

Expression Searches: Transcriptomics and Proteomics

Learning Objectives

- Review the types of expression searches in VEuPathDB.
- Use the RNA-Seq differential expression, fold change and percentile searches to explore gene expression in midgut and proventriculus expression of *Trypanosoma brucei* infections.
- Use the Proteomics peptides and quantitative searches to explore expression at the protein level.

Transcript expression or the abundance of an mRNA, can be determined in the laboratory with several different techniques including RNA-sequence, microarray, and RT-PCR. VEuPathDB supports these data types with several searches (see table below). For RNA sequence data, expression values are graphed on gene pages and mapped reads can be visualized in the genome browser. Proteomics experiments can determine the peptide sequence by mass spectrometry or the abundance value via e.g. isobaric tagging for a peptide in a sample. Each data type is available on gene record pages but only the mass spec peptides have data that can be viewed in JBrowse. Using the search strategy system, it's easy to delve deep into a specific data set and to take advantage of several types of data when combining search results in the strategy system.

Search	Description	RNA-seq	Micro-array	Proteomics
Differential Expression	Statistical analysis of studies whose experimental design includes biological replicates. A differential expression search finds genes based on fold change difference between two samples with a user defined p-value cutoff. Only pairwise comparisons can be made with this search.	✓		
Fold Change	Expression differences between samples are calculated but statistical analyses are not performed. A fold change search finds genes whose expression value differs between samples without considering statistical parameters. This search offers a form of differential expression analysis when the experimental design did not include replicates and allows for comparing groups of samples, e.g. find genes whose expression is up-regulated in the liver time course (2, 24, 36, and 54 hours) vs the control (0 hours).	✓	✓	
Percentile	For each sample in an experiment, each genes' expression value is sorted from lowest to highest and a percentile rank is determined. For example, a percentile search can find genes whose expression is in the highest 10% of expression values within a sample.	✓	✓	
Sense/Antisense	For strand-specific RNA sequence, expression values are determined in the sense and antisense direction. This search finds genes that exhibit simultaneous changes in sense and antisense transcripts. For example, you can look for genes with increasing antisense transcripts and decreasing sense transcripts, as might occur when antisense transcription suppresses sense transcription.	✓		
Splice-site Location	This trypanosome-specific search takes advantage of the 'splice-leader' RNA-seq data which determines transcript abundance within the polycistronic mRNA using splice-leader specific primers. This search identified genes whose 5' splice site location varies between samples.	✓		

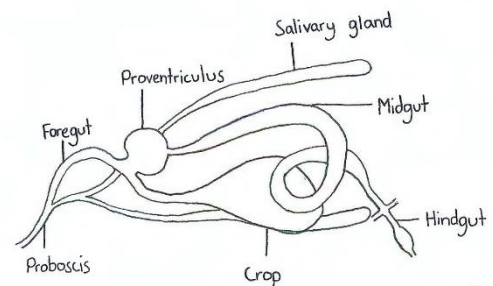
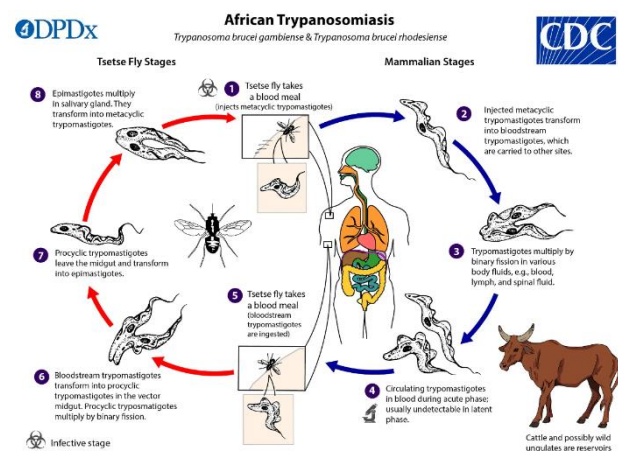
Metacycle	The MetaCycle package detects rhythmic signals from large scale time-series data, such as circadian rhythms within expression time courses, using either ARSER or JTK-Cycle. This search returns genes whose rhythmic signals match the conditions (period and amplitude range) you specify. The search will return the corresponding period, amplitude and p-value of genes that meet your search criteria.	✓	✓	
Similarity	The similarity search returns genes whose expression profile within the experiment follow a similar pattern as the gene you specify.	✓	✓	
Direct Comparison	Microarray data for two samples is often collected on the same glass slide. For these experiments, the direct comparison search returns genes whose expression varies between samples in pairwise comparisons.		✓	
Co-expression	Meta-analysis across multiple microarray experiments defined a co-expression network. This search returns genes within the co-expression network of your gene(s) of interest.		✓	
Mass spec. evidence	Peptides from proteomics experiments are mapped to a reference genome enabling searches for genes based on the mapping			✓
Quantitative mass spec. evidence	Quantitative proteomic experiments produce abundance values for proteins identified in the sample analyzed. (excel spreadsheets of gene/protein IDs and expression intensity values)			✓
Post-translational modification (PTM)	PTM data from proteomics experiments, excel spreadsheets of gene/protein IDs, location and type of modification, are associated with genes in VEuPathDB, enabling searches for genes based on the type and number of the PTM.			✓

1. Find genes that are up-regulated in the midgut compared to the Proventriculus stages of *Trypanosoma* infection. TriTrypDB.org

The life cycle of *Trypanosoma brucei* is split between the tsetse fly and the human host. The fly stage includes a migration from the midgut to the proventriculus and eventually to the salivary gland where it can infect a host during a blood meal. The cycle in the fly takes approximately 3 weeks. The erythrocytic stages are well studied compared to the liver stages.

