

## 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)**

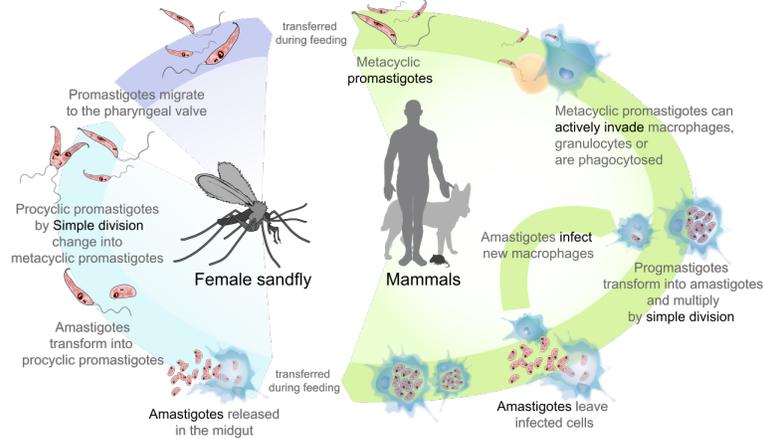
*PTM data from proteomics experiments are loaded on genes enabling searches for genes based on the type and number of the PTM.*

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

- 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](https://commons.wikimedia.org/wiki/File:Leishmaniasis_life_cycle_diagram_en.svg)

- 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.

## Identify Genes based on Mass Spec. Evidence

10 selected, out of 151

add these | clear these | select only these  
select all | clear all

amasti

- Leishmania
  - Leishmania donovani
    - Leishmania donovani BPK282A1
      - Promastigote and amastigote stage proteomes (MHOM/IN/80/Dd8) (Nirujogi et al.)
        - amastigote
        - promastigote
  - Leishmania infantum
    - Leishmania infantum JPCM5
      - Promastigote and Amastigote Phosphoproteomes (donovani) (Tsigankov et al.)
        - amastigote phosphopeptides
        - promastigote phosphopeptides
      - Promastigote and amastigote proteomes (MHOM/MA/67/ITMAP-263) (Brotherton et al.)
        - amastigote by 1DE, LC-MS/MS
        - amastigote by 2DE, LC-MS/MS, pH6-11
        - amastigote by 2DE, LC-MS/MS, pH6-9
        - promastigote by 2DE, LC-MS/MS, pH6-11
        - promastigote by 2DE, LC-MS/MS, pH6-9
        - promastigote by 2DE, LC-MS/MS, temp and pH control
        - promastigote by 2DE, LC-MS/MS, temp and pH stressed
        - promastigote secretome
  - Leishmania mexicana
    - Leishmania mexicana MHOM/GT/2001/U1103
      - Intracellular Amastigotes (MNYC/BZ/62/M379) (Paape et al.)
        - amastigotes (FACS sorted, LC-MS/MS)
  - Trypanosoma
    - Trypanosoma cruzi
      - Trypanosoma cruzi CL Brener Esmeraldo-like

- c. Keep the default search parameters and click on the Get Answer button.

Minimum Number of Unique Peptide Sequences

1

Apply min # peptide sequences / sample OR across samples

Per Sample

Advanced Parameters

Get Answer

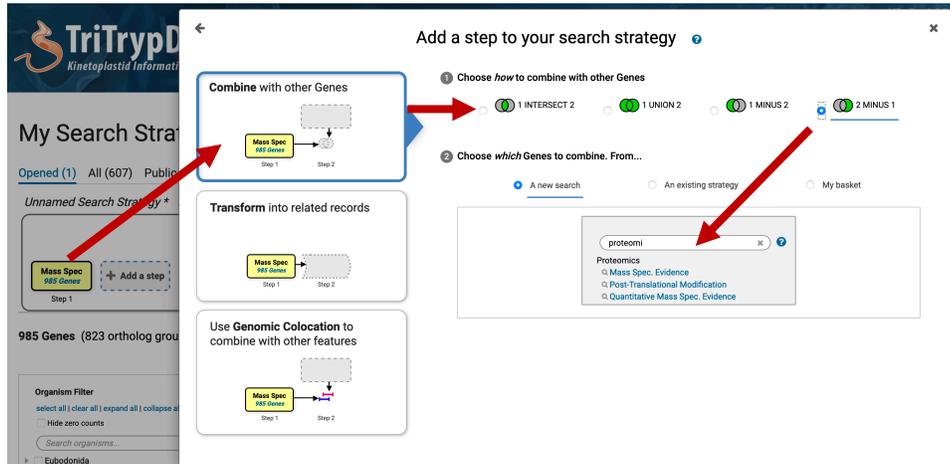
- d. How many genes did you get?

