

Exploring Expression data

1. Exploring Transcript Expression data in *Plasmodium vivax*.

The data set “Intraerythrocytic developmental cycle of three isolates (Bozdech et al.)” contains the transcriptional profile (microarray) throughout the 48-h intraerythrocytic cycle of three distinct *P. vivax* isolates. The data were integrated into PlasmoDB and two searches that query the data were created – one search queries the expression percentile rank of each gene and one search queries the data based on expression fold change.

Types of searches:

Fold Change: Search for genes which are differentially expressed between two sets of samples. No statistics are applied in this analysis

Fold Change with p-Value: Search for genes which are differentially expressed between 2 samples. Fisher's Exact Test p-values for each gene/sample comparison were precomputed.

Percentile: Search for genes based on the rank of expression w/in an experiment.

- Find all *P. vivax* genes that are expressed in the 80th or greater percentile in the late stages (40 and 43 hours) of the intraerythrocytic cycle of Patient 1.
 - How many genes did you find?

Identify Genes by:

- Expand All | Collapse All
- Text, IDs, Organism
- Genomic Position
- Gene Attributes
- Protein Attributes
- Protein Features
- Similarity/Pattern
- Transcript Expression
- EST Evidence
- SAGE Tag Evidence
- Microarray Evidence
- RNA Seq Evidence
- ChIP on Chip Evidence
- TF Binding Site Evidence
- Protein Expression
- Cellular Location

Identify Genes based on Microarray Evidence

Filter Data Sets: Type keyword(s) to filter

Legend: Add, Associate, DC, Direct Co., FC, Fold Chan..., FD, Fold Differ..., P, Percentile, S, Similarity, SA, Similarity...

Organism	Data Set	Choose a search
<i>P. berghei</i> ANKA	DO2 Mutant Transcript Profile (Mair et al.)	FC
<i>P. berghei</i> ANKA	Transcript Profiling of Developmental Stages - High Producer (HP/HPe) (Andrew Waters)	P
<i>P. falciparum</i> 3D7	eQTL for HB3, Dd2 and 34 progeny (Gonzales et al.)	ADD, S, SA
<i>P. falciparum</i> 3D7	Life cycle expression data (3D7) (Le Roch et al.)	FC, P
<i>P. falciparum</i> 3D7	Erythrocytic expression time series (3D7, DO2, HB3) (Bozdech et al. and Linas et al.)	FC, P, S
<i>P. falciparum</i> 3D7	Asexual blood stage transcriptomes of clonal strains (Rovira-Graells et al.)	FC, FD
<i>P. falciparum</i> 3D7	Gametocyte stages I-V transcriptomes (Young et al.)	FC, P
<i>P. falciparum</i> 3D7	Invasion pathway knockouts (Stubbs et al.)	FC, P
<i>P. falciparum</i> 3D7	Two Sir2 KO lines expression profiling (Tonkin et al.)	FC, P
<i>P. falciparum</i> 3D7	Three isogenic lines w/ CQ Treatment: expression profiles (Jiang et al.)	FC, P
<i>P. vivax</i> Sal-1	Intraerythrocytic developmental cycle of three isolates (Bozdech et al.)	FC, P
<i>P. yoelii</i> yoelii 17X	Liver, mosquito and blood stage expression profiles (Tarun et al.)	DC, P

Identify Genes based on P.v. Intraerythrocytic Infection Cycle (percentile)

Experiment: Patient 1

Samples: 9 hours, 13 hours, 17 hours, 20 hours, 23 hours, 29 hours, 35 hours, 40 hours, 43 hours, select all | clear all

