

## 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.*

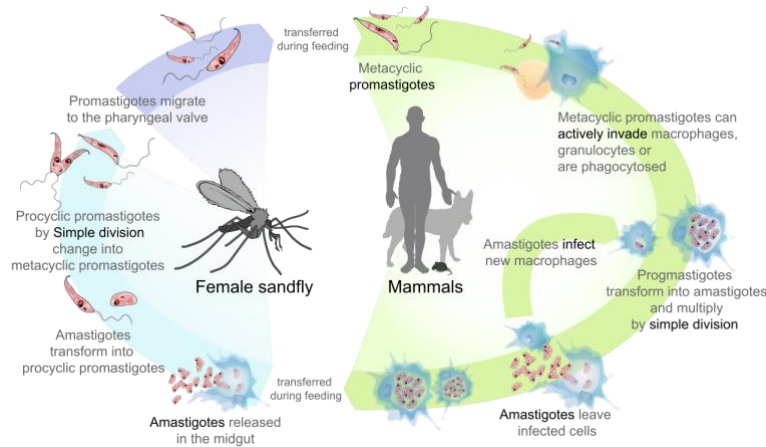
The exercises below explore the different categories and searches available for proteomics in VEuPathDB.

### 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

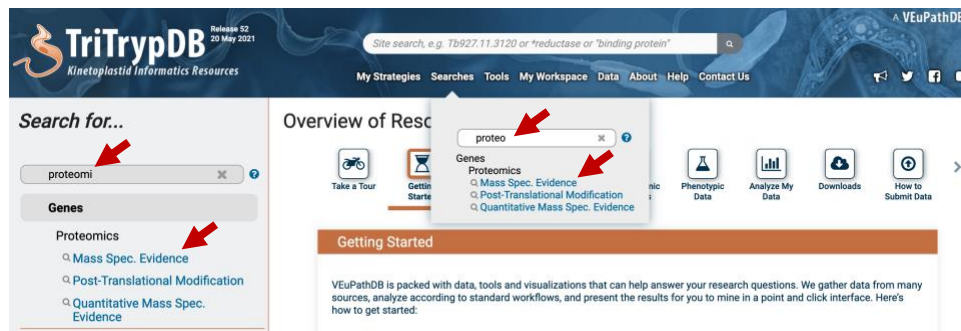
- Find genes that have peptide evidence from metacyclic stages but not amastigote or promastigote stages of *Leishmania infantum*.

