

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). A filter bar shows "crypto" and "39 results (filtered from a total of 572)". Below the results, 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) 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. It has a legend with four options: "Lung uninfected" (selected), "Lung infected", "Brain uninfected", and "Brain infected". An orange circle labeled "4" is to the left of the legend.
- Step 5:** The "Comparator Sample" section is shown. It has a legend with four options: "Lung uninfected", "Lung infected" (selected), "Brain uninfected", and "Brain infected". 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 has 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 has 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

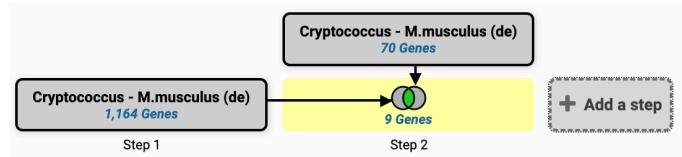
Direction: up-regulated

7

fold difference >= 2

adjusted P value less than or equal to 0.1

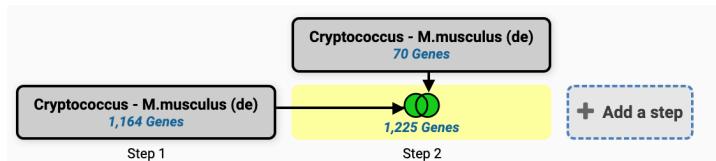
Run Step



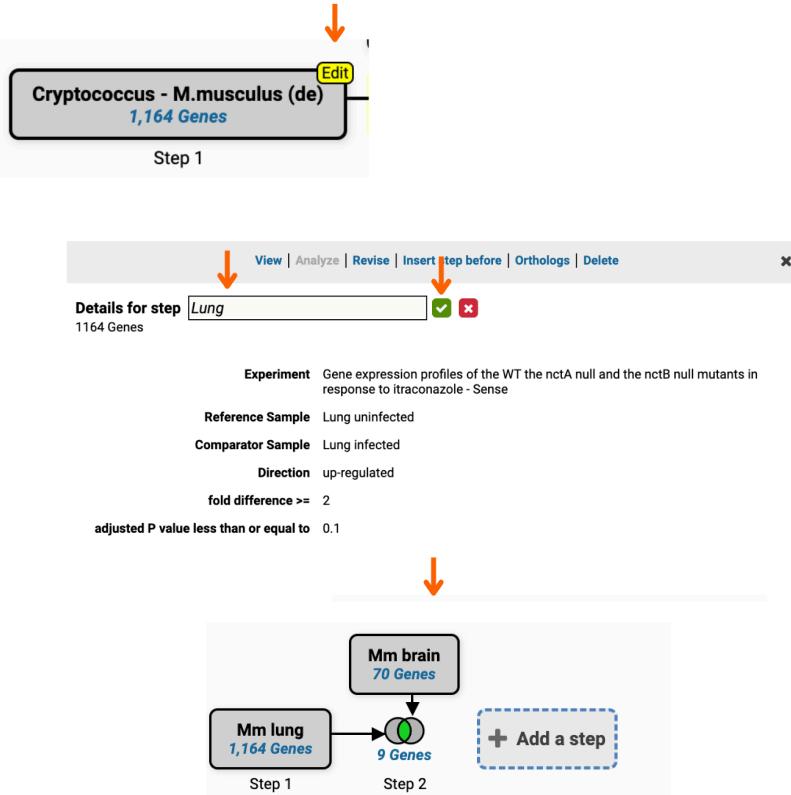
This search strategy identified genes that are upregulated by 2 fold in the lung and also in the brain samples. Do the results make sense?

(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.



