

## Exploring transcriptomics datasets in FungiDB

Learning objectives:

- Query host-pathogen RNA-Seq data in HostDB and FungiDB.
- Create a proteomics query and save this strategy to your account.

Transcriptomics datasets can be analyzed using Fold Change (FC), Differential Expression (DE), Percentile (P), and Sense/Antisense searches (SA).

**Percentile (P):** For each Experiment and Sample, genes are ranked by expression level (e.g., search for low/high levels of gene expression).

**Fold change (FC):** Find genes with changes in gene expression when statistical analysis is not available (e.g., no replicates). After selecting samples, you have the option to take the average, minimum, or maximum expression value within each group. If choosing only one sample from a group, the selected 'operation' will not affect your results. Time-series experiments will offer an extra parameter called "Global min/max" which allows you to filter your results further. Finally, you can choose the directionality and the magnitude of the difference (e.g., up/down regulates, fold difference of 2, etc.)

**Differential Expression (DE).** This search uses DESeq2 analysis results. You can choose the directionality and the magnitude of the difference by setting both fold change and adjusted p values. For example, selecting up-regulated genes with a fold difference of 2 and an adjusted p-value cut off 0.1 will only show results where the comparator is twice that of the reference with an adjusted p-value of 0.1 or less.

**Sense/antisense (SA).** This search is applied to stranded datasets. You can find genes that exhibit simultaneous changes in sense and antisense transcripts in the Comparison sample relative to the Reference Sample. For example, you could look for genes showing increasing antisense transcripts and decreasing sense transcripts, as might occur when antisense transcription suppresses sense transcription. The search will perform all pairwise comparisons between the chosen Comparison samples and the chosen Reference samples.

**MetaCycle.** This search is applied to circadian datasets. For each study/experiment, you can choose either ARSER or JTK\_Cycle method for detecting rhythmic signals. The search will return the corresponding period, amplitude, and p-value.

In this exercise we will use HostDB.org to query host (*Mus musculus C57BL6J*) and pathogen (*Cryptococcus neoformans*) RNA-Seq data generated by Li et al. 2019. The authors used animal infection model (mouse) understand the mechanism of *C. neoformans* infection in the brain and lungs. Next, we will use FungiDB.org to identify differentially expressed (upregulated) genes in *Cryptococcus neoformans*.

The next block of exercises will be carried out in HostDB.org

- **Identify genes upregulated in the lung samples.**

1. Navigate to the “RNA-Seq evidence” search from the “Searches” menu at the top of the site.
2. Filter datasets on “crypto” and click on the “DE” button for the dataset titled “Transcriptional landscape of Cryptococcus-host response (Li et al. 2019)” for *Mus musculus* C57BL6J.
3. Choose to examine the sense strand.
4. Select reference sample: Lung uninfected.
5. Select comparator sample: Lung infected.
6. Look for up-regulated genes.
7. Select magnitude of upregulation: 2 fold and click on the “Get answer” button.

The screenshot shows the "Identify Genes based on RNA-Seq Evidence" search interface. The process is numbered 1 through 7:

- Step 1:** A search bar contains "rna". Below it, a sidebar lists "Genes", "Gene models", and "Transcriptomics" sections, each with several options like "Gene Model Characteristics", "Unannotated Intron Junctions", "Microarray Evidence", and "RNA-Seq Evidence". An orange circle labeled "1" is positioned to the right of the sidebar.
- Step 2:** The main search results page is shown. It has a legend with various filters (e.g., Expression, Quantitative Phenotype, Similarity, Splice Site Loc, Differential Expression, Fold Change, MetaCycle, Percentile, SenseAntisense) and a filter bar set to "crypto". Below the legend, two datasets are listed: "Mus musculus C57BL6J" and "Mus musculus C57BL6J". An orange arrow points down from the sidebar to this section. An orange circle labeled "2" is to the right of the results.
- Step 3:** A modal or dropdown menu is open, showing two options: "Gene expression profiles of the WT the nctA null and the nctB null mutants in response to itraconazole - Sense" (selected, indicated by a pink dot) and "Gene expression profiles of the WT the nctA null and the nctB null mutants in response to itraconazole - Antisense". An orange circle labeled "3" is to the left of the menu.
- Step 4:** The "Reference Sample" section is shown, with a legend for "Lung uninfected" (pink dot) and "Lung infected" (white circle). An orange circle labeled "4" is to the left of the legend.
- Step 5:** The "Comparator Sample" section is shown, with a legend for "Lung uninfected" (white circle), "Lung infected" (pink dot), "Brain uninfected" (white circle), and "Brain infected" (white circle). An orange circle labeled "5" is to the left of the legend.
- Step 6:** The "Direction" section shows a dropdown menu set to "up-regulated". An orange circle labeled "6" is to the left of the dropdown.
- Step 7:** The "fold difference >=" section shows a text input field containing "2". An orange circle labeled "7" is to the left of the input field.
- Step 8:** The "adjusted P value less than or equal to" section shows a text input field containing "0.1".
- Step 9:** At the bottom right, there is an orange arrow pointing right and a "Get Answer" button.