Maximum expression percentile: 80

Maximum expression percentile: 100

Matches Any or All Selected Samples? any

Protein Coding Only: protein coding

Advanced Parameters

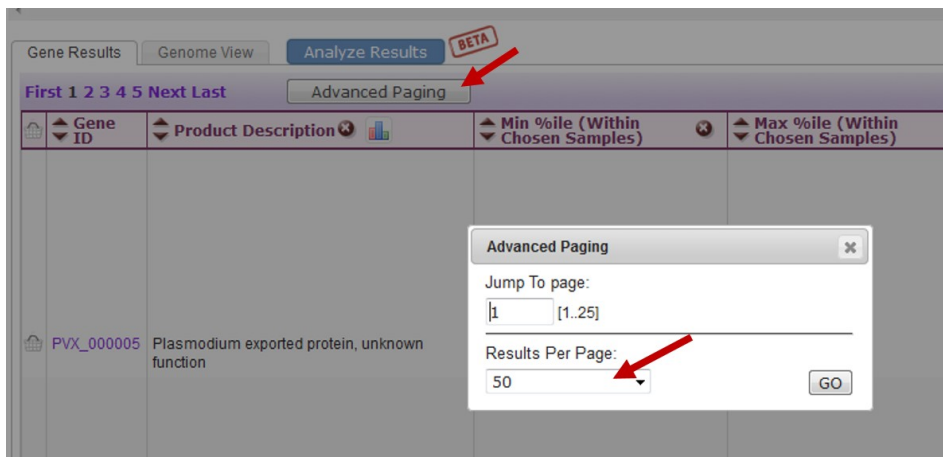
Get Answer

Patient1 80+ 1213 Genes Step 1

Add Step

Dataset information is linked below the Get Answer button

- Can you find genes in the result set that are expressed at low levels in the early time point but highly expressed in the late time points? Display 50 genes in your result list:



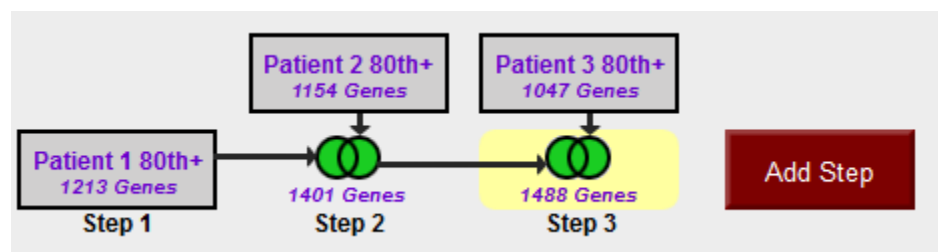
Scroll down your result list and look at the graph in the column labeled **“Pv-iRBC Patients %ile – Graph”**. Many genes returned by the search are expressed at the same level across all time points but some have low expression early and high expression later. For example:

PVX_000980 - *Plasmodium yoelii* blood stage membrane protein ag-1

PVX_001100 - RAD protein (Pv-fam-e)

PVX_001680 - Phist protein (Pf-fam-b)

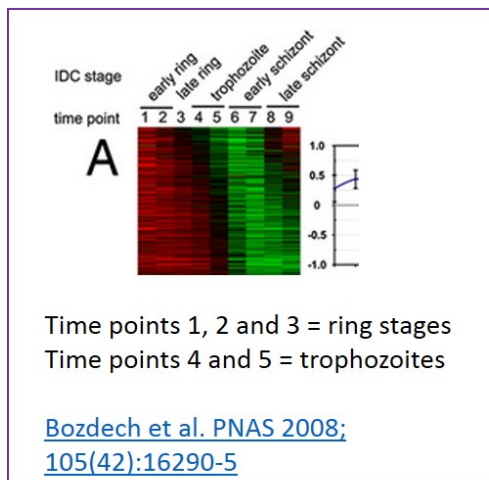
- (Optional) Find genes that are expressed in the 80th percentile or greater in the late stages of the intraerythrocytic cycle of **all patients**.



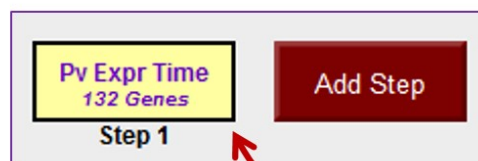
<http://plasmodb.org/plasmo/im.do?s=bec5bf555d632cc2>

b. Find *P. vivax* genes whose expression is upregulated 3-fold from ring to trophozoite stage in Patient 1.

- Use the fold change search associated with the data set used in exercise 1a. The search page is divided into left and right sections with parameters on the left and help information on the right. As you modify parameters on the left side note the dynamic help on the right side.
- Experiment = Patient 1
- Genes = Protein coding
- Regulation = up-regulated
- Fold change ≥ 3
- Reference Samples = 9, 13 and 17 hour time points that correspond to ring stages
- Comparison Samples = 20 and 23 hour time points that correspond to trophozoites.



