

Exploring Transcriptomic data

1. Exploring RNA sequence data in *Plasmodium falciparum*.

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

- a. Find all genes in *P. falciparum* that are up-regulated during the later stages of the intraerythrocytic cycle.
 - Use the fold change search for the data set “**Transcriptome during intraerythrocytic development (Bartfai et al.)**”. For this data set, synchronized Pf3D7 parasites were assayed by RNA-seq at 8 time-points during the iRBC cycle. We want to find genes that are up-regulated in the later time points (30, 35, 40 hours) using the early time points (5, 10, 15, 20, 25 hours) as reference.

The image shows two screenshots of the Plasmodb.org website. The left screenshot shows the 'Search for...' sidebar with 'Genes' expanded and 'RNA-Seq Evidence' selected. The right screenshot shows the 'Identify Genes based on RNA-Seq Evidence' main panel. In this panel, the 'Filter Data Set' dropdown is set to 'development', and the 'Data Set' dropdown is set to 'Transcriptome during intraerythrocytic development (Bartfai et al.)'. The 'Choose a Search' section shows 'Fold Change' selected. Below this, there are input fields for 'Up or down regulated' (set to 'Up') and 'Reference Sample' (set to '5, 10, 15, 20, 25'). A red arrow points from the 'RNA-Seq Evidence' option in the sidebar to the 'Identify Genes based on RNA-Seq Evidence' panel. Another red arrow points from the 'Fold Change' search option to the 'Up or down regulated' input field.

- There are a number of parameters to manipulate in this search. As you modify parameters on the left side note the dynamic help on the right side. See screenshots.
- **Direction:** the direction of change in expression. **Choose up-regulated.**
- **Fold Change** \geq the intensity of difference in expression needed before a gene is returned by the search. **Choose 12** but feel free to modify this.
- **Reference Sample:** the samples that will serve as the reference when comparing expression between samples. **choose 5, 10, 15, 20, 25**
- **Between each gene's AVERAGE expression value:** This parameter appears once you have chosen two Reference Samples and defines the operation applied to reference samples.

Fold change is calculated as the ratio of two values (upregulated ratio = expression in comparison)/(expression in reference). When you choose multiple samples to serve as reference, we generate one number for the fold change calculation by using the minimum, maximum, or average. **Choose average**

- **(or a Floor of 10 reads):** This parameter defines a lower limit of aligned reads for a gene to avoid unreliable fold change calculations. (Low numbers of aligned reads means low expression but the low values may be may be technically inaccurate. Dividing by small numbers creates large numbers. $2000\text{FPKM}/10 = 200$; $2000/0.1 = 20,000$) If a gene has fewer than 10 aligned reads, it is assigned 10 reads before the fold change calculation is made. **Leave this as default at 10 reads.**
- **Comparison Sample:** the sample that you are comparing to the reference. In this case you are interested in genes that are up-regulated in later time points **choose 30, 35, 40**
- **And its AVERAGE expression value:** This parameter appears once you have chosen two Comparison Samples and defines the operation applied to comparison samples. See explanation above. **Choose average**

Identify Genes based on P. falciparum 3D7 Transcriptome during intraerythrocytic development RNASeq (fold change) [Tutorial](#)

For the **Experiment** Transcriptome during intraerythrocytic development scaled unstranded

return **protein coding** **Genes**

that are **up-regulated**

with a **Fold change** ≥ 12

between each gene's **average** expression value

(or a **Floor** of 10 reads (1 FPKM))

in the following **Reference Samples**

☒ Hour 5
☒ Hour 10
☒ Hour 15
☒ Hour 20
☒ Hour 25
[select all](#) | [clear all](#)

and its **maximum** expression value

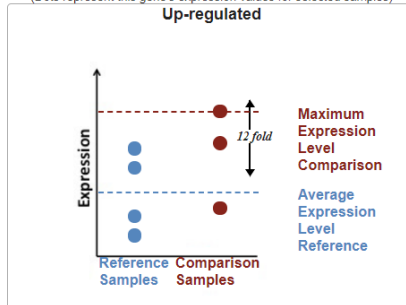
(or the **Floor** selected above)

in the following **Comparison Samples**

☐ Hour 20
☐ Hour 25
☒ Hour 30
☒ Hour 35
☒ Hour 40
[select all](#) | [clear all](#)

Example showing one gene that would meet search criteria

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



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:

$$\text{fold change} = \frac{\text{maximum expression level in comparison}}{\text{average expression level in reference}^*}$$

and returns genes when **fold change** ≥ 12 .

To narrow the window, use the maximum reference value, or average or minimum comparison value. To broaden the window, use the minimum reference value.

See the detailed help for this search.

* or FPKM Floor, whichever is greater

Get Answer

3D7 iRBC RNA-Seq (fc)
969 Genes

+ Add a step

Step 1



- b. For the genes returned by the search, how does the RNA-sequence data compare to microarray data?
- Hint: PlasmoDB contains data from a similar experiment that was analyzed by microarray instead of RNA sequencing. This experiment is called: **Erythrocytic expression time series (3D7, DD2, HB3) (Bozdech et al. and Linas et al.)**. IDC 48 hr Marray – Expr Graph shows normalized expression values. To directly compare the data for genes returned by the RNA-seq search that you just ran, add the column called “Pf-iRBC 48hr - Graph”.

The screenshot displays the PlasmoDB web interface. At the top, there's a navigation bar with 'My Strategies' and buttons for 'New', 'Opened (1)', 'All', 'Basket', 'Public Strategies (27)', and 'Help'. Below this, a 'Hide search strategy panel' button is visible. The main content area shows a search strategy named '3D7 iRBC RNASeq (fc)' with 976 genes. A table of results is shown, with columns for various Plasmodium species and their respective gene counts. A 'Select Columns' dialog box is open, allowing users to choose columns for the results. The dialog box has a search bar and a list of columns to select. The 'Add Columns' button is highlighted with a red circle. To the right of the dialog box, there are several graphs. The top graph is titled '3D7 iRBC RNASeq - fpkm Graph' and shows FPKM values for a specific gene. Below it, there are two more graphs: '3D7 iRBC RNASeq - fpkm Graph' and 'IDC 48 hr Marray - Expr Graph'. The bottom graph shows normalized expression values for a specific gene. The graphs are used to compare RNA-seq data with microarray data.





OPTIONAL: You can also run a fold change search using this experiment to compare results on a genome scale. Add a step to your strategy and intersect your current results (genes upregulated 12 fold in later IDC time periods) with a fold change search using the “Erythrocytic expression time series (3D7, Dd2, HB3) (Bozdech et al. and Linas et al.)” experiment (under microarray evidence). Configure it similarly to the RNA-seq experiment although you will probably need to make the fold change smaller (try 2 or 3) due to the decreased dynamic range of microarrays compared to RNA-seq.







← Add a step to your search strategy ⓘ

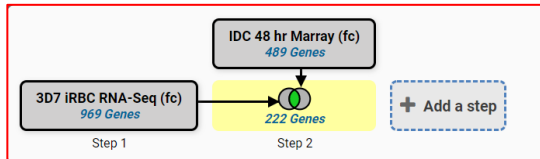
Search for Genes by Microarray Evidence

The results will be  intersected with  the results of Step 1.


