Exploring proteomics data in VEuPathDB Resources

Data from proteomics experiments are integrated into VEuPathDB resources under three categories:

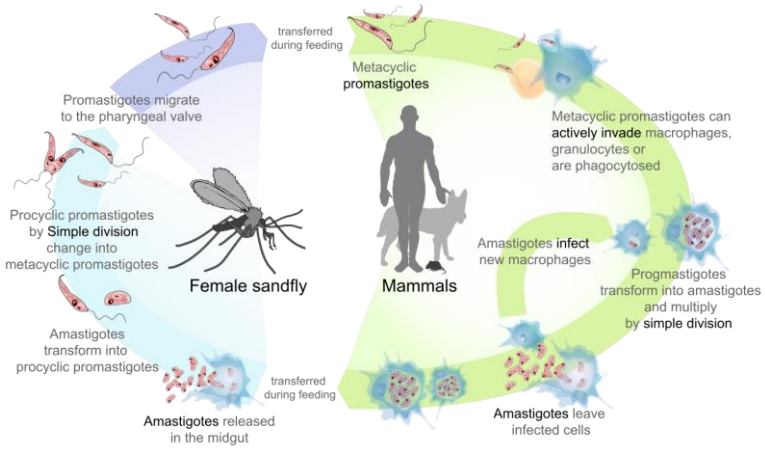
1. **Mass spec. evidence -** Peptides from proteomics experiments are mapped to a reference genome enabling searches for genes based on that mapping.
2. **Quantitative mass spec. evidence -** Data from quantitative proteomic experiments are loaded and made available for searching based on fold change or differential expression.
3. **Post-translational modification (PTM) -** PTM data from proteomics experiments are loaded on genes enabling searches for genes based on the type and number of the PTM.

The exercises below explore the different categories and searches available for proteomics in VEuPathDB.

Learning objectives:

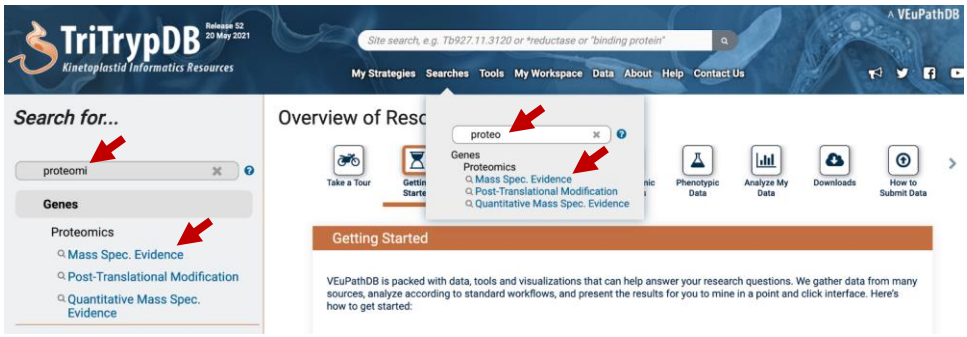
- Understand the different categories of proteomics data
- Learn how to run searches to identify genes based on peptide evidence
- Learn how to identify differentially expressed genes based on quantitative data
- Learn how to identify genes with different PTMs

1. **Find genes that have peptide evidence from metacyclic stages but not amastigote or promastigote stages of *Leishmania infantum*. Note: for this exercise use http://tritrypdb.org**

