

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 liver stage *Plasmodium berghei* 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.	✓		

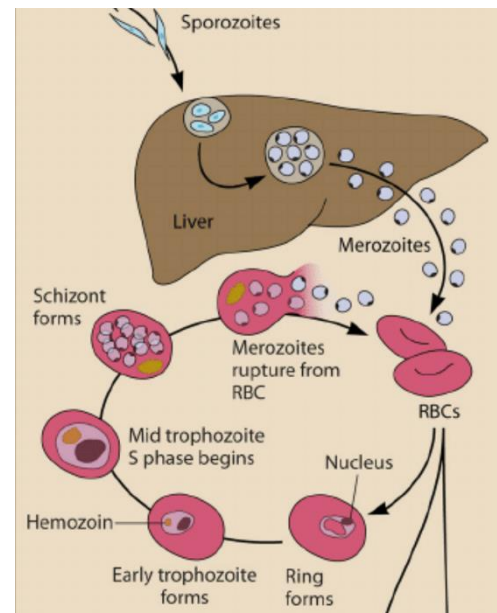
Metacycle	The MetaCycle package detects rhythmic signals from large scale time-series data, such as circadian rhythms within expression time courses, using either ARSER or JTK-Cycle. This search returns genes whose rhythmic signals match the conditions (period and amplitude range) you specify. The search 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 later liver stages of *Plasmodium* infection. [PlasmoDB.org](https://plasmodb.org)

The life cycle of *Plasmodium* is split between the sexual mosquito stage and the asexual host phase. The host phase includes a 6-7-day asymptomatic liver stage which ends with the release of merozoites into the bloodstream where they infect erythrocytes. The erythrocytic stages are well studied compared to the liver stages.

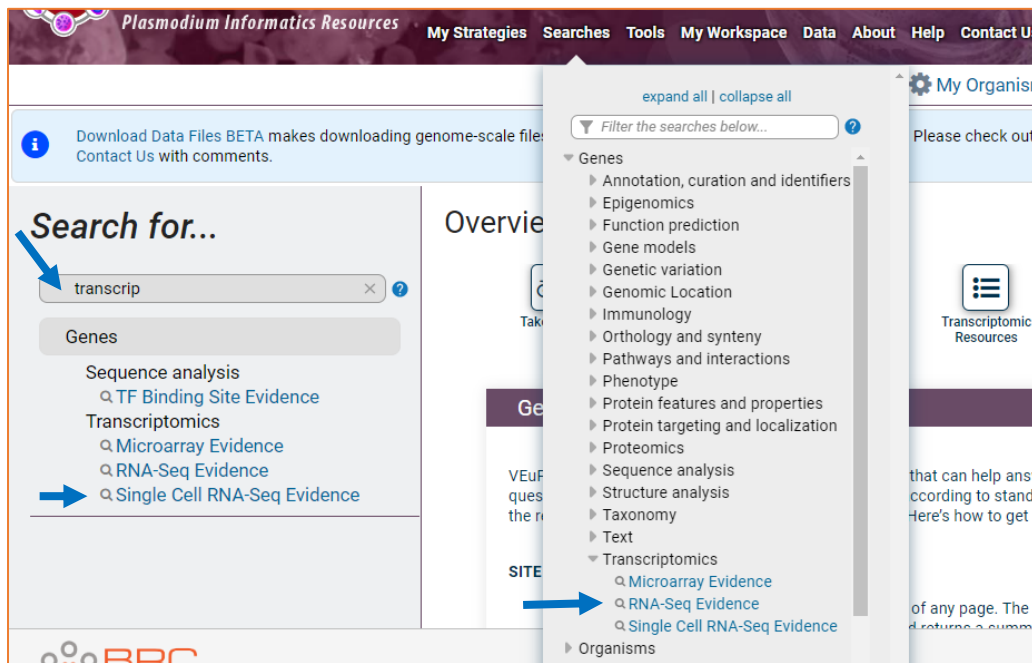
PlasmoDB contains RNA seq data from a study in the rodent model *Plasmodium berghei*, that includes a time course of liver infection as well as sporozoite and merozoite samples for comparison. ([Caledlari et al. 2019](#)) Seven samples were assayed in triplicate for RNA sequence:

1. Sporozoites
2. 6 hr liver infection
3. 24 hr liver infection
4. 48 hr liver infection
5. 54 hr liver infection
6. 60 hr liver infection
7. Merozoites (detached cells).



