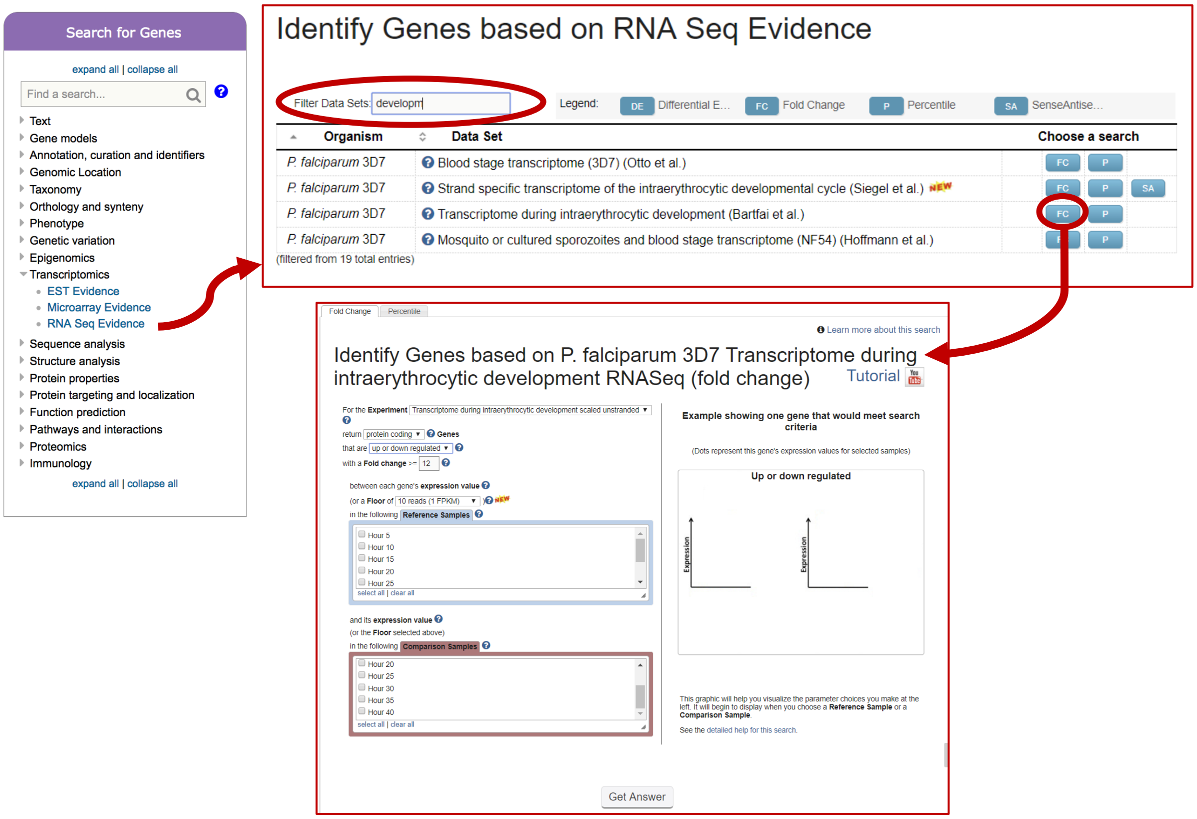
Exploring Transcriptomic data

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.

- Hint: 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.



Search for Genes

expand all | collapse all

Find a search...

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

expand all | collapse all

Identify Genes based on RNA Seq Evidence

Filter Data Sets: developm

Legend: DE Differential E... FC Fold Change P Percentile SA SenseAntise...

Organism	Data Set	Choose a search
P. falciparum 3D7	Blood stage transcriptome (3D7) (Otto et al.)	FC P
P. falciparum 3D7	Strand specific transcriptome of the intraerythrocytic developmental cycle (Siegel et al.) NEW	FC P SA
P. falciparum 3D7	Transcriptome during intraerythrocytic development (Bartfai et al.)	FC P
P. falciparum 3D7	Mosquito or cultured sporozoites and blood stage transcriptome (NF54) (Hoffmann et al.)	FC P

(filtered from 19 total entries)

Fold Change | Percentile

Learn more about this search

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 or down regulated

with a **Fold change** >= 12

between each gene's **expression value**

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

in the following **Reference Samples**

□ Hour 5
□ Hour 10
□ Hour 15
□ Hour 20
□ Hour 25

select all | clear all

and its **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)

This graphic will help you visualize the parameter choices you make at the left. It will begin to display when you choose a **Reference Sample** or a **Comparison Sample**.

See the detailed help for this search.

Get Answer

- 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>=** the intensity of difference in expression needed before a gene is returned by the search. Choose 12 but feel free to modify this.

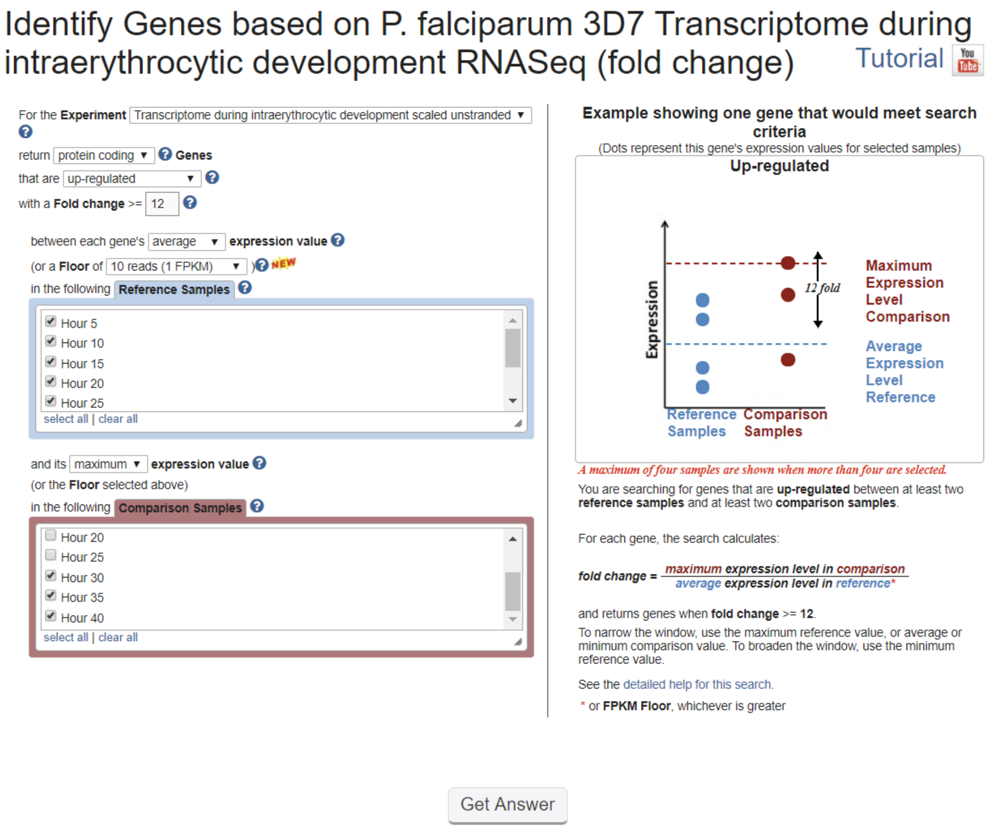
- **Reference Sample**: the samples that will serve as the reference when comparing expression between samples. choose 5, 10, 15, 20, 25

- **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. 2000FPKM/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) ▼) NEW

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)

Up-regulated
Expression
12 fold
Maximum Expression Level Comparison
Average Expression Level Reference
Reference Samples
Comparison 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:

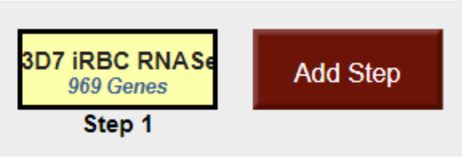
$$\textit{fold change} = \frac{\textit{maximum expression level in comparison}}{\textit{average expression level in reference}^{*}}$$

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.

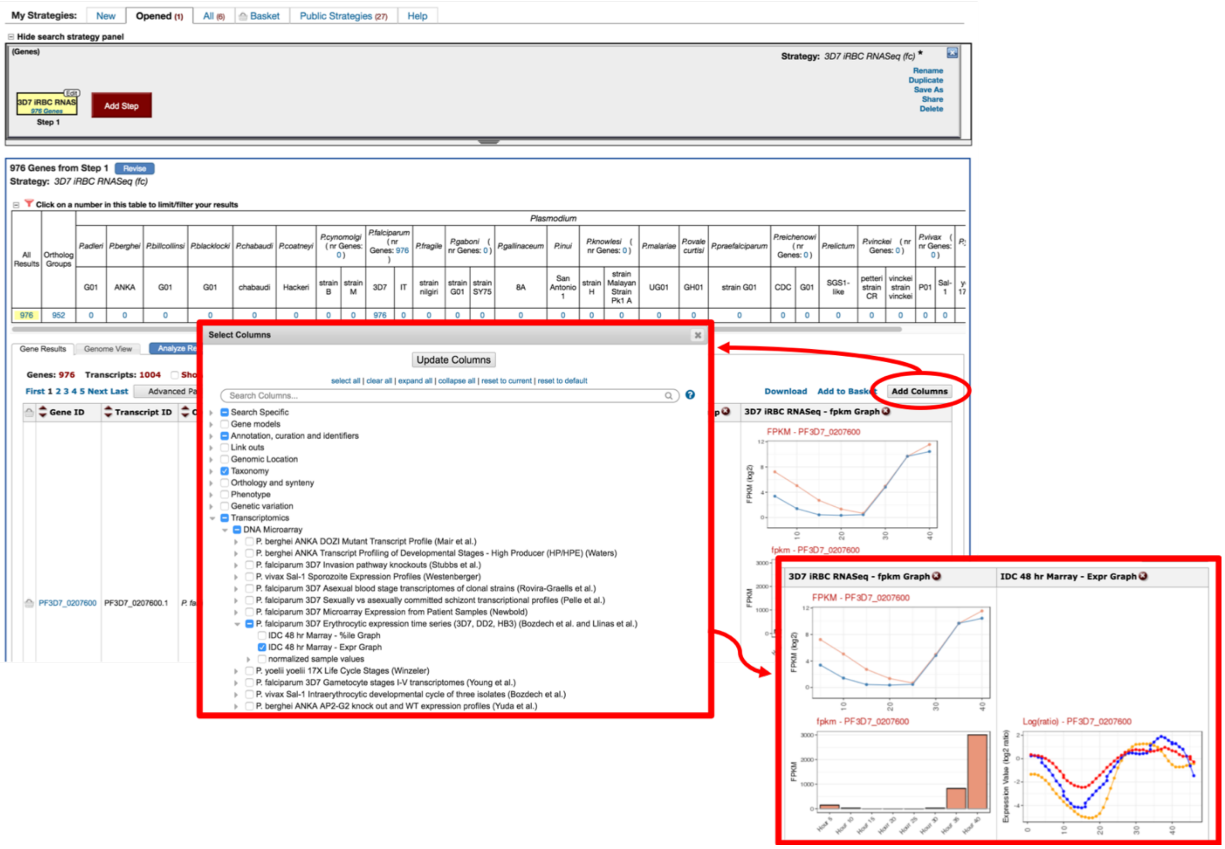
See the detailed help for this search.

* or **FPKM Floor**, whichever is greater



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".



Strategy: *3D7 iRBC RNASeq (fc)* *

976 Genes from Step 1

Strategy: *3D7 iRBC RNASeq (fc)*

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

All Results	Ortholog Groups	*Plasmodium*																												
		P.adleri	*P.berghei*	*P.billcollinsi*	*P.blacklocki*	*P.chabaudi*	*P.coatneyi*	*P.cynomolgi* (nr Genes: 0)		*P.falciparum* (nr Genes: 976)		*P.fragile*	*P.gaboni* (nr Genes: 0)		*P.gallinaceum*	*P.inui*	*P.knowlesi* (nr Genes: 0)		*P.malariae*	*P.ovale curtisi*	*P.praefalciparum*	*P.reichenowi* (nr Genes: 0)		*P.relictum*	*P.vinckei* (nr Genes: 0)		*P.vivax* (nr Genes: 0)		*P.*
		G01	ANKA	G01	G01	chabaudi	Hackeri	strain B	strain M	3D7	IT	strain nilgiri	strain G01	strain SY75	8A	San Antonio 1	strain H	strain Malayan Strain Pk1 A	UG01	GH01	strain G01	CDC	G01	SGS1-like	petteri strain CR	vinckei strain vinckei	P01	Sal-1	y 17
976	952	0	0	0	0	0	0	0	0	976	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Genes: 976 Transcripts: 1004

