Exploring transcriptomics & proteomics datasets in FungiDB

Transcriptomics

Learning objectives:

- Query host-pathogen RNA-Seq data in HostDB and FungiDB, respectively.
- 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 in 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.

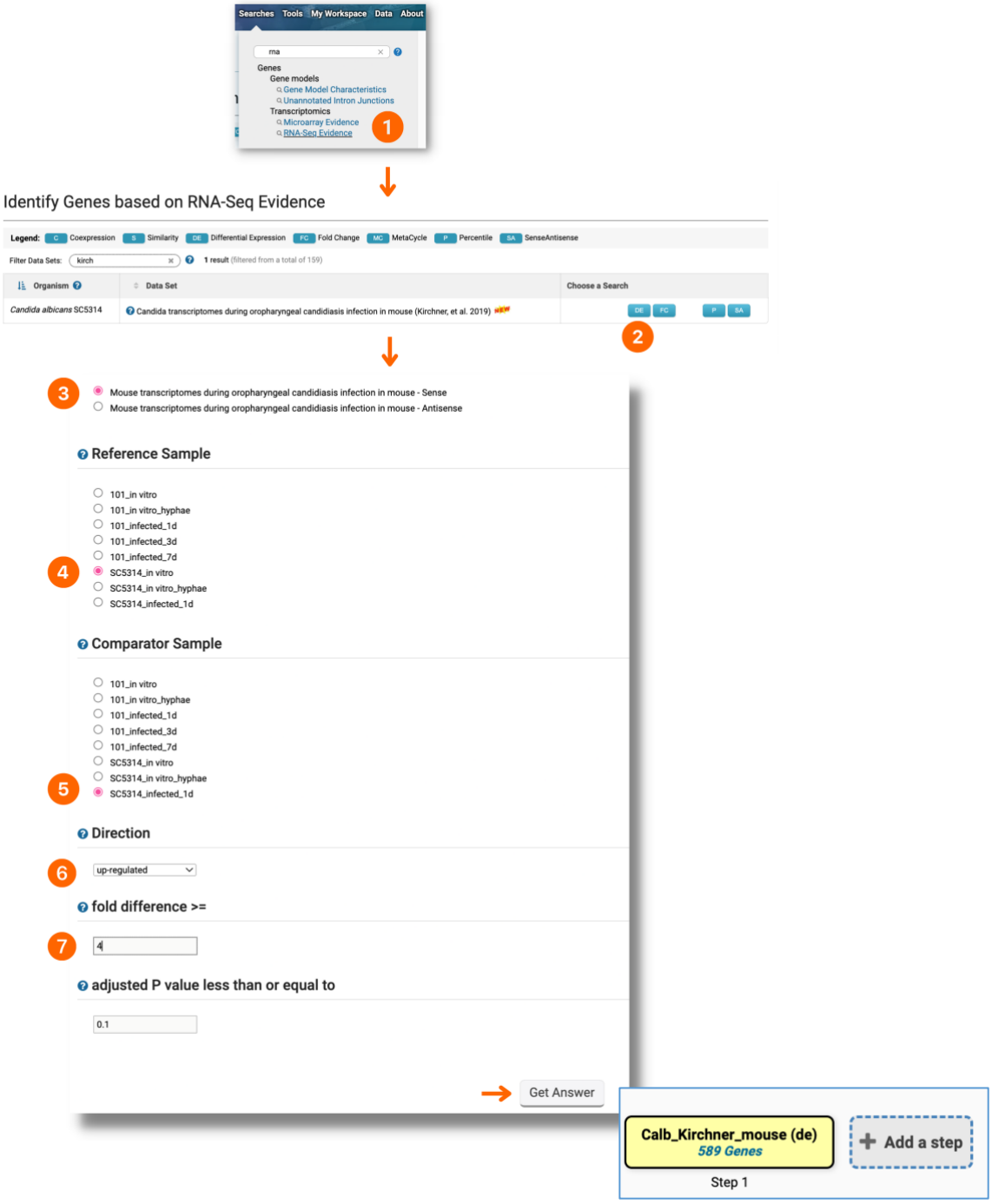
In this exercise we will query host (mouse) and pathogen (*Candida albicans*) RNA-Seq data generated by Kirchner et al. 2019. The authors used the experimental model of oropharyngeal candidiasis in mice to understand the interaction of *C. albicans* with the host at mucosal surfaces *in vivo*. Two *C. albicans* strains were used in this study – SC5314 (virulent lab strain) and the persistent strain 101.

Objectives:

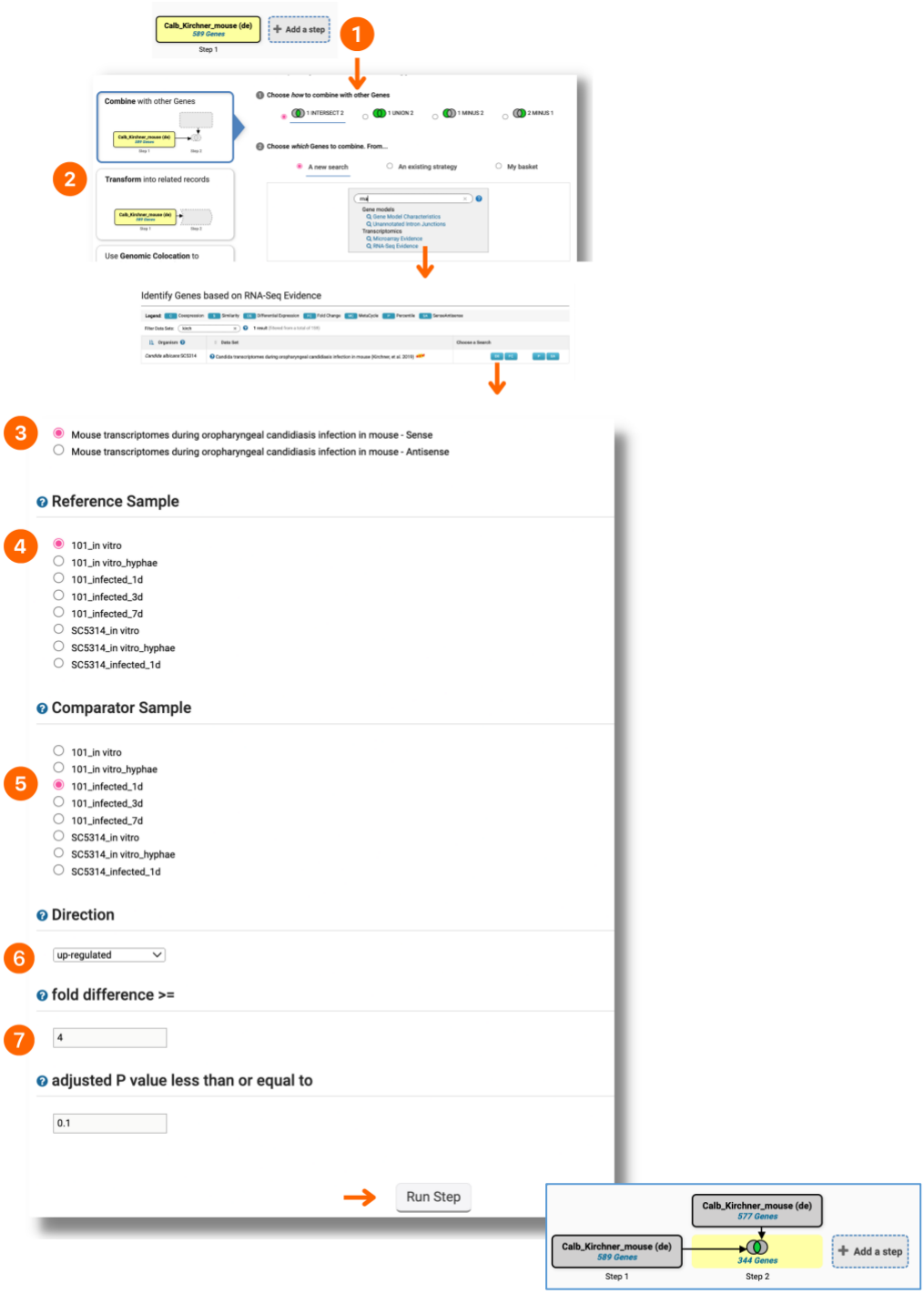
1. Identify differentially expressed genes in the virulent SC5314 strain compared to strain 101 using FungiDB.
2. Identify genes upregulated in mouse in response to the infection with SC5314 but not strain 101.

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

- **Identify genes that are up-regulated in SC5314 at 1d of infection**.
    1. Navigate to the RNA-Seq Evidence search and filter RNA-Seq datasets for "Kirch".
    2. Click on the "DE" button.
    3. Choose to examine the sense strand.
    4. Select reference sample: SC5314_in vitro.
    5. Select comparator sample: SC5314_infected_1d.
    6. Look for up-regulated genes.
    7. Select magnitude of upregulation: 4 fold.

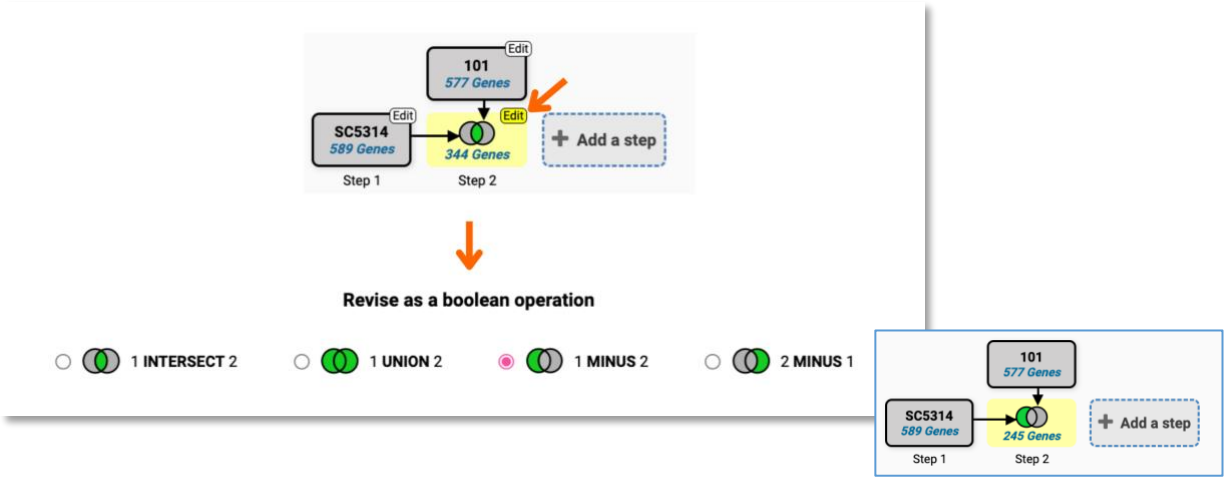


- **Identify genes that are up-regulated in SC5314 but not 101 persistent strain at 1d of infection**.
    1. Click on the “Add Step” button
    2. Navigate to the RNA-Seq Evidence search, filter for “Kirch” to quickly identify the dataset and click on the “DE” button.
    3. Choose to examine the sense strand.
    4. Select reference sample: 101_in vitro.
    5. Select comparator sample: 101_infected_1d.
    6. Look for up-regulated genes.
    7. Select magnitude of upregulation: 4 fold.

