

# Host Response

## Background

- **HostDB** (<https://hostdb.org/hostdb/app/>) contains bioinformatic data related to hosts of eukaryotic pathogens
- This exercise explores host responses by running a search strategy in HostDB to ask the question- What host genes are differentially expressed in *Candida* infection?
- To answer this question, we will use data from three different experiments that examine fold change in gene expression after infection by *Candida* spp. In each dataset we will choose genes that are differentially expressed in uninfected vs. infected samples.

## Objectives

- Explore host responses by running a search strategy in HostDB
  - Add steps in a search strategy
  - Revise steps in a search strategy
1. Start on the **HostDB home page**: <https://hostdb.org/hostdb/app>
  2. Initiate a search for “RNA-Seq Evidence” (**Searches** in header menu or left sidebar)
  3. Find genes from the first experiment as follows
    - a. Filter datasets (red box below) to find the experiment “*Mouse transcriptomes during oropharyngeal candidiasis infection (Kirchner, et al. 2019)*” (red arrow)
    - b. In that dataset, choose the FC (fold change) option (red circle)

The screenshot shows the HostDB web application interface. At the top, there is a navigation bar with the HostDB logo, release information (Release 68, 7 May 2024), a search bar, and a menu with links: My Strategies, Searches, Tools, My Workspace, Data, About, Help, and Contact Us. Below the navigation bar, there is a section titled 'Identify Genes based on RNA-Seq Evidence'. This section includes a legend with tabs for S (Similarity), DE (Differential Expression), FC (Fold Change), W (WGCNA), P (Percentile), and SA (SenseAntisense). The 'Filter Data Sets' field is set to 'oropharyngeal', showing 2 results (filtered from a total of 55). The results table has two columns: 'Organism' and 'Data Set'. The first row shows 'Mus musculus C57BL6J' and 'Characterization of transcriptome dynamics in response to contact with host cells (Liu et al. 2015)'. The second row shows 'Mus musculus C57BL6J' and 'Mouse transcriptomes during oropharyngeal candidiasis infection (Kirchner, et al. 2019)'. A red arrow points to the second row. To the right of the table, there is a 'Choose a Search' section with buttons for DE, FC, P, and SA. The 'FC' button is circled in red.

- c. Configure the search to return all genes that are up or down regulated with a Fold Change  $\geq 2$ 
  - i. Reference Samples (blue): naive
  - ii. Comparison Samples (red): all infected samples up to 3 days

Differential Expression
**Fold Change**
Percentile
SenseAntisense

## Identify Genes based on M. musculus C57BL6J Mouse transcriptomes during oropharyngeal candidiasis infection RNA-Seq (fold change)

Configure Search
Learn More
View Data Sets Used

Reset values to default

For the **Experiment**

☒ Mouse transcriptomes during oropharyngeal candidiasis infection - Sense  
☐ Mouse transcriptomes during oropharyngeal candidiasis infection - Antisense

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**

☒ naive  
☐ 101\_infected\_9h  
☐ 101\_infected\_1d  
☐ 101\_infected\_3d  
☐ 101\_infected\_7d  
☐ SC5314\_infected\_9h  
☐ SC5314\_infected\_1d  
☐ SC5314\_infected\_3d  
☐ SC5314\_infected\_7d

select all | clear all

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

in the following **Comparison Samples**

☐ naive  
☒ 101\_infected\_9h  
☒ 101\_infected\_1d  
☒ 101\_infected\_3d  
☐ 101\_infected\_7d  
☒ SC5314\_infected\_9h  
☒ SC5314\_infected\_1d  
☒ SC5314\_infected\_3d  
☐ SC5314\_infected\_7d

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**

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

For each gene, the search calculates:

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

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

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 one reference sample and at least two comparison samples.