First 1 2 3 4 5 Next Last

Gene ID	Transcript ID
PF3D7_0207600	PF3D7_0207600.1

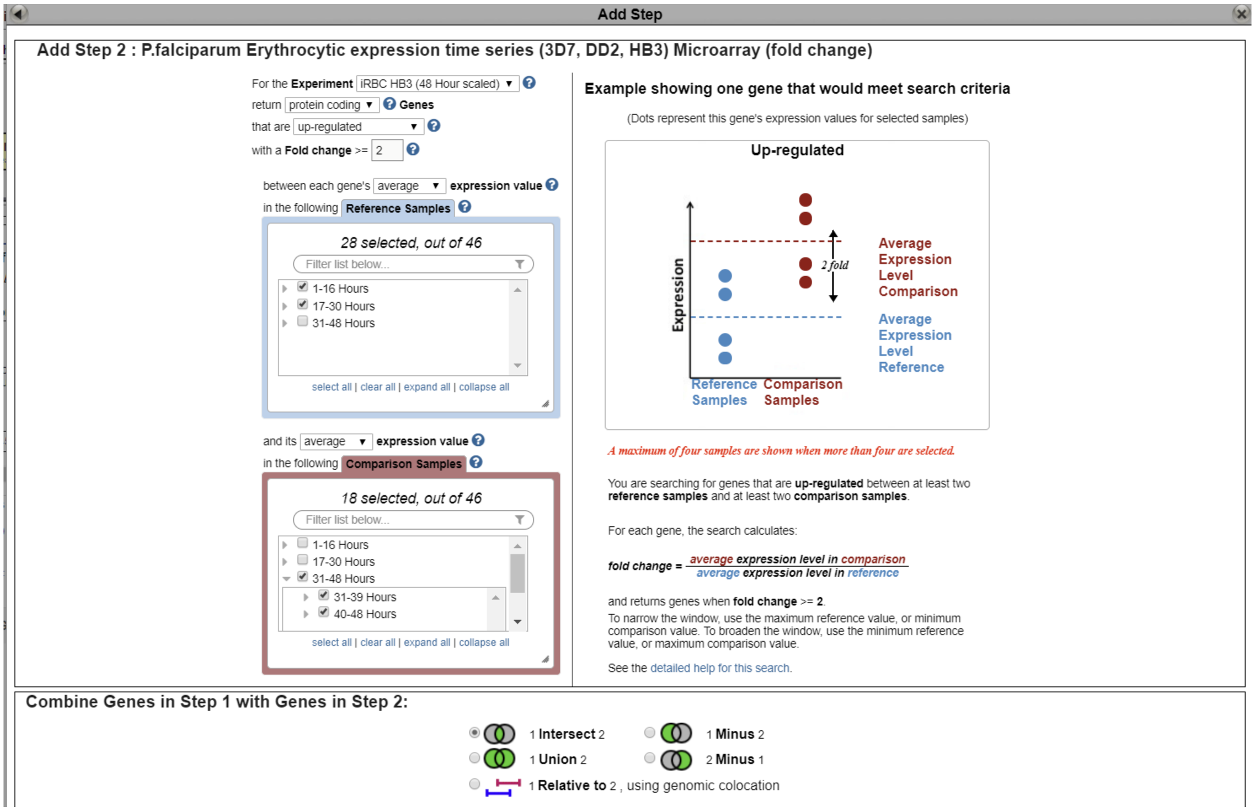
Select Columns

- ▸ ⊟ Search Specific
- ▸ □ Gene models
- ▸ ⊟ Annotation, curation and identifiers
- ▸ □ Link outs
- ▸ □ Genomic Location
- ▸ ☑ Taxonomy
- ▸ □ Orthology and synteny
- ▸ □ Phenotype
- ▸ □ Genetic variation
- ▾ ⊟ Transcriptomics
  - ▾ ⊟ DNA Microarray
    - ▸ □ P. berghei ANKA DOZI Mutant Transcript Profile (Mair et al.)
    - ▸ □ P. berghei ANKA Transcript Profiling of Developmental Stages - High Producer (HP/HPE) (Waters)
    - ▸ □ P. falciparum 3D7 Invasion pathway knockouts (Stubbs et al.)
    - ▸ □ P. vivax Sal-1 Sporozoite Expression Profiles (Westenberger)
    - ▸ □ P. falciparum 3D7 Asexual blood stage transcriptomes of clonal strains (Rovira-Graells et al.)
    - ▸ □ P. falciparum 3D7 Sexually vs asexually committed schizont transcriptional profiles (Pelle et al.)
    - ▸ □ P. falciparum 3D7 Microarray Expression from Patient Samples (Newbold)
    - ▾ ⊟ P. falciparum 3D7 Erythrocytic expression time series (3D7, DD2, HB3) (Bozdech et al. and Llinas et al.)
      - □ IDC 48 hr Marray - %ile Graph
      - ☑ IDC 48 hr Marray - Expr Graph
      - ▸ □ normalized sample values
    - ▸ □ P. yoelii yoelii 17X Life Cycle Stages (Winzeler)
    - ▸ □ P. falciparum 3D7 Gametocyte stages I-V transcriptomes (Young et al.)
    - ▸ □ P. vivax Sal-1 Intraerythrocytic developmental cycle of three isolates (Bozdech et al.)
    - ▸ □ P. berghei ANKA AP2-G2 knock out and WT expression profiles (Yuda et al.)

3D7 iRBC RNASeq - fpkm Graph

IDC 48 hr Marray - Expr Graph

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 Step 2 : P.falciparum Erythrocytic expression time series (3D7, DD2, HB3) Microarray (fold change)

For the **Experiment** iRBC HB3 (48 Hour scaled) ▾

return protein coding ▾ **Genes**

that are up-regulated ▾

with a **Fold change** >= 2

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

in the following **Reference Samples**

28 selected, out of 46

Filter list below...

- ☑ 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**

18 selected, out of 46

Filter list below...

- □ 1-16 Hours
- □ 17-30 Hours
- ☑ 31-48 Hours
  - ☑ 31-39 Hours
  - ☑ 40-48 Hours

select all | clear all | expand all | collapse all

Example showing one gene that would meet search criteria

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

Up-regulated
Expression
2 fold
Average Expression Level Comparison
Average Expression Level Reference
Reference Samples
Comparison 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:

$$\textit{fold change} = \frac{\textit{average expression level in comparison}}{\textit{average expression level in reference}}$$

and returns genes when **fold change** >= **2**.

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

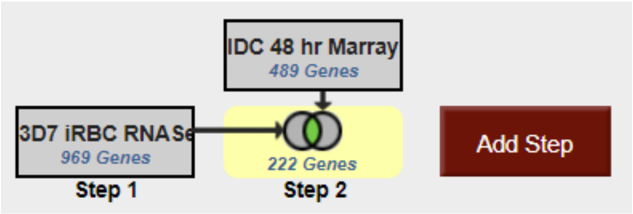
See the detailed help for this search.

Combine Genes in Step 1 with Genes in Step 2:

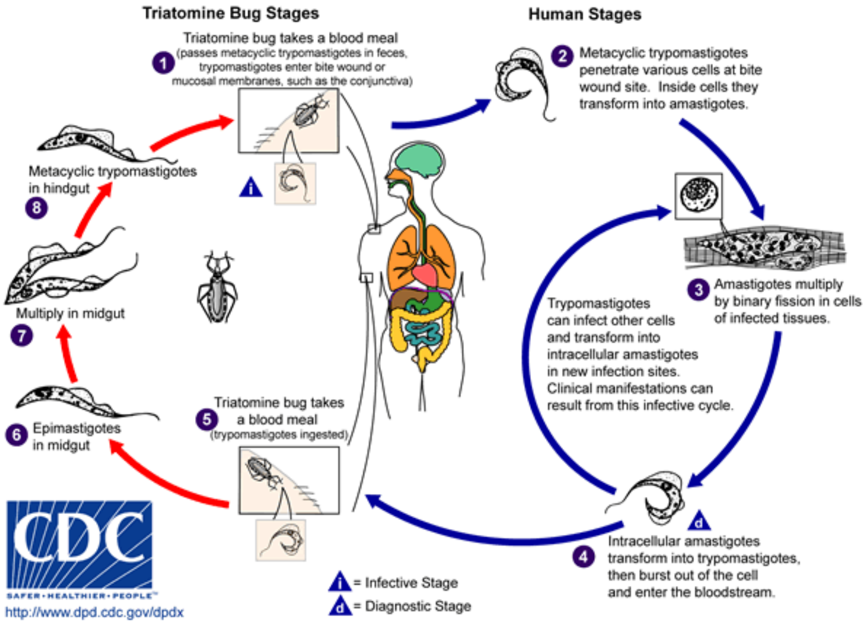
● 1 **Intersect** 2 ○ 1 **Minus** 2

○ 1 **Union** 2 ○ 2 **Minus** 1

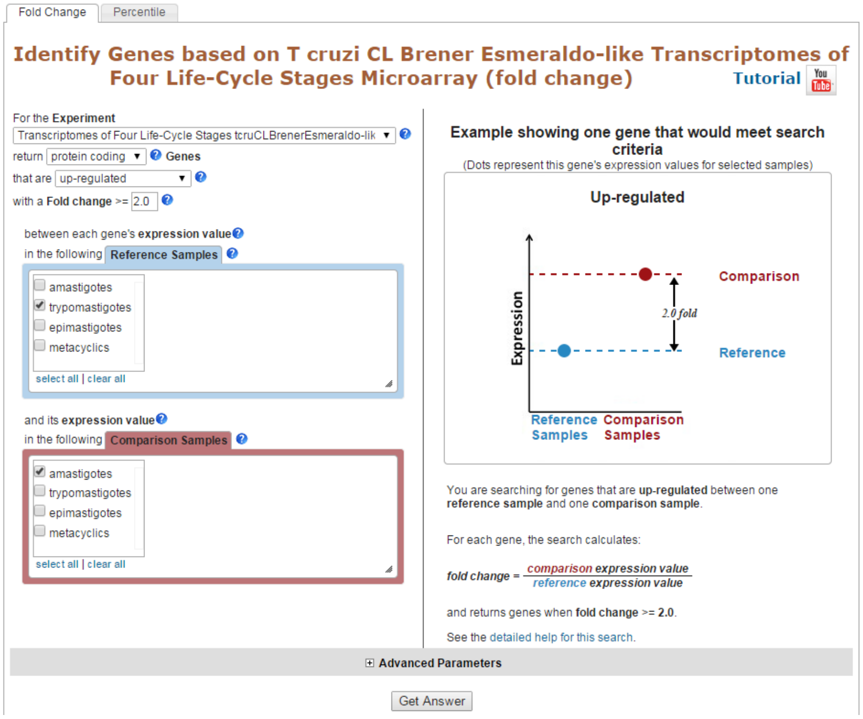
○ 1 **Relative to** 2 , using genomic colocation



2. *Optional (come back if time).* Exploring microarray data in TriTrypDB.
   Note: For this exercise use http://www.tritrypdb.org


Triatomine Bug Stages
Human Stages
1 Triatomine bug takes a blood meal (passes metacyclic trypomastigotes in feces, trypomastigotes enter bite wound or mucosal membranes, such as the conjunctiva)
2 Metacyclic trypomastigotes penetrate various cells at bite wound site. Inside cells they transform into amastigotes.
3 Amastigotes multiply by binary fission in cells of infected tissues.
Trypomastigotes can infect other cells and transform into intracellular amastigotes in new infection sites. Clinical manifestations can result from this infective cycle.
4 Intracellular amastigotes transform into trypomastigotes, then burst out of the cell and enter the bloodstream.
5 Triatomine bug takes a blood meal (trypomastigotes ingested)
6 Epimastigotes in midgut
7 Multiply in midgut
8 Metacyclic trypomastigotes in hindgut
i
d
i = Infective Stage
d = Diagnostic Stage
CDC
SAFER·HEALTHIER·PEOPLE™
http://www.dpd.cdc.gov/dpdx

a. 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.)**.



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 tcruCLBrenerEsmeraldo-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:

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

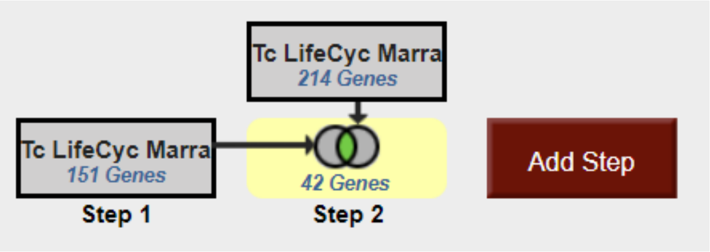
and returns genes when **fold change** >= **2.0**.

See the detailed help for this search.

- Select the direction of regulation, your reference sample and your comparison sample. For the fold change keep the default value 2.

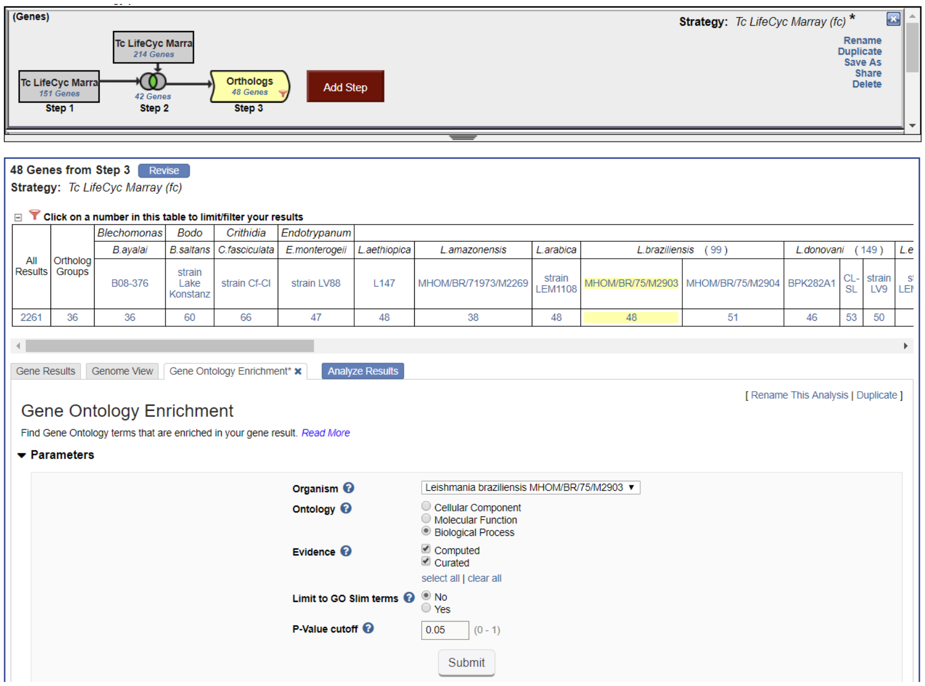
- How many genes did you find? Do the results seem plausible?

- 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? (*Hint*: add a step and transform your results into orthologs in all other organisms in TriTrypDB (select all for the ortholog transform).

  How many orthologs exist in *L. braziliensis* MHOM/BR/75/M2903? (*Hint*: look at the filter table between the strategy panel and your result list. Click on the number in the table under a species to view results from a specific species). Explore your results. Scan the product descriptions for this list of genes. Did you find anything interesting? Perhaps a GO enrichment analysis would support your ideas.



