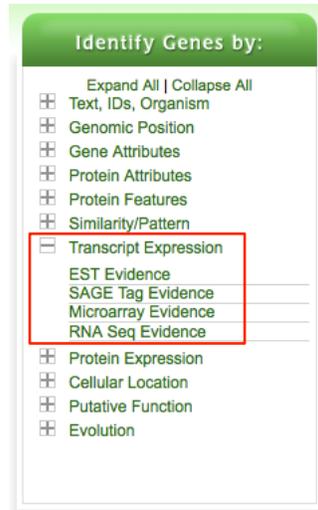


Exploring Proteomics Data (Exercise 5)

1. Find *Giardia* genes that are differentially regulated (up or down) at 6 and 18 hours post interaction with host cells. (<http://giardiadb.org>)
 - What types of data are available in GiardiaDB that provide information about transcript expression? Explore the various categories of transcription data.



- For host interaction, GiardiaDB currently has a microarray dataset from Staffan Svård's group called: "Transcript Profiling of Host-parasite Interactions". You can find this search by selecting "Microarray Evidence" under the "Transcript Expression" section.

Identify Genes based on Microarray Evidence		
Filter Data Sets: <input type="text" value="Type keyword(s) to filter"/>	Legend: Dc Direct Co...	Fc Fold Chan... P Percentile
^ Organism	◇ Data Set	Choose a search
G. Assemblage A isolate WB	● Stress Response in Trophozoites - Temperature Concentration (Adrian Hehl)	Dc P
G. Assemblage A isolate WB	● Stress Response in Trophozoites - Incubation Time (Adrian Hehl)	Dc P
G. Assemblage A isolate WB	● Encystation in trophozoites (Morf et al.)	Dc Fc P
G. Assemblage A isolate WB	● Transcript Profiling of Host-parasite Interactions (Staffan Svård)	Fc P

- Note that if you want to find genes that are changing in expression select the fold change (Fc) query.
- The configurable search page allows you to define the type of expression profile you are interested in. In this case, we are interested in genes that are differentially regulated between the Caco 6 and 18 samples (comparison samples) compared to the Caco 1.5, TYDK 1.5, TYDK 6 and TYDK 18 samples (reference).

Identify Genes based on G.I. Host-Parasite Interaction (fold change) Tutorial

For the Experiment **Host Parasite Interaction**
 return **protein coding** Genes
 that are **up or down regulated**
 with a **Fold change >= 2**

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

- DMEM 1.5
- TYDK 1.5
- TYDK 6
- TYDK 18
- Caco 1.5

and its **average** expression value
 in the following **Comparison Samples**

- TYDK 6
- TYDK 18
- Caco 1.5
- Caco 6
- Caco 18

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

Up or down regulated

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

For each gene, the search calculates:

$$\text{fold change}_{\text{up}} = \frac{\text{average expression value in comparison samples}}{\text{average expression value in reference samples}}$$

$$\text{fold change}_{\text{down}} = \frac{\text{average expression value in reference samples}}{\text{average expression value in comparison samples}}$$

and returns genes when $\text{fold change}_{\text{up}} \geq 2$ or $\text{fold change}_{\text{down}} \geq 2$.

See the detailed help for this search.

- How many genes did you identify?
- What do the expression profile graphs look like? Do they agree with your search parameters?
- How many of these genes are expressed in the lower 30% percentile from the trophozoite RNAseq experiment available in GiardiaDB? (hint: add a step and go to the RNAseq section - configure the percentile query for “(WB) Trophozoite Transcriptome sense strand” to return genes expressed between 0 and 30 percentile).

Add Step

Add Step 2 : G Assemblage A isolate WB (WB) Trophozoite Transcriptome RNASeq (percentile)

Experiment **(WB) Trophozoite Transcriptome sense strand**

Samples WB

Minimum expression percentile

Maximum expression percentile

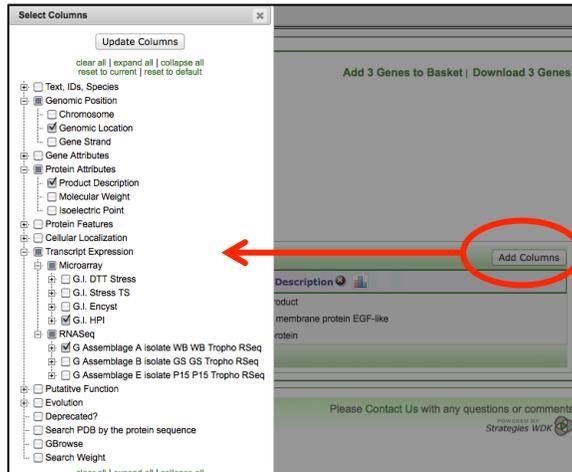
Matches Any or All Selected Samples? **any**

Protein Coding Only: **protein coding**

Combine Genes in Step 1 with Genes in Step 2:

- 1 Intersect 2
- 1 Union 2
- 1 Relative to 2, using genomic colocation
- 1 Minus 2
- 2 Minus 1

Look at your results and add the columns for expression graphs from the experiments you searched. Do the graphs coincide with what you searched?



- Notice that one of the genes has a high level of antisense transcription. Could this be indicating something interesting about how this gene is regulated?

2. Find *Cryptosporidium* genes that are upregulated during exystation based on RT PCR evidence. (<http://cryptodb.org>)

- Configure the fold change search of the *Cryptosporidium parvum* RT-PCR data to identify genes that are upregulated by at least 1.5 fold between the minimal expression level at time points 24-72 hrs compared to the maximal expression level at time points 2-12hrs.
- How many genes did you get? What happens if you revise this search to

Identify Genes based on C.p. post-infection semi-quantitative Real Time PCR (fold change)

Tutorial

For the Experiment: C.parvum expression profiles based on RT-PCR

return protein coding Genes

that are up-regulated

with a Fold change \geq 1.5

between each gene's maximum expression value

In the following Reference Samples

2Hrs
 6Hrs
 12Hrs
 24Hrs
 36Hrs
[select all](#) | [clear all](#)

and its minimum expression value

In the following Comparison Samples

12Hrs
 24Hrs
 36Hrs
 48Hrs
 72Hrs
[select all](#) | [clear all](#)

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

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:

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

and returns genes when fold change \geq 1.5. This calculation creates the narrowest window of expression values in which to look for genes that meet your fold change cutoff. To broaden the window, use the average or minimum reference value, or average or maximum comparison value.

[See the detailed help for this search.](#)

compare the average expression values between the selected time points.

- Which gene has the highest fold difference in expression? You can sort the columns by clicking on the up or down arrow in the fold change column.

Gene ID	Organism	Product Description	Fold Change	Chosen Ref (log2)	Chosen Comp (log2)	Cp RT PCR - graph
cpg3_3090	<i>C. parvum Iowa II</i>	conserved hypothetical protein	1.8	0.06	0.87	
cpg1_800	<i>C. parvum Iowa II</i>	protein disulfide isomerase, signal peptide plus possible ER retention motif	1.7	0.05	0.8	
cpg2_4190	<i>C. parvum Iowa II</i>	DEXDc+HELICc, mu308/POLc like SFII DNA helicase, no polymerase domain	1.7	0.07	0.86	

- Can you find out anything about the function of this gene? (hint: you can add additional columns from the “putative function” section, of you can visit the gene page and explore it further).