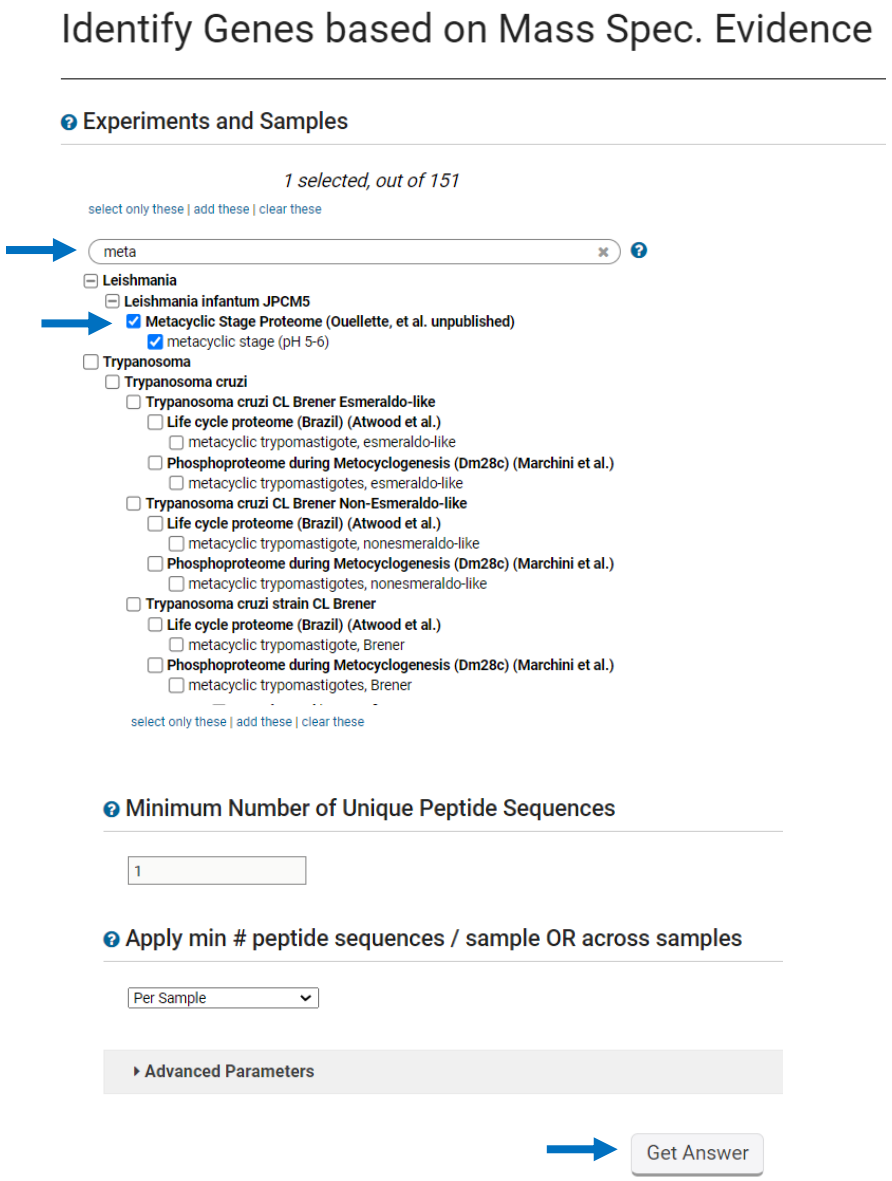


Life cycle of Leishmania. https://commons.wikimedia.org/wiki/File:Leishmaniasis_life_cycle_diagram_en.svg

a. Navigate to the mass spec. evidence search. This search returns genes whose protein products mapped to peptides found in proteomics experiments.

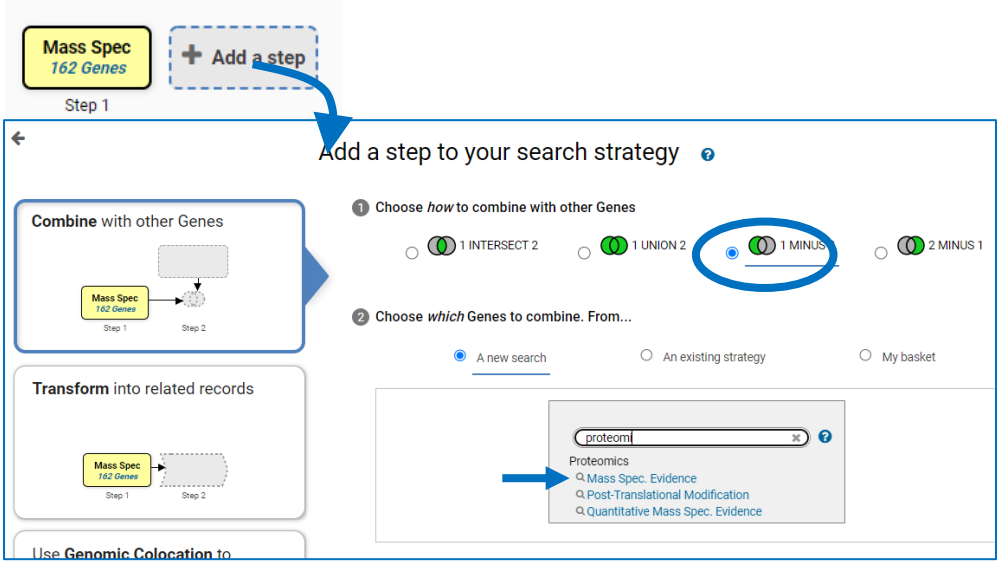


b. Filter the experiment and sample tree by typing a word in the filter box. Select all ***L. infantum*** samples that come from the metacyclic stages. Keep the default search parameters and click on the Get Answer button.

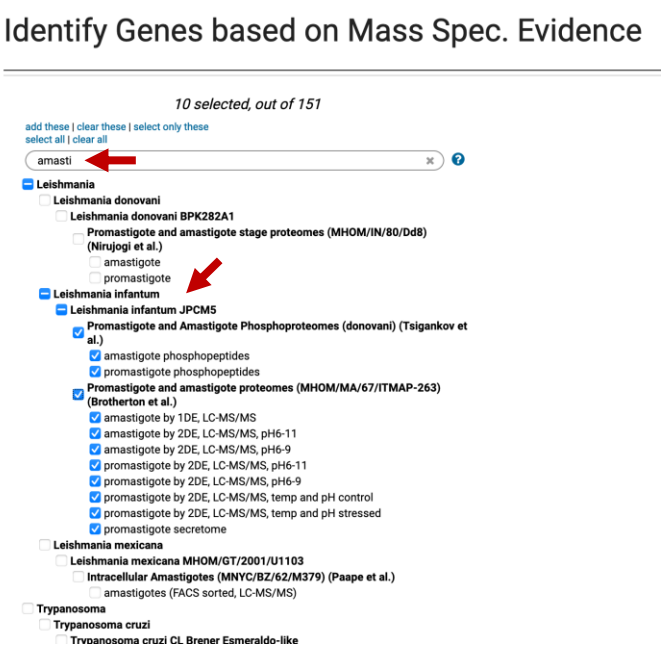


c. How many genes did you get?

