Functional Genomics I
Exploring transcriptomics, proteomics, phenomics and metabolic pathways

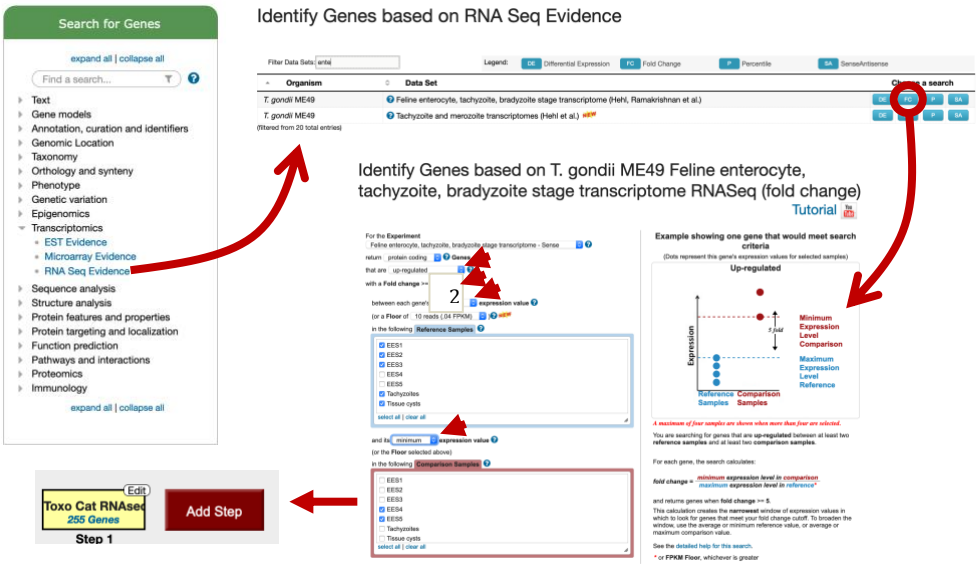
1. Find *T. gondii* genes expressed in late enteroepithelial stages

Toxoplasma gondii is a zoonotic pathogen for which felids serve as definitive hosts. In cats, the parasite undergoes several rounds of asexual replication before entering the sexual cycle which gives rise to oocysts that are shed into the environment. These then sporulate and become infective to humans and livestock. To understand the genes involved in the parasite development in the felid host and identify potential intervention targets, we designed a transcriptomic approach to compare the cat intestinal stages with the well characterized tachyzoites that mediate acute infection and tissue cysts that are responsible for chronic infection. Cats were infected with *T. gondii* CZ clone H3 tissue cysts from mouse brain and the intestinal stages were sampled at day 3, 5 and 7 post infection. As an input sample, we also collected tissue cysts from mouse brain. In vitro cultivated tachyzoites were also harvested. Total RNA was extracted, enriched for mRNA and used for cDNA synthesis. RNA-Seq was then performed to describe the transcriptomic repertoire of each developmental stage. RNA-seq datasets from each time point post inoculation with bradyzoites in kittens were subjected to cluster analysis and assigned to five enteroepithelial developmental stages (EES) according to their profile.

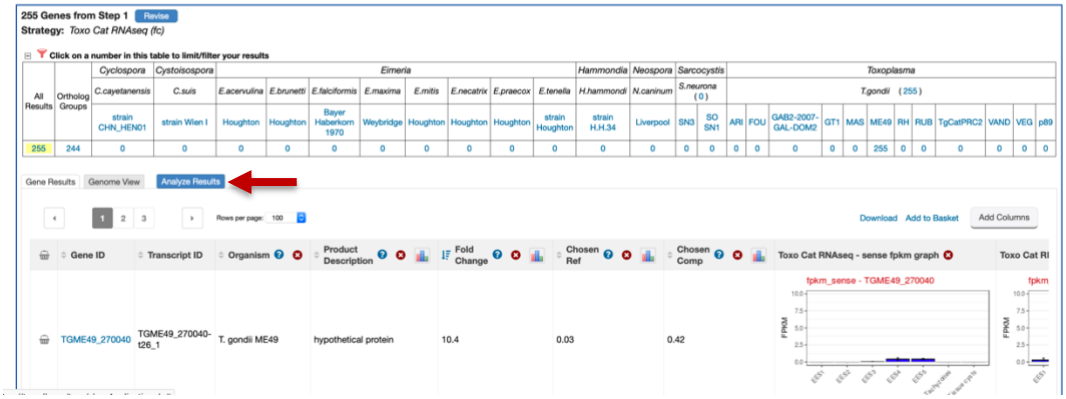
Cat enteroepithelial stages:

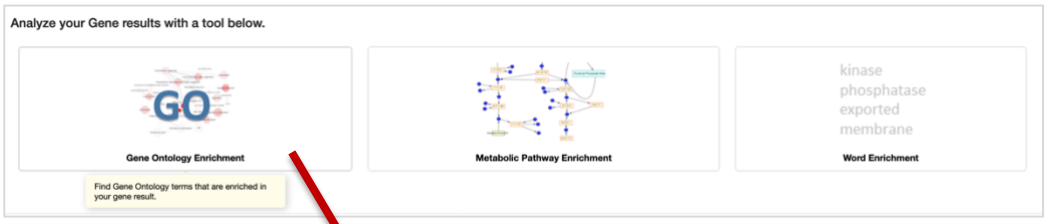
- EES1 = very early enteroepithelial stages
- EES2 = early enteroepithelial stages
- EES3 = mixed enteroepithelial stages
- EES4 = late enteroepithelial stages
- EES5 = very late enteroepithelial stages

