FungiDB: Exploring transcriptomics & proteomics datasets

Learning objectives:

- Create search strategy examining RNA-Seq data using sense/antisense search
- Create a proteomics query and save this strategy to your account.

1. Running and RNA-Seq evidence search

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 were ranked by expression level. This search enables you to find genes w/in a specified range of values. This search does allow you to search for genes with low levels of expression, however, care should be taken when making conclusions because many factors may contribute to a negative result.

Fold change (FC). Differential expression is based on the number of unique reads for each gene. 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. For example, selecting up-regulated with a fold difference of 2 will only show results where the comparator is twice that of the reference.

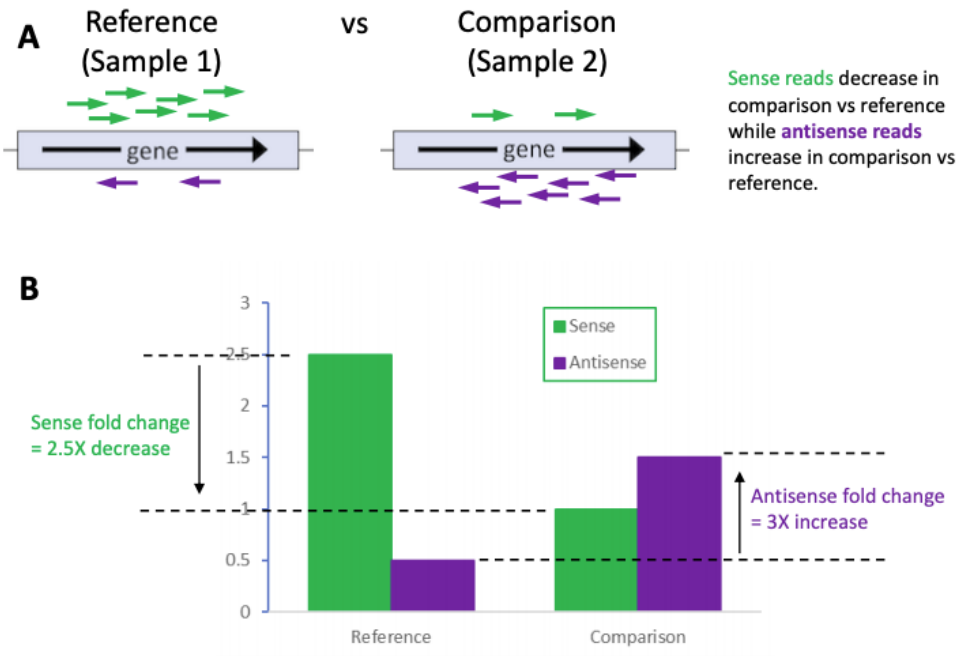
Differential Expression (DE). Find genes that are differentially expressed in a mRNASeq experiment as determined by DESeq analysis. Differential expression is based on the number of unique reads for each gene. Please choose the experiment. If only one "Experiment" is shown, this field will be selected for you. 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 with a fold difference of 2 and an adjusted p-value cut off of 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). 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.

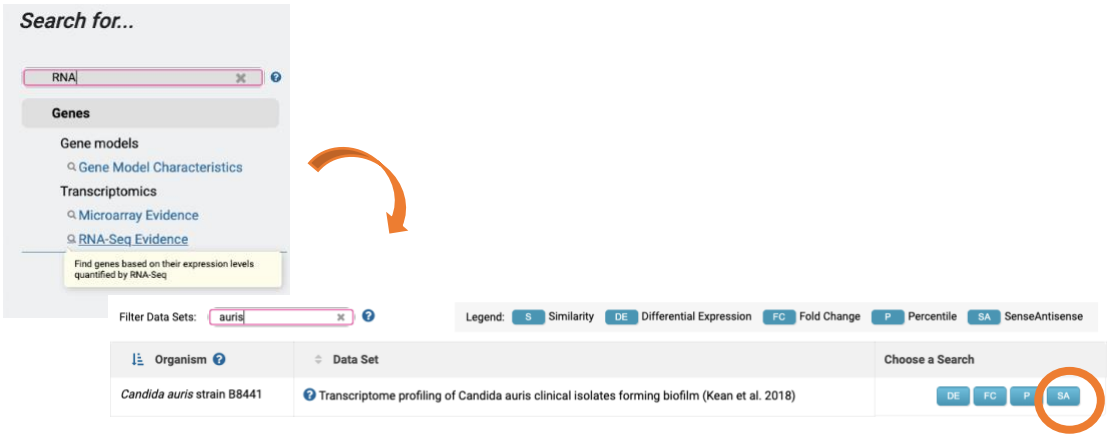
Candida auris presents a serious global health threat. Clinical isolates of *C. auris* are often resistant to multiple antifungal drugs and cells have the capacity to form adherent biofilm communities on a range of clinically important substrates.

Let’s take advantage of FungiDB sense and antisense query to identify genes that may be regulated by anti-sense transcripts during biofilm formation in the clinical isolate NCPF8973, which was isolated in a hospital in Slough, UK (Jan 2016).