Filter Data Sets: Type keyword(s) to filter	Legend:	AGS	Associatio...	DC	Direct Co...	FC	Fold Chan...	FD	Fold Differ...
		P	Percentile	S	Similarity	SA	Similarity ...		
Organism	Data Set								
<i>P. berghei</i> ANKA	DOZI Mutant Transcript Profile (Mair et al.)					DC		P	
<i>P. berghei</i> ANKA	Transcript Profiling of Developmental Stages - High Producer (HP/HPE) (Andrew Waters)							P	
<i>P. falciparum</i> 3D7	eQTL for HB3, Dd2 and 34 progeny (Gonzales et al.)			AGS				B	SA
<i>P. falciparum</i> 3D7	Life cycle expression data (3D7) (Le Roch et al.)					FC		P	
<i>P. falciparum</i> 3D7	Erythrocytic expression time series (3D7, DD2, HB3) (Bozdech et al. and Linas et al.)					FC		P	S
<i>P. falciparum</i> 3D7	Asexual blood stage transcriptomes of clonal strains (Rovira-Graells et al.)					FC	FD		
<i>P. falciparum</i> 3D7	Gametocyte stages I-V transcriptomes (Young et al.)					FC		P	
<i>P. falciparum</i> 3D7	Invasion pathway knockouts (Stubbs et al.)					FC		P	
<i>P. falciparum</i> 3D7	Two Sir2 KO lines expression profiling (Tonkin et al.)					FC		P	
<i>P. falciparum</i> 3D7	Three isogenic Lines w/ CQ Treatment: expression profiles (Jiang et al.)					FC		P	
<i>P. vivax</i> Sal-1	Intraerythrocytic developmental cycle of three isolates (Bozdech et al.)					FC		P	
<i>P. yoelii</i> yoelii 17X	Liver, mosquito and blood stage expression profiles (Tarun et al.)					DC		P	



Identify Genes based on P.v. Intraerythrocytic Infection Cycle (fold change)

Tutorial

For the Experiment: Patient 1
return: protein coding Genes
that are: up-regulated
with a Fold change ≥ 3

between each gene's minimum expression value
in the following Reference Samples

and its maximum expression value
in the following Comparison Samples

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

Up-regulated

Maximum Comparison
Minimum Reference

You are searching for genes that are up-regulated between at least two reference samples and at least two comparison samples.

For each gene, the search calculates:

$$\text{fold change} = \frac{\text{maximum expression value in comparison samples}}{\text{minimum expression value in reference samples}}$$

and returns genes when fold change ≥ 3 . This calculation creates the broadest window of expression values in which to look for genes that meet your fold change cutoff. To narrow the window, use the average or maximum reference value, or average or minimum comparison value.

See the detailed help for this search.