Use this data set to determine what genes are upregulated at least 4 fold (p-value ≤ 0.001) at 48 hr post infection vs the sporozoite stage.

- Navigate to the RNA seq search page and find the data set called **Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)**. 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: ☒ S Similarity ☒ DE Differential Expression ☒ FC Fold Change ☒ P Percentile ☒ SA SenseAntisense

Filter Data Sets: 4 results (filtered from a total of 54)

Organism	Data Set	Choose a Search
<i>Plasmodium berghei</i> ANKA	Transcriptome during early and mid-stage <i>P. berghei</i> liver infection (Toro-Moreno and Sylvester et al.)	DE FC P
<i>Plasmodium berghei</i> ANKA	Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)	DE FC P
<i>Plasmodium cynomolgi</i> strain M	Liver stage hypnozoite vs schizont transcriptomes (primary culture) (Voorverg-van der Wel et al.)	DE FC P
<i>Plasmodium vivax</i> P01	Sporozoite transcriptome in different microenvironments (Roth et al.)	DE FC P

- Arrange the differential expression search to return genes that are at least 4 fold up-regulated in the 48-hour liver infection compared to sporozoites with a p-value of $p < 0.001$.

Differential Expression
Fold Change
Percentile

Identify Genes based on P. berghei ANKA Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) RNA-Seq (Differential Expression)

Reset values

Experiment

☒ Ex-erythrocytic stage transcriptomes (sporozoite liver time course and detached cells) unstranded

Reference Sample

☒ sporozoite
☐ Liver 6h
☐ Liver 24h
☐ Liver 48h
☐ Liver 54h
☐ Liver 60h
☐ DC

Comparator Sample

☐ sporozoite
☐ Liver 6h
☐ Liver 24h
☒ Liver 48h
☐ Liver 54h
☐ Liver 60h
☐ DC

Direction

fold difference >=

adjusted P value less than or equal to

Pber ex-erythro RNAseq (de)
1,331 Genes
Step 1

+ Add a step

Get Answer

- How many genes were returned by the search? Do you believe these results? To convince yourself, you could browse the product description column. Are there clues that these genes are liver-specific?
- 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.

Unnamed Search Strategy *

Pber ex-erythro RNAseq (de)
1,331 Genes
Step 1

1,331 Genes (1,291 ortholog groups)

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

Details for step Pber ex-erythro RNAseq (de)
1331 Genes

Experiment Ex-erythrocytic stage transcriptomes (sporozoite liver t cells) unstranded

Reference Sample sporozoite

Comparator Sample Liver 48h

Direction up-regulated

fold difference >= 4

adjusted P value less than or equal to 0.0001

Give this search a weight

Pber ex-erythro RNAseq (de)
1,151 Genes
Step 1

- e. What other properties would you expect of a late liver stage gene/protein? Since the next step is to emerge from the hepatocyte, these genes may have proteolytic activity. Intersect your RNA seq search with a GO term search to see if any of your genes are annotated with proteolytic or peptidase activity. ([GO:0008233 peptidase activity](#) [GO:0006508 proteolysis](#)) How many genes have these activities?

Pber ex-erythro RNAseq (de)
1,151 Genes
Step 1

Add a step to your search strategy

Combine with other Genes

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

GO

Function prediction
Q: GO Term
Text
Q: Text (product name, notes, etc.)

← Add a step to your search strategy ?