TriTrypDB contains RNA seq data from samples taken during the establishment of infection and progression of *Trypanosoma brucei brucei* through its insect host. ([Naguleswaran et al. 2021](#)) This data set includes a time course of midgut samples from 3 to 28 days, as well as proventriculus and salivary gland samples.

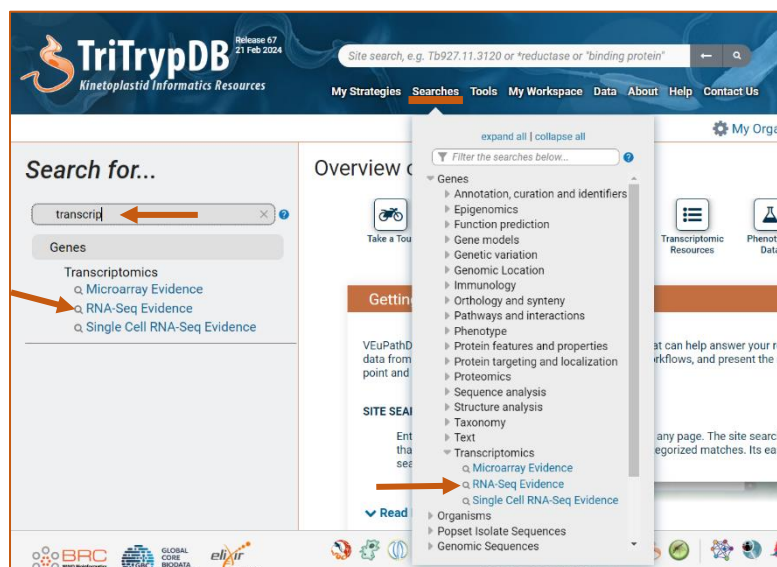
- Midgut Day 3
- Midgut Day 7



- Midgut Day 11
- Midgut Day 15
- Midgut Day 28
- Proventriculus Day 28
- Salivary Gland Day 28
- This gene is not statistical <https://tritrypdb.org/tritrypdb/app/record/gene/Tb927.8.8280>

Use this data set to determine what genes are upregulated at least 2 fold (p-value ≤ 0.001) at Midgut Day 28 vs the proventriculus Day 28.

- Navigate to the RNA seq search page and find the data set called **Transcriptome of *T. brucei* from midgut (days 3-28), proventriculus and salivary glands (Naguleswaran et al 2021)**. Searches are available from the Search For... menu on the left side of the home page, as well as the Searches drop down menu in the header.



Identify Genes based on RNA-Seq Evidence

Legend: PQ Quantitative Phenotype SSL Splice Site Loc DE Differential Expression FC Fold Change MC MetaCycle P Percentile SA SenseAntisense

Filter Data Sets: midgut 2 results (filtered from a total of 45)

Organism	Data Set	Choose a Search
<i>Trypanosoma brucei</i> brucei TREU927	Procyclic and bloodstream form transcriptomes and ribosome profiling (Jensen et al.)	DE FC P SA
<i>Trypanosoma brucei</i> brucei TREU927	Transcriptome of <i>T. brucei</i> from midgut (days 3-28), proventriculus and salivary glands (Naguleswaran et al 2021)	DE FC P SA

- b. Arrange the differential expression search to return genes that are at **least 2-fold up-regulated** in the Midgut Day 28 sample compared to the Proventriculus Day 28 with a p-value of $p < 0.001$.

Differential Expression

Fold ChangePercentileSenseAntisense

Identify Genes based on T. brucei brucei TREU927 Transcriptome of T brucei from midgut (days 3-28), proventriculus and salivary glands RNA-Seq (Differential Expression)