- Identify genes that are also upregulated in the brain samples.

1. Click on the “Add Step” button.
2. Navigate to the “RNA-Seq Evidence” search, filter for “crypto” to quickly identify the same dataset and click on the “DE” button.
3. Choose to examine the sense strand.
4. Select reference sample: Brain uninfected.
5. Select comparator sample: Brain infected.
6. Look for up-regulated genes.
7. Select magnitude of upregulation: 2 fold.

**2**

Choose how to combine with other Genes  
 2 INTERSECT 3    2 UNION 3    2 MINUS 3    3 MINUS 2

Choose which Genes to combine. From...  
 A new search    An existing strategy    My basket

Filter Data Set: crypto   4 results (Filtered from a total of 46)

Gene models:  
 Gene Model Characteristics  
 Transcriptomics  
 RNA-Seq Evidence  
 Single Cell RNA-Seq Evidence

Organism: Bos taurus breed Hereford  
Data Set: Host cell transcriptome in bovine cells infected with Cryptosporidium parvum (Widmer et al.)

Macaca fascicularis REF  
Data Set: Transcriptional landscape of Cryptococcus-host response (Li et al. 2019)

Mus musculus C57BL/6J  
Data Set: IFN-gamma independent host response to intestinal C. parvum infection (Mendoza Cavazos et al. 2022)

Mus musculus C57BL/6J  
Data Set: Transcriptional landscape of Cryptococcus-host response (Li et al. 2019)

**3**

Experiment  
 Gene expression profiles of the WT the nctA null and the nctB null mutants in response to itraconazole - Sense  
 Gene expression profiles of the WT the nctA null and the nctB null mutants in response to itraconazole - Antisense

**4**

Reference Sample  
 Lung uninfected  
 Lung infected  
 Brain uninfected  
 Brain infected

**5**

Comparator Sample  
 Lung uninfected  
 Lung infected  
 Brain uninfected  
 Brain infected

**6**

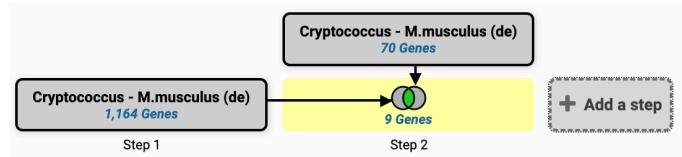
Direction: up-regulated

**7**

fold difference >= 2

adjusted P value less than or equal to 0.1

Run Step



This search strategy identified genes that are upregulated by 2 fold in the lung and also in the brain samples. Do the results make sense?

(optional) How can you adjust the search strategy to return a sum of all genes upregulated in both samples? Hint: Use the “union” Boolean operator. Here is how:

1. Hover over the Boolean step and click on the Edit button in yellow.
2. Change the selection to the “1 union 2” option.
3. Click on the “Revise” button.

**Details for step Combine Gene results**

9 Genes

**Revise as a boolean operation**

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

---

**Revise as a span operation**

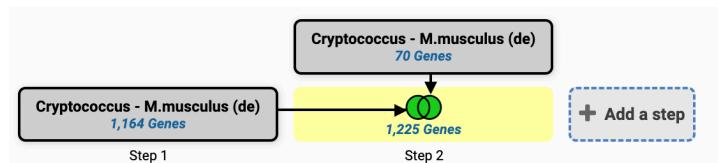
1 RELATIVE TO 2, using genomic colocation

