalent one is FBbt, or the anatomy ontology. If you head to the vocabulary page (top header button), you can see all the ones they use. If you click down further and check out their anatomy ontology, you can see pages specifically devoted to individual terms (we’ll call these classes later). This organizes information about phenotypes, alleles, etc.



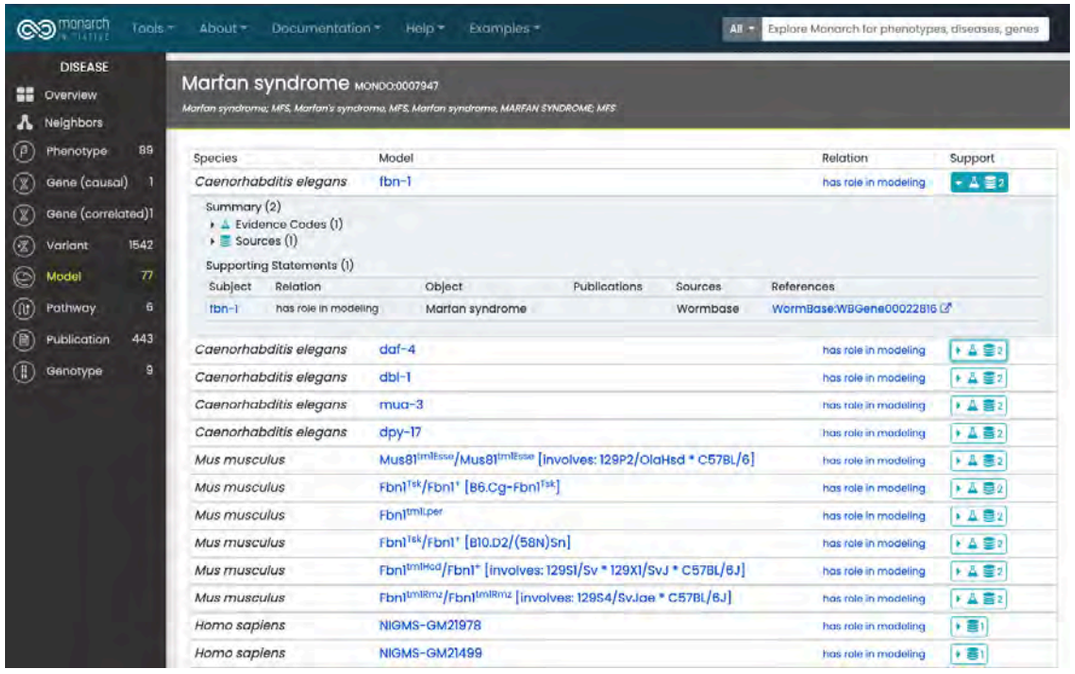
When you look at Drosophila's anatomy ontology a very large portion of it is neuronal terms which is no coincidence. Ontologies are built and change according to a fields knowledge and needs and this one reflects the community's use and efforts towards connectome mapping. A good example of this is Virtual Fly brain which used anatomy ontology terms to organize imagery.



Anatomy ontologies are a foundation for a phenotype ontology (you have define correct structures before you can define when structures go wrong).The Monarch Initiative is an integrative data platform connecting phenotypes to genotypes across species. It bridges basic research and applied research.

We have created or currently contribute to many essential bio-ontologies that together enable sophisticated and semantically integrated computational analysis across gene, genotype, variant, disease, and phenotype data. We have developed algorithms and tools that are in use by multiple communities for tasks including the identification of animal models of human disease through phenotypic similarity, phenotype-driven computational support for differential diagnostics, and translational research.

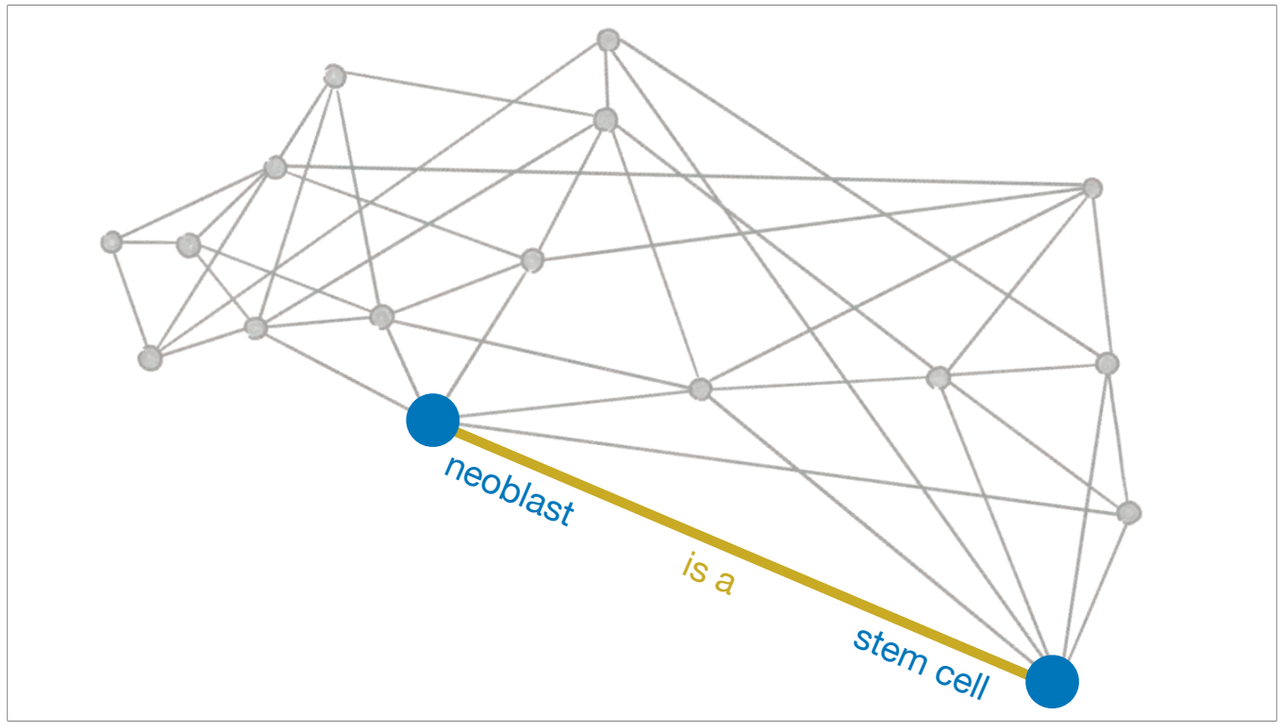
LEFT (1) Human coding genes with disease/phenotype mutations.
RIGHT (2) Model organism orthologs of human coding genes with associated phenotypes
CENTER (3) Combined this makes ~82 percent of human coding genes and this is using 33 sources of data.



