

## Transcriptomics: RNA sequence and microarray data searches

### Learning Objectives

- Review the types of expression searches in VEuPathDB
- Use the differential expression, fold change and percentile search to explore gene expression
- Compare the expression searches to reveal advantages and disadvantages of each search

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) and for RNA seq, expression is graphed on gene pages and can be visualized in the genome browser. 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
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.	✓	

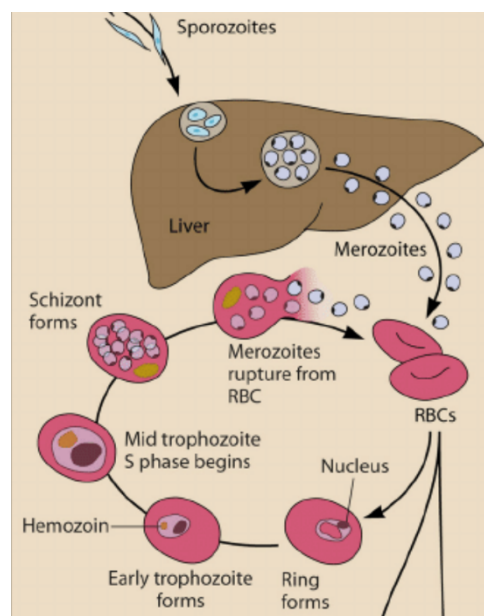
Metacycle	The MetaCycle package detects rhythmic signals from large scale time-series data, such as circadian rhythms within expression time courses, using either ARSER or JTK-Cycle. This search returns genes whose rhythmic signals match the conditions (period and amplitude range) you specify. The search 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.		✓
Coexpression	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.		✓

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 stage 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. ([Caldelari et al. 2019](#)) Seven samples were assayed in triplicate for RNA sequence:

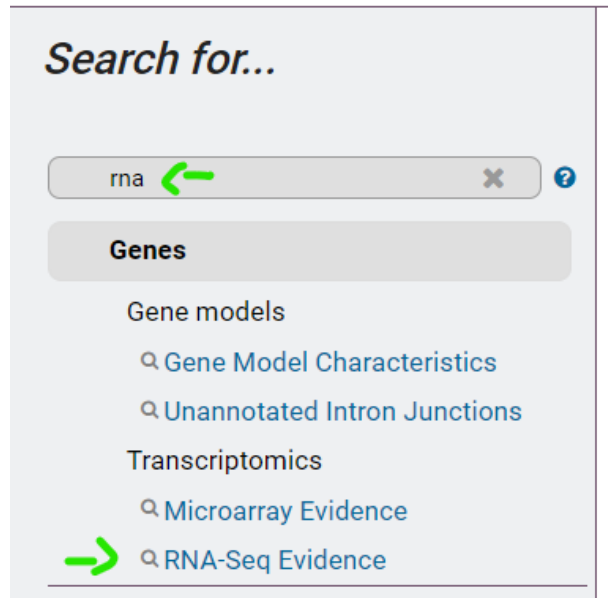
1. Sporozoites
2. 6 hr liver infection
3. 24 hr liver infection
4. 48 hr liver infection
5. 54 hr liver infection
6. 60 hr liver infection
7. Merozoites (detached cells).



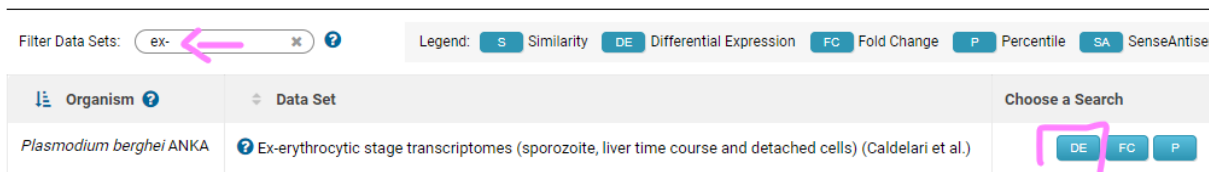
**Use this data set to determine what genes are upregulated at least 4 fold ( $p\text{-value} \leq 0.001$ ) at 48 hr post infection (liver stage) vs the sporozoite (stage injected in the host by the mosquito vector).**

- a. 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



- b. Arrange the differential expression search to return genes that are at least 4 fold up-regulated in the 48-hour liver infection (comparator) vs sporozoites (reference),  $p \leq 0.001$ .

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

☒ 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

up-regulated

### fold difference >=

4

### adjusted P value less than or equal to

0.001

COMMUNITY CH

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

+ Add a step

1,331 Genes (1,291 ortholog groups)

Revise this search

Organism Filter

select all | clear all | expand all | collapse all

☐ Hide zero counts

Search organisms...

☐ Hepatocytis sp. ex Pilocolobus tephrosceles 2019

0

☒ Plasmodium

1,331

select all | clear all | expand all | collapse all

☐ Hide zero counts

Hide Organism Filter

Gene Results

Genome View

Analyze Results

Genes: 1,331

Transcripts: 1,333

☐ Show Only One Transcript Per Gene

◀

1

2

3

...