Configure Search

Learn More

View Data Sets Used

↻ Reset values to default

Experiment

☒ Tbrucei EATRO1125 RNAseq - Sense

☐ Tbrucei EATRO1125 RNAseq - Antisense

Reference Sample

☐ Midgut Day 3

☐ Midgut Day 7

☐ Midgut Day 11

☐ Midgut Day 15

☐ Midgut Day 28

☒ Proventriculus Day 28

☐ Salivary Gland Day 28

Comparator Sample

☐ Midgut Day 3

☐ Midgut Day 7

☐ Midgut Day 11

☐ Midgut Day 15

☒ Midgut Day 28

☐ Proventriculus Day 28

☐ Salivary Gland Day 28

Direction

up-regulated

fold difference >=

2

adjusted P value less than or equal to

0.001

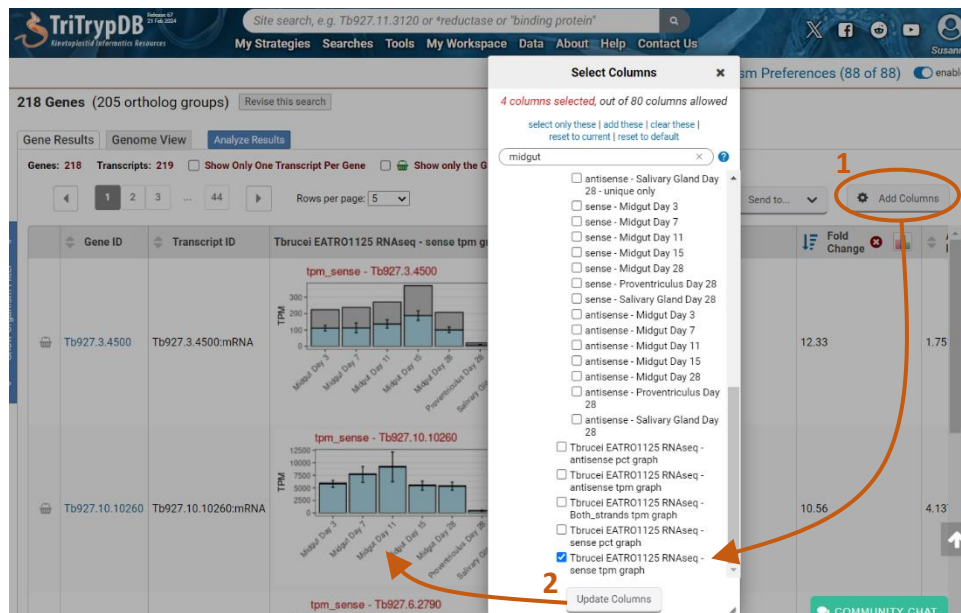
Get Answer

Tbrucei EATRO1125 RNAseq (d...
216 Genes

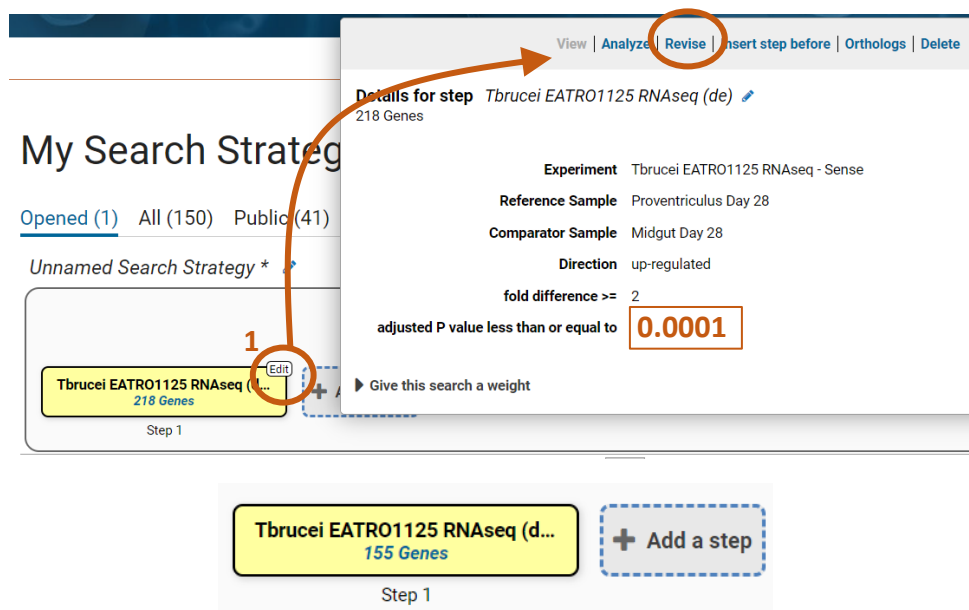
+ Add a step

Step 1

- c. How many genes were returned by the search? Do you believe these results? To convince yourself, you could turn on a column with the TPM graph and check if Midgut day 28 is at least 2-fold higher than proventriculus. (The column with the graph might already be showing in the result table.)



- d. Increase the statistical stringency of the search from $p \leq 0.001$ to $p < 0.0001$. How many genes are returned by the search now? Hint: revise the search and change the p-value. Hover over the yellow search box until the Edit icon appears. Click the Edit icon and choose revise from the options panel.



- e. Analyze this gene result for enriched GO terms in the Biological Process ontology. The Analyze Results tab offers tools for performing a GO, metabolic pathway and word enrichments. GO terms are a form of annotation that provide information about a gene's biological function, molecular function and cellular location. A GO Enrichment is a statistical analysis that determines the probability that the 155 genes in the result share a term compared to the genome. What Biological processes are shared by the gene set?