Strategy: *Tc LifeCyc Marray (fc)* *

48 Genes from Step 3 Revise

Strategy: *Tc LifeCyc Marray (fc)*

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

		Blechomonas	*Bodo*	*Crithidia*	*Endotrypanum*									
All Results	Ortholog Groups	*B.ayalai*	*B.saltans*	*C.fasciculata*	*E.monterogeii*	*L.aethiopica*	*L.amazonensis*	*L.arabica*	*L.braziliensis* (99)		*L.donovani* (149)			*L.e*
		B08-376	strain Lake Konstanz	strain Cf-Cl	strain LV88	L147	MHOM/BR/71973/M2269	strain LEM1108	MHOM/BR/75/M2903	MHOM/BR/75/M2904	BPK282A1	CL-SL	strain LV9	st LEI
2261	36	36	60	66	47	48	38	48	48	51	46	53	50	

Gene Results | Genome View | Gene Ontology Enrichment* ✖ | Analyze Results

[Rename This Analysis | Duplicate]

Gene Ontology Enrichment

Find Gene Ontology terms that are enriched in your gene result. *Read More*

Parameters

Organism ? Leishmania braziliensis MHOM/BR/75/M2903 ▼

Ontology ?
○ Cellular Component
○ Molecular Function
● Biological Process

Evidence ?
☑ Computed
☑ Curated
select all | clear all

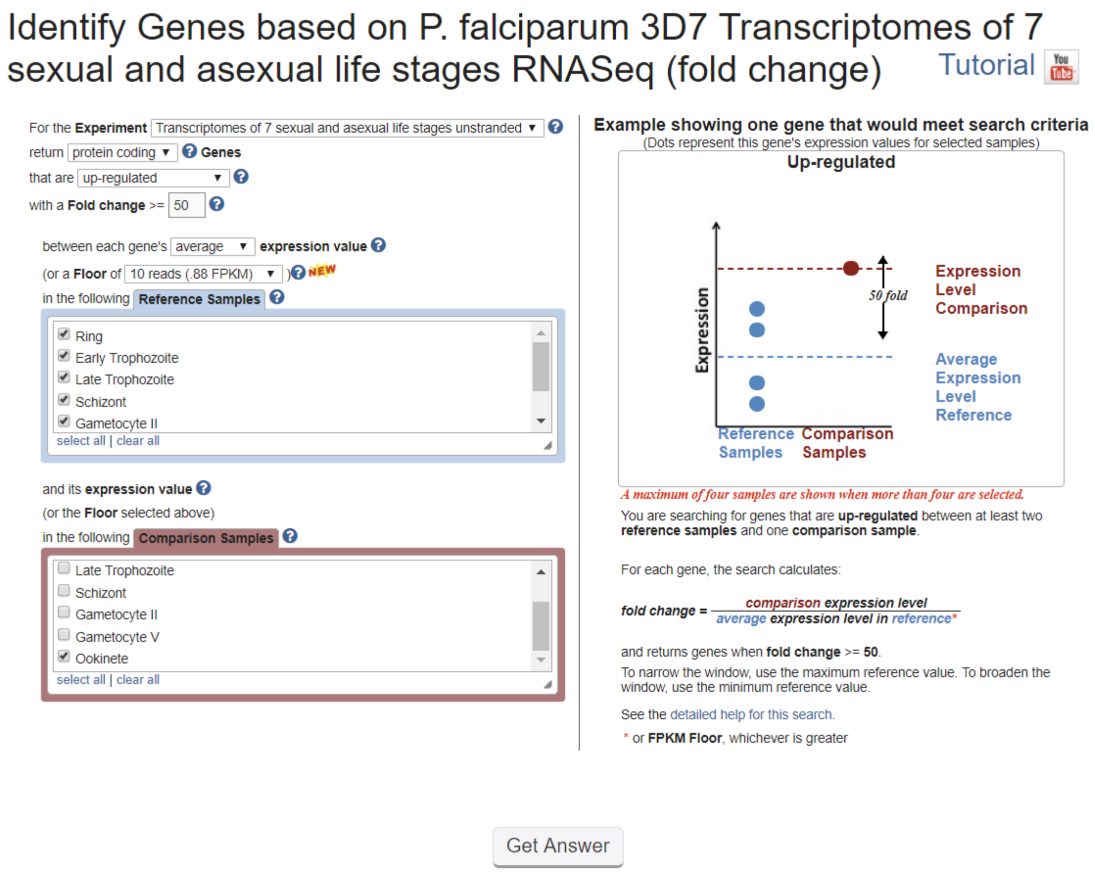
Limit to GO Slim terms ?
● No
○ Yes

P-Value cutoff ? 0.05 (0 - 1)

Submit

3. Finding genes based on RNAseq evidence and inferring function of hypothetical genes.
   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)

For the **Experiment** Transcriptomes of 7 sexual and asexual life stages unstranded ▾
return protein coding ▾ **Genes**
that are up-regulated ▾
with a **Fold change** >= 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:

fold change = *comparison expression level* / *average expression level in reference**

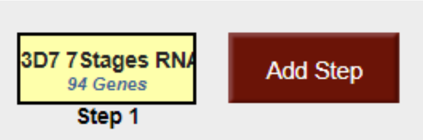
and returns genes when **fold change** >= **50**.

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

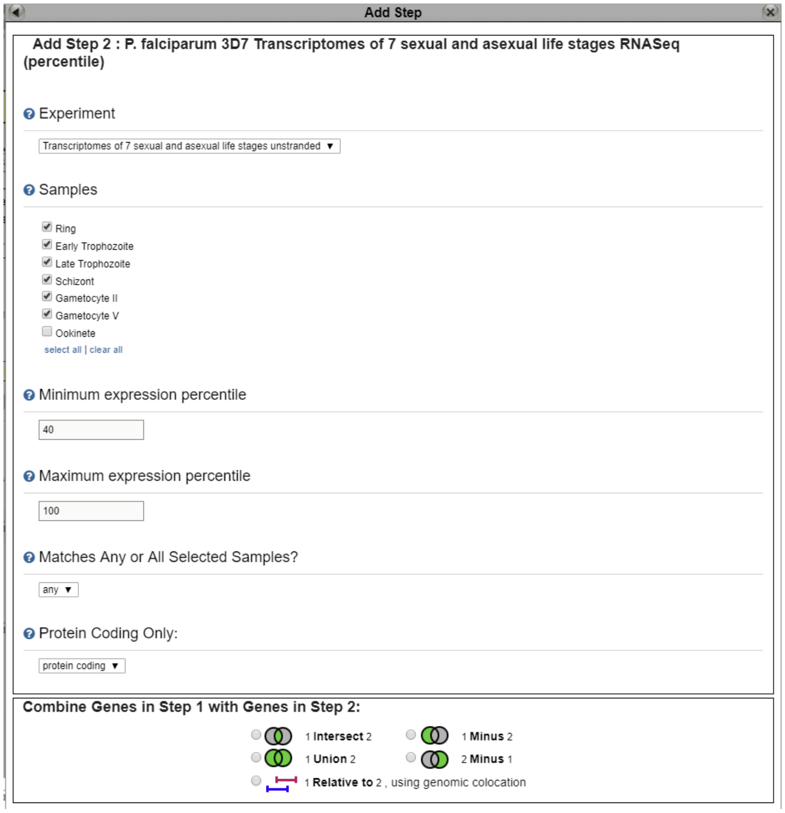
See the detailed help for this search.

* or **FPKM Floor**, whichever is greater

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?



- *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 Step

Add Step 2 : P. falciparum 3D7 Transcriptomes of 7 sexual and asexual life stages RNASeq (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 ▼

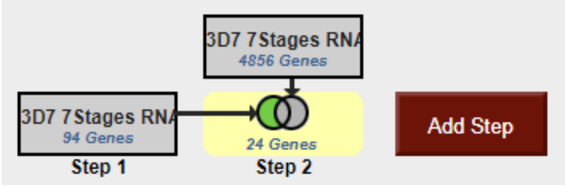
Protein Coding Only:

protein coding ▼

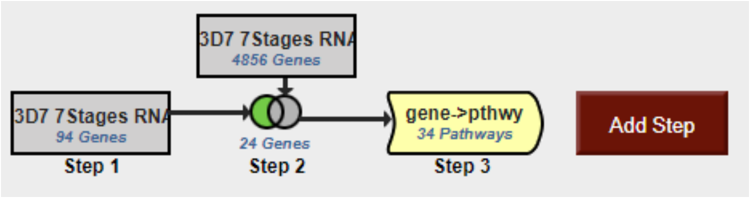
Combine Genes in Step 1 with Genes in Step 2:

○ 1 **Intersect** 2 ○ 1 **Minus** 2
○ 1 **Union** 2 ○ 2 **Minus** 1
○ 1 **Relative to** 2 , using genomic colocation

- *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?*

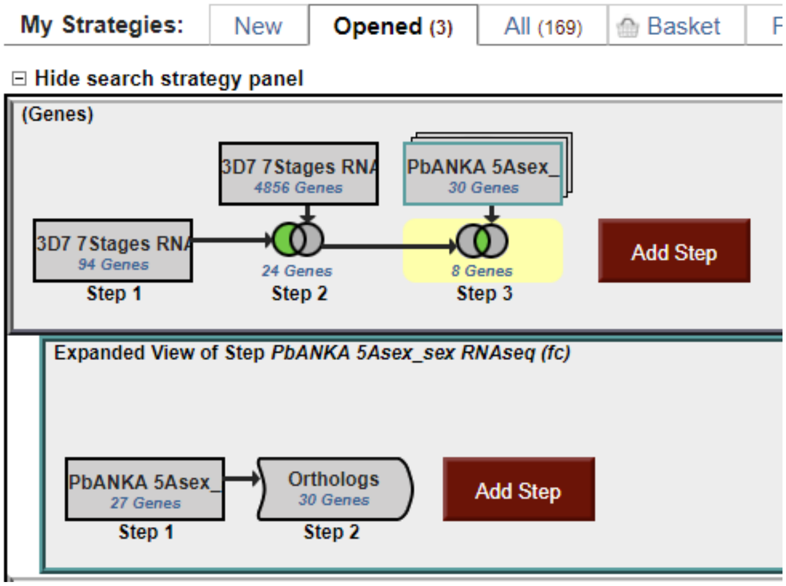


c. Which metabolic pathways are represented in this gene list? *Hint: add a step and transform results to pathways*. How does this result compare to running a pathways enrichment on step 2?



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 a 10 fold up regulation compared to the other samples?

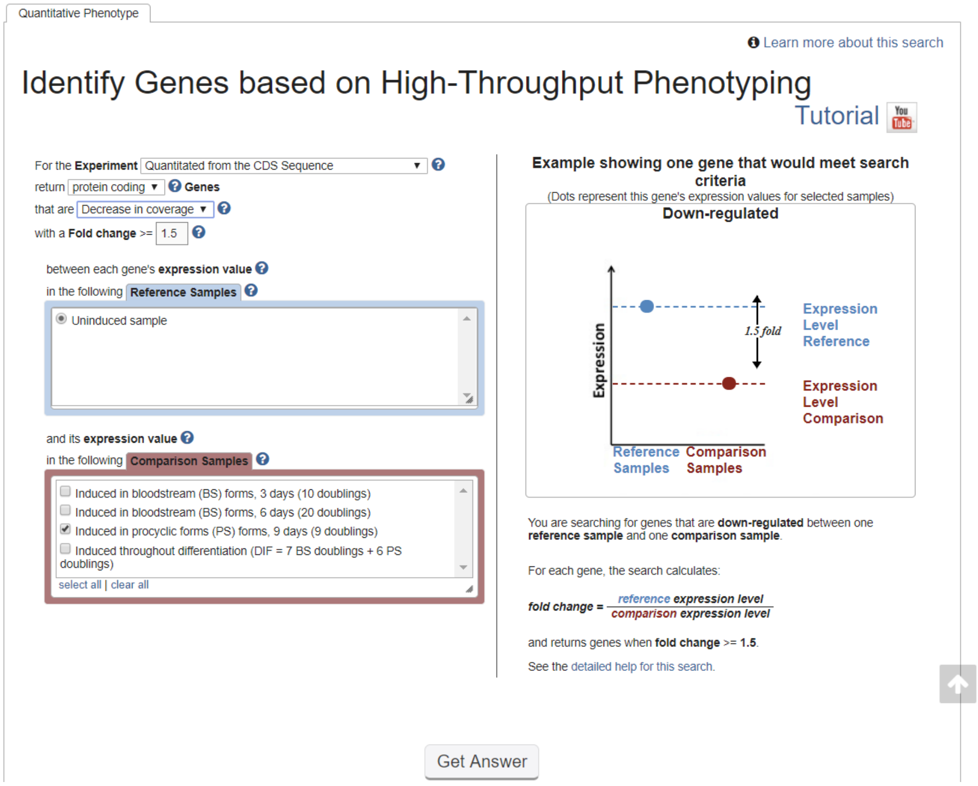
e. PlasmoDB also has an experiment examining gene expression during sexual development in *Plasmodium berghei* (rodent malaria). Can you determine if there are genes that are up-regulated in both human and rodent ookinetes (compared to all other stages)? *Hint:* start by deleting the last step you added 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". Why did you get 0 results? *Hint: we are comparing results from different organisms …. Click the edit link in either step and choose orthologs to transform to appropriate organism. Try it both ways … do you get the same number of genes? Why does the strategy make a nested strategy when you transform the last step and not when you transform the second to last step?*



High-throughput phenotyping searches.

4. **Find genes that are essential in procyclics but not in blood form *T. brucei*. Note that this search uses the same search form as RNASequencing searches but it is NOT RNASequencing. Read the search description at the bottom of the High Throughput Sequencing search page for information about this assay.**
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)

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

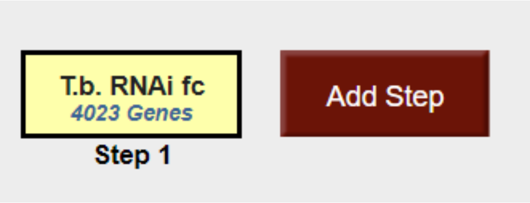
For each gene, the search calculates:

fold change = *reference expression level* / *comparison expression level*

and returns genes when **fold change** >= **1.5**.

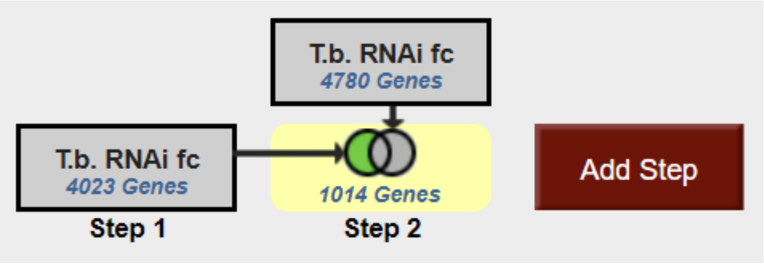
See the detailed help for this search.

Get Answer



- Next add a step and run the same search except this time select the “induced bloodstream form” samples.

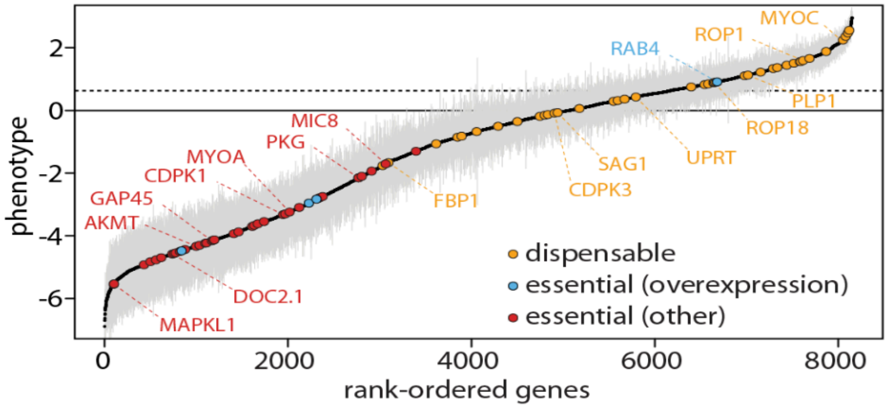
- How did you combine the results? Remember you want to find genes that are essential in procyclics and not in blood form.