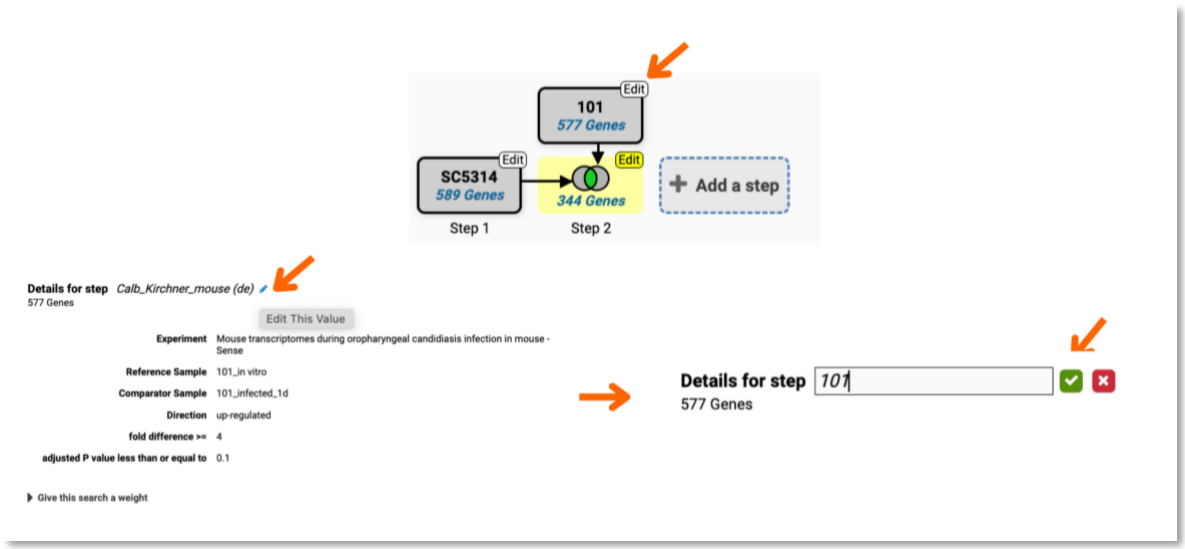


The default setting of the Boolean operators was set to the “intersect” option, which returns genes that are up-regulated by 4 fold in both strains.

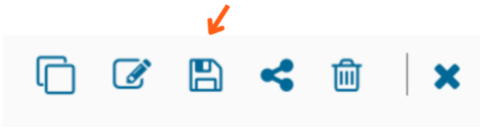
- **Change the search criteria to display genes upregulated in SC5314 only.**



Note: you can rename steps to keep track of the datasets/search results:



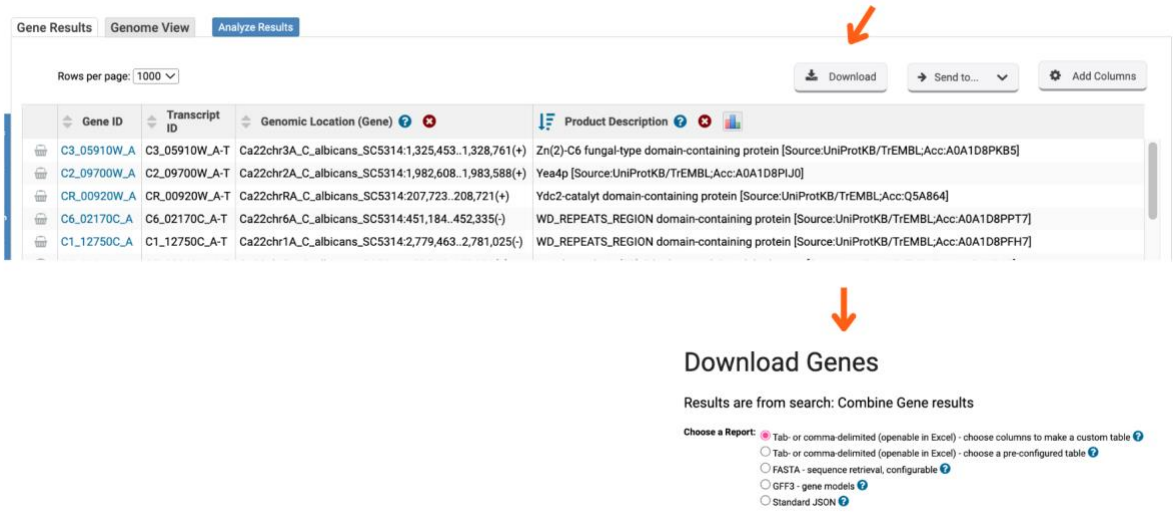
Save the strategy by clicking on the floppy disk icon on the right.



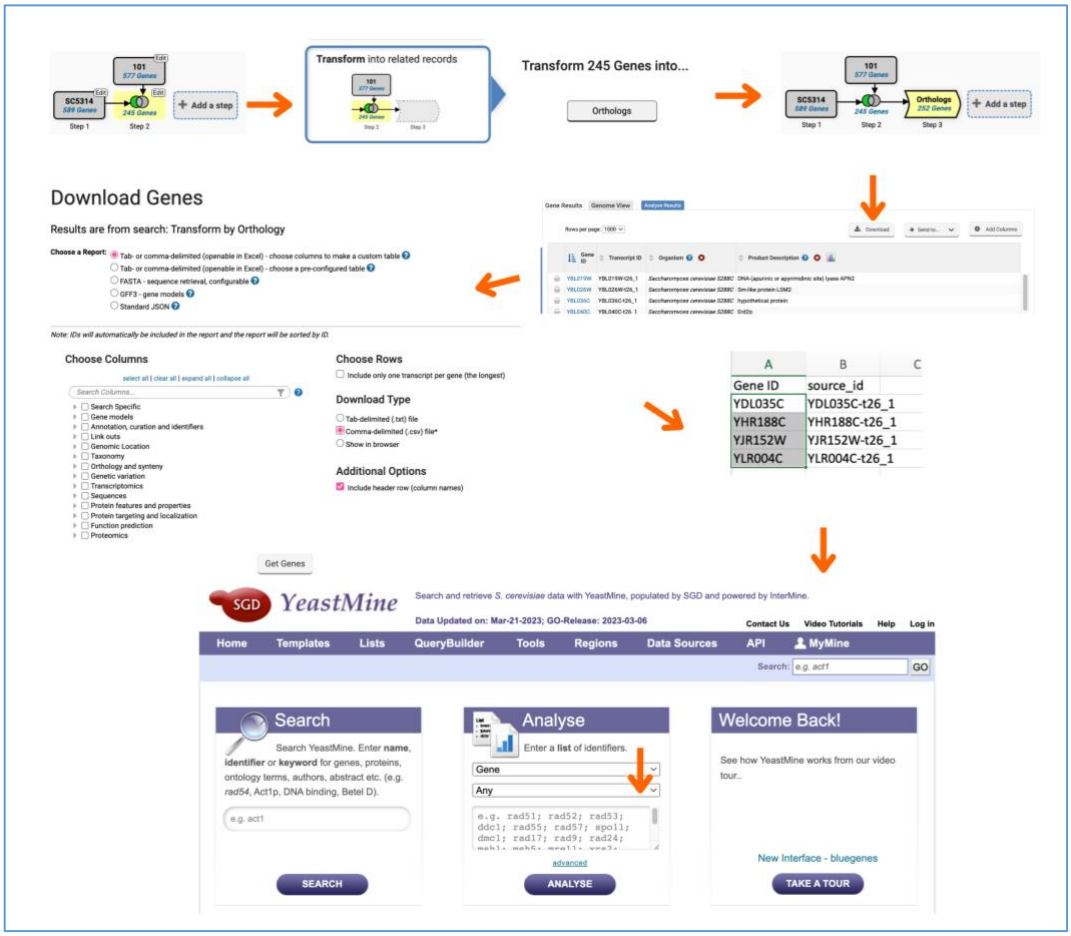
In summary, this strategy identified genes up-regulated in SC5314 when infecting mice at 1d while subtracting any genes that are also up-regulated in strain 101.