Organism

62 selected, out of 62
select all | clear all | expand all | collapse all

Filter list below... ? ☐ Reference only

- ☒ Haemoproteidae
- ☒ Plasmodiidae

Evidence

- ☒ Curated
- ☒ Computed

select all | clear all

Limit to GO Slim terms

☐ Yes
☒ No

GO Term or GO ID

GO:0008233 : peptidase activity : 4 x GO:0006508 : proteolysis : 5 x

Pber ex-erythro RNAseq (de)
1,151 Genes

Step 1

GO Term
8,519 Genes

59 Genes

Step 2

+ Add a step

Run Step

2. **Using the same data set, find genes that are upregulated 4 fold in any liver stage compared to sporozoites.** Hint: use the Fold change search to compare the 6, 24, 48, 54 and 60-hour time points to sporozoites.
 - a. Navigate to the RNA Seq search page and choose the Fold Change search for the **Ex-erythrocytic (Caldelari et al 2019)** data set as in 1a above.
 - b. Arrange the fold change search to return genes that are up-regulated in the average expression across the liver stages compared to the sporozoites. (do not use the DC sample)

For the **Experiment**

☒ Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) unstranded

return **Genes**

that are

with a **Fold change** \geq

between each gene's **expression value**

(or a **Floor** of)

in the following **Reference Samples**

☒ sporozoite
☐ Liver 6h
☐ Liver 24h
☐ Liver 48h
☐ Liver 54h
☐ Liver 60h
☐ DC

select all | clear all

and its **expression value**

(or the **Floor** selected above)

in the following **Comparison Samples**

☐ sporozoite
☒ Liver 6h
☒ Liver 24h
☒ Liver 48h
☒ Liver 54h
☒ Liver 60h
☐ DC

select all | clear all

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

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 4.

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.

Get Answer

Pber ex-erythro RNAseq (fc)
 2,098 Genes

+ Add a step

Step 1

Gene ID	Transcript ID	Organism	Product Description	Fold Change	Chosen Ref (floor)	Chosen Comp (floor)	Pber ex-erythro RNAseq - tpm Graph
PBANKA_1003000	PBANKA_1003000.1	Plasmodium berghei ANKA	liver specific protein 2	1057.7	0.53 (1.09)	1149.01	
PBANKA_0518900	PBANKA_0518900.1	Plasmodium berghei ANKA	conserved Plasmodium membrane protein, unknown function	838.2	0.66 (5.36)	4495.2	
PBANKA_1203800	PBANKA_1203800.1	Plasmodium berghei ANKA	DnaJ protein, putative	609.5	57.36	34965.19	

[Download](#)
[Add to Basket](#)
[Add Columns](#)

- c. Explore your results. Did the search return more genes or fewer genes than the differential expression search that employed a statistical analysis along with fold change values?
- d. Use the Add Columns to turn on the TPM graph for the 'Ex-erythrocytic stages' data set. Notice the error bars for the DNAJ protein PBANKA_1203800. Would this gene be returned by the Differential Expression search that applies statistics before returning genes?
- e. Use the **Percentile search** to determine which genes in this result are also expressed in the top 10% of genes in the merozoite (detached cells) sample AND remove them from the result? Hint: Add a

step to the strategy that intersects your current result with search that returns the 90-100th percentile genes of the merozoite sample and choose 1 minus 2 for the operator.

Step 1
Pber ex-erythro RNAseq (fc)
2,098 Genes

Add a step to your search strategy

Combine with other Genes

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

RNA-Seq

Gene models
Long Read Evidence
Transcriptomics
RNA-Seq Evidence
Single Cell RNA-Seq Evidence

Search for Genes by RNA-Seq Evidence

The results will be intersected with the results of Step 1.

