



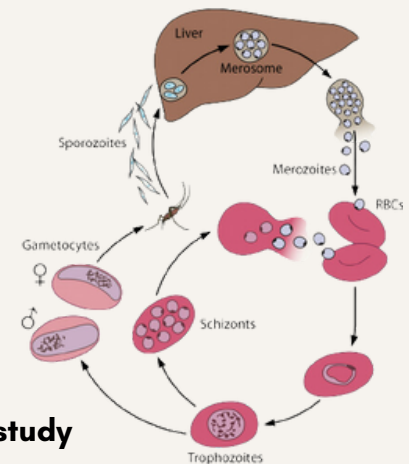
## The search question

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.**

[Image credit](#)

### What platform is best suited for this question?



This knowledgebase has data and informatics resources for *Haemoproteus*, *Hepatocystis*, *Plasmodium*.

### What search is best suited for this question?

- ▼ Transcriptomics
  - Microarray Evidence
  - RNA-Seq Evidence
  - Single Cell RNA-Seq Evidence

**RNA Seq Evidence** offers a **Differential Expression search** that finds genes based on fold change difference between two samples (pairwise comparisons) with a user defined p-value cutoff and a **Fold Change search** suitable for comparing groups of samples.

### What is a possible search strategy that will answer this question?

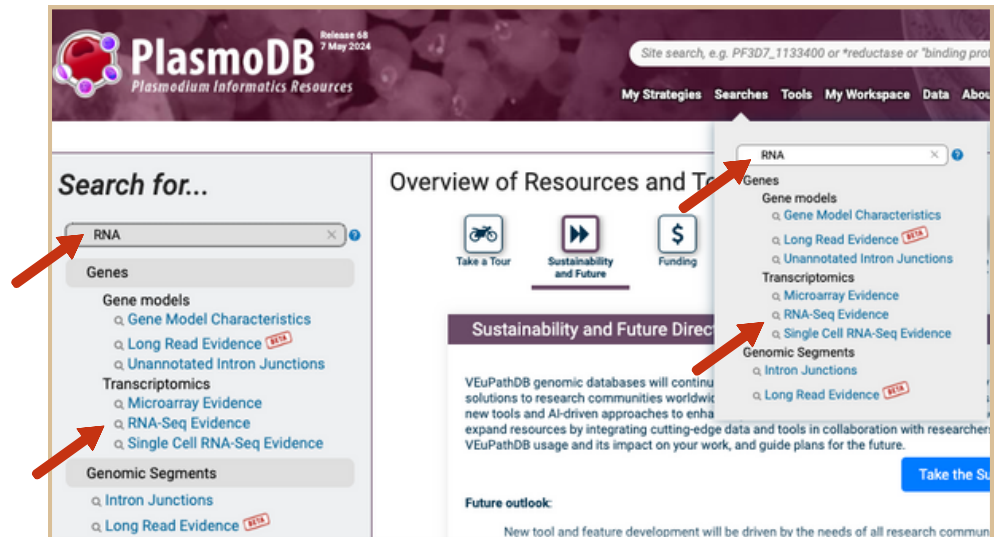
Part 1: Use the RNA seq **Differential Expression** search to identify genes up-regulated in late liver vs. sporozoite stages, adjusting the stringency of the search as needed.

Part 2: Use the **Fold Change** search to identify genes upregulated in any liver stage compared to sporozoites.

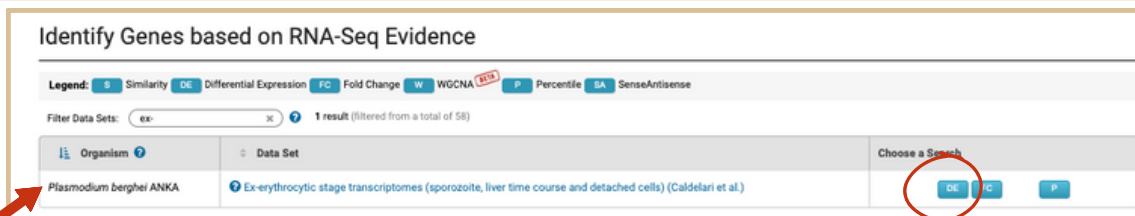
## A. Determine what genes are up-regulated at least 4 fold (p-value $\leq 0.001$ ) at 48 hr post infection vs. the sporozoite stage