Mass Spec  
985 Genes

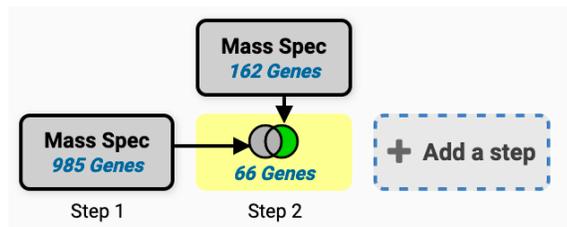
+ Add a step

Step 1

- e. Can you remove these results from any gene 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?



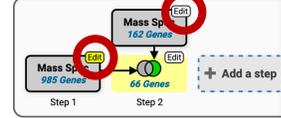
- g. How can you increase the stringency of your results? One way is to increase the minimum number of unique peptides. The default returns any gene with a minimum of one peptide. What happens if you change this to a minimum of 5 peptides in both steps?
- Click on the edit button
  - Click on the revise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

## My Search Strategies

Opened (1) All (607) Public (41) Help

Unnamed Search Strategy \*



66 Genes (63 ortholog groups)

View | Analyze | **Revise** | Insert step before | Orthologs | Delete

**Details for step** Mass Spec

985 Genes

**Experiments and Samples** amastigote phosphopeptides, promastigote phosphopeptides, amastigote by 1DE, LC-MS/MS, amastigote by 2DE, LC-MS/MS, pH6-11, amastigote by 2DE, LC-MS/MS, pH6-9, promastigote by 2DE, LC-MS/MS, pH6-11, promastigote by 2DE, LC-MS/MS, pH6-9, promastigote by 2DE, LC-MS/MS, temp and pH control, promastigote by 2DE, LC-MS/MS, temp and pH stressed, promastigote secretome

Minimum Number of Unique Peptide Sequences 1

Apply min # peptide sequences / sample OR across samples Per Sample

Organism Filter  
select all | clear all  
 Hide zero counts

### Revise your step

**Experiments and Samples**

10 selected, out of 151

select all | clear all | expand all | collapse all

Filter list below...

- Leishmania
- Trypanosoma

select all | clear all | expand all | collapse all

**Minimum Number of Unique Peptide Sequences**

5

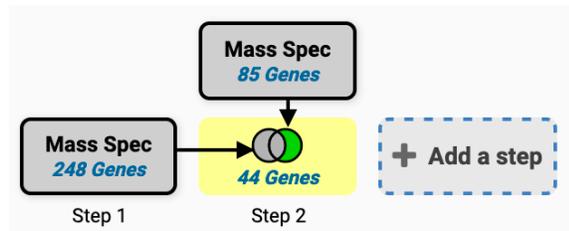
**Apply min # peptide sequences / sample OR across samples**

Per Sample

Advanced Parameters

**Revise**

- 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)

Search for... Identify Genes based on Quantitative Mass Spec. Evidence

quant

Filter Data Sets:  Legend: FC Fold Change

Organism	Data Set	Choose a Search
<i>Plasmodium falciparum</i> 3D7	Long-lived merozoite proteome (Kumar et al.)	FC
<i>Plasmodium falciparum</i> 3D7	Proteome and phosphoproteome during intraerythrocytic development (Quantitative) (Pease et al.)	FC
<i>Plasmodium falciparum</i> 3D7	Apicoplast and ER Proteomes (Quantitative)(Dd2) (Boucher et al)	FC