Legend: S Similarity DE Differential Expression FC Fold Change P Percentile SA SenseAntisense

Filter Data Sets: liver 4 results (filtered from a total of 54)

Organism	Data Set	Choose a Search
Plasmodium berghei ANKA	Transcriptome during early and mid-stage P. berghei liver infection (Toro-Moreno and Sylvester et al.)	DE FC P
Plasmodium berghei ANKA	Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)	DE FC P

Experiment
Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) unstranded

Samples

sporozoite
Liver 6h
Liver 24h
Liver 48h
Liver 54h
Liver 60h
DC
select all | clear all

Minimum expression percentile
90

Maximum expression percentile
100

Matches Any or All Selected Samples?
any

Protein Coding Only:
protein coding

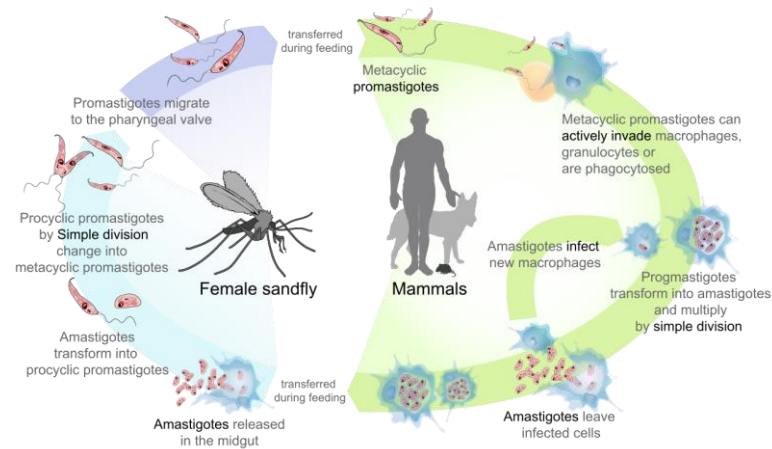
Step 1
Pber ex-erythro RNAseq (fc)
2,098 Genes

Step 2
Pber ex-erythro RNAseq (%ile)
499 Genes

1,833 Genes

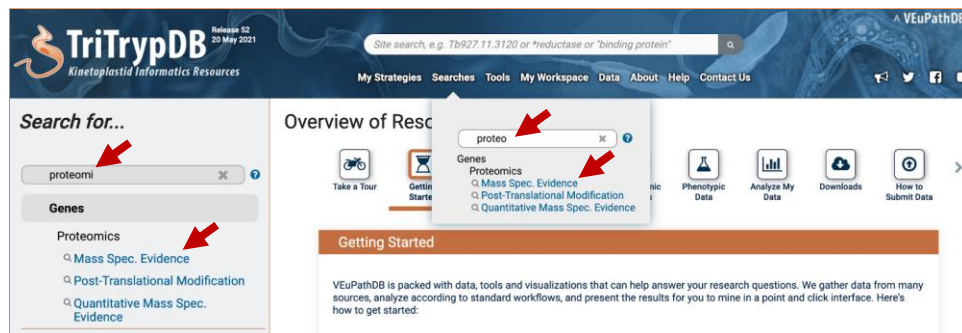
Add a step

3. 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 152
[select only these](#) | [add these](#) | [clear these](#)

infan

×

?

☐ Reference only

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

- How many genes did you get?
- 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.

Mass Spec
162 Genes

+ Add a step

Step 1

← Add a step to your search strategy ⓘ

Combine with other Genes

Mass Spec
162 Genes

Step 1

Step 2

Transform into related records

Mass Spec
162 Genes

Step 1

Step 2

Use Genomic Colocation to

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

proteom

Proteomics

Mass Spec. Evidence

Post-Translational Modification

Quantitative Mass Spec. Evidence

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

10 selected, out of 152

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

infan

☐ Reference only

☒ 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

Mass Spec
162 Genes

Step 1

Mass Spec
985 Genes

66 Genes

Step 2

+ Add a step

- f. 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)
- 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 settings that we used above return 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.