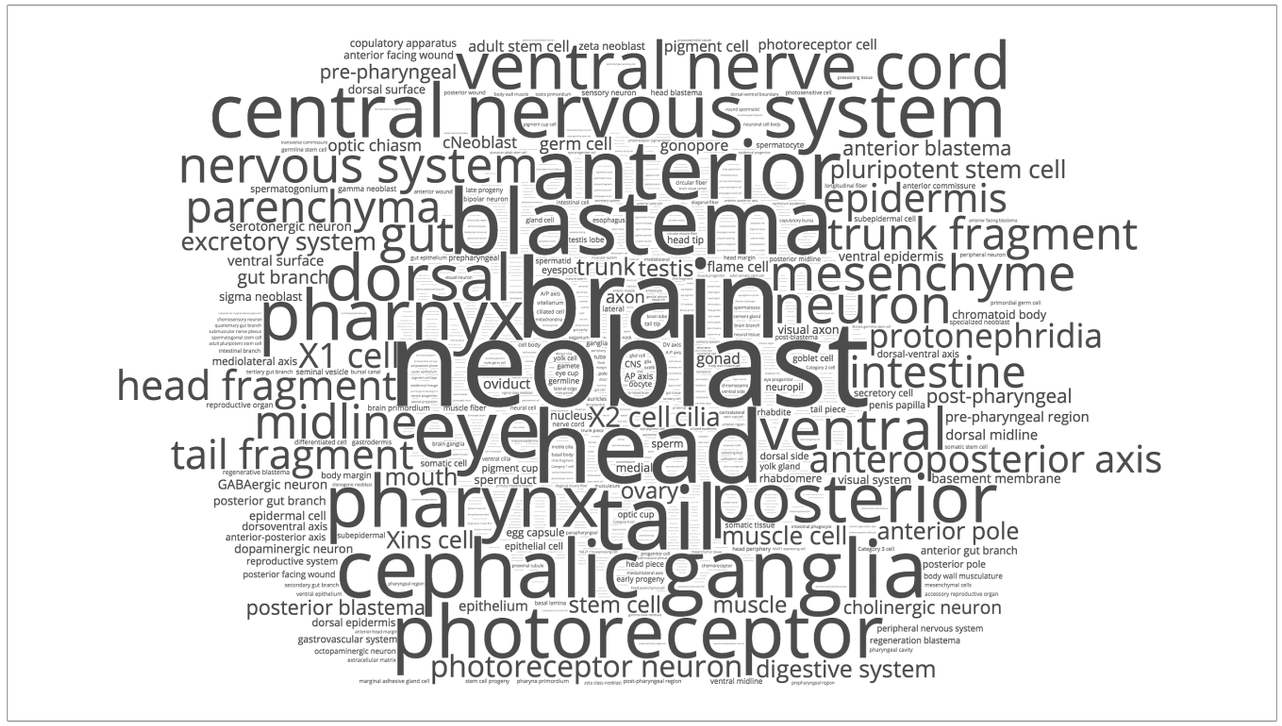
Monarch's browser looks like this. You can see integration of information from model organism in the "Model" tab. In this case we can see C. Elegans, mouse and human related genes and mutations that have a role in modeling Marfan's syndrome. Each of these links back to the source- so for example the C. Elegans fbn-1 entry links back to Wormbase. Clicking through to the Phenotype tab, you'd see things like "flat cornea" as a phenotype, which is based on the anatomical term Cornea. The important points here are that multiple organism ontologies help power aggregators that allow comparative work.

?

Up next : What is an ontology?



Ontologies are a relational network where nodes in the graph are called CLASSES (blue dots). Classes are connected by relationships. It’s like a sentence you can read directionally as THIS is RELATED to THAT. A good example of this for Schmidtea mediterranea is a ‘neoblast’ is a ‘stem cell’. So the first thing you need building an ontology are the nodes or classes. How did we do that?



We did not have to come up with them out of thin air. Anatomical words are use in publications, so we surveyed the literature between 2005-2019, ~200 papers and recorded every instance of an anatomical term used in them. Not all of these ended up as classes, some ended up as synonyms, some were more appropriate to be used in writing definitions for classes (both are still searchable in the ontology).

Classes are annotated with metadata

Once we decided what the name or label of each class would be, then each of those classes got annotated with metadata. Those with heavy lines around them are REQUIRED while the others are optional. 1) LABEL(required) : This is literally the term and is a human-readable description of the concept. 2) CLASS ID (required) : A computer-readable ID, 3)SYNONYMS: any words that could be a synonym, 4) DEFINITION (required): A definition for the concept with a 5)DEFINITION DATABASE CROSS REFERECE (def_dbxref, required): This is the publication that defined the term. 5) COMMENT : an optional field to house any notes or clarification of use that does not belong in definition. 6)DEPICTED BY: representative imagery of the structure with a 7) DEPICTED BY COMMENT: which tells users what they are looking at. The final annotation we have in our ontology is 8) DATABASE CROSS REFERENCE (dbxref): In our case this holds all the ids of the papers that used the term from our literature survey and it holds the IDs of related or homologous classes in other anatomy ontologies (this increases interoperability)

