

## 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 RNA-Seq evidence search interface with the following elements and numbered steps:

- Step 1:** A dropdown menu is open, showing the navigation path: **Genes** > **Gene models** > **Transcriptomics** > **RNA-Seq Evidence**. An arrow points from this menu to the search filter.
- Filter Data Sets:** A search bar contains the text "crypto", resulting in "4 results (filtered from a total of 46)".
- Table of Results:**

Organism	Data Set	Choose a Search
<i>Bos taurus</i> breed Hereford	Host cell transcriptome in bovine cells infected with <i>Cryptosporidium parvum</i> (Widmer et al.)	DE FC P SA
<i>Macaca fascicularis</i> REF	Transcriptional landscape of <i>Cryptococcus</i> -host response (Li et al. 2019)	DE FC P SA
<i>Mus musculus</i> C57BL6J	IFN-gamma independent host response to intestinal <i>C. parvum</i> infection (Mendoza Cavazos et al 2022)	DE FC P SA
<i>Mus musculus</i> C57BL6J	Transcriptional landscape of <i>Cryptococcus</i> -host response (Li et al. 2019)	DE FC P SA
- Step 2:** An arrow points from the "DE" button in the last row of the table to the "Transcriptional landscape of *Cryptococcus*-host response - Sense" radio button.
- Step 3:** The "Transcriptional landscape of *Cryptococcus*-host response - Sense" radio button is selected.
- Step 4:** Under the "Reference Sample" section, the "H99 - log phase" radio button is selected.
- Step 5:** Under the "Comparator Sample" section, the "Lung - M. musculus" radio button is selected.
- Step 6:** Under the "Direction" section, the "up-regulated" dropdown menu is selected.
- Step 7:** Under the "fold difference >=" section, the value "2" is entered in the input field.
- adjusted P value less than or equal to:** The value "0.1" is entered in the input field.
- Get Answer:** A button at the bottom right of the form.

- **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**

Combine with other Genes

Transform into related records

Use Genomic Colocation to combine with other features

**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

ma

Gene models

- Q Gene Model Characteristics
- Transcriptomics
- Q RNA-Seq Evidence
- Q Single Cell RNA-Seq Evidence

Filter Data Sets:  4 results (filtered from a total of 46)

Organism	Data Set	Choose a Search
Bos taurus breed Hereford	Host cell transcriptome in bovine cells infected with Cryptosporidium parvum (Widmer et al.)	DE FC2 FC3 SA
Macaca fascicularis REF	Transcriptional landscape of Cryptococcus-host response (Li et al. 2019)	DE FC2 FC3 SA
Mus musculus C57BL6J	IFN-gamma independent host response to intestinal C. parvum infection (Mendoza Cavazos et al. 2022)	DE FC2 FC3 SA
Mus musculus C57BL6J	Transcriptional landscape of Cryptococcus-host response (Li et al. 2019)	DE FC2 FC3 SA
Mus musculus C1		DE FC2 FC3 SA

**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

**Reference Sample**

☐ Lung uninfected
 ☐ Lung infected
 ☒ Brain uninfected
 ☐ Brain infected

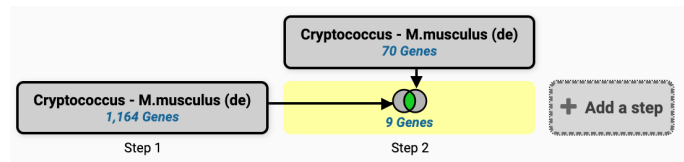
**Comparator Sample**

☐ Lung uninfected
 ☐ Lung infected
 ☐ Brain uninfected
 ☒ Brain infected

**Direction**

**fold difference >=**

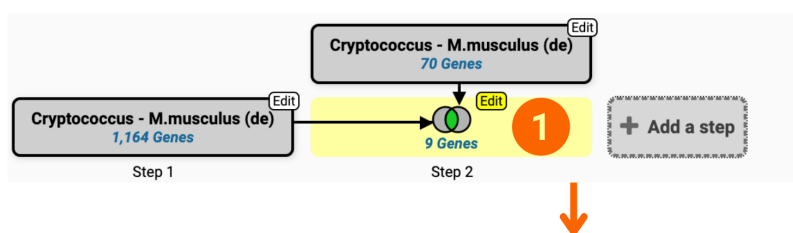
**adjusted P value less than or equal to**