The image shows a workflow diagram at the top with three steps: 'Step 1' (Mass Spec, 162 Genes), 'Step' (Mass Spec, 985 Genes), and 'Step' (Mass Spec, 66 Genes). Arrows indicate a sequence from Step 1 to Step, and then to the final Step. Each step has an 'Edit' button. A dashed box labeled '+ Add a step' is also present.

Below the diagram is a 'Details for step' popup for 'Mass Spec' (162 Genes). It shows the following settings:

- 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

At the bottom of the popup is a 'Revise' button, which is circled in orange.

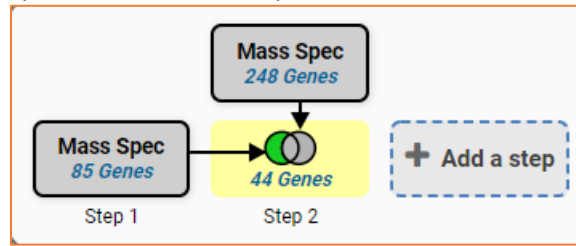
Below the popup is the 'Revise your step' configuration page. It has tabs for 'Configure Search', 'Learn More', and 'View Data Sets Used'. A 'Reset values to default' button is at the top.

The 'Experiments and Samples' section shows '1 selected, out of 152' with a filter list below it containing 'Leishmania' and 'Trypanosoma'. A 'Reference only' checkbox is also present.

The 'Minimum Number of Unique Peptide Sequences' section has a text input field with the value '5' entered, indicated by an orange arrow. Below it is a section for 'Apply min # peptide sequences / sample OR across samples' with a dropdown menu set to 'Per Sample'.

At the bottom of the configuration page is an 'Advanced Parameters' section and a 'Revise' button, which is circled in orange.

h. How did this change your results? Would you consider these results more stringent?



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: 2 results (filtered from a total of 5)

Organism	Data Set	Choose a Search
<i>Plasmodium berghei</i> ANKA	Proteome of ApiAP2 double vs single knockout (Modrzynska et al)	DCC
<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)

For the Experiment

☒ Apicoplast and ER Proteomes (Quantitative)(Dd2)

return protein coding Genes

that are up-regulated

with a **Fold change** \geq

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:

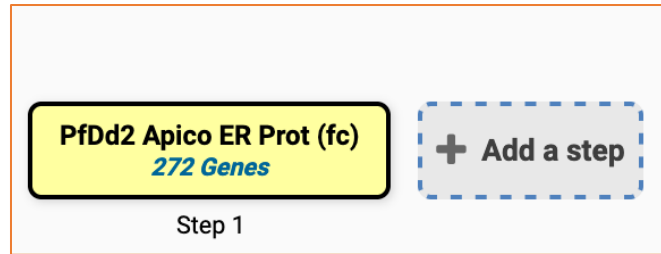
$$\text{fold change} = \frac{\text{comparison expression value}}{\text{reference expression value}}$$

and returns genes when **fold change** \geq 1.5.