Trypanosoma EATRO1125 RNAseq (d...
155 Genes

Step 1

155 Genes (150 ortholog groups) [Revise this search](#)

Gene Results **Genome View** **Gene Ontology Enrichment*** [Analyze Results](#)

Genes: 155 Transcripts: 156 ☐ Show Only One Transcript Per Gene ☐ Show Only One GO

Analyze your Gene results with a tool below.

Gene Ontology Enrichment

Metabolic Pathway Enrichment

kinase
phosphatase
exported
membrane

Word Enrichment

Gene Results **Genome View** **Gene Ontology Enrichment*** [Analyze Results](#)

[Rename This Analysis | Duplicate]

Gene Ontology Enrichment

Find Gene Ontology terms that are enriched in your gene result. [Read More](#)

▼ Parameters

Organism [?](#) Trypanosoma brucei brucei TREU927 ▼

Ontology [?](#) ☒ Biological Process
☐ Cellular Component
☐ Molecular Function

Evidence [?](#) ☒ Computed
☒ Curated
[select all](#) | [clear all](#)

Limit to GO Slim terms [?](#) ☒ No
☐ Yes

P-Value cutoff [?](#) (0 - 1)

[Submit](#)

Analysis Results:

[Open in Revigo](#) [Show Word Cloud](#) [Download](#)

GO ID	GO Term	Genes in the bkgd with this term	Genes in your result with this term	Percent of bkgd genes in your result	Fold enrichment	Odds ratio	P-value	Benjamini	Bonferroni
GO:0044281	small molecule metabolic process	487	38	7.8	4.43	5.95	2.27e-15	1.77e-12	1.77e-12
GO:0019752	carboxylic acid metabolic process	188	23	12.2	6.95	8.97	1.41e-13	5.17e-11	1.10e-10
GO:0043436	oxoacid metabolic process	191	23	12.0	6.84	8.81			
GO:0006082	organic acid metabolic process	194	23	11.9	6.73	8.65			
GO:0008152	metabolic process	2807	91	3.2	1.84	3.14			
GO:0071704	organic substance metabolic process	2592	81	3.1	1.77	2.70			
GO:0006091	generation of precursor metabolites and energy	94	13	13.8	7.85	9.69	8.31e-9	9.25e-7	6.47e-6

The genome has 'this many genes' annotated with this GO term

Your result has 'this many genes' annotated with this GO term

Revigo – offers further analysis of your list of GO terms

Creates a word cloud of the GO Term descriptions

The probability of this GO term appearing in a list of 155 genes compared to the whole genome.

2. Using the same data set, find genes that are upregulated 2-fold in any midgut stage compared to proventriculus.

 - a. Navigate to the RNA Seq search page and choose the **Fold Change** search for the **Transcriptome of *T. brucei* from midgut (days 3-28), proventriculus and salivary glands (Naguleswaran et al 2021)** data set as in 1a above.
 - b. Arrange the fold change search to return **protein coding** genes that are **up-regulated** in the **average** expression across the midgut samples (comparator) compared to the proventriculus sample (reference).

For the Experiment

☒ Transcriptome of *T. brucei* from midgut (days 3-28), proventriculus and salivary glands - Sense

☐ Transcriptome of *T. brucei* from midgut (days 3-28), proventriculus and salivary glands - Antisense

return protein coding Genes

that are up-regulated

with a Fold change $\geq$ 2

between each gene's average expression value

(or a Floor of 10 reads)

in the following Reference Samples

☐ Midgut Day 3
☐ Midgut Day 7
☐ Midgut Day 11
☐ Midgut Day 15
☐ Midgut Day 28
☒ Proventriculus Day 28
☐ Salivary Gland Day 28

select all clear all

and its average expression value

(or the Floor selected above)

in the following Comparison Samples

☒ Midgut Day 3
☒ Midgut Day 7
☒ Midgut Day 11
☒ Midgut Day 15
☒ Midgut Day 28
☐ Proventriculus Day 28
☐ Salivary Gland Day 28

select all clear all

Get Answer

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

Up-regulated

A maximum of four samples are shown when more than four are selected.

For each gene, the search calculates:

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

and returns genes when fold change $\geq$ 2.

You are searching for genes that are **up-regulated** between one **reference sample** and at least two **comparison samples**.

To narrow the window, use the minimum comparison value. To broaden the window, use the maximum comparison value.

Tbrucei EATRO1125 RNAseq (fc)
725 Genes

+ Add a step

Step 1

725 Genes (649 ortholog groups) Revise this search

Some Genes in your result have Transcripts that did not meet the search criteria. Explore

Gene Results Genome View Analysis Results

Genes: 725 Transcripts: 731 ☐ Show Only One Transcript Per Gene ☐ Show only the Genes in my basket.