Filter Data Sets: ⓘ



Legend:  Similarity  Direct Comparison  Fold Change  Percentile



Organization	Data Set	Choose a Search
Plasmodium falciparum 3D7	Erythrocytic expression time series (3D7, DD2, HB3) (Bozdech et al. and Linas et al.)	 
Plasmodium knowlesi strain H	Intraerythrocytic cycle expression profile: in vitro and ex vivo (Pkn0 PK1(A*)) (Lapp et al.)	 
Plasmodium vivax P01	Intraerythrocytic developmental cycle of three isolates (Bozdech et al.)	 




Similarity Fold Change Percentile


For the Experiment  IRBC HB3 (48 Hour scaled) ⓘ

return  protein coding  Genes

that are  up-regulated 


with a Fold change \geq ⓘ


between each gene's average  expression value

in the following  Reference Samples ⓘ

Note: You must select at least 1 values for this parameter.
28 selected, out of 46

☒ 1-16 Hours
☒ 17-30 Hours
☒ 31-48 Hours
select all | clear all | expand all | collapse all

and its average  expression value

in the following  Comparison Samples ⓘ

Note: You must select at least 1 values for this parameter.
18 selected, out of 46

☒ 1-16 Hours
☒ 31-48 Hours
☒ 31-39 Hours
☒ 40-48 Hours
select all | clear all | expand all | collapse all

Run Step

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.

For each gene, the search calculates:

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

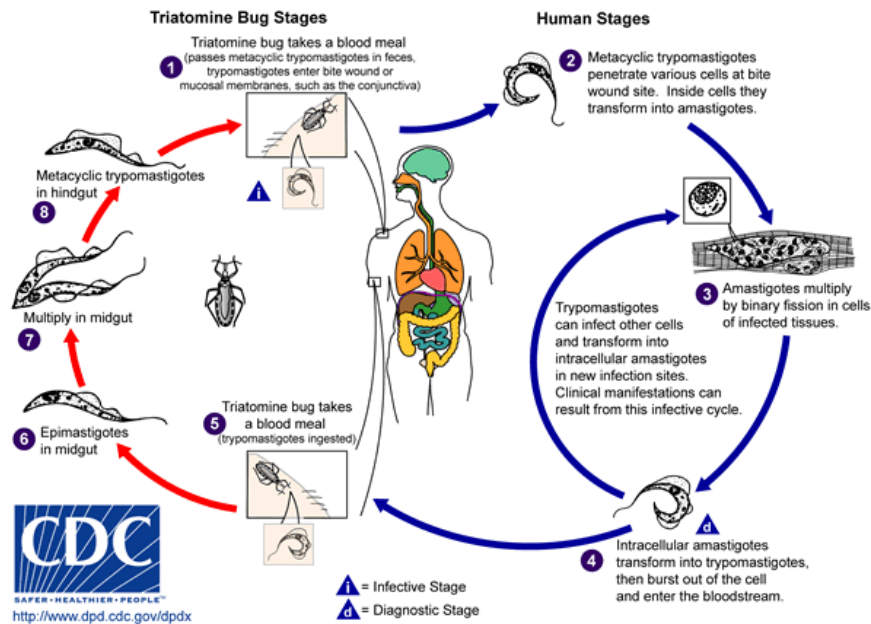
and returns genes when fold change \geq 2.

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

To narrow the window, use the maximum reference value, or minimum comparison value.
To broaden the window, use the maximum reference value, or maximum comparison value.

2. Exploring microarray data in TriTrypDB.

Note: For this exercise use <http://www.tritrypdb.org>



- Find *T. cruzi* protein coding genes that are upregulated in amastigotes compared to trypomastigotes. Go to the transcript expression section then select **microarray**. Choose the fold change (FC) search for the data set called: **Transcriptomes of Four Life-Cycle Stages (Minning et al.)**.

Fold Change Percentile

Identify Genes based on *T. cruzi* CL Brener Esmeraldo-like Transcriptomes of Four Life-Cycle Stages Microarray (fold change)

For the Experiment
 Transcriptomes of Four Life-Cycle Stages trcuCLBrenerEsmeraldo-lik

return protein coding Genes
 that are up-regulated
 with a Fold change ≥ 2.0

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

☐ amastigotes
☒ trypomastigotes
☐ epimastigotes
☐ metacyclics

select all | clear all

and its expression value
 in the following Comparison Samples

☒ amastigotes
☐ trypomastigotes
☐ epimastigotes
☐ metacyclics

select all | clear all

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

Up-regulated

You are searching for genes that are up-regulated between one reference sample and one comparison sample.

For each gene, the search calculates:

$$\text{fold change} = \frac{\text{comparison expression value}}{\text{reference expression value}}$$

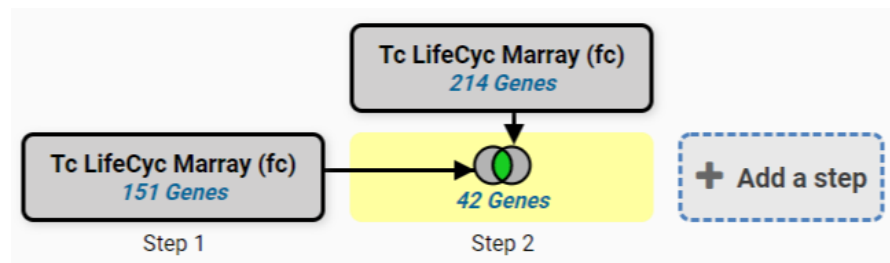
and returns genes when fold change ≥ 2.0 .

See the detailed help for this search.

Advanced Parameters

Get Answer

- Select the direction of regulation, your reference sample and your comparison sample. For the fold change keep the default value 2.
- How many genes did you find? Do the results seem plausible?
- Are any of these genes also up-regulated in the replicative insect stage compared to the transmissive insect stage? How can you find this out? (*Hint*: add a step and run a microarray search comparing expression of epimastigotes to metacyclics).



- Do these genes have orthologs in other kinetoplastids? Transform your results into orthologs in all other organisms in TriTrypDB. This can be done by adding a step, or by editing a step, as shown in the screenshots.

Details for step Combine Gene results

42 Genes

Revise as a boolean operation

1 INTERSECT 2 1 UNION 2 1 MINUS 2 2 MINUS 1

Revise as a span operation

1 RELATIVE TO 2, using genomic colocation

Ignore one of the inputs

IGNORE 2 IGNORE 1

Revise

TcChr24-S:392,078..392,896(+) Mitochondrial outer membrane protein porin, p

TcChr26-S:87,624..88,388(+) fatty acid desaturase, putative (fragment)

TcChr26-S:708,934..710,118(-) hypothetical protein, conserved


Orthologs 2,792 Genes


- ## My Search Strategies


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


- a. Find all genes in *P. falciparum* that are up-regulated at least 50-fold in ookinetes compared to other stages: “**Transcriptomes of 7 sexual and asexual life stages (Lopez-Barragan et al.)**”. For this search select “average” for the operation applied on the reference samples.