Strategy URL:
https://fungidb.org/fungidb/app/workspace/strategies/import/802d9f2b606fc1fa

Note: this data can be exported and FungiDB offers several download options that can be access by clicking on the Download button located about the results table.



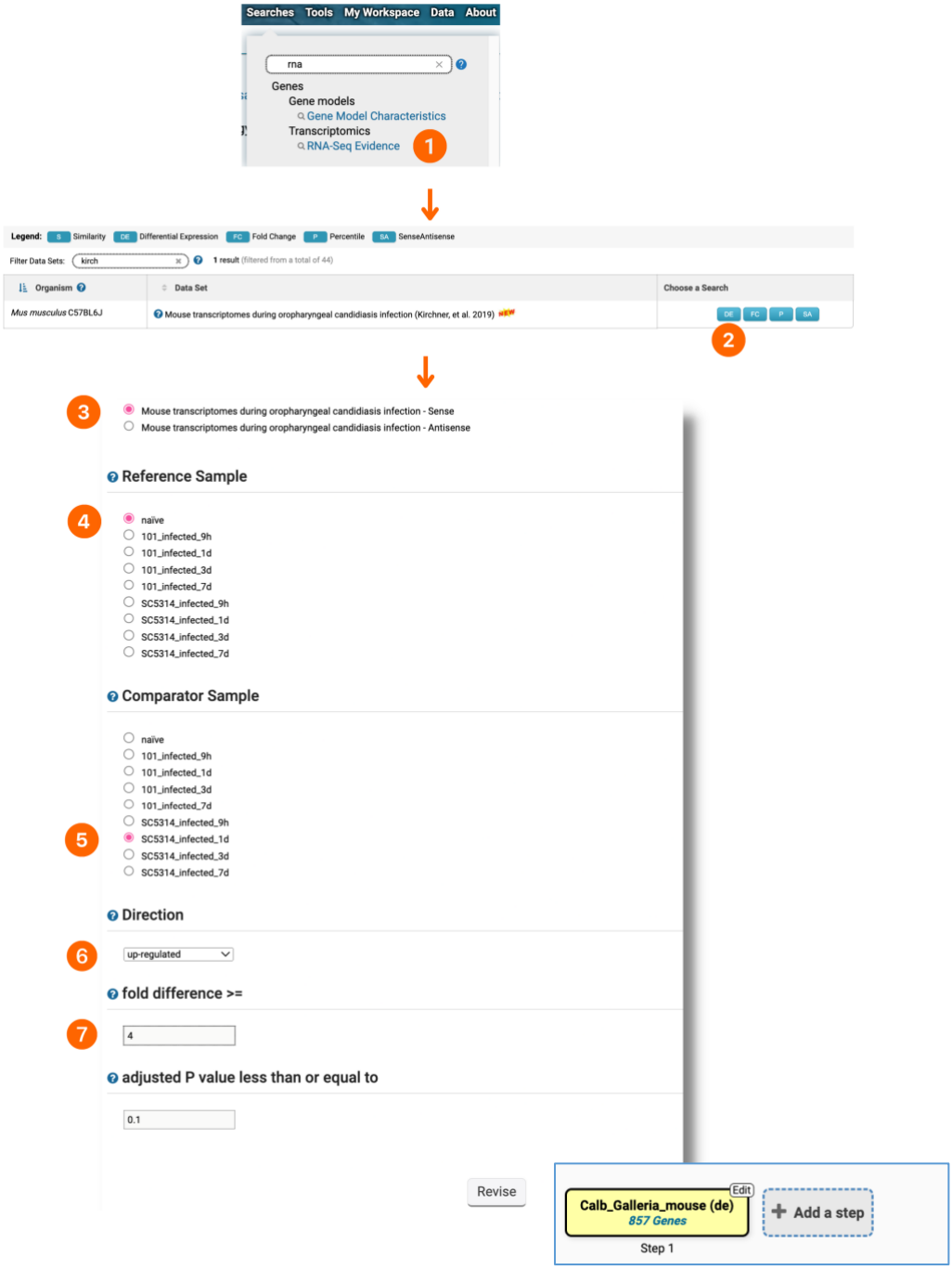
You can also export yeast orthologs of *C. albicans* genes into Yeast mine. YeastMine enables rapid retrieval and manipulation of curated biological data on yeast, which you can use to make predictions about orthologs in fungal pathogens. Here is an outline of the workflow extracting Gene IDs compatible with SGD searches:



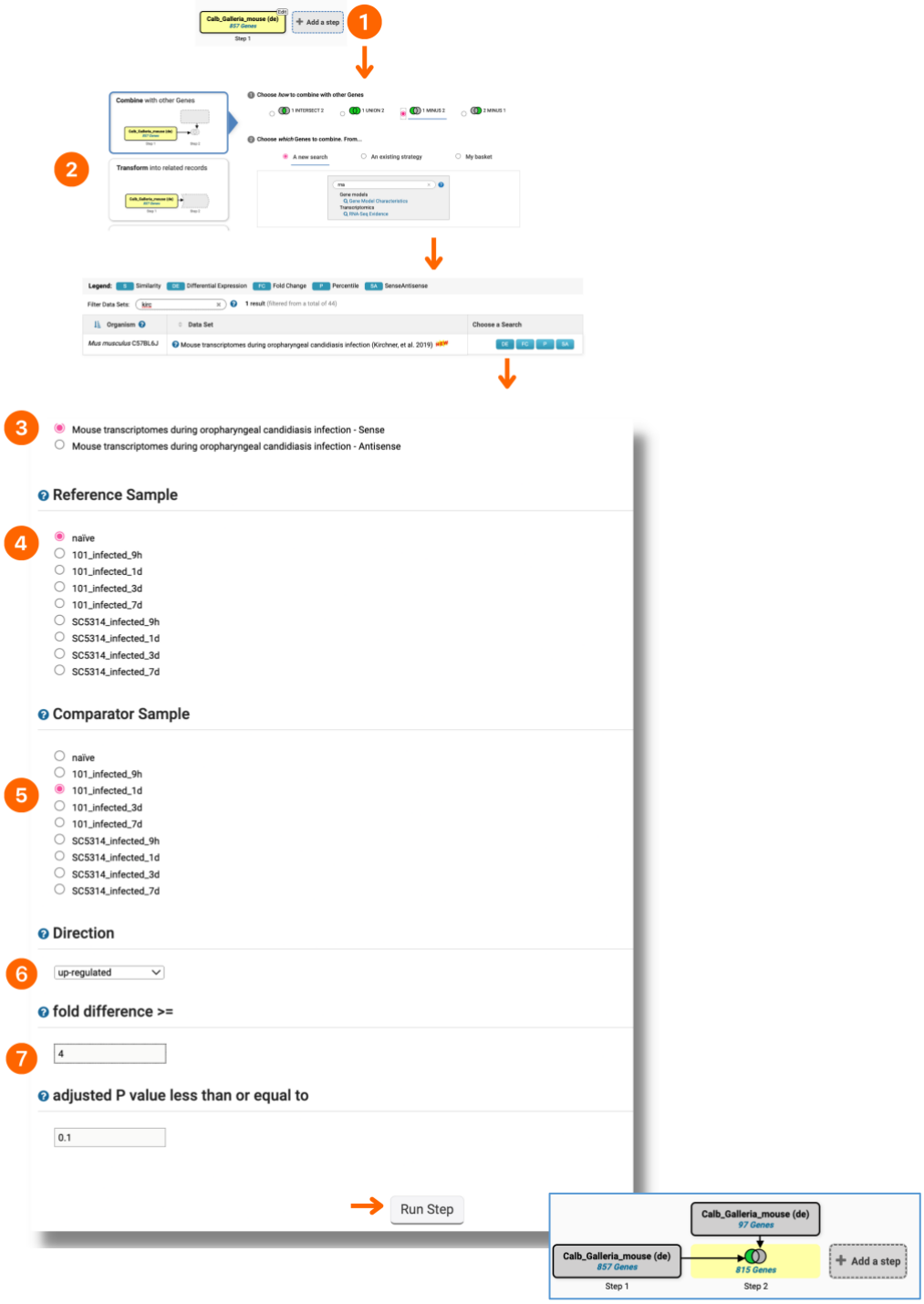
Next, we will identify gene up-regulated in mice when infected with SC5314 and 101 and select for SC5314-specific responses.

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

- **Identify genes that are up-regulated in mice infected with SC5314 at 1d**.
    1. Navigate to the RNA-Seq Evidence search and filter RNA-Seq datasets for "Kirch".
    2. Click on the "DE" button.
    3. Choose to examine the sense strand.
    4. Select reference sample: naïve.
    5. Select comparator sample: SC5314_infected_1d.
    6. Look for up-regulated genes.
    7. Select magnitude of upregulation: 4 fold.

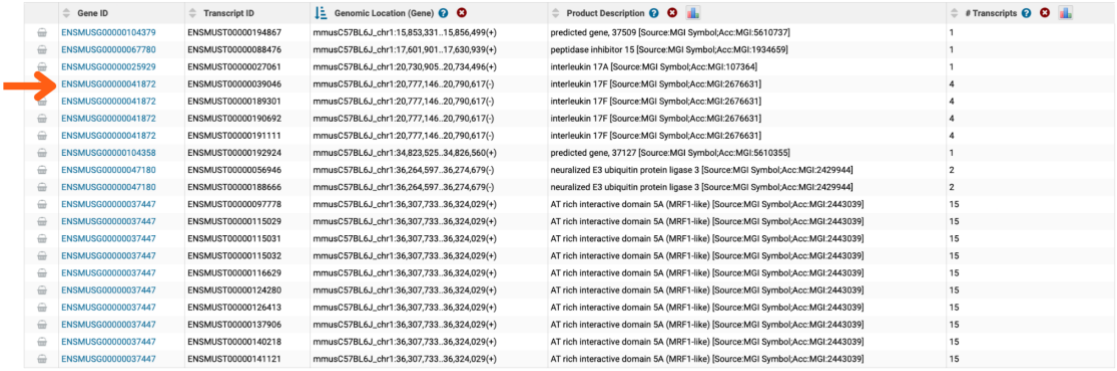


- **Identify genes that are up-regulated in SC5314 but not 101 persistent strain at 1d of infection**.
    1. Click on the “Add Step” button.
    2. Navigate to the RNA-Seq Evidence search, select “1 minus 2” Boolean operator, filter for “Kirch” to quickly identify the dataset and click on the “DE” button.
    3. Choose to examine the sense strand.
    4. Select reference sample: naïve.
    5. Select comparator sample: 101_infected_1d.
    6. Look for up-regulated genes.
    7. Select magnitude of upregulation: 4 fold.