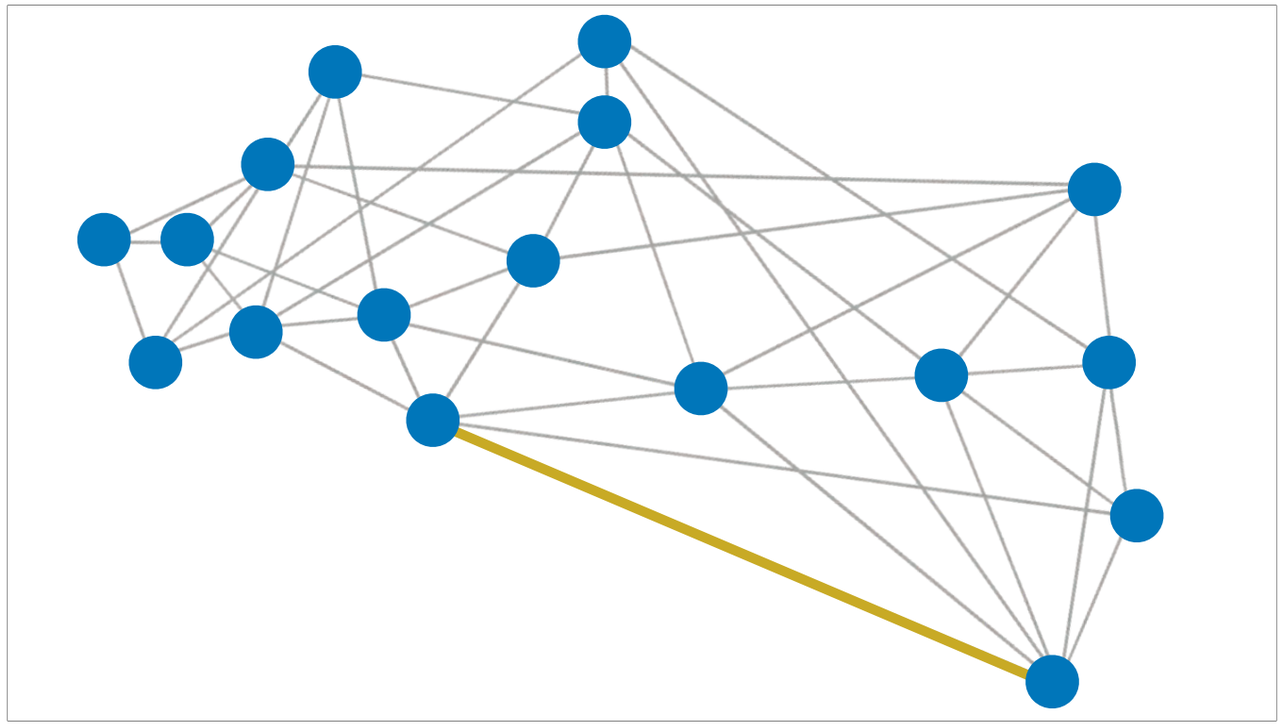
No need to rebuild the wheel and split hairs

Some classes are imported from other ontologies, others are constructed

Some of the classes from the literature search were specific to S.med, like “neoblast” and we had to generate all the class annotations for it. However, there are many anatomy ontologies, and why make something multiple times? For generic high level categorical terms , we could repurpose or flat out use these terms from other ontologies like the Gene Ontology (GO), The Uber Anatomy Ontology (UBERON), or my personal favorite the Biological Spatial Ontology (BSPO). When we looked at some terms we saw we could construct them by taking existing pieces from inside our ontology and outside of it to create more granular terms. These are called composite terms.

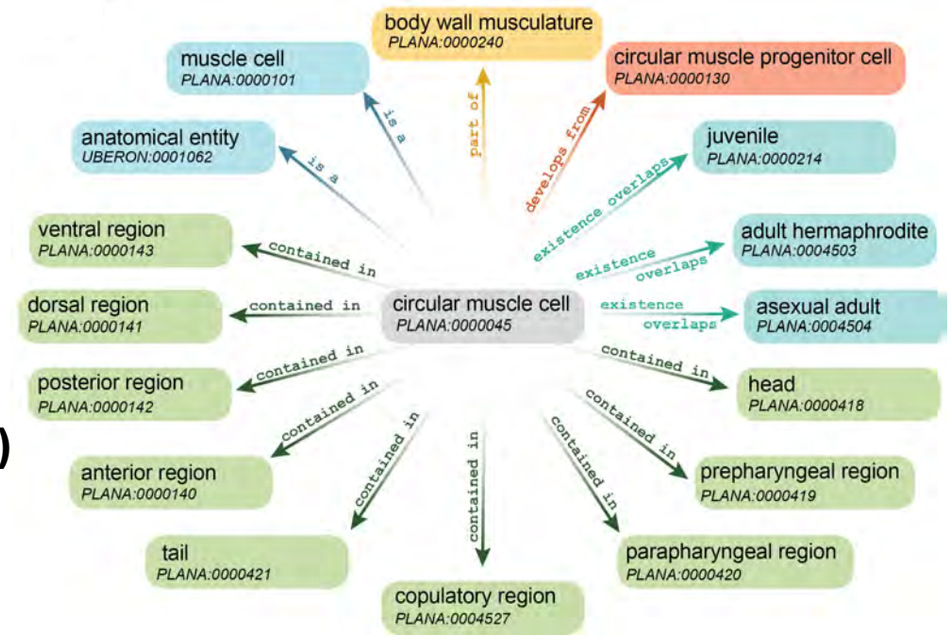
?