Identify Genes based on P. falciparum 3D7 Transcriptomes of 7 sexual and asexual life stages RNASeq (fold change) Tutorial


For the Experiment Transcriptomes of 7 sexual and asexual life stages unstranded 


return protein coding  Genes

that are up-regulated 

with a Fold change \geq 50 

between each gene's average  expression value 

(or a Floor of 10 reads (88 FPKM)  NEW)

in the following Reference Samples 

☒ Ring


☒ Early Trophozoite

☒ Late Trophozoite


☒ Schizont

☒ Gametocyte II

[select all](#) | [clear all](#)

and its expression value 

(or the Floor selected above)

in the following Comparison Samples 

☐ Late Trophozoite

☐ Schizont

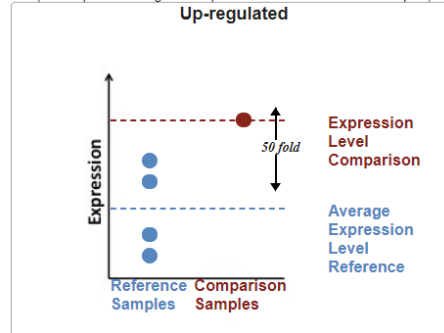
☐ Gametocyte II

☐ Gametocyte V

☒ Ookinete

[select all](#) | [clear all](#)

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



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 one **comparison sample**.

For each gene, the search calculates:

$$\text{fold change} = \frac{\text{comparison expression level}}{\text{average expression level in reference}^*}$$

and returns genes when **fold change \geq 50**.

To narrow the window, use the maximum reference value. To broaden the window, use the minimum reference value.

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

* or FPKM Floor, whichever is greater

[Get Answer](#)

- b. The above search will give you all genes that are up-regulated by 50 fold in ookinetes compared to the average expression level of other stages. Despite the high fold change, some genes in the list may be highly expressed in the other stages. How can you remove genes from the list that are highly expressed in the other stages?

3D7 7Stages RNA-Seq (fc)
31 Genes

+ Add a step



Step 1

- *Hint: Add a search for genes based on RNA Seq evidence from the same experiment, but this time select the percentile search: P.f. seven stages - RNA Seq (percentile). What minimal percentile values should you choose? 40 – 100%? How does setting the any / all samples impact the result Which would be better in this case?*




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.

Filter Data Sets: ?

Legend: **DE** Differential Expression **FC** Fold Change **P** Percentile **SA** SenseAntisense

 Organism ?	Data Set	Choose a Search
<i>Plasmodium falciparum</i> 3D7	? Transcriptomes of 7 sexual and asexual life stages (Lopez-Barragan et al.)	FC P
↓ Show All Data Sets ↓		

Fold Change Percentile

? Experiment

Transcriptomes of 7 sexual and asexual life stages unstranded ▾

? Samples

- ☒ Ring
 - ☒ Early Trophozoite
 - ☒ Late Trophozoite
 - ☒ Schizont
 - ☒ Gametocyte II
 - ☒ Gametocyte V
 - ☐ Ookinete
- [select all](#) | [clear all](#)

? Minimum expression percentile

40 

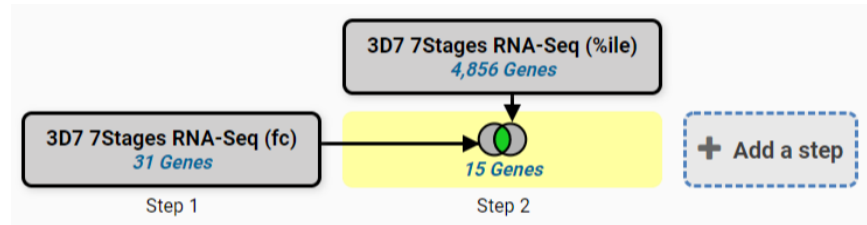
? Maximum expression percentile

100

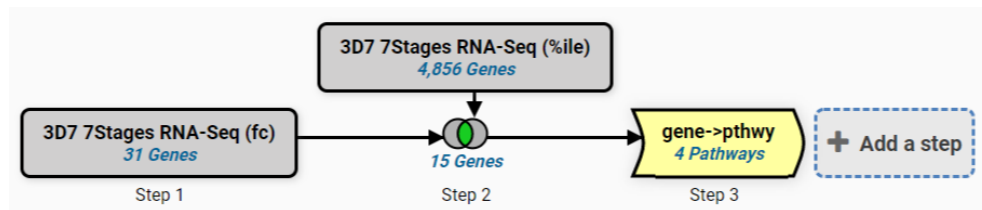
? Matches Any or All Selected Samples?

any ▾

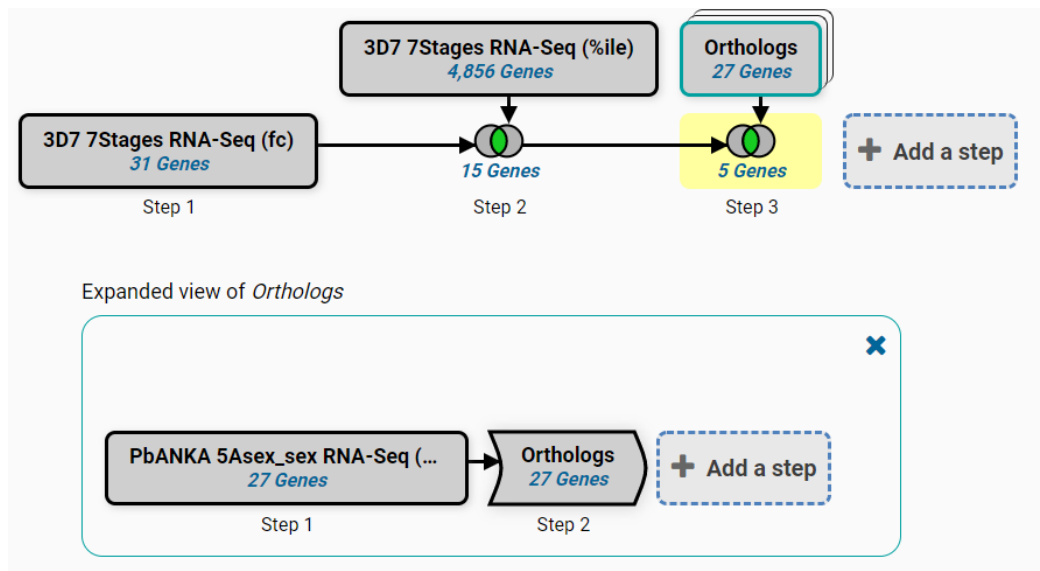
- *Hint II: Try changing the operator from average to maximum for the set of non-ookinete stages in your initial fold change search. What does this do? How do the resulting genes compare with the two step strategy you generated in the first hint? Which hint do you think works better?*



- c. Which metabolic pathways are represented in this gene list? *Hint: add a step and transform results to pathways.* How does this result compare to running a pathways enrichment on step 2?



