Exploring proteomics data in VEuPathDB Resources

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

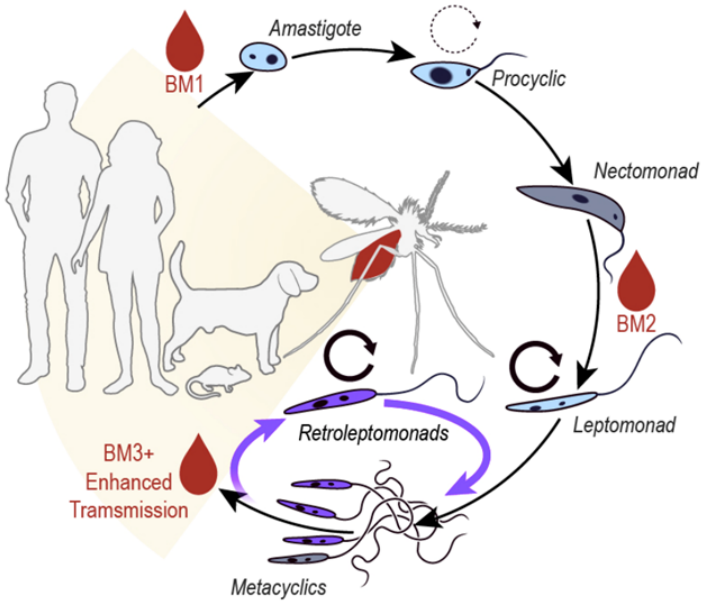
1. **Mass spec. evidence**
   Peptides from proteomics experiments are mapped to a reference genome enabling searches for genes based on that mapping.
2. **Quantitative mass spec. evidence**
   Data from quantitative proteomic experiments are loaded and made available for searching based on fold change 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:

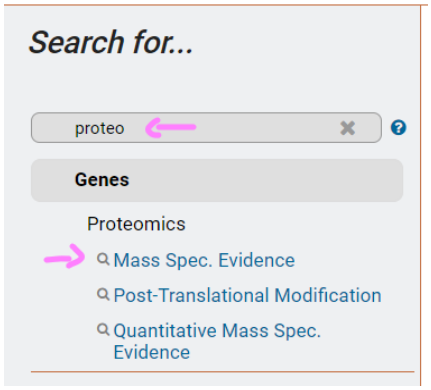
- Understand the different categories of proteomics data
- Learn how to run searches to identify genes based on peptide evidence
- Learn how to identify differentially expressed genes based on quantitative data
- Learn how to identify genes with different PTMs

