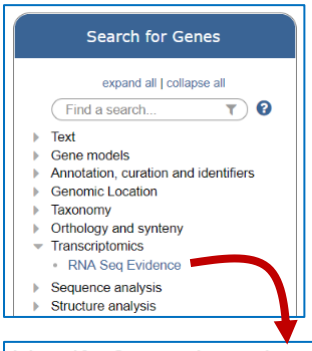
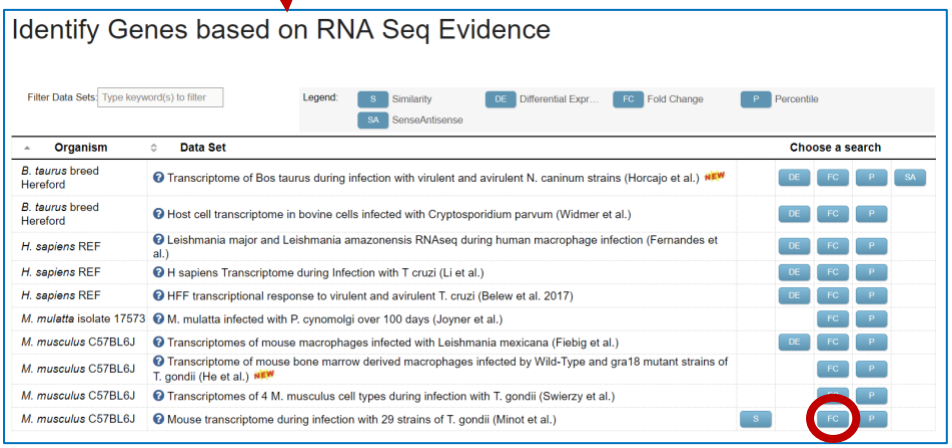
Host Response

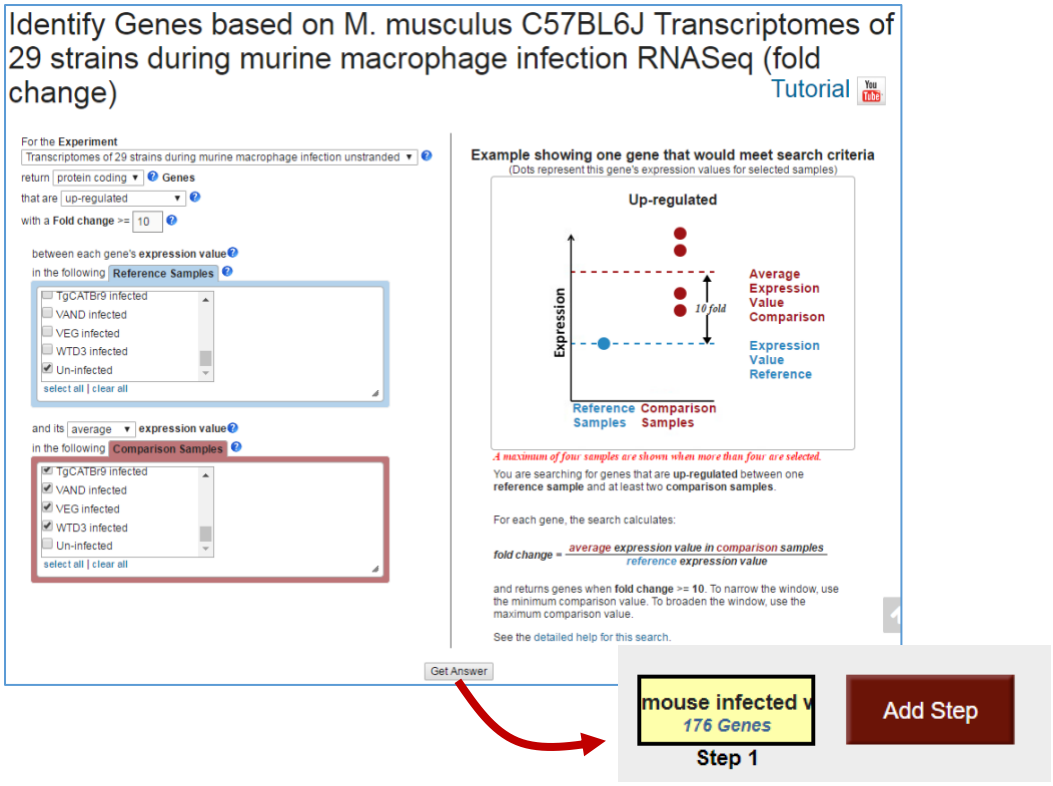
1. Find host genes that are upregulated in infected mouse cells compared to uninfected ones. For this exercise use http://hostdb.org

   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.