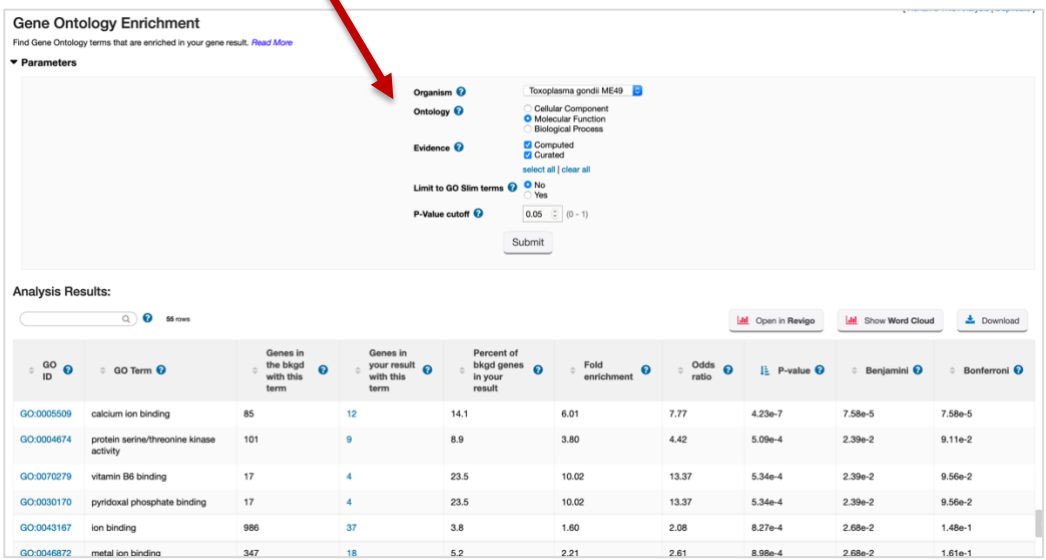
- Navigate to the RNAseq searches and identify the experiment of cat enterocyte stages. Configure the search to identify call *T. gondii* genes that are upregulated by at least 2-fold in late and very late enteroepithelial stages (EES4 and EES5) compared to all other stages available from this experiment.



- What kinds of genes did this search identify? How can you determine if your results are enriched for a particular function? Try clicking on Analyze Results and explore the GO enrichment tool.



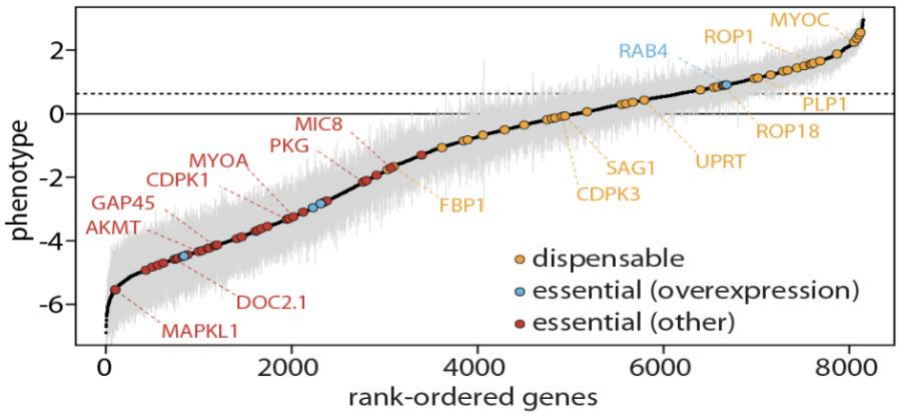


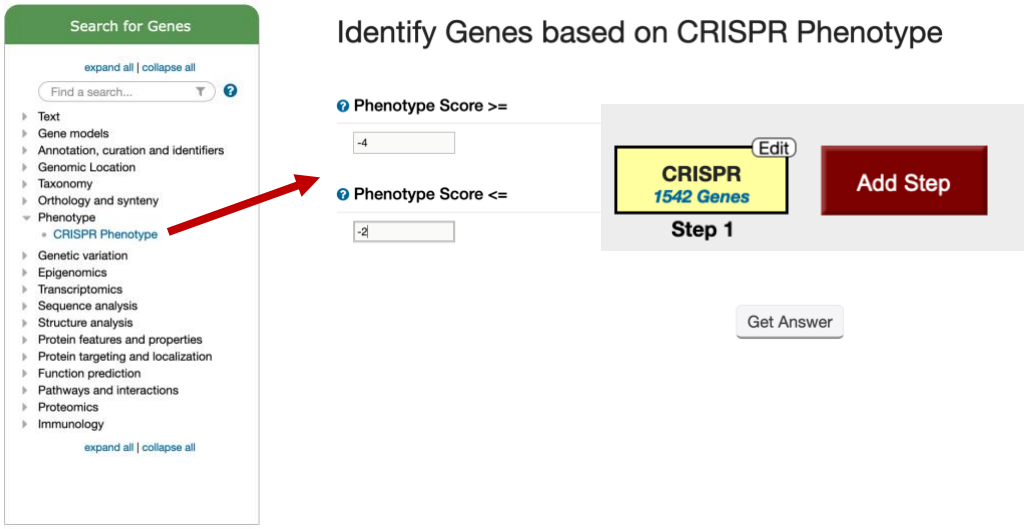


2. Finding genes based on high throughput mutagenesis and fitness analysis.

In EuPathDB we have a variety of studies where genome scale phenotypic analyses were carried out. In this exercise we'll use ToxoDB.org and look at fitness following CRISPR mutagenesis. You could also explore phenotyping studies in PlasmoDB or FungiDB if you prefer, the principles are the same.