Up next : Relationships

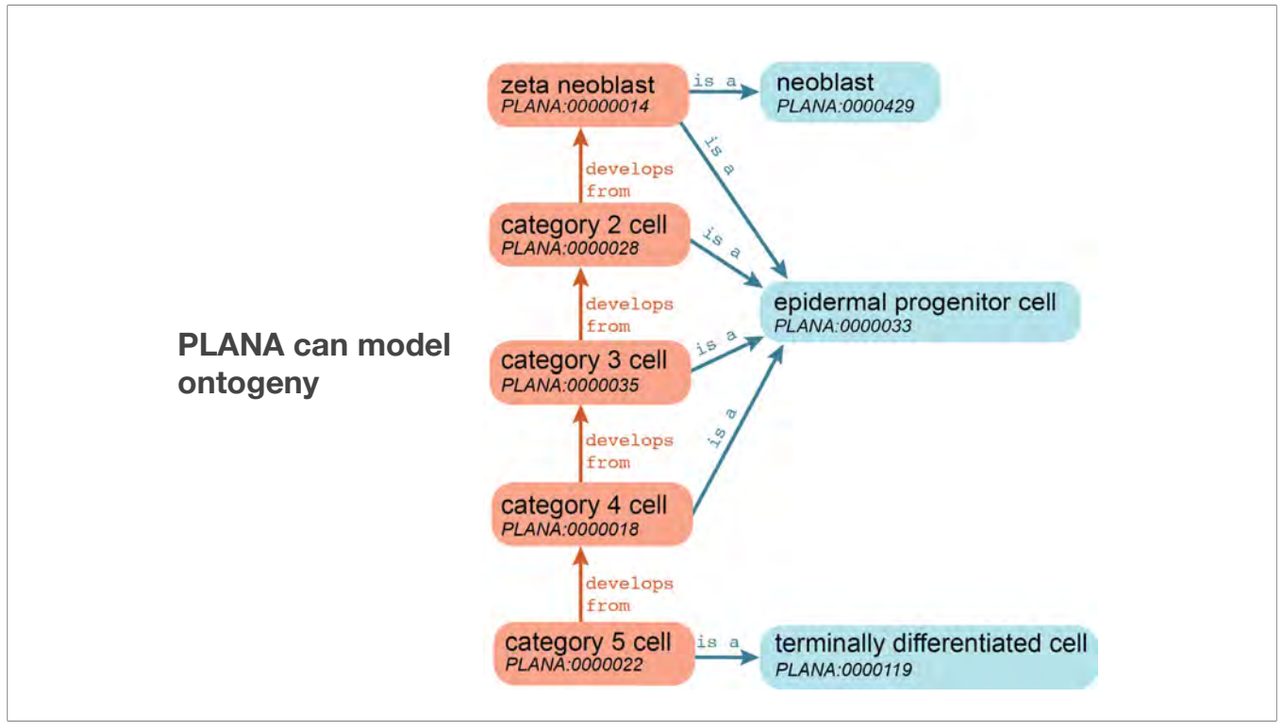


Now that I’ve told you how our nodes are constructed, this leaves the relationships between them to talk about. Luckily,

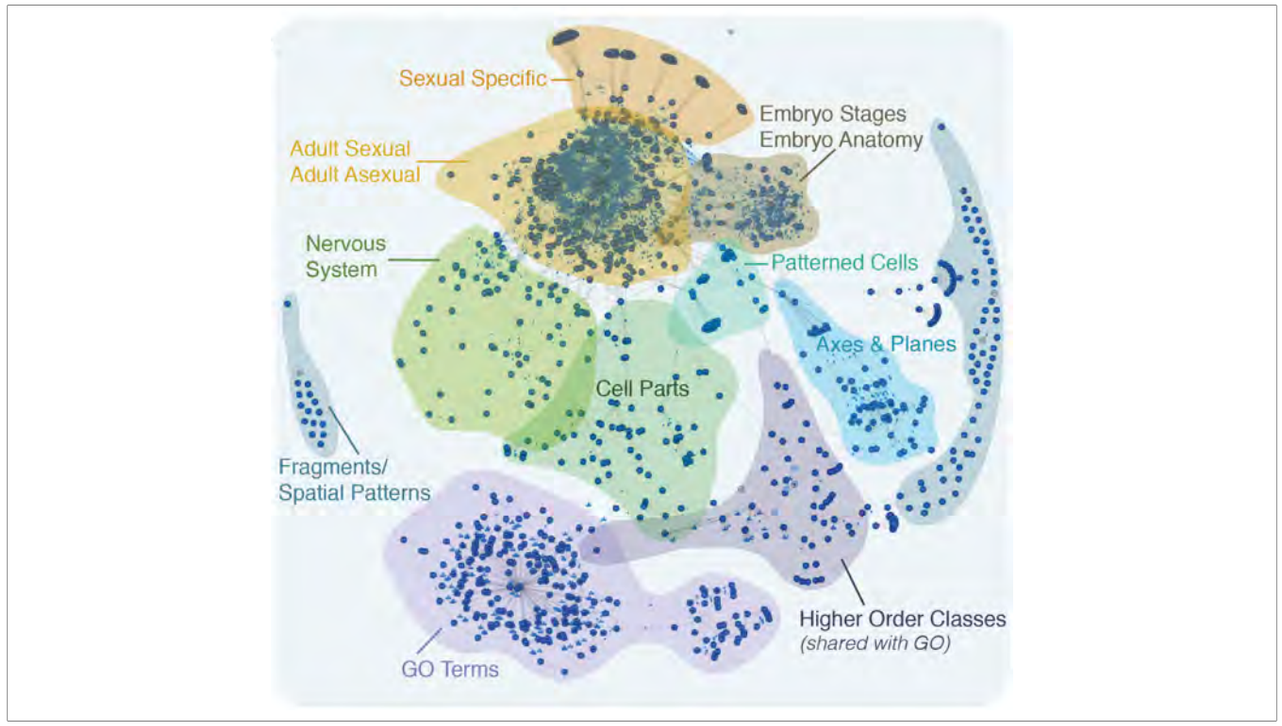
PLANA uses 14 relationships from the Relationship Ontology (RO)



Luckily, we didn’t have to make relationships up either. In fact, it’s better for interoperability between ontologies if the relationships are the same and used the same. For this we use the relationship ontology (RO). Relationships cover categorical, spatial and developmental connections and as is pointed out with the empty one, we can add as we need. As long as the relationships are from RO.



When we take a look at how relationships are hooked together we can do things like model ontogeny where we can show differentiation of stem cells in Smed.



When we look at the relationships and classes together our PLANA ontology looks a bit like this (for now). This is a living structure and is subject to change as our knowledge changes and as our needs as fieled change.