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

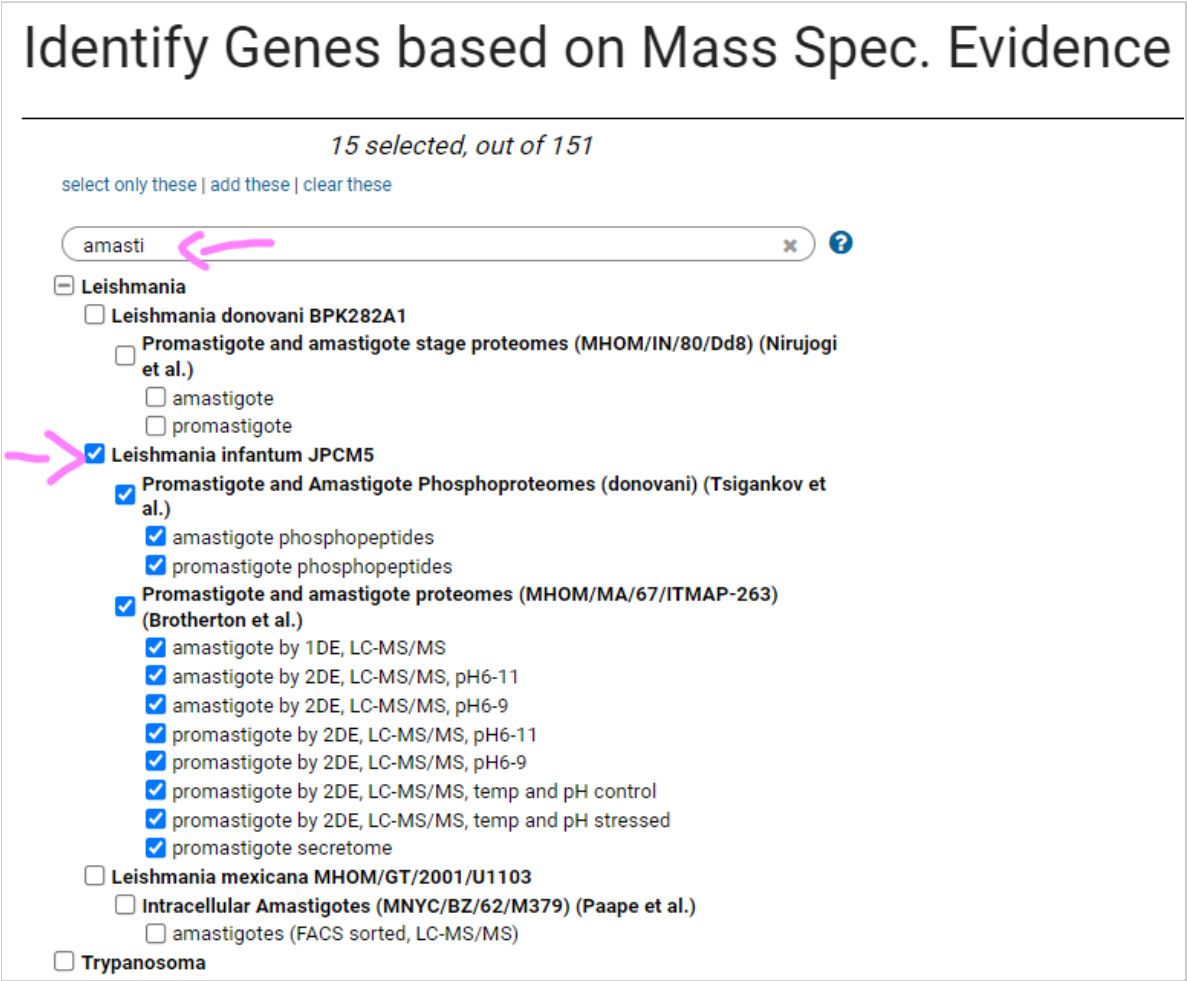


Serafim, T. D., Coutinho-Abreu, I. V., Oliveira, F., Meneses, C., Kamhawi, S., & Valenzuela, J. G. (2018). Sequential blood meals promote *Leishmania* replication and reverse metacyclogenesis augmenting vector infectivity. *Nature microbiology*, *3*(5), 548–555. https://doi.org/10.1038/s41564-018-0125-7

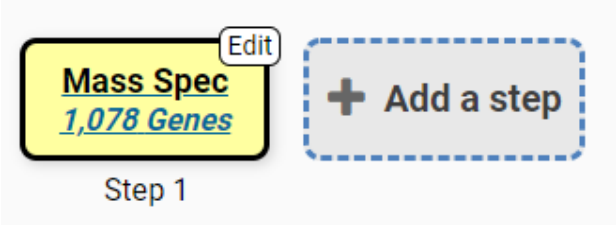
   a. Navigate to the mass spec. evidence search. How did you find it? You can use the search filter on the left of the home page or in the searches menu at the top of the page. Filter the searches by typing a word in the filter box.



b. Select all *L. infantum* samples that come from the amastigote or promastigote stages. Note that you can filter the samples with key words like amastigote.
c. Keep the default search parameters and click on the Get Answer button.



d. How many genes did you get?