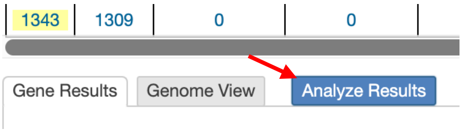
5. Finding genes based on high throughput mutagenesis and fitness analysis.

In EuPathDB we have a variety of studies where genome scale phenotypic analyses were carried out. In this exercise we'll use ToxoDB.org 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 > than -2 so setting the phenotype score <= -4 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?

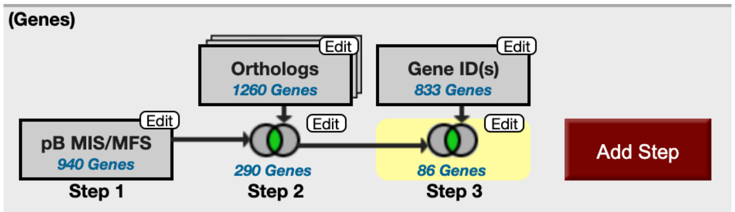


- Can you find additional evidence that these genes are essential? One way is to use the analysis tools to assess biological process and go function. Are the results what you would expect?

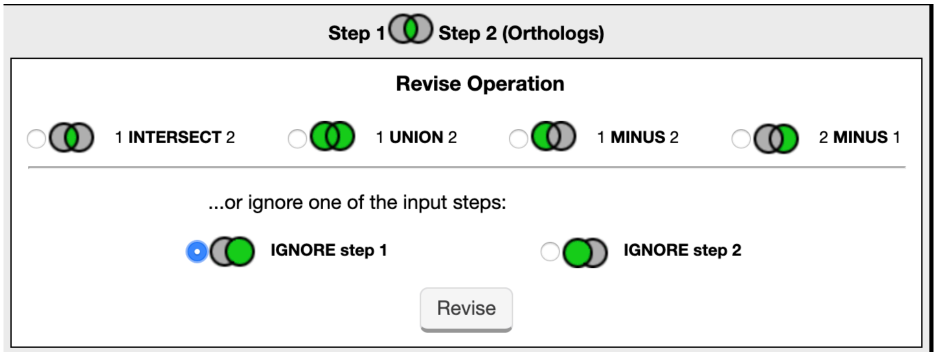


- Try adding columns to show additional data or intersecting these results with other queries, perhaps expression queries, to further assess this list. NOTE: this experiment was done in GT1 while all *T. gondii* functional data in ToxoDB is mapped to ME49 so an ortholog transform to ME49 is required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

- Optional, try intersecting your results with the results from the previous exercise or one of the experiments in PlasmoDB. NOTE that we don't make this easy ☺. Due to technical reasons, most of these searches are not available in EuPathDB and we can't currently generate orthologs between component websites. I accomplished this by re-running the CRISPR search in EuPathDB, doing an ortholog transform to *P. falciparum* 3D7 and downloading the results as a list of IDs. I then went to PlasmoDB and ran a strategy intersecting the results of the two high throughput assays (incorrectly called curated phenotype searches ☹). Then added an id search using the list I downloaded from the CRISPR results. The following strategy shows this effort (https://tinyurl.com/ThreeOrgPheno).



- click the edit link by the Boolean operator in step 2 and try excluding the first or second step and compare your results.

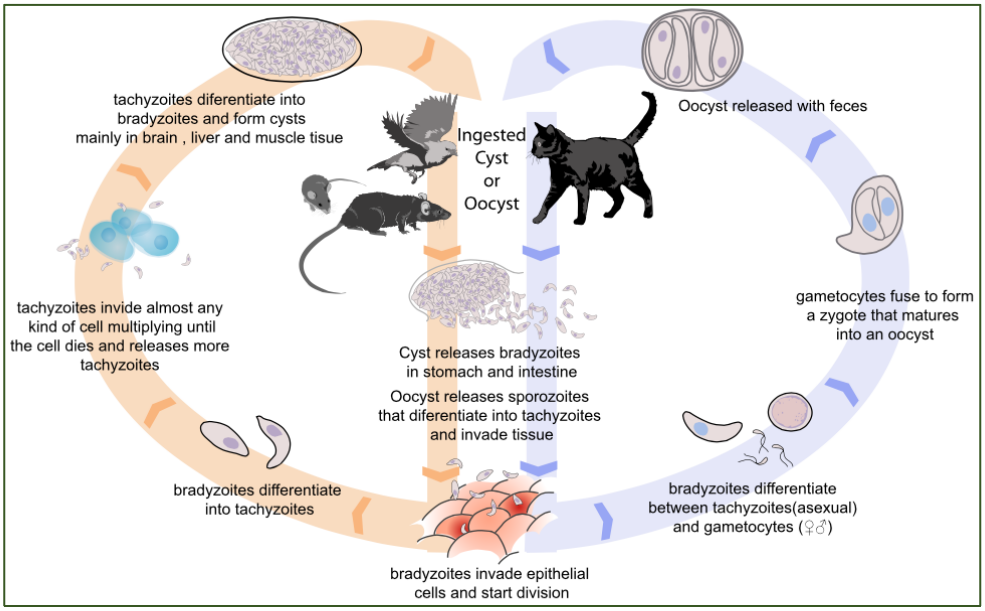


Try ignoring each of the first two steps. Which gives you the most results when the other is ignored? Can you think of plausible reasons for this difference?

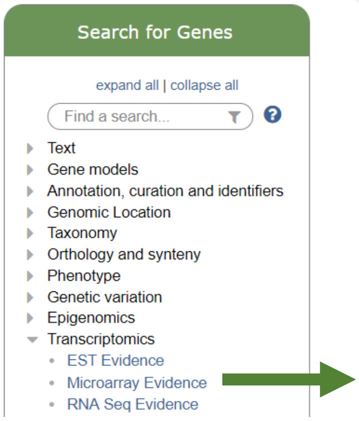
- You've now identified a core set of genes conserved across three Apicomplexan species that are likely to be essential. One criterion you might want in a parasite drug target would be that it not be conserved in humans as you wouldn't want the drug to impact the host. Are any of these genes also not conserved in humans (or more stringently mammals)? *Hint: add an Orthology Phylogenetic Profile search.*

6. *Optional (come back if time).* Finding oocyst expressed genes in *T. gondii* based on microarray evidence.

Note: For this exercise use http://toxodb.org


tachyzoites diferentiate into
bradyzoites and form cysts
mainly in brain , liver and muscle tissue
Ingested
Cyst
or
Oocyst
Oocyst released with feces
tachyzoites invide almost any
kind of cell multiplying until
the cell dies and releases more
tachyzoites
gametocytes fuse to form
a zygote that matures
into an oocyst
Cyst releases bradyzoites
in stomach and intestine
Oocyst releases sporozoites
that diferentiate into tachyzoites
and invade tissue
bradyzoites differentiate
into tachyzoites
bradyzoites differentiate
between tachyzoites(asexual)
and gametocytes (♀♂)
bradyzoites invade epithelial
cells and start division

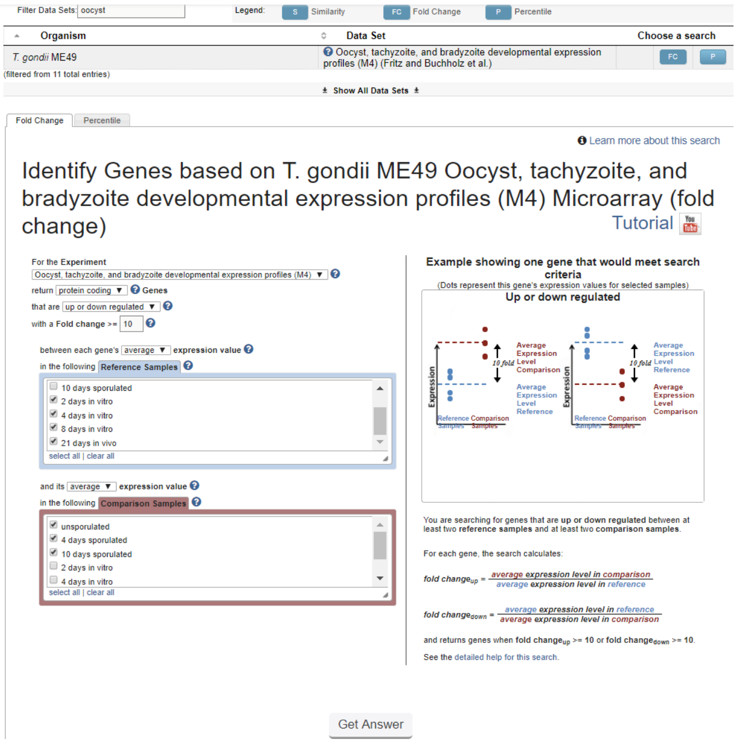
a. Find genes that are expressed at 10 fold higher levels in one of the oocyst stages than in any other stage in the "**Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4) (John Boothroyd)**" microarray experiment.


Search for Genes

expand all | collapse all

Find a search...

- Text
- Gene models
- Annotation, curation and identifiers
- Genomic Location
- Taxonomy
- Orthology and synteny
- Phenotype
- Genetic variation
- Epigenomics
- Transcriptomics
  - EST Evidence
  - Microarray Evidence
  - RNA Seq Evidence


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

Organism	Data Set	Choose a search
T. gondii ME49	Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4) (Fritz and Buchholz et al.)	FC P

(filtered from 11 total entries)

Show All Data Sets

Fold Change | Percentile

Learn more about this search

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

Tutorial

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

Expression
10 fold
Average Expression Level Comparison
Average Expression Level Reference
Reference Samples
Comparison Samples

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:

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

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

and returns genes when **fold change$_{up}$** >= **10** or **fold change$_{down}$** >= **10**.

See the detailed help for this search.

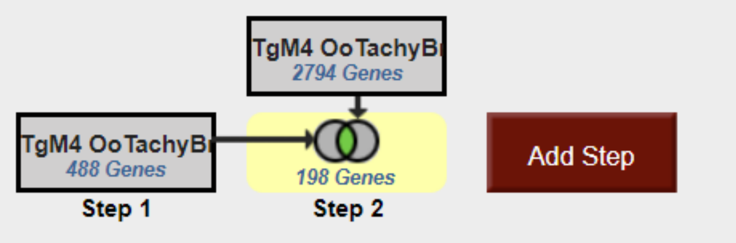
Get Answer

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 … ie 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).

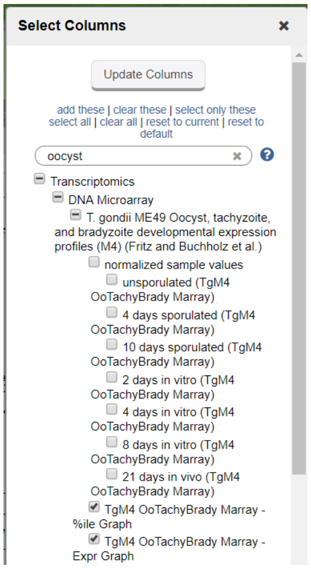
- Select the 4 **non-oocyst** samples.

- 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).

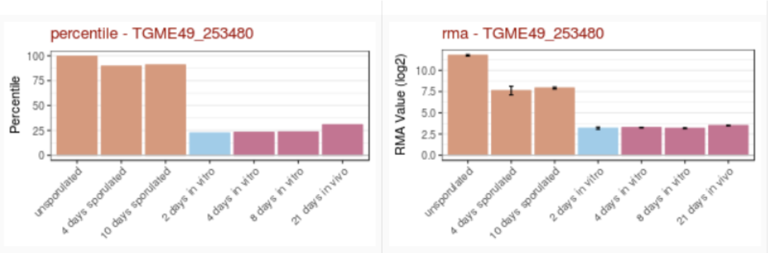

Select Columns ×

Update Columns

add these | clear these | select only these
select all | clear all | reset to current | reset to default

oocyst

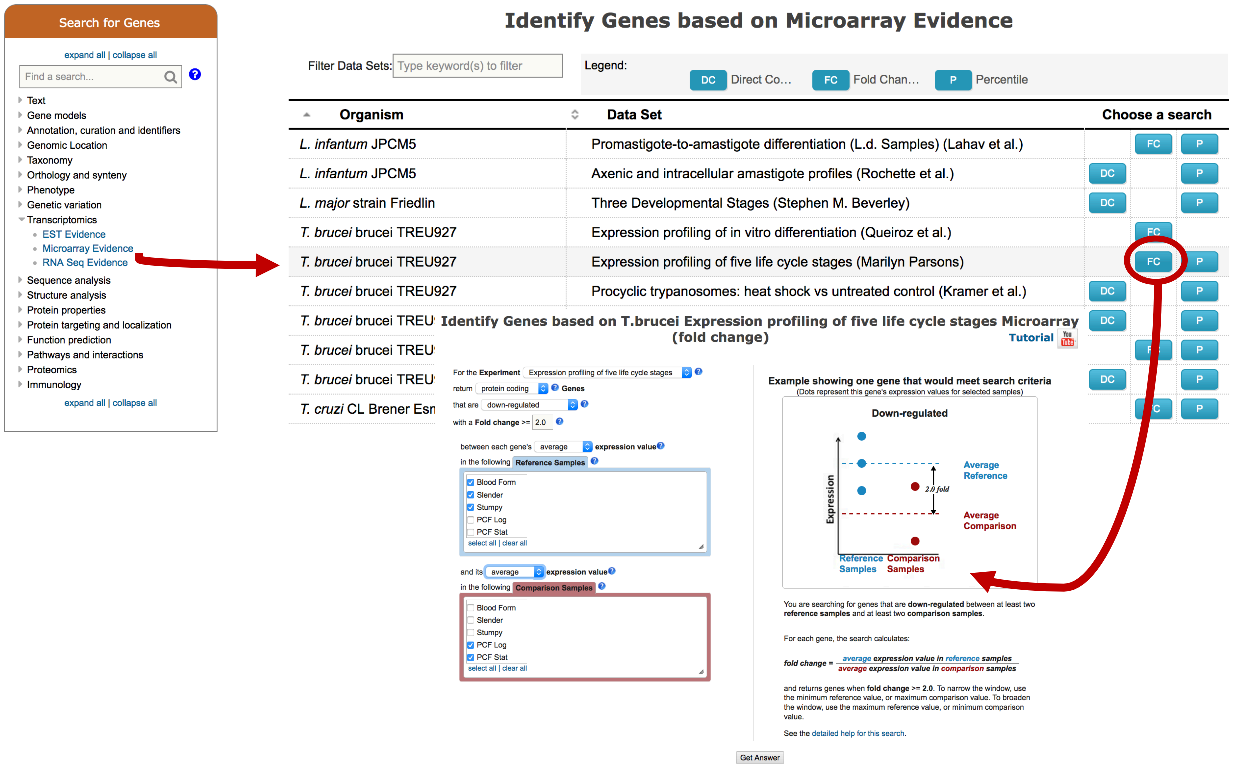
⊟ Transcriptomics
⊟ DNA Microarray
⊟ T. gondii ME49 Oocyst, tachyzoite, and bradyzoite developmental expression profiles (M4) (Fritz and Buchholz et al.)
□ normalized sample values
□ unsporulated (TgM4 OoTachyBrady Marray)
□ 4 days sporulated (TgM4 OoTachyBrady Marray)
□ 10 days sporulated (TgM4 OoTachyBrady Marray)
□ 2 days in vitro (TgM4 OoTachyBrady Marray)
□ 4 days in vitro (TgM4 OoTachyBrady Marray)
□ 8 days in vitro (TgM4 OoTachyBrady Marray)
□ 21 days in vivo (TgM4 OoTachyBrady Marray)
☑ TgM4 OoTachyBrady Marray - %ile Graph
☑ TgM4 OoTachyBrady Marray - Expr Graph



7. 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.