204 Primary Literature Papers

21 Other Ontologies

862 Classes

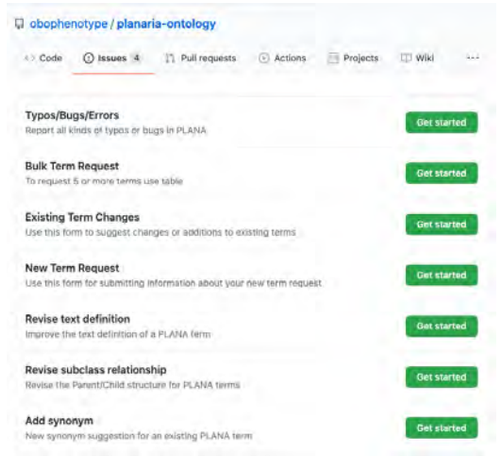
15 Relationships

13746 Axioms

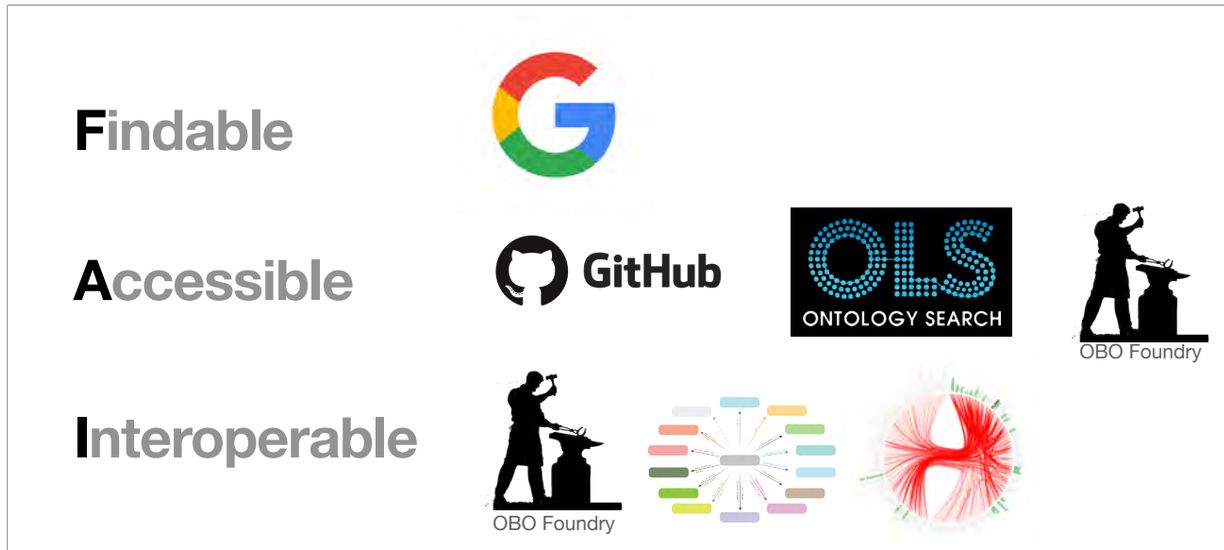
The whole structure of PLANA includes these numbers/ metrics… for now.

PLANA is a living structure

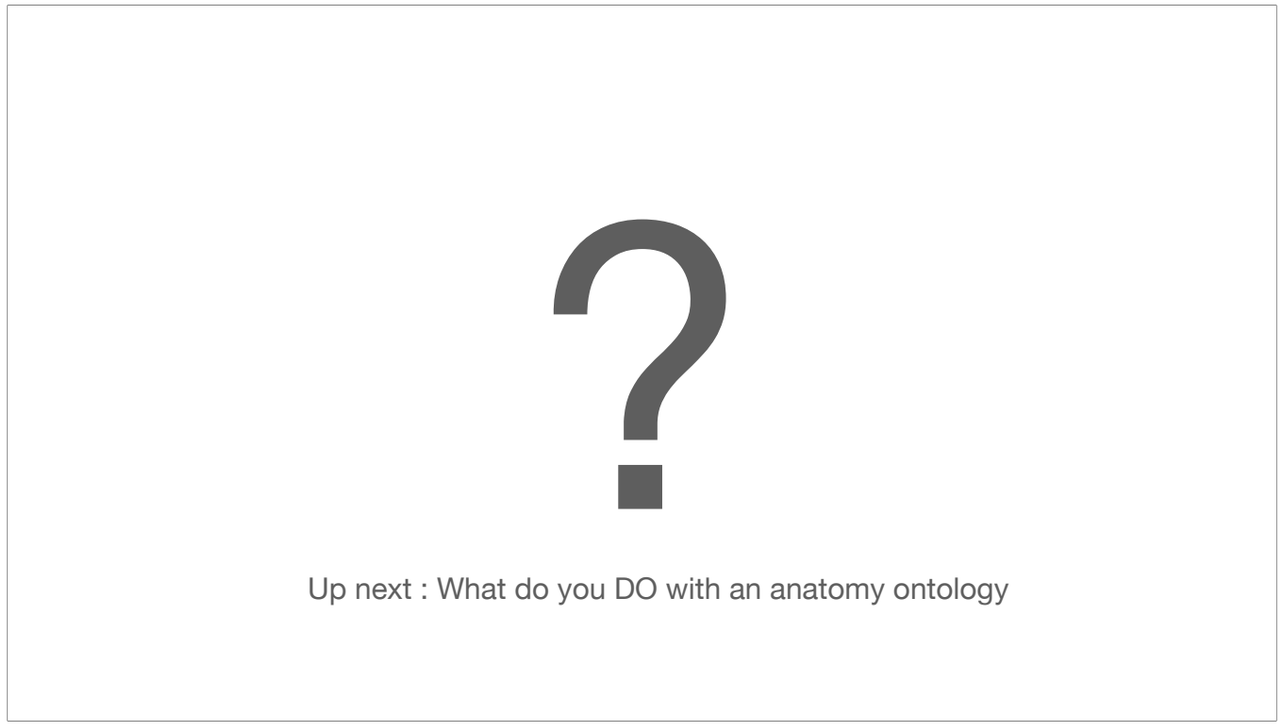
that will grow and change with the field



I say for now because PLANA is a living structure/document and is subject to change as our knowledge changes and as our needs as fieled change. Right now changes to the ontology can be suggested through our Git Hub Repo Issue Tracker. The way this works is, someone publishes new information (like Guidepost cells from Peter Reddien's lab) and can request that the term be added. Or if someone finds a mistake or would like to edit any of the annotations, they can. They submit a request, dialog is open on the page for a week and the change gets incorporated at the end of that week if there is majority consensus among those participating. Conflicts at the end of the week get brought up to a third party expert for decisions and persistent discussions will be Brought up in the tools session of the International Planarian Meeting.



One of the things we are most proud of with PLANA, is that it adheres to FAIR practices. It’s finable through a quick google search, it’s FREELY accessible through Git Hub, The Ontology Lookup Service and the OBO Foundry, because of using RO relations and adhering to other OBO Foundry guidelines PLANA is interoperable. And it’s reproducible, you could reproduce the whole thing for S.med (Not sure why you’d want to-but you could), but more importantly you can clone the ontology and repurpose for closely related species as a starting point.