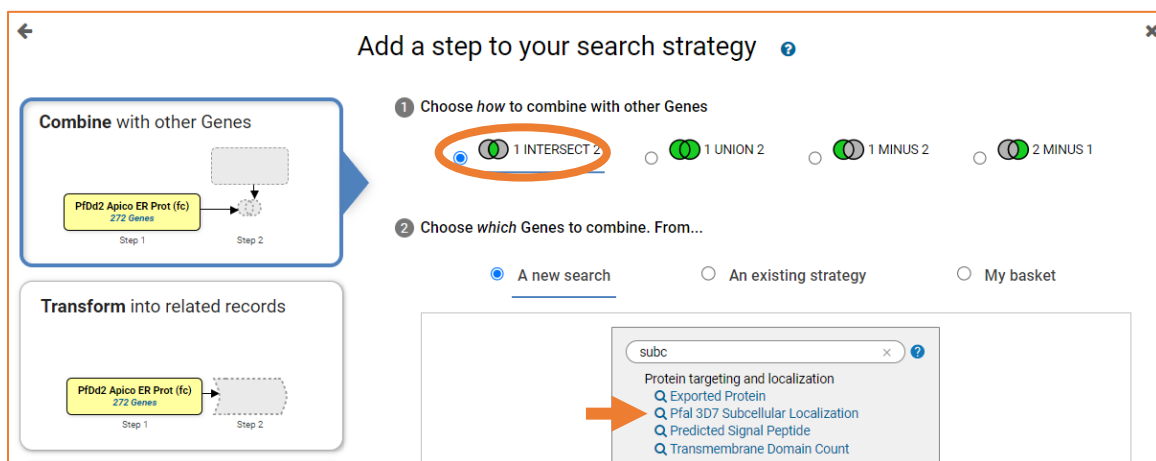
You are searching for genes that are **up-regulated** between one **reference sample** and one **comparison sample**.

Get Answer

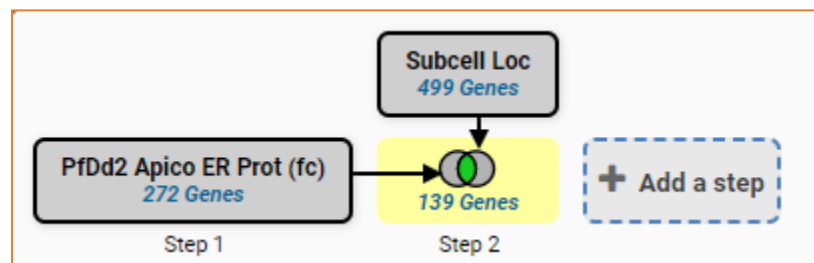
COMMUN



- 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'



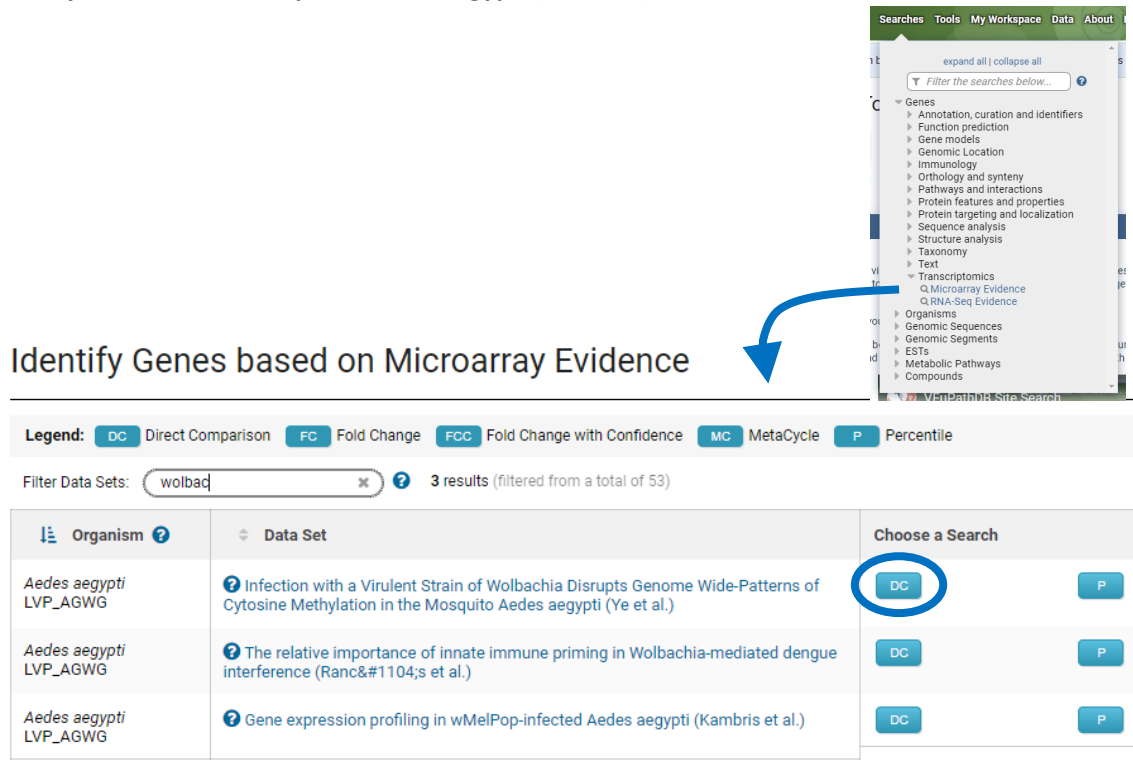
- e. Make sure Apicoplast localization is selected and click on the Run Step button. How many genes did you identify? Are you more confident that these genes are apicoplast genes? How would you use the PlasmoDB tools to boost your confidence in these so called apicoplast genes?