Search for Genes

expand all | collapse all

Find a search...

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

expand all | collapse all

Identify Genes based on Microarray Evidence

Filter Data Sets: Type keyword(s) to filter

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

Organism	Data Set	Choose a search
L. infantum JPCM5	Promastigote-to-amastigote differentiation (L.d. Samples) (Lahav et al.)	FC P
L. infantum JPCM5	Axenic and intracellular amastigote profiles (Rochette et al.)	DC P
L. major strain Friedlin	Three Developmental Stages (Stephen M. Beverley)	DC P
T. brucei brucei TREU927	Expression profiling of in vitro differentiation (Queiroz et al.)	FC
T. brucei brucei TREU927	Expression profiling of five life cycle stages (Marilyn Parsons)	FC P
T. brucei brucei TREU927	Procyclic trypanosomes: heat shock vs untreated control (Kramer et al.)	DC P
T. brucei brucei TREU		DC P
T. brucei brucei TREU		F P
T. brucei brucei TREU		DC P
T. cruzi CL Brener Esn		C P

Identify Genes based on T.brucei Expression profiling of five life cycle stages Microarray (fold change)

Tutorial

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
select all | clear all

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

☐ Blood Form
☐ Slender
☐ Stumpy
☑ PCF Log
☑ PCF Stat
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 **down-regulated** between at least two **reference samples** and at least two **comparison samples**.

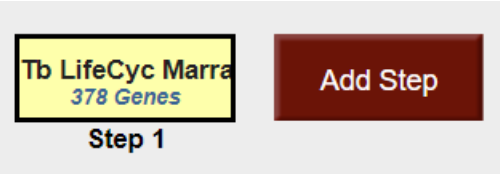
For each gene, the search calculates:

fold change = *average* ***expression value in*** *reference* ***samples*** / *average* ***expression value in*** *comparison* ***samples***

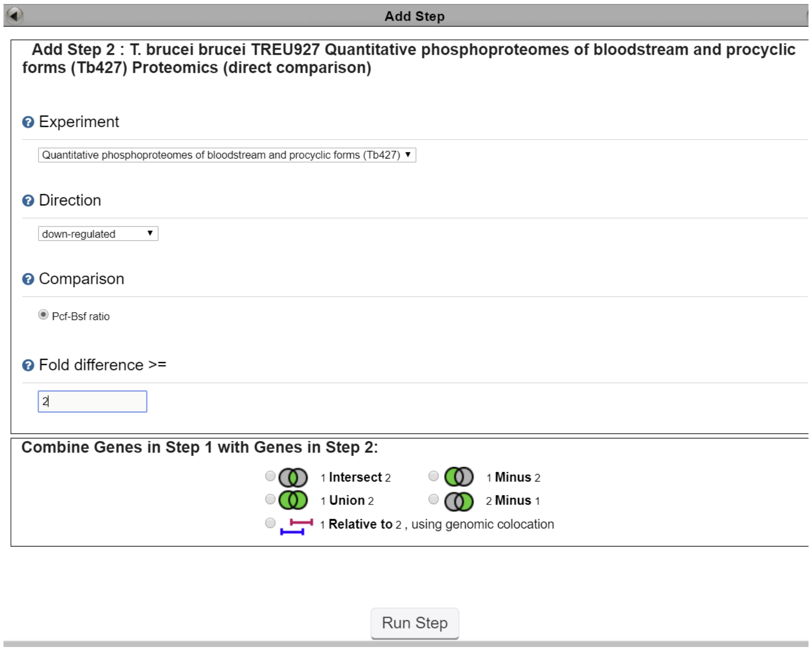
and returns genes when **fold change >= 2.0**. To narrow the window, use the minimum reference value, or maximum comparison value. To broaden the window, use the maximum reference value, or minimum comparison value.

See the detailed help for this search.

Get Answer



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 Step

Add Step 2 : T. brucei brucei TREU927 Quantitative phosphoproteomes of bloodstream and procyclic forms (Tb427) Proteomics (direct comparison)

Experiment

Quantitative phosphoproteomes of bloodstream and procyclic forms (Tb427) ▾

Direction

down-regulated ▾

Comparison

● Pcf-Bsf ratio

Fold difference >=

2

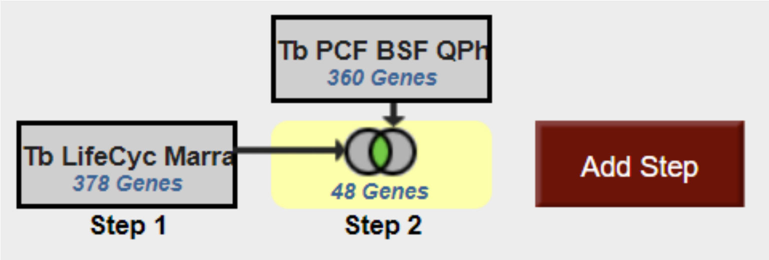
Combine Genes in Step 1 with Genes in Step 2:

○ 1 **Intersect** 2 ○ 1 **Minus** 2

○ 1 **Union** 2 ○ 2 **Minus** 1

○ 1 **Relative to** 2 , using genomic colocation

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



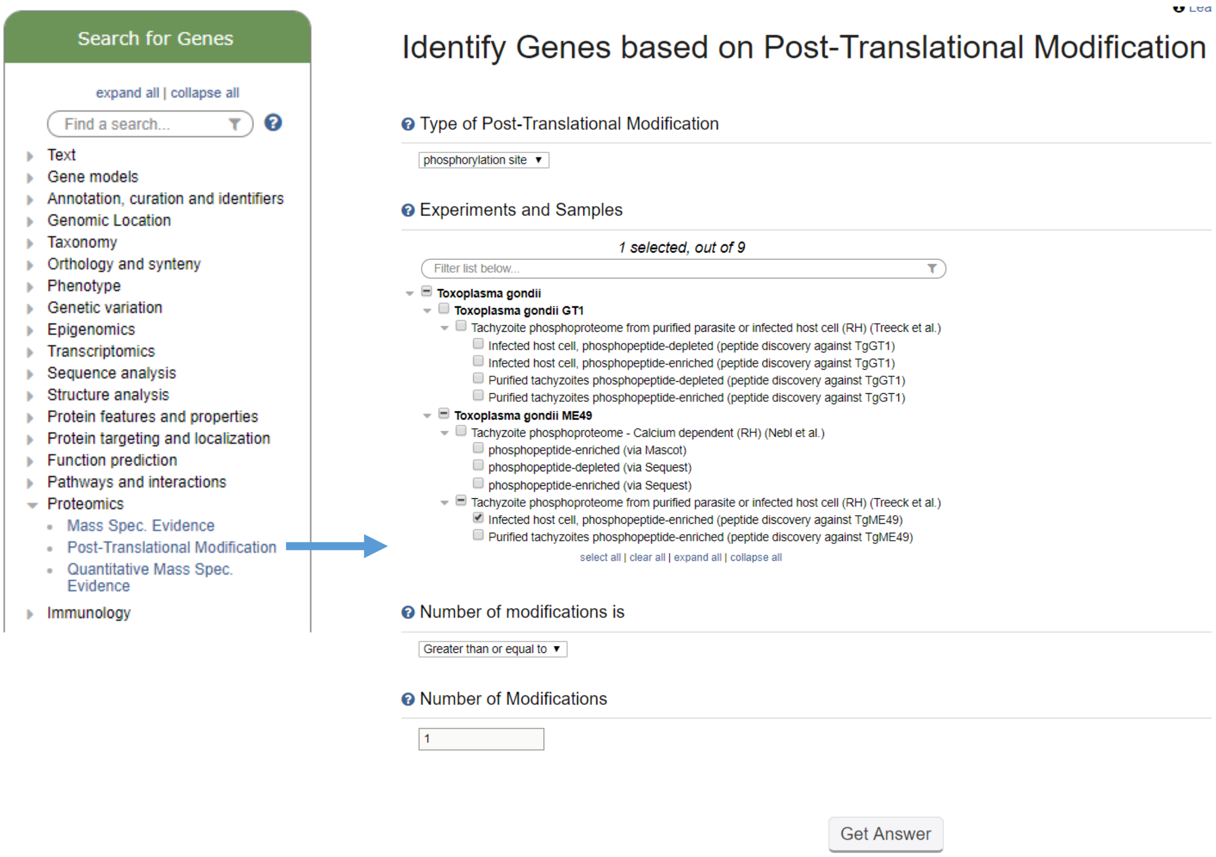
d. Try changing directions and compare up-regulated genes/proteins. (*Hint:* revise the existing strategy … you might want to duplicate it so you can keep both). When you change one of the steps but not the other do you have any genes in the intersection? Why might this be?

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).

8. **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.

8a. 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.)**"



Identify Genes based on Post-Translational Modification

Type of Post-Translational Modification

phosphorylation site ▼

Experiments and Samples

1 selected, out of 9

Filter list below...

- ⊟ **Toxoplasma gondii**
  - □ **Toxoplasma gondii GT1**
    - □ Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treeck et al.)
      - □ Infected host cell, phosphopeptide-depleted (peptide discovery against TgGT1)
      - □ Infected host cell, phosphopeptide-enriched (peptide discovery against TgGT1)
      - □ Purified tachyzoites phosphopeptide-depleted (peptide discovery against TgGT1)
      - □ Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgGT1)
  - ⊟ **Toxoplasma gondii ME49**
    - □ Tachyzoite phosphoproteome - Calcium dependent (RH) (Nebl et al.)
      - □ phosphopeptide-enriched (via Mascot)
      - □ phosphopeptide-depleted (via Sequest)
      - □ phosphopeptide-enriched (via Sequest)
    - ⊟ Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treeck et al.)
      - ☑ Infected host cell, phosphopeptide-enriched (peptide discovery against TgME49)
      - □ Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgME49)

Number of modifications is

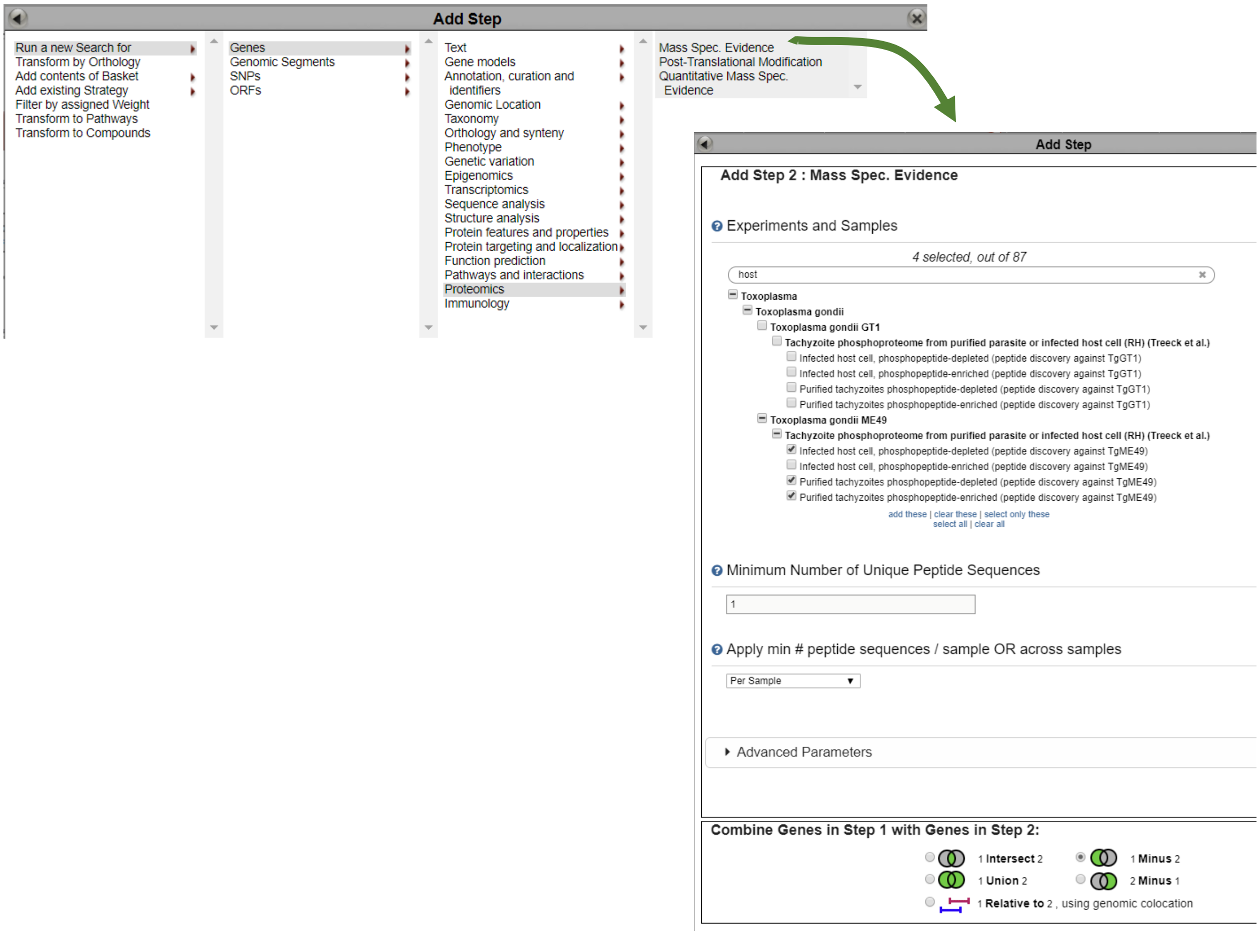
Greater than or equal to ▼

Number of Modifications

1

8b. 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 Step 2 : Mass Spec. Evidence