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". A legend at the top includes "Gene models" (Gene Model Characteristics, Unannotated Intron Junctions) and "Transcriptomics" (Microarray Evidence, RNA-Seq Evidence). An orange circle labeled "1" is positioned above the search bar.
- Step 2:** The results page shows 7 results from a total of 174. The "Choose a Search" section has "DE" selected. An orange circle labeled "2" is positioned next to the search filters.
- Step 3:** The "Sense" radio button is selected under "Transcriptional landscape of Cryptococcus-host response". An orange circle labeled "3" is positioned next to the radio buttons.
- Step 4:** The "H99 - log phase" radio button is selected under "Reference Sample". An orange circle labeled "4" is positioned next to the radio buttons.
- Step 5:** The "H99 - log phase" radio button is selected under "Comparator Sample". An orange circle labeled "5" is positioned next to the radio buttons.
- Step 6:** The "up-regulated" dropdown menu is selected under "Direction". An orange circle labeled "6" is positioned next to the dropdown.
- Step 7:** The "2" input field is selected under "fold difference >=". An orange circle labeled "7" is positioned next to the input field.
- Step 8:** The "0.1" input field is selected under "adjusted P value less than or equal to".
- Step 9:** A "Get Answer" button is located at the bottom right.

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

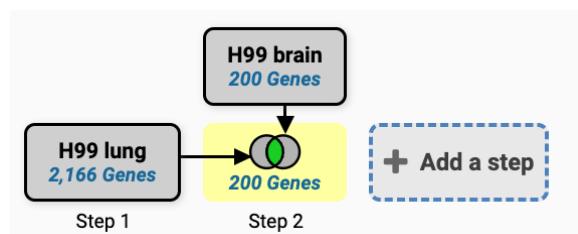
Repeat the steps from above.

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.

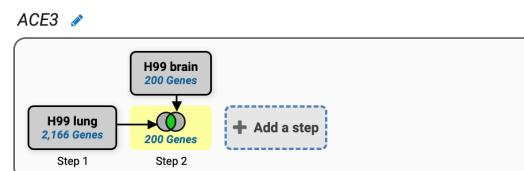
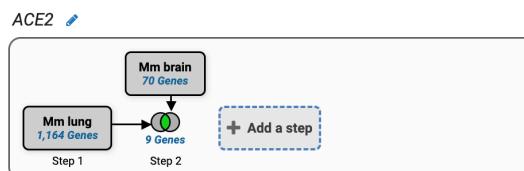
How would you identify genes that are only upregulated in the lung but not brain? (Hint: use Boolean operators).

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.



- Next, create a new RNA-Seq query using fungal response to macrophages data from *Candida albicans* and *Clavispora lusitaniae* and merge results with the ACE 3 search strategy (lung and brain responses).

- ***Candida albicans***

1. Navigate to the “RNA-Seq Evidence” search and filter RNA-Seq datasets for “macro”.
2. Click on the “DE” button for the dataset titled “Transcriptional responses to macrophage phagocytosis” for *Candida albicans* SC5314.
3. Choose to examine the sense strand.
4. Select reference sample: SC5314.
5. Select comparator sample: SC5314 + mac.
6. Look for up-regulated genes.
7. Select 2 fold and click on the “Get Answer” button.

The screenshot shows the RNA-Seq Evidence search interface. A vertical sequence of numbered steps (1 through 7) is overlaid on the interface, with orange arrows indicating the flow between them.