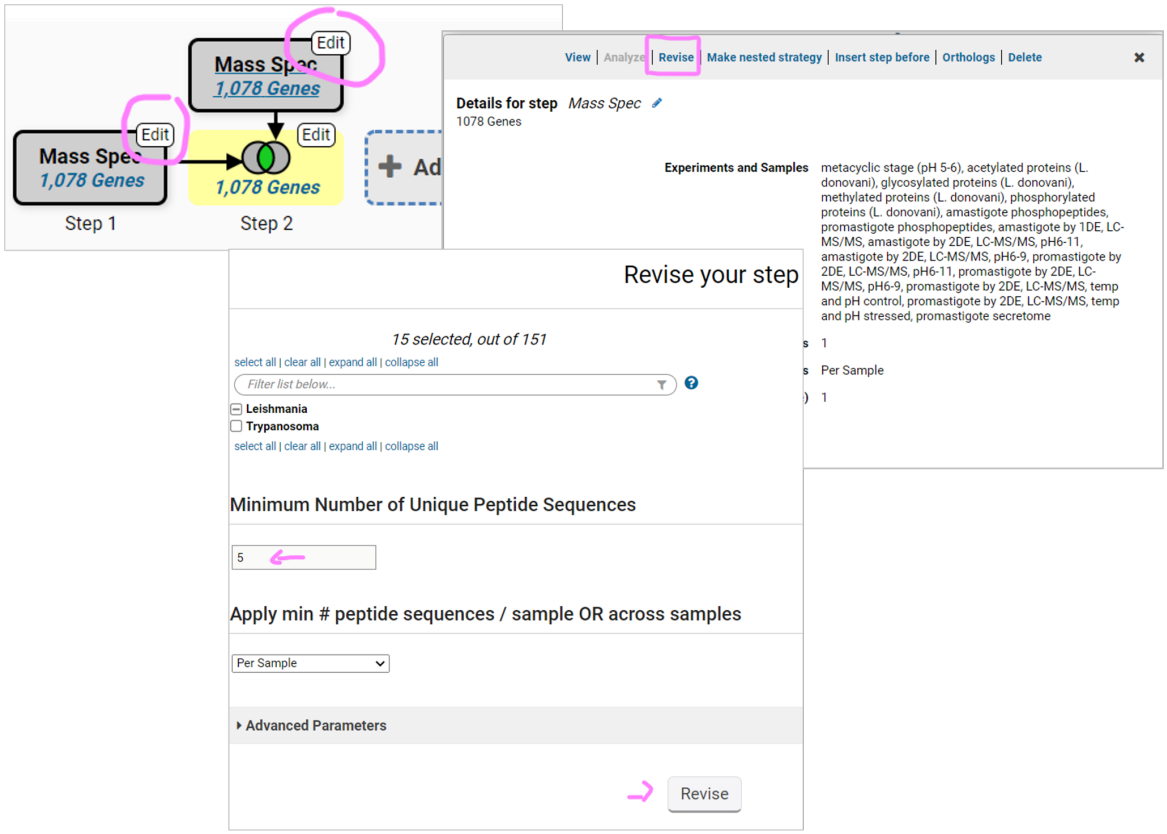
e. Extend the strategy to return genes with protein expression in metacyclic stage but not the amastigote or promastigoe stage. (Remove the above results from genes with peptide evidence from the metacyclic stage of *L. infantum*? Try the following:
    - Click on add step
    - Select how to combine the results

- Find and click on the mass spec. evidence search
- Select the metacyclic stage proteome data and click on the Get Answer button.

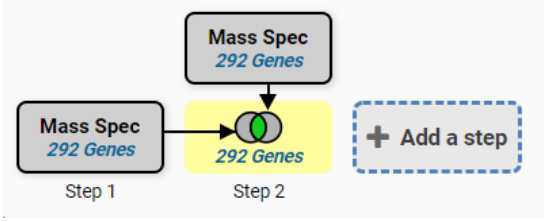
f. How many genes did you get? Explore the results, do they make sense from a biological standpoint?

g. How can you increase the stringency of your results? One way is to increase the minimum number of unique peptides. The default returns any gene with a minimum of one peptide. What happens if you change this to a minimum of 5 peptides in both steps?

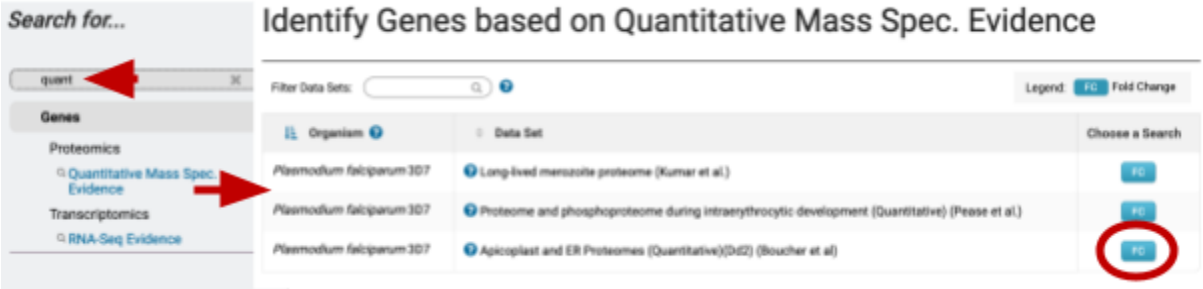
- Click on the edit button
- Click on the revise option in the popup
- Change the value from 1 to 5 and click on the Revise button.
- Remember you need to do this for each step.