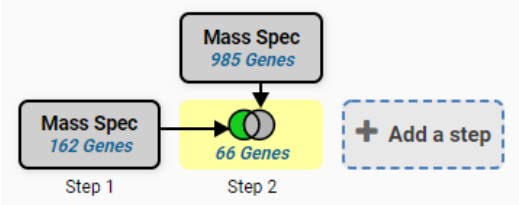
d. Now subtract the genes that have protein expression in the amastigote and promastigote stages. Add a step to your strategy that returns amastigote and promastigote genes and choose the 1 minus 2 operator to combine the searches.



e. Choose all the *L. infantum* samples labeled amastigote and promastigote and run the search

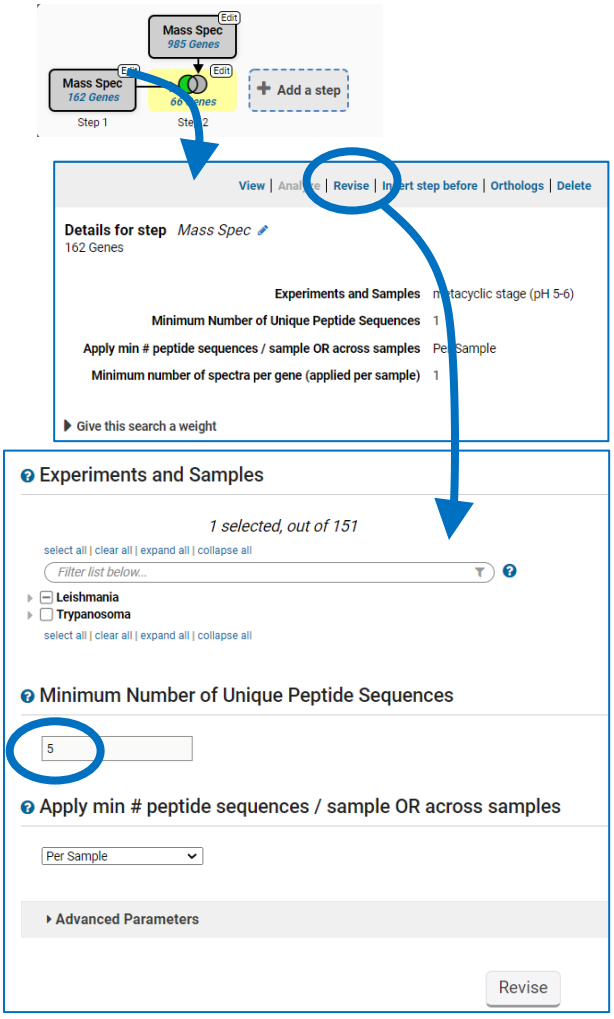


f. Explore the results, do they make sense from a biological standpoint? What does the enrichment analysis of cellular component terms show? Visit the gene pages of some of your results. There you can view mapped peptides and data from other experiments.

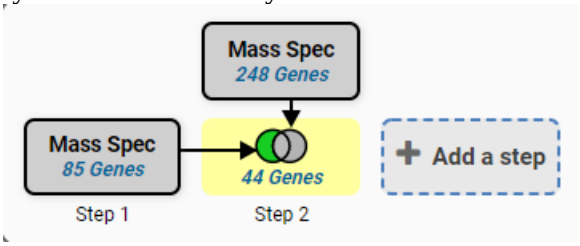


g. How can you increase the stringency of your results? One way is to increase the minimum number of unique peptides that are required to map to a gene before it is returned by the search. The default returns any gene with a minimum of one peptide.

-Click on the edit button
-Click on the revise option in the popup
-Change the value from 1 to 5 and click on the Revise button.
-Remember you need to do this for each step.

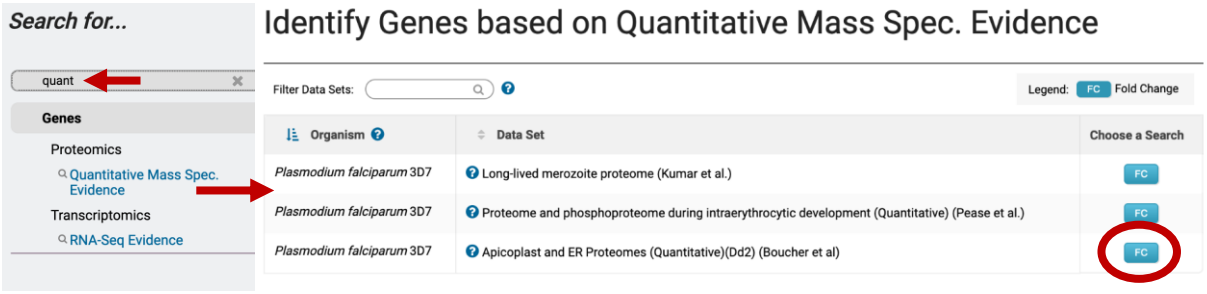


h. How did this change your results? Would you consider these results more stringent?