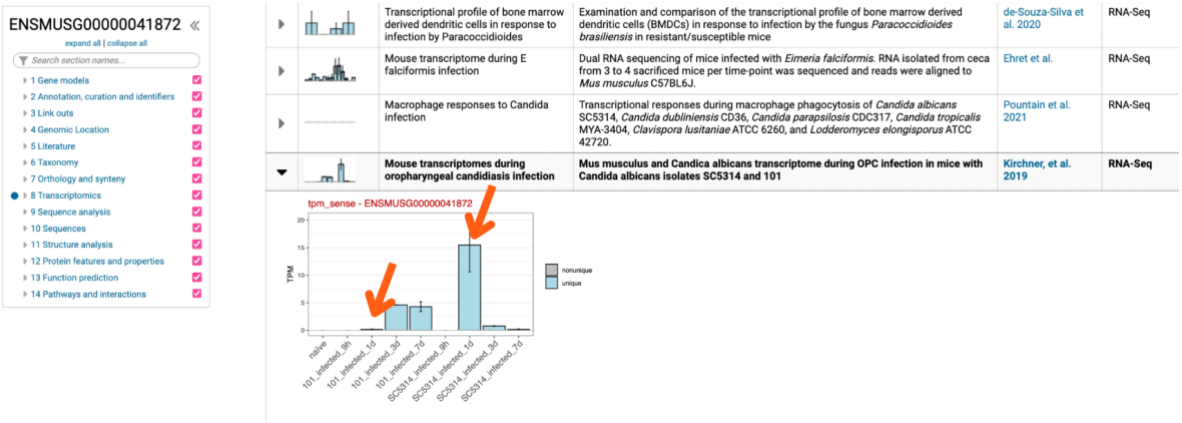
- **Examine the results in HostDB:**

    1. Click on the Gene ID link for “interleukin 17F” and navigate to the transcriptomics expression section.



Gene ID	Transcript ID	Genomic Location (Gene)	Product Description	# Transcripts
ENSMUSG00000104379	ENSMUST00000194867	mmusC57BL6J_chr1:15,853,331..15,856,499(+)	predicted gene, 37509 [Source:MGI Symbol;Acc:MGI:5610737]	1
ENSMUSG00000067780	ENSMUST00000088476	mmusC57BL6J_chr1:17,601,901..17,630,939(+)	peptidase inhibitor 15 [Source:MGI Symbol;Acc:MGI:1934659]	1
ENSMUSG00000025929	ENSMUST00000027061	mmusC57BL6J_chr1:20,730,905..20,734,496(+)	interleukin 17A [Source:MGI Symbol;Acc:MGI:107364]	1
ENSMUSG00000041872	ENSMUST00000039046	mmusC57BL6J_chr1:20,777,146..20,790,617(-)	interleukin 17F [Source:MGI Symbol;Acc:MGI:2676631]	4
ENSMUSG00000041872	ENSMUST00000189301	mmusC57BL6J_chr1:20,777,146..20,790,617(-)	interleukin 17F [Source:MGI Symbol;Acc:MGI:2676631]	4
ENSMUSG00000041872	ENSMUST00000190692	mmusC57BL6J_chr1:20,777,146..20,790,617(-)	interleukin 17F [Source:MGI Symbol;Acc:MGI:2676631]	4
ENSMUSG00000041872	ENSMUST00000191111	mmusC57BL6J_chr1:20,777,146..20,790,617(-)	interleukin 17F [Source:MGI Symbol;Acc:MGI:2676631]	4
ENSMUSG00000104358	ENSMUST00000192924	mmusC57BL6J_chr1:34,823,525..34,826,560(+)	predicted gene, 37127 [Source:MGI Symbol;Acc:MGI:5610355]	1
ENSMUSG00000047180	ENSMUST00000056946	mmusC57BL6J_chr1:36,264,597..36,274,679(-)	neuralized E3 ubiquitin protein ligase 3 [Source:MGI Symbol;Acc:MGI:2429944]	2
ENSMUSG00000047180	ENSMUST00000188666	mmusC57BL6J_chr1:36,264,597..36,274,679(-)	neuralized E3 ubiquitin protein ligase 3 [Source:MGI Symbol;Acc:MGI:2429944]	2
ENSMUSG00000037447	ENSMUST00000097778	mmusC57BL6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source:MGI Symbol;Acc:MGI:2443039]	15
ENSMUSG00000037447	ENSMUST00000115029	mmusC57BL6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source:MGI Symbol;Acc:MGI:2443039]	15
ENSMUSG00000037447	ENSMUST00000115031	mmusC57BL6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source:MGI Symbol;Acc:MGI:2443039]	15
ENSMUSG00000037447	ENSMUST00000115032	mmusC57BL6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source:MGI Symbol;Acc:MGI:2443039]	15
ENSMUSG00000037447	ENSMUST00000116629	mmusC57BL6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source:MGI Symbol;Acc:MGI:2443039]	15
ENSMUSG00000037447	ENSMUST00000124280	mmusC57BL6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source:MGI Symbol;Acc:MGI:2443039]	15
ENSMUSG00000037447	ENSMUST00000126413	mmusC57BL6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source:MGI Symbol;Acc:MGI:2443039]	15
ENSMUSG00000037447	ENSMUST00000137906	mmusC57BL6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source:MGI Symbol;Acc:MGI:2443039]	15
ENSMUSG00000037447	ENSMUST00000140218	mmusC57BL6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source:MGI Symbol;Acc:MGI:2443039]	15
ENSMUSG00000037447	ENSMUST00000141121	mmusC57BL6J_chr1:36,307,733..36,324,029(+)	AT rich interactive domain 5A (MRF1-like) [Source:MGI Symbol;Acc:MGI:2443039]	15

In summary, we identified genes upregulated in response to SC5314 infection. Notice that the interleukin 17F response is much stronger at 1d in response SC5314 infection. This is consistent with mouse response to *C. albicans* strain 101 being delayed compared to strain SC5314. Now, you may want to go back and look at gene enrichment signatures in fungi to learn more about SC5314 and 101-driven responses.



Strategy URL: https://hostdb.org/hostdb/app/workspace/strategies/import/de6763c0b7f9916c