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

## 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**

☒ **Promastigote (al.)**

☒ amastigote

☒ promastigote

**Promastigote (Brotherton)**

☒ amastigote

☒ 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**

**Mass Spec**  
985 Genes

+ Add a step

Step 1

**Minimum Number of Unique Peptide Sequences**

1

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

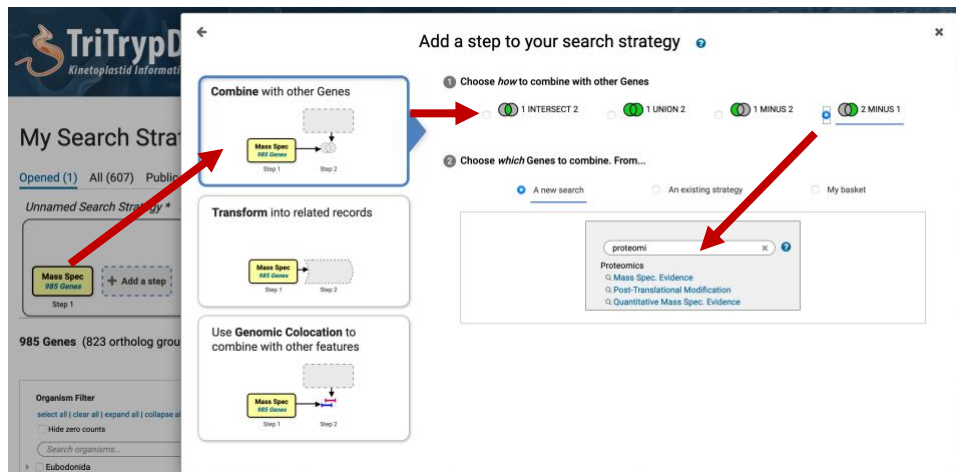
Per Sample

Advanced Parameters

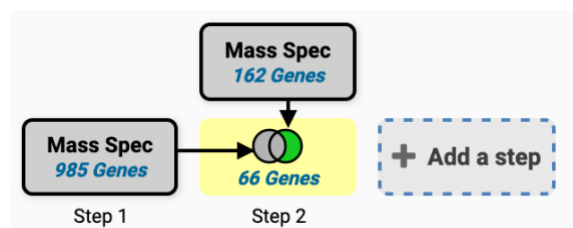
Get Answer

- d. How many genes did you get?
- 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?



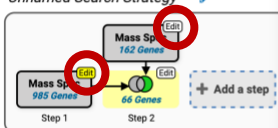
- g. How can you increase the stringency of your results? One way is to increase the minimum number of unique peptides. The default returns any gene with a minimum of one peptide. What happens if you change this to a minimum of 5 peptides in both steps?
- Click on the edit button
  - Click on the revise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

Organism Filter  
select all | clear all | expand all | collapse all  
☐ Hide zero counts

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

Revise your step

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

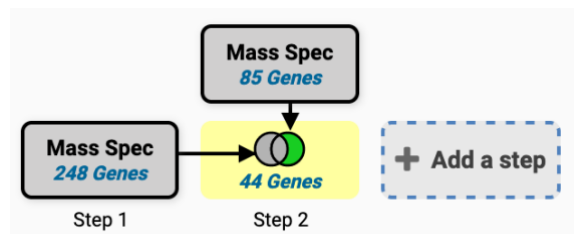
Apply min # peptide sequences / sample OR across samples

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...

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

- 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 >= 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:

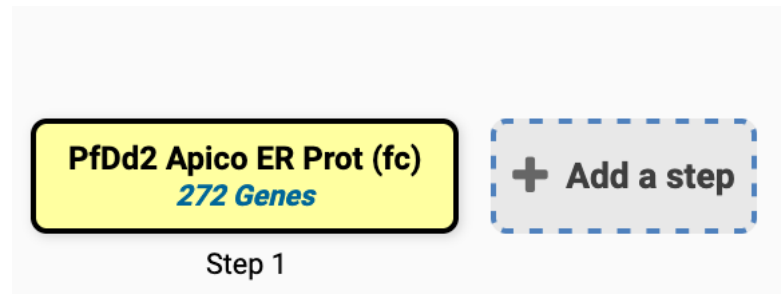
$$\text{fold change} = \frac{\text{comparison expression value}}{\text{reference expression value}}$$

and returns genes when fold change >= 1.5.

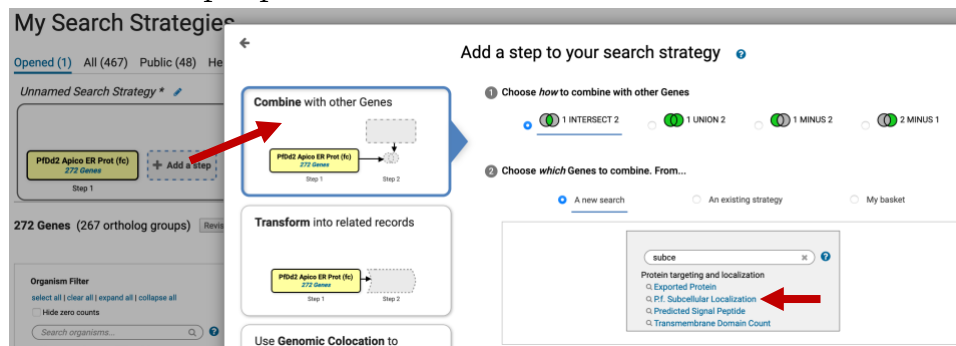
You are searching for genes that are up-regulated between one reference sample and one comparison sample.