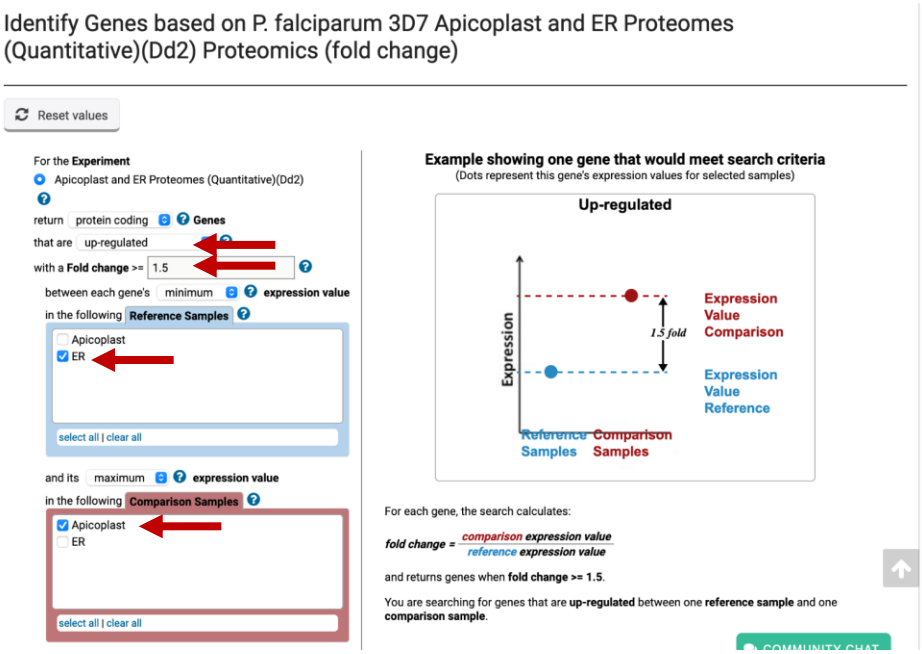


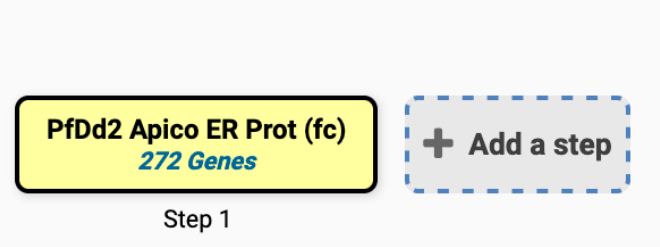
2. Find genes in *Plasmodium falciparum* that are present at a higher concentration in the apicoplast compared to the endoplasmic reticulum (ER). Note for this exercise use https://plasmodb.org

a. Go to the quantitative mass spec evidence searches
b. Select the experiment called Apicoplast and ER Proteomes (Quantitative)(Dd2) (Boucher et al)

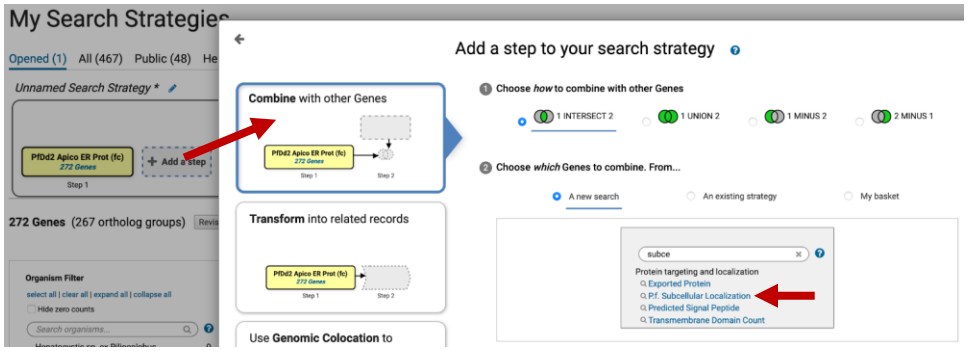


c. Configure this search to return all genes that are upregulated by 1.5 fold in the apicoplast sample compared to the ER sample.

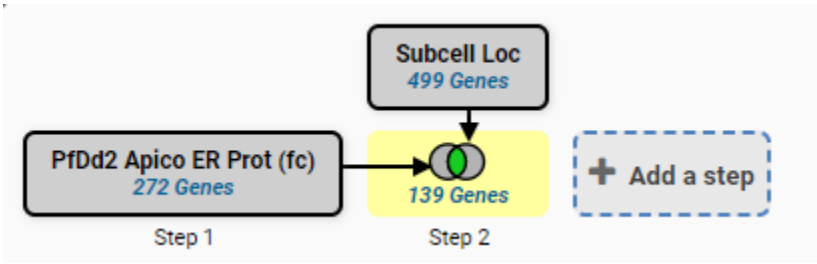




d. Can you further limit your results by leveraging available subcellular localization data? PlasmoDB has a data set that returns genes with the transit peptides that mediate protein targeting to the apicoplast. Click on the add step button and find the subcellular localization search