- Step 1:** The search bar at the top contains "rna". Below it, under "Genes", there is a section for "Gene models" with options like "Gene Model Characteristics", "Unannotated Intron Junctions", "Transcriptomics", "Microarray Evidence", and "RNA-Seq Evidence". The "RNA-Seq Evidence" option is highlighted with a red circle.
- Step 2:** An arrow points down to the main search interface. The legend at the top includes "Coexpression", "Similarity", "Differential Expression", "Fold Change", "MetaCycle", "Percentile", and "SenseAntisense". The "Filter Data Sets" dropdown is set to "macro". The search results table shows two entries: "Candida albicans SC5314" and "Candida albicans SC5314". The second entry has a blue question mark icon and the text "Transcriptional analysis of sorted subpopulations of macrophages infected with C. albicans (Munoz et al. 2019)". Another blue question mark icon next to it indicates "Transcriptional responses to macrophage phagocytosis (Pountain et al. 2021)". A red arrow points to this second entry.
- Step 3:** An arrow points down to the "Reference Sample" section. It shows two radio buttons: "SC5314 + mac" (selected) and "SC5314".
- Step 4:** An arrow points down to the "Comparator Sample" section. It shows two radio buttons: "SC5314 + mac" (selected) and "SC5314".
- Step 5:** An arrow points down to the "Direction" section, which has a dropdown menu set to "up-regulated".
- Step 6:** An arrow points down to the "fold difference >=" section, which has a text input field containing "2".
- Step 7:** An arrow points down to the "adjusted P value less than or equal to" section, which has a text input field containing "0.1".

At the bottom right of the interface is a "Get Answer" button.

- ***Clavispora lusitaniae***

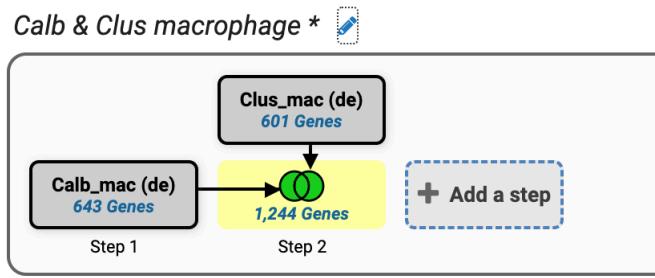
1. “Add Step” and navigate to the “RNA-Seq Evidence” search and filter RNA-Seq datasets for “macro”.
2. Click on the “DE” button for the dataset titled “Transcriptional responses to macrophage phagocytosis” for *C. lusitaniae* ATCC 42720.
3. Choose to examine the sense strand.
4. Select reference sample: SC5314.
5. Select comparator sample: SC5314 + mac.
6. Look for up-regulated genes.
7. Select 2 fold and click on the “Get Answer” button.

The screenshot shows the RNA-Seq Evidence search interface. A vertical orange arrow on the left indicates the flow from step 1 to step 7. Numbered circles (1-7) are placed on specific UI elements to guide the user:

- Step 1:** A screenshot of the search bar with "rna" typed in. The "RNA-Seq Evidence" option is highlighted with a red circle.
- Step 2:** The search results page for "macro". It shows two entries: "Transcriptional analysis of sorted subpopulations of macrophages infected with *C. albicans* (Munoz et al. 2019)" and "Transcriptional responses to macrophage phagocytosis (Pountain et al. 2021)". The second entry is highlighted with a red arrow.
- Step 3:** The "Choose a Search" section with the "DE" button highlighted with a red circle.
- Step 4:** The "Reference Sample" section with the "ATCC 42720 + mac" radio button highlighted with a red circle.
- Step 5:** The "Comparator Sample" section with the "ATCC 42720 + mac" radio button highlighted with a red circle.
- Step 6:** The "Direction" dropdown menu set to "up-regulated", highlighted with a red circle.
- Step 7:** The "fold difference >=" input field containing the value "2", highlighted with a red circle.

