

# Mining Transcriptomics Data

## Genes up-regulated in liver stages of *Plasmodium* infection

### Learning objectives

- Review the types of expression searches in VEuPathDB
- Use the differential expression and fold change searches to explore gene expression in liver stage *Plasmodium* infections

### Transcriptomics data in VEuPathDB

Transcript expression, or the abundance of an mRNA, can be determined in the laboratory with several different techniques including RNA sequencing, microarray, and RT-PCR. VEuPathDB supports these data types with several searches, as shown in the table below. For RNA sequence data, expression values are graphed on gene pages and mapped reads 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	Microarray
Differential Expression	<b>Statistical</b> analysis of studies whose experimental design includes <b>biological replicates</b> . Differential expression (DE) methods rely on replicates to estimate how much gene expression naturally varies within a condition. A DE search finds genes based on fold change difference between two samples with a user defined p-value cutoff. <b>Only pairwise comparisons</b> can be made with this search.	✓	
Fold Change	Expression differences between samples are calculated but <b>statistical analyses are not performed</b> . 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 <b>did not include replicates</b> and <b>allows for comparing groups of samples</b> , 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' <b>expression value is sorted from lowest to highest</b> and a <b>percentile rank</b> 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 follows 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.		✓

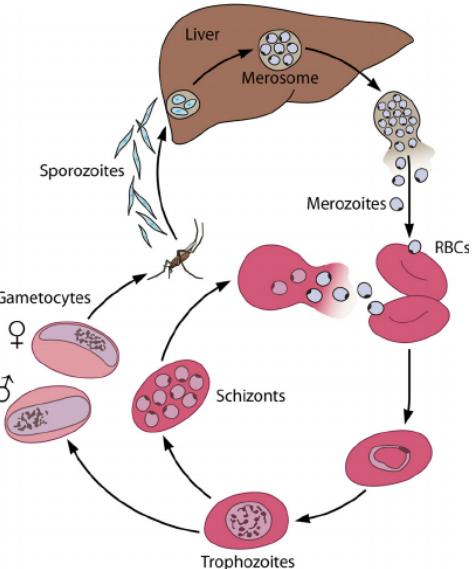
## Introduction to the exercise

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 ([Caledari et al. 2019](#)).

Seven samples were assayed in triplicate for RNA sequence data:

- Sporozoites
- 6 hr liver infection



- 24 hr liver infection
- 48 hr liver infection
- 54 hr liver infection
- 60 hr liver infection
- Merozoites (detached cells)

The purpose of this two-part exercise is to use data from this study to explore gene up-regulation in liver stages of *Plasmodium* infection.

## Part 1: Determine what genes are up-regulated at least 4 fold ( $p\text{-value} \leq 0.001$ ) at 48 hr post infection vs. the sporozoite stage

1. Navigate to the appropriate search for this question
  - a. Go to [PlasmoDB.org](https://plasmodb.org)
  - b. From the sidebar or header, search for “RNA” and click on **RNA-Seq Evidence**

The screenshot shows the PlasmoDB homepage. A red arrow points to the search bar where "RNA" has been typed. Another red arrow points to the search results dropdown, which lists "RNA-Seq Evidence" under the "Genes" category. The main content area displays an "Overview of Resources and Tools" with sections for "Take a Tour", "Sustainability and Future Directions", and "Funding". A sidebar on the right provides information about VEuPathDB genomic databases and their future outlook.

2. Choose appropriate experiments/ samples & parameters within the search
  - a. Find the data set called “Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)”
  - b. Select the DE (differential expression) search

The screenshot shows the "Identify Genes based on RNA-Seq Evidence" search interface. A red arrow points to the "Organism" dropdown menu, which is set to "Plasmodium berghei ANKA". Another red circle highlights the "DE" (Differential Expression) button in the "Choose a Search" row. The search results table shows one result for the specified dataset.

Identify Genes based on RNA-Seq Evidence	
Legend: S Similarity DE Differential Expression FC Fold Change W WGNA P Percentile SA SenseAntisense	
Filter Data Sets: ex- (dropdown) 1 result (filtered from a total of 58)	
Organism (dropdown): Plasmodium berghei ANKA	Data Set (dropdown): Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)
Choose a Search	
DE FC P	

3. Configure 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  $p < 0.001$ .

How many genes did you get? Do you believe these results? To convince yourself, you could browse the product description column of the results table. Are there clues that these genes are liver-specific?

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

Configure Search Learn More View Data Sets Used

Reset values to default

**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**

Up-regulated

**fold difference  $\geq$**

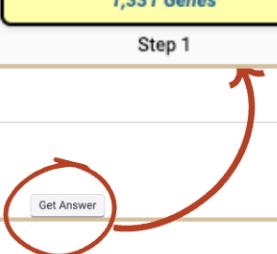
**adjusted P value less than or equal to**

**Up-reg Liver 48h vs Sporozoites \*** 

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

**+ Add a step**

Step 1



4. Increase the statistical stringency of the search from  $p \leq 0.001$  to  $p < 0.0001$ .

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.