- d. What happens if you revise the first step and modify the fold difference to a lower value - 10 for example? Compare results when you also modify the “between each genes” parameter. What happens if you set this to maximum? Which value do you think is most stringent for ensuring at 10 fold up regulation compared to the other samples?
- e. PlasmoDB also has an experiment examining gene expression during sexual development in *Plasmodium berghei* (rodent malaria). Can you determine if there are genes that are up-regulated in both human and rodent ookinetes (compared to all other stages)? *Hint:* start by deleting the last step you added in this exercise (transform to pathways). To do this click on edit then delete in the popup. Next, add steps for the *P. berghei* experiments “P berghei ANKA 5 asexual and sexual stage transcriptomes RNASeq”. Note that you will have to use a nested strategy or by running a separate strategy then combining both strategies.



4. Find genes that are essential in procyclics but not in blood form *T. brucei*.
Note: for this exercise use <http://TriTrypDB.org>.
 - Find the query for High Throughput Phenotyping. Think about how to set up this query (*Hint*: you will have to set up a two-step strategy). Remember you can play around with the parameters but there is no one correct way of setting them up –

Quantitative Phenotype

Learn more about this search

Identify Genes based on High-Throughput Phenotyping

Tutorial

For the **Experiment** Quantitated from the CDS Sequence

return protein coding Genes

that are Decrease in coverage

with a **Fold change** ≥ 1.5

between each gene's **expression value**

in the following **Reference Samples**

Uninduced sample

and its **expression value**

in the following **Comparison Samples**

☐ Induced in bloodstream (BS) forms, 3 days (10 doublings)
☐ Induced in bloodstream (BS) forms, 6 days (20 doublings)
☒ Induced in procyclic forms (PS) forms, 9 days (9 doublings)
☐ Induced throughout differentiation (DIF = 7 BS doublings + 6 PS doublings)

select all | clear all

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

Down-regulated

You are searching for genes that are **down-regulated** between one **reference sample** and one **comparison sample**.

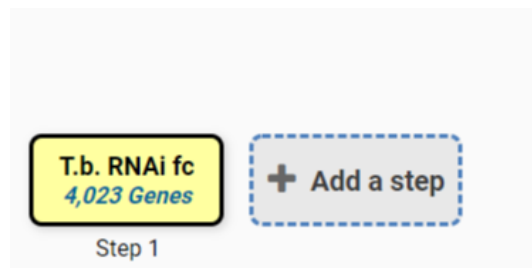
For each gene, the search calculates:

$$\text{fold change} = \frac{\text{reference expression level}}{\text{comparison expression level}}$$

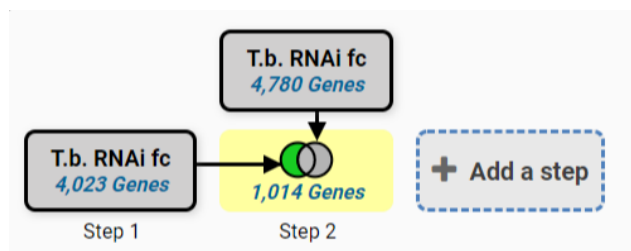
and returns genes when **fold change** ≥ 1.5 .

See the detailed help for this search.

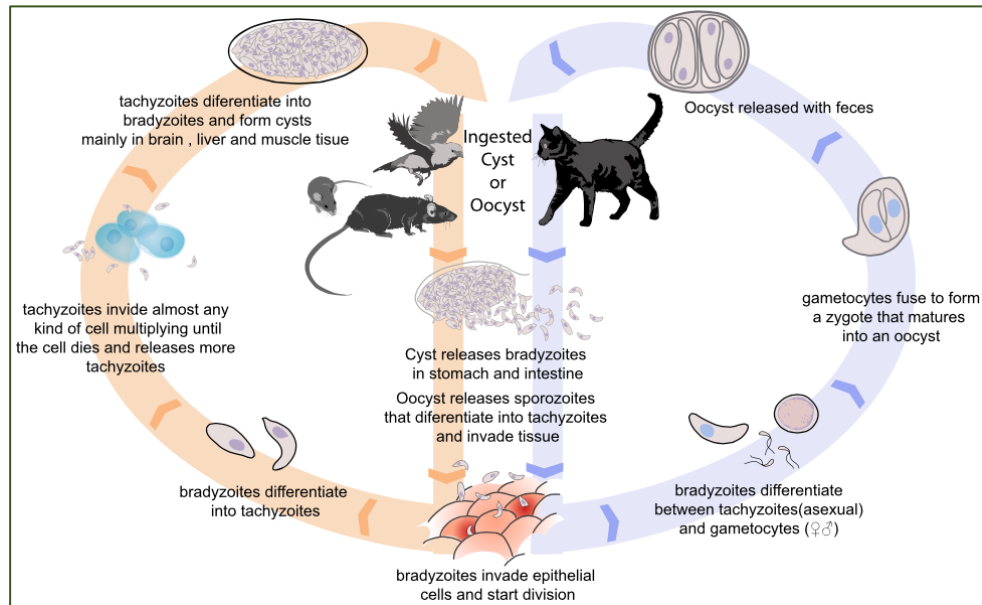
Get Answer



- Next add a step and run the same search except this time select the “induced bloodstream form” samples.
- How did you combine the results? Remember you want to find genes that are essential in procyclics and not in blood form.



5. Finding oocyst expressed genes in *T. gondii* based on microarray evidence.
Note: For this exercise use <http://toxodb.org>



- a. Find genes that are expressed at 10 fold higher levels in one of the oocyst stages than in any other stage in the “Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4) (Fritz and Buchholz)” microarray experiment.

Search for...

expand all | collapse all

Filter the searches below...

Genes

- Annotation, curation and identifiers
- Epigenomics
- Function prediction
- Gene models
- Genetic variation
- Genomic Location
- Immunology
- Orthology and synten
- Pathways and interactions
- Phenotype
- Protein features and properties
- Protein targeting and localization
- Proteomics
- Sequence analysis
- Structure analysis
- Taxonomy
- Text
- Transcriptomics
 - Microarray Evidence
 - RNA-Seq Evidence

Filter Data Sets | oocyst | Legend: S Similarity FC Fold Change P Percentile

Organism: T. gondii ME49 (filtered from 11 total entries)

Data Set: Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4) (Fritz and Buchholz et al.)

Choose a search: FC P

Show All Data Sets

Learn more about this search

Identify Genes based on T. gondii ME49 Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4) Microarray (fold change)

For the Experiment: Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4)

return | protein coding | Genes

that are up or down regulated

with a Fold change >= 10

between each gene's [average] expression value

in the following Reference Samples

- ☐ 10 days sporulated
- ☒ 2 days in vitro
- ☒ 4 days in vitro
- ☒ 8 days in vitro
- ☒ 21 days in vivo

select all | clear all

and its [average] expression value

in the following Comparison Samples

- ☒ unsporulated
- ☒ 4 days sporulated
- ☒ 10 days sporulated
- ☐ 2 days in vitro
- ☐ 4 days in vitro

select all | clear all

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:

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

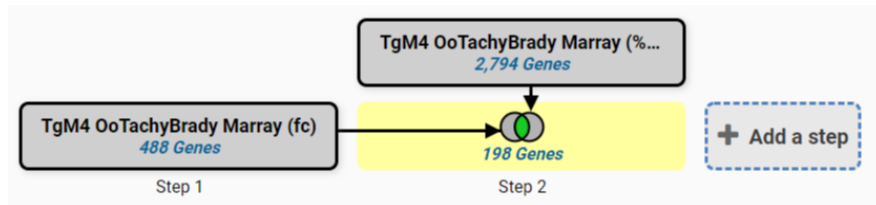
$$\text{fold change}_{\text{down}} = \frac{\text{average expression level in reference}}{\text{average expression level in comparison}}$$

and returns genes when fold change_{up} >= 10 or fold change_{down} >= 10.