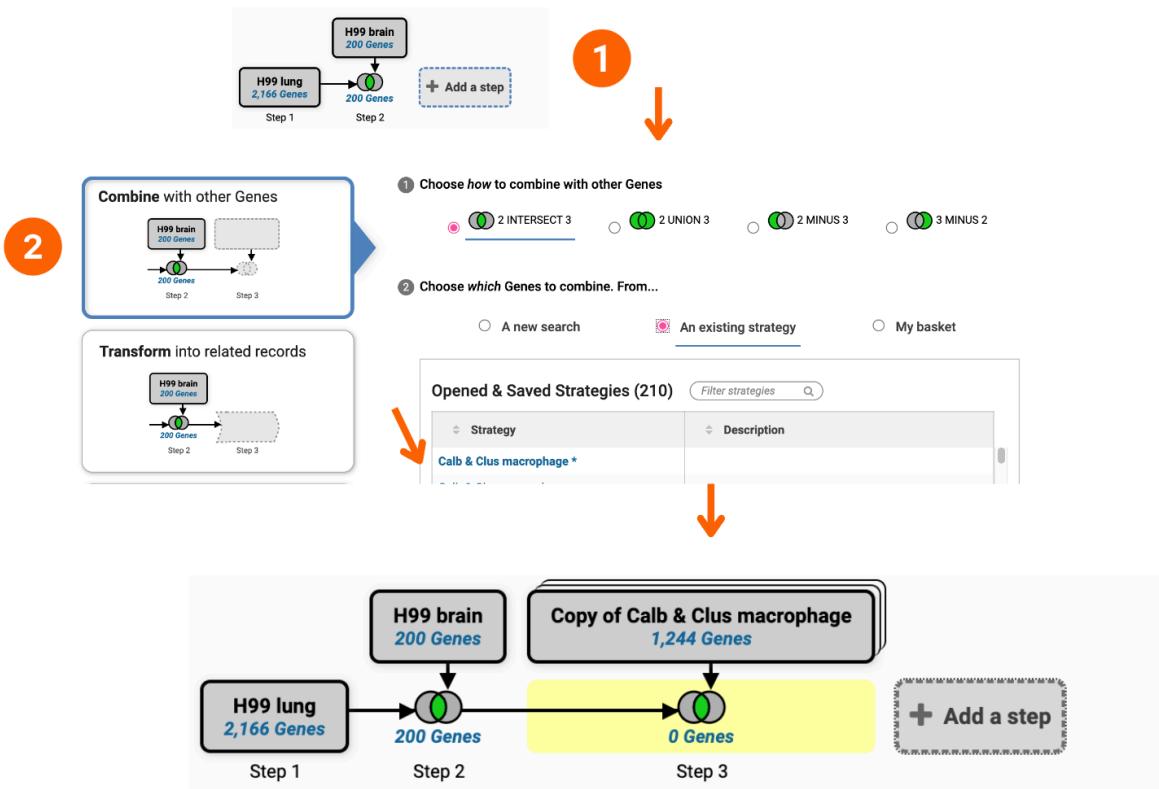
At the bottom right, there is a "Run Step" button with an orange arrow pointing to it.

Name this strategy as "C Alb & Clus macrophage" and save it to your account.



This strategy identifies genes upregulated in *C. albicans* and *C. lusitaniae* in response to macrophages.

- In the next step, merge this strategy with ACE3 to determine all genes in *Cryptococcus* that may be relevant to fungal responses to host (lung, brain, macrophage).
 1. Navigate to the ACE3 strategy and click on the “Add Step” button.
 2. Run “Combine with other Genes” search, choose the “An existing strategy” option, and select the “C Alb & Clus macrophage” strategy from the list.

