Enrichment analysis

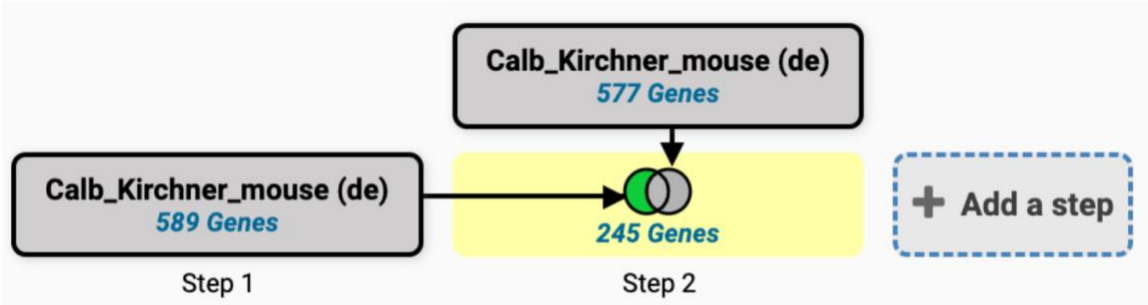
The enrichment analysis tools can be accessed under the blue Analyze Results tab and it includes Gene Ontology, Metabolic Pathway, and Word Enrichment tools. The three types of analysis apply Fisher's Exact test to evaluate ontology terms, over-represented pathways, and product description terms. Enrichment 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.

In the exercise above, you examined host response to infection with Candida strains. In this section, we will learn how to perform enrichment analysis on the fungal component of the Kirchner et al dataset.

To begin, click on the search strategy link below:
https://fungidb.org/fungidb/app/workspace/strategies/import/34f998f05745cbc3
This strategy identifies genes up-regulated in SC5314 at 1d post infection (Step 1) and subtracts up-regulated genes in common with the persistent 101 strain (Step2).

Reference Sample	SC5314_in vitro
Comparator Sample	SC5314_infected_1d
Direction	up-regulated
fold difference >=	4
adjusted P value less than or equal to	0.1

Reference Sample	101_in vitro
Comparator Sample	101_infected_1d
Direction	up-regulated
fold difference >=	4
adjusted P value less than or equal to	0.1



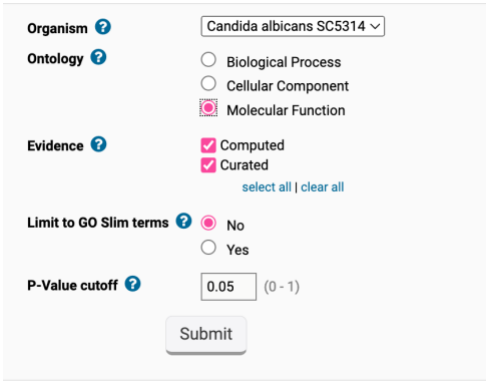
- **Perform GO Enrichment analysis (Molecular function)**
  1. Select Step 2 results (they will become highlighted in yellow).
  2. Click on the Analyze Results button located about the gene results table.

GO 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.

3. Select the “Molecular function” option.

4. Run the enrichment analysis on both computed and curated evidence (GO terms) and leave other parameters at default.

   Note: 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



5. Click on the “Submit” button.

- Examine your results. ooking at the enriched terms, do they make sense in terms of what you know about the Kirchner et al. 2019 dataset?