COMMUNITY CHAT

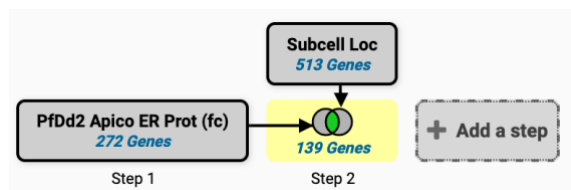
- d. How many genes did your search return?



- 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. 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*

### Identify Genes based on Quantitative Mass Spec. Evidence

Filter Data Sets: <input type="text"/>			Legend: <span>DC</span> Direct Comparison <span>FC</span> 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)	<span>FC</span>	
<i>Aspergillus fischeri</i> NRRL 181	Proteomics changes in response to human serum (Wiedner et al. 2013)	<span>FC</span>	
<i>Aspergillus fumigatus</i> Af293	Proteomics changes in response to human serum (Wiedner et al. 2013)	<span>FC</span>	
<i>Aspergillus fumigatus</i> Af293	<i>Aspergillus fumigatus</i> response to hypoxia (Barker et al. 2012) <span>NEW</span>	<span>FC</span>	
<i>Aspergillus fumigatus</i> Af293	Development stage specific proteome (Suh et al.)	<span>FC</span>	
<i>Aspergillus fumigatus</i> Af293	Adaptive mechanisms of <i>Aspergillus fumigatus</i> /conidia to nutrient restriction Quant (Andjo et al.)	<span>FC</span>	
<i>Cryptococcus neoformans</i> var. <i>grubii</i> H99	Secretome profiling of Pka-1 regulated proteins in <i>Cryptococcus neoformans</i> (16, 48, 72, and 120 hr post inoculation) (Geddes et al.)	<span>FC</span>	
<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.)	<span>DC</span>	
<i>Neurospora crassa</i> OR74A	Circadian time course data from wild type and delta csp-1 (Hurley et al.)	<span>DC</span>	

*neoformans* (Geddes et al.)”

- Configure the direct comparison search to identify genes that are upregulated

### 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

by 3 fold in galactose media

- How many genes did you get?
- 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	<p>expm_val - CNAG_01579</p>
CNAG_03710	CNAG_03710-126.1	<i>Cryptococcus neoformans</i> var. <i>grubii</i> H99	hypothetical protein	21.25	<p>expm_val - CNAG_03710</p>



- 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. A search strategy named 'Unnamed Search Strategy' is shown on the left, with a step 'Protein kinase A-regulated prot... 28 Genes'. A red arrow points to the 'Add a step' button. The main panel shows the configuration for a new step:

- Add a step to your search strategy**
- 1 Choose how to combine with other Genes**: Options are INTERSECT 2, UNION 2, MINUS 2 (selected), and MINUS 1.
- 2 Choose which Genes to combine. From...**: Options are 'A new search' (selected), 'An existing strategy', and 'My basket'.
- Direction**: A dropdown menu set to 'up-regulated'.
- Comparison**: Radio buttons for 'PGAL7::PKA1 + glucose' (selected) and 'PGAL7::PKA1 + galactose'.
- Fold difference >=**: A text input field set to '1.5'.
- Run Step** button.

- 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

phosph

**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) (Nebi 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 Mod  
2,266 genes

2,266 Genes (1,531 ortholog groups)

Organism Filter  
select all | clear all | expand all | collapse all  
Hide zero counts  
Search organisms  
Enkaidae  
Sarcocystidae  
select all | clear all | expand all | collapse all  
Hide zero counts