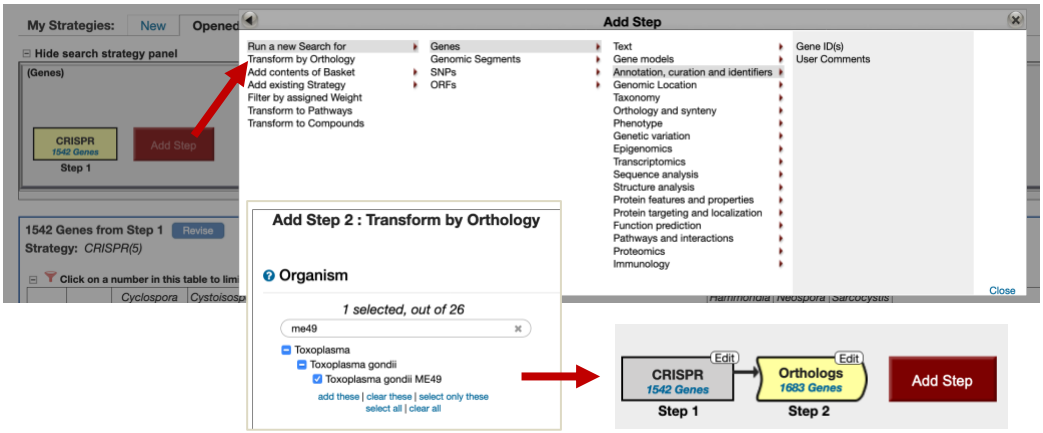
- Navigate to the CRISPR phenotype search. Note that this search form is quite simple just requiring a range of fitness values. The defaults return all genes not limiting the search at all. This is only useful in as much as it tells you which genes were assayed which is nearly the entire genome. The tricky bit is deciding where to make the cutoffs. Again, the description on the search form is very helpful in this regard (as is the link to the paper … remember these phenotypes were assayed under specific conditions so just because a particular gene doesn't show a phenotype doesn't mean it wouldn't in other conditions (or infecting an actual host). The plot showing the phenotype score (fitness) is particularly useful. Red points along the plot are genes known to be essential under these conditions while yellow are known to be expendable. This will help you determine where to set the values. The last essential gene has a fitness score just >= than -4 so setting the phenotype score <= -2 would provide a pretty stringent search but still return more than 1000 genes. Try it. Do you get the expected results based on the number of genes returned?



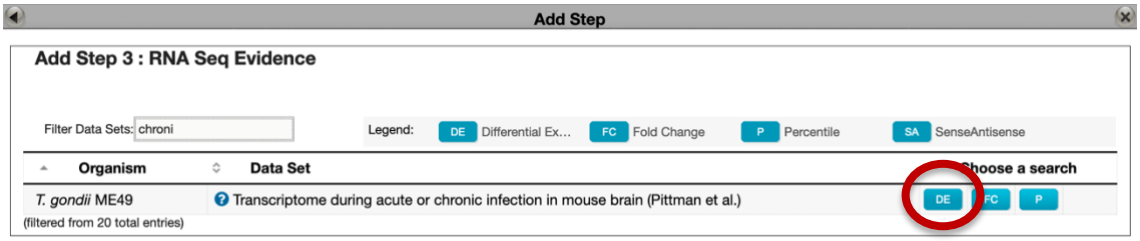


- Can you find additional evidence that these genes are essential? One way is to use the analysis tools to assess biological process and go function. Are the results what you would expect?

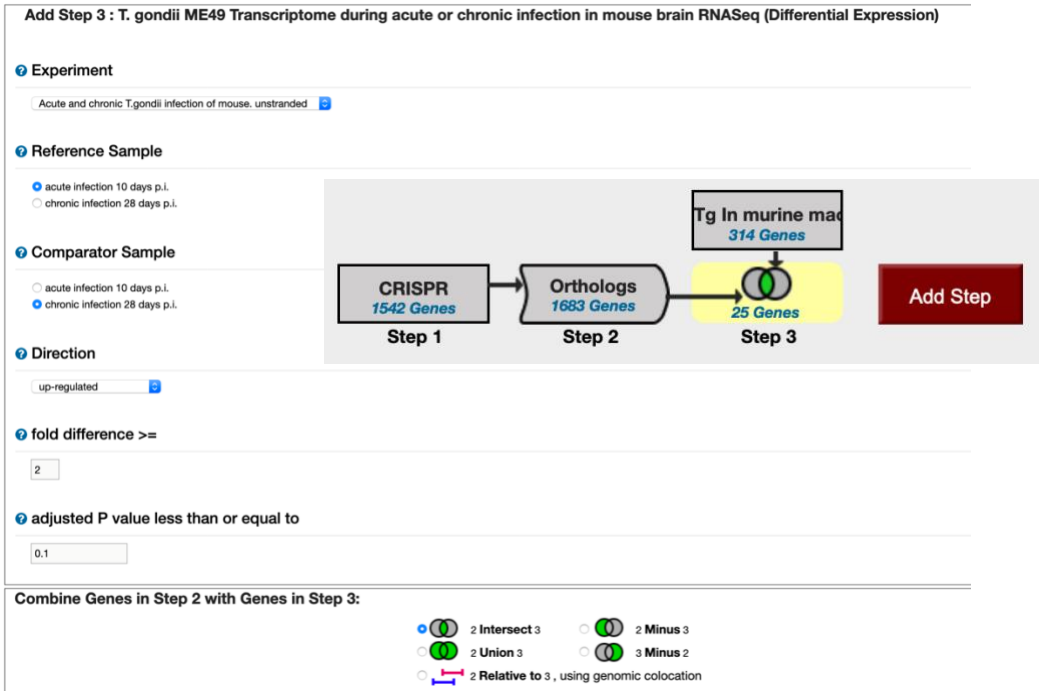
- Try adding columns to show additional data or intersecting these results with other queries, perhaps expression queries, to further assess this list. NOTE: this experiment was done in GT1 while all *T. gondii* functional data in ToxoDB is mapped to ME49 so an ortholog transform to ME49 is required before adding any additional functional studies.
- To do this, click on add step and select the Transform to orthologs option and select *T. gondii* ME49 to transform to.