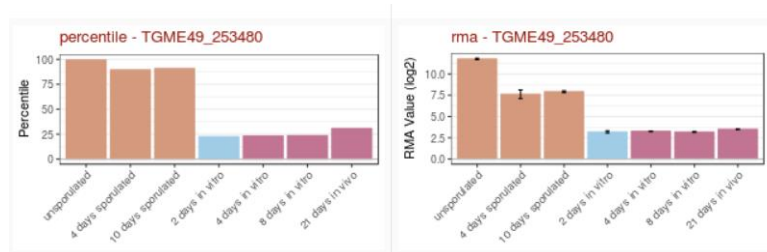
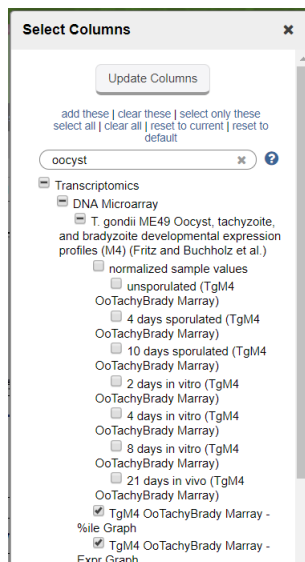
See the detailed help for this search.

Get Answer

- b. Add a step to limit this set of genes to only those for which all the non-oocyst stages are expressed below 50th percentile ... i.e., likely not expressed at those stages. (Hint: after you click on add step find the same experiment under microarray expression and chose the percentile search).
- Select the 4 **non-oocyst** samples.
 - We want all to have less than 50th percentile so set **minimum percentile to 0** and **maximum percentile to 50**.
 - Since we want all of them to be in this range, choose **ALL** in the "Matches Any or All Selected Samples".



- To view the graphs in the final result table, turn on the columns called “TgM4 OoTachyBrady Marray - Expr Graph” and “TgM4 OoTachyBrady Marray - %ile Graph” (inside the “*T. gondii* ME49 Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4) (Fritz and Buchholz et al.)” Microarray).



6. Comparing RNA abundance and Protein abundance data.

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

In this exercise we will compare genes that show differential RNA abundance levels between procyclic and blood form stages in *T. brucei* with genes that show differential protein abundance in these same stages.

- a. Find genes that are down-regulated 2-fold in procyclic form cells. Go to the search page for Genes by Microarray Evidence and select the fold change search for the “Expression profiling of five life cycle stages (Marilyn Parsons)” experiment and configure the search to return protein-coding genes that are down-regulated 2 fold in procyclic form (PCF) relative to the Blood Form reference sample. Since there are two PCF samples, it is reasonable to choose both and average them.

Identify Genes based on Microarray Evidence

Search for...

expand all | collapse all

Filter the searches below...

- Genes
 - Annotation, curation and identifiers
 - Function prediction
 - Gene models
 - Genetic variation
 - Genomic Location
 - Immunology
 - Orthology and synteny
 - Pathways and interactions
 - Phenotype
 - Protein features and properties
 - Protein targeting and localization
 - Proteomics
 - Sequence analysis
 - Structure analysis
 - Taxonomy
 - Text
 - Transcriptomics
 - Microarray Evidence**
 - RNA-Seq Evidence

Filter Data Sets:

Legend: DC Direct Co... FC Fold Chan... P Percentile

Organism	Data Set	Choose a search	
<i>L. infantum</i> JPCM5	Promastigote-to-amastigote differentiation (L.d. Samples) (Lahav et al.)	<input type="button" value="FC"/>	<input type="button" value="P"/>
<i>L. infantum</i> JPCM5	Axenic and intracellular amastigote profiles (Rochette et al.)	<input type="button" value="DC"/>	<input type="button" value="P"/>
<i>L. major</i> strain Friedlin	Three Developmental Stages (Stephen M. Beverley)	<input type="button" value="DC"/>	<input type="button" value="P"/>
<i>T. brucei</i> brucei TREU927	Expression profiling of in vitro differentiation (Queiroz et al.)	<input type="button" value="FC"/>	<input type="button" value="P"/>
<i>T. brucei</i> brucei TREU927	Expression profiling of five life cycle stages (Marilyn Parsons)	<input type="button" value="FC"/>	<input type="button" value="P"/>
<i>T. brucei</i> brucei TREU927	Procyclic trypanosomes: heat shock vs untreated control (Kramer et al.)	<input type="button" value="DC"/>	<input type="button" value="P"/>
<i>T. brucei</i> brucei TREU	Identify Genes based on T.brucei Expression profiling of five life cycle stages Microarray (fold change)	<input type="button" value="FC"/>	<input type="button" value="P"/>
<i>T. brucei</i> brucei TREU		<input type="button" value="FC"/>	<input type="button" value="P"/>
<i>T. cruzi</i> CL Brener ESR		<input type="button" value="DC"/>	<input type="button" value="P"/>

For the Experiment: Expression profiling of five life cycle stages

return: protein coding ☒ Genes

that are: down-regulated ☒

with a Fold change >= 2.0

between each gene's average expression value

in the following **Reference Samples**

☒ Blood Form
☒ Slender
☒ Stumpy
☐ PCF Log
☐ PCF Stat

and its average expression value

in the following **Comparison Samples**

☐ Blood Form
☐ Slender
☐ Stumpy
☒ PCF Log
☒ PCF Stat

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 **down-regulated** between at least two **reference samples** and at least two **comparison samples**.

For each gene, the search calculates:

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

and returns genes when fold change >= 2.0. To narrow the window, use the minimum reference value, or maximum comparison value. To broaden the window, use the maximum reference value, or minimum comparison value.

See the detailed help for this search.

Tb LifeCyc Marray (fc)
378 Genes


+ Add a step

Step 1

- b. Add a step to compare with quantitative protein expression. Select protein expression then “Quantitative Mass Spec Evidence” and the "Quantitative phosphoproteomes of bloodstream and procyclic forms (Tb427) (Urbaniak et al.)" experiment. Configure this search to return genes that are down-regulated in procyclic form relative to blood form.

← Add a step to your search strategy ⓘ

Search for Genes by Quantitative Mass Spec. Evidence

The results will be  intersected with | ▾ the results of Step 2.

