

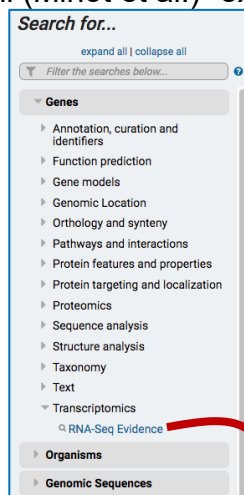
Host Response

Learning objectives:

- Exploring host responses by running a search strategy in HostDB
- Add steps in a search strategy
- Performing a GO enrichment analysis
- Revising steps in a search strategy

1. Find host genes that are upregulated in infected mouse cells compared to uninfected ones. For this exercise use <http://hostdb.org>

- a. HostDB has data from a published study that performed a comparative transcriptome analysis of 29 different strains of *Toxoplasma gondii* and the murine macrophages infected with them. We loaded the parasite component of the data in ToxoDB and the host component in HostDB. Go to HostDB.org and navigate to the “Transcriptomics” section then select “RNA Seq Evidence”. Select the fold change query for the “Mouse transcriptomes during infection with 29 strains of *T. gondii* (Minot et al.)” experiment.



Identify Genes based on RNA-Seq Evidence		
Filter Data Sets: <input type="text"/>		
Legend: <input checked="" type="checkbox"/> Similarity <input checked="" type="checkbox"/> Differential Expression <input checked="" type="checkbox"/> Fold Change <input checked="" type="checkbox"/> Percentile <input checked="" type="checkbox"/> SenseAntisense		
Organism	Data Set	Choose a Search
Bos taurus breed Hereford	Host cell transcriptome in bovine cells infected with Cryptosporidium parvum (Widmer et al.)	DE FC P
Bos taurus breed Hereford	Transcriptome of Bos taurus during infection with virulent and avirulent N. caninum strains (Hercjajo et al.)	DE FC P SA
Homo sapiens REF	HFF transcriptional response to virulent and avirulent T. cruzi (Belew et al. 2017)	DE FC P
Homo sapiens REF	Transcriptomes of 46 malaria-infected Gambian children (Lee et al.)	FC P SA
Homo sapiens REF	H sapiens Transcriptome during infection with T. cruzi (Li et al.)	DE FC P
Homo sapiens REF	Leishmania major and Leishmania amazonensis RNA-Seq during human macrophage infection (Fernandes et al.)	DE FC P
Macaca mulatta isolate 17573	M. mulatta infected with P. cynomolgi over 100 days (Joyner et al.)	FC P
Mus musculus C57BL/6J	Mouse transcriptome during infection with 29 strains of T. gondii (Minot et al.)	S FC P
Mus musculus C57BL/6J	Transcriptomes of 4 M. musculus cell types during infection with T. gondii (Swierzy et al.)	FC P
Mus musculus C57BL/6J	Mouse transcriptome during early and late chronic infection with T. gondii (Garfoot et al.)	DE FC P
Mus musculus C57BL/6J	Transcriptional analysis of sorted subpopulations of mouse macrophages infected with C. albicans (Munoz et al.)	DE FC P
Mus musculus C57BL/6J	Transcriptome of mouse bone marrow derived macrophages infected by Wild-Type and grn18 mutant strains of T. gondii (He et al.)	FC P
Mus musculus C57BL/6J	Transcriptomes of mouse macrophages infected with Leishmania mexicana (Fiebig et al.)	DE FC P

- b. Configure the search to return genes that are up-regulated at least 10-fold across all strains in the experiment compared to the uninfected control. Make sure to select upregulated. In the example below a fold change of 10 was selected and the “**average**” operation was applied on the comparison samples.

Identify Genes based on M. musculus C57BL6J Mouse transcriptome during infection with 29 strains of T. gondii RNA-Seq (fold change)

For the Experiment

Mouse transcriptome during infection with 29 strains of T. gondii unstranded

return protein coding Genes

that are up-regulated

with a Fold change ≥ 10

between each gene's average expression value (or a Floor of 10 reads)

in the following Reference Samples

- ☐ TgCATBr9 infected
- ☐ TgCATBr9 infected
- ☐ VAND infected
- ☐ VEG infected
- ☐ WTD3 infected
- ☒ Un-infected

select all | clear all

and its average expression value (or the Floor selected above)

in the following Comparison Samples

- ☒ TgCATBr9 infected
- ☒ TgCATBr9 infected
- ☒ VAND infected
- ☒ VEG infected
- ☒ WTD3 infected
- ☐ Un-infected

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

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

For each gene, the search calculates:

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

and returns genes when fold change ≥ 10 .

