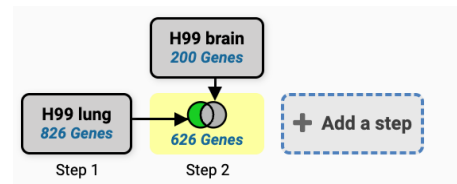# FungiDB: Performing GO Enrichment analysis

Learning objectives:
- Perform a GO enrichment analysis
- Create complex search strategy using both FungiDB and SGD

In this exercise, we will come back to the two strategies created previously (ACE2 and ACE3).

- **In FungiDB: Perform gene enrichment analysis on genes upregulated during lung infection _only in Cryptococcus neoformans._**
  1. Use Boolean operator "1 minus 2" to identify gene uniquely expressed the lung samples.
 2. Modify the first search step to return gene upregulated by 8 fold.

Strategy URL:
https://fungidb.org/fungidb/app/workspace/strategies/import/88d4eabdafd0b451

- **In HostDB: Perform gene enrichment analysis on genes upregulated during lung infection _only in Mus musculus._**
  1. Repeat steps above for HostDB.org strategy.

Strategy URL:
https://hostdb.org/hostdb/app/workspace/strategies/import/9976d08eede6ceb9
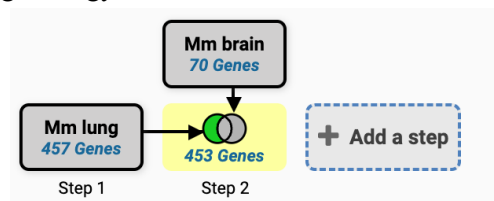
The enrichment analysis tools can be accessed under the blue Analyze Results tab:
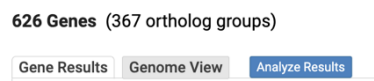

The enrichment analysis includes Gene Ontology, Metabolic Pathway, and Word Enrichment tools. The three types of analysis apply Fisher's Exact test to evaluate ontology terms, over-represented pathways, and product description terms. Enrichment is carried out using a Fisher's Exact test with the background defined as all genes from the organism being queried. P-values corrected for multiple testing are provided using both the Benjamini-Hochberg false discovery rate method and the Bonferroni method.

- **Deploy GO enrichment analysis by clicking on the "Gene Ontology Enrichment" button.**

GO

Analyze your Gene results with a tool below.

| Gene Ontology Enrichment | Metabolic Pathway Enrichment | Word Enrichment |
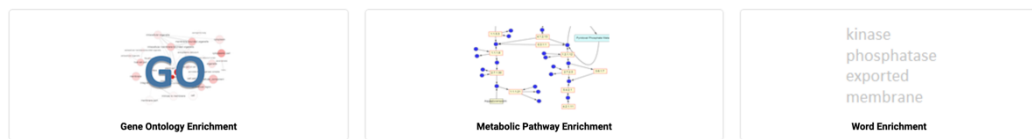
enrichment analysis can be performed on the following ontology groups: molecular function, cellular component, and biological processes. Also, other parameters allow users to limit their analysis on either "Curated" or "Computed" annotations, or both. Those with a GO evidence code inferred from electronic annotation (IEA) are denoted "Computed", while all others have some degree of curation. The default P-value is set to 0.05 but can be adjusted manually.

When the GO Slim option is chosen, both the genes of interest and the background are limited to GO terms that are part of the generic GO Slim subset.

- **Perform GO enrichment analysis (Biological Process) at FungiDB and HostDB using default settings.**

Analysis Results:

43 rows

GO ID	GO Term	Genes in the bkgd with this term	Genes in your result with this term	Percent of bkgd genes in your result	Fold enrichment	Odds ratio	P-value	Benjamini	Bonferroni
GO:0071840	cellular component organization or biogenesis	457	29	6.3	1.50	1.62	1.59e-2	3.88e-1	1.00e+0
GO:0055085	transmembrane transport	350	25	7.1	1.69	1.84	6.09e-3	1.93e-1	1.00e+0
GO:0042254	ribosome biogenesis								
GO:0022613	ribonucleoprotein complex biogenesis								
GO:0044085	cellular component biogenesis								

GO ID	GO Term	Genes in the bkgd with this term	Genes in your result with this term	Percent of bkgd genes in your result	Fold enrichment	Odds ratio	P-value	Benjamini	Bonferroni
GO:0006952	defense response	1366	83	6.1	4.37	5.94	2.03e-31	8.17e-28	8.17e-28
GO:0006954	inflammatory response	464	49	10.6	7.60	9.80	7.59e-29	1.53e-25	3.05e-25
GO:0051707	response to other organism	1437	77	5.4	3.86	5.04	1.72e-25	1.85e-22	6.94e-22
GO:0043207	response to external biotic stimulus	1440	77	5.3	3.85	5.03	1.97e-25	1.85e-22	7.92e-22
GO:0002376	immune system process	2203	96	4.4	3.14	4.27	2.30e-25	1.85e-22	9.26e-22

The results table includes several additional statistical measurements:
- **Fold enrichment** - The ratio of the proportion of genes in the list of interest with a specific GO term over the proportion of genes in the background with that term.
- **Odds ratio** - Determines if the odds of the GO term appearing in the list of interest are the same as that for the background list.
- **P-value** - Assumptions under a null hypothesis, the probability of getting a result that is equal or greater than what was observed.
- **Benjamini-Hochburg false discovery rate** - A method for controlling false discovery rates for type 1 errors.
- **Bonferroni adjusted P-values** - A method for correcting significance based on multiple comparisons.

The GO enrichment table can be opened in Revigo, viewed as a word cloud (produced via the GO Summaries R package) or downloaded.

Notice that the table contains columns with GO IDs and GO terms along with the number of genes in the background and those specific to the RNA-Seq analysis results presented (linked in blue).

- Examine GO enrichment analysis results. What kinds of GO terms are enriched?

Note: you can sort genes in your results using the sort options within a column:

- Visualize the results in Revigo by clicking on the Revigo button above the results table and leaving other parameters at default. Click the Start Revigo button below the results set and then select scatterplot.

Bubble color corresponds to the FungiDB/HostDB-provided p-value (see legend in upper right-hand corner)

Bubble size represent the frequency of the GO term in the underlying database.
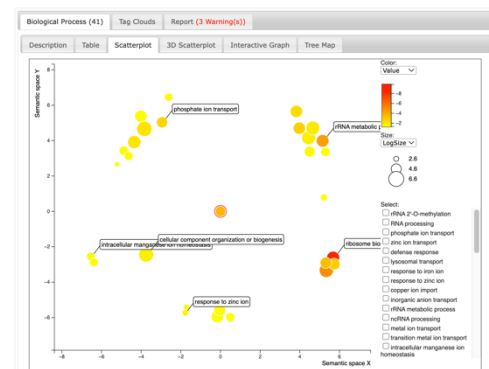
The table tab provides a detailed overview of the GO terms, P-values and also parent GO terms used to describe a group of related GO terms (http://geneontology.org/docs/ontology-relations/)

Additional resources:

More info on Fischer's exact test:
http://udel.edu/~mcdonald/statfishers.html
Some more info about Odds ratios:
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2938757/
False discovery rates and P value correction:
http://brainder.org/2011/09/05/fdr-corrected-fdr-adjusted-p-values/