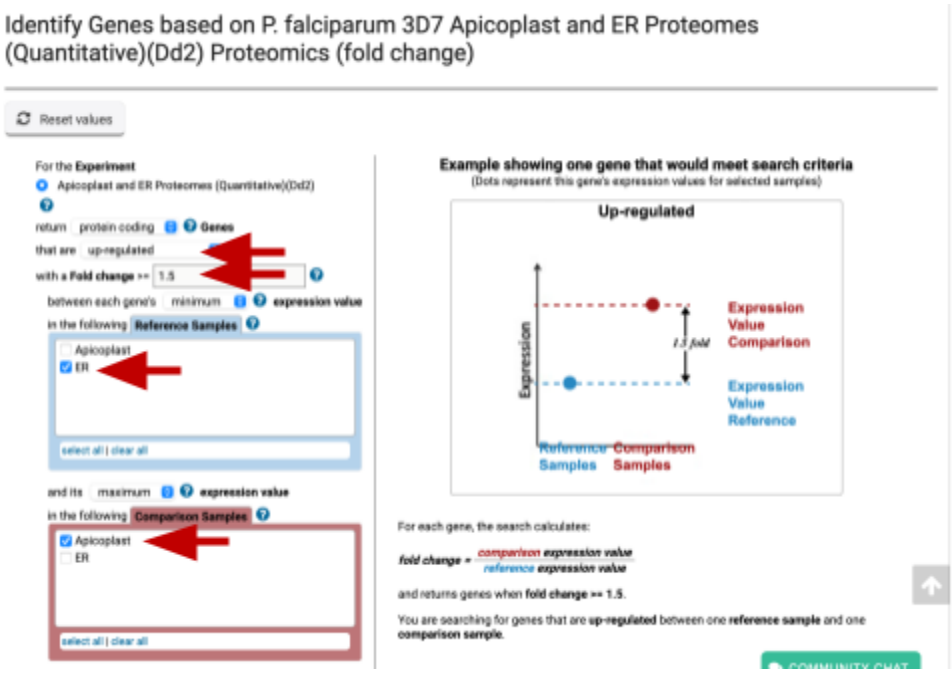
h. How did this change your results? Would you consider these results more stringent?

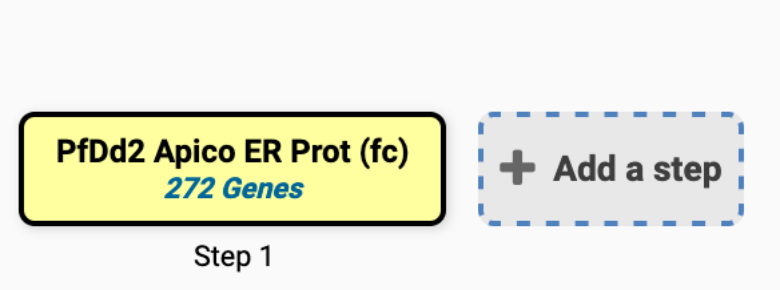


2. **Find genes in *Plasmodium falciparum* that are quantitatively present at a higher concentration in the apicoplast compared to the endoplasmic reticulum (ER). Note for this exercise use https://plasmodb.org**
    a. Go to the quantitative mass spec evidence searches
    b. Select the experiment called Apicoplast and ER Proteomes (Quantitative)(Dd2) (Boucher et al)



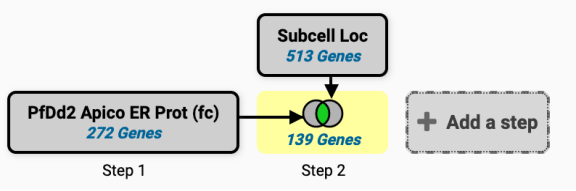
    c. Configure this search to return all genes that are upregulated by 1.5 fold in the apicoplast sample compared to the ER sample
    d. How many genes did your search return?





e. Can you further limit your results by leveraging available subcellular localization data?
   - Click on the add step button and find the subcellular localization search

   - Make sure Apicoplast localization is selected and click on the Run Step button. How many genes did you identify?