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

Identify Genes based on P. falciparum 3D7 Apicoplast and ER Proteomes (Quantitative)(Dd2) Proteomics (fold change)

Reset values

For the Experiment  
 Apicoplast and ER Proteomes (Quantitative)(Dd2)

return protein coding  Genes  
 that are up-regulated  
 with a Fold change  $\geq 1.5$   
 between each gene's minimum expression value  
 in the following Reference Samples  
 Apicoplast  
 ER  
 and its maximum expression value  
 in the following Comparison Samples  
 Apicoplast  
 ER

**Example showing one gene that would meet search criteria**  
 (Dots represent this gene's expression values for selected samples)

**Up-regulated**

For each gene, the search calculates:  

$$\text{fold change} = \frac{\text{comparison expression value}}{\text{reference expression value}}$$
 and returns genes when  $\text{fold change} \geq 1.5$ .  
 You are searching for genes that are up-regulated between one reference sample and one comparison sample.

- d. How many genes did your search return?

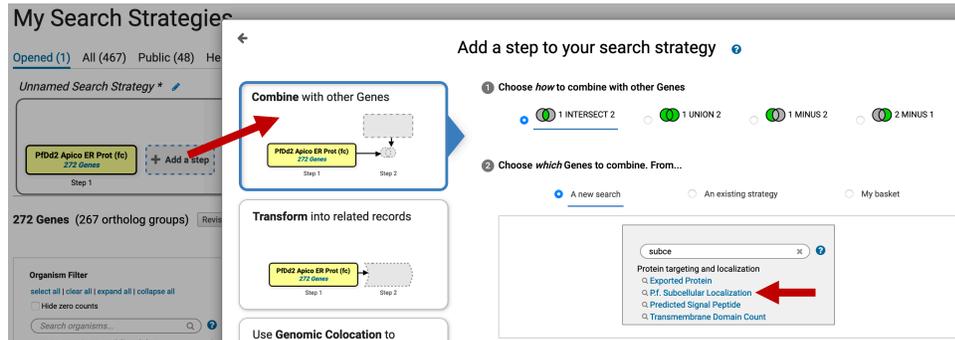
**PfDd2 Apico ER Prot (fc)**  
**272 Genes**

+ Add a step

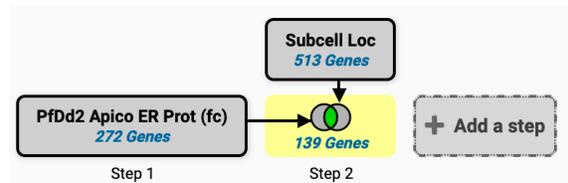
Step 1

- 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. I 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.

- 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

Organism	Data Set	Choose a Search
<i>Aspergillus clavatus</i> NRRL 1	Proteomics changes in response to human serum (Wiedner et al. 2013)	FC
<i>Aspergillus fischeri</i> NRRL 181	Proteomics changes in response to human serum (Wiedner et al. 2013)	FC
<i>Aspergillus fumigatus</i> Af293	Proteomics changes in response to human serum (Wiedner et al. 2013)	FC
<i>Aspergillus fumigatus</i> Af293	<i>Aspergillus fumigatus</i> response to hypoxia (Barker et al. 2012)	FC
<i>Aspergillus fumigatus</i> Af293	Development stage specific proteome (Suh et al.)	FC
<i>Aspergillus fumigatus</i> Af293	Adaptive mechanisms of <i>Aspergillus fumigatus</i> /conidia to nutrient restriction Quant (Andjo et al.)	FC
<i>Cryptococcus neoformans</i> var. grubii H99	Secretome profiling of Pka-1 regulated proteins in <i>Cryptococcus neoformans</i> (16, 48, 72, and 120 hr post inoculation) (Geddes et al.)	FC
<i>Cryptococcus neoformans</i> var. grubii H99	Analysis of the protein kinase A-regulated proteome of <i>Cryptococcus neoformans</i> (Geddes et al.)	DC
<i>Neurospora crassa</i> OR74A	Circadian time course data from wild type and delta csp-1 (Hurley et al.)	DC

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

Identify Genes based on *C. neoformans* var. *grubii* H99 Analysis of the protein kinase A-regulated proteome of *Cryptococcus neoformans* Proteomics (direct comparison)

Analysis of the protein kinase A-regulated proteome of *Cryptococcus neoformans*

**Direction**

**Comparison**  
 PGAL7:PKA1 + glucose  
 PGAL7:PKA1 + galactose

**Fold difference >=**

[Get Answer](#)

c. How many genes did you get?