- How many of these genes are upregulated in *in vivo* chronic stages of *T. gondii*?
    - Click on add step
    - Select the RNAseq searches under the Transcriptomics category
    - Find the experiment with chronic stages and run a search based on differentially expressed genes (DE).

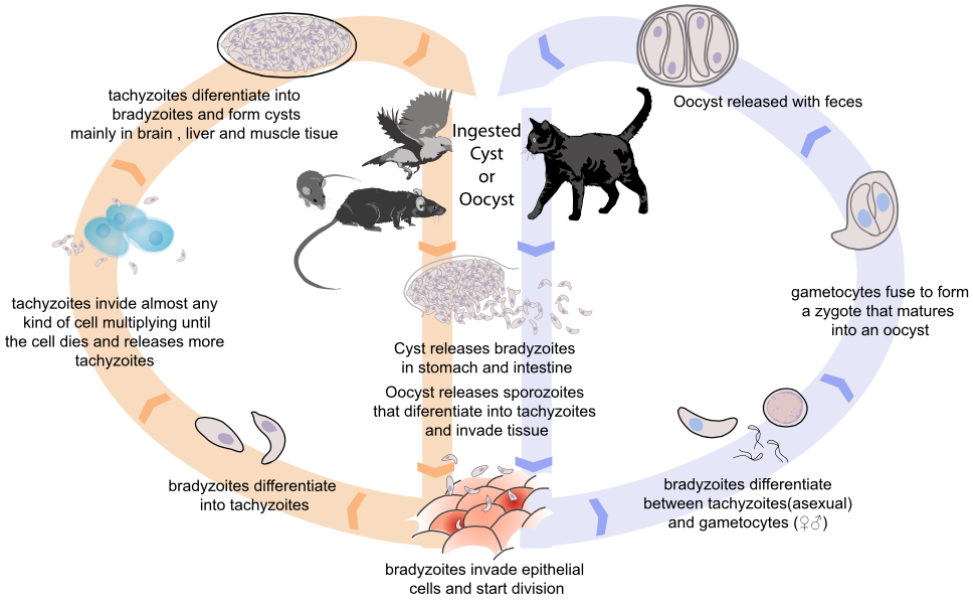


    - Intersect genes that are 2-fold upregulated in chronic stages compared to acute stages.

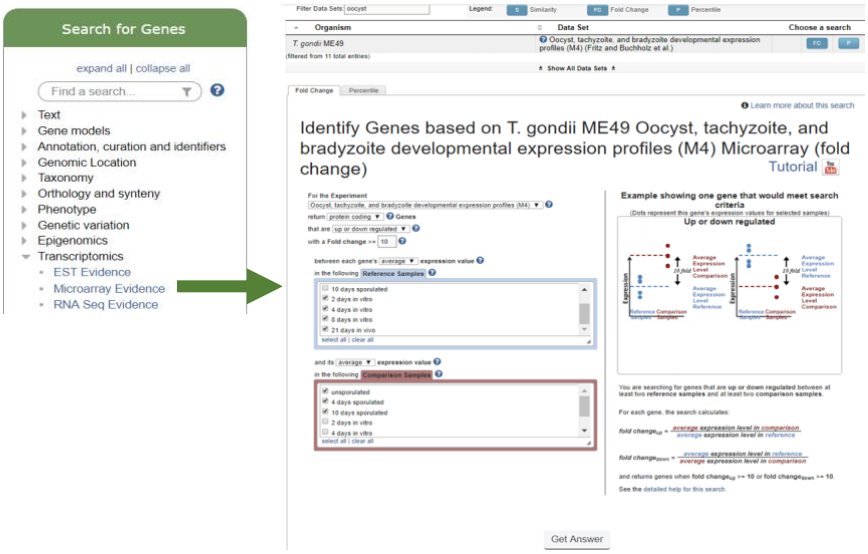


- What do these results look like? Do you find any interesting genes?

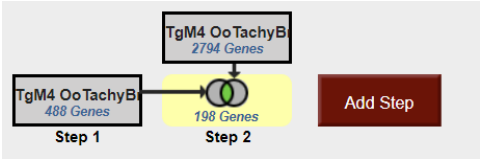
3. Finding oocyst expressed genes in *T. gondii* based on microarray evidence.



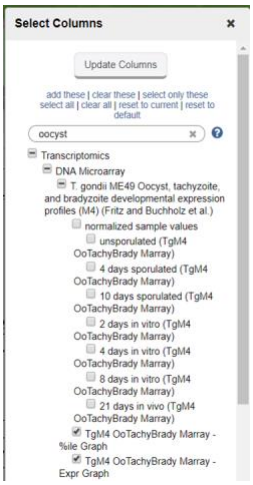
a. Find genes that are expressed at 10 fold higher levels in one of the oocyst stages than in any other stage in the "**Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4) (John Boothroyd)**" microarray experiment.

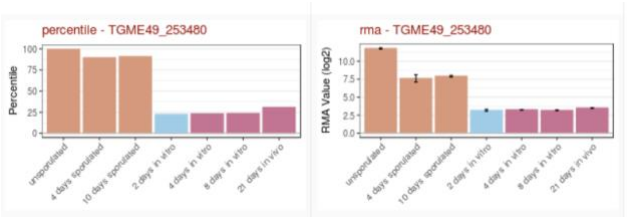


b. Add a step to limit this set of genes to only those for which all the non-oocyst stages are expressed below 50th percentile … ie likely not expressed at those stages. *(Hint*: after you click on add step find the same experiment under microarray expression and chose the percentile search).