Filter Data Sets: ⓘ Legend: DC Direct Comparison DCC Direct Confidence Comparison GR Quantitative Ratio FC Fold Change

Organism ⓘ	Data Set	Choose a Search
<i>Trypanosoma brucei</i> brucei TREU927	Quantitative phosphoproteomes of bloodstream and procyclic forms (Tb427) (Urbaniak et al.)	DC

↓ Show All Data Sets ↓

Direct Comparison

Experiment

Quantitative phosphoproteomes of bloodstream and procyclic forms (Tb427) ▾

Direction

down-regulated ▾

Comparison

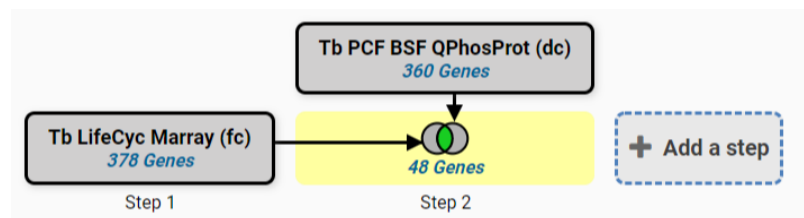
☒ Pcf-Bsf ratio

Fold difference >=

2

Run Step

- c. How many genes are in the intersection? Does this make sense? Make certain that you set the directions correctly.



- d. Try changing directions and compare up-regulated genes/proteins. (Hint: revise the existing strategy ... you might want to duplicate it so you can keep both). When you change one of the steps but not the other do you have any genes in the intersection? Why might this be?

- e. Can you think of ways to provide more confidence (or cast a broader net) in the microarray step? (*Hint*: you could insert steps to restrict based on percentile or add a RNA Sequencing step that has the same samples).

7. Find genes with evidence of protein phosphorylation in intracellular *Toxoplasma* tachyzoites.

For this exercise use <http://www.toxodb.org>

Phosphorylated peptides can be identified by searching the appropriate experiments in the [Mass Spec Evidence](#) search page.

- 7a. Find all genes with evidence of protein phosphorylation in intracellular tachyzoites. Navigate to the Post-Translational Modification search. Select the “Infected host cell, phosphopeptide-enriched (peptide discovery against TgME49)” sample under the experiment called “Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treeck et al.)”

The screenshot displays the Toxodb.org search interface. On the left, a sidebar titled 'Search for...' contains a list of search categories. The 'Post-Translational Modification' category is highlighted with a red circle and a blue arrow pointing to the main search results area. The main area is titled 'Identify Genes based on Post-Translational Modification'. It features several filters: 'Type of Post-Translational Modification' set to 'phosphorylation site', 'Experiments and Samples' showing '1 selected, out of 9' results, 'Number of modifications is' set to 'Greater than or equal to', and 'Number of Modifications' set to '1'. The search results list several experiments, including 'Toxoplasma gondii GT1' and 'Toxoplasma gondii ME49'. The 'Toxoplasma gondii ME49' experiment is expanded, showing a list of samples. The sample 'Infected host cell, phosphopeptide-enriched (peptide discovery against TgME49)' is selected, indicated by a checkmark. At the bottom right, there is a 'Get Answer' button.


- 7b. Remove all genes with phosphorylation evidence from purified tachyzoites and the phosphopeptide depleted fractions.

Hint: Use the Mass Spec Evidence search to access the tachyzoite and depleted fractions. Subtract (1 minus 2) these results from your first search.



Add a step to your search strategy ?

Search for Genes by Mass Spec. Evidence

The results will be  subtracted from |  the results of Step 1.

Experiments and Samples

*Note: You must select at least 1 values for this parameter.
3 selected, out of 92*

[select all](#) | [clear all](#) | [expand all](#) | [collapse all](#)

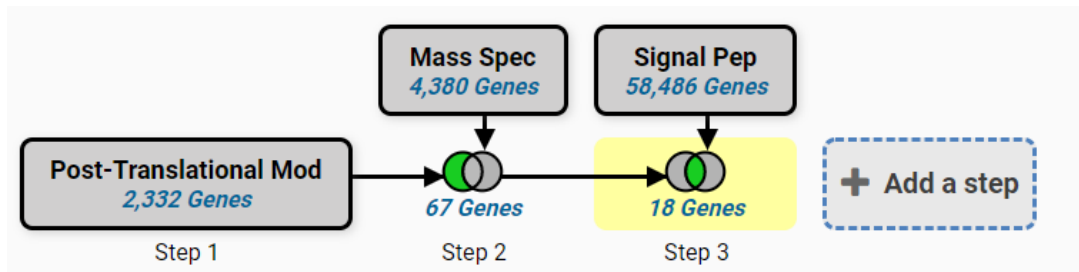
Filter list below...

- ☐ Eimeria
- ☐ Neospora
- ☒ Toxoplasma
 - ☒ Toxoplasma gondii
 - ☐ Toxoplasma gondii GT1
 - ☒ Toxoplasma gondii ME49
 - ☐ *T. gondii* Proteome During Infection in Homo sapiens (GT1 ME49 RH VEG) (Krishna et al.)
 - ☐ Extracellular vesicles (RH) (Wowk et al.)
 - ☐ Mitochondrial Matrix Proteome (Seidi and Mueller-Wong et al.)
 - ☒ Monomethylarginine Proteomics (RH) (Yakubu et al.)
 - ☐ monomethylarginine
 - ☐ Mouse brain bradyzoite proteomics time course (Garfoot et al.)
 - ☐ N-terminal Peptides (RH) (Dogga et al.)
 - ☒ Oocyst Partially Sporulated Proteome (VEG) (Possenti et al.)
 - ☐ Oocyst proteome
 - ☒ Oocyst proteome (M4 Typell) (Wastling)
 - ☐ Oocyst peptides
 - ☐ Oocyst proteome - Fractionated (M4 type II) (Fritz et al.)
 - ☒ TAILS peptides (Coffey et al.)
 - ☐ TAILS N-terminal proteomics
 - ☐ Tachyzoite Intra- and Extracellular Lysine-Acetylomes (RH) (Jeffers and Xue)
 - ☒ Tachyzoite Rhoptry proteome (RH) (Bradley et al.)
 - ☐ purified rhoptries
 - ☐ Tachyzoite Ubiquitome (Silmon de Monerri et al.)
 - ☐ Tachyzoite conoid proteome (RH) (Hu et al.)
 - ☐ Tachyzoite membrane and cytosolic proteomes (RH) (Dybas et al.)
 - ☐ Tachyzoite phosphoproteome - Calcium dependent (RH) (Nebi et al.)
 - ☒ Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treeck et al.)
 - ☒ Infected host cell, phosphopeptide-depleted (peptide discovery against TgME49)
 - ☐ Infected host cell, phosphopeptide-enriched (peptide discovery against TgME49)
 - ☒ Purified tachyzoites phosphopeptide-depleted (peptide discovery against TgME49)
 - ☒ Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgME49)
 - ☐ Tachyzoite secretome (RH) (Zhou et al.)
 - ☐ Tachyzoite subcellular fractions (Moreno)
 - ☐ Tachyzoite total proteome (RH) (Wastling)

[select all](#) | [clear all](#) | [expand all](#) | [collapse all](#)

7d. Explore your results. What kinds of genes did you find? *Hint: use the Product description word column or perform a GO enrichment analysis of your results.*

7e. Are any of these genes likely to be secreted? Hint: add a step searching for genes with secretory signal peptides.



7f. Pick one or two of the hypothetical genes in your results and visit their gene pages. Can you infer anything about their function? Hint: explore the protein and expression sections.

7g. What about polymorphism data? Go back to your strategy and add columns for SNP data found under the population biology section. Explore the gene page for the gene that has the most number of non-synonymous SNPs. Hint: you can sort the columns by clicking on the up/down arrows next to the column names.