Dataset reference: Kirchner et al. 2019 DOI: 10.3389/fimmu.2019.00330

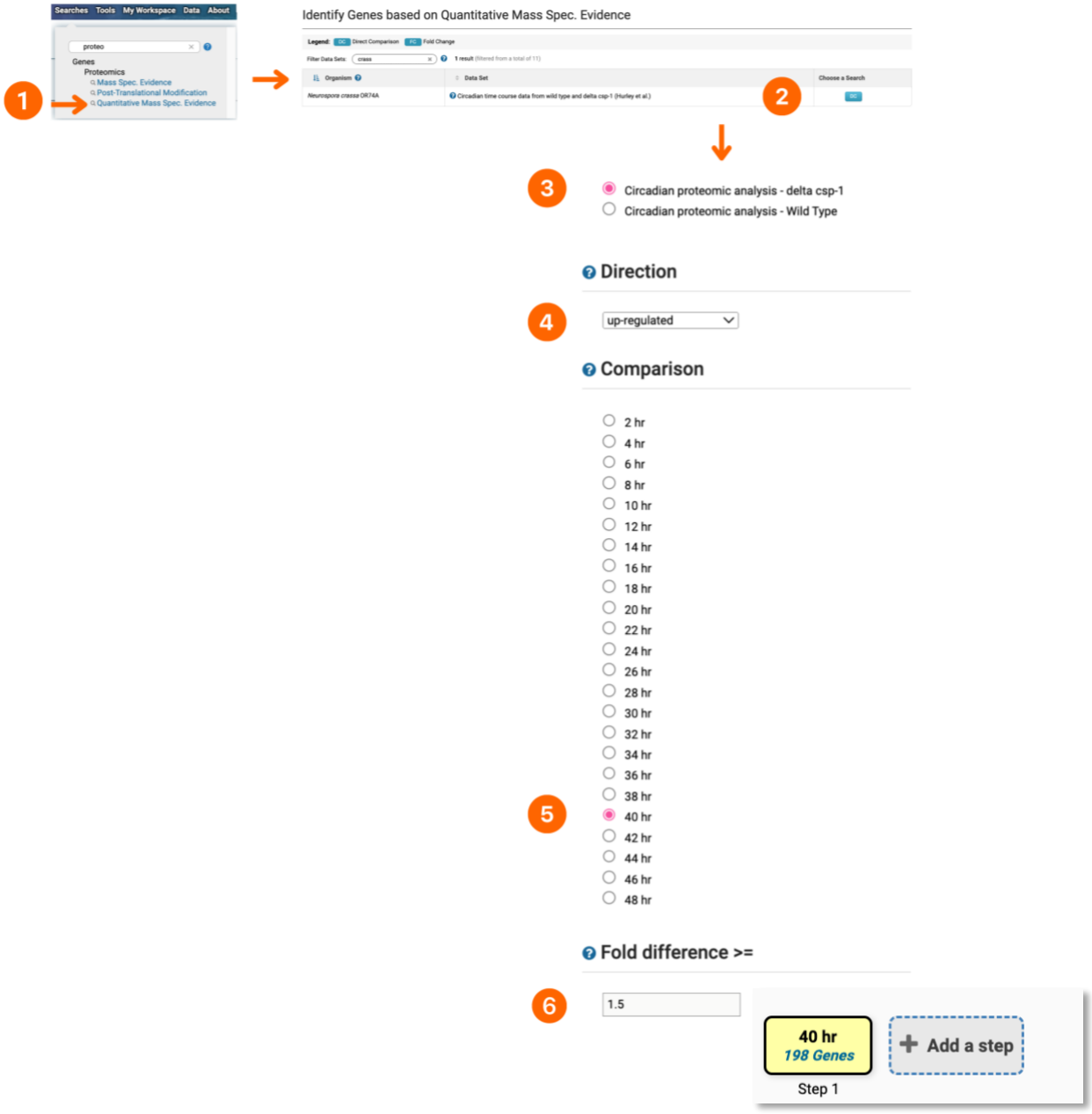
Proteomics

Learning objectives:

- Query proteomics data for *N. crassa* (e.g., genes upregulated between 40 and 46hr of incubation) and map results to *N. crassa* knockout phenotypes.

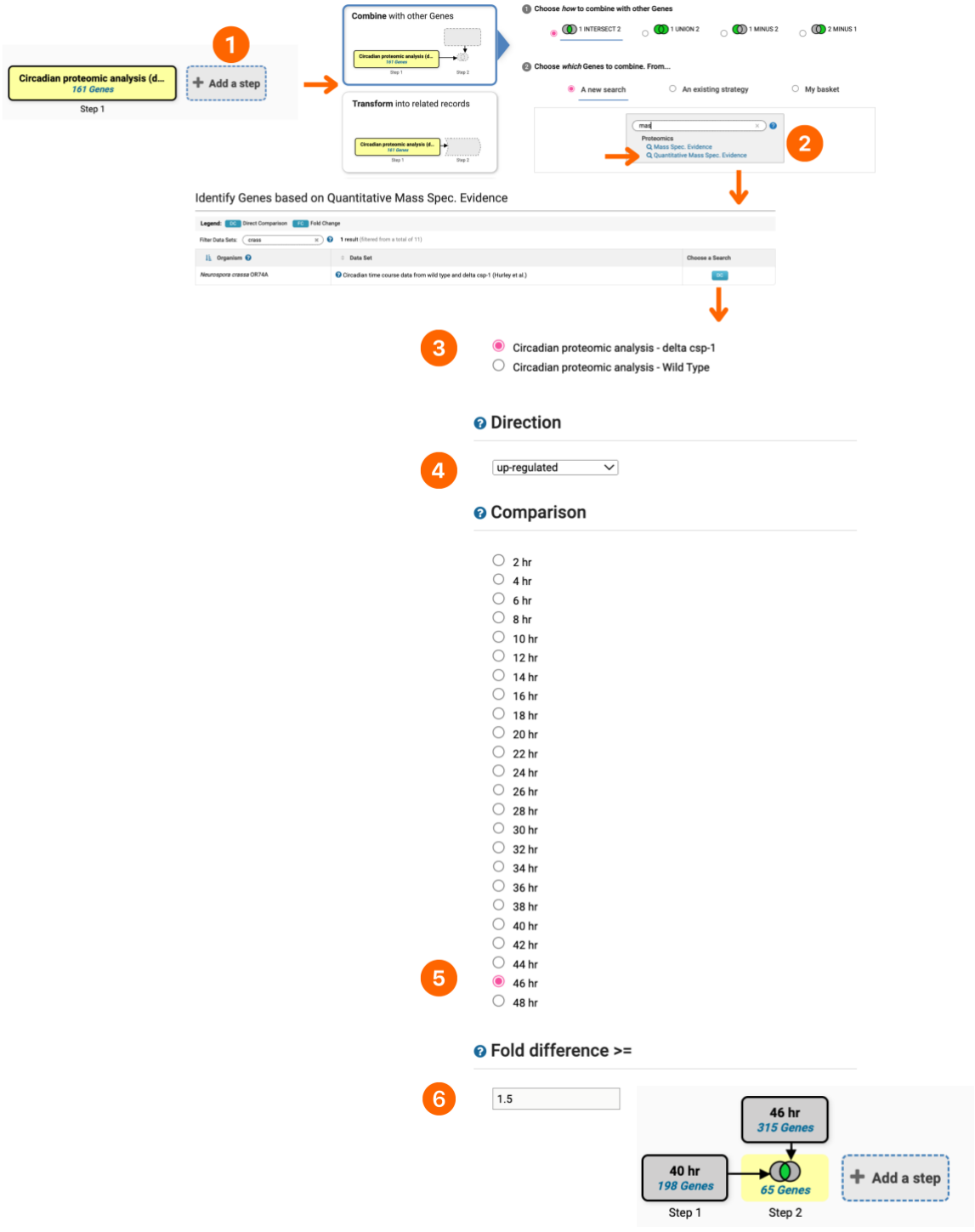
- **Identify proteins expressed in culture at 40 hr.**