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 [Edit](#)  
9 Genes

**2** 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

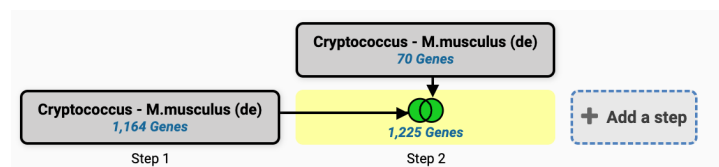
---

Ignore one of the inputs

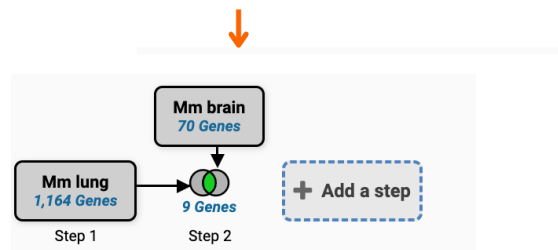
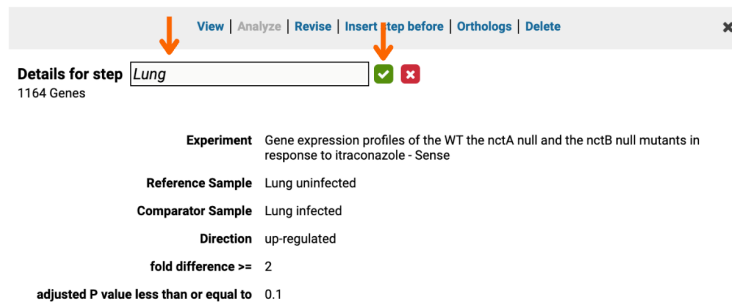
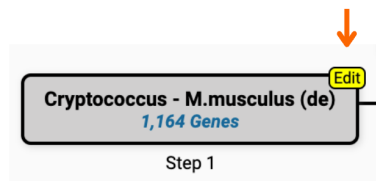
☐ IGNORE 2
 ☐ IGNORE 1

---

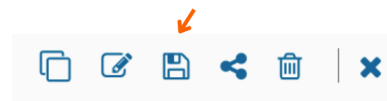
**3** [Revise](#)



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 both lung and brain samples in response to infection with *Cryptococcus neoformans* H99 in mice.

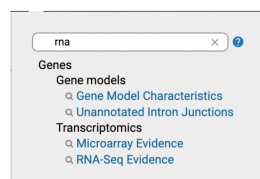
There are additional ways to continue analyze this data (e.g., GO enrichment, metabolic pathways, etc.) and we will look at these approaches to data analysis in the subsequent modules.

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.



#### Identify Genes based on RNA-Seq Evidence

Legend: ☒ Coexpression ☒ Similarity ☒ Differential Expression ☒ Fold Change ☒ MetaCycle ☒ Percentile ☒ SenseAntisense

Filter Data Sets:  7 results (filtered from a total of 174)

Organism	Data Set	Choose a Search
<i>Cryptococcus neoformans</i> var. <i>grubii</i> H99	Gene expression variation in response to biologically relevant stresses (Yu et al. 2020)	<input type="button" value="DE"/> <input type="button" value="FC"/> <input type="button" value="P"/> <input type="button" value="SA"/>
<i>Cryptococcus neoformans</i> var. <i>grubii</i> H99	Transcriptional landscape of Cryptococcus-host response (Li et al. 2019)	<input checked="" type="button" value="DE"/> <input type="button" value="FC"/> <input type="button" value="P"/> <input type="button" value="SA"/>

- 3
- ☒ Transcriptional landscape of Cryptococcus-host response - Sense  
☐ Transcriptional landscape of Cryptococcus-host response - Antisense

#### Reference Sample

- 4
- ☐ Lung - M. musculus  
☐ Brain - M. musculus  
☐ Lung - M. fascicularis  
☒ H99 - log phase

#### Comparator Sample

- 5
- ☒ Lung - M. musculus  
☐ Brain - M. musculus  
☐ Lung - M. fascicularis  
☐ H99 - log phase

#### Direction

6

#### fold difference >=

7

#### adjusted P value less than or equal to

Get Answer

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

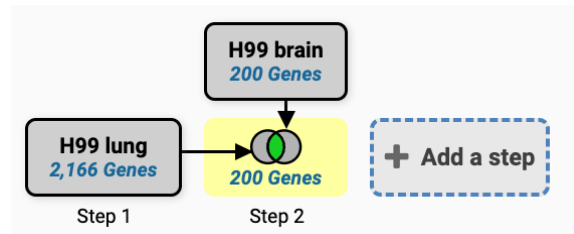
Repeat the steps from above.

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 analyzing a host-pathogen dataset addressing the mechanism of *Cryptococcus* infection in different organs (lung and brain) using animal model for human fungal disease.