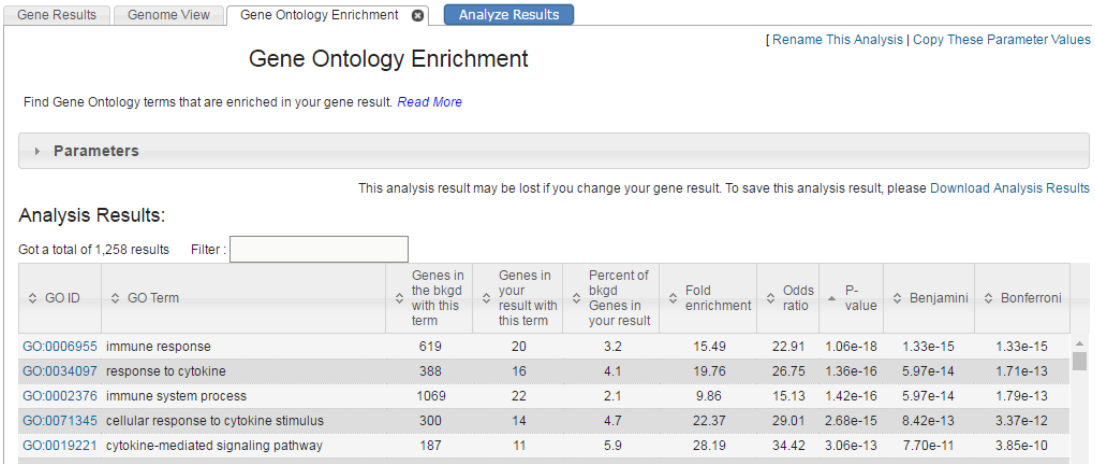




   b. Configure the search to return genes that are up-regulated at least 10-fold across all strains in the experiment compared to the uninfected control. Make sure to select upregulated. In the example below a fold change of 10 was selected and the **"average"** operation was applied on the comparison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

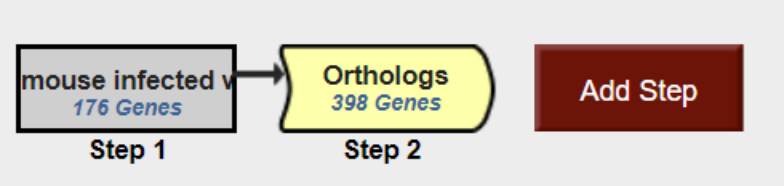


c. What are the functional characteristics of the genes in this result? What kinds of GO terms are enriched? Does the host immune response appear to be turned on? Is there a particular cellular location that is common in this group of genes?
Hint: click on the "Analyze Results" tab and perform a GO enrichment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