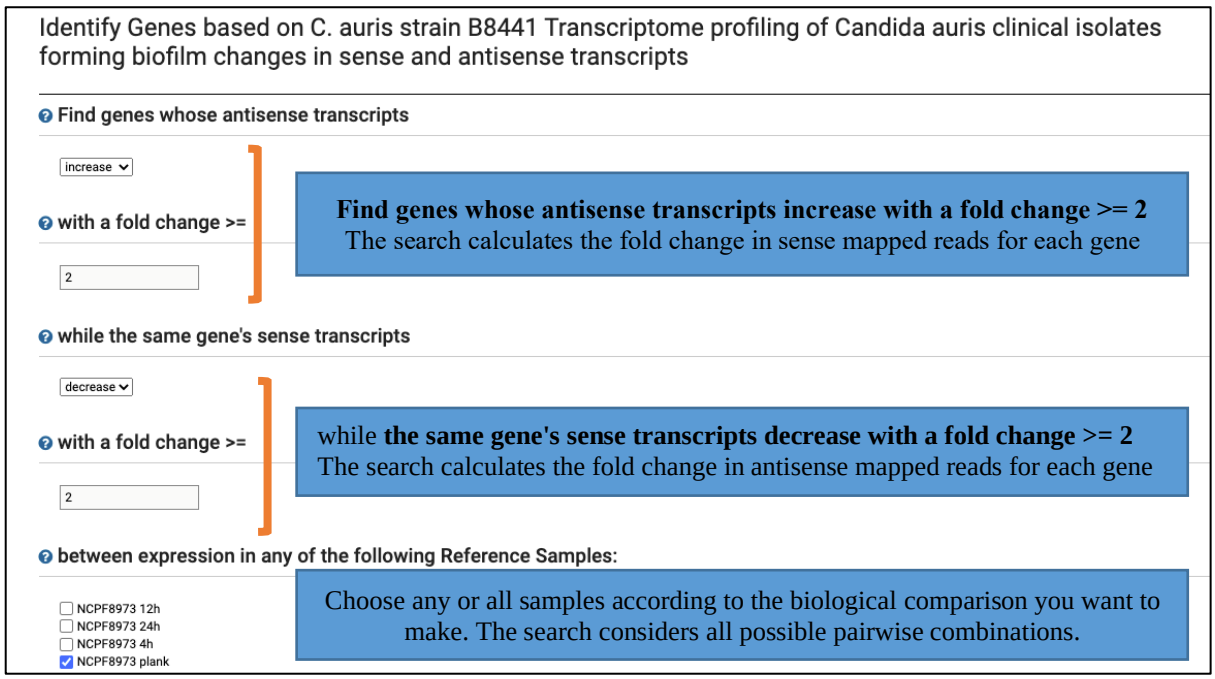
The VEuPathDB Sense/Antisense search is available for any RNA sequence data set based on strand specific data. For each gene, the search compares the fold change in sense reads to the fold change in antisense reads (A) and returns genes that have simultaneous changes in sense vs antisense expression between two samples (B). Versatile search parameters allow for choosing the direction and magnitude of the change in sense and antisense transcripts. The search result page lists genes and indicates the paired Reference -> Comparison samples that meet the criteria, and offers two graphs to help interpret the search results (TPM vs sample bar graph, antisense FC vs sense FC scatter plot).

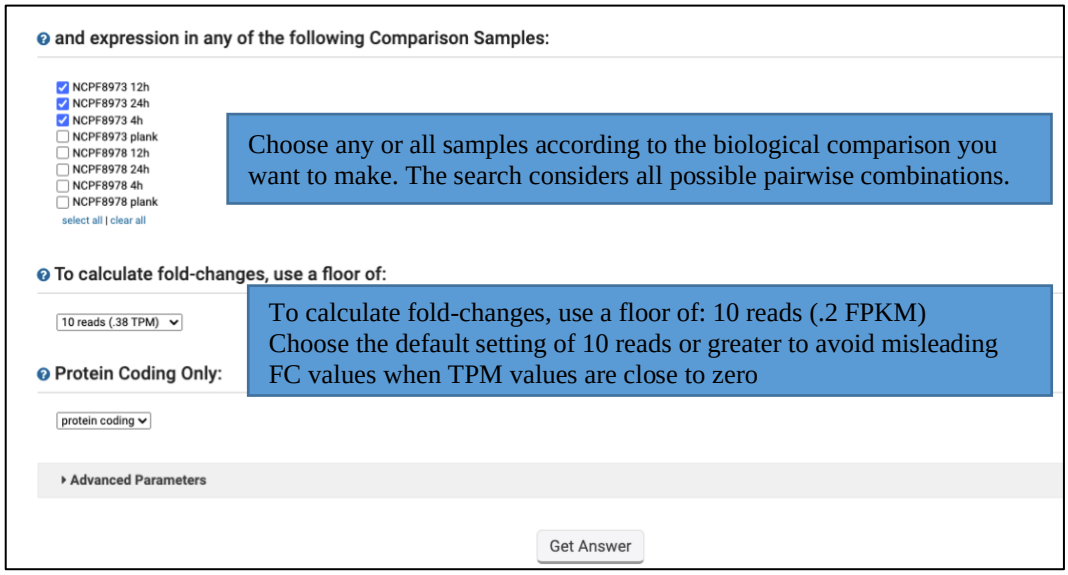


- Find the "*Transcriptome profiling of Candida auris clinical isolates forming biofilm*" RNA-Seq dataset and deploy SA query.
  Hint: Filter on "auris" to quickly ring up RNA-Seq datasets for *C. auris*.