- Select the 4 **non-oocyst** samples.
- We want all to have less than 50th percentile so set ***minimum percentile*** **to 0** and ***maximum percentile*** **to 50**.
- Since we want all of them to be in this range, choose **ALL** in the "***Matches Any or All Selected Samples***".
- To view the graphs in the final result table, turn on the columns called "TgM4 OoTachyBrady Marray - Expr Graph" and "TgM4 OoTachyBrady Marray - %ile Graph" (inside the "T. gondii ME49 Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4) (Fritz and Buchholz et al.)" Microarray).



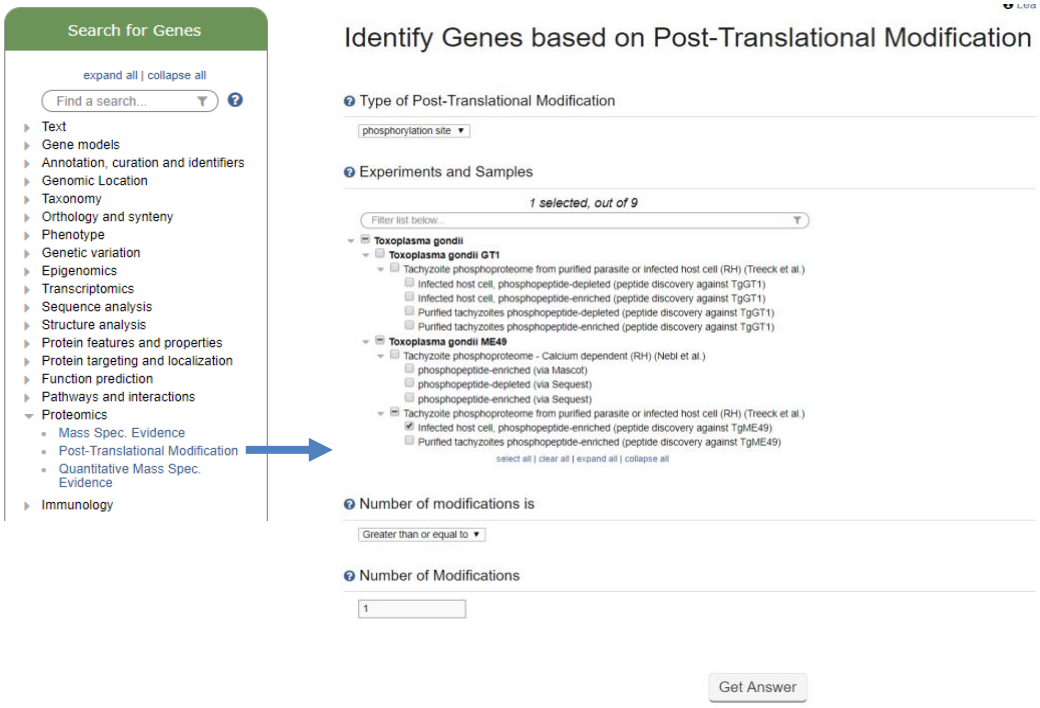


4. Find genes with evidence of protein phosphorylation in intracellular *Toxoplasma* tachyzoites. For this exercise use http://www.toxodb.org

Phosphorylated peptides can be identified by searching the appropriate experiments in the Mass Spec Evidence search page.

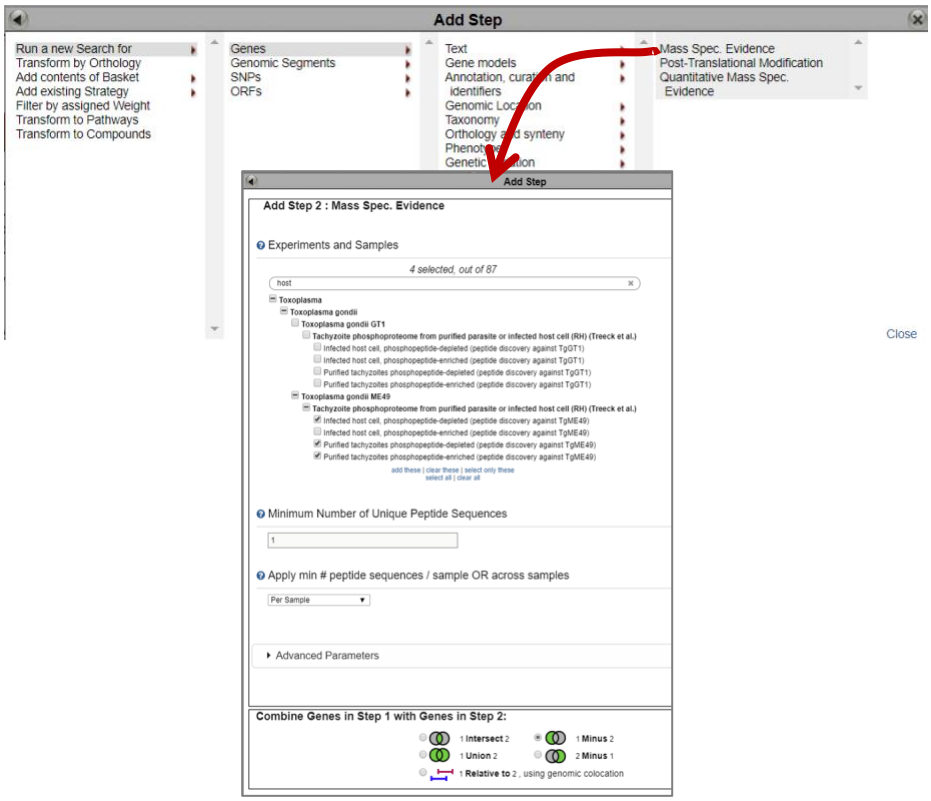
4a. Find all genes with evidence of protein phosphorylation in intracellular tachyzoites. Navigate to the Post-Translational Modification search. Select the "**Infected host cell, phosphopeptide-enriched** (peptide discovery against **TgME49**)" sample under the experiment called "**Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treeck et al.)**"

4b. Remove all genes with phosphorylation evidence from purified tachyzoites and the

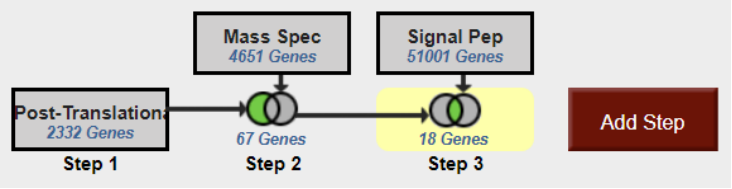


phosphopeptide depleted fractions.