So great, we cleaned out the anatomical closet for S.med. Now what?!

PLANA is a framework for organizing data

We built the Planarian Anatomy Gene Expression Database with it

www.planosphere.stowers.org

173
S.mediterranea papers
from 2005-2019

83,595
instances of expression
across life cycle stages
and evidence types

Searchable by:
Publication
Transcript ID
Anatomical Term
Ontology Relationship

Ontologies are a framework and a tool. We used our PLANA tool to build another tool, the Planarian Anatomy Gene Expression Database (PAGE for short). It is available at www.planosphere.stowers.org. At present it covers 173 publications holds >80,000 instances of expression across life cycle stage and evidence types. It's searchable by Publication, Transcript ID, Anatomical Term and Ontology Relationship. For example you can search for all transcripts that have been described as being expressed in any anatomical structure in the head. This search returns >20,000 instances of expression. (If things are going well here- do a live search)

PLoS Genet. 2011 Aug; 7(8): e1002226.
Published online 2011 Aug 11. doi: 10.1371/journal.pgen.1002226
PMCID: PMC3154955
PMID: 21852957

dlx and *sp6-9* Control Optic Cup Regeneration in a Prototypic Eye

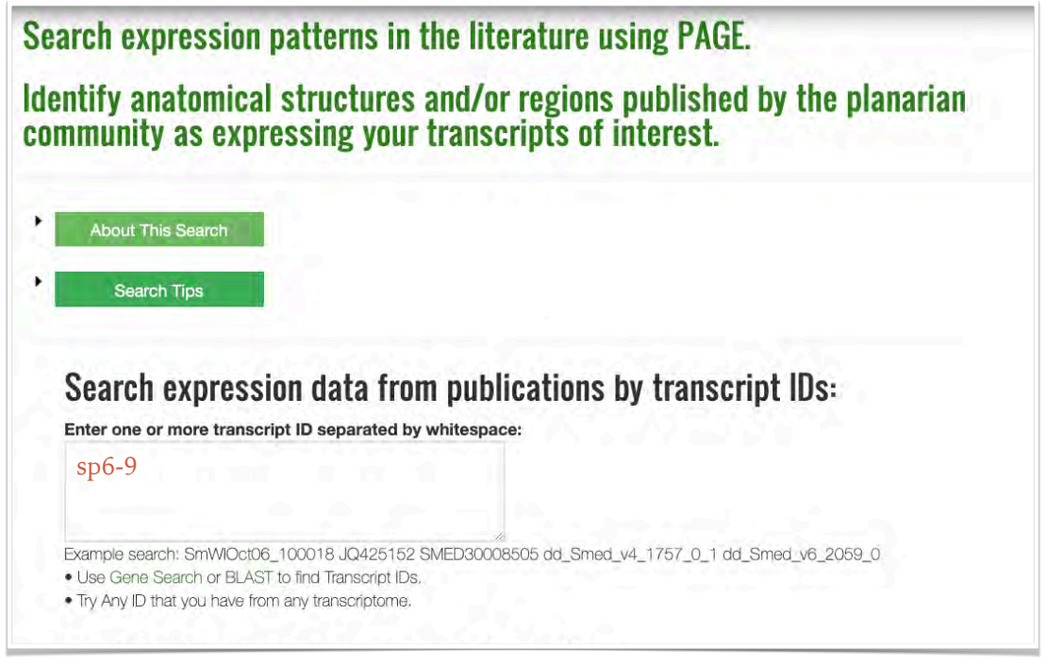
Sylvain W. Lapan [1, 2] and Peter W. Reddien [1, 2, 3, *]

Abstract

Go to:

Optic cups are a structural feature of diverse eyes, from simple pit eyes to camera eyes of vertebrates and cephalopods. We used the planarian prototypic eye as a model to study the genetic control of optic cup formation and regeneration. We identified two genes encoding transcription factors, *sp6-9* and *dlx*, that were expressed in the eye specifically in the optic cup and not the photoreceptor neurons. RNAi of these genes prevented formation of visible optic cups during regeneration. Planarian regeneration requires an adult proliferative cell population with stem cell-like properties called the neoblasts. We found that optic cup formation occurred only after migration of progressively differentiating progenitor cells from the neoblast population. The eye regeneration defect caused by *dlx* and *sp6-9* RNAi can be explained by a failure to generate these early optic cup progenitors. Dlx and Sp6-9 genes function as a module during the development of diverse animal appendages, including vertebrate and insect limbs. Our work reveals a novel function for this gene pair in the development of a fundamental eye component, and it utilizes these genes to demonstrate a mechanism for total organ regeneration in which extensive cell movement separates new cell specification from organ morphogenesis.

Let’s go through how PAGE was built and what it does in a little more detail. Annotators read papers and recorded instances of expression. Like in this paper sp-6 is expressed in the eye.



You can use page to search sp-6 by the transcript id used in the paper.

An example instance of expression in PAGE

sp6-9 —*expressed in*→ **eye (PLANA:0000036)**

- whole animal (PLANA:0000136)
- adult asexual (PLANA:0004504)
- fluorescent *in situ* hybridization (ECO:0001047)
- PMID:21852957

An instance of annotation for sp-6 from that paper looks like this:

- Recorded as being in the whole, adult asexual animal,
- The experiment types come from another ontology, evidence and conclusion ontology.
- And the pubmed ID is noted.

Observation Curated from Publication

Seq ID	Expressed In	Experimental Details	PMID
sp6-9	eye	AA	21852957
sp6-9	cephalic ganglia	AA	21852957
sp6-9	trail cell	AA	21852957
sp6-9	ventral region of the whole animal	AA	21852957
sp6-9	head margin	AA	21852957

This is a table of results of observations from the literature. Here you see the sp6-9 expression observation I have mentioned at top. sp6-9 is expressed in the eye. There are other observations from the same paper. sp6-9 is also expressed in the cephalic ganglia, trail cells etc. Each of these anatomical structures are PLANA ontology term and these are not all of the results… just the ones for this paper.

