

Complex strategies with Genomic Colocation Exercise 14

14.1 Divergent genes with similar expression profiles.

Note: for this exercise use <http://plasmodb.org>.

Identify genes that meet these four criteria:

1. are located within 1000 bp of each other
2. are divergently transcribed,
3. are expressed maximally at day 30 of the iRBC cycle +/- 8 hrs and,
4. show at least a 3-fold increase in expression.

- Hint: first use the “Genes bases on Microarray Evidence” -> “Erythrocytic expression time series (3D7,DD2, & HB3) (Bozdech et al. and Linas et al.)” -> “**Fc**” search.

Fold Change
Percentile
Similarity

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

For the Experiment IRBC HB3 (48 Hour scaled) ?

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 ?

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

- 1-16 Hours
- 17-30 Hours
- 17-23 Hours
- 24-30 Hours
- 31-48 Hours

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

and its maximum ? expression value ?

in the following Comparison Samples ?

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

- 1-16 Hours
- 17-30 Hours
- 17-23 Hours
- 24-30 Hours
- 31-48 Hours
- 31-39 Hours

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

Advanced Parameters

Weight ?

Global min / max in selected time points ? Maximum ?

Example showing one gene that would meet search criteria

(Dots represent this gene's expression values for selected samples)

Up-regulated

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

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](#).

- Add a step that is the same as the first step and select the genomic colocation (1 relative to 2) operation.

- Set up the form to identify those genes that are transcribed on the opposite strand that have their starts located within 1000 bp of another genes start.
- If you are having difficulty setting this up, you can see the strategy at: <http://plasmodb.org/plasmo/im.do?s=0ebfe58b1c9b42cc>. Cut and paste the link into your browser if the hyperlink does not work.
- Turn on the “Pf-iRBC 48hr - Graph” column to assess how well the pairs of genes compare in terms of expression. The pairs of genes are located one above the other in the result table if sorted by location.
- Note that you could do similar types of experiments to look at potential co-regulation / shared enhancers / divergent promoters with other sorts of data such as:
 - Genes by ChiP-chip peaks in ToxoDB.
 - DNA motifs for transcription factor binding sites.
 - Of course other expression queries.
 - Etc ...
- The screenshot below shows one way (there are MANY) to configure the genome colocation form to identify genes that are divergently transcribed located with their start within 1000 bp of each other.

Combine Step 1 and Step 2 using relative locations in the genome

You had **684 Genes** in your Strategy (Step 1). Your new **Genes** search (Step 2) returned **684 Genes**.

"Return each whose overlaps the of a Gene in Step 2 and is on

(684 Genes in Step 1)

Exact

Upstream: bp

Downstream: bp

Custom:

begin at: bp

end at: bp

(684 Genes in Step 2)

Exact

Upstream: bp

Downstream: bp

Custom:

begin at: bp

end at: bp

[Close](#)

14.2 Finding possible oocyst expressed genes based on DNA motifs.

Note: for this exercise use <http://toxodb.org>

In exercise 13.4 you defined a number of *T. gondii* genes that are preferentially expressed in the oocyst stages. How can you use this information to expand the number of possible oocyst regulated genes? One possibility is to try and define

14.3. Identifying conserved DNA elements upstream of genes

The goal of this exercise is to identify a DNA element in the upstream region of similarly regulated genes.

a. Identify genes that are up-regulated in malaria sporozoites compared to blood stage parasites. Examine the list of searchable experiments on the PlasmoDB microarray search page: Identify Genes based on Microarray Evidence. Can you identify an experiment that would give you this answer? (hint: look at *Plasmodium* species other than *P. falciparum*, ie. *P. yoelii* [Liver, mosquito and blood stage expression profiles (Tarun et al.) (direct comparison)])

Organism
Data Set
Choose a search

P. yoelii yoelii 17X
Liver, mosquito and blood stage expression profiles (Tarun et al.)
Do
P

[Show All Data Sets](#)

Direct Comparison
Percentile

Identify Genes based on P.y. Liver Stages (fold change)

Direction

Samples

Fold difference >=

Protein Coding Only:

Advanced Parameters

b. How many genes did you find? What you are interested in is looking at the nucleotide sequence upstream of the start sites of these genes. How can you do this in bulk? PlasmoDB has a sequence retrieval tool that allows you to download results of your searches in bulk. This includes a tool that allows you to specify the sequence you want.

(Genes)
Strategy: Py Expression(3) *

Py Expression
57 Genes
Step 1

Add Step

[Rename](#)
[Duplicate](#)
[Save As](#)
[Share](#)
[Delete](#)

57 Genes from Step 1
Strategy: Py Expression(3)

[Add 57 Genes to Basket](#) | [Download 57 Genes](#)

Click on a number in this table to limit/filter your results

All Results	Ortholog Groups	Plasmodium											
		<i>P.berghei</i>	<i>P.chabaudi</i>	<i>P.cynomolgi</i>	<i>P.falciparum</i>	(nr Genes: 0)	<i>P.gallinaceum</i>	<i>P.knowlesi</i>	<i>P.reichenowi</i>	<i>P.vivax</i>	<i>P.yoelii</i> (nr Genes: 57)		
		ANKA	chabaudi	strain B	3D7	IT	8A	strain H	Dennis	Sal-1	yoelii 17XNL	yoelii 17X	yoelii YM
57	57	0	0	0	0	0	0	0	0	0	0	57	0

Gene Results
Genome View

First 1 2 3 Next Last
Advanced Paging
Add Columns

Gene ID	Product Description	Fold Change	Py-Liver Stages - Graph
PY17X_0523600	conserved Plasmodium protein, unknown function	34.55	

- c. After you click on “Download ### Genes”, you are offered a drop down menu of options. Explore these; which one will allow you to specify the sequence to download. (hint: Configurable FASTA)

- d. Define the sequence you want to retrieve. For this exercise retrieve 500 nucleotides up-stream of the start of translation.

- e. The next step is to take this sequence and run it through a DNA motif finder such as MEME (<http://meme.sdsc.edu/meme/intro.html>). To speed up this process we have pre-run the motif finder and results are presented here:

g. How many times did this motif occur in the genome? How many of them are in the upstream region of genes? Can you find all *P. yoelii* genes that are within 1000 nucleotides downstream of the motif? (hint: use the genomic colocation option when combining searches).

Genomic Colocation ? ↻

Combine Step 1 and Step 2 using relative locations in the genome
 You had **1257 Genomic Segments** in your Strategy (Step 1). Your new **Genes** search (Step 2) returned **7774 Genes**.

"Return each whose **upstream region** overlaps the **exact region** of a Genomic Segment in Step 1 and is on

(7774 Genes in Step)

Exact

Upstream: bp

Downstream: bp

Custom:

begin at: bp

end at: bp

(1257 Genomic Segments in Step)

Exact

Upstream: bp

Downstream: bp

Custom:

begin at: bp

end at: bp

h. Do these genes have orthologs in other *Plasmodium* species? (hint: add a step to your search strategy and transform the results to their orthologs).

Add Step ✕

Run a new Search for

Transform by Orthology

Add contents of Basket

Add existing Strategy

Filter by assigned Weight

Add Step 4 : Transform by Orthology

Organism

- Plasmodium berghei
- Plasmodium chabaudi
- Plasmodium falciparum
- Plasmodium knowlesi
- Plasmodium vivax
- Plasmodium yoelii

Systemic Orthologs Only?

Population Biology

[Close](#)

Optional: add a step and do the motif search on these orthologs to find out how many of them also contain the motif.