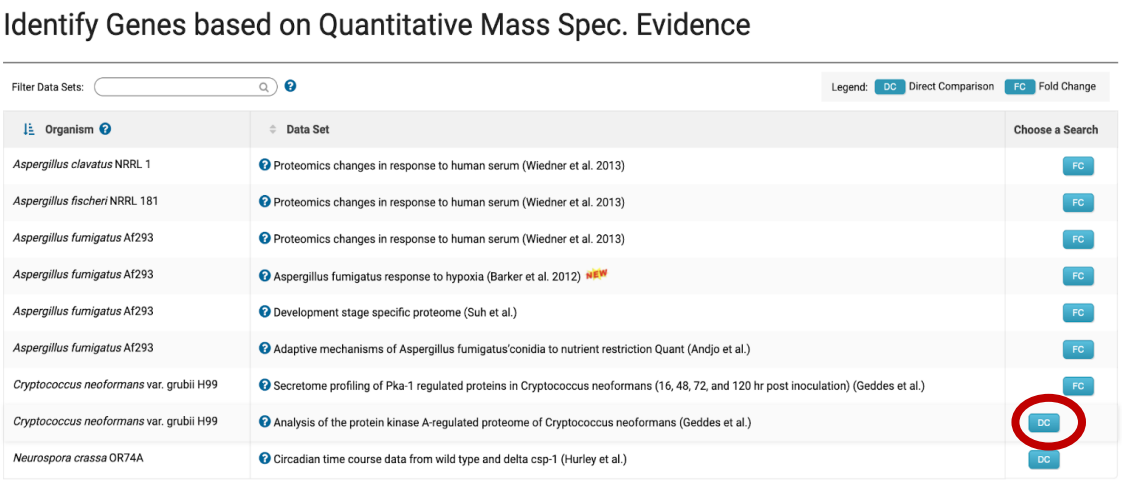
e. Make sure Apicoplast localization is selected and click on the Run Step button. How many genes did you identify? Are you more confident that these genes are apicoplast genes? How would you use the PlasmoDB tools to boost your confidence in these so called apicoplast genes?



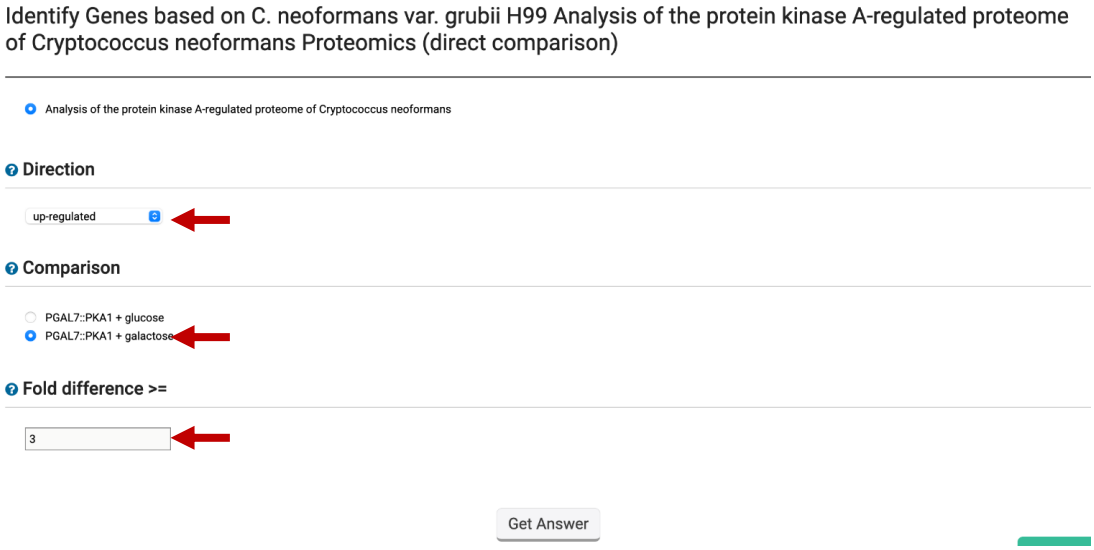
3. Identify *Cryptococcus neoformans* genes that are upregulated in a protein kinase A dependent (PKA) manner and not in a non-PKA dependent manner. Note for this exercise use https://fungidb.org

The expression of virulence factors in *C. neoformans*, including capsule and melanin, is in part regulated by the cyclic-AMP/protein kinase A (cAMP/PKA) signal transduction pathway. *C. neoformans* PGAL7::PKA1 strain can be used to induce the PKA pathway in galactose media and repress the pathway in glucose media.

a. Go to the quantitative proteomic search section and find the experiment called “Analysis of the protein kinase A-regulated proteome of Cryptococcus neoformans (Geddes et al.)”