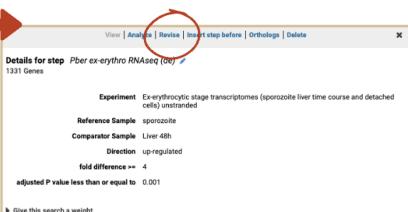
How many genes are returned by the search now?

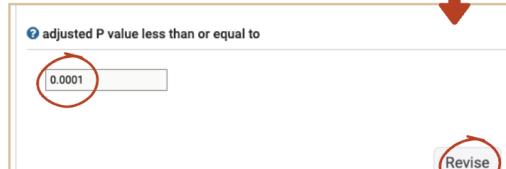
**Up-reg Liver 48h vs Sporozoites \*** 

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

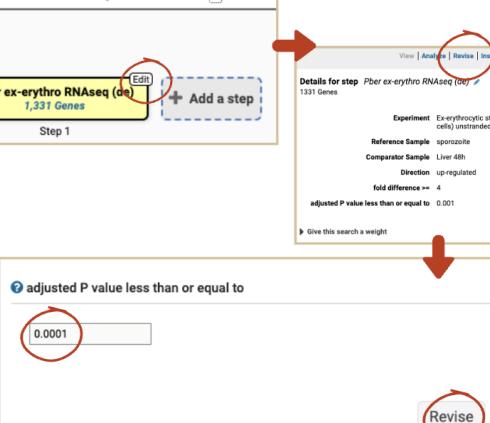
**+ Add a step**

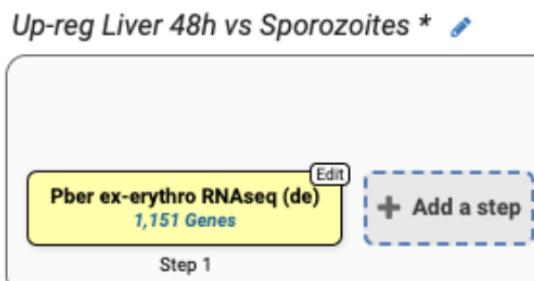
Step 1





adjusted P value less than or equal to





5. 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. How could you integrate functional annotation into your search strategy?
6. Add a step to your search to add functional annotation. 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 are up-regulated in late liver stages compared to sporozoites AND have been annotated with proteolytic or peptidase activity?

Up-reg Liver 48h vs Sporozoites \*

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

Step 1

+ Add a step

Add a step to your search strategy

Choose how to combine with other Genes

- 1 INTERSECT 2
- 1 UNION 2
- 1 MINUS 2
- 2 MINUS 1

Choose which Genes to combine. From...

A new search    An existing strategy    My basket

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

Search for Genes by GO Term

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

Configure Search    Learn More    View Data Sets Used

Reset values to default

Organism

62 selected out of 42

select all | clear all | collapse all

Other filters:  Human and Malaria  Plasmodium

Evidence

Curated    Computed

Limit to GO Slim terms

Yes    No

GO Term or GO ID

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

GO Term wildcard search

Run Step

Up-reg Liver 48h vs Sporozoites \*

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

Step 1

GO Term  
8,519 Genes

Step 2

+ Add a step

## Part 2: Determine what genes are up-regulated 4 fold in any liver stage compared to sporozoites

1. Navigate to the same dataset that was used in the first part of this exercise- "Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)"
2. For this question, select the FC (fold change) search. A fold change search finds genes whose expression value differs between samples without considering statistical parameters. It allows for comparing groups of samples.

Identify Genes based on RNA-Seq Evidence

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

Filter Data Sets: Ex... 1 result (filtered from a total of 58)

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

3. Configure the fold change search to return genes that are 4 fold up-regulated in the average expression across all the liver stages compared to the sporozoites.

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 (fold change)

Configure Search Learn More View Data Sets Used

For the Experiment  
 Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) unstranded  
 protein coding  
 Genes  
 that are Up-regulated  
 with a Fold change >= 4  
 between each gene's Average expression value  
 (or a Floor of 10 reads)  
 in the following Reference Samples  
 sporozoite  
 Liver 6h  
 Liver 24h  
 Liver 48h  
 Liver 54h  
 Liver 60h  
 select all | clear all

return protein coding Genes that are Up-regulated with a Fold change >= 4 between each gene's Average expression value (or a Floor of 10 reads) in the following Reference Samples sporozoite Liver 6h Liver 24h Liver 48h Liver 54h Liver 60h select all | clear all

and its Average expression value (or the Floor selected above) in the following Comparison Samples  
 sporozoite  
 Liver 6h  
 Liver 24h  
 Liver 48h  
 Liver 54h  
 Liver 60h  
 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

Average Expression Value Comparison

Up-ref Liver vs. Sporozoites \*

Pber ex-erythro RNAseq (fc) 2,098 Genes  + Add a step

Step 1

How many genes did you get?

