

Exploring proteomics data in VEuPathDB Resources

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

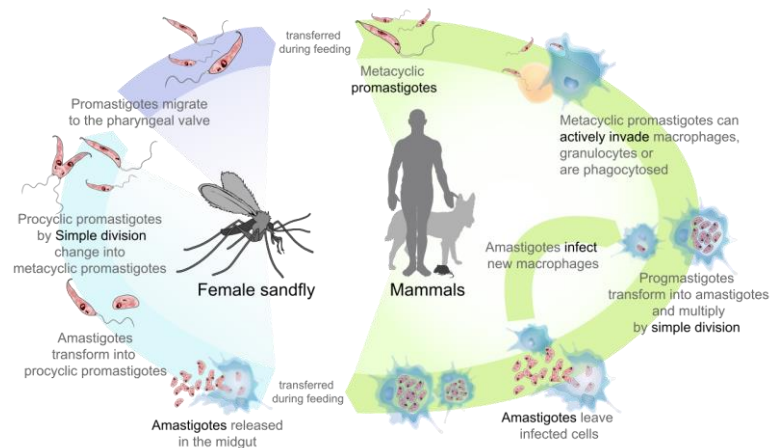
1. **Mass spec. evidence** - Peptides from proteomics experiments are mapped to a reference genome enabling searches for genes based on that mapping.
2. **Quantitative mass spec. evidence** - Data from quantitative proteomic experiments are loaded and made available for searching based on fold change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

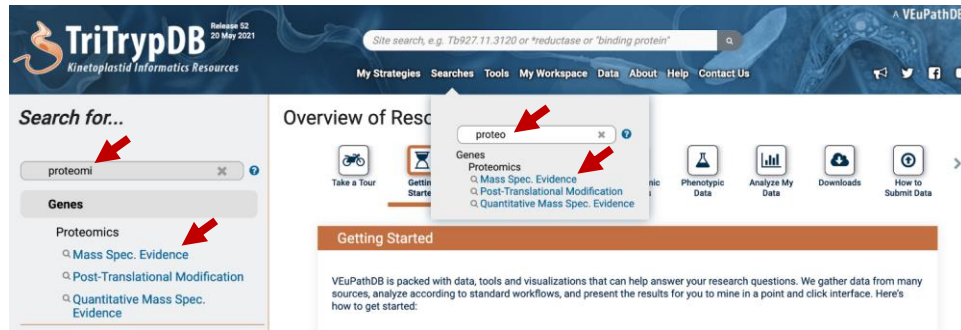
- Understand the different categories of proteomics data
- Learn how to run searches to identify genes based on peptide evidence
- Learn how to identify differentially expressed genes based on quantitative data
- Learn how to identify genes with different PTMs

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.

Identify Genes based on Mass Spec. Evidence

Experiments and Samples

1 selected, out of 151

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

- ☐ Leishmania
 - ☐ Leishmania infantum JPCM5
 - ☒ Metacyclic Stage Proteome (Ouellette, et al. unpublished)
 - ☒ metacyclic stage (pH 5-6)
- ☐ Trypanosoma
 - ☐ Trypanosoma cruzi
 - ☐ Trypanosoma cruzi CL Brener Esmeraldo-like
 - ☐ Life cycle proteome (Brazil) (Atwood et al.)
 - ☐ metacyclic trypomastigote, esmeraldo-like
 - ☐ Phosphoproteome during Metocyclogenesis (Dm28c) (Marchini et al.)
 - ☐ metacyclic trypomastigotes, esmeraldo-like
 - ☐ Trypanosoma cruzi CL Brener Non-Esmeraldo-like
 - ☐ Life cycle proteome (Brazil) (Atwood et al.)
 - ☐ metacyclic trypomastigote, nonesmeraldo-like
 - ☐ Phosphoproteome during Metocyclogenesis (Dm28c) (Marchini et al.)
 - ☐ metacyclic trypomastigotes, nonesmeraldo-like
 - ☐ Trypanosoma cruzi strain CL Brener
 - ☐ Life cycle proteome (Brazil) (Atwood et al.)
 - ☐ metacyclic trypomastigote, Brener
 - ☐ Phosphoproteome during Metocyclogenesis (Dm28c) (Marchini et al.)
 - ☐ metacyclic trypomastigotes, Brener