Rows per page 5

Gene	Transcript	Gene	Transcript	Gene	Transcript	Gene	Transcript	Gene	Transcript
Tb927.10.10260	Tb927.10.10260.mRNA	Trypanosoma brucei brucei TREU927	EP1 procyclin	14.5	459.82	6681.52			
Tb11.v5.1046	Tb11.v5.1046.1	Trypanosoma brucei brucei TREU927	elongation factor 1-alpha, putative	14.1	30.13	424.58			
Tb927.6.510	Tb927.6.510.mRNA	Trypanosoma brucei brucei TREU927	GPEET procyclin	13.6	175.44	2387.99			

Download Send Add Columns

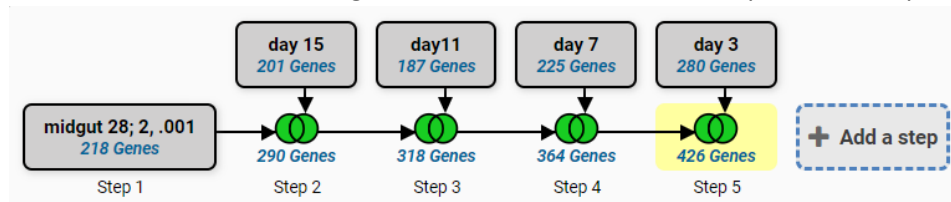
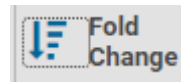
tpm_sense - Tb927.10.10260

tpm_sense - Tb11.v5.1046

tpm_sense - Tb927.6.510

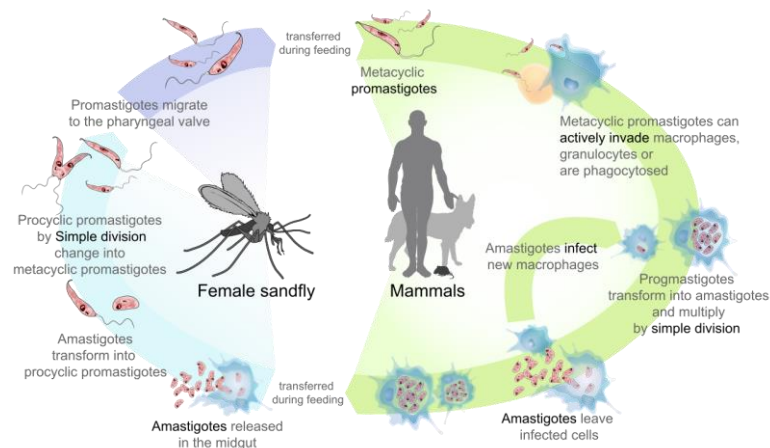
- c. Explore your results. Did the search return more genes or fewer genes than the differential expression search? How confident are you that this gene list?

- d. Use the Add Columns to turn on the TPM graph for the 'Ex-erythrocytic stages' data set. Notice the error bars and fold change values for the Tb927.6.510 GPEET procyclin gene. Would GPEET procyclin gene be returned by the Differential Expression search for with pairwise comparisons of Midgut Days 7, 11, 15, 28 with proventriculus? (You may need to sort the Fold Change column in the results, to display the gene with the highest FC value at the top. Then GPEET is about the 4th gene down)
- e. How would you use the DE search to find genes upregulated in any midgut sample compared to proventriculus? Hint: Use the strategies to sum the results of several pairwise comparisons.



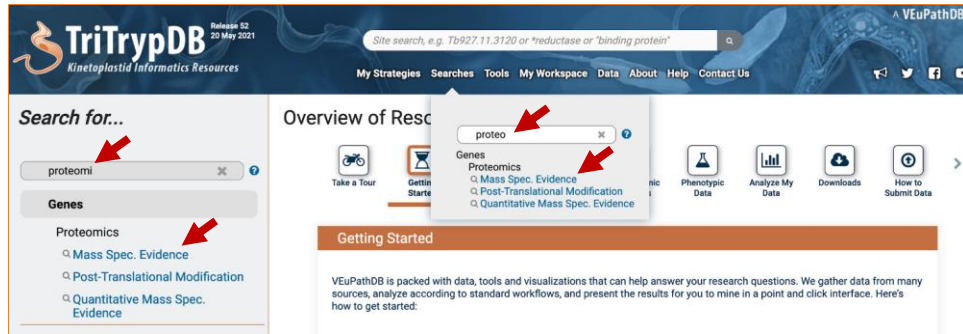
<https://tritrypdb.org/tritrypdb/app/workspace/strategies/import/d432e56011aae73e>

3. Find *Leishmania infantum* genes that have protein expression evidence from metacyclic stages but not amastigote or promastigote stages. 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 152
select only these | add these | clear these

infan

☐ Leishmania

☐ Leishmania infantum JPCM5

☒ Metacyclic Stage Proteome (Ouellette, et al. unpublished)

☒ metacyclic stage (pH 5-6)

☐ Post-translationally modified proteins during differentiation (Rosenzweig et al.)

☐ acetylated proteins (L. donovani)

☐ glycosylated proteins (L. donovani)

☐ methylated proteins (L. donovani)

☐ phosphorylated proteins (L. donovani)

☐ 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

Minimum Number of Unique Peptide Sequences

1

Apply min # peptide sequences / sample OR across samples

Per Sample

Advanced Parameters

Mass Spec
162 Genes

+ Add a step

Step 1

Get Answer

- c. 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.