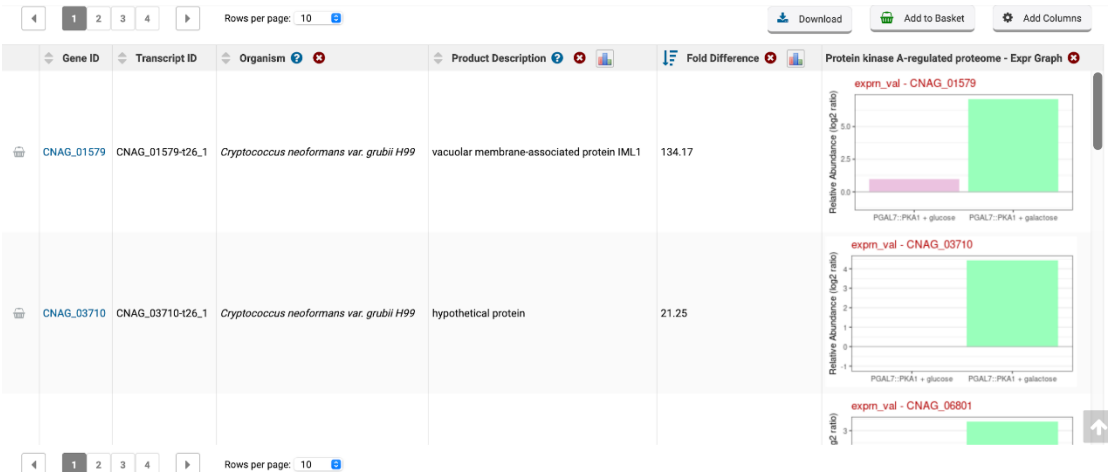


b. Configure the direct comparison search to identify genes that are upregulated by 3 fold in galactose media

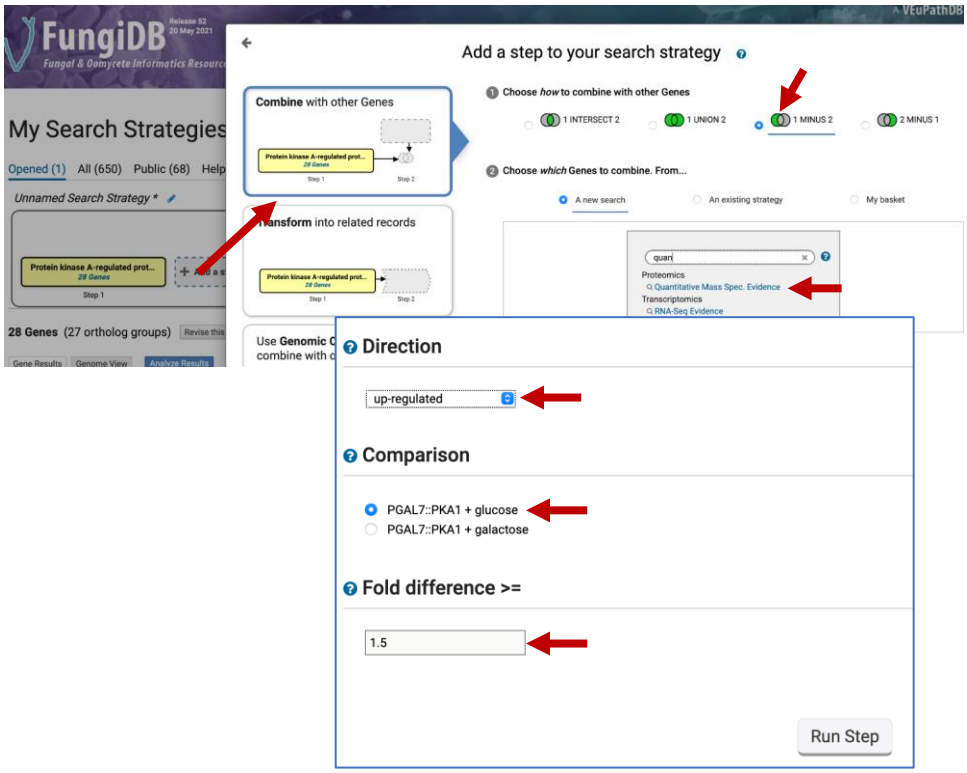


c. How many genes did you get?

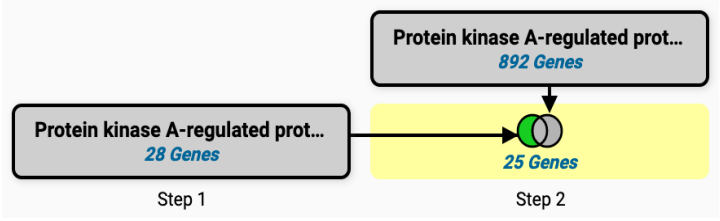
d. Explore your results. Do the expression graphs meet the criteria you selected?



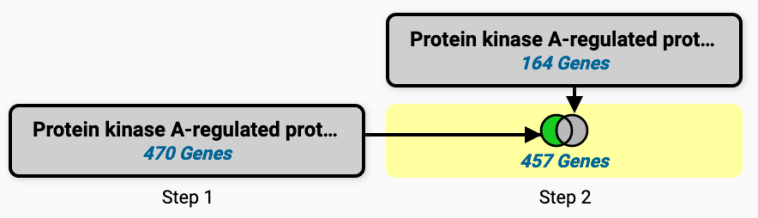
e. Add a step and remove from this list any gene that is upregulated by 1.5 fold in glucose media.



f. How many genes did you get?