Get Answer

- iii. Click on Get Answer. This opens up a new search strategy. How many genes did it return?
- iv. Name the search strategy, e.g., “Genes differentially expressed in Candida infection”

My Search Strategies
Opened (1) All (6) Public (0) Help

Genes diff exp in Candida infection

Candida albicans infection (2)
Add a step

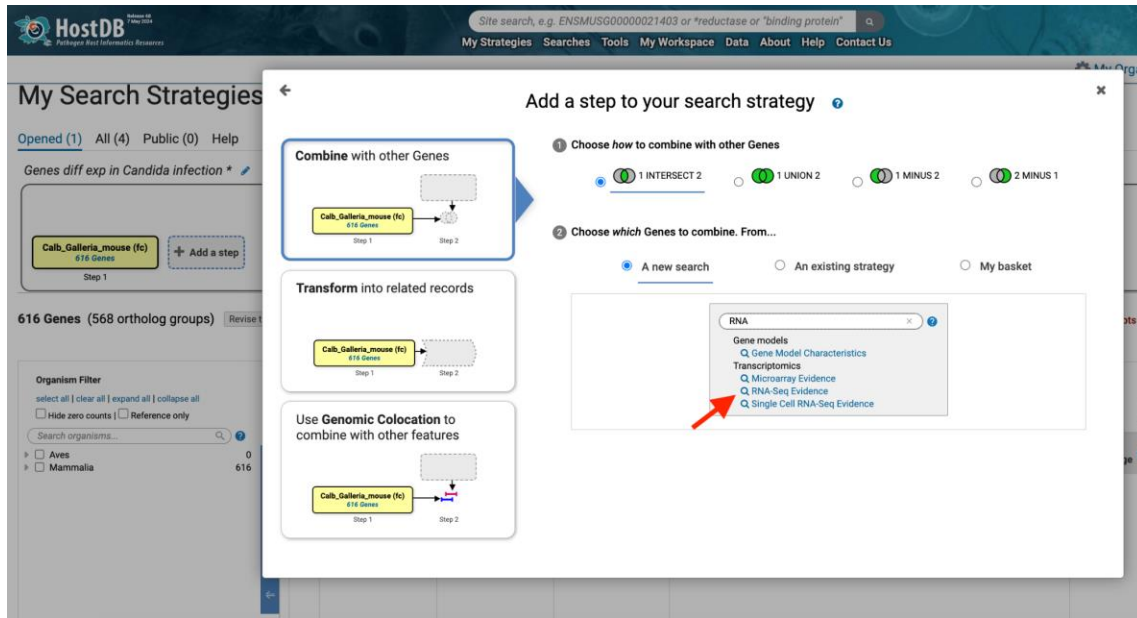
616 Genes (568 ortholog groups)

Organism Filter  
select all (3) out of 3 (collapse all)  
☐ Check gene quality ☐ Reference only  
☐ Search organisms  
☐ Add  
☐ Materials

Gene Results  
Genes: 616 Transcripts: 1,749  
☐ Show Only One Transcript Per Gene ☐ Show only the Genes in my basket  
Rows per page: 25

Gene ID	Transcript ID	Organism	Product Description	Fold Change	Chosen Ref (Row)	Chosen Comp (Rows)	Candida albicans - sense log2 graph	Candida albicans - antisense log2 graph
ENSMUSG00000000001	ENSMUST00000000001	Mouse Mus musculus C57BL/6J	5100 calcium binding protein A9 (paragardhin B) [Bosaurus B9] [Symbol: A9] [MGI:138945]	4174.9	2.22	5257.31		

4. Add data from a second RNA-Seq experiment by clicking on “Add step” in the strategy



- Find the experiment “*Fresh whole blood infected with Candida albicans (Sieber et al. 2019)*” and select the fold-change search.
- Configure the search to return all genes that are differentially regulated (up or down) by at least 2-fold when comparing 0 min and uninfected samples (reference) to 15-240 min infected samples (comparison).
- Click on Run Step

FOR THE EXPERIMENT

● Fresh whole blood infected with *Candida albicans* 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

☒ 0\_min  
☐ 15\_min  
☐ 30\_min  
☐ 60\_min  
☐ 120\_min  
☐ 240\_min  
☒ uninfected\_240\_min

select all | clear all

and its average expression value (or the Floor selected above) in the following Comparison Samples

☐ 0\_min  
☒ 15\_min  
☒ 30\_min  
☒ 60\_min  
☒ 120\_min  
☒ 240\_min  
☐ uninfected\_240\_min

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

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

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 results did you get?

## My Search Strategies

Opened (1) All (4) Public (0) Help

Genes diff exp in Candida infection \*

0 Genes (0 ortholog groups)

- e. Is this result of 0 genes surprising? What can you do to change this? Hint: think about the host organism for each experiment, are they the same or different? What can you do to compare across species?

## My Search Strategies

Opened (1) All (4) Public (0) Help

Genes diff exp in Candida infection \*

0 Genes (0 ortholog groups)

Organism Filter

select all | clear all | expand all | collapse all

☐ Hide zero counts ☐ Reference only

Search organisms...