Search Details for: sp6-9

Result count: 15
specimenType: any | lifecycleType: any | evidenceType: any

Reference Sequence Information			Observation Curated from Publication					
ID	Description	Gene Models	Seq ID	Expressed In	Experimental Details			PMID
SMED30011934	sp6-9	SMESG000050160.1	JN983830.1	pigment cup cell				28976975
SMED30011934	sp6-9	SMESG000050160.1	JN983830.1	trail cell				28976975
SMED30011934	sp6-9	SMESG000050160.1	dd_Smed_v6_17385_0_1	head				28171748
SMED30011934	sp6-9	SMESG000050160.1	Contig37509	eye				27612384
SMED30011934	sp6-9	SMESG000050160.1	Contig37509	eye				27612384
SMED30011934	sp6-9	SMESG000050160.1	Contig37509	neuron				27612384
SMED30011934	sp6-9	SMESG000050160.1	Contig37509	neuron				27612384
SMED30011934	sp6-9	SMESG000050160.1	SmedASXL_006176	head				27034770
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	eye				25254346
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	central nervous system				25254346
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	eye				21852957
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	cephalic ganglia				21852957
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	trail cell				21852957
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	ventral region of the whole animal				21852957
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	head margin				21852957

Here’s all 15 observations. Let’s take a closer look at them

Search Details for: sp6-9

Result count: 15

specimenType: any | lifecycleType: any | evidenceType: any

Reference Sequence Information			Observation Curated from Publication				
ID	Description	Gene Models	Seq ID	Expressed In	Experimental Details		PMID
SMED30011934	sp6-9	SMESG000050160.1	JN983830.1	pigment cup cell			28976975
SMED30011934	sp6-9	SMESG000050160.1	JN983830.1	trail cell			28976975
SMED30011934	sp6-9	SMESG000050160.1	dd_Smed_v6_17385_0_1	head			28171748
SMED30011934	sp6-9	SMESG000050160.1	Contig37509	eye			7612384
SMED30011934	sp6-9	SMESG000050160.1	Contig37509	eye			27612384
SMED30011934	sp6-9	SMESG000050160.1	Contig37509	neuron			27612384
SMED30011934	sp6-9	SMESG000050160.1	Contig37509	neuron			27612384
SMED30011934	sp6-9	SMESG000050160.1	SmedASXL_006176	head			27034770
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	eye			25254346
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	central nervous system			25254346
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	eye			21852957
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	cephalic ganglia			21852957
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	trail cell			21852957
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	ventral region of the whole animal			21852957
SMED30011934	sp6-9	SMESG000050160.1	sp6-9	head margin			21852957

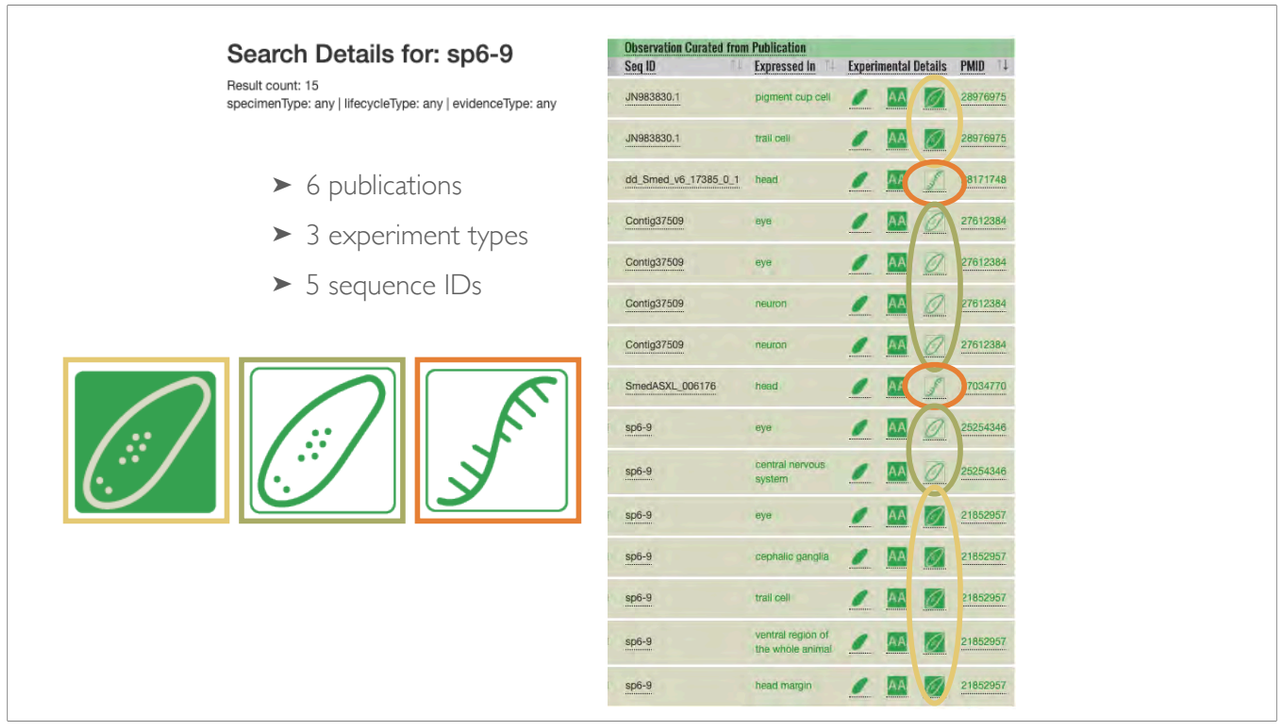
First starting at the far right we can see they are from 6 different publications

Search Details for: sp6-9

Result count: 15
specimenType: any | lifecycleType: any | evidenceType: any

- 6 publications
- 3 experiment types
- 5 sequence IDs

Observation Curated from Publication			
Seq ID	Expressed In	Experimental Details	PMID
JN983830.1	pigment cup cell		28976975
JN983830.1	trail cell		28976975
dd_Smed_v6_17385_0_1	head		28171748
Contig37509	eye		27612384
Contig37509	eye		27612384
Contig37509	neuron		27612384
Contig37509	neuron		27612384
SmedASXL_006176	head		27034770
sp6-9	eye		25254346
sp6-9	central nervous system		25254346
sp6-9	eye		21852957
sp6-9	cephalic ganglia		21852957
sp6-9	trail cell		21852957
sp6-9	ventral region of the whole animal		21852957
sp6-9	head margin		21852957