Experiments and Samples

4 selected, out of 87

host

- Toxoplasma
  - Toxoplasma gondii
    - □ Toxoplasma gondii GT1
      - □ Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treeck et al.)
        - □ Infected host cell, phosphopeptide-depleted (peptide discovery against TgGT1)
        - □ Infected host cell, phosphopeptide-enriched (peptide discovery against TgGT1)
        - □ Purified tachyzoites phosphopeptide-depleted (peptide discovery against TgGT1)
        - □ Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgGT1)
    - Toxoplasma gondii ME49
      - 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)

add these | clear these | select only these
select all | clear all

Minimum Number of Unique Peptide Sequences

1

Apply min # peptide sequences / sample OR across samples

Per Sample

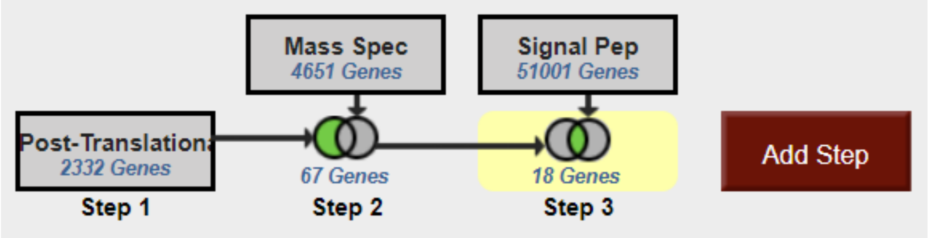
Advanced Parameters

Combine Genes in Step 1 with Genes in Step 2:

○ 1 **Intersect** 2 ● 1 **Minus** 2
○ 1 **Union** 2 ○ 2 **Minus** 1
○ 1 **Relative to** 2 , using genomic colocation

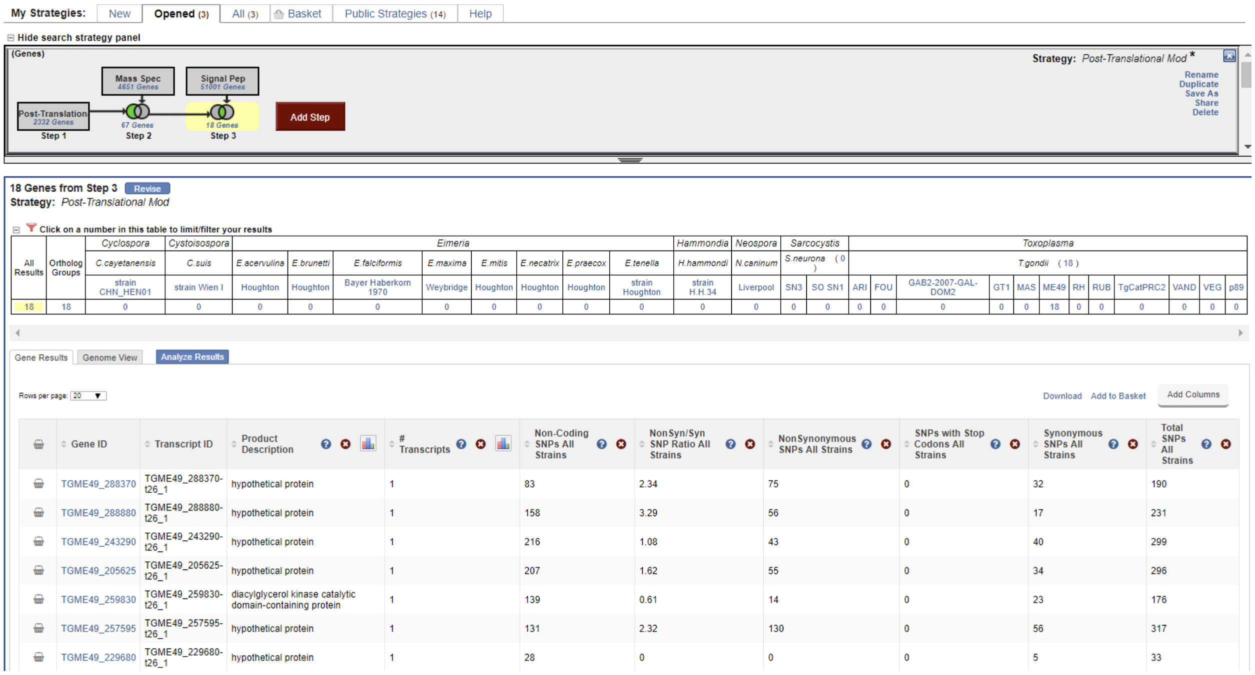
8d. Explore your results. What kinds of genes did you find? *Hint: use the Product description word column or perform a GO enrichment analysis of your results.*

8e. Are any of these genes likely to be secreted? Hint: add a step searching for genes with secretory signal peptides.



8f. 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.

8g. 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 highet number of non-synonymous SNPs. Hint: you can sort the columns by clicking on the up/down arrows next to the column names.



Strategy: *Post-Translational Mod* *

18 Genes from Step 3

Strategy: *Post-Translational Mod*

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

		Cyclospora	*Cystoisospora*	*Eimeria*								*Hammondia*	*Neospora*	*Sarcocystis*		*Toxoplasma*											
All Results	Ortholog Groups	*C.cayetanensis*	*C.suis*	*E.acervulina*	*E.brunetti*	*E.falciformis*	*E.maxima*	*E.mitis*	*E.necatrix*	*E.praecox*	*E.tenella*	*H.hammondi*	*N.caninum*	*S.neurona* (0)		*T.gondii* (18)											
		strain CHN_HEN01	strain Wien I	Houghton	Houghton	Bayer Haberkorn 1970	Weybridge	Houghton	Houghton	Houghton	strain Houghton	strain H.H.34	Liverpool	SN3	SO SN1	ARI	FOU	GAB2-2007-GAL-DOM2	GT1	MAS	ME49	RH	RUB	TgCatPRC2	VAND	VEG	p89
18	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0

Rows per page: 20

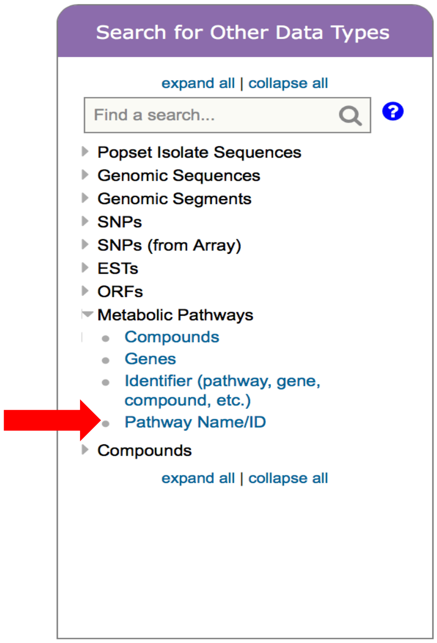
	Gene ID	Transcript ID	Product Description	# Transcripts	Non-Coding SNPs All Strains	NonSyn/Syn SNP Ratio All Strains	NonSynonymous SNPs All Strains	SNPs with Stop Codons All Strains	Synonymous SNPs All Strains	Total SNPs All Strains
	TGME49_288370	TGME49_288370-t26_1	hypothetical protein	1	83	2.34	75	0	32	190
	TGME49_288880	TGME49_288880-t26_1	hypothetical protein	1	158	3.29	56	0	17	231
	TGME49_243290	TGME49_243290-t26_1	hypothetical protein	1	216	1.08	43	0	40	299
	TGME49_205625	TGME49_205625-t26_1	hypothetical protein	1	207	1.62	55	0	34	296
	TGME49_259830	TGME49_259830-t26_1	diacylglycerol kinase catalytic domain-containing protein	1	139	0.61	14	0	23	176
	TGME49_257595	TGME49_257595-t26_1	hypothetical protein	1	131	2.32	130	0	56	317
	TGME49_229680	TGME49_229680-t26_1	hypothetical protein	1	28	0	0	0	5	33

9. Find and explore the metabolic pathway for glycolysis. For this exercise use http://plasmodb.org

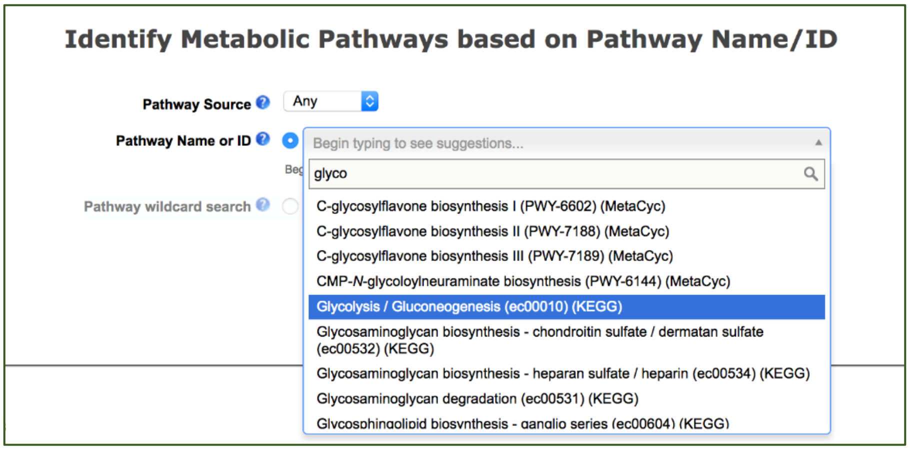
Navigate to the search page for Identify Metabolic Pathways based on Pathway Name/ID.

– Metabolic pathway and compound searches are available under the "Identify Other Data Types" head on the home page. You can find metabolic pathways based on the pathway name, genes involved in the pathway, or compounds involved in the pathway. Search for the **glycolysis** pathway using the Pathway Name/ID option.

– This search is equipped with a type-ahead function for choosing the metabolic pathway name. Begin typing glycolysis and then choose the pathway name from the list that appears.



a. Examine the Glycolysis / Gluconeogenesis pathway.



Identify Metabolic Pathways based on Pathway Name/ID

Pathway Source: Any

Pathway Name or ID: ● Begin typing to see suggestions...

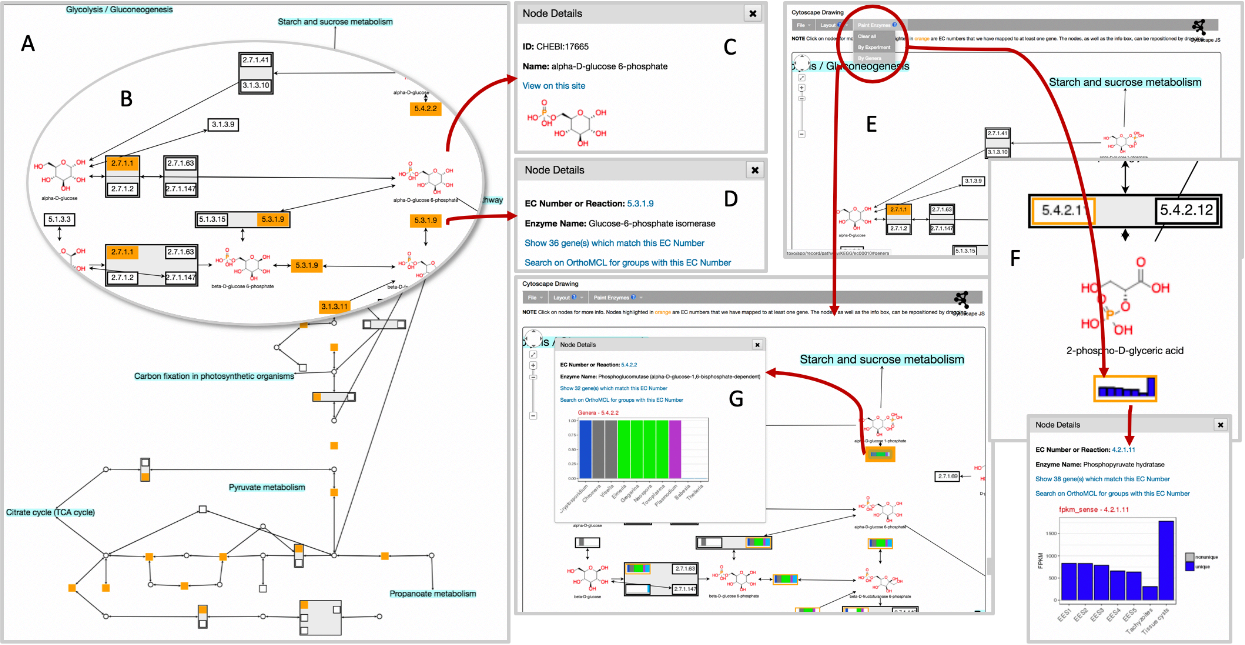
glyco

- C-glycosylflavone biosynthesis I (PWY-6602) (MetaCyc)
- C-glycosylflavone biosynthesis II (PWY-7188) (MetaCyc)
- C-glycosylflavone biosynthesis III (PWY-7189) (MetaCyc)
- CMP-*N*-glycoloylneuraminate biosynthesis (PWY-6144) (MetaCyc)
- Glycolysis / Gluconeogenesis (ec00010) (KEGG)
- Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate (ec00532) (KEGG)
- Glycosaminoglycan biosynthesis - heparan sulfate / heparin (ec00534) (KEGG)
- Glycosaminoglycan degradation (ec00531) (KEGG)
- Glycosphingolipid biosynthesis - ganglio series (ec00604) (KEGG)

Pathway wildcard search ○

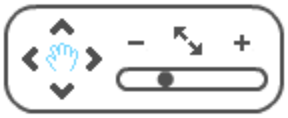
– The search takes you straight to the record page for the Glycolysis / Gluconeogenesis (ec00010) metabolic pathway from KEGG. The overview section of the record page contains an interactive graphical representation of the pathway. The pathway map and the legend can be repositioned.

A. Initial pathway view is zoomed out.
B. Zoom in to see more details including EC numbers and metabolite structures.
C. Click on a metabolite structure to get additional information.
D. Click on the EC number to get more info about the enzyme including links to retrieve all genes in the database assigned to this EC number.