The screenshot shows the 'Add a step to your search strategy' dialog. On the left, there are three panels: 'Combine with other Genes', 'Transform into related records', and 'Use Genomic Colocation to'. The 'Combine with other Genes' panel shows a diagram of two sets of genes being combined. The 'Transform into related records' panel shows a search for 'proteom' with results like 'Mass Spec. Evidence' and 'Post-Translational Modification'. The 'Add a step to your search strategy' dialog has two main sections: '1 Choose how to combine with other Genes' and '2 Choose which Genes to combine. From...'. In the first section, the '1 MINUS 2' operator is selected and circled in orange. In the second section, the 'A new search' option is selected, and a search box contains the text 'proteom'.

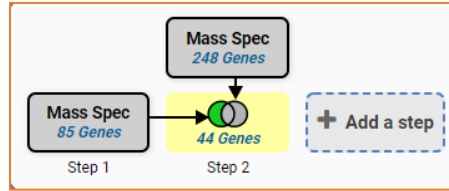
- d. Choose all the *L. infantum* samples labeled amastigote and promastigote and run the search

The screenshot shows the search results for 'infan'. The results are filtered to 10 selected out of 152. The search results are organized into a tree structure under the 'Leishmania' category. The 'Promastigote and Amastigote Phosphoproteomes (donovani)' and 'Promastigote and amastigote proteomes (MHOM/MA/67/ITMAP-263) (Brotherton et al.)' sections are expanded, showing various sample types. An orange arrow points from the search results to a diagram showing the final search strategy. The diagram shows a 'Mass Spec 162 Genes' box (Step 1) and a 'Mass Spec 985 Genes' box (Step 2) connected by a plus sign. A third box labeled '66 Genes' is shown below the plus sign, and an 'Add a step' button is to the right.

- e. Visit the gene pages of some of your results. There you can view mapped peptides and data from other experiments.
(https://tritrypdb.org/tritrypdb/app/record/gene/LINF_340044200#ProteinExpressionPBrowse)
- f. 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 settings that we used above return any gene with a minimum of one peptide. Edit each search to look for genes that are supported by 5 peptide sequences instead of just one.

The image shows a workflow in the tritrypdb application. At the top, a sequence of steps is shown: 'Step 1: Mass Spec 162 Genes' and 'Step 2: Mass Spec 985 Genes'. An arrow points from the 'Revise' button in the top navigation bar to a 'Details for step' panel for 'Mass Spec 162 Genes'. This panel shows settings for '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), and 'Minimum number of spectra per gene (applied per sample)' (1). Below this, a 'Revise your step' dialog is open. It has tabs for 'Configure Search', 'Learn More', and 'View Data Sets Used'. Under 'Experiments and Samples', it shows '1 selected, out of 152' with a filter dropdown set to 'Leishmania'. Under 'Minimum Number of Unique Peptide Sequences', a text input field contains the number '5', with an arrow pointing to it from the left. Under 'Apply min # peptide sequences / sample OR across samples', a dropdown menu is set to 'Per Sample'. At the bottom right of the dialog, a 'Revise' button is circled in orange.

- g. How did this change your results? Are you more or less confident in saying that these genes are expressed at the protein level?



4. 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...

Genes

- Proteomics
 - Quantitative Mass Spec. Evidence
- Transcriptomics
 - RNA-Seq Evidence

Identify Genes based on Quantitative Mass Spec. Evidence

Legend: DCC Direct Confidence Comparison FC Fold Change

Filter Data Sets: 5 rows

Organism	Data Set	Choose a Search
Plasmodium berghei ANKA	Proteome of ApiAP2 double vs single knockout (Modrzynska et al)	DCC
Plasmodium falciparum 3D7	Long-lived merozoite proteome (Kumar et al.)	FC
Plasmodium falciparum 3D7	Proteome and phosphoproteome during intraerythrocytic development (Quantitative) (Pease et al.)	FC
Plasmodium falciparum 3D7	PfCRK4 regulated proteome at 29 and 37 hpi (quantitative) (Ganter et al.)	DCC
Plasmodium falciparum 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 to default

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 average expression value

in the following **Reference Samples**

- ☐ Apicoplast
- ☒ ER

select all clear all

and its average 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**.

Get Answer

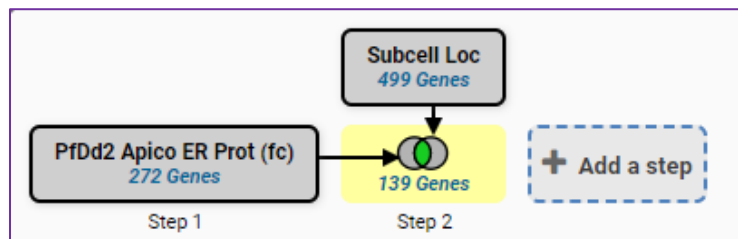
PfDd2 Apico ER Prot (fc)
272 Genes

Step 1

+ Add a step

- d. Can you leveraging other data about apicoplast biology to validate your results? For example, it is well known that proteins with transit peptides are targeted to the apicoplast. PlasmoDB has a search that returns genes with the transit peptides. Add a step to your strategy that increases the stringency of evidence for these genes being apicoplast genes. The search is called 'Pfal 3D7 Subcellular Localization'

- 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?