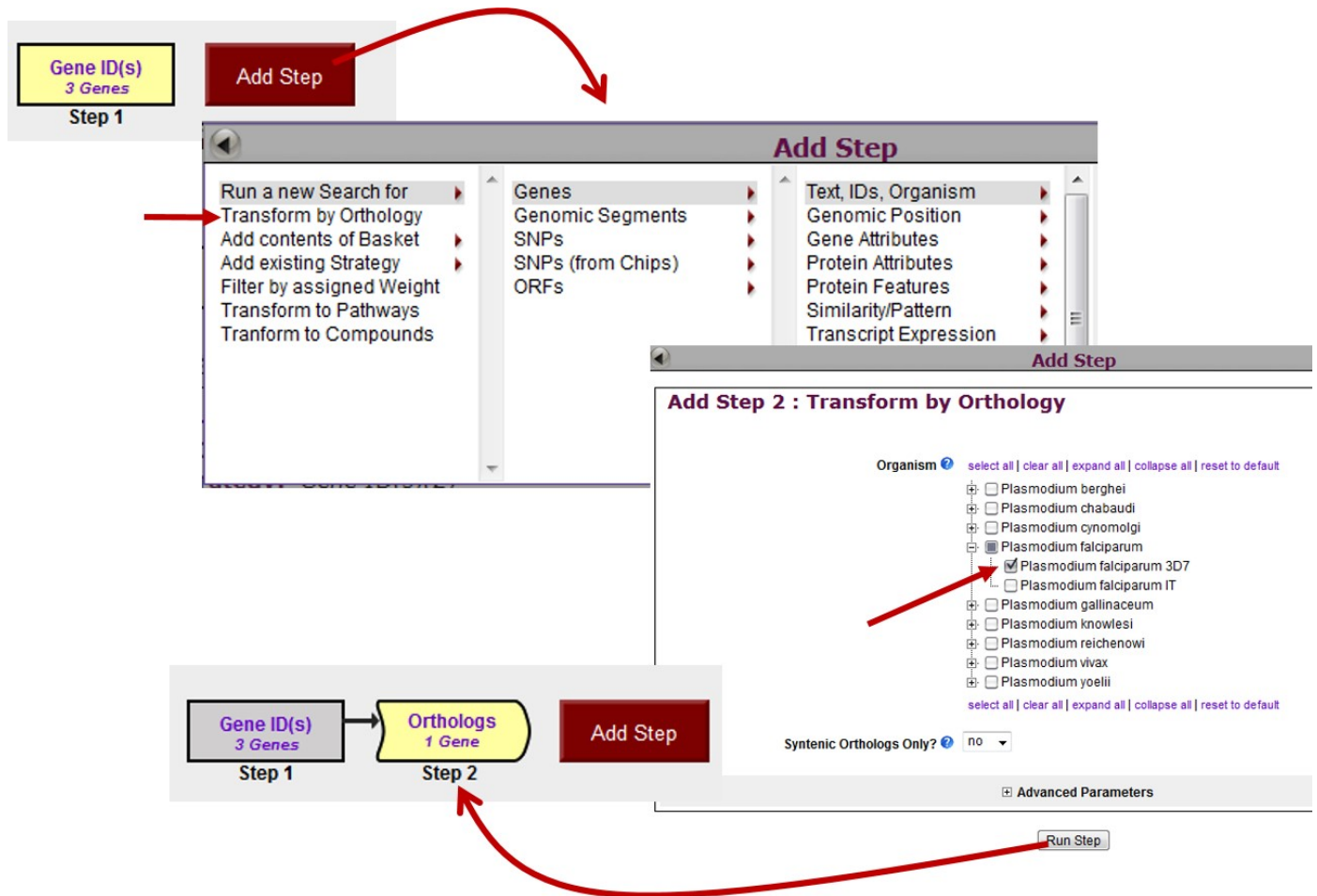
Get Answer

c. Do the *P. falciparum* 3D7 orthologs of these genes show the same expression profile?

- In exercise 1a we looked at the expression profiles of three genes, PVX_000980, PVX_001100, and PVX_001680. Run a Gene ID search to pull up the three *P. vivax* genes. Cut and paste or type the IDs into the Gene ID search.

The screenshot shows a web interface for identifying genes. On the left, a sidebar titled 'Identify Genes by:' contains a list of criteria: Text, IDs, Organism, Text (product name, notes, etc.), Gene ID(s), Organism, User Comments, Having Updated Annotation at GeneDB, Reagent Availability, Genomic Position, Gene Attributes, Protein Attributes, and Protein Features. A red arrow points from 'Gene ID(s)' in the sidebar to the main search area. The main area is titled 'Identify Genes based on Gene ID(s)'. It has a 'Gene ID input set' section with a radio button selected for 'Enter a list of IDs or text:'. Below this is a text input field containing the gene IDs: PVX_000980, PVX_001100, and PVX_001680. There are also options for 'Upload a text file:', 'Copy from My Basket:', and 'Copy from My Strategy:'. At the bottom of the main area is a 'Get Answer' button. A red arrow points from this button to a 'Step 1' box in a lower section. The 'Step 1' box contains the text 'Gene ID(s)' and '3 Genes'. To the right of the 'Step 1' box is a red 'Add Step' button.

- Use the Transform by Orthology function to transform the *P. vivax* genes into their *P. falciparum* orthologs.
- Do all of your *P. vivax* genes have orthologs in *P. falciparum* 3D7?
- Go to the gene page for the *P. falciparum* gene returned by the ortholog transform. Does this gene have the same expression profile as its *P. vivax* ortholog – low in early but high in late intraerythrocytic stages? What lines of evidence support your conclusion? Go to the expression section on the *P. falciparum* gene page and look at these data tracks:
 - Erythrocytic expression time series (3D7, DD2, HB3)
 - Overlay of Intraerythrocytic Expression Profiles
 - Life cycle expression data (3D7)



d. Find *P. vivax* genes that are likely actively translated in the trophozoite stage.

