Mining Transcriptomics Data on VEuPathDB: An Exercise
Genes up-regulated in liver stages of *Plasmodium* infection

VEuPathDB
Eukaryotic Pathogen, Vector & Host
Informatics Resources

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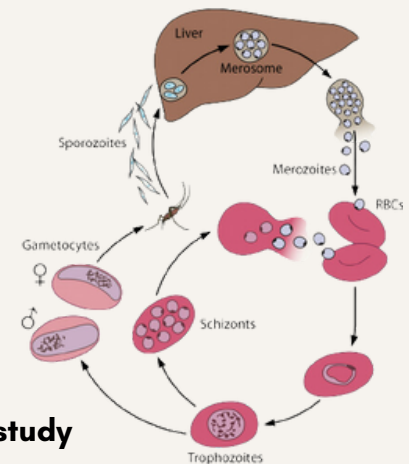
The purpose of this two-part exercise is to use data from this study to explore gene up-regulation in liver stages of *Plasmodium* infection.

Image credit

What platform is best suited for this question?

PlasmoDB
Plasmodium Informatics Resources

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.

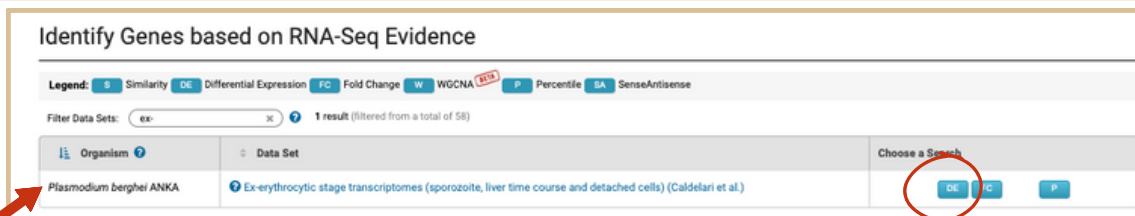Part 2: Use the **Fold Change** search to identify genes upregulated in any liver stage compared to sporozoites.

A1. Navigate to the appropriate search

- Navigate to PlasmoDB.org
- From the sidebar or header, search for "RNA" and click on RNA-Seq Evidence
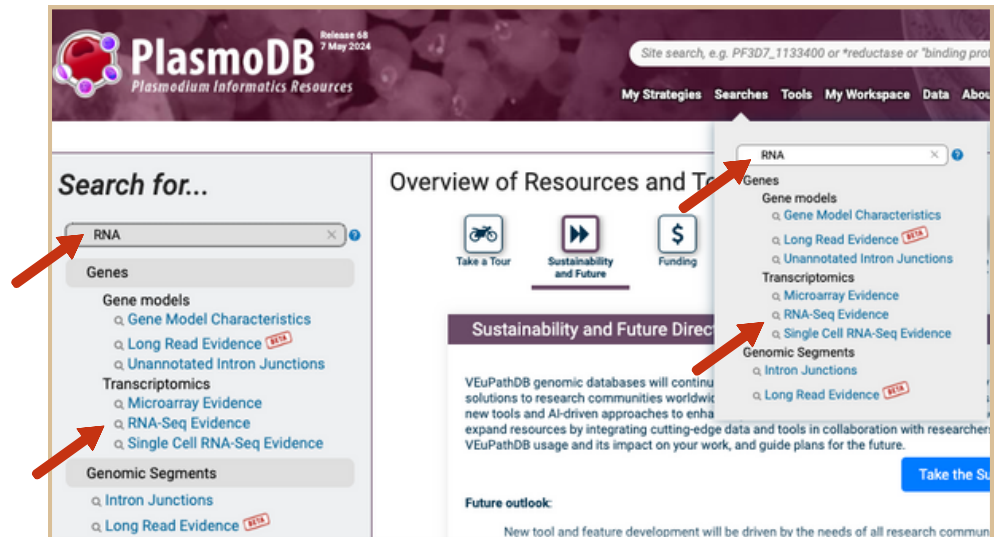
A2. Choose appropriate experiments/ samples & parameters within the search

- 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.
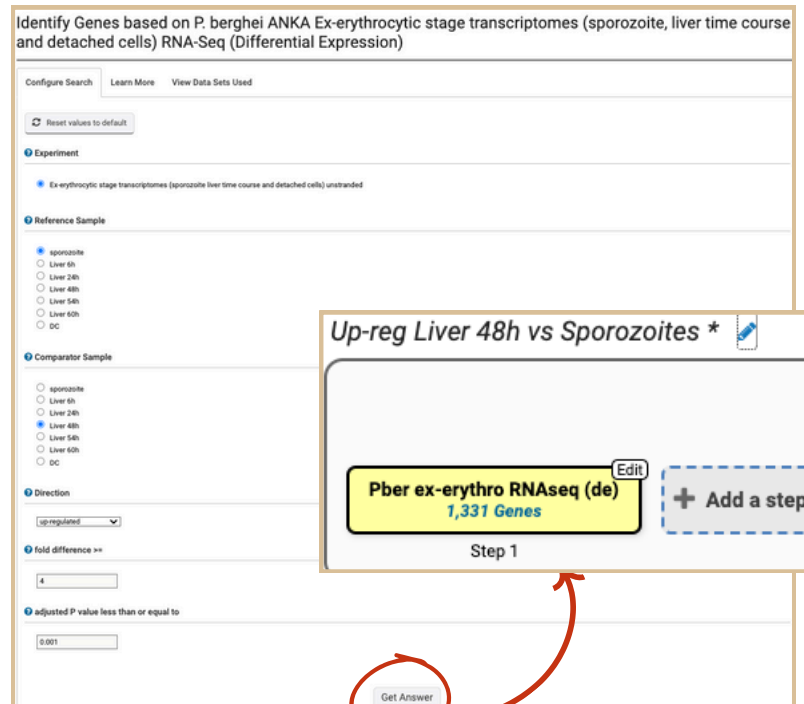
How many genes did you get?

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 <= 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?

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)

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

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.

How many genes did you get?

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

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?

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 *

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

+ Add a step

Step 1

Add a step to your search strategy

Combine with other Genes

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

Transform into related records

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

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

Gene models
Q Gene Model Characteristics
Q Long Read Evidence
Q Unannotated Intron Junctions
Transcriptomics
Q Microarray Evidence
Q RNA-Seq Evidence
Q Single Cell RNA-Seq Evidence

Add a step to your search strategy

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: cald ✕ ? 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.)	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
select all | clear all

? **Minimum expression percentile**

90

? **Maximum expression percentile**

100

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

any ⌄

? **Protein Coding Only:**

protein coding ⌄

Run 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)?

*Up-ref Liver vs. Sporozoites *

Pber ex-erythro RNAseq (%ile)
499 Genes

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

265 Genes

+ Add a step

Step 1   Step 2

Questions? Comments? Write to
help@veupathdb.org

7