67

▶

Rows per page: 20

Download

Add to Basket

Add Columns

Gene ID	Transcript ID	Organism	Product Description	Fold Change	Adjusted P value
<div></div> PBANKA_1003000	PBANKA_1003000.1	Plasmodium berghei ANKA	liver specific protein 2	1325.8	1.022E-90
<div></div> PBANKA_1237900	PBANKA_1237900.1	Plasmodium berghei ANKA	mitochondrial-processing peptidase subunit alpha, putative	722.63	2.433E-10
<div></div> PBANKA_1358600	PBANKA_1358600.1	Plasmodium berghei ANKA	isocitrate dehydrogenase [NADP], mitochondrial, putative	607.2	9.105E-10

- c. 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?
- 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.

**My Search Strategies**

Opened (1) All (1) Public (51) Help

Unnamed Search Strategy \*

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

**Details for step** *Pber ex-erythro RNAseq (de)*

1331 Genes

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

**Reference Sample** sporozoite

**Comparator Sample** Liver 48h

**Direction** up-regulated

**fold difference >=** 4

**adjusted P value less than or equal to** 0.001

► Give this search a weight

- 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**

← Add a step to your search strategy ?

**Combine with other Genes**

Step 1 → Step 2

**Transform into related records**

Step 1 → Step 2

- Choose *how* to combine with other Genes
 

☒ 1 INTERSECT 2
 ☐ 1 UNION 2
 ☐ 1 MINUS 2
- Choose *which* Genes to combine. From...
 

☒ A new search
 ☐ An existing strategy

Function prediction  
  
 Text

← Add a step to your search strategy

56 selected, out of 56

select all | clear all | expand all | collapse all

Filter list below... ?

- ☒ Hepatocystis sp. ex Piliocolobus tephrosceles 2019 [Reference]
- ☒ Plasmodium

select all | clear all | expand all | collapse all

**? Evidence**

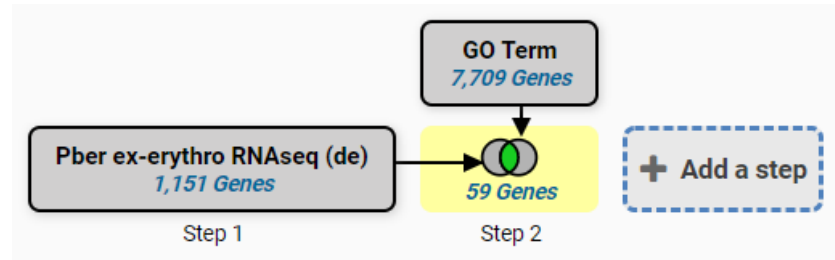
- ☒ Curated
- ☒ Computed

select all | clear all

**? Limit to GO Slim terms**

- ☐ Yes
- ☒ No

**? GO Term or GO ID**



2. 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.

## Identify Genes based on RNA-Seq Evidence

Filter Data Sets:    Legend:  Similarity  Differential Expression  Fold Change  Percentile  SenseAntisense

Organism	Data Set	Choose a Search
<i>Plasmodium berghei</i> ANKA	Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)	<input type="button" value="DE"/> <input type="button" value="FC"/> <input type="button" value="P"/>

- 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. Click on Get Answer.

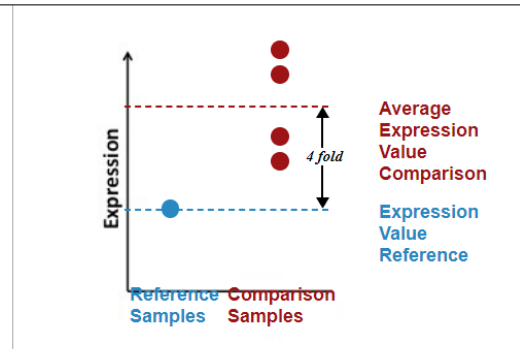
Identify Genes based on *P. berghei* ANKA Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) RNA-Seq (fold change)

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

 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



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

For each gene, the search calculates:

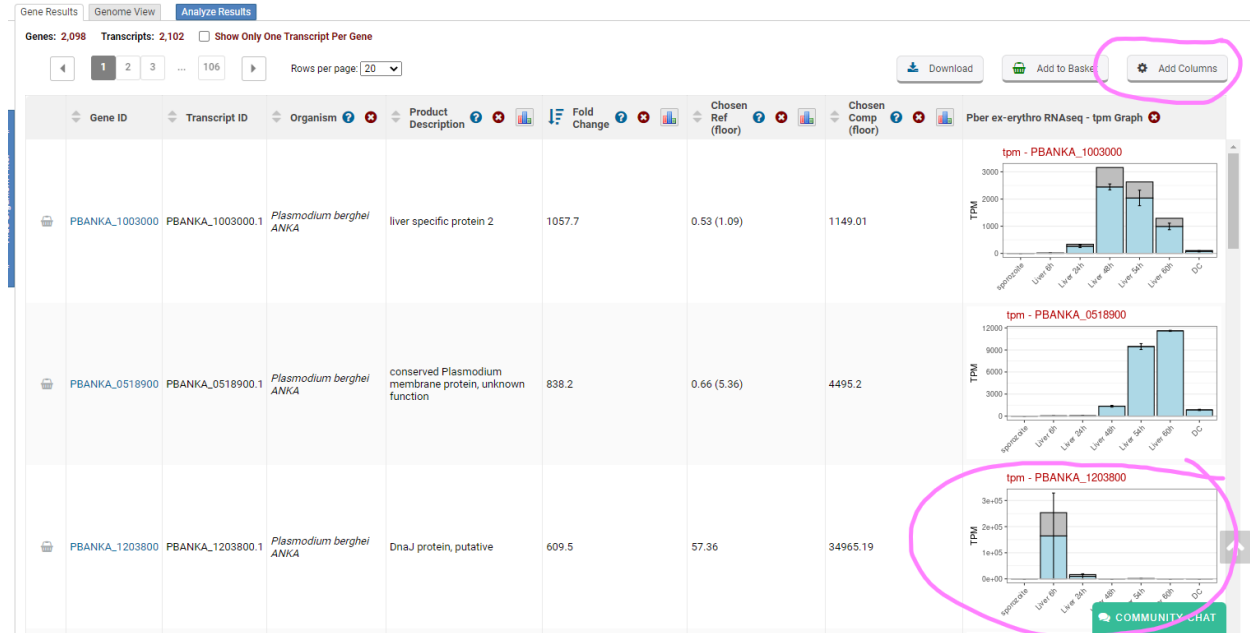
$$\text{fold change} = \frac{\text{average expression value in comparison}}{\text{reference expression value}}$$

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.

- c. Explore your results. Did the search return more genes or fewer genes than the differential expression search?
- 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?



3. Use the Percentile search to determine what genes in this result list are also expressed in the top 10% of genes in the merozoite (detached cells) sample? Hint: Add a step to the strategy that intersects your current result with search that returns the 90-100<sup>th</sup> percentile genes of the merozoite sample.

Edit

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

Step 1

+

Add a step



← Add a step to your search strategy ?

### Combine with other Genes

Step 1

Step 2

### Transform into related records

Step 1

Step 2

### Use Genomic Colocation to

1 Choose *how* to combine with other Genes

☒ 1 INTERSECT 2 ☐ 1 UNION 2

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

☒ A new search ☐ An existing strat

RNA

Gene models

- Gene Model Characteristics
- Unannotated Intron Junctions

Transcriptomics

- Microarray Evidence
- RNA-Seq Evidence

## Search for Genes by RNA-Seq Evidence

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

Filter Data Sets: EX- ?

Legend: Similarity Differential Expression Fold Change Percentile SenseAntisense

Organism ?	Data Set	Choose a Search
<i>Plasmodium berghei</i> ANKA	Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)	



- 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.)'



**Search for...**

micro ← × ?

**Genes**

Transcriptomics

Microarray Evidence ←

## Identify Genes based on Microarray Evidence


Filter Data Sets: wolbachia × ?

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

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.

## 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)

 Reset values

### ? 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

 COMMUNITY CHAT

- c. Intersect your search result with another search that returns genes upregulated 2 fold in muscle infected vs muscle 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**

← Add a step to your search strategy ?

**Combine with other Genes**

Step 1 Step 2

**Transform into related records**

Step 1 Step 2

1 Choose *how* to combine with other Genes

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

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

☒ A new search
 ☐ An existing strategy

×
?

Transcriptomics  
Q. Microarray Evidence

← Add a step to your search strategy

Direct Comparison Percentile

**? Experiment**

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

**? Direction**

up or down 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?

Wolbachia infection in head an... 827 Genes

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

Wolbachia infection in head an... 394 Genes Step 2

+ Add a step

394 Genes (355 ortholog groups)

Gene Results Genome View **Analyze Results**

Analyze your Gene results with a tool below.

GO

Gene Ontology Enrichment

### Gene Ontology Enrichment

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

▼ Parameters

Organism [?](#) Aedes aegypti LVP\_AGWG

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

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

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

P-Value cutoff [?](#) 0.05 (0 - 1)

Submit

Analysis Results:

77 rows

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

GO ID <a href="#">?</a>	GO Term <a href="#">?</a>	Genes in the bkgd with this term <a href="#">?</a>	Genes in your result with this term <a href="#">?</a>	Percent of bkgd genes in your result <a href="#">?</a>	Fold enrichment <a href="#">?</a>	Odds ratio <a href="#">?</a>	P-value <a href="#">?</a>	Benjam <a href="#">?</a>
GO:0003824	catalytic activity	3624	162	4.5	1.62	2.38	9.32e-14	1.57e-11
GO:0003824	catalytic activity	3624	162	4.5	1.62	2.38	9.32e-14	1.57e-11