- 3 different experiment types
    - fluorescent in situ hybridization
    - RNA seq
    - and colorimetric in situ hybridization

Search Details for: sp6-9

Result count: 15
specimenType: any | lifecycleType: any | evidenceType: any

- 6 publications
- 3 experiment types
- 5 sequence IDs

Observation Curated from Publication			
Seq ID	Expressed In	Experimental Details	PMID
JN983830.1	pigment cup cell		28976975
JN983830.1	trail cell		28976975
dd_Smed_v6_17385_0_1	head		28171748
Contig37509	eye		27612384
Contig37509	eye		27612384
Contig37509	neuron		27612384
Contig37509	neuron		27612384
SmedASXL_006176	head		27034770
sp6-9	eye		25254346
sp6-9	central nervous system		25254346
sp6-9	eye		21852957
sp6-9	cephalic ganglia		21852957
sp6-9	trail cell		21852957
sp6-9	ventral region of the whole animal		21852957
sp6-9	head margin		21852957

- and 5 different sequence IDs were used in those 6 papers!!!

What just happened?

★ Search took seconds

★ Every paper, even those using different experiment types, confirms that sp6-9 is expressed in some part of the eye, the eye, or more generally the head

★ Observations for the same sequence (sp6-9) were returned even though different IDs were used in the publications

★ mapping of a dozen datasets done by Eric Ross

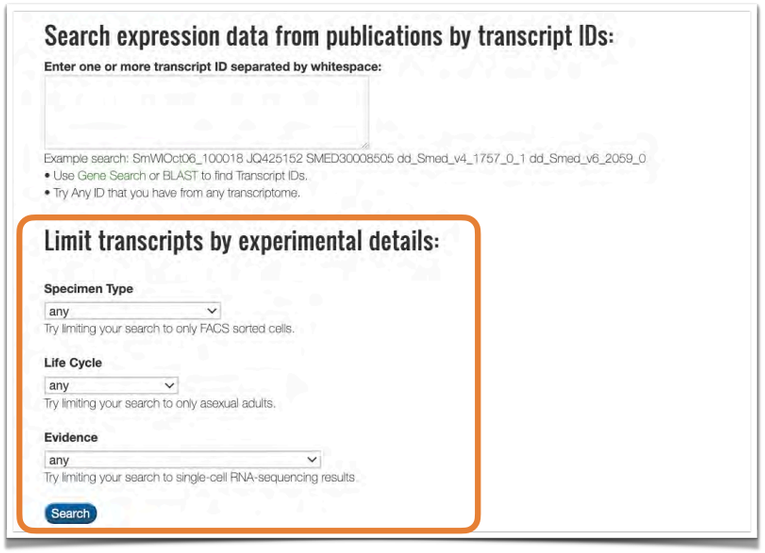
Observation Curated from Publication

Seq ID	Expressed In	Experimental Details	PMID
JN983830.1	pigment cup cell		28976975
JN983830.1	trail cell		28976975
dd_Smed_v6_17385_0_1	head		28171748
Contig37509	eye		27612384
Contig37509	eye		27612384
Contig37509	neuron		27612384
Contig37509	neuron		27612384
SmedASXL_006176	head		27034770
sp6-9	eye		25254346
sp6-9	central nervous system		25254346
sp6-9	eye		21852957
sp6-9	cephalic ganglia		21852957
sp6-9	trail cell		21852957
sp6-9	ventral region of the whole animal		21852957
sp6-9	head margin		21852957

lets take a minute to talk about what happened here. First, the search took seconds.

- how long would it take you to FINDand read 6 papers about a transcript you are interested in? *probably longer than seconds*
- every paper, even those using different experiment types, confirms that sp6-9 is expressed in some part of the eye, the eye, or more generally in the head.
- observations for the same sequence (sp6-9) were returned even though different IDs where used in the publications
    - this is made possible by the mapping of a dozen different datasets (eric ross)
- I know i helped to build this tool, but i think it is pretty amazing.

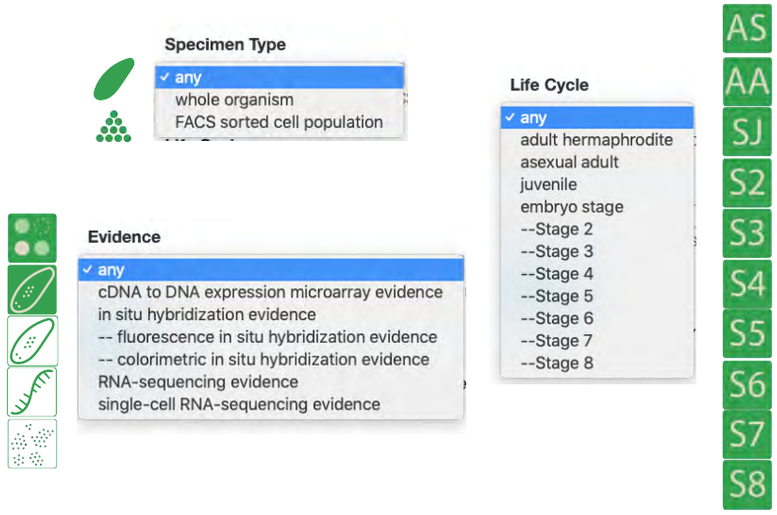
Searches can be limited by experiment details



In our PAGE searches, here is our search by transcript, you can limit your results by these experimental details.

- Specimen type
- Life cycle
- Evidence

Choose details that are important to you.



When doing a search you can limit the result to the details that matter the most to you.
- perhaps you are only interested in FACS sorted cells
- and only RNAseq or single cell rna seq experiments

well you can do that.

While collecting information about each observation we have assembled information from
2 specimen types
- Whole organism
- Facs sorted cells

5 experiment types, or evidence
- Microarray
- Fluorescent in situ hybridization
- Colormetric in situ hybridization
- RNA seq
- Single cell rna seq

10 life cycle types
- Adult sexual and asexual
- Sexual juvenile

- 7 different embryonic stages 2-8

PLANA is currently being used for:

Constructing a phenotype ontology

Annotation of 3D EM data


Sofia Robb
Erin Davies
Nico Matzengolu
Eric Ross
Viraj Doddihal
Mol Mir
Melainia McClain
Alejandro Sánchez Alvarado