**Protein kinase A-regulated prot...**  
 28 Genes

[+ Add a step](#)

Step 1

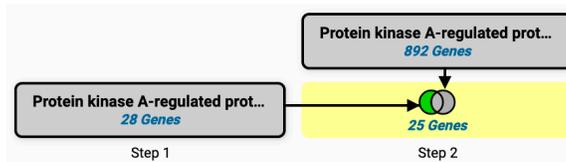
d. Explore your results. Do the expression graphs meet the criteria you selected?

Gene ID	Transcript ID	Organism	Product Description	Fold Difference	Protein kinase A-regulated proteome - Expr Graph
CNAG_01579	CNAG_01579-126_1	<i>Cryptococcus neoformans</i> var. <i>grubii</i> H99	vacuolar membrane-associated protein IML1	134.17	
CNAG_03710	CNAG_03710-126_1	<i>Cryptococcus neoformans</i> var. <i>grubii</i> H99	hypothetical protein	21.25	
CNAG_06801					

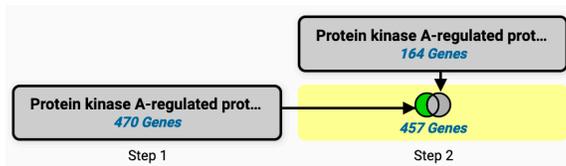
- e. Add a step and remove from this list any gene that is upregulated by 1.5 fold in glucose media.

The screenshot shows the FungiDB search strategy editor. The main window is titled 'Add a step to your search strategy'. It has two main sections: '1 Choose how to combine with other Genes' and '2 Choose which Genes to combine. From...'. In the first section, '1 MINUS 2' is selected. In the second section, 'A new search' is selected. A search box contains 'quant' and a dropdown menu shows 'Proteomics', 'Quantitative Mass Spec. Evidence', 'Transcriptomics', and 'RNA-Seq Evidence'. A 'Direction' dialog box is open, showing 'up-regulated' selected, 'PGAL7::PKA1 + glucose' selected for comparison, and '1.5' entered for fold difference. A 'Run Step' button is at the bottom right of the dialog.

- f. How many genes did you get?



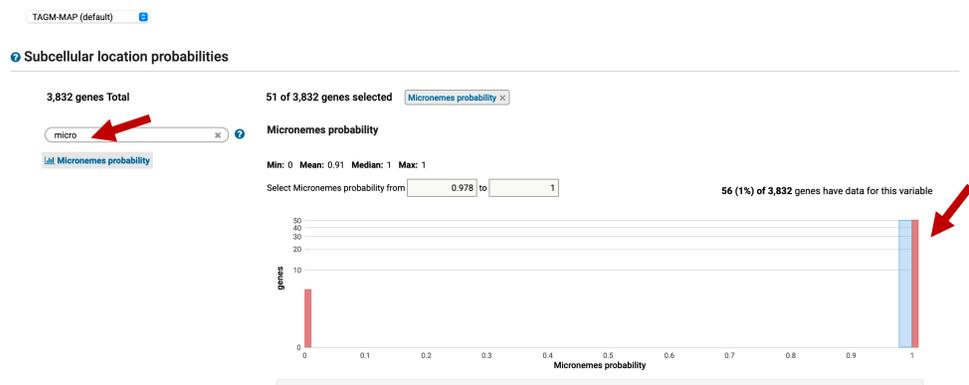
- g. Can you reconfigure the above searches to identify genes that are downregulated as opposed to upregulated? Did your results change?





- Add a step and locate the Protein Targeting and Localization searches. Select the one called Localization by LOPIT Mass Spec.

- Filter the localization categories using the word microneme. Select all genes with a probability of 1 (or close to 1).



- Explore your results.