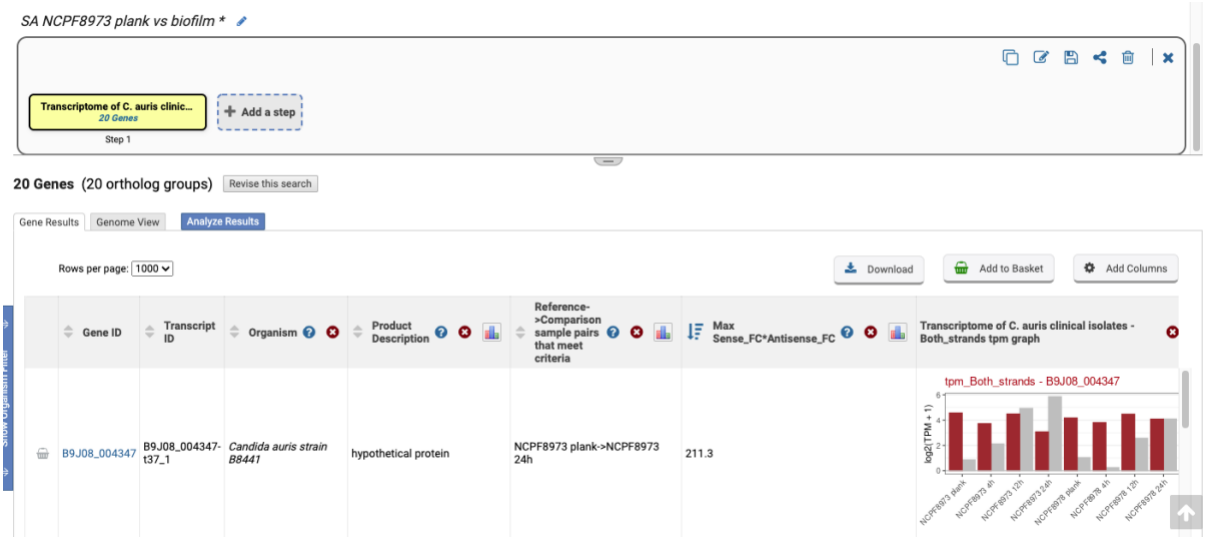
- Configure the RNA-Seq search to return protein coding genes whose *sense transcripts decrease 2 fold* while *their antisense transcripts increase 2 fold* between any time points (NCPF8973 plank (reference) vs biofilm at 4, 12, 24 hr (comparison)).





Advanced Parameters – Change in antisense fraction>= 0.5 Leave as default. The fraction of total transcripts that are antisense must change (up or down) by at least this amount. Enter a number between 0 and 1. Larger fractions will select for genes with a large fraction of antisense in one sample and small fraction of antisense in the other sample.

- Explore the results



In the results table, four search specific columns are present:

- Reference -> Comparison sample pairs that meet criteria,
- Max Sense_FC*Antisense_FC,
- Expression graph (TPM vs Sample)
- Strand specific fold change graph (log2 antisense vs log2 sense).
-

Reference->Comparison sample pairs that meet criteria – Because we arranged the search to query several samples, the search returns all possible combinations of reference and comparison that meet the criteria (2X increase antisense, 2X decrease in sense). For the gene, B9J08_004830, three sample combinations meet the criteria.

B9J08_004830	B9J08_004830-t37_1	*Candida auris strain B8441*	hypothetical protein	NCPF8973 plank->NCPF8973 12h, NCPF8973 plank->NCPF8973 24h, NCPF8973 plank->NCPF8973 4h

Max Sense_FC*Antisense_FC – This column is an in-house metric derived from multiplying 'Max sense strand fold change' and 'Max antisense fold change' for that gene. This metric can be used to roughly rank the genes according to the magnitude of differential between sense and antisense transcription.

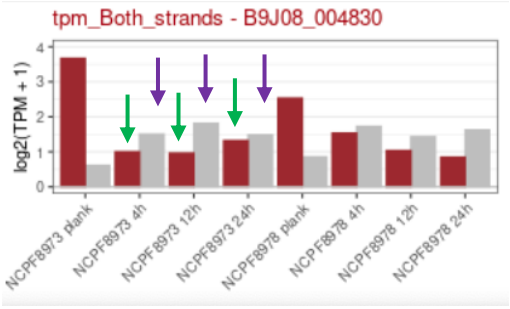
B9J08_000095	B9J08_000095-t37_1	*Candida auris strain B8441*	hypothetical protein	NCPF8973 plank->NCPF8973 24h	76.4
B9J08_004830	B9J08_004830-t37_1	*Candida auris strain B8441*	hypothetical protein	NCPF8973 plank->NCPF8973 12h, NCPF8973 plank->NCPF8973 24h, NCPF8973 plank->NCPF8973 4h	59

Both_strandsTPM graph – This graph, TPM vs sample, is a visual representation of the simultaneous changes detected by the

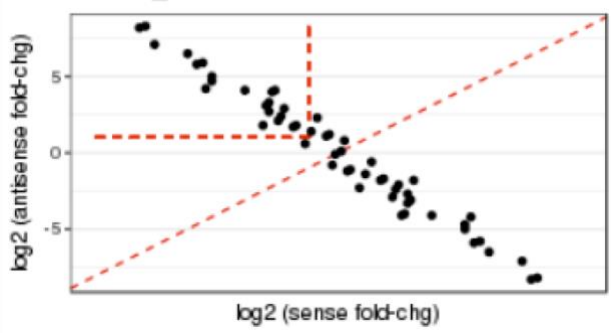
search. For example, when comparing the following samples:

- NCPF8973 plank->NCPF8973 12h,
- NCPF8973 plank->NCPF8973 24h,
- NCPF8973 plank->NCPF8973 4h

the sense transcripts decrease (green arrows) while the antisense transcripts increase (purple arrows). This graph is also available on the gene page in the transcriptomics section.



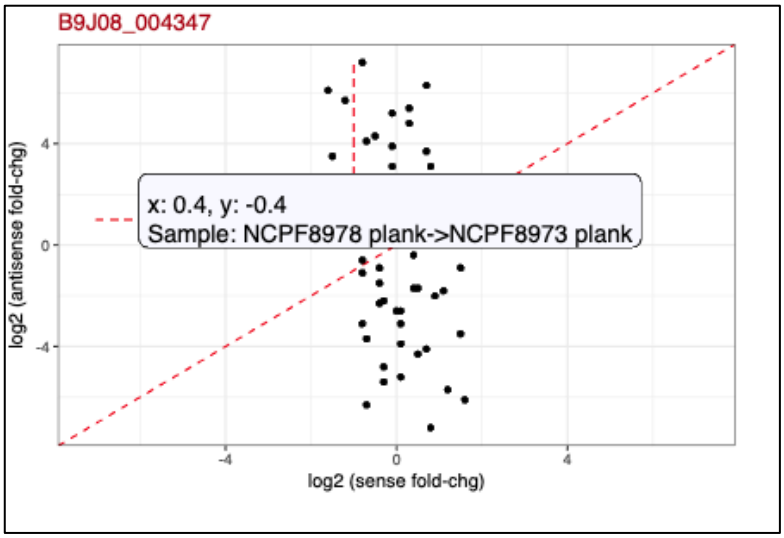
Strand_specific fold change graph – Each point shows the sense vs antisense fold change between a pair of samples.