ToxoDB Toxoplasma Informatics Resources

Site search, e.g. *TOXO49_239250* or "hydroxase or" "binding protein"

My Strategies Searches Tools My Workspace Data About Help Contact Us

My Search Strategies

Opened (1) All (2) Public (17) Help

Unnamed Search Strategy

18 Genes (18 ortholog groups)

Gene Results Genome View Endpoint Results

Rows per page: 20

Download Add to Booklet Add Columns

Organism Filter

select all Clear all (required all) Collapse all

Hide zero counts

Search expression

18

18 Genes (18 ortholog groups)

Gene ID	Transcript ID	Gene Group (Representative gene)	Genomic Location (Gene)	Product Description	# Transcripts	Typed OnChyHydrolysis Memory - Expr Graph	Non-Coding SNPs All Strains	Non-Coding SNPs All Strains	Non-Coding SNPs All Strains	Non-Coding SNPs All Strains	Non-Coding SNPs All Strains	Non-Coding SNPs All Strains	Non-Coding SNPs All Strains	Non-Coding SNPs All Strains
TOXO49_237955	TOXO49_237955	TOXO49_237955:CDL1	TOXO49_237955:CDL1	hypothetical protein	1		131	2.32	130	0	56			
TOXO49_219640	TOXO49_219640	TOXO49_219640:CDL1	TOXO49_219640:CDL1	hypothetical protein	1		264	2.5	85	0	34			

8. Find *T. gondii* genes expressed in late enteroepithelial stages

Toxoplasma gondii is a zoonotic pathogen for which felids serve as definitive hosts. In cats, the parasite undergoes several rounds of asexual replication before entering the sexual cycle which gives rise to oocysts that are shed into the environment. These then sporulate and become infective to humans and livestock. To understand the genes involved in the parasite development in the felid host and identify potential intervention targets, we designed a transcriptomic approach to compare the cat intestinal stages with the well characterized tachyzoites that mediate acute infection and tissue cysts that are responsible for chronic infection. Cats were infected with *T. gondii* CZ clone H3 tissue cysts from mouse brain and the intestinal stages were sampled at day 3, 5 and 7 post infection. As an input sample, we also collected tissue cysts from mouse brain. In vitro cultivated tachyzoites were also harvested. Total RNA was extracted, enriched for mRNA and used for cDNA synthesis. RNA-Seq was then performed to describe the transcriptomic repertoire of each developmental stage. RNA-seq datasets from each time point post inoculation with bradyzoites in kittens were subjected to cluster analysis and assigned to five enteroepithelial developmental stages (EES) according to their profile.

Cat enteroepithelial stages:

- EES1 = very early enteroepithelial stages
- EES2 = early enteroepithelial stages
- EES3 = mixed enteroepithelial stages
- EES4 = late enteroepithelial stages
- EES5 = very late enteroepithelial stages
- Navigate to the RNAseq searches and identify the experiment of cat enterocyte stages. Configure the search to identify call *T. gondii* genes that are upregulated by at least 2-fold in late and very late enteroepithelial stages (EES4 and EES5) compared to all other stages available from this experiment.

Identify Genes based on RNA-Seq Evidence

Filter Data Sets: [?](#) Legend: **DE** Differential Expression **FC** Fold Change **P** Percentile **SA** SenseAntisense

Organism [?](#) Data Set Choose a Search

Toxoplasma gondii ME49 [?](#) Feline enterocyte, tachyzoite, bradyzoite stage transcriptome (Hehl, Ramakrishnan et al.) **DE** **FC** **P** **SA**

↓ Show All Data Sets ↓

Differential Expression Fold Change Percentile SenseAntisense

Identify Genes based on T. gondii ME49 Feline enterocyte, tachyzoite, bradyzoite stage transcriptome RNA-Seq (fold change)

For the Experiment
 [?](#)

return [?](#) **Genes** [?](#)

that are [?](#)

with a Fold change >= [?](#)

between each gene's [?](#) expression value
 (or a Floor of [?](#))

in the following **Reference Samples** [?](#)

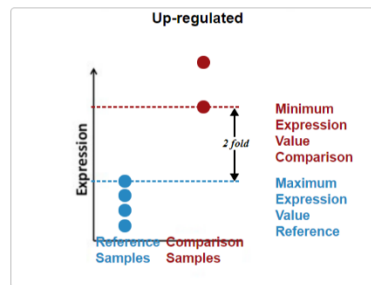
☒ EES3
☐ EES4
☐ EES5
☒ Tachyzoites
☒ Tissue cysts

and its [?](#) expression value
 (or the Floor selected above)

in the following **Comparison Samples** [?](#)

☐ EES1
☐ EES2
☐ EES3
☒ EES4
☒ EES5
☒ Tachyzoites

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



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

For each gene, the search calculates:

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

and returns genes when fold change >= 2.

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

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 maximum reference value, or average or maximum comparison value.

Get Answer

Toxo Cat RNAseq (fc)
238 Genes

Step 1

+ Add a step

- What kinds of genes did this search identify? How can you determine if your results are enriched for a particular function? Try clicking on Analyze Results and explore the GO enrichment tool.

My Search Strategies

Search Strategy: 4535, Public (17), 19/10/2020

Organism: Toxoplasma gondii

Search Strategy: 4535, Public (17), 19/10/2020

2281 Genes (2281 of 2281 genes) (20/10/2020)

Organism: Toxoplasma gondii

Search Strategy: 4535, Public (17), 19/10/2020

2281 Genes (2281 of 2281 genes) (20/10/2020)

Search Strategy: 4535, Public (17), 19/10/2020

2281 Genes (2281 of 2281 genes) (20/10/2020)

Analyze your Gene results with a tool below.



Organism: Toxoplasma gondii ME49

Ontology: ☐ Cellular Component ☐ Molecular Function ☒ Biological Process

Evidence: ☒ Computed ☒ Curated [select all](#) | [clear all](#)

Limit to GO Slim terms: ☒ No ☐ Yes

P-Value cutoff: 0.05 (0 - 1)

Submit

Analysis Results:

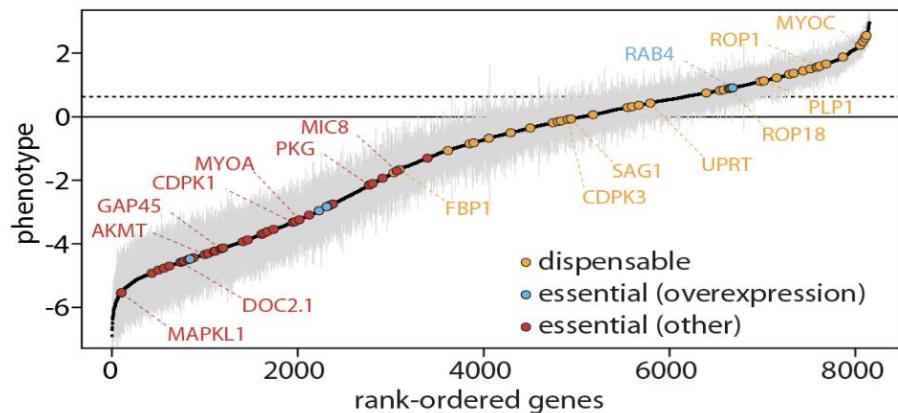
GO ID	GO Term	Genes in the list with this term	Genes in your result with this term	Percent of list genes in your result	Fold enrichment	Odds ratio	P-value	Benjamini	Bonferroni
GO:0016070	phosphorylation	192	11	5.7	2.47	2.76	4.35e-3	3.28e-1	5.96e-1
GO:0006468	protein phosphorylation	167	10	6.0	2.58	2.90	4.85e-3	3.28e-1	6.57e-1
GO:0016044	cysteine biosynthetic processes	1	1	100.0	43.09	infinity	3.52e-2	4.20e-1	1.00e+0
GO:0016040	cysteine biosynthetic processes via cystathionine	1	1	100.0	43.09	infinity	3.52e-2	4.20e-1	1.00e+0
GO:0006012	galactose metabolic process	1	1	100.0	43.09	infinity	3.52e-2	4.20e-1	1.00e+0