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

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

Get Answer

mouse infected w/ 29 Tg strain...
210 Genes

+ Add a step

Step 1

- c. What are the functional characteristics of the genes in this result? What kinds of GO terms are enriched? Does the host immune response appear to be turned on? Is there a particular cellular location that is common in this group of genes?
Hint: click on the “Analyze Results” tab and perform a GO enrichment analysis for the biological process ontology.

Gene Ontology Enrichment
Find Gene Ontology terms that are enriched in your gene result. [Read More](#)

Parameters

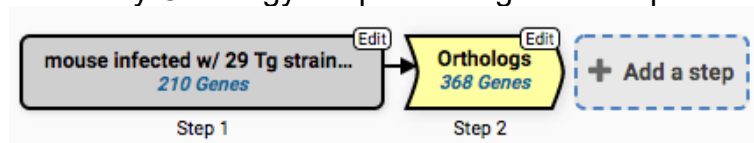
Analysis Results:

1797 rows

Open in Revigo Show Word Cloud Download

GO ID	GO Term	Genes in the bkgd with this term	Genes in your result with this term	Percent of bkgd genes in your result	Fold enrichment	Odds ratio	P-value	Benjamini	Bonferroni
GO:0051239	regulation of multicellular organismal process	1496	55	3.7	5.22	7.11	7.10e-25	2.29e-21	2.29e-21
GO:0071310	cellular response to organic substance	958	45	4.7	6.67	8.74	1.77e-24	2.84e-21	5.69e-21
GO:0070887	cellular response to chemical stimulus	1209	48	4.0	5.64	7.42	5.28e-23	5.66e-20	1.70e-19
GO:0048583	regulation of response to stimulus	2005	60	3.0	4.25	5.85	1.17e-22	9.45e-20	3.78e-19
GO:0002376	immune system process	1073	45	4.2	5.95	7.73	1.73e-22	1.08e-19	5.57e-19
GO:0050896	response to stimulus	6524	109	1.7	2.37	4.18	2.01e-22	1.08e-19	6.47e-19

- d. Expand the result set to include human orthologs/paralogs of these genes. *Hint:* add a “Transform by Orthology” step choosing Homo sapiens.



- e. Does this set of human genes also show enriched GO terms? What, if any, are the enriched GO terms?
- f. Do any of these human genes also have peptide evidence for their expression during infection? *Hint:* add a step and explore the proteomics data “Human Proteome During T. gondii infection”

Experiments and Samples

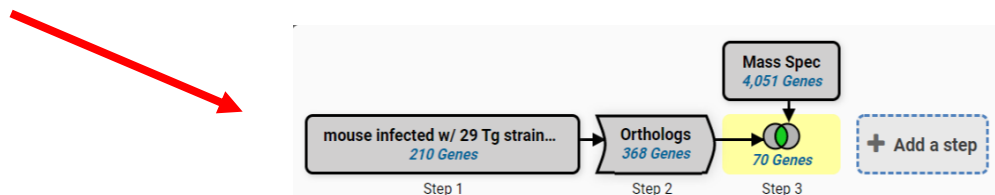
Note: You must select at least 1 values for this parameter.
7 selected, out of 14

select all | clear all | expand all | collapse all

Filter list below...

- ☒ Homo sapiens
 - ☒ Homo sapiens REF
 - ☒ Human Proteome During Infection with 4 strains of T. gondii and one strain of N. caninum (Wastling)
 - ☒ 16 hour infection of H.sapiens cells (GT1)
 - ☒ 16 hour infection of H.sapiens cells (ME49)
 - ☒ 16 hour infection of H.sapiens cells (VEG)
 - ☒ 36 hour infection of H.sapiens cells (RH)
 - ☒ 36 hour infection of H.sapiens cells (ncLIV)
 - ☒ 44 hour infection of H.sapiens cells (ME49)
 - ☒ 44 hour infection of H.sapiens cells (VEG)
 - ☐ Giardia secretome IEC infection (Maayeh et al.)
 - ☐ Human Erythrocyte Phosphoproteome during infection with P. fal 3D7 schizonts (2012) (Lasonder et al.)
 - ☐ Enriched schizont phospho-proteins (2012)

select all | clear all | expand all | collapse all



2. Identify host genes that are differentially regulated in multiple infection models (e.g. different hosts and parasites).