5. **Find *Aedes aegypti* genes that are upregulated in both head and muscle during infection with *Wolbachia*.** The *Wolbachia* strain wMelPop, which reduces longevity in *Drosophila melanogaster*, has been introduced into the Dengue virus mosquito vector, *Aedes aegypti* as a strategy to reduce disease transmission. VectorBase has a **microarray** data set that compared *Wolbachia* infected and uninfected mosquito head and muscle. This exercise uses VectorBase.org.
- a. Navigate to the microarray search and choose the **Direct Comparison** search for the dataset titled '**Infection with a Virulent Strain of *Wolbachia* Disrupts Genome Wide-Patterns of Cytosine Methylation in the Mosquito *Aedes aegypti* (Ye et al.)**'

My Strategies **Searches** Tools My Workspace Data About Help Contact Us

My Organism Preferences (88 of 88) enabled

Identify Genes based on

Legend: DC Direct Comparison FC Fold Change MC MetaCycle P Percentile

Filter Data Sets: wolba 3 results (filtered from a total of 53)

Organism	Data Set	Choose a Search
<i>Aedes aegypti</i> LVP_AGWG	Infection with a Virulent Strain of Wolbachia Disrupts Genome Wide-Patterns of Cytosine Methylation in the Mosquito <i>Aedes aegypti</i> (Ye et al.)	DC P
<i>Aedes aegypti</i> LVP_AGWG	The relative importance of innate immune priming in Wolbachia-mediated dengue interference (Rancès et al.)	DC P
<i>Aedes aegypti</i> LVP_AGWG	Gene expression profiling in wMelPop-infected <i>Aedes aegypti</i> (Kambris et al.)	DC P

b. Initiate a search that returns genes that are upregulated 2 fold in infected head vs uninfected.

Experiment

Infection with a Virulent Strain of Wolbachia Disrupts Genome Wide-Patterns of Cytosine Methylation in the Mosquito *Aedes aegypti*

Direction

up-regulated

Comparison

head infected v head uninfected

muscle infected v muscle uninfected

Fold difference >=

2.0

Protein Coding Only:

protein coding

Wolbachia infection in head an...
695 Genes
Step 1

Get Answer

c. Intersect your search result with another search that returns genes upregulated 2-fold in muscle vs uninfected. Your combined result will be genes that are upregulated in head and muscle in response to *Wolbachia* infection.

Wolbachia infection in head an... 695 Genes

Step 1

Add a step to your search strategy

Combine with other Genes

Choose how to combine with other Genes

INTERSECT 2

Choose which Genes to combine. From...

A new search

RNA

Gene models

Gene Model Characteristics

Transcriptomics

Microarray Evidence

RNA-Seq Evidence

Experiment

Infection with a Virulent Strain of Wolbachia Disrupts Genome Wide-Patterns of Cytosine Methylation

Direction

up-regulated

Comparison

muscle infected v muscle uninfected

Fold difference >=

2.0

Protein Coding Only:

protein coding

Run Step

Wolbachia infection in head an... 827 Genes

Wolbachia infection in head an... 695 Genes

Step 1

Step 2

394 Genes

- d. Determine enriched Molecular Function GO terms for the upregulated genes. Make sure you are viewing the combined result (the Step 2 result will be highlighted yellow) and click Analyze Result to open the Enrichment Tool. What gene functions are shared by the combined result? What biological role can you envision for these mosquito genes during the *wolbachia* infection?

Wolbachia infection in head an... 827 Genes

Wolbachia infection in head an... 695 Genes

394 Genes (355 ortholog groups)

Gene Results Genome View **Analyze Results**

Genes: 394 Transcripts: 759 ☐ Show Only One Transcript Per Gene

Rows per page: 1000

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

Gene ID Transcript Genomic Location

Analyze your Gene results with a tool

GO

Gene Ontology Enrichment

Gene Ontology Enrichment

Find Gene Ontology terms that are enriched in your gene result. [Read More](#)

Parameters

Organism:

Ontology: ☐ Biological Process ☒ Molecular Function ☐ Cellular Component

Evidence: ☒ Computed ☒ Curated [select all](#) [clear all](#)

Limit to GO Slim terms: ☒ No ☐ Yes

P-Value cutoff: (0 - 1)

Analysis Results:

81 rows

GO ID	GO Term	Genes in the bgkd with this term	Genes in your result with this term	Percent of bgkd genes in your result	Fold enrichment	Odds ratio	P-value	Benjamini
GO:0003824	catalytic activity	3701	165	4.5	1.62	2.40	5.15e-14	1.22e-11
GO:0016787	hydrolase activity	1639	96	5.9	2.12	2.74	6.77e-14	1.22e-11
GO:0004175	endopeptidase activity	472	45	9.5	3.46	4.18	1.93e-13	1.87e-11
GO:0004252	serine-type endopeptidase activity	346	38	11.0	3.98	4.83	2.21e-13	1.87e-11

- e. Modify the strategy to find genes that are upregulated in head but not in muscle. Do these genes have any enriched Molecular Function GO terms?