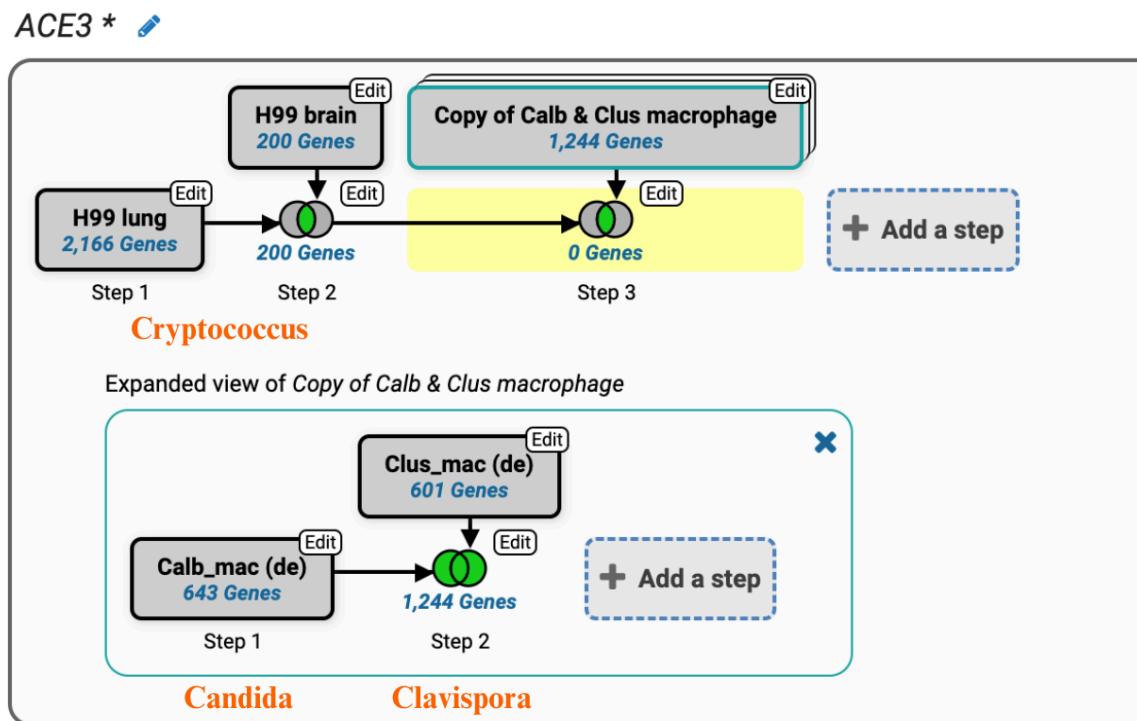


Note that you have 0 results, and the Step 3 has multiple search windows. Why did the last search step return 0 results and what can be done to address this issue within the strategy itself?