- Go to the RNAseq searches in HostDB. How many different pathogen infections are available? How many host organisms?
- Start by running a search using the “M. musculus C57BL6J Transcriptomes of 4 M. musculus cell types during infection with T. gondii (Swierzy et al.)”. Identify all genes that are differentially regulated (up or down) in all infected cell types compared to uninfected cell types.

Identify Genes based on M. musculus C57BL6J Transcriptomes of 4 M. musculus cell types during infection with T. gondii RNA-Seq (fold change)

For the Experiment
[Transcriptomes of 4 M. musculus cell types during infection with T. gondii unstranded]

return [protein coding] Genes
that are [Up or down regulated]

with a Fold change \geq 2

between each gene's [average] expression value
(or a Floor of 10 reads)

In the following [Reference Samples]

☐ SKM1C5 Uninfected
☐ SKM1C5 Infected
☒ Astrocytes Uninfected
☐ Astrocytes Infected
☒ Fibroblasts Uninfected
☐ Fibroblasts Infected

select all | clear all

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

In the following [Comparison Samples]

☐ SKM1C5 Uninfected
☐ SKM1C5 Infected
☒ Astrocytes Uninfected
☐ Astrocytes Infected
☐ Fibroblasts Uninfected
☒ Fibroblasts Infected

select all | clear all

Get Answer

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

Up or down regulated

For each gene, the search calculates:

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

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

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

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

- How many of these results are also differentially regulated in a *Leishmania* infection model? Try adding a step and running a search for genes that are differentially regulated in “M. musculus C57BL6J Transcriptomes of mouse macrophages infected with *Leishmania mexicana* (Fiebig et al.)”.

Add a step to your search strategy

The results will be [Intersected with] the results of Step 1.

Filter Data Sets: [Search]

Legend: [B] Similarity [DE] Differential Expression [FC] Fold Change [P] Percentile [SA] Sense/Antisense

Organism: [Mus musculus C57BL6J] Data Set: [Transcriptomes of mouse macrophages infected with Leishmania mexicana (Fiebig et al.)] Choose a Search: [DE] [FC] [P]

Show All Data Sets

Differential Expression Fold Change Percentile

For the Experiment
[Transcriptomes of mouse macrophages infected with Leishmania mexicana unstranded]

return [protein coding] Genes
that are [Up or down regulated]

with a Fold change \geq 2

between each gene's [average] expression value
(or a Floor of 10 reads)

In the following [Reference Samples]

☒ uninfected
☐ infected

select all | clear all

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

In the following [Comparison Samples]

☐ uninfected
☒ infected

select all | clear all

Run Step

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

Up or down regulated

For each gene, the search calculates:

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

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

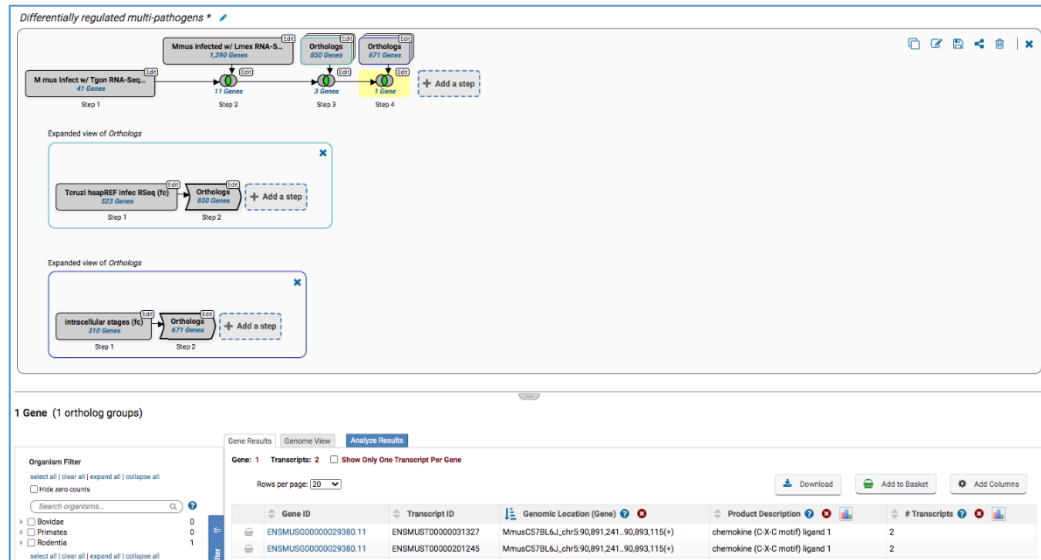
and returns genes when $\text{fold change}_{\text{up}} \geq 2$ or $\text{fold change}_{\text{down}} \geq 2$.