### A1. Navigate to the appropriate search

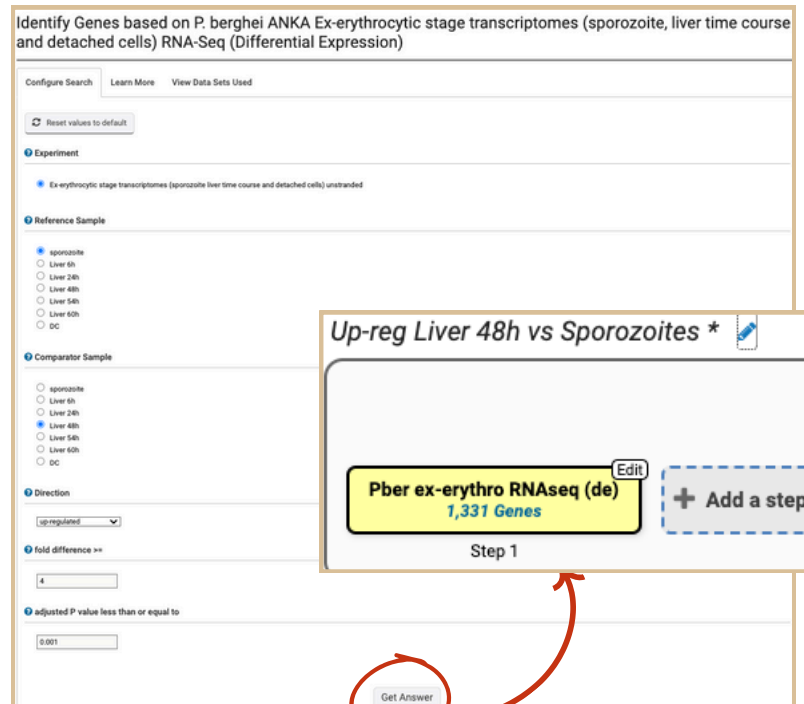
- Navigate to [PlasmoDB.org](https://plasmodb.org)
- From the sidebar or header, search for “RNA” and click on RNA-Seq Evidence



### A2. Choose appropriate experiments/ samples & parameters within the search



- Find the data set called “Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)”
- Select the DE (differential expression) search
- 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. Are there clues that these genes are liver-specific?

### A3. Make the search more stringent

- 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.

The first screenshot shows a search step titled "Up-reg Liver 48h vs Sporozoites \*". The step is "Pber ex-erythro RNAseq (de)" with 1,331 Genes. An "Edit" button is circled in red. A red arrow points to the second screenshot, which shows the "Details for step" panel. The panel lists search criteria: Experiment "Ex-erythrocytic stage transcriptomes (sporozoite liver time course and detached cells) unstranded", Reference Sample "sporozoite", Comparator Sample "Liver 48h", Direction "up-regulated", fold difference  $\geq 4$ , and adjusted P value less than or equal to 0.001. A red circle highlights the "Revise" button in the top navigation bar. A red arrow points to the third screenshot, which shows the "adjusted P value less than or equal to" input field set to 0.0001. A red circle highlights the "Revise" button in the bottom right corner.

**How many genes are returned by the search now?**

The screenshot shows the search step "Up-reg Liver 48h vs Sporozoites \*". The step is "Pber ex-erythro RNAseq (de)" with 1,151 Genes. The "Edit" button is circled in red. A red arrow points to the second screenshot, which shows the "Details for step" panel. The panel lists search criteria: Experiment "Ex-erythrocytic stage transcriptomes (sporozoite liver time course and detached cells) unstranded", Reference Sample "sporozoite", Comparator Sample "Liver 48h", Direction "up-regulated", fold difference  $\geq 4$ , and adjusted P value less than or equal to 0.001. A red circle highlights the "Revise" button in the top navigation bar. A red arrow points to the third screenshot, which shows the "adjusted P value less than or equal to" input field set to 0.0001. A red circle highlights the "Revise" button in the bottom right corner.