The screenshot above shows a gene with a consistent negative correlation between sense and antisense expression, i.e. log2(antisenseFC) ~ -log2(senseFC); or as antisenseFC increases, the senseFC decreases. Clicking on the graph opens a new tab with an enlarged image which retrieves sample pair details on mouseover.

The results page lists the genes, along with each pair of samples (Reference->Comparison), that meet the criteria. It also shows a graph depicting all pairwise fold-changes in antisense and sense transcripts that is useful in identifying which samples meet the criteria that you selected AND whether this pattern is observed across all samples.

- Click on the strand specific fold change graph to examine the results more closely:



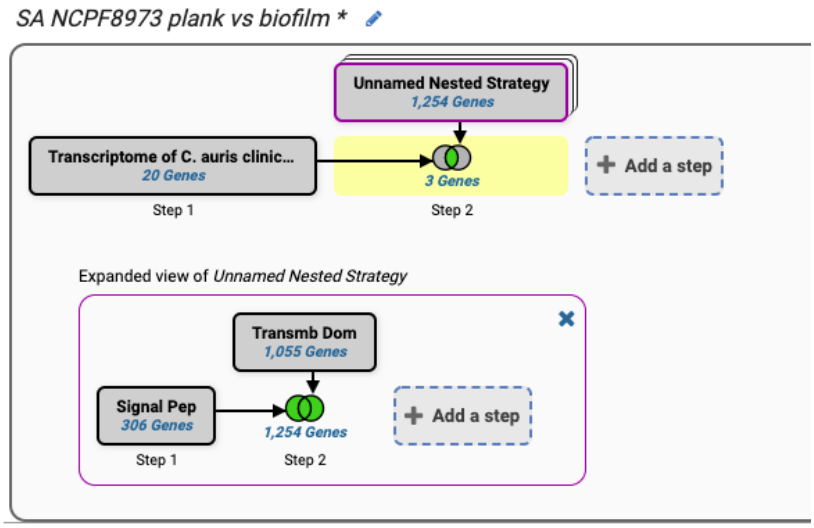
Each point shows the sense fold-change and the antisense fold-change between a pair of samples.

A point on the red dotted diagonal line represents a pair of samples where the antisense fold-change and the sense fold-change are the same.

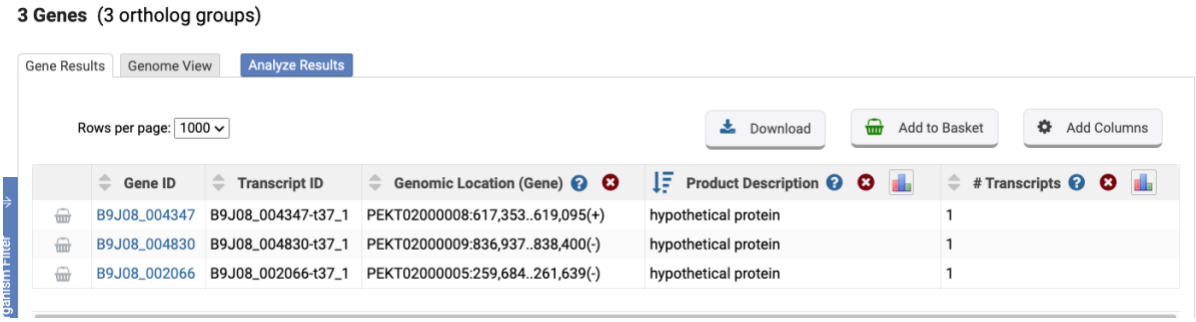
Sample pairs that meet the fold-change criteria are represented by points located within the corner of the graph that is closed off with a vertical and a horizontal red dotted line.
Note that some pairs meet the fold-change criteria but do not meet the change in antisense fraction, which is an advanced parameter. If you want to identify more genes, then you should decrease the value of this parameter.

On the Results page, you can click on the graph to open a new browser window containing an enlarged graph. The enlarged graph can be moused over to show which sample comparison is represented by each point.

- Add a step to determine if any of the genes have a predicted signal peptide and/or transmembrane domain or both (Hint: you will need to create a nested strategy)



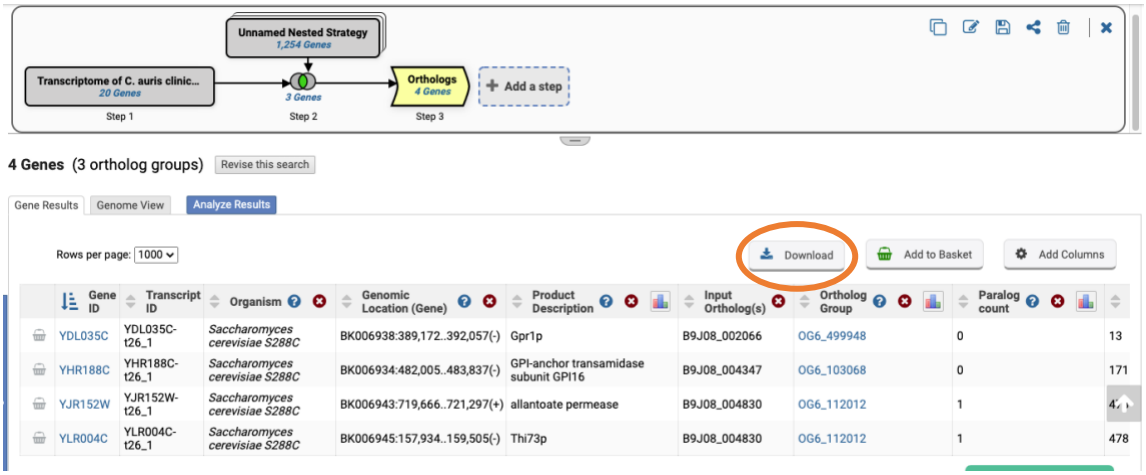
- Although all genes are hypothetical, how can you learn more about their function?