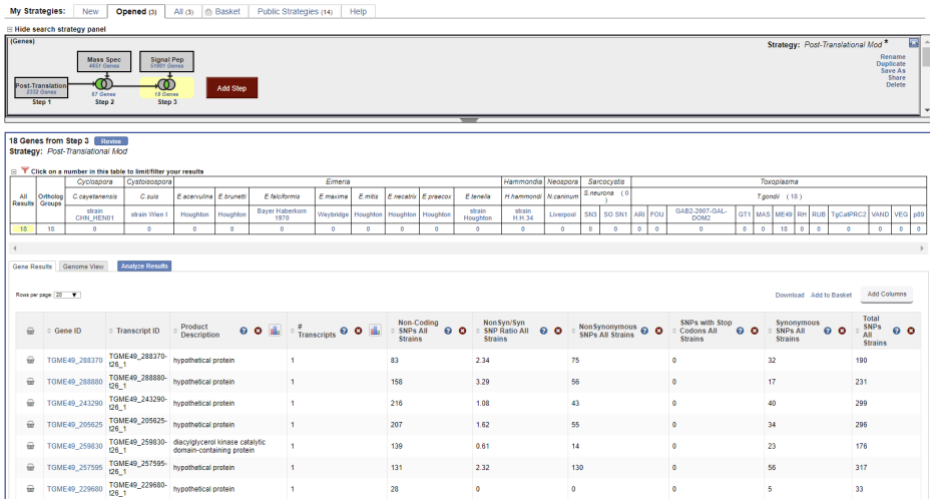
Hint: Use the Mass Spec Evidence search to access the tachyzoite and depleted fractions. Subtract (1 minus 2) these results from your first search.



4c. Explore your results. What kinds of genes did you find? *Hint: use the Product description word column or perform a GO enrichment analysis of your results.*



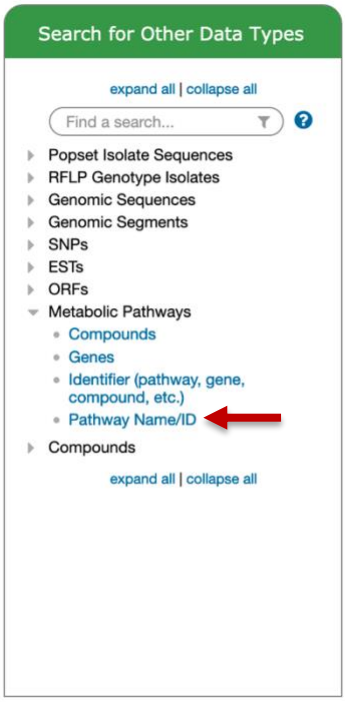
4d. Are any of these genes likely to be secreted? Hint: add a step searching for genes with secretory signal peptides.



4e. Pick one or two of the hypothetical genes in your results and visit their gene pages. Can you infer anything about their function? Hint: explore the protein and expression sections.

4f. What about polymorphism data? Go back to your strategy and add columns for SNP data found under the population biology section. Explore the gene page for the gene that has the highest number of non-synonymous SNPs. Hint: you can sort the columns by clicking on the up/down arrows next to the column names.

5. Find and explore the metabolic pathway for glycolysis.

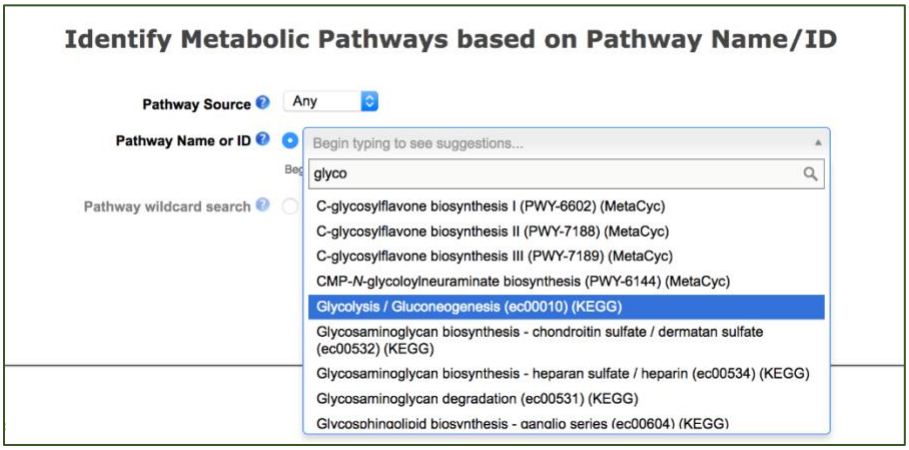


Navigate to the search page for Identify Metabolic Pathways based on Pathway Name/ID.

- Metabolic pathway and compound searches are available in the "Identify Other Data Types" section on the home page. You can find metabolic pathways based on the pathway name, genes involved in the pathway, or compounds involved in the pathway. Search for the **glycolysis** pathway using the Pathway Name/ID option.