---

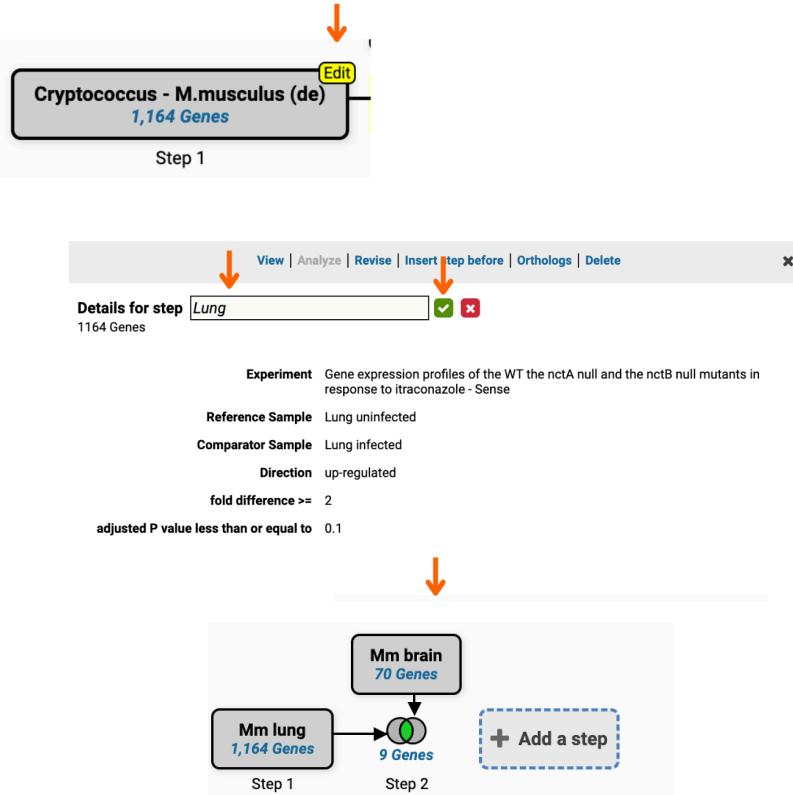
**Ignore one of the inputs**

IGNORE 2    IGNORE 1

**3** Revise



Note: you can rename steps to keep track of the samples and search results:



Save the strategy to your account under the name “ACE2”. We will come back to it in the module on enrichment analysis.



If this is your first time using HostDB.org you can use your FungiDB login created earlier. You only need one account to use all genomics VEuPathDB sites.

In summary, this strategy identified genes upregulated in both lung and brain samples in response to infection with *Cryptococcus neoformans* H99 in mice.

There are additional ways to continue analyze this data (e.g., GO enrichment, metabolic pathways, etc.) and we will look at these approaches to data analysis in the subsequent modules.

Strategy URL: <https://hostdb.org/hostdb/app/workspace/strategies/import/d701a66b6540c903>

**The next block of exercises will be carried out in FungiDB.org**

- Next, identify genes that are upregulated in *C. neoformans* H99 when infecting mouse lungs.
  1. Navigate to the “RNA-Seq Evidence” search and filter RNA-Seq datasets for “crypto”.
  2. Click on the “DE” button for the dataset titled “Transcriptional landscape of Cryptococcus-host response (Li et al. 2019)” for *Cryptococcus neoformans* var. *grubii* H99.
  3. Choose to examine the sense strand.
  4. Select reference sample: H99 -log phase.
  5. Select comparator sample: Lung – M. musculus.
  6. Look for up-regulated genes.
  7. Select magnitude of upregulation: 2 fold.

The screenshot shows the "Identify Genes based on RNA-Seq Evidence" search interface. The steps are numbered 1 through 7:

- Step 1:** The search bar contains "rna". The legend indicates "DE" is selected. A red circle labeled "1" is positioned above the search bar.
- Step 2:** The results page shows two datasets: "Gene expression variation in response to biologically relevant stresses (Yu et al. 2020)" and "Transcriptional landscape of Cryptococcus-host response (Li et al. 2019)". A red circle labeled "2" is positioned next to the second dataset.
- Step 3:** The "Reference Sample" section is shown, with "H99 - log phase" selected. A red circle labeled "3" is positioned next to the "Reference Sample" heading.
- Step 4:** The "Comparator Sample" section is shown, with "Lung - M. musculus" selected. A red circle labeled "4" is positioned next to the "Comparator Sample" heading.
- Step 5:** The "Direction" dropdown is set to "up-regulated". A red circle labeled "5" is positioned next to the "Direction" heading.
- Step 6:** The "fold difference >=" input field contains "2". A red circle labeled "6" is positioned next to the "fold difference >=" label.
- Step 7:** The "adjusted P value less than or equal to" input field contains "0.1". A red circle labeled "7" is positioned next to the "adjusted P value less than or equal to" label.

A "Get Answer" button is located at the bottom right of the form.

- Identify genes that are also upregulated in *C. neoformans* H99 when infecting mouse lungs.

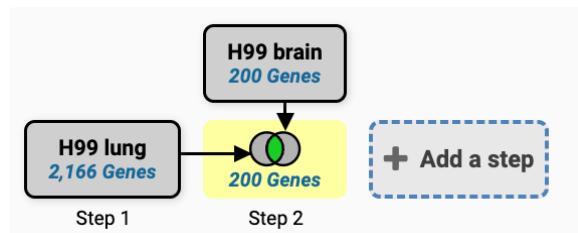
Repeat the steps from above.

Select the reference sample: H99 -log phase.

Select comparator sample: Brain – M. musculus.

Look for up-regulated genes.

Select magnitude of upregulation: 2 fold.



The strategy above identifies Cneo genes that are upregulated during infection of the lung and brain.

How would you identify genes that are only upregulated in the lung but not brain? (Hint: use Boolean operators).

Save the strategy to your account under the name “ACE3”. We will come back to it in the module on enrichment analysis.

Strategy URL:

<https://fungidb.org/fungidb/app/workspace/strategies/import/11bdb6cf129ecb7f>

In summary, in this exercise we used VEuPathDB.org sites (HostDB and FungiDB) to create RNA-Seq evidence queries. These queries examined host-pathogen interactions and transcriptomics changes occurring during *Cryptococcus* infection in different organs (lung and brain) using mouse and an animal model for human fungal disease.