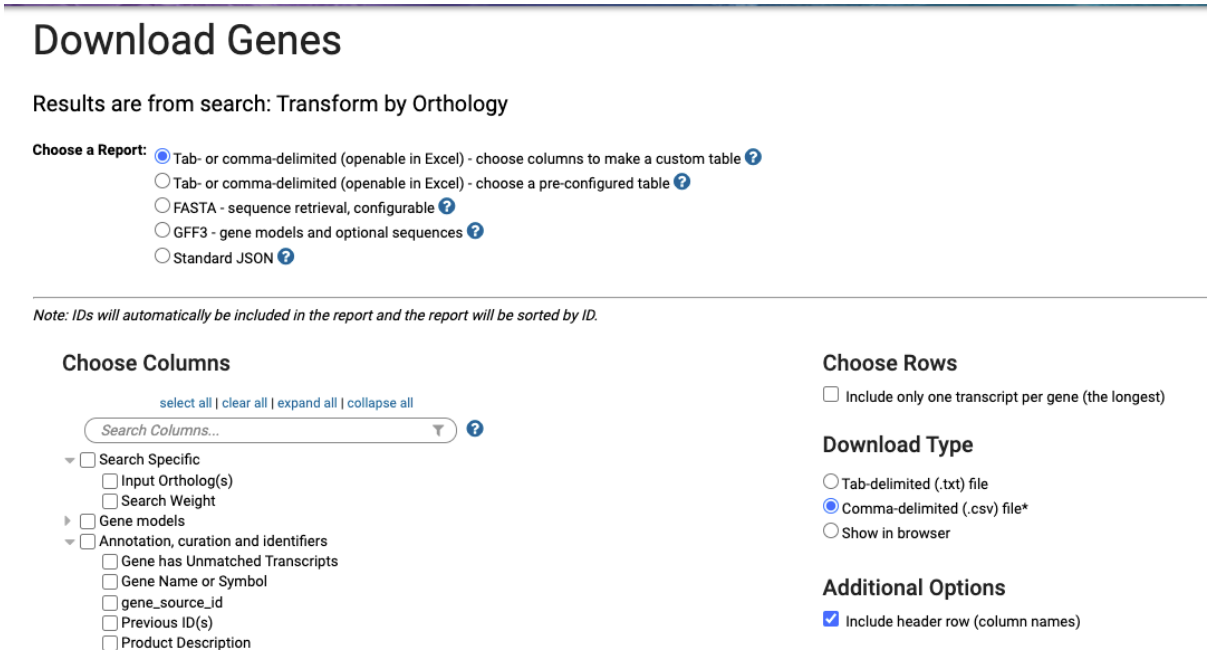


- Click on the GeneID ending with 4347 and navigate to the Orthology and synteny section within the gene record page. What is known about orthologs of this gene in other species?

- Some orthologs are annotated as "GPI transamidase". Does make sense in terms of the biology and the process examined in your search? (More about the Glycosylphosphatidylinositol biosynthetic pathway: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6056829/)

- Add a step to transform these genes into orthologs in *S. cerevisiae* S288c and extract the list by clicking on the Download button above the results table