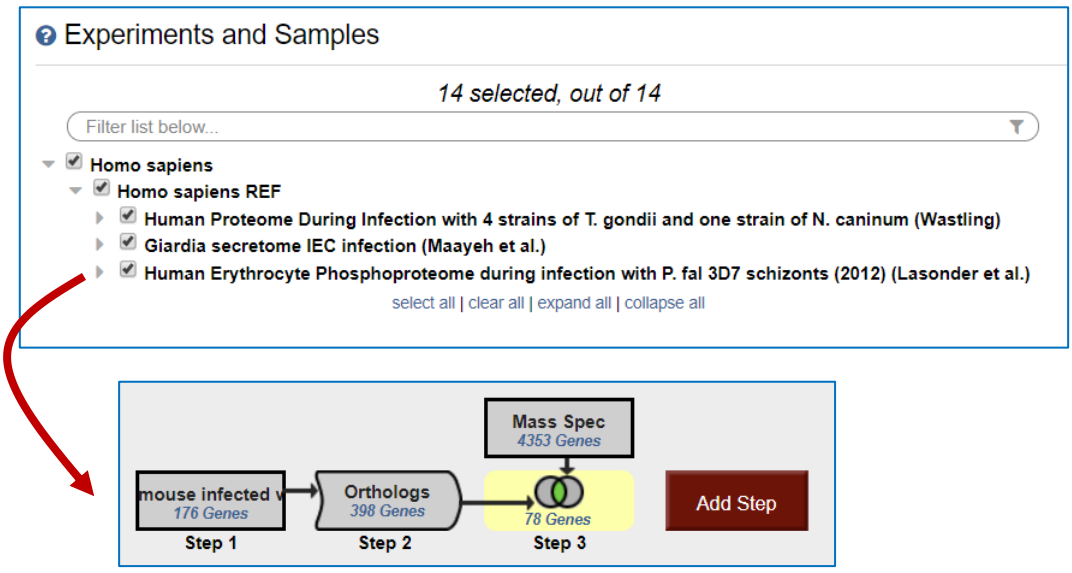
GO ID	GO Term	Genes in the bkgd with this term	Genes in your result with this term	Percent of bkgd Genes in your result	Fold enrichment	Odds ratio	P-value	Benjamini	Bonferroni
GO:0006955	immune response	619	20	3.2	15.49	22.91	1.06e-18	1.33e-15	1.33e-15
GO:0034097	response to cytokine	388	16	4.1	19.76	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.

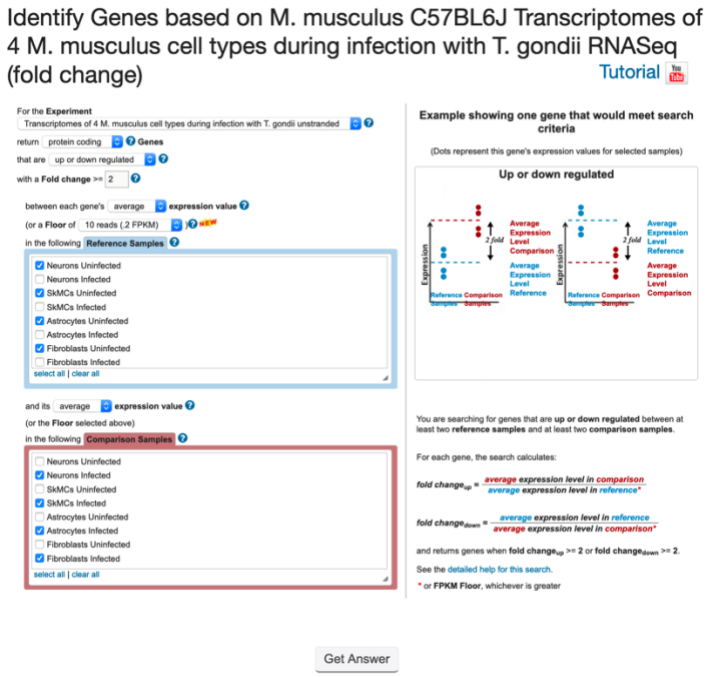


e. Does this set of human genes also show enriched GO terms? What, if any, are the enriched GO terms?

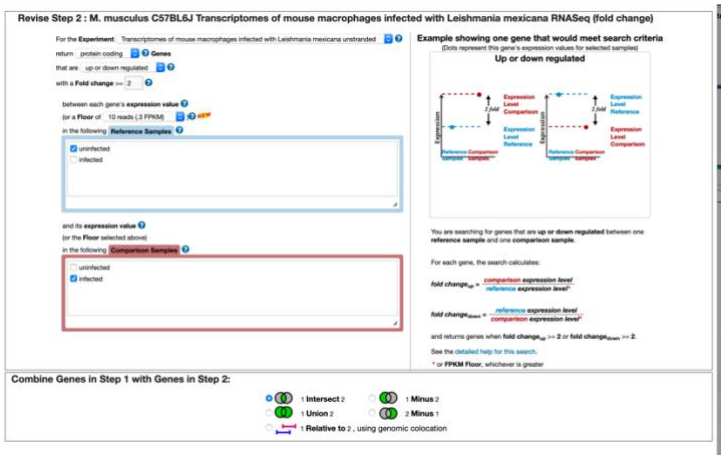
f. Do any of these human genes also have peptide evidence for their expression during infection? *Hint:* add a step and explore the proteomics data "Human Proteome During T. gondii infection"



