

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.

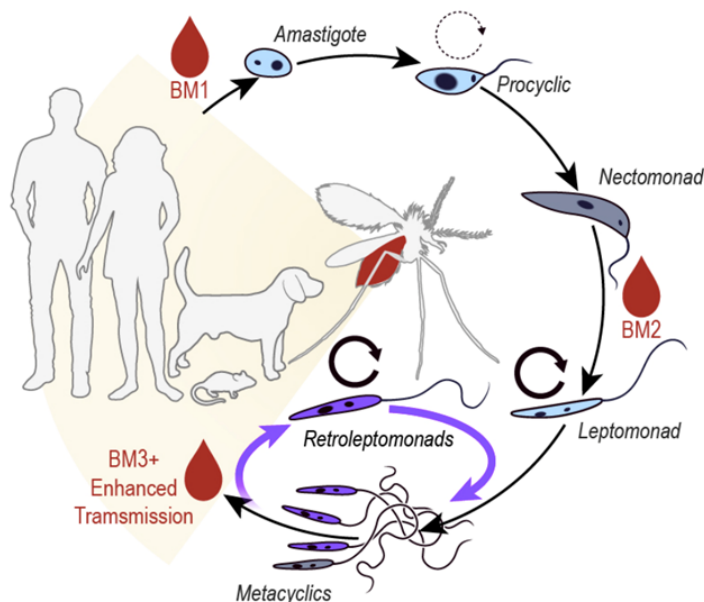
The exercises below explore the different categories and searches available for proteomics in VEuPathDB.

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

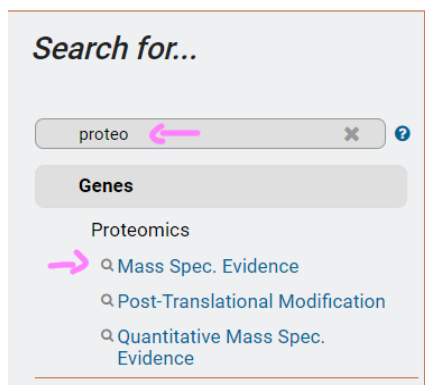
1. Find genes that have peptide evidence from metacyclic stages but not amastigote or promastigote stages of *Leishmania infantum*.

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.

Identify Genes based on Mass Spec. Evidence

15 selected, out of 151

[select only these](#) | [add these](#) | [clear these](#)

amasti ✕ ?

☐ **Leishmania**

- ☐ **Leishmania donovani BPK282A1**
 - ☐ Promastigote and amastigote stage proteomes (MHOM/IN/80/Dd8) (Nirujogi et al.)
 - ☐ amastigote
 - ☐ promastigote
- ☒ **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 MHOM/GT/2001/U1103**
 - ☐ Intracellular Amastigotes (MNYC/BZ/62/M379) (Paape et al.)
 - ☐ amastigotes (FACS sorted, LC-MS/MS)
- ☐ **Trypanosoma**

- d. How many genes did you get?

Edit

Mass Spec

1,078 Genes

+

Add a step

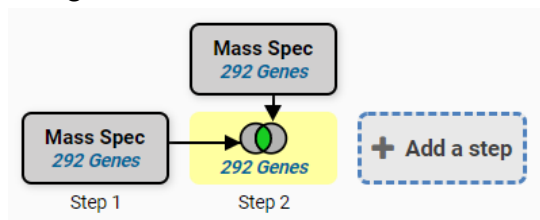
Step 1

- e. Extend the strategy to return genes with protein expression in metacyclic stage but not the amastigote or promastigote 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.

The screenshot shows the 'Revise your step' dialog box in the software. The dialog has a title bar with 'View | Analyze | Revise | Make nested strategy | Insert step before | Orthologs | Delete'. The main content area shows 'Details for step Mass Spec' with '1078 Genes'. Below this, there is a section titled 'Experiments and Samples' with a list of experiments: 'metacyclic stage (pH 5-6), acetylated proteins (L. donovani), glycosylated proteins (L. donovani), methylated proteins (L. donovani), phosphorylated proteins (L. donovani), 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'. The 'Revise' button is highlighted in the top right corner of the dialog.

- 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...

quant

Genes

Proteomics

- Quantitative Mass Spec. Evidence

Transcriptomics

- RNA-Seq Evidence

Identify Genes based on Quantitative Mass Spec. Evidence

Filter Data Sets:

Legend: FC Fold Change

Organism	Data Set	Chose a Search
Plasmodium falciparum 3D7	1 Long-lived merozoite proteome (Kumar et al.)	FC
Plasmodium falciparum 3D7	2 Proteome and phosphoproteome during intraerythrocytic development (Quantitative) (Pease et al.)	FC
Plasmodium falciparum 3D7	3 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
- d. How many genes did your search return?