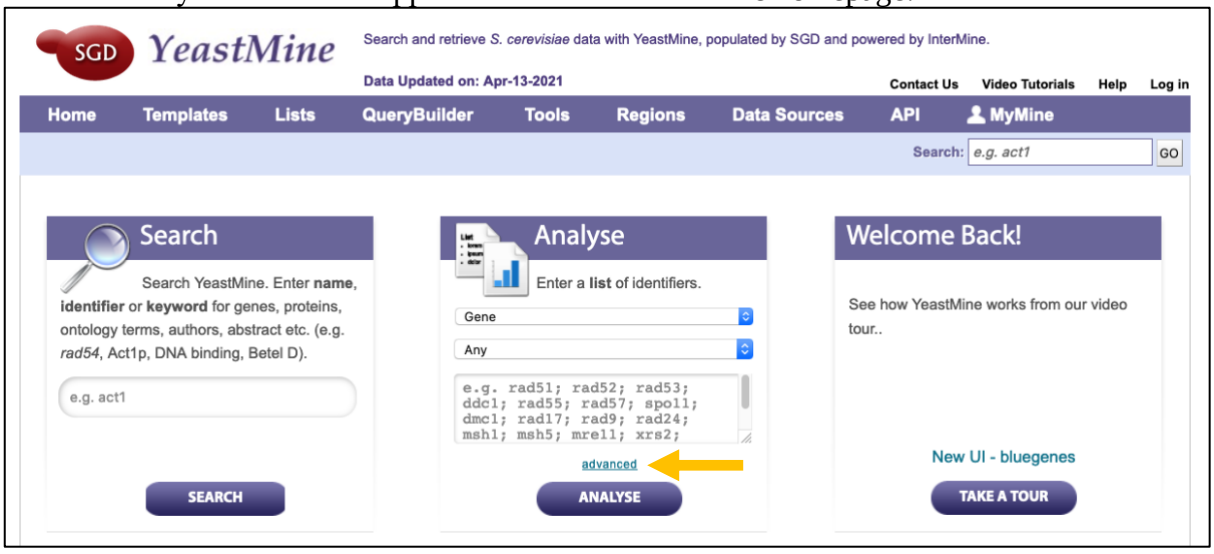




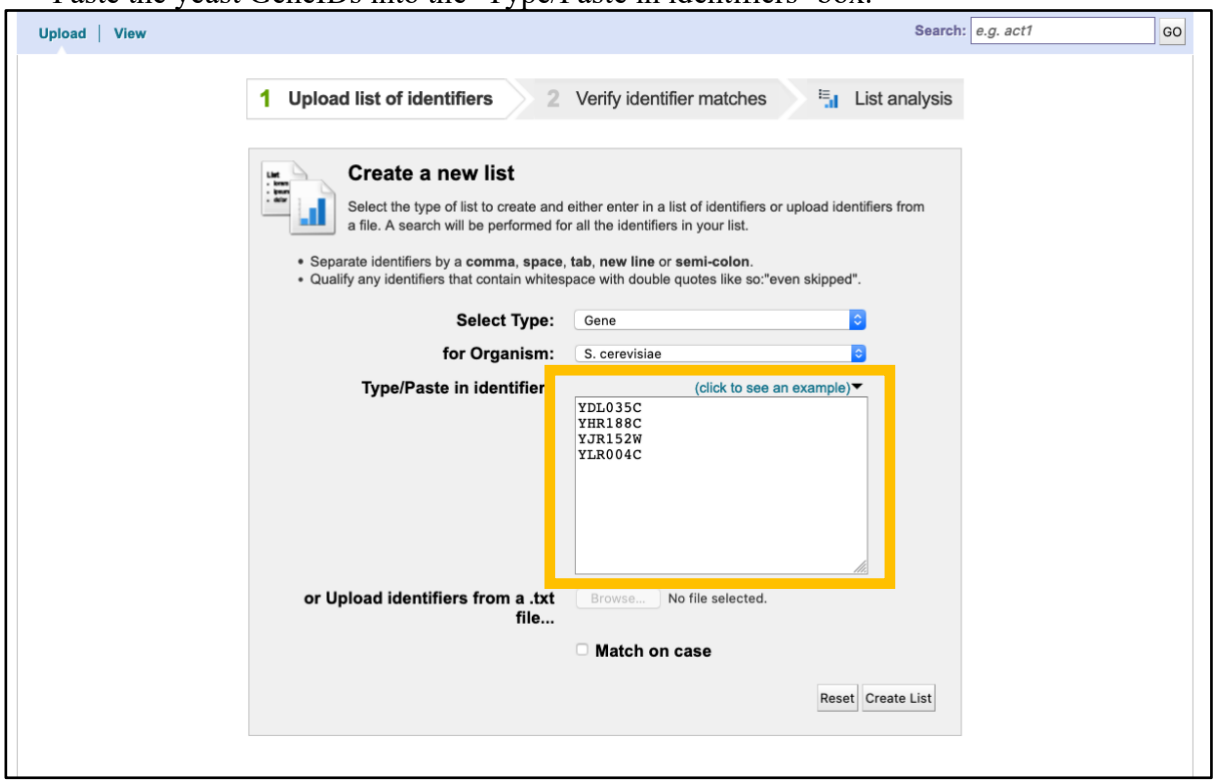
- Save and then open the comma-delimited file in Excel and copy the list of S. cerevisiae GeneIDs

	A	B	C
	Gene ID	source_id	
	YDL035C	YDL035C-t26_1	
	YHR188C	YHR188C-t26_1	
	YJR152W	YJR152W-t26_1	
	YLR004C	YLR004C-t26_1	

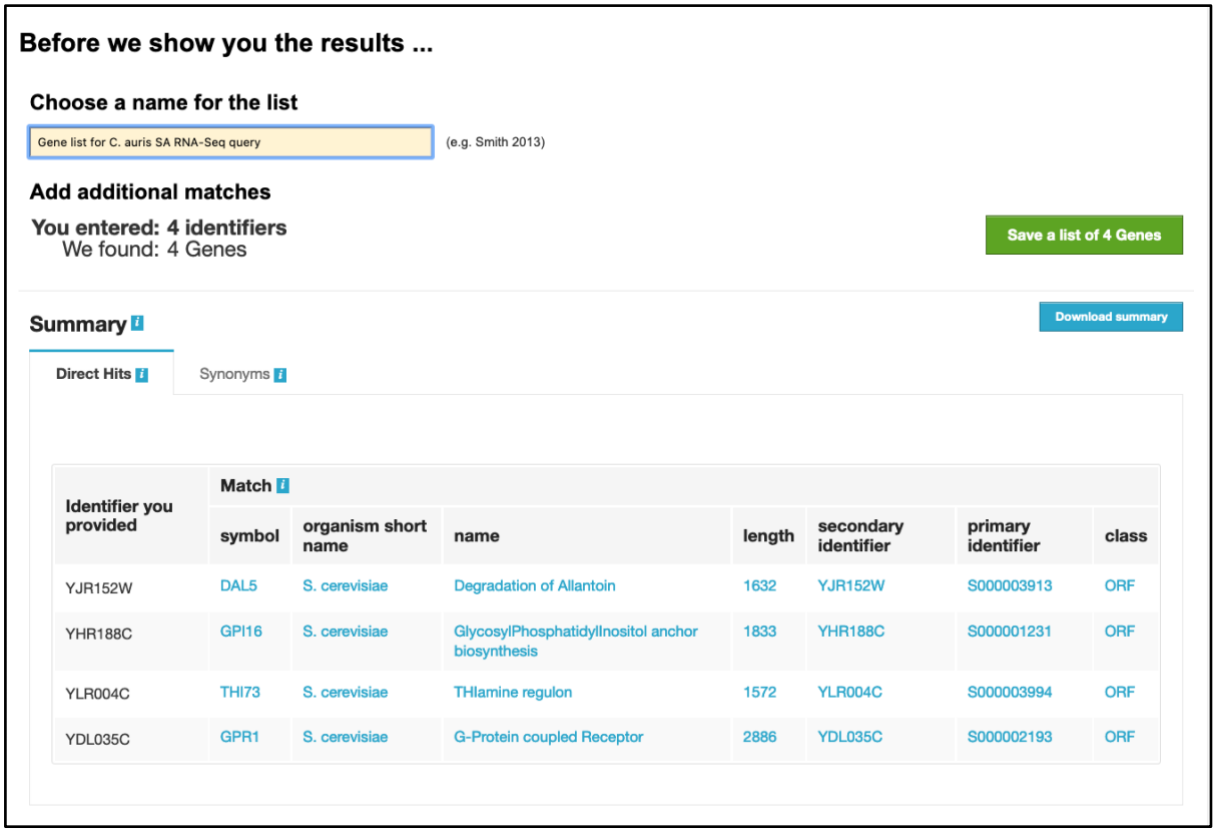
- Navigate to SGD's YeastMine (https://yeastmine.yeastgenome.org)
- To do a query specifically for *S. cerevisiae* genes, click on the blue 'advanced' link in the "Analyse" box in the upper center of the YeastMine homepage.