☐ Aves ☐ Mammalia

Gene Results Genome V

☐ Show only the Genes in

Download

Details for step Calb\_Galleria\_mouse (fc) 616 Genes

Experiment Mouse transcriptomes during oropharyngeal candidiasis infection - Sense

Direction up or down regulated

Reference Samples naive

Operation Applied to Reference Samples average

Comparison Samples 101\_infected\_9h, 101\_infected\_1d, 101\_infected\_3d, SC5314\_infected\_9h, SC5314\_infected\_1d, SC5314\_infected\_3d

Operation Applied to Comparison Samples average

fold difference >= 2

Floor = 10 reads

Protein Coding Only: protein coding

Give this search a weight

- f. Edit the first step to convert to orthologs, choosing *Homo Sapiens* as the species. Run step.

## Add a step to your search strategy

Your Genes from Step 1 will be converted into Orthologs

Configure Search Learn More View Data Sets Used

Reset values to default

Organism

1 selected, out of 10

select only these | add these | clear these

☐ Reference only

☒ Homo sapiens

☐ Mammalia

☐ Primates

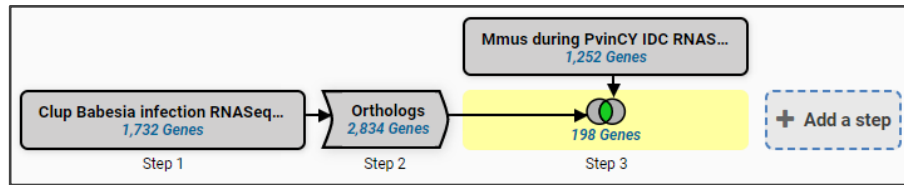
☒ Hominidae

☒ Homo sapiens REF [Reference]

Syntenic Orthologs Only?

no

Run Step



- Use the same logic above and add data from a third RNA-Seq experiment “Fresh whole blood infected with *Candida glabrata* (Sieber et al. 2019)”. Configure the search as before (shown below) and run step.

### Add a step to your search strategy

☒ Fresh whole blood infected with *Candida glabrata* unstranded  
 return  Genes  
 that are   
 with a **Fold change**   
 between each gene's  expression value  
 (or a **Floor** of )  
 in the following **Reference Samples**:

- ☐ 0\_min
- ☐ 15\_min
- ☐ 30\_min
- ☐ 60\_min
- ☐ 120\_min
- ☐ 240\_min
- ☒ uninfected\_240\_min

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

and its  expression value  
 (or the **Floor** selected above)  
 in the following **Comparison Samples**:

- ☐ 0\_min
- ☒ 15\_min
- ☒ 30\_min
- ☒ 60\_min
- ☒ 120\_min
- ☒ 240\_min
- ☐ uninfected\_240\_min

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

(dots represent this genes expression values for selected samples)  
**Up or down regulated**

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

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 genes were differentially expressed in all three experiments? Are there any genes with interesting functions in the list of results? Hint: Analyze results and try a GO enrichment.

### My Search Strategies

Opened (1) All (4) Public (0) Help

Genes diff exp in Candida infection \*

```

graph LR
    S1[Calb_Galleria_mouse (fc)  
616 Genes] --> S2[Orthologs  
1,237 Genes]
    S2 --> S3[blood Calbicans (fc)  
875 Genes]
    S2 --> S4[blood Cglabrata (fc)  
1,235 Genes]
    S2 & S3 --> I1[108 Genes]
    S2 & S4 --> I2[77 Genes]
    S4 --> AS[+ Add a step]
  
```

77 Genes (66 ortholog groups)

Some Genes in your basket

Organism Filter:      
☐ Hide zero counts ☐ Reference only

Search organisms:

Gene Results: 77 Genes 406 Transcripts ☐ Show Only One Transcript Per Gene ☐ Show only the Genes in my basket.

Rows per page: 20

Gene ID	Transcript ID	Genomic Location (Gene)	Product Description
ENSG00000159189	ENST00000374637	hsapREF_chrl:22,643,633..22,648,110(+)	complement C1q C cha
ENSG00000159189	ENST00000374639	hsapREF_chrl:22,643,633..22,648,110(+)	complement C1q C cha
ENSG00000159189	ENST00000374640	hsapREF_chrl:22,643,633..22,648,110(+)	complement C1q C cha