Wolbachia infection in head an... 827 Genes

Wolbachia infection in head an... 695 Genes

394 Genes

Step 1 Step 2

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

Details for step Combine Gene results

394 Genes

Revise as a boolean operation

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

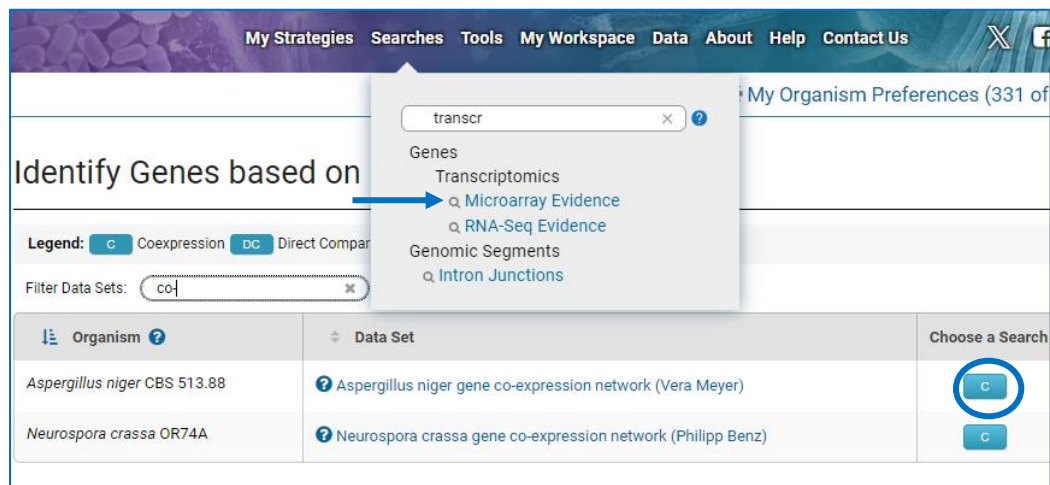
Wolbachia infection in head an... 827 Genes

Wolbachia infection in head an... 695 Genes

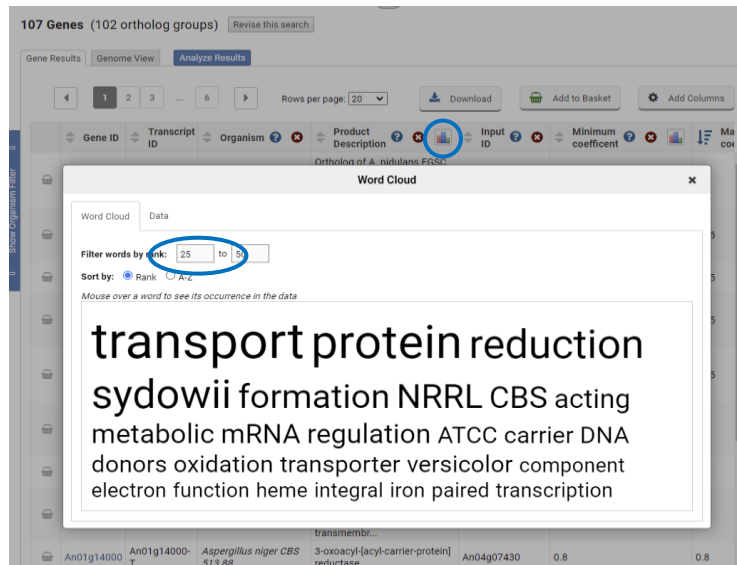
301 Genes

Step 1 Step 2

6. **Find genes that are likely co-expressed with An04g07430, an *Aspergillus niger* protein coding gene with little functional annotation.** By finding genes that are expressed at the same time as An04g07430, we may find clues about its function and the biological processes that it participates in. This exercise uses [FungiDB](#).
 - a. Navigate to the microarray searches in FungiDB and choose the Co-expression search for the data set titled ***Aspergillus niger* gene co-expression network (Vera Meyer)**. [Schape et al Nucleic Acids Research 2019](#). These data are the results of a meta-analysis of 155 publicly available transcriptomics analyses for *A. niger*, which were used to generate a genome-level co-expression network and sub-networks for >9,500 genes.
 - b. Run the search to find the co-expression network for An04g07430.



- c. What genes share the co-expression profile of An04g07430? Four of these genes have a correlation coefficient of 0.85. What are these genes? Can you tell from the product descriptions in the result table? The Orthologs table or the Attributes and Protein Browser section on their respective gene pages might offer more clues.
- d. Scan the product description column for genes with known functions. Use the Column Histogram tool to view a word cloud of the product descriptions in the column. Set the rank range to 25-50. What words occur most often in the product descriptions of An04g07430 co-expressed genes?



- e. Run the enrichment analyses for Molecular Function, Cellular Component and Biological Processes. Do these provide information about what this group of co-expressed genes might be doing?