- Paste the yeast GeneIDs into the ‘Type/Paste in identifiers’ box.



- Click on the Create List button give your gene list a name and click on the “Save a list of 4 Genes” button.

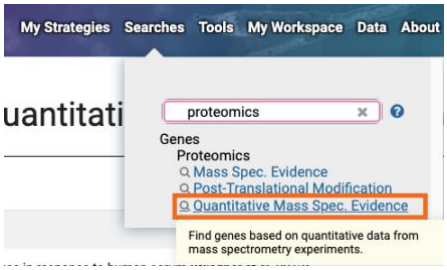


Identifier you provided	Match						
	symbol	organism short name	name	length	secondary identifier	primary identifier	class
YJR152W	DAL5	S. cerevisiae	Degradation of Allantoin	1632	YJR152W	S000003913	ORF
YHR188C	GPI16	S. cerevisiae	GlycosylPhosphatidylInositol anchor biosynthesis	1833	YHR188C	S000001231	ORF
YLR004C	THI73	S. cerevisiae	THIamine regulon	1572	YLR004C	S000003994	ORF
YDL035C	GPR1	S. cerevisiae	G-Protein coupled Receptor	2886	YDL035C	S000002193	ORF

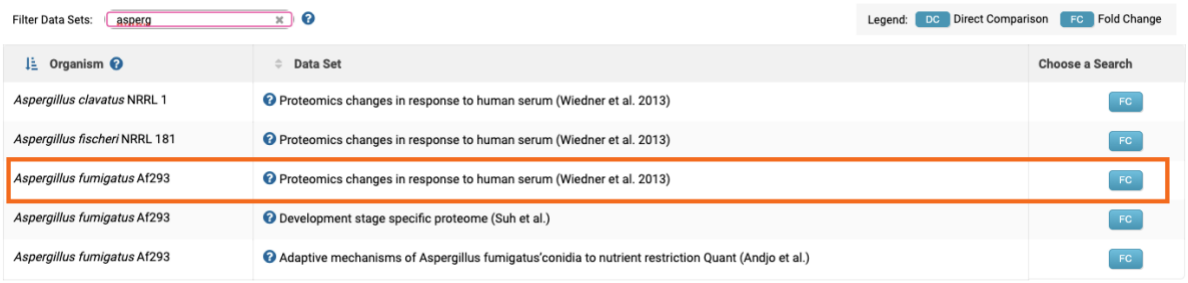
The yeast orthologs of *C. auris* genes that may be regulated by the antisense transcripts are now ready for analysis in YeastMine. YeastMine enables rapid retrieval and manipulation of curated biological data on yeast, which you can use to make predictions about orthologs in fungal pathogens. For more information on how to use YeastMine to answer complex biological questions, see the "Search Strategies in SGD" section of the exercise book.

2. Running a proteomics search

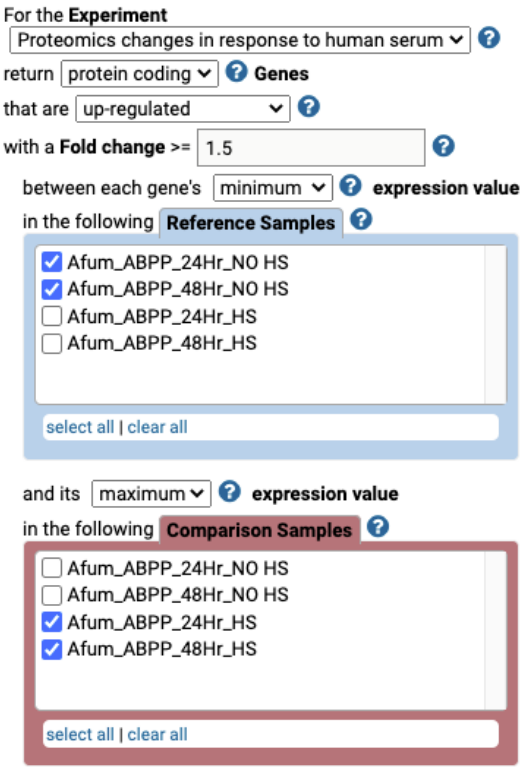
Now, let's switch gears and create a proteomics search using the "Identify Genes based on Quantitative Mass Spec. Evidence" query.