OPTIONAL exercises below concerning transcriptomics and proteomics. Result numbers in images may not match the current database

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.)**'

Identify Genes based on Microarray Evidence



The screenshot shows the VectorBase search interface. At the top, there is a navigation bar with links: Searches, Tools, My Workspace, Data, and About. Below this is a search bar with the text 'Filter the searches below...'. A dropdown menu is open, showing a list of search categories: Genes, Annotation, curation and identifiers, Function prediction, Gene models, Genomic Location, Immunology, Orthology and synteny, Pathways and interactions, Protein features and properties, Protein targeting and localization, Sequence analysis, Structure analysis, Taxonomy, Text, Transcriptomics, Q.Microarray Evidence, Q.RNA-Seq Evidence, Organisms, Genomic Sequences, Genomic Segments, ESTs, Metabolic Pathways, and Compounds. A blue arrow points from the 'Q.Microarray Evidence' category to the 'Identify Genes based on Microarray Evidence' section.

Below the search bar, there is a legend for search types: DC (Direct Comparison), FC (Fold Change), FCC (Fold Change with Confidence), MC (MetaCycle), and P (Percentile). The 'Filter Data Sets' section shows the search term 'wolbach' and the results: 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 <i>Wolbachia</i> 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 <i>Wolbachia</i> -mediated dengue interference (Ranc#1104;s 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**.

Direct Comparison Percentile

Identify Genes based on A. aegypti LVP_AGWG Infection with a Virulent Strain of Wolbachia Disrupts Genome Wide-Patterns of Cytosine Methylation in the Mosquito Aedes aegypti Microarray (direct comparison)

Configure Search Learn More View Data Sets Used

Reset values to default

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

Get Answer

Wolbachia infection in head an...
695 Genes

+ Add a step

Step 1

- 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

+ Add a step

Step 1

Add a step to your search strategy

Combine with other Genes

Choose how to combine with other Genes

1 INTERSECT 2 UNION 2 1 MINUS 2 2 MINUS 1

Choose which Genes to combine. From...

A new search An existing strategy

RNA

Gene models
Q Gene Model Characteristics
Transcriptomics
Q Microarray Evidence
Q RNA-Seq Evidence

Direct Comparison Percentile

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

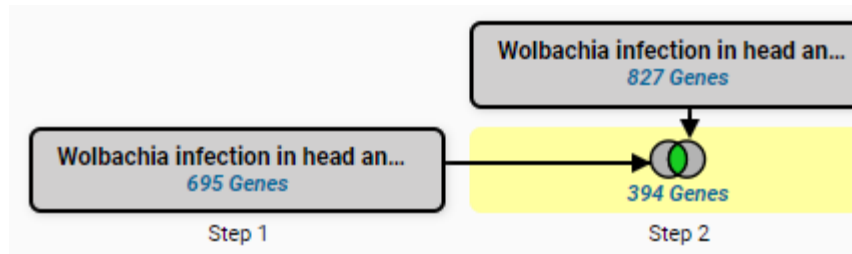
Fold difference >=

2.0

Protein Coding Only:

protein coding

Run Step



- 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?