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

Minimum Number of Unique Peptide Sequences

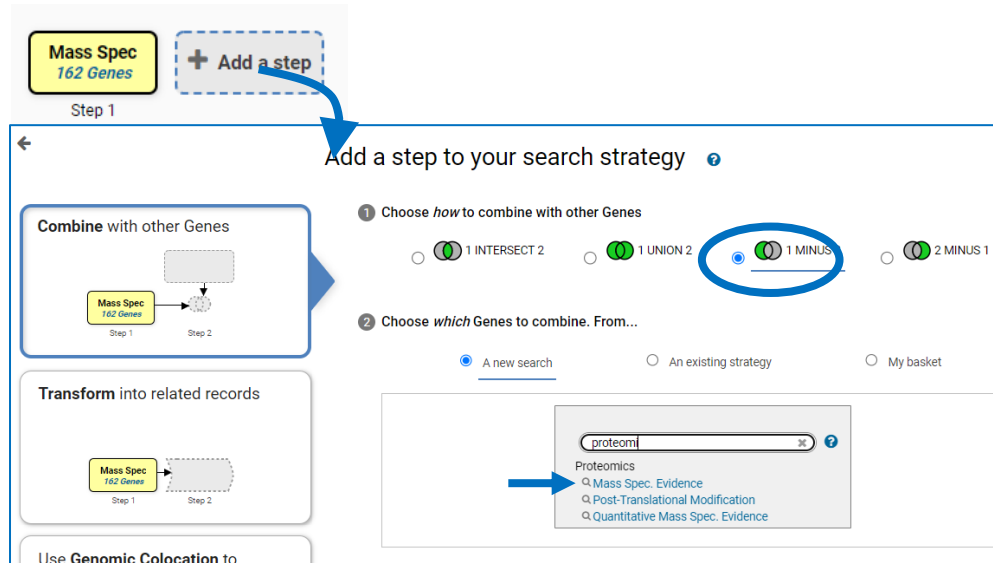
Apply min # peptide sequences / sample OR across samples

► Advanced Parameters

c. How many genes did you get?

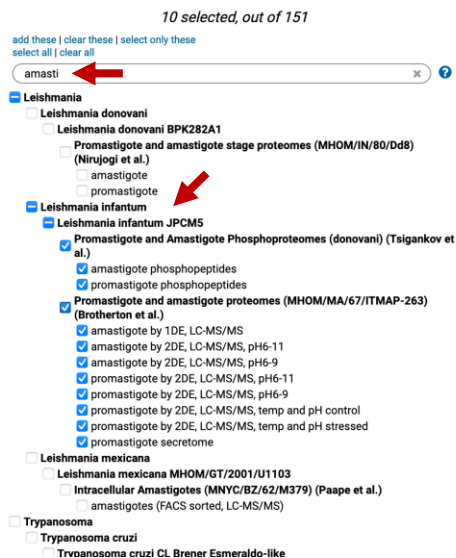


d. Now subtract the genes that have protein expression in the amastigote and promastigote stages. Add a step to your strategy that returns amastigote and promastigote genes and choose the 1 minus 2 operator to combine the searches.

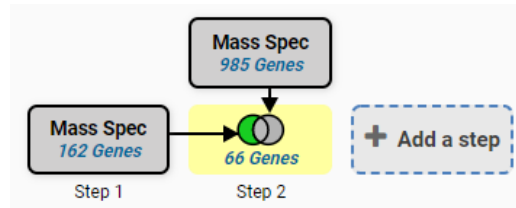


e. Choose all the *L. infantum* samples labeled amastigote and promastigote and run the search

Identify Genes based on Mass Spec. Evidence



- 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.

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

Details for step *Mass Spec* [Edit](#)

162 Genes

Experiments and Samples *metacyclic stage (pH 5-6)*

Minimum Number of Unique Peptide Sequences 1

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

Minimum number of spectra per gene (applied per sample) 1

► Give this search a weight

Experiments and Samples

1 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

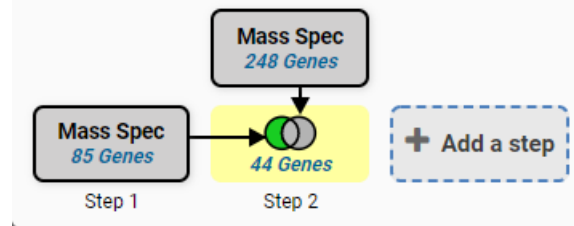
Apply min # peptide sequences / sample OR across samples

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 present at a higher concentration in the apicoplast compared to the endoplasmic reticulum (ER). Note for this exercise use <https://plasmodb.org>

- Go to the quantitative mass spec evidence searches
- 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	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

- 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 ←

select all | clear all

and its maximum ? expression value

in 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

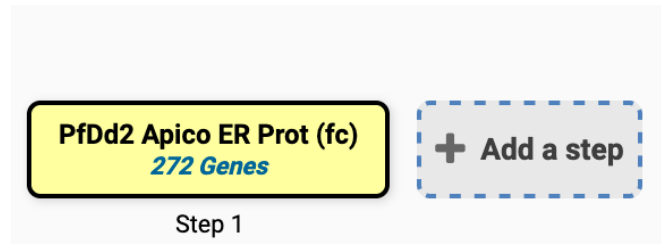
For each gene, the search calculates:

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

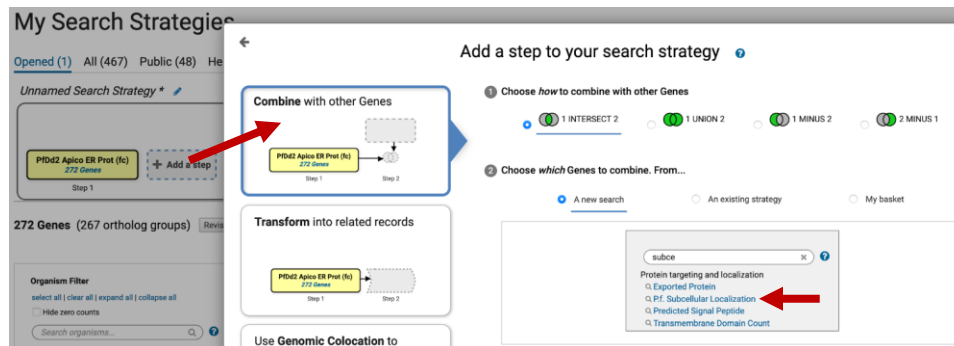
and returns genes when fold change \geq 1.5.

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

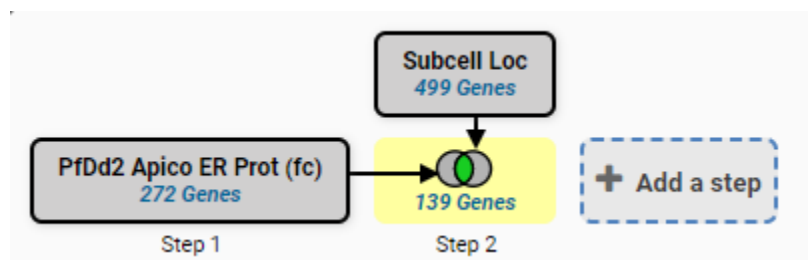
↑ COMMUNITY CHAT



- 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.)”

Identify Genes based on Quantitative Mass Spec. Evidence

Filter Data Sets: <input type="text"/>		Legend: DC Direct Comparison FC Fold Change
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) NEW	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 >=

- 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?

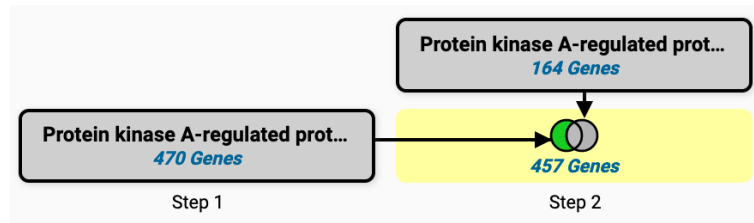
Gene ID	Transcript ID	Organism	Product Description	Fold Difference	Protein kinase A-regulated proteome - Expr Graph
CNAG_01579	CNAG_01579-426.1	Cryptococcus neoformans var. grubii H99	vacuolar membrane-associated protein IML1	134.17	
CNAG_03710	CNAG_03710-426.1	Cryptococcus neoformans var. grubii H99	hypothetical protein	21.25	

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.

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

Identify Genes based on Post-Translational Modification

Type of Post-Translational Modification

phosphorylation site

Experiments and Samples

12 selected, out of 12

select all | clear all | expand all | collapse all

Filter list below...

☒ *Toxoplasma gondii*

☒ *Toxoplasma gondii* GT1

☒ *Toxoplasma gondii* ME49

select all | clear all | expand all | collapse all

Number of modifications is

Greater than or equal to

Number of Modifications

5

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

Post-Translational Mod
3,212 Genes

+ Add a step

Step 1

3,212 Genes (1,794 ortholog groups) [Revise this search](#)

Gene Results [Genome View](#) [Analyze Results](#)

Rows per page: 1000

Download Send to... Add Columns

Gene ID	Transcript ID	Total Modified Residues	Total Modifications By Type
TGGT1_311230	TGGT1_311230-t26_1	149	phosphorylation site:149
TGGT1_253750	TGGT1_253750-t26_1	126	phosphorylation site:126
TGGT1_209000	TGGT1_209000-t26_1	122	phosphorylation site:122
TGGT1_254940	TGGT1_254940-t26_1	121	phosphorylation site:121
TGGT1_291180	TGGT1_291180-t26_1	116	phosphorylation site:116
TGGT1_232080	TGGT1_232080-t26_1	116	phosphorylation site:116

- 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.

← Add a step to your search strategy

Combine with other Genes

Post-Translational Mod (Step 1) → [] (Step 2)

1 Choose *how* to combine with other Genes

☒ 1 INTERSECT 2 ☐ 1 UNION 2 ☐ 1 MINUS 2 ☐ 2 MINUS 1

2 Choose *which* Genes to combine. From...

☒ A new search ☐ An existing strategy ☐ My basket

Transform into related records

Post-Translational Mod (Step 1) → [] (Step 2)

Use Genomic Colocation to combine with other features

local

Genomic Location

- Genomic Location
- Genomic Location (Non-nuclear)
- Proximity to Centromeres
- Proximity to Telomeres

Protein targeting and localization

- Localization by LOPIT Mass Spec
- Predicted Signal Peptide
- Transmembrane Domain Count

- 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