- This search is equipped with a type-ahead function for choosing the metabolic pathway name. Begin typing glycolysis and then choose the pathway name from the list that appears.

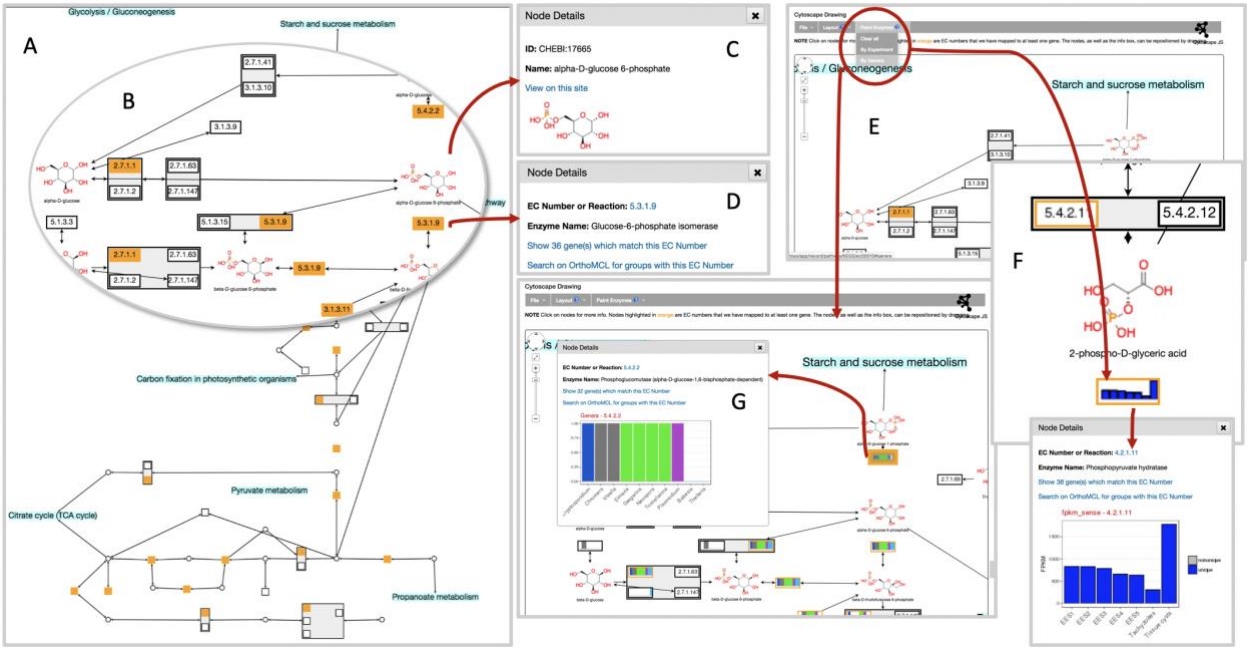
a. Examine the Glycolysis / Gluconeogenesis pathway.



- The search takes you straight to the record page for the Glycolysis / Gluconeogenesis (ec00010) metabolic pathway from KEGG. The overview section of the record page contains an interactive graphical representation of the pathway. The pathway map and the legend can be repositioned.

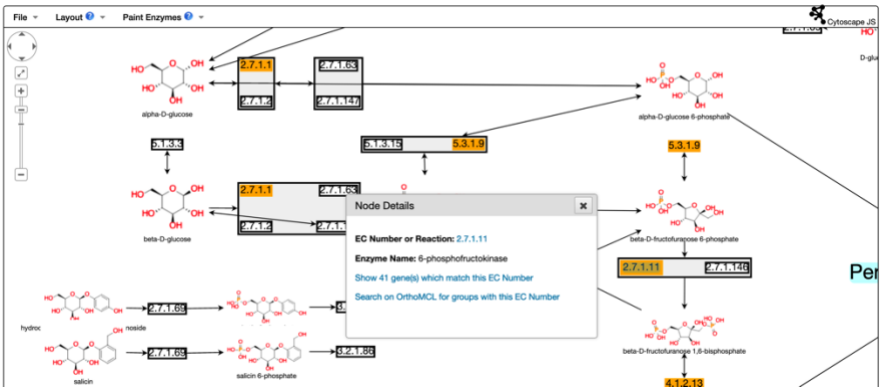
A. Initial pathway view is zoomed out.
B. Zoom in to see more details including EC numbers and metabolite structures.

C. Click on a metabolite structure to get additional information.
D. Click on the EC number to get more info about the enzyme including links to retrieve all genes in the database assigned to this EC number.

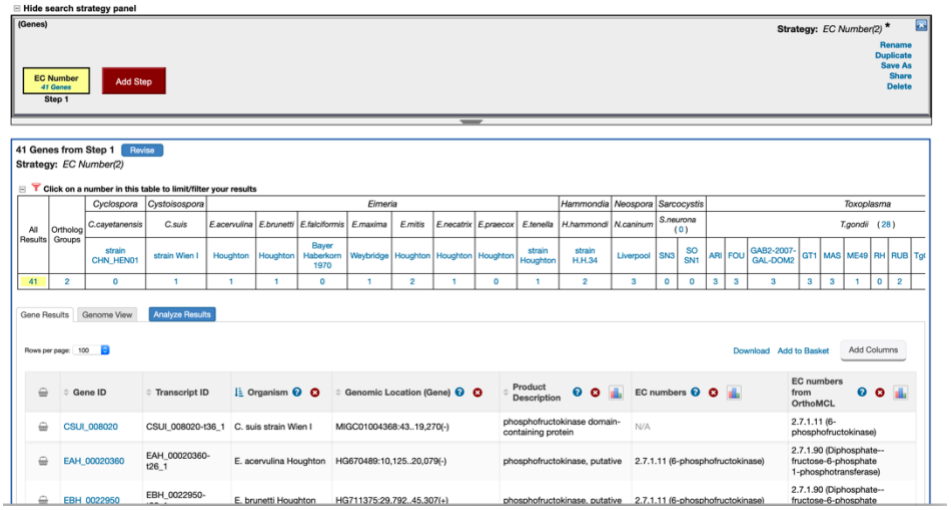


E. The drop-down menu under the heading "Paint Enzymes" allows you paint the pathway based on experiments or based on phyletic pattern.
F. Painting pathway by experiment provides a graphical representation of experimental results. Click on the graph to see more details.
G. Painting pathway based on phyletic pattern provides a graphical representation of phyletic distribution. Clicking on the phyletic pattern graphic provides additional information.