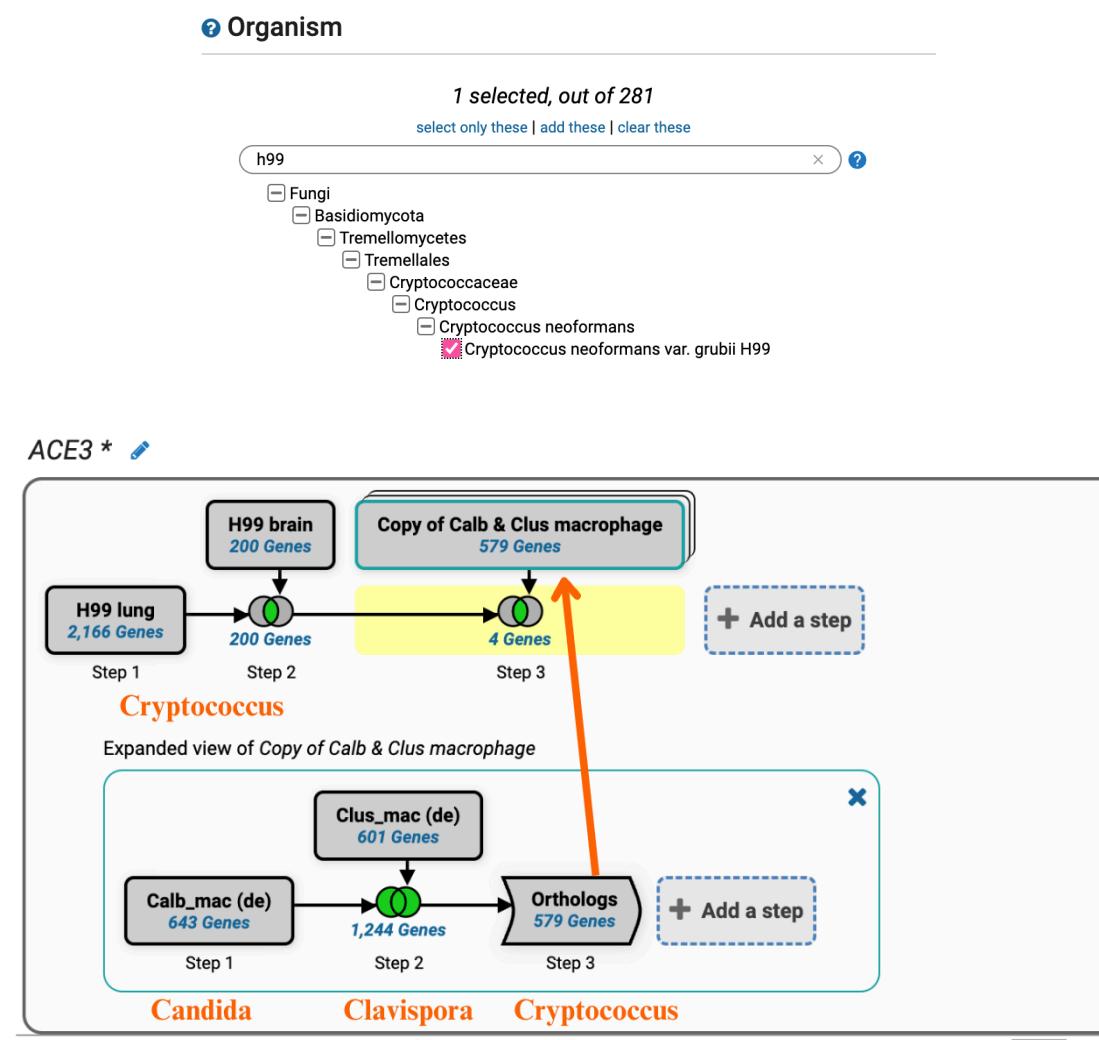
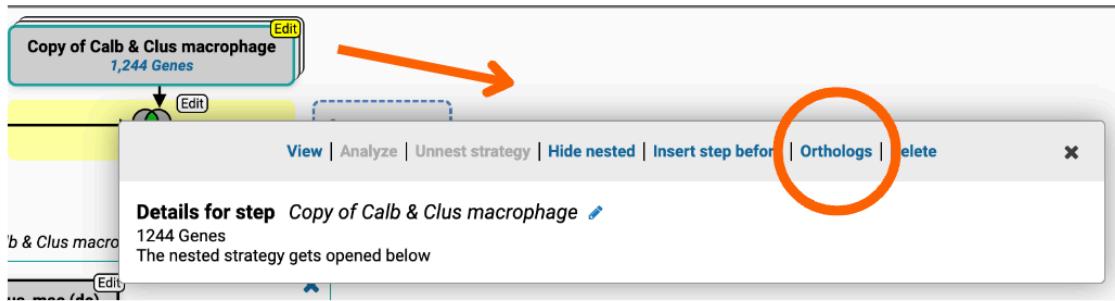
Hover over the last step and click on the edit button, and choose the “Show nested” option to display all search details:



This strategy uses datasets that return gene lists from three organisms: *Cryptococcus neoformans* (Step 1 and 2), and *Candida albicans* and *Clavispora lusitaniae* (Step 3).



How can you reconcile these differences? (Hint: use the “Orthologs” option to convert gene records in the nest strategy into *C. neoformans* gene IDs).



In summary, the multi-step nested strategy we created here combines knowledge from four different datasets addressing fungal pathogen response to host (lung and brain samples from *Cryptococcus*, macrophage experiments from *Candida* and *Clavispora*). The strategy uses orthologous transformation approach to bridge records from different genomes and identifies genes that are consistently up-regulated 2 fold across all datasets.

Strategy URL:

<https://fungidb.org/fungidb/app/workspace/strategies/import/4644792af2d8a913>