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?**

## A4. 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 )

The image shows a multi-step process in a web application for gene set analysis. The main window is titled "Up-reg Liver 48h vs Sporozoites \*". It contains a search strategy diagram with two steps: "Pber ex-erythro RNAseq (de)" (1,151 Genes) and "GO Term" (8,519 Genes). The results are shown as "59 Genes".

The "Add a step to your search strategy" dialog is open, showing options to combine genes (Intersect, Union, Minus) and choose which genes to combine (A new search, An existing strategy, My basket). The "A new search" option is selected, and a search box is visible with the text "GO".

The "Search for Genes by GO Term" dialog is also open, showing the search configuration. The results will be "intersected with" the results of Step 1. The search is configured for the organism "Haemoproteidae" and "Plasmodidae". The search terms are "GO:0008233 : peptidase activity" and "GO:0006508 : proteolysis". The "Run Step" button is highlighted.

How many genes are up-regulated in late liver stages compared to sporozoites AND have been annotated with proteolytic or peptidase activity?

## B. Determine what genes are up-regulated 4 fold in any liver stage compared to sporozoites

### B1. Choose the appropriate search

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

**Identify Genes based on RNA-Seq Evidence**

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

Filter Data Sets: ex- 1 result (filtered from a total of 58)

Organism: **Plasmodium berghei ANKA**

Data Set: **Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)**

Choose a Search: **FC** **P**

**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

Reset values to default

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

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

Up-ref Liver vs. Sporozoites \*

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

Step 1

Get Answer

**How many genes did you get?**

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

## B2. Explore search results

Gene Results | Genome View | **Analyze Results**

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

Rows per page: 20

Download Send to... Add Columns

| Gene ID        | Transcript ID    | Organism                | Product Description                                     | Fold Change | Chosen Ref (floor) | Chosen Comp (floor) |
|----------------|------------------|-------------------------|---|-------------|--------------------|---------------------|
| PBANKA_1203800 | PBANKA_1203800.1 | 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              |

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** ✕

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) (Toenhake 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.)
- ☐ P. vivax 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.)
- ☐ P. vivax - 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 | Genome View | **Analyze Results**

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

Rows per page: 20

Download Send to... Add Columns

| Organism                | Product Description                                     | Fold Change | Chosen Ref (floor) | Chosen Comp (floor) | Pber ex-erythro RNAseq - tpm Graph |
|-------------------------|---|-------------|--------------------|---------------------|------------------------------------|
| Plasmodium berghei ANKA | conserved Plasmodium membrane protein, unknown function | 838.2       | 0.66 (5.36)        | 4495.2              |                                    |
| Plasmodium berghei ANKA | DnaJ protein, putative                                  | 609.5       | 57.36              | 34965.19            |                                    |

### B3. 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

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

**Step 1**

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

**Add a step**

**Add a step to your search strategy**

**Combine with other Genes**

1 INTERSECT 2

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

**Transform into related records**

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

**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, W WGCNA, P Percentile, SA SenseAntisense

Filter Data Sets: **cat** 1 result (filtered from a total of 58)

**Organism:** Plasmodium berghel ANKA

**Data Set:** Ex-erythrocytic stage transcriptomes (sporozoite, liver time course and detached cells) (Caldelari et al.)

**Choose a Search:** DE, FC, **P**

**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 6h
- ☐ Liver 24h
- ☐ Liver 48h
- ☐ Liver 54h
- ☐ Liver 60h
- ☒ DC

**Minimum expression percentile:** 90

**Maximum expression percentile:** 100

**Matches Any or All Selected Samples?:** any

**Protein Coding Only:** protein coding

**Run Step**

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

**Step 1**

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

**Step 2**

Pber ex-erythro RNAseq (%ile)  
499 Genes

**Add a step**

**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)?**



Questions? Comments? Write to  
help@eveupathdb.org