9. Finding genes based on high throughput mutagenesis and fitness analysis.

In EuPathDB we have a variety of studies where genome scale phenotypic analyses were carried out. In this exercise we'll use [ToxoDB.org](https://toxodb.org/) and look at fitness following CRISPR mutagenesis. You could also explore phenotyping studies in PlasmoDB or FungiDB if you prefer, the principles are the same.

- Navigate to the CRISPR phenotype search. Note that this search form is quite simple just requiring a range of fitness values. The defaults return all genes not limiting the search at all. This is only useful in as much as it tells you which genes were assayed which is nearly the entire genome. The tricky bit is deciding where to make the cutoffs. Again, the description on the search form is very helpful in this regard (as is the link to the paper ... remember these phenotypes were assayed under specific conditions so just because a particular gene doesn't

show a phenotype doesn't mean it wouldn't in other conditions (or infecting an actual host). The plot showing the phenotype score (fitness) is particularly useful. Red points along the plot are genes known to be essential under these conditions



while yellow are known to be expendable. This will help you determine where to set the values. The last essential gene has a fitness score just \geq than -4 so setting the phenotype score \leq -2 would provide a pretty stringent search but still return more than 1000 genes. Try it. Do you get the expected results based on the number of genes returned?

Search for...

expand all | collapse all

Filter the searches below...

Genes

- Annotation, curation and identifiers
- Epigenomics
- Function prediction
- Gene models
- Genetic variation
- Genomic Location
- Immunology
- Orthology and synteny
- Pathways and interactions
- Phenotype
 - CRISPR Phenotype**
- Protein features and properties
- Protein targeting and localization
- Proteomics
- Sequence analysis
- Structure analysis
- Taxonomy
- Text
- Transcriptomics

Identify Genes based on CRISPR Phenotype

Phenotype Score \geq

-4

Phenotype Score \leq

-2

CRISPR
1,542 Genes

+ Add a step

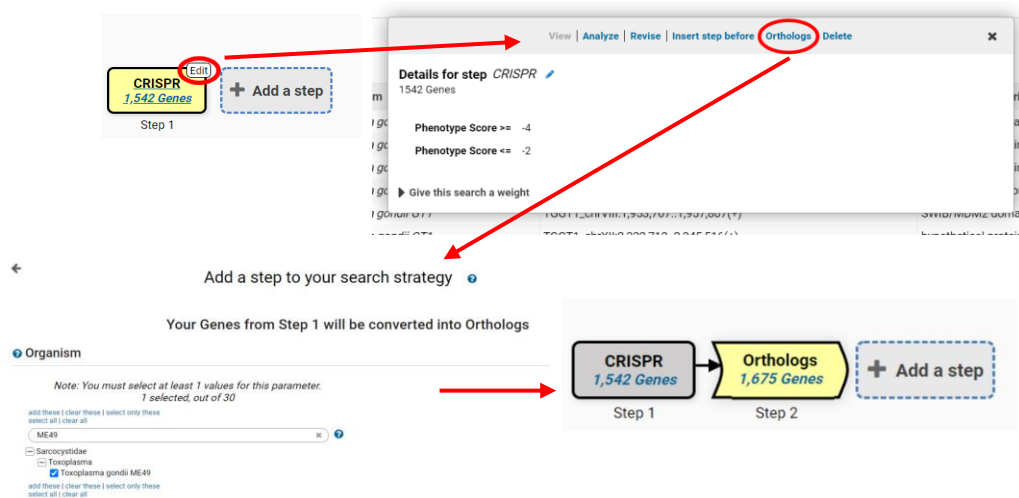
Step 1

Get Answer

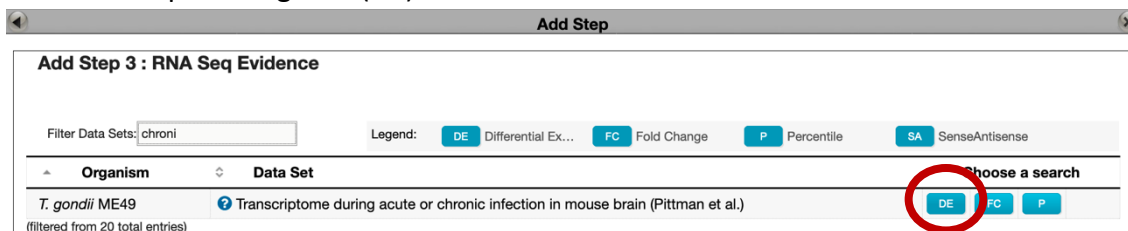
- Can you find additional evidence that these genes are essential? One way is to use the analysis tools to assess biological process and go function. Are the results what you would expect?



- Try adding columns to show additional data or intersecting these results with other queries, perhaps expression queries, to further assess this list. NOTE: this experiment was done in GT1 while all *T. gondii* functional data in ToxoDB is mapped to ME49 so an ortholog transform to ME49 is required before adding any additional functional studies.
- To do this, click on add step and select the Transform to orthologs option and select *T. gondii* ME49 to transform to.



- How many of these genes are upregulated in *in vivo* chronic stages of *T. gondii*?
 - Click on add step
 - Select the RNAseq searches under the Transcriptomics category
 - Find the experiment with chronic stages and run a search based on differentially expressed genes (DE).



- Intersect genes that are 2-fold upregulated in chronic stages compared to acute stages.
- What do these results look like? Do you find any interesting genes?

Add Step 3 : T. gondii ME49 Transcriptome during acute or chronic infection in mouse brain RNASeq (Differential Expression)

Experiment

Reference Sample
☒ acute infection 10 days p.i.
☐ chronic infection 28 days p.i.

Comparator Sample
☐ acute infection 10 days p.i.
☒ chronic infection 28 days p.i.

Direction

fold difference >=

adjusted P value less than or equal to

Combine Genes in Step 2 with Genes in Step 3:

☒ 2 Intersect 3
 ☐ 2 Minus 3
☐ 2 Union 3
 ☐ 3 Minus 2
☐ 2 Relative to 3, using genomic colocation

CRISPR
 1,542 Genes
 Step 1

Orthologs
 1,675 Genes
 Step 2

Tg In murine macrophages RN...
 314 Genes
 Step 3

+ Add a step