A
B
C
D
E
F
G
Glycolysis / Gluconeogenesis
Starch and sucrose metabolism
Carbon fixation in photosynthetic organisms
Pyruvate metabolism
Citrate cycle (TCA cycle)
Propanoate metabolism
2.7.1.41
3.1.3.10
3.1.3.9
2.7.1.1
2.7.1.2
2.7.1.63
2.7.1.147
5.1.3.3
5.1.3.15
5.3.1.9
5.4.2.2
3.1.3.11
alpha-D-glucose
beta-D-glucose 6-phosphate
Node Details
ID: CHEBI:17665
Name: alpha-D-glucose 6-phosphate
View on this site
EC Number or Reaction: 5.3.1.9
Enzyme Name: Glucose-6-phosphate isomerase
Show 36 gene(s) which match this EC Number
Search on OrthoMCL for groups with this EC Number
Cytoscape Drawing
Paint Enzymes
Clear all
By Experiment
By Genera
EC Number or Reaction: 5.4.2.2
Genera - 5.4.2.2
5.4.2.1
5.4.2.12
2-phospho-D-glyceric acid
EC Number or Reaction: 4.2.1.11
Enzyme Name: Phosphopyruvate hydratase
Show 38 gene(s) which match this EC Number
Search on OrthoMCL for groups with this EC Number
fpkm_sense - 4.2.1.11
FPKM
EES1
EES2
EES3
EES4
EES5

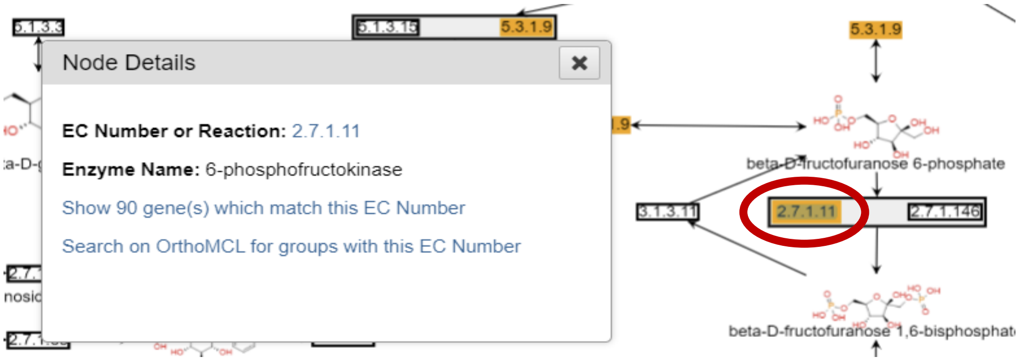
E. The drop-down menu under the heading “Paint Enzymes” allows you paint the pathway based on experiments or based on phyletic pattern.
F. Painting pathway by experiment provides a graphical representation of experimental results. Click on the graph to see more details.

G. Painting pathway based on phyletic pattern provides a graphical representation of phyletic distribution. Clicking on the phyletic pattern graphic provides additional information.



- Use the Tool Box to move (drag) the map and individual nodes. Zoom in and out to help explore the map.
- What do the rectangles with numbers like 2.7.1.11 represent?
- What is the difference between the rectangular nodes that are orange and those that are not?
- Why are some enzymes grouped?
- Find the node representing 6-phosphofructokinase (EC number = 2.7.1.11). You may need to zoom and reposition the map to find the node.

- Click on the 2.7.1.11 node to open a popup with information about this enzyme.



Node Details

EC Number or Reaction: 2.7.1.11

Enzyme Name: 6-phosphofructokinase

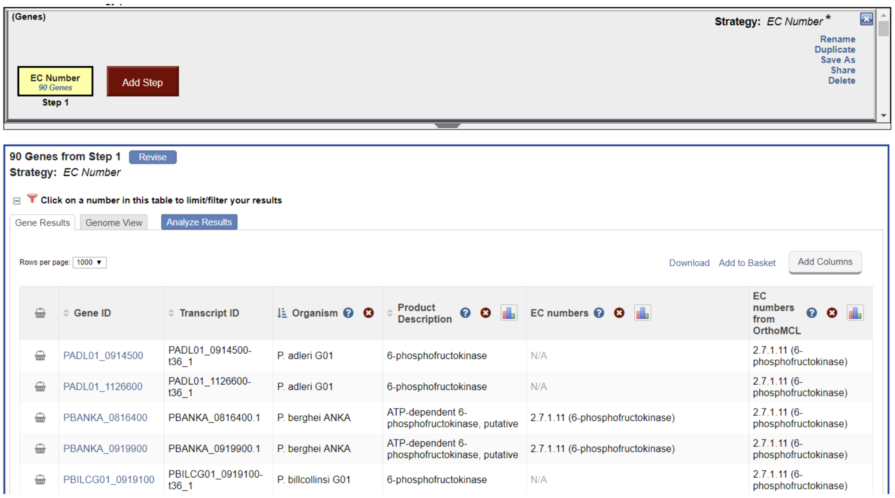
Show 90 gene(s) which match this EC Number

Search on OrthoMCL for groups with this EC Number

5.1.3.3
5.1.3.15
5.3.1.9
5.3.1.9
beta-D-fructofuranose 6-phosphate
3.1.3.11
2.7.1.11
2.7.1.146
beta-D-fructofuranose 1,6-bisphosphate

- How many genes in the database matched this EC number?

- Try the link ‘Search for Gene(s) by EC Number’. Where did you end up? What do the 90 genes in the result list represent? Is 6-phosphofructokinase unique to *P. falciparum*? Notice the two columns called “EC numbers” and “EC numbers from OrthoMCL”. What do these columns represent?



(Genes)

Strategy: *EC Number* *

90 Genes from Step 1 Revise

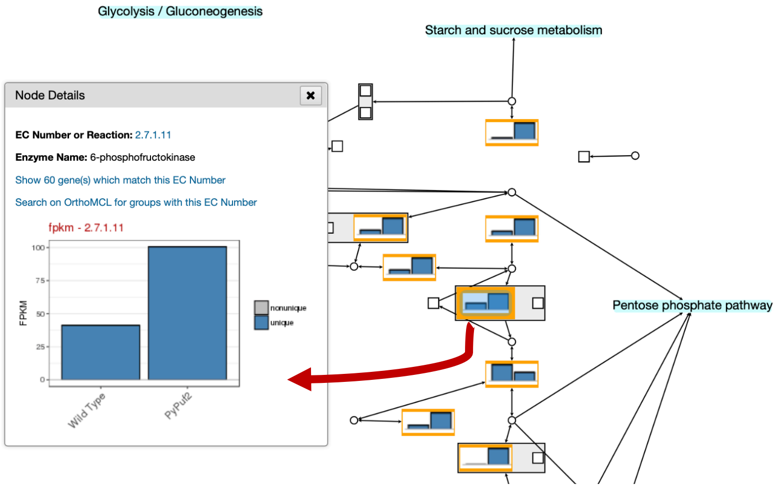
Strategy: *EC Number*

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

	Gene ID	Transcript ID	Organism	Product Description	EC numbers	EC numbers from OrthoMCL
	PADL01_0914500	PADL01_0914500-t36_1	P. adleri G01	6-phosphofructokinase	N/A	2.7.1.11 (6-phosphofructokinase)
	PADL01_1126600	PADL01_1126600-t36_1	P. adleri G01	6-phosphofructokinase	N/A	2.7.1.11 (6-phosphofructokinase)
	PBANKA_0816400	PBANKA_0816400.1	P. berghei ANKA	ATP-dependent 6-phosphofructokinase, putative	2.7.1.11 (6-phosphofructokinase)	2.7.1.11 (6-phosphofructokinase)
	PBANKA_0919900	PBANKA_0919900.1	P. berghei ANKA	ATP-dependent 6-phosphofructokinase, putative	2.7.1.11 (6-phosphofructokinase)	2.7.1.11 (6-phosphofructokinase)
	PBILCG01_0919100	PBILCG01_0919100-t36_1	P. billcollinsi G01	6-phosphofructokinase	N/A	2.7.1.11 (6-phosphofructokinase)

– Use your Browser's back button to return to the Glycolysis pathway record page and open the Paint Experiment menu. Choose the experiment "Salivary gland sporozoite transcriptomes: WT vs Pfu2-KO". Be patient while the graphs appear in place of the EC numbers.

- Does 6-phosphofructokinase appear to be expressed in salivary gland sporozoites? What enzymes in this pathway are affected in knockouts of Pfu2?



Node Details

EC Number or Reaction: 2.7.1.11

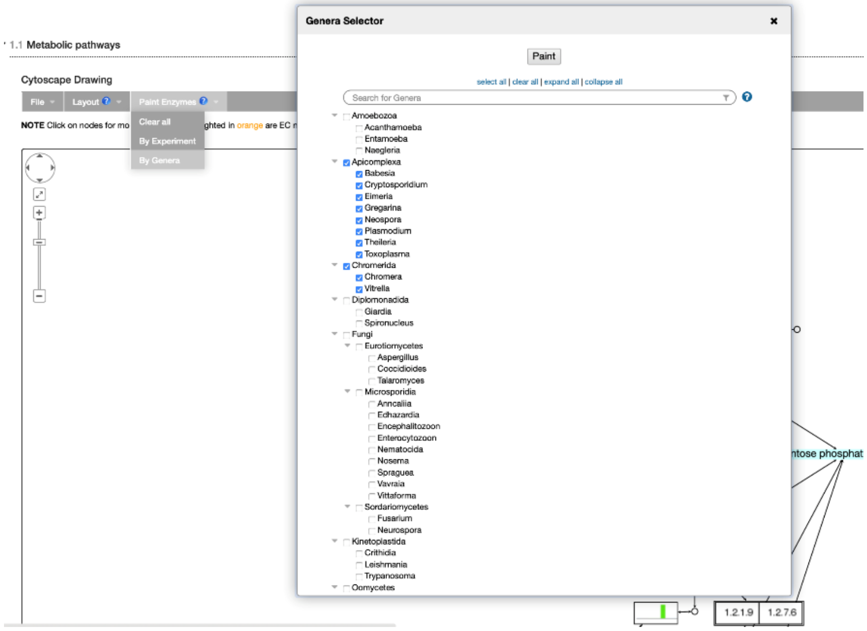
Enzyme Name: 6-phosphofructokinase

Show 60 gene(s) which match this EC Number

Search on OrthoMCL for groups with this EC Number

Glycolysis / Gluconeogenesis
Starch and sucrose metabolism
Pentose phosphate pathway
fpkm - 2.7.1.11
FPKM
100
75
50
25
0
Wild Type
PyPuf2
nonunique
unique

- Use the Paint Genera option to determine whether 6-phosphofructokinase has orthologs across Apicomplexa and Chromerida.



1.1 Metabolic pathways

Cytoscape Drawing

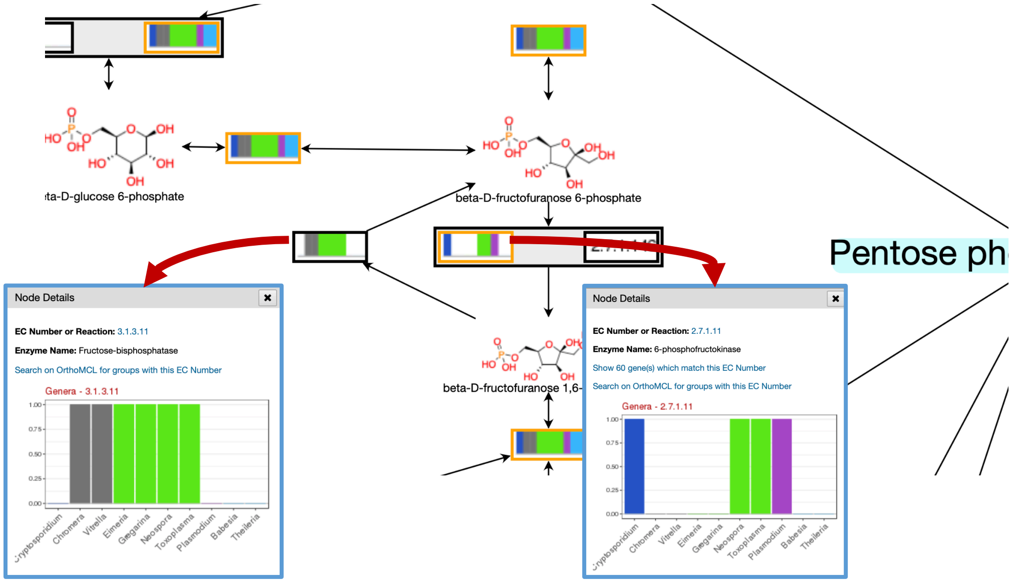
NOTE Click on nodes for mo … ghted in orange are EC n

Genera Selector

Search for Genera

- ☐ Amoebozoa
  - ☐ Acanthamoeba
  - ☐ Entamoeba
  - ☐ Naegleria
- ☑ Apicomplexa
  - ☑ Babesia
  - ☑ Cryptosporidium
  - ☑ Eimeria
  - ☑ Gregarina
  - ☑ Neospora
  - ☑ Plasmodium
  - ☑ Theileria
  - ☑ Toxoplasma
- ☑ Chromerida
  - ☑ Chromera
  - ☑ Vitrella
- ☐ Diplomonadida
  - ☐ Giardia
  - ☐ Spironucleus
- ☐ Fungi
  - ☐ Eurotiomycetes
    - ☐ Aspergillus
    - ☐ Coccidioides
    - ☐ Talaromyces
  - ☐ Microsporidia
    - ☐ Anncaliia
    - ☐ Edhazardia
    - ☐ Encephalitozoon
    - ☐ Enterocytozoon
    - ☐ Nematocida
    - ☐ Nosema
    - ☐ Spraguea
    - ☐ Vavraia
    - ☐ Vittaforma
  - ☐ Sordariomycetes
    - ☐ Fusarium
    - ☐ Neurospora
- ☐ Kinetoplastida
  - ☐ Crithidia
  - ☐ Leishmania
  - ☐ Trypanosoma
- ☐ Oomycetes

- What about the enzyme that catalyzes the reverse reaction (Fructose-bisphosphatase)?



Node Details

EC Number or Reaction: 3.1.3.11

Enzyme Name: Fructose-bisphosphatase

Search on OrthoMCL for groups with this EC Number

Node Details

EC Number or Reaction: 2.7.1.11

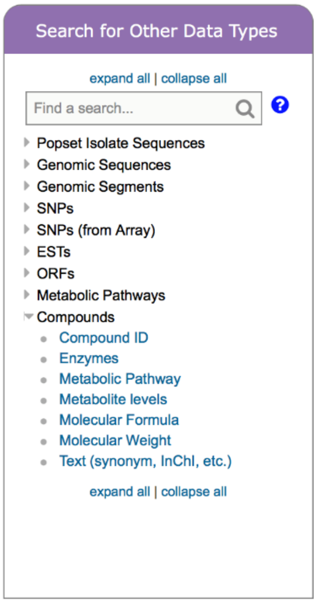
Enzyme Name: 6-phosphofructokinase

Show 60 gene(s) which match this EC Number

Search on OrthoMCL for groups with this EC Number

10. **Find and explore the compound record page for phosphoenolpyruvate (phosphoenolpyruvic acid or PEP).**

Compound records are accessed by running a compound search available under the "Identify Other Data Types" heading on the home page. For example, compounds may be retrieved by ID, text, metabolic pathway, molecular formula, molecular weight and metabolite levels. Compound records can also be accessed from the metabolic pathway legend after clicking on a compound (blue circle) in the map.