Gene Results  
Genome View  
Analysis Results

Rows per page: 20

Gene ID	Transcript ID	Modified Residues	Total Modified Residues	Modifications By Type	Product Description
TG0T1_311230	TG0T1_311230	S:130, S:166, S:174, S:231, S:240, S:243, T:248, T:330, S:404, T:405, S:408, S:422, S:870, S:920...	129	phosphorylation site:128	hypothetical protein
TG0T1_253750	TG0T1_253750	Y:129, S:131, T:170, S:213, S:222, S:242, T:243, S:273, S:278, S:293, S:296, S:300, S:304, S:305...	115	phosphorylation site:115	PLU-1 family protein
TG0T1_209000	TG0T1_209000	S:1303, S:1348, S:1350, S:1354, S:1359, S:1364, S:1401, S:1405, S:1410, S:1412, T:1419, S:1506, T...	110	phosphorylation site:113	HECT-domain (ubiquitin transferase) domain-containing protein
TG0T1_254940	TG0T1_254940	S:103, S:137, S:143, S:254, S:678, S:705, S:769, S:791, S:809, S:812, T:814, S:835, S:838, S:841...	106	phosphorylation site:106	MF4G domain-containing protein
TG0T1_291180	TG0T1_291180	S:198, S:248, S:255, S:259, S:263, S:265, S:266, T:270, S:271, T:289, S:364, S:368, T:341, S:342...	103	phosphorylation site:103	hypothetical protein
TG0T1_233080	TG0T1_233080	S:320, T:440, S:442, S:445, S:503, S:568, S:699, S:990, S:1042, S:1085, S:1088, S:1091, S:1131, T...	103	phosphorylation site:103	hypothetical protein

- How many of these phosphorylated genes are also localized to the microneme organelles?

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

Opened (1) All (1) Public (9) Help

Unnamed Search Strategy \*

Post-Translational Mod 2,266 Genes

Step 1

2,266 Genes (1,531 ortholog groups)

Organism Filter

select all | clear all | expand all | collapse all

Hide zero counts

Search organisms...

Emmeridae 0

Add a step to your search strategy

1 Choose how to combine with other Genes

1 INTERSECT 2 UNION 2 1 MINUS 2 2 MINUS 1

2 Choose which Genes to combine. From...

A new search An existing strategy My basket

local

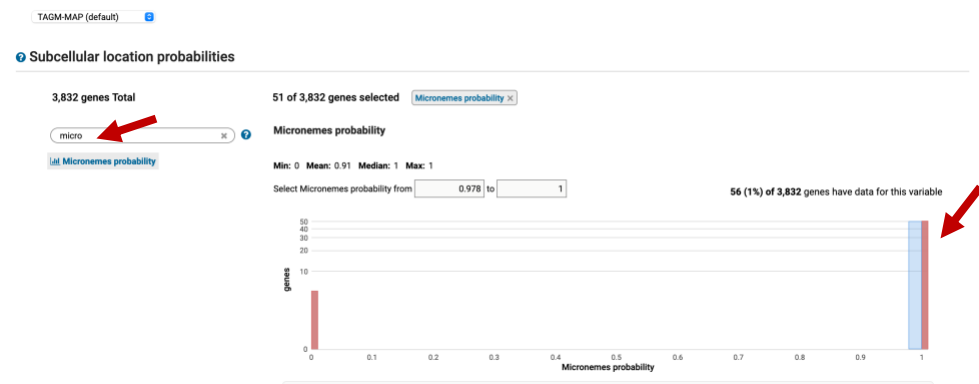
Protein targeting and localization

Localization by LOPIT Mass Spec

Predicted Signal Peptide

Transmembrane Domain Count

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



- Explore your results.

Opened (1) All (1) Public (9) Help

Unnamed Search Strategy \*

Post-Translational Mod 2,266 Genes

Step 1

LOPIT 51 Genes

Step 2

5 Genes (5 ortholog groups)

Organism Filter

select all | clear all | expand all | collapse all

Hide zero counts

Search organisms...

Emmeridae 0

Sarcocystidae 5

Gene Results Genome View Analyze Results

Rows per page: 20

Download Add to Basket Add Columns

Gene ID	Transcript ID	Genomic Location (Gene)	Product Description	# Transcripts
TGME49_221180	TGME49_221180-i26_1	TGME49_chrII:108,336..114,763(-)	hypothetical protein	1
TGME49_205680	TGME49_205680-i26_1	TGME49_chrVIIa:1,214,495..1,219,931(-)	hypothetical protein	1
TGME49_226020	TGME49_226020-i26_1	TGME49_chrX:1,862,972..1,869,285(-)	transporter, major facilitator family protein	1
TGME49_245490	TGME49_245490-i26_1	TGME49_chrXII:2,306,394..2,309,968(+)	microneme protein MIC8	1