Analyze your Gene results with a tool

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
- Limit to GO Slim terms:**
 - ☒ No
 - ☐ Yes
- P-Value cutoff:** (0 - 1)

Submit

Analysis Results:

81 terms

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

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	3.40	4.18	1.93e-13	1.87e-11
GO:0004252	serine-type endopeptidase activity	346	38	11.0	3.99	4.83	2.21e-13	1.67e-11

3. 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](#).

- b. Navigate to the microarray searches in FungiDB and choose the Coexpression search for the data set titled ***Aspergillus niger* gene co-expression network (Vera Meyer)**. [Schape et al Nucleic Acids Research 2019](#). This 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.
- c. Run the search to find the co-expression network for An04g07430.

Identify Genes based on Microarray Evidence

Filter Data Sets: Legend: **C** Coexpression **DC** Direct Comparison **FC** Fold Change **P** Percentile

Organism	Data Set	Choose a Search
<i>Aspergillus fumigatus</i> Af293	Response to hypoxia (Barker et al. 2012)	FC P
<i>Aspergillus niger</i> CBS 513.88	Aspergillus niger gene co-expression network (Vera Meyer)	C
<i>Candida albicans</i> SC5314	Antifungal Benzimidazole Derivative Response (Steffen Rupp)	DC P

Identify Genes based on Coexpression

Reset values

Gene ID input set

☒ Enter a list of IDs or text:

☐ Upload a text file: No file chosen
Maximum size 1GB. The file should contain the list of IDs.

☐ Copy from My Basket: 0 records will be copied from your basket.

☐ Copy from My Strategy: Pyrimidine metabolism (ec00240) (KEGG) (60 records)

Correlation

Positive Correlation

Spearman coefficient (greater or equal to)

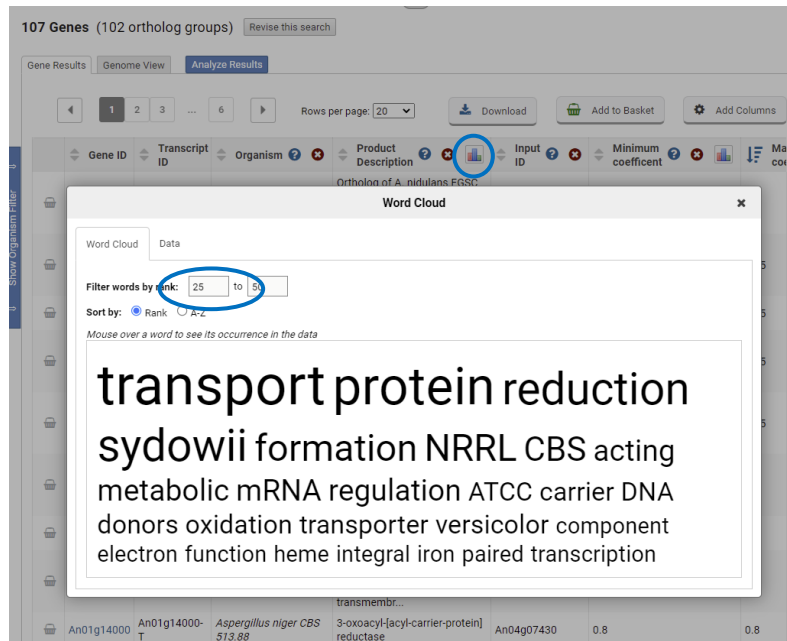
Get Answer

Coexpression
107 Genes

+ Add a step

Step 1

- d. What genes share the co-expression profile of An04g07430? Several genes have a correlation coefficient of 0.85. What are these genes? Visit their gene pages to learn more.
- e. 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?



- f. 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?