

Complex strategies with Genomic Co-location (Exercise 12)

12.1 Divergent genes with similar expression profiles.

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

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

- Hint: use the “Genes bases on Microarray Evidence” -> “*Intraerythrocytic Infection Cycle (DeRisi)*” -> “[P.f. Intraerythrocytic Infection Cycle \(exprn timing\)](#)” search.

Revise Step 1 : P.f. Intraerythrocytic Infection Cycle (exprn timing)

Query based on

Timing

Expression time +/-

Max/Min Expression Ratio >=

Maximum Expression percentile >=

Microarray time course(s) to query

Give this search a weight

Give this search a name

- Add a step that is the same as the first step. Note that you could copy the first step and then add an existing strategy to avoid setting the parameters again.
- Select the genomic collocation (Relative to ... using relative genomic locations) 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=6bb7de2c16ec5379>
- Turn on the “Pf-iRBC expr profile graph (GS array)” column to assess how well the pairs of genes (should be located one above the other if sorted by location) compare in terms of expression.
- 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 next page shows one way (there are MANY) to configure the genome collocation 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)

Region

Gene

Exact

Upstream: bp

Downstream: bp

Custom:

begin at: bp

end at: bp

(684 Genes in Step)

Region

Gene

Exact

Upstream: bp

Downstream: bp

Custom:

begin at: bp

end at: bp

12.2 Finding possible oocyst expressed genes based on DNA motifs.

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

In exercise 11.3 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 common elements in promoter or 5'UTR regions (ie. 5' to the start of the genes). For this you will have to be able to retrieve 5' sequence from all of the genes in the oocyst list. How would you do this? (hint: click on download genes then select FASTA format from the drop down menu). The amount of upstream sequence you retrieve is up to you.

After you have your sequences you will need to run them through a DNA pattern finder like MEME (http://meme.nbcr.net/meme4_6_1/intro.html). Results from a submission to MEME could take up to several hours so for your convenience 300 nucleotides upstream of all the oocyst results were analyzed using MEME – results can be visualized here:

<http://ws.nbcr.net/app1307935081632/meme.html#overview>

Can you take one of the generated motifs and find additional genes in *T. gondii* that contain this motif in their upstream regions? What do your results look like? Did you get too many or too few results? How would you modify the motif to change your results?