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 ≥ 1.5

between each gene's minimum expression value is the following

Reference Samples

- Apicoplast
- ☒ ER

select all | clear all

and its maximum expression value is the following

Comparison Samples

- ☒ Apicoplast
- ER

select all | clear all

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

Up-regulated

Expression

Expression Value Comparison

Expression Value Reference

1.5 fold

Reference Samples Comparison Samples

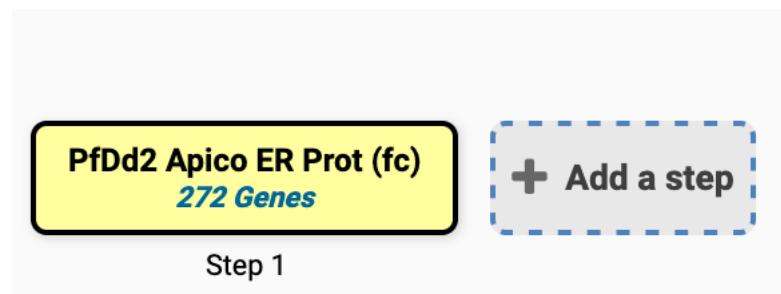
For each gene, the search calculates:

$$\text{fold change} = \frac{\text{comparison expression value}}{\text{reference expression value}}$$

and returns genes when fold change ≥ 1.5 .

You are searching for genes that are up-regulated between one reference sample and one comparison sample.

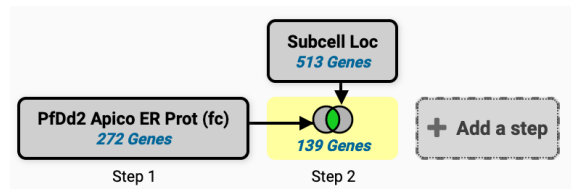
CONTINUE



- 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

Filter Data Sets:

Legend: **DC** Direct Comparison **FC** Fold Change

Organism	Data Set	Choose a Search
<i>Cryptococcus neoformans</i> var. <i>grubii</i> H99	Analysis of the protein kinase A-regulated proteome of <i>Cryptococcus neoformans</i> (Geddes 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

☒ up-regulated

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

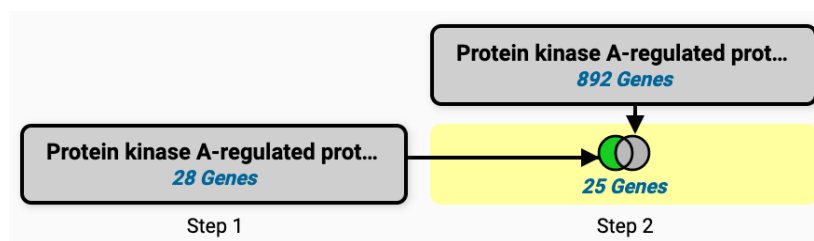
- d. Explore your results. Do the expression graphs meet the criteria you selected?

Gene ID	Transcript ID	Organism	Product Description	Fold Difference	Protein kinase A-regulated proteome - Expr Graph
CNAG_01579	CNAG_01579-t26.1	<i>Cryptococcus neoformans</i> var. <i>grubii</i> H99	vacuolar membrane-associated protein IML1	134.17	
CNAG_03710	CNAG_03710-t26.1	<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.



- 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.

- 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.
- Next make sure to change the number of modifications to 5.
- How many genes did you return? Which gene has the highest number of phosphorylation sites? Hint, examine the column called total modified residues.

5 selected, out of 9

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

phospho

Toxoplasma gondii

Toxoplasma gondii GT1

☒ Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Trecek et al.)

☐ Infected host cell, phosphopeptide-depleted (peptide discovery against TgGT1)

☒ Infected host cell, phosphopeptide-enriched (peptide discovery against TgGT1)

☐ Purified tachyzoites phosphopeptide-depleted (peptide discovery against TgGT1)

☒ Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgGT1)

Toxoplasma gondii ME49

☒ Tachyzoite phosphoproteome - Calcium dependent (RH) (Nebel et al.)

☒ phosphopeptide-enriched (via Mascot)

☐ phosphopeptide-depleted (via Sequest)

☒ phosphopeptide-enriched (via Sequest)

☒ Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Trecek et al.)

☒ Infected host cell, phosphopeptide-enriched (peptide discovery against TgME49)

☐ Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgME49)

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

Number of modifications is

Greater than or equal to

Number of Modifications

5

Post-Translational Modification Search

3,266 Genes (1,531 ortholog groups)

Organism Filter: select all | clear all | expand all | collapse all

File attachments

Download | Remove | Add to basket | Add columns

Gene ID	Transcript ID	Modified Residues	Modifications By Type	Protein Description
TG0211_011230	TG0211_011230	5136, 5166, 5174, 5221, 5343, 5343, 7248, 7396, 5404, 7405, 7515, 5408, 5503, 5870, 5903...	phosphorylation site128	Hypothetical protein
TG0211_293750	TG0211_293750	4125, 5131, 7170, 5213, 5222, 5342, 7343, 5275, 5276, 5293, 5294, 5385, 5384, 5385...	phosphorylation site115	PLU1 family protein
TG0211_293000	TG0211_293000	51305, 51348, 51350, 51354, 51355, 51354, 51481, 51485, 51413, 51417, 71416, 51556, 7...	phosphorylation site113	HISTubrain (ubiquitin transferase) domain-containing protein
TG0211_254940	TG0211_254940	5183, 5127, 5143, 5254, 5478, 5708, 5784, 5791, 5889, 5812, 6253...	phosphorylation site155	SW43 domain-containing protein
TG0211_291180	TG0211_291180	5116, 5246, 5295, 5295, 5363, 5363, 5364, 7276, 5271, 7395, 5384, 5388, 7341, 5342...	phosphorylation site103	Hypothetical protein
TG0211_233080	TG0211_233080	51026, 7440, 5442, 5465, 5903, 5948, 5498, 5990, 51842, 51844, 51038, 51061, 51030, 7...	phosphorylation site103	Hypothetical protein