Search for Other Data Types

expand all | collapse all

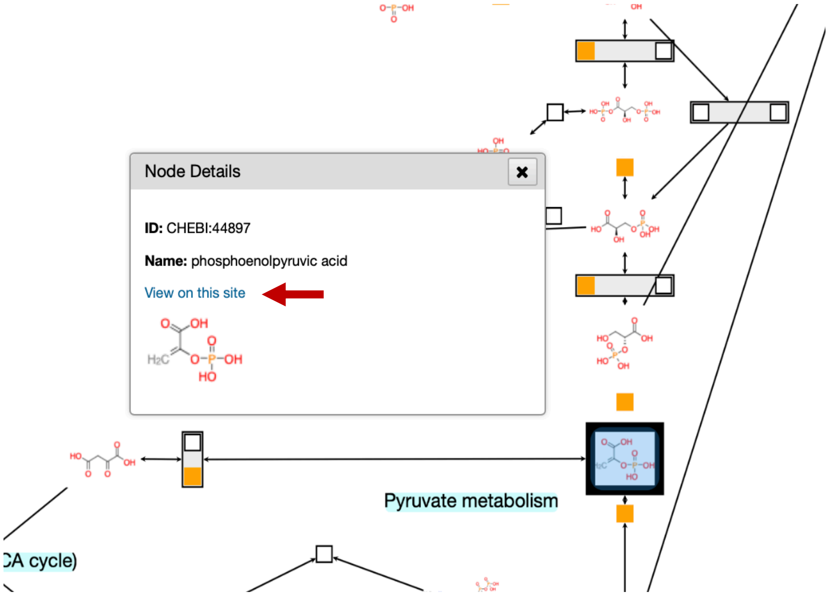
Find a search...

- Popset Isolate Sequences
- Genomic Sequences
- Genomic Segments
- SNPs
- SNPs (from Array)
- ESTs
- ORFs
- Metabolic Pathways
- Compounds
  - Compound ID
  - Enzymes
  - Metabolic Pathway
  - Metabolite levels
  - Molecular Formula
  - Molecular Weight
  - Text (synonym, InChI, etc.)

expand all | collapse all

- Choose one of these searches and retrieve the PEP record page.

- Alternatively, you can reach the PEP record page via a metabolic pathway where it is present as a substrate or a product of an enzymatic reaction (ie. glycolysis). Click on the node representing a compound



Node Details

ID: CHEBI:44897

Name: phosphoenolpyruvic acid

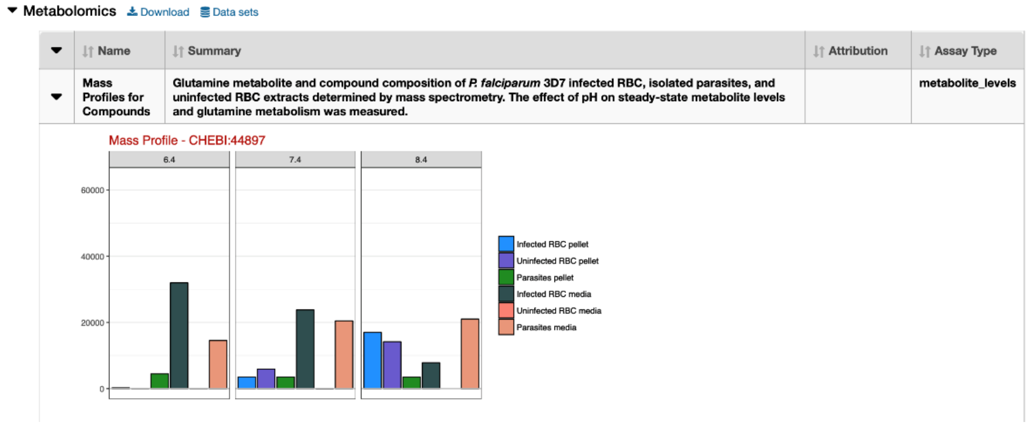
View on this site

- Which method did you use to get to the PEP record page? What compound name worked the best?

- Examine the PEP record page.

- What data sections do you see?

- Under which conditions is PEP present at highest concentrations? (Hint: navigate to the Metabolomics section)



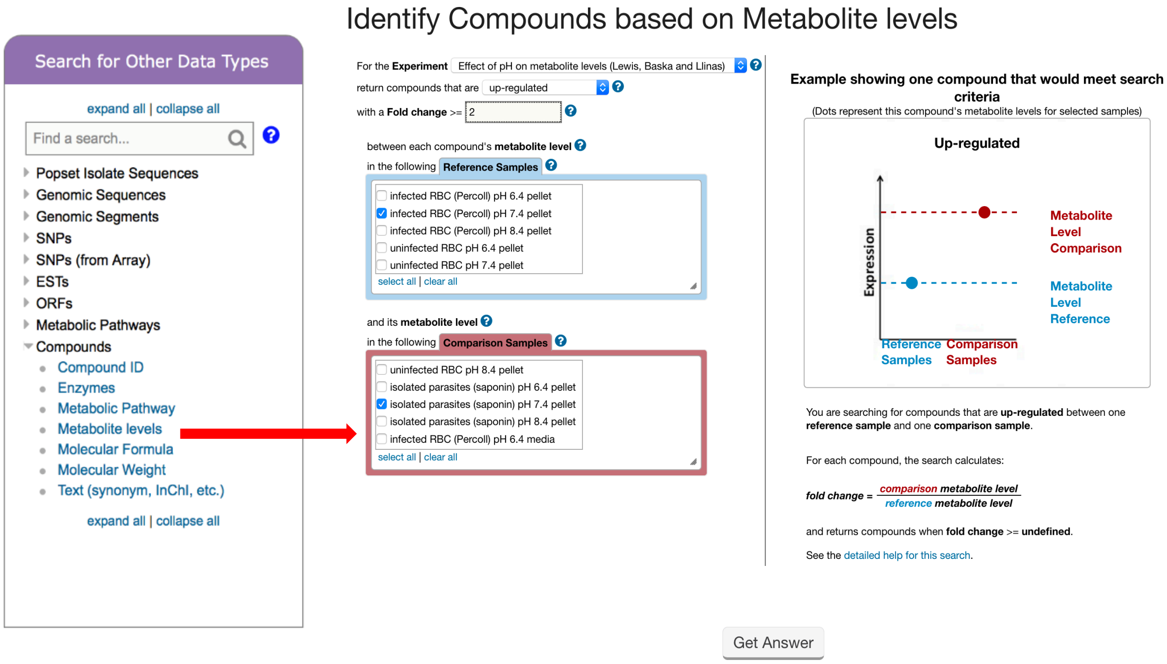
Metabolomics Download Data sets

	Name	Summary	Attribution	Assay Type
	Mass Profiles for Compounds	Glutamine metabolite and compound composition of *P. falciparum* 3D7 infected RBC, isolated parasites, and uninfected RBC extracts determined by mass spectrometry. The effect of pH on steady-state metabolite levels and glutamine metabolism was measured.		metabolite_levels

Mass Profile - CHEBI:44897

11. Identify metabolites (compounds) that are 20-fold enriched at pH7.4 in saponin lysed infected red blood cell (iRBCs) pellets compared the pH7.4 percoll pellet.

This requires running a metabolite levels search (2-fold enriched in saponin pellet compared to the percoll pellet as the reference).



Search for Other Data Types

expand all | collapse all

Find a search...

- Popset Isolate Sequences
- Genomic Sequences
- Genomic Segments
- SNPs
- SNPs (from Array)
- ESTs
- ORFs
- Metabolic Pathways
- Compounds
  - Compound ID
  - Enzymes
  - Metabolic Pathway
  - Metabolite levels
  - Molecular Formula
  - Molecular Weight
  - Text (synonym, InChI, etc.)

expand all | collapse all

Identify Compounds based on Metabolite levels

For the **Experiment** Effect of pH on metabolite levels (Lewis, Baska and Llinas)

return compounds that are up-regulated

with a **Fold change** >= 2

between each compound's **metabolite level**

in the following **Reference Samples**

- □ infected RBC (Percoll) pH 6.4 pellet
- ☑ infected RBC (Percoll) pH 7.4 pellet
- □ infected RBC (Percoll) pH 8.4 pellet
- □ uninfected RBC pH 6.4 pellet
- □ uninfected RBC pH 7.4 pellet

select all | clear all

and its **metabolite level**

in the following **Comparison Samples**

- □ uninfected RBC pH 8.4 pellet
- □ isolated parasites (saponin) pH 6.4 pellet
- ☑ isolated parasites (saponin) pH 7.4 pellet
- □ isolated parasites (saponin) pH 8.4 pellet
- □ infected RBC (Percoll) pH 6.4 media

select all | clear all

Example showing one compound that would meet search criteria

(Dots represent this compound's metabolite levels for selected samples)

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

For each compound, the search calculates:

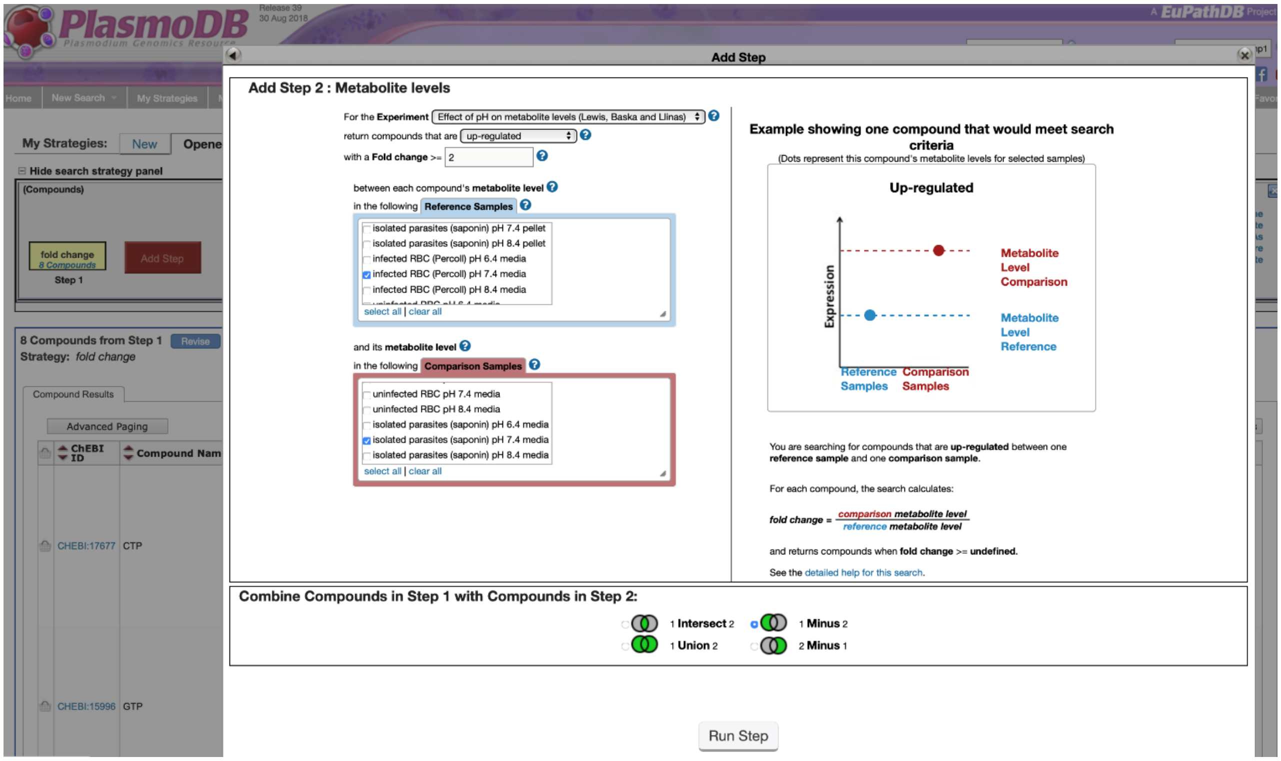
fold change = $\frac{\textit{comparison metabolite level}}{\textit{reference metabolite level}}$

and returns compounds when **fold change** >= **undefined**.

See the detailed help for this search.

Get Answer

- How many compounds did you get?
- How many of these compounds (metabolites) are NOT enriched by 2-fold in the pH7.4 saponin media fraction compared to the percoll media as reference?



PlasmoDB Plasmodium Genomics Resource Release 39 30 Aug 2018

Home | New Search | My Strategies

My Strategies: New | Opene

Hide search strategy panel

(Compounds)

fold change 8 Compounds Step 1 | Add Step

8 Compounds from Step 1 Revise

Strategy: *fold change*

Compound Results

Advanced Paging

ChEBI ID	Compound Nam
CHEBI:17677	CTP
CHEBI:15996	GTP

Add Step

Add Step 2 : Metabolite levels

For the **Experiment** Effect of pH on metabolite levels (Lewis, Baska and Llinas)

return compounds that are up-regulated

with a **Fold change** >= 2

between each compound's **metabolite level**

in the following **Reference Samples**

- □ isolated parasites (saponin) pH 7.4 pellet
- □ isolated parasites (saponin) pH 8.4 pellet
- □ infected RBC (Percoll) pH 6.4 media
- ☑ infected RBC (Percoll) pH 7.4 media
- □ infected RBC (Percoll) pH 8.4 media

select all | clear all

and its **metabolite level**

in the following **Comparison Samples**

- □ uninfected RBC pH 7.4 media
- □ uninfected RBC pH 8.4 media
- □ isolated parasites (saponin) pH 6.4 media
- ☑ isolated parasites (saponin) pH 7.4 media
- □ isolated parasites (saponin) pH 8.4 media

select all | clear all

Example showing one compound that would meet search criteria

(Dots represent this compound's metabolite levels for selected samples)

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

For each compound, the search calculates:

fold change = $\frac{\textit{comparison metabolite level}}{\textit{reference metabolite level}}$

and returns compounds when **fold change** >= **undefined**.

See the detailed help for this search.

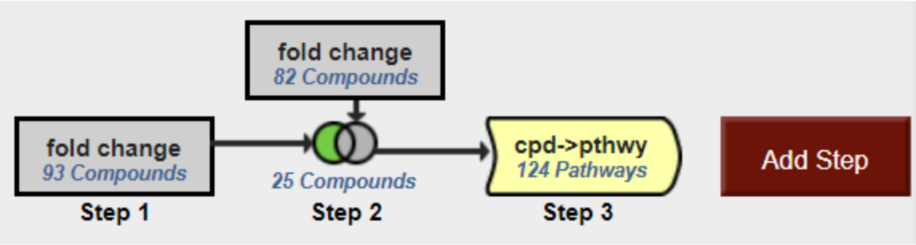
Combine Compounds in Step 1 with Compounds in Step 2:

○ 1 Intersect 2 ● 1 Minus 2

○ 1 Union 2 ○ 2 Minus 1

Run Step

To which metabolic pathways do these compounds belong? Click Add Step and transform the results to metabolic pathways.


fold change
82 Compounds
fold change
93 Compounds
Step 1
25 Compounds
Step 2
cpd->pthwy
124 Pathways
Step 3
Add Step