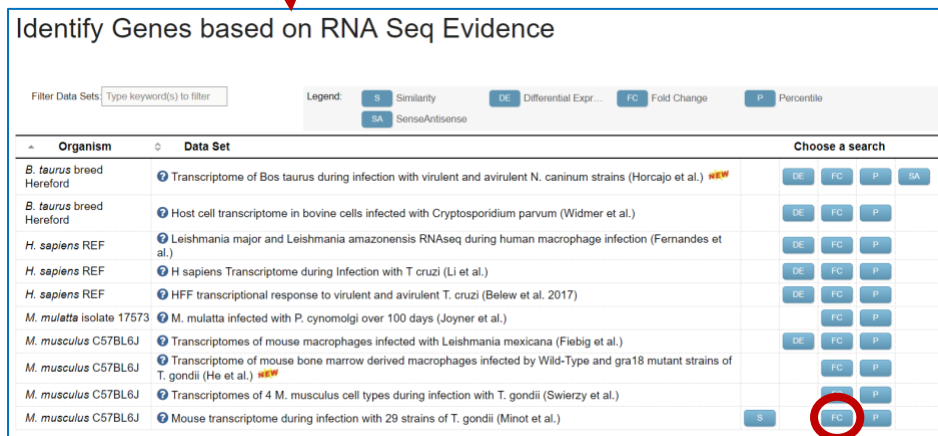
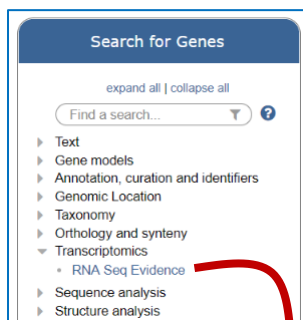


Host Response

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

- 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.



- 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 Transcriptomes of 29 strains during murine macrophage infection RNASeq (fold change) Tutorial

For the Experiment
Transcriptomes of 29 strains during murine macrophage infection unstranded

return protein coding Genes
that are up-regulated
with a Fold change ≥ 10

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

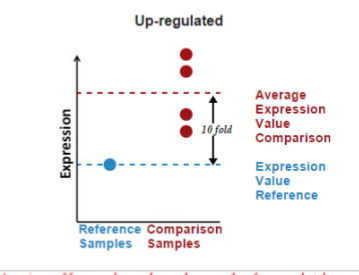
☐ TgCATBr0 infected
☐ VAND infected
☐ VEG infected
☐ WTD3 infected
☒ Un-infected

and its average expression value
in the following Comparison Samples

☒ TgCATBr0 infected
☒ VAND infected
☒ VEG infected
☒ WTD3 infected
☐ Un-infected

[Get Answer](#)

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 one reference sample and at least two comparison samples.

For each gene, the search calculates:
$$\text{fold change} = \frac{\text{average expression value in comparison samples}}{\text{reference expression value}}$$

and returns genes when fold change ≥ 10 . To narrow the window, use the minimum comparison value. To broaden the window, use the maximum comparison value.
See the detailed help for this search.

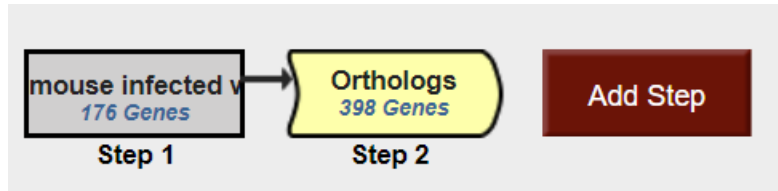
mouse infected v
 176 Genes
 Step 1

Add Step

- 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 Results Genome View Gene Ontology Enrichment Analyze Results										
[Rename This Analysis Copy These Parameter Values]										
Gene Ontology Enrichment										
Find Gene Ontology terms that are enriched in your gene result. Read More										
Parameters										
This analysis result may be lost if you change your gene result. To save this analysis result, please Download Analysis Results										
Analysis Results:										
Got a total of 1,258 results Filter: <input type="text"/>										
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:0006955	immune response	619	20	3.2	15.49	22.91	1.06e-18	1.33e-15	1.33e-15	
GO:0034097	response to cytokine	388	16	4.1	19.76	26.75	1.36e-16	5.97e-14	1.71e-13	
GO:0002376	immune system process	1069	22	2.1	9.86	15.13	1.42e-16	5.97e-14	1.79e-13	
GO:0071345	cellular response to cytokine stimulus	300	14	4.7	22.37	29.01	2.68e-15	8.42e-13	3.37e-12	
GO:0019221	cytokine-mediated signaling pathway	187	11	5.9	28.19	34.42	3.06e-13	7.70e-11	3.85e-10	

- 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

14 selected, out of 14

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 (Maayeh et al.)
 - ☒ Human Erythrocyte Phosphoproteome during infection with P. fal 3D7 schizonts (2012) (Lasonder et al.)

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

Workflow diagram showing Step 1: mouse infected v (176 Genes) leading to Step 2: Orthologs (398 Genes), which leads to Step 3: Mass Spec (4363 Genes) and 78 Genes, with an 'Add Step' button.