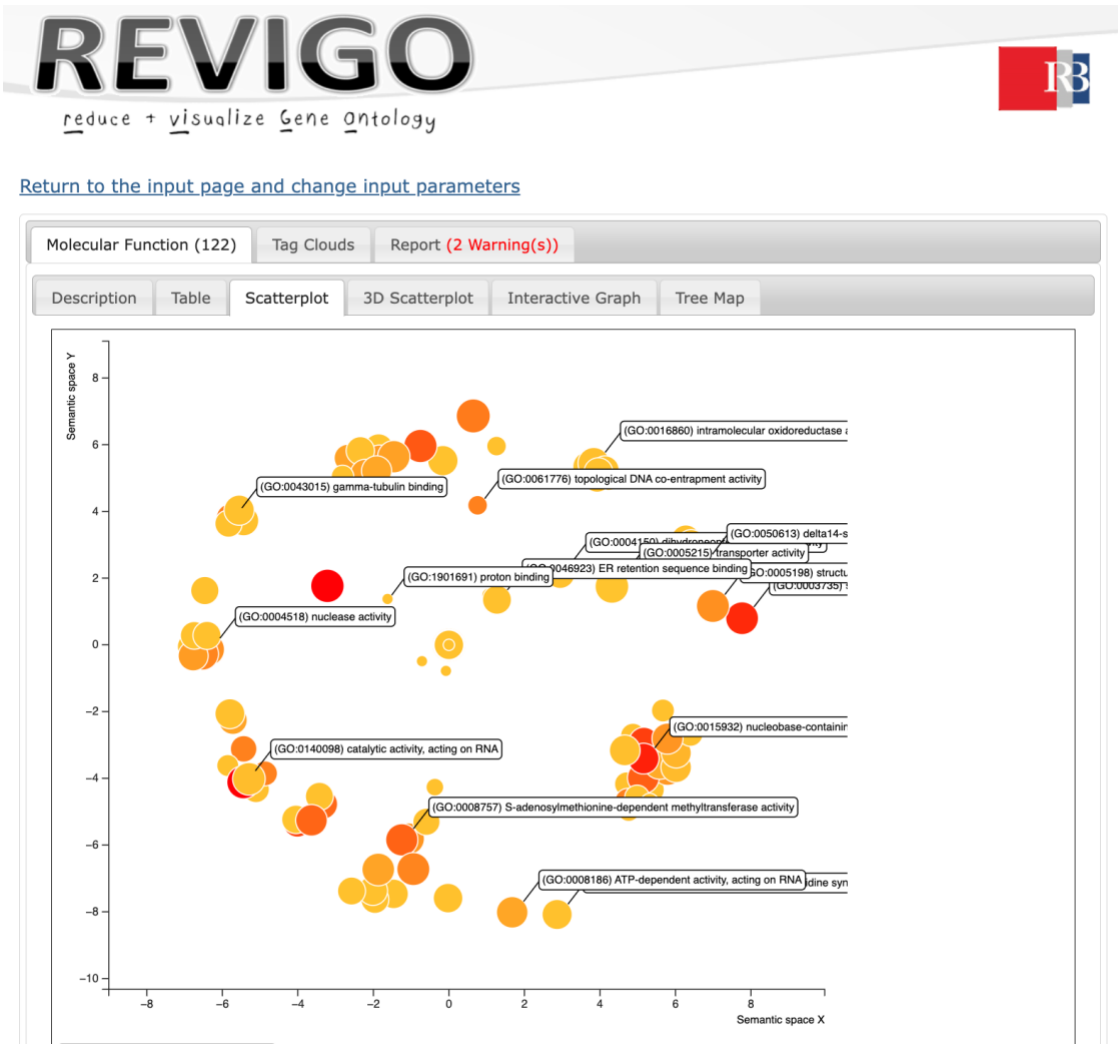
GO ID	GO Term	Genes in the bkgd with this term	Genes in your result with this term	Percent of bkgd genes in your result	Fold enrichment	Odds ratio	P-value	Benjamini	Bonferroni
GO:0140098	catalytic activity, acting on RNA	262	28	10.7	2.70	3.15	1.19e-6	6.18e-4	6.18e-4
GO:0140640	catalytic activity, acting on a nucleic acid	403	36	8.9	2.26	2.62	2.56e-6	6.68e-4	1.34e-3
GO:0015932	nucleobase-containing compound transmembrane transporter activity	40	9	22.5	5.68	7.28	1.92e-5	2.26e-3	1.00e-2
GO:0005347	ATP transmembrane transporter activity	23	7	30.4	7.69	10.90	1.99e-5	2.26e-3	1.04e-2
GO:0015215	nucleotide transmembrane transporter activity	32	8	25.0	6.31	8.33	2.48e-5	2.26e-3	1.29e-2
GO:0003735	structural constituent of ribosome	166	19	11.4	2.89	3.32	2.60e-5	2.26e-3	1.36e-2

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.

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

- **Visualize enrichment in REVIGO.**
  1. Click on the “Open in Revigo” button and follow prompts to complete this step.



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/).
More about REVIGO:
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021800