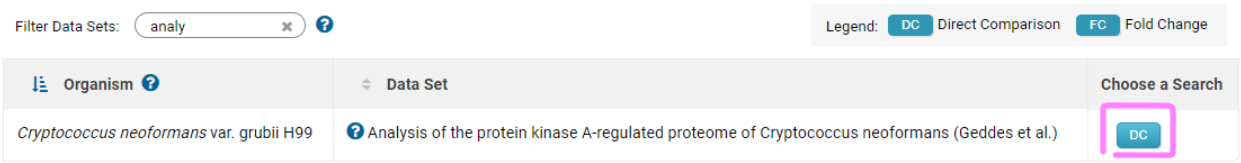


3. Identify *Cryptococcus neoformans* genes that are upregulated in a protein kinase A dependent (PKA) manner and not in a non-PKA dependent manner.
Note for this exercise use https://fungidb.org

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.)

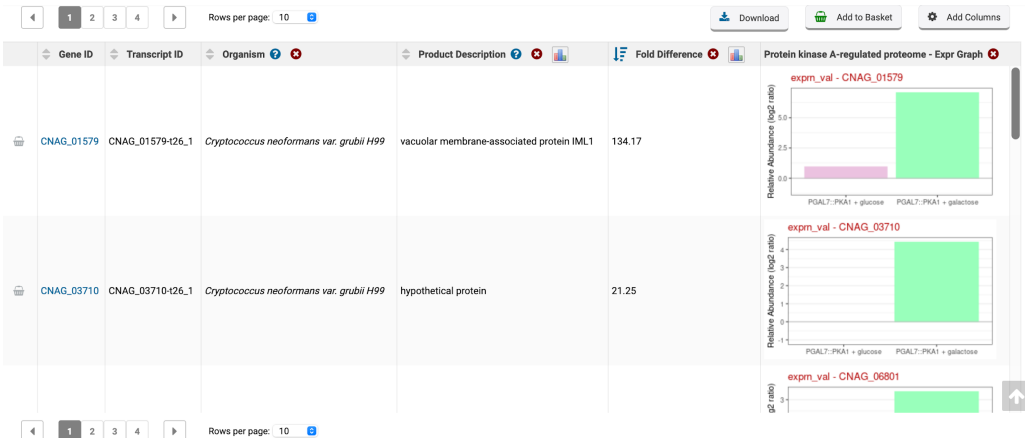
Identify Genes based on Quantitative Mass Spec. Evidence



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

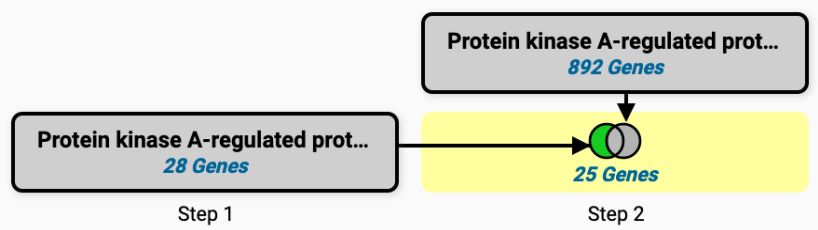
c. How many genes did you get?

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

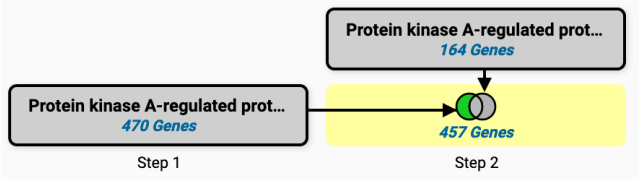


e. Add a step and remove from this list any gene that is upregulated by 1.5 fold in glucose media.

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

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

   a. Find all genes that have at least 5 phosphorylation events from all available phosphoproteomic experiments. Navigate to the Post-Translational Modification search. Filter the samples using the key word *phospho* then select all the samples that are phosphopeptide enriched.
   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.