Did the search return more genes or fewer genes than the differential expression search?

4. Explore search results. 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?

Select Columns X

**6 columns selected, out of 80 columns allowed**

[select only these](#) | [add these](#) | [clear these](#) | [reset to current](#) | [reset to default](#)

**TPM graph** ?

- Pf3D7 iRBC cycle RNA-Seq - sense tpm graph
- P. falciparum 3D7 Intraerythrocytic development cycle transcriptome (2018) (Toenake et al.)
- Pfal IDC 2018 RNA-Seq - antisense tpm graph
- Pfal IDC 2018 RNA-Seq - Both\_strands tpm graph
- Pfal IDC 2018 RNA-Seq - sense tpm graph
- P. vivax P01 Patient isolates cultured through the intraerythrocytic development cycle (Rangel et al.)
- Pviv 3-patient IDC RNASEq - tpm Graph
- P. falciparum 3D7 Intraerythrocytic development cycle transcriptome by DAFT-Seq (3D7, HB3, IT, 2020) (Chappell et al. 2020)
- Pfal IDC 3D7, Hb3, IT RNASEq - antisense tpm graph
- Pfal IDC 3D7, Hb3, IT RNASEq - Both\_strands tpm graph
- Pfal IDC 3D7, Hb3, IT RNASEq - sense tpm graph
- P. berghei ANKA Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)
- Pber ex-erythro RNAseq - tpm Graph
- P. falciparum 3D7 Oocyst and salivary gland sporozoite transcriptome comparison in P. falciparum (Lindner et al.)
- Pf3D7 Oocyst vs SG sporo RNAseq - antisense tpm graph
- Pf3D7 Oocyst vs SG sporo RNAseq - Both\_strands tpm graph
- Pf3D7 Oocyst vs SG sporo RNAseq - sense tpm graph
- P. yoelii yoelii 17XNL 2023 Oocyst and salivary gland sporozoite transcriptome comparison in P. yoelii yoelii 17XNL (Lindner et al. 2019)
- Py17XNL2023 Oocyst vs salivary sporo RNAseq - antisense tpm graph
- Py17XNL2023 Oocyst vs salivary sporo RNAseq - Both\_strands tpm graph
- Py17XNL2023 Oocyst vs salivary sporo RNAseq - sense tpm graph
- P. vivax P01 Transcriptome of P. vivax salivary gland sporozoites (Muller et al.)
- Pviv - tpm Graph
- P. falciparum 3D7 Polysomal and steady-state asexual stage transcriptomes (Bunnik et al.)
- Pf3D7 Polysomal/Steady iRBC RNA-Seq - tpm Graph

[Update Columns](#)

Gene Results Gene View Analyze Results

Gene: 2,098 Transcript: 2,102  Show Only One Transcript Per Gene  Show only the Genes in my basket.

[Download](#) [Send to...](#) [Add Columns](#)

Organism	Product Description	Fold Change	Chosen Ref	Chosen Comp	Chosen Comp (Floor)	Pber ex-erythro RNAseq - tpm Graph
Isosmoldum bergheli ANKA	conserved Plasmodium membrane protein, unknown function	838.2	0.64 (3.24)	4495.2		
Isosmoldum bergheli ANKA	DnaJ protein, putative	609.5	57.36	34965.19		

5. Add a step to the search strategy to determine what genes in this result are also represented in the top 10% of genes in the merozoite (aka detached cells or DC) sample.

How many genes are up-regulated in liver stages compared to sporozoites and also represented in the top 10% of genes expressed in detached cells (merozoites)?

**Up-ref Liver vs. Sporozoites \***

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

Add a step

**Add a step to your search strategy**

Combine with other Genes

Choose how to combine with other Genes:

- 1 INTERSECT 2 (selected)
- 1 UNION 2
- 1 MINUS 2
- 2 MINUS 1

Choose which Genes to combine. From...:
 

- A new search (selected)
- An existing strategy
- My basket

**Search for Genes by RNA-Seq Evidence**

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

Legend: Similarity Differential Expression Fold Change WGCNA Percentile RNA-Seq Evidence

Filter Data Sets: valid | Result (Based on a total of 50)

Organism: Data Set: Choose a Search

Plasmodium falciparum (ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Cardozo et al.)

Differential Expression | Fold Change | Percentile

Configure Search | Learn More | View Data Sets Used

Reset values to default

Experiment:
 

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

Samples:
 

- sporozoite
- Liver 1h
- Liver 24h
- Liver 48h
- Liver 60h
- DC (selected)

select all | clear all

Minimum expression percentile: 90

Maximum expression percentile: 100

Matches Any or All Selected Samples? any

Protein Coding Only: protein coding

**Up-ref Liver vs. Sporozoites \***

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

Step 1

Step 2

285 Genes

Add a step

Run Step