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 RNASeq**”. Identify all genes that are differentially regulated (up or down) in all infected cell types compared to uninfected cell

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

Tutorial

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 ≥ 2

between each gene's: average expression value (or a Floor of 10 reads (2 FPKM))

in the following Reference Samples:

☒ Neurons Uninfected
☐ Neurons Infected
☒ SKMCs Uninfected
☐ SKMCs 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:

☐ Neurons Uninfected
☒ Neurons Infected
☐ SKMCs Uninfected
☒ SKMCs Infected
☐ Astrocytes Uninfected
☒ Astrocytes Infected
☐ Fibroblasts Uninfected
☒ Fibroblasts 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 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 $\text{fold change}_{\text{up}} \geq 2$ or $\text{fold change}_{\text{down}} \geq 2$.

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

* or FPKM Floor, whichever is greater

Get Answer

types.

- 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* RNASeq”

Revise Step 2: M. musculus C57BL6J Transcriptomes of mouse macrophages infected with Leishmania mexicana RNASeq (fold change)

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 ≥ 2

between each gene's expression value (or a Floor of 10 reads (2 FPKM))

in the following Reference Samples:

☒ uninfected
☐ infected

and its expression value (or the Floor selected above)

in the following Comparison Samples:

☒ uninfected
☐ infected

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 one reference sample and one comparison sample.

For each gene, the search calculates:

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

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

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

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

* or FPKM Floor, whichever is greater

Combine Genes in Step 1 with Genes in Step 2:

☒ 1 Intersect 2
☐ 1 Union 2
☐ 1 Relative to 2, using genomic colocation

☐ 2 Intersect 1
☐ 2 Union 1
☐ 2 Relative to 1, using genomic colocation

☐ 1 Minus 2
☐ 2 Minus 1

- 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>

The screenshot shows the HostDB search strategy interface. At the top, there's a navigation bar with 'My Strategies: New Opened (7) All (7) Basket Public Strategies Help'. Below this is a 'Hide search strategy panel' button. The main area displays a search strategy titled 'Differentially regulated multi-pathogens'. The strategy consists of four steps: Step 1 (M. mus infect w/ 150 Genes), Step 2 (Orthologs 1778 Genes), Step 3 (Orthologs 438 Genes), and Step 4 (Orthologs 1 Gene). An 'Add Step' button is visible. Below the strategy steps is an 'Expanded View of Step Orthologs' section showing 'Intracellular stage 150 Genes' and 'Orthologs 438 Genes'. The bottom section, '3 Genes from Step 3', shows a table of results for the strategy 'Differentially regulated multi-pathogens'. The table has columns for 'All Results', 'Ortholog Groups', 'Bos', 'Homo', 'Macaca', and 'Mus'. The 'All Results' column shows 3 results. The 'Ortholog Groups' column shows 'B. taurus', 'H. sapiens', 'M. mulatta', and 'M. musculus'. The 'Bos' column shows 'Breed Hereford', 'REF', 'Isolate 17573', and 'CS7BL6J'. The 'Homo' column shows 'REF', 'Isolate 17573', and 'CS7BL6J'. The 'Macaca' column shows 'Isolate 17573' and 'CS7BL6J'. The 'Mus' column shows 'CS7BL6J'. Below the table, there's a 'Gene Results' section with 'Genes: 3' and 'Transcripts: 13'. A 'Show Only One Transcript Per Gene' checkbox is present. The 'Rows per page' is set to 100. The 'Download', 'Add to Basket', and 'Add Columns' buttons are visible. The table below shows the following data:

Gene ID	Transcript ID	Genomic Location (Gene)	Product Description	# Transcripts
ENSMUSG00000029380.11	ENSMUST00000031327	MmusCS7BL6J_ch5:90,891,241..90,893,115(+)	chemokine (C-X-C motif) ligand 1	2

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

1 selected, out of 14

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 (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](#)

Minimum Number of Unique Peptide Sequences

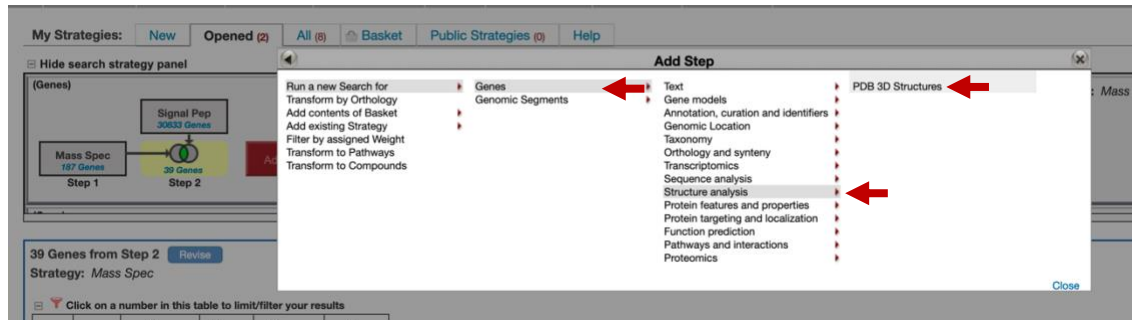
1

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 Step 3 : PDB 3D Structures

Organism

4 selected, out of 4

Filter list below...

- ☒ Bos
- ☒ Homo
- ☒ Macaca
- ☒ Mus

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

With similarity to PDB Proteins from

- ☒ Archaea
- ☒ Bacteria
- ☒ Eukaryota - all
- ☒ Eukaryota - only Chordata

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

BLAST P-value less than 10 to the

-20

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