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

- Continue adding steps from other pathogen infections. For example, try the “Host cell transcriptome in bovine cells infected with *Cryptosporidium parvum*”. How many results in common did you get? If your answer is zero, did you remember to transform these results from bovine to mouse?

Here is an example search strategy to explore:

<https://hostdb.org/hostdb/im.do?s=4d0a7299510641cf>



3. **Identify host genes from a *Plasmodium* infection that are phosphorylated, secreted and have similarity to a 3D structure in the PDB database.**
- To do this you can start by looking at the available proteomics data. Can you find an experiment that identifies phosphoproteins?

Identify Genes based on Mass Spec. Evidence

Experiments and Samples

Note: You must select at least 1 values for this parameter.
1 selected, out of 14

select all | clear all | expand all | collapse all

Filter list below...

- ▼ Homo sapiens
 - ▼ Homo sapiens REF
 - Human Proteome During Infection with 4 strains of *T. gondii* and one strain of *N. caninum* (Wastling)
 - Giardia secretome IEC infection (Maayah et al.)
 - Human Erythrocyte Phosphoproteome during infection with *P. fal* 3D7 schizonts (2012) (Lasonder et al.)
 - Enriched schizont phospho-proteins (2012)

select all | clear all | expand all | collapse all

Minimum Number of Unique Peptide Sequences

1

Apply min # peptide sequences / sample OR across samples

Per Sample

- How many of these gene are also predicted be secreted? To figure this out add a step and search for genes that have a secretory signal peptide.
- Using the same logic as above, add another step to identify any gene with similarity to any structure in the PDB database.

← Add a step to your search strategy ⓘ

Combine with other Genes

Step 2 Step 3

Transform into related records

Step 2 Step 3

Use Genomic Colocation to combine with other features

Step 2 Step 3

1 Choose *how* to combine with other Genes

☒ 2 INTERSECT 3
 ☐ 2 UNION 3
 ☐ 2 MINUS 3
 ☐ 3 MINUS 2

2 Choose *which* Genes to combine. From...

☒ A new search
 ☐ An existing strategy
 ☐ My basket

expand all | collapse all

Filter the searches below...

- Annotation, curation and identifiers
- Function prediction
- Gene models
- Genomic Location
- Orthology and syteny
- Pathways and interactions
- Protein features and properties
- Protein targeting and localization
- Proteomics
- Sequence analysis
- Structure analysis
 - PDB 3D Structures
- Taxonomy
- Text
- Transcriptomics

← Add a step to your search strategy ⓘ

Search for Genes by PDB 3D Structures

The results will be ☐ intersected with ☐ the results of Step 2.

📘 Organism

Note: You must select at least 1 values for this parameter.
4 selected, out of 4

select all | clear all | expand all | collapse all

Filter list below...

- ☒ Bovidae
- ☒ Primates
- ☒ Rodentia

select all | clear all | expand all | collapse all

📘 With similarity to PDB Proteins from

- ☒ Archaea
- ☒ Bacteria
- ☒ Eukaryota - all
- ☒ Eukaryota - only Metazoa

select all | clear all

📘 BLAST P-value less than 10 to the

-20

Signal Pep
30,833 Genes

PDB 3D Struc
60,617 Genes

Mass Spec
187 Genes

39 Genes

25 Genes

Step 1 Step 2 Step 3

+ Add a step