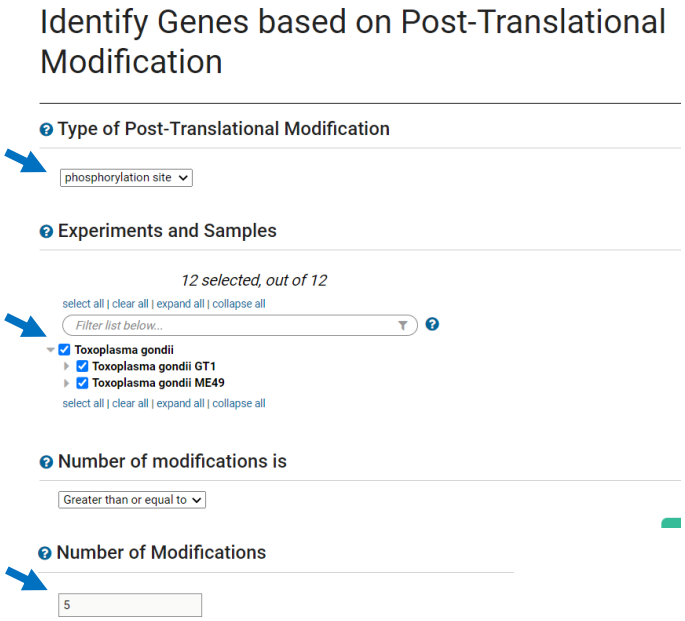
g. Can you reconfigure the above searches to identify genes that are downregulated as opposed to upregulated? Did your results change?



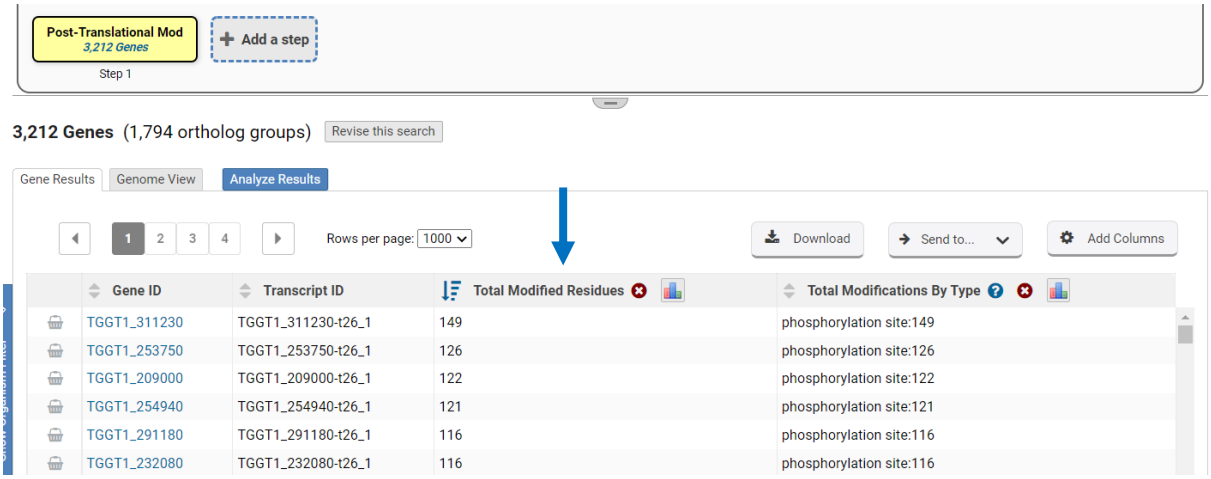
4. Find genes with evidence of protein phosphorylation in intracellular *Toxoplasma* tachyzoites. Note for this exercise use https://toxodb.org

Although phosphorylated peptides can be identified by searching the appropriate experiments in the Mass Spec Evidence search page, VEuPathDB also contains a search that specifically access proteomics data that determined post-translational modifications. Find all genes that have at least 5 phosphorylation events from all available phosphoproteomic experiments.