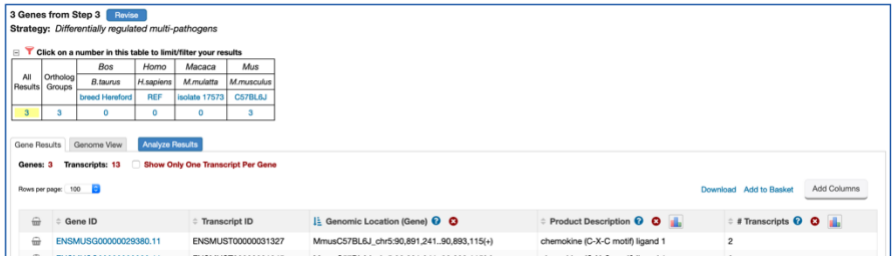
2. Identify host genes that are differentially regulated in multiple infection models (e.g. different hosts and parasites).
    - Go to the RNAseq searches in HostDB. How many different pathogen infections are available? How many host organisms?
    - Start by running a search using the "M. musculus 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 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"

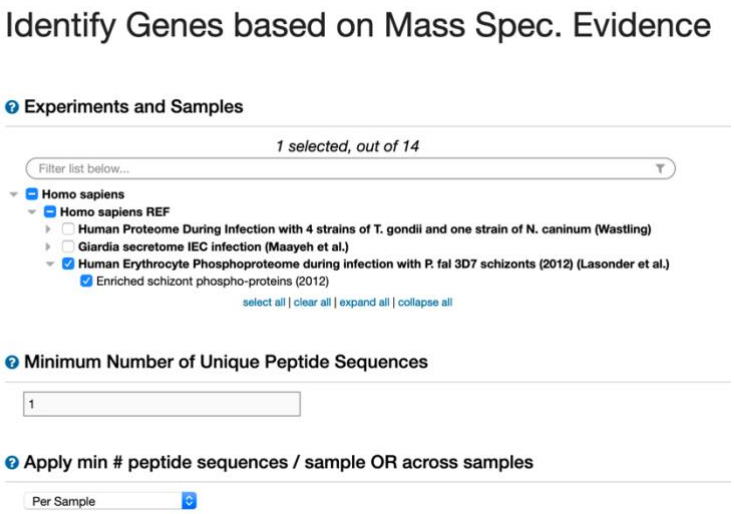


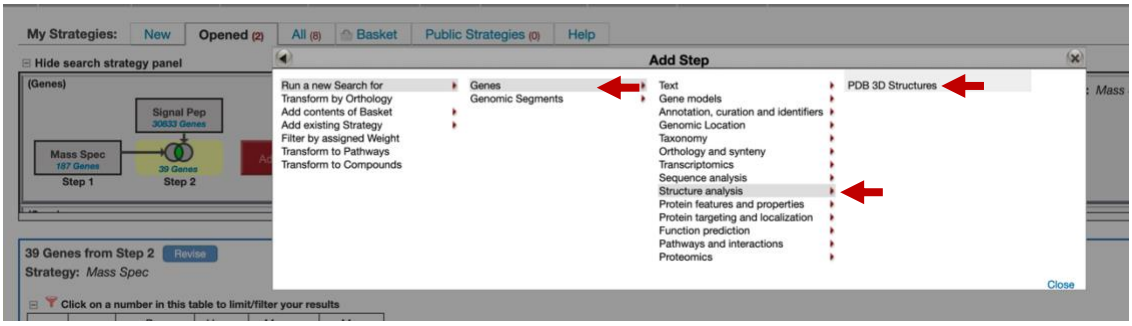