Few functional data sets addressing transcript expression are available for *P. vivax*. We can take advantage of data collected from *P. falciparum* to find genes with certain expression characteristics and then transform those results to their *P. vivax* orthologs. This strategy infers expression information on *P. vivax* genes for which there is little functional data. The “**Polysomal and steady-state asexual stage transcriptomes (Bunnik et al.)**” data set profiles *P. falciparum* polysomally associated RNA as a measure of actively translated transcripts. Let’s run that search and transform the results to *P. vivax* orthologs.

- Navigate to the search page “Identify Genes based on RNA-Seq Evidence”.
- How many data sets describe expression data from *P. vivax*? How many are from *P. falciparum*?
- Focus on the Percentile search for the “**Polysomal and steady... Bunnik**” data set and arrange the parameters to return genes that are expressed in the 80th or greater percentile for the polysomal trophozoite stage.
- How many genes did you get?
- Notice that the genes returned are *P. falciparum* genes.

Identify Genes based on RNA Seq Evidence

Filter Data Sets: Type keyword(s) to filter Legend: FC Fold Cha... FCpV Fold Cha... P Percentile

Organism	Data Set	FC	FCpV	P
<i>P. berghei</i> ANKA	5 asexual and sexual stage transcriptomes (Hoeijmakers et al.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>P. falciparum</i> 3D7	Transcriptome during intraerythrocytic development (Bartfai et al.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>P. falciparum</i> 3D7	Transcriptomes of 7 sexual and asexual life stages (Lopez-Barragan et al.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>P. falciparum</i> 3D7	Strand specific transcriptomes of 4 life cycle stages (Lopez-Barragan et al.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>P. falciparum</i> 3D7	NSR-seq Transcript Profiling of malaria-infected pregnant women and children (Vignali et al.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>P. falciparum</i> 3D7	Blood stage transcriptome (3D7) (Otto et al.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>P. falciparum</i> 3D7	Polysomal and steady-state asexual stage transcriptomes (Bunnik et al.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>P. falciparum</i> 3D7	Intraerythrocytic cycle transcriptome (3D7) (Hoeijmakers et al.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<i>P. yoelii</i> yoelii 17				

Identify Genes by:

- Text, IDs, Organism
- Genomic Position
- Gene Attributes
- Protein Attributes
- Protein Features
- Similarity/Pattern
- Transcript Expression
- EST Evidence
- SAGE Tag Evidence
- Microarray Evidence
- RNA Seq Evidence
- ChIP on Chip Evidence
- TF Binding Site Evidence
- Protein Expression

Identify Genes based on *P. falciparum* 3D7 Polysomal and steady-state asexual stage transcriptomes RNASeq (percentile)

Experiment: Polysomal and steady-state asexual stage transcriptomes Polysomal and steady-state asexual cel

Samples:

- ☐ Steady_state ring 0h
- ☐ Polysomal ring 0h
- ☐ Steady_state troph 18h
- ☒ Polysomal troph 18h
- ☐ Steady_state schiz 36h
- ☐ Polysomal schiz 36h

select all | clear all

Minimum expression percentile: 80

Maximum expression percentile: 100

Matches Any or All Selected Samples?: any

Protein Coding Only: protein coding

Advanced Parameters

Get Answer

PF3D7 Polysomal 1136 Genes Step 1

Add Step

- Use the transform by orthology function to convert the *P. falciparum* genes to their *P. vivax* orthologs.

PF3D7 Polysomal 1136 Genes Step 1

Add Step

Run a new Search for

- Transform by Orthology
- Add contents to Basket
- Add existing Strategy
- Filter by assigned Weight
- Transform to Pathways
- Transform to Compounds

Add Step 2 : Transform by Orthology

Organism: select all | clear all | expand all | collapse all | reset to default

- ☐ Plasmodium berghei
- ☐ Plasmodium chabaudi
- ☐ Plasmodium cynomolgi
- ☐ Plasmodium falciparum
- ☐ Plasmodium gallinaceum
- ☐ Plasmodium knowlesi
- ☐ Plasmodium reichenowi
- ☒ Plasmodium vivax
- ☒ Plasmodium vivax Sal-1
- ☐ Plasmodium yoelii

select all | clear all | expand all | collapse all | reset to default

Syntenic Orthologs Only? no

Advanced Parameters

Run Step

PF3D7 Polysomal 1136 Genes Step 1

Orthologs 1095 Genes Step 2

Add Step