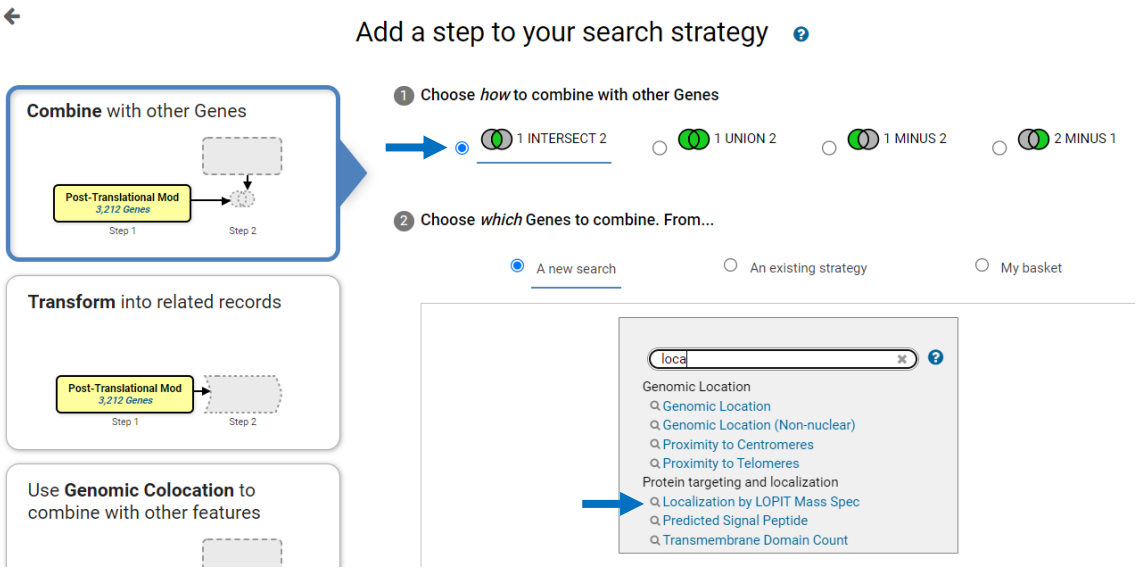
a. Navigate to the Post-Translational Modification search.
b. Next make sure to change the number of modifications to 5.



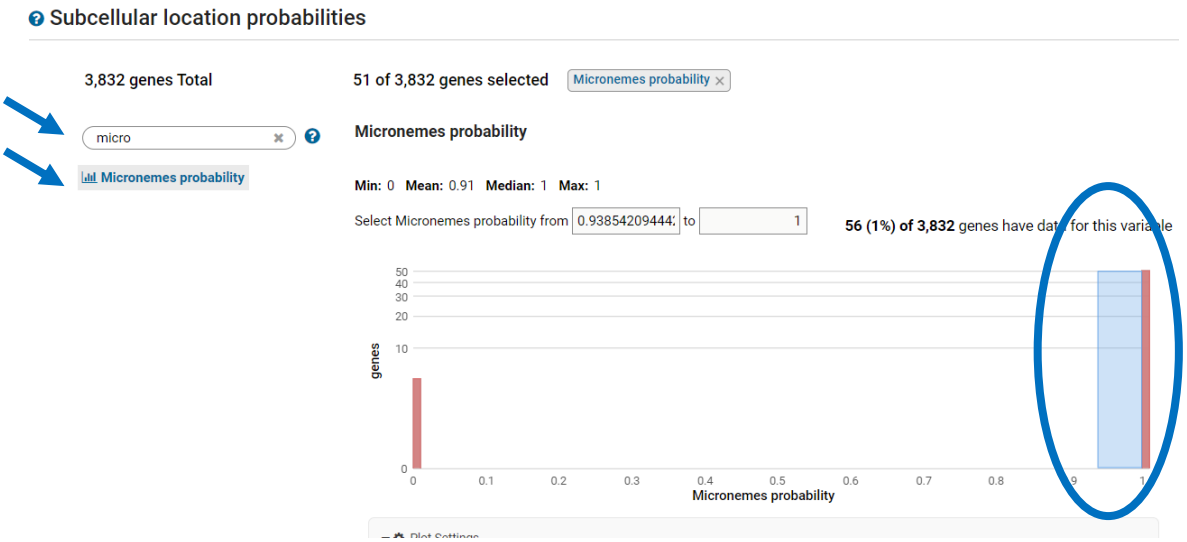
c. How many genes did you return? Which gene has the highest number of phosphorylation sites? Hint, examine the column called total modified residues.

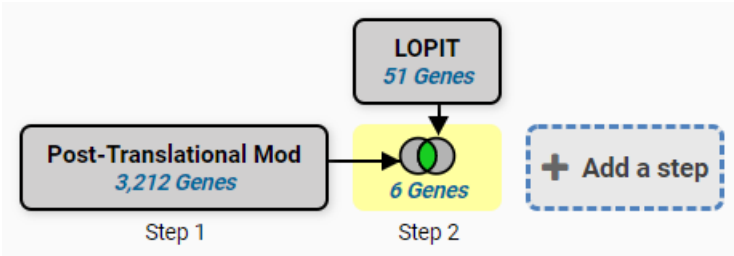


d. How many of these phosphorylated genes are also localized to the microneme organelles? ToxoDB has hyperLOPIT data (Hyperplexed Localisation of Organelle Proteins by Isotope Tagging),) a spatial proteomics method that simultaneously captures the steady-state subcellular association of thousands of proteins. The technique reveals the probability that a protein is present in a specific cellular location (fraction). Use this data and search to find which phosphorylated genes are localized to the microneme.



e. Filter the localization categories using the word microneme. Select all genes with a probability of 1 (or close to 1) and click Run Step.


Subcellular location probabilities
3,832 genes Total
51 of 3,832 genes selected
Micronemes probability
micro
Micronemes probability
Min: 0 Mean: 0.91 Median: 1 Max: 1
Select Micronemes probability from 0.93854209444; to 1
56 (1%) of 3,832 genes have dat for this variable
genes
Micronemes probability


LOPIT
51 Genes
Post-Translational Mod
3,212 Genes
6 Genes
Add a step
Step 1
Step 2