1. Navigate to the “Quantitative Mass Spec. Evidence” search.
2. Click on the “DC” button for Hurley et al. 2019 dataset.
3. Select delta-csp1 mutant.
4. Choose to look for up-regulated genes.
5. Set Comparison to 40hr.
6. Leave the fold difference parameter at default.



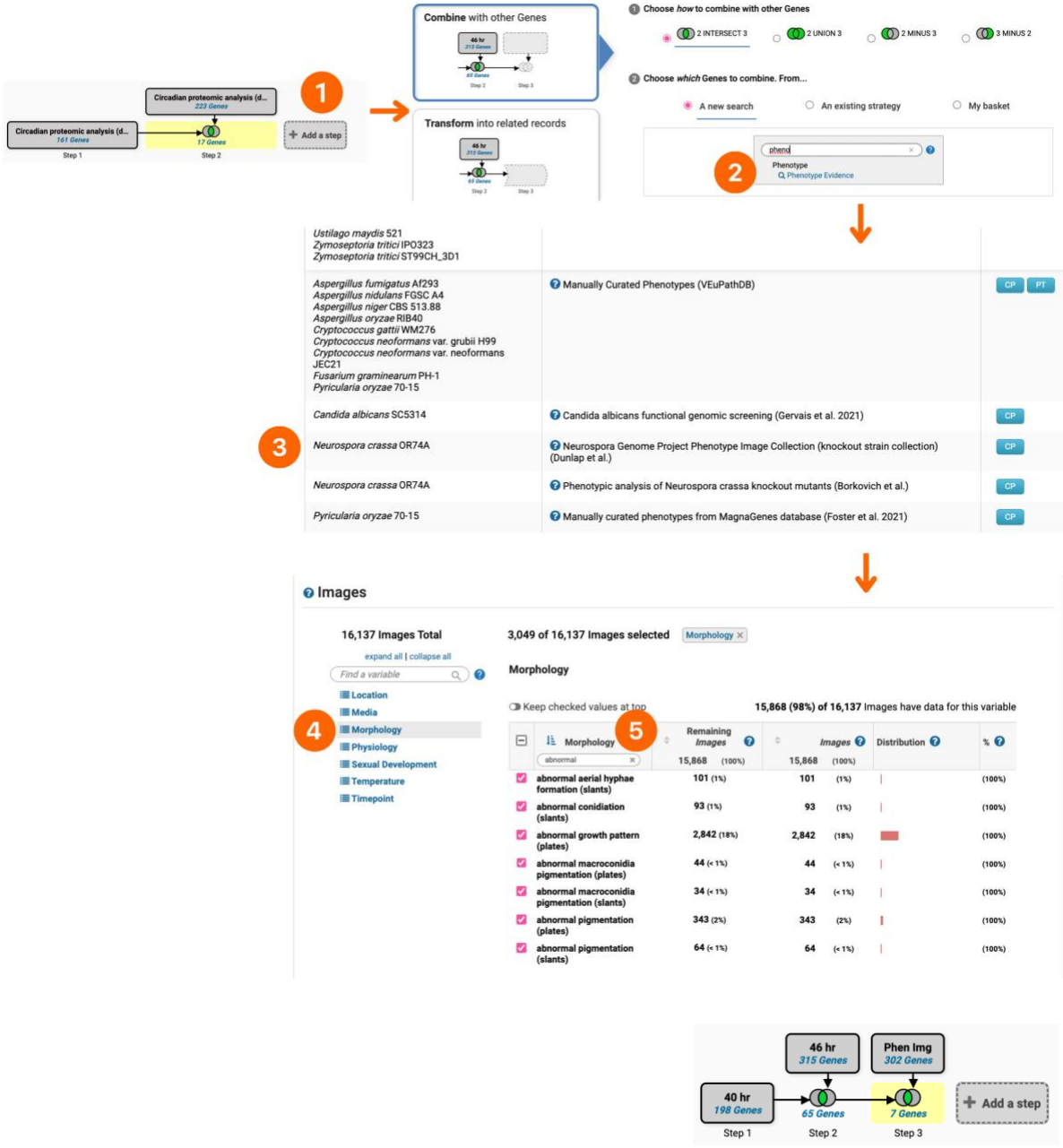
- **Identify proteins expressed in culture at 46 hr.**

1. Click on the “Add step”.
2. Select the “Combine with other Genes” search and navigate to the Quantitative Mass. Spec Evidence search.
3. Click on the “DC” button for Hurley et al. 2019 dataset.
4. Select WT sample.
5. Choose to look for up-regulated genes.
6. Set Comparison to 46hr.



- **Identify genes required for normal growth morphology in *N. crassa*.**

    1. Click on the “Add step”.
    2. Select the “Combine with other Genes” search and navigate to the Phenotype Evidence search.
    3. Click on the curated phenotypes (CP) button to investigate records from Neurospora Genome Project Phenotype Image Collection (Dunlap et al.).
    4. Navigate to the “Morphology” section.
    5. Filter on “abnormal” and select all annotated abnormal phenotypes.



Strategy URL:
https://fungidb.org/fungidb/app/workspace/strategies/import/0cae335cef282483
Reference: Hurley et al. 2018 DOI: 10.1016/j.cels.2018.10.014