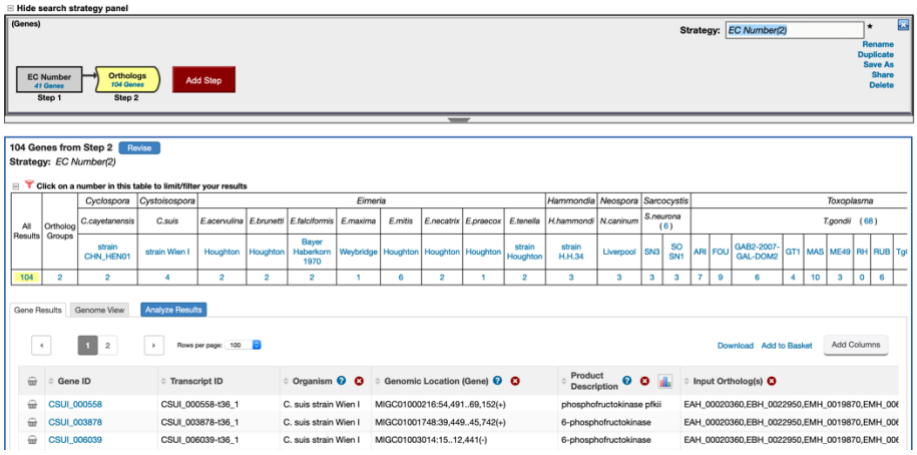
- What do the rectangles with numbers like 2.7.1.11 represent – zoom in closer to see EC numbers?
- What is the difference between the rectangular nodes that are orange and those that are not?
- Why are some enzymes grouped?
- Click on the 2.7.1.11 node to open a popup with information about this enzyme.



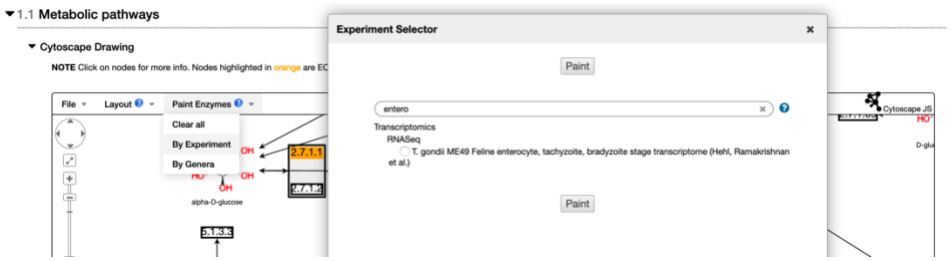
- How many genes in the database matched this EC number?
- Try the link ‘Show 41 gene(s) which match this EC Number’. Where did you end up? What do the 41 genes in the result list represent? Is 6-phosphofructokinase unique to *Toxoplasma*? Notice the two columns called “EC numbers” and “EC numbers from OrthoMCL”. What do these columns represent?



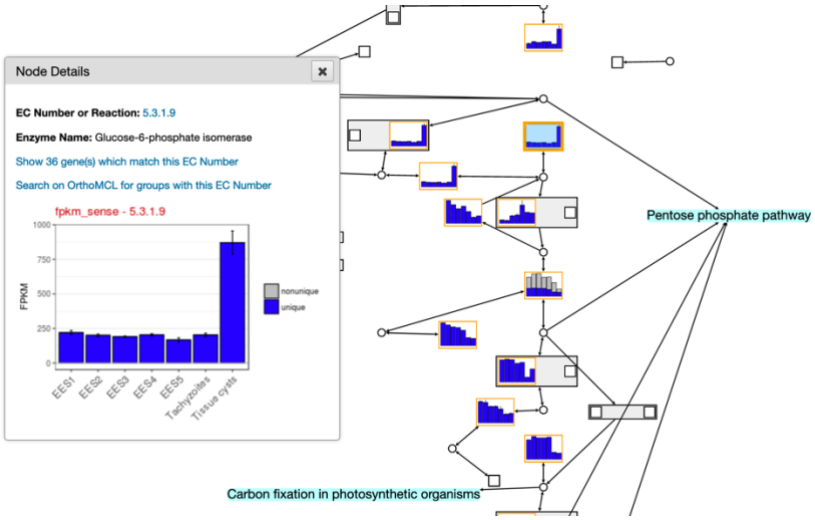
- Is this enzyme missing from some organisms? Do you think this is possible? What step can you add to confirm this? (*hint: try an ortholog transform*)



- Use your Browser’s back button to return to the Glycolysis pathway record page and open the Paint Experiment menu. Choose the experiment “T. gondii ME49 Feline enterocyte, tachyzoite, bradyzoite stage transcriptome (Hehl, Ramakrishnan et al.)”. Be patient while the graphs appear in place of the EC numbers.



- Do you see enzymes that appear to be differentially regulated? Note you can click on the graph to see a larger image.



- Use the Paint Genera option to paint the pathway with orthologs from across Apicomplexa and Chromerida. Explore the results.