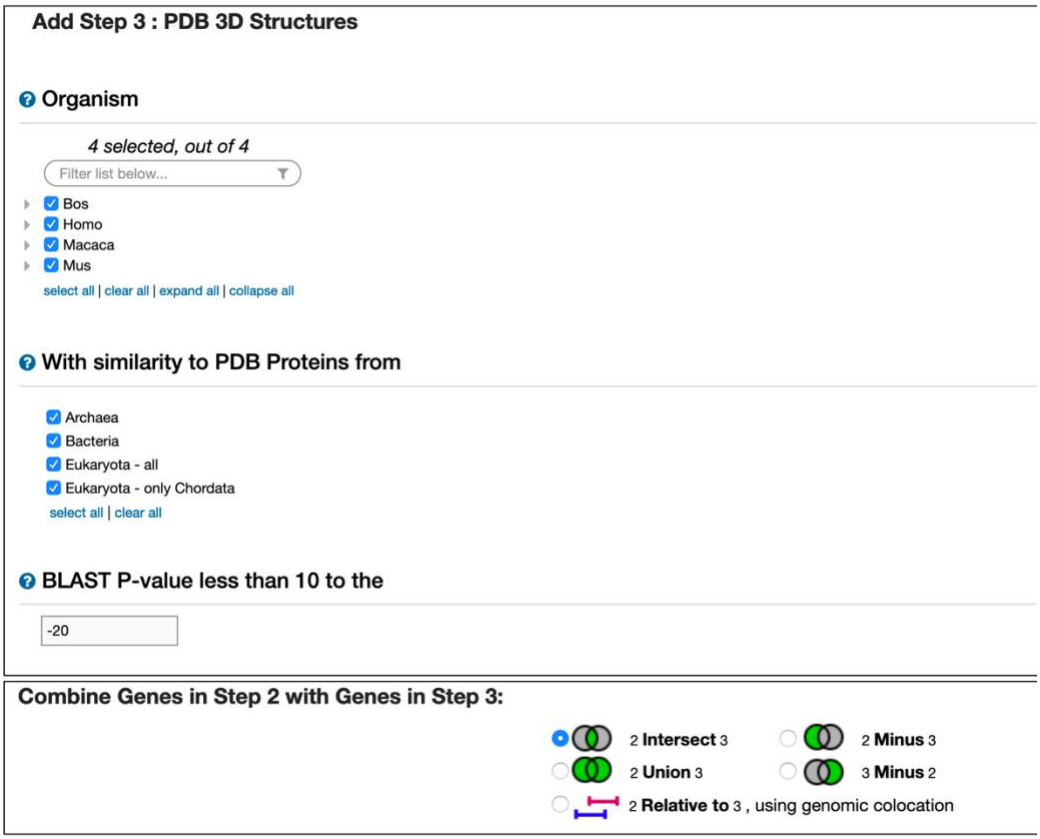
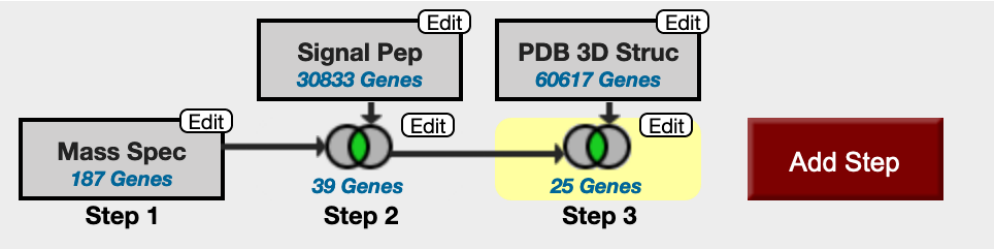
- Continue adding steps from other pathogen infections. For example, try the “Host cell transcriptome in bovine cells infected with Cryptosporidium parvum”. How many results in common did you get? If your answer is zero, did you remember to transform these results from bovine to mouse?
  Here is an example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



- 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.