- Select the dataset titled "Proteomics changes in response to human serum (Wiedner et al. 2013)" for *Aspergillus fumigatus* Af293



- Create a search looking for up-regulated genes when *A. fumigatus* is exposed to blood media



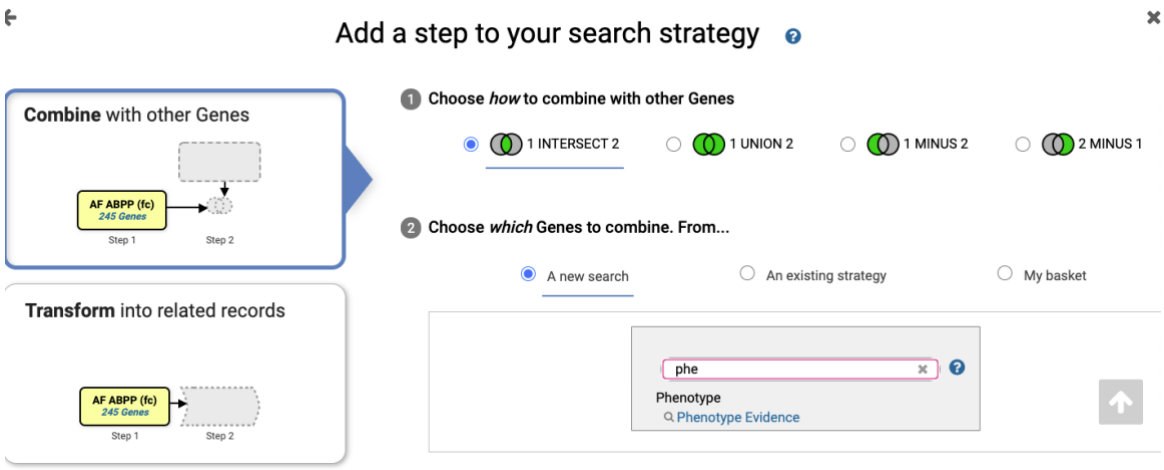
After selecting samples, you have the option to take the average, minimum, or maximum expression value within each group. (If choosing only one sample from a group, the selected 'operation' will not affect your results).

Time series experiments will offer an extra parameter called "Global min/max" which allows you to filter your results further.

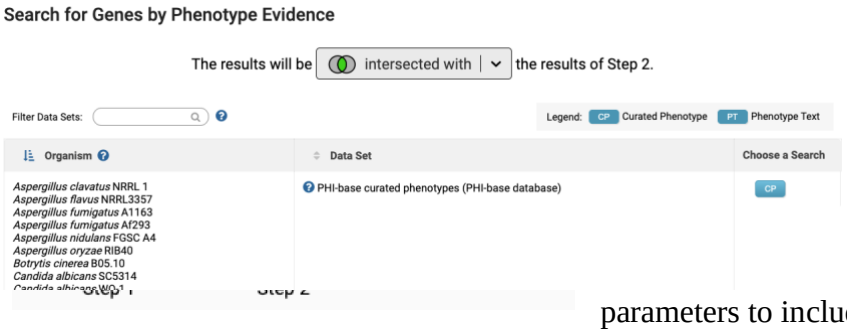
Finally, you can choose the directionality and the magnitude of the difference. For example, selecting up-regulated with a fold difference of 2 will only show results where the comparator is twice that of the reference.

Note: Any values less than the floor of 0.01 are raised to 0.01 to compute the fold difference.

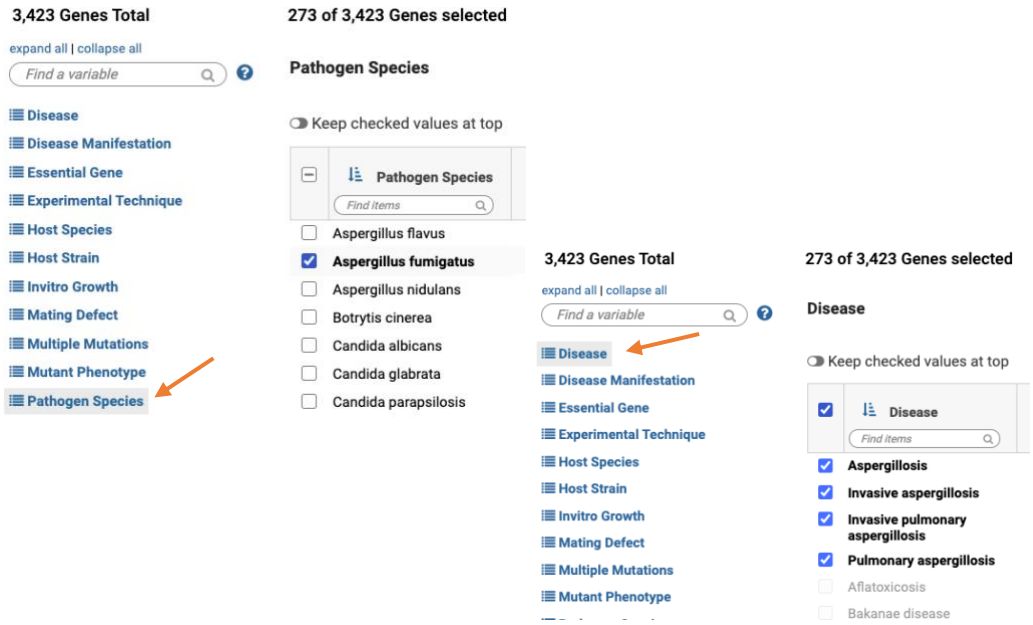
- How can you search for genes that have annotations related to virulence or disease phenotype?
    - Add a step and select the Phenotype Evidence search



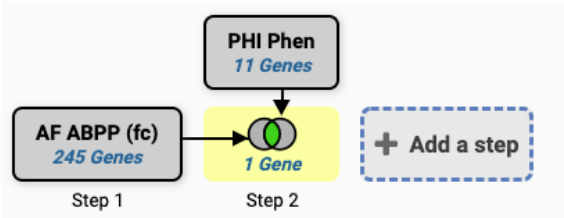
- Click on CP within the dataset titled “PHI-base curated phenotypes (PHI-base database)”



- Set search parameters to include Pathogen Species > Aspergillus fumigatus and then choose all annotations under the Disease metadata parameter



- How can you find genes that are also annotated with hypervirulence phenotype? Hint: Click on Edit the step and add a new metadata parameter for “increased virulence (hypervirulence)”, which is located under the “Mutant Phenotype” option on the left.



- Name and save your strategy.