

Performing GO Enrichment analysis

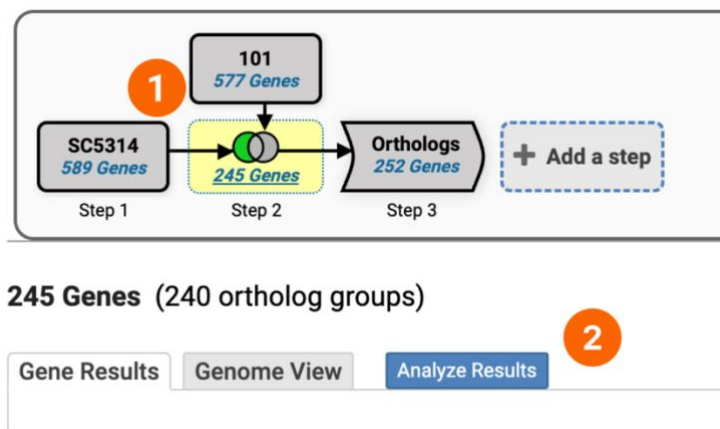
Learning objectives:

- Perform a GO enrichment analysis
- Create complex search strategy using both FungiDB and SGD
- Use a previously created search strategy to perform Gene Ontology enrichment analysis on genes upregulated (identified by RNA-Seq) in *C. albicans* SC5314 only.

Strategy URL:

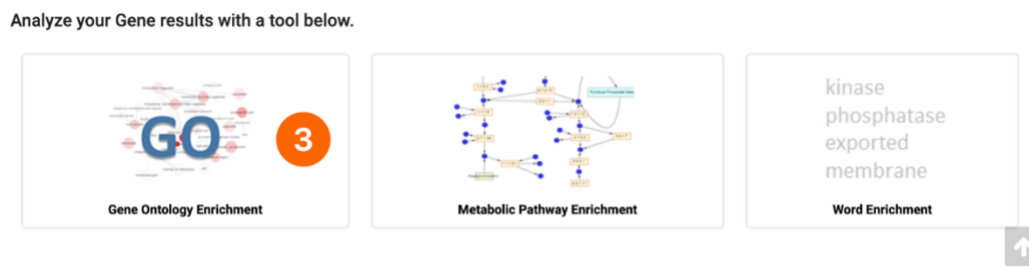
<https://fungidb.org/fungidb/app/workspace/strategies/import/802d9f2b606fc1fa>

1. Click on the Step 2 to identify upregulated gene in *C. albicans* SC5314 only.
2. Click on the “Analyze Results” tab to bring up enrichment analysis options.



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.

3. Deploy GO enrichment analysis by clicking on the “Gene Ontology Enrichment” button.



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.

Organism [?](#) Candida albicans SC5314

Ontology [?](#) ☒ Molecular Function
☐ Biological Process
☐ Cellular Component

Evidence [?](#) ☒ Computed
☒ Curated
[select all](#) | [clear all](#)

Limit to GO Slim terms [?](#) ☒ No
☐ Yes

P-Value cutoff [?](#) 0.05 (0 - 1)

[Submit](#)

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.

4. Perform GO enrichment analysis (Biological Process) at default selection criteria.

Organism [?](#) Candida albicans SC5314

Ontology [?](#) ☐ Molecular Function
☒ Biological Process
☐ Cellular Component

Evidence [?](#) ☒ Computed
☒ Curated
[select all](#) | [clear all](#)

Limit to GO Slim terms [?](#) ☒ No
☐ Yes

P-Value cutoff [?](#) 0.05 (0 - 1)

[Submit](#)

Analysis Results: [?](#) 244 rows

[Open in Revigo](#) [Show Word Cloud](#) [Download](#)

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:0042273	ribosomal large subunit biogenesis	558	67	12.0	3.03	4.20	1.08e-17	1.68e-14	1.68e-14
GO:0000470	maturation of LSU-rRNA	440	55	12.5	3.16	4.20	3.31e-15	2.59e-12	5.17e-12
GO:0000463	maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	432	53	12.3	3.10	4.07	2.62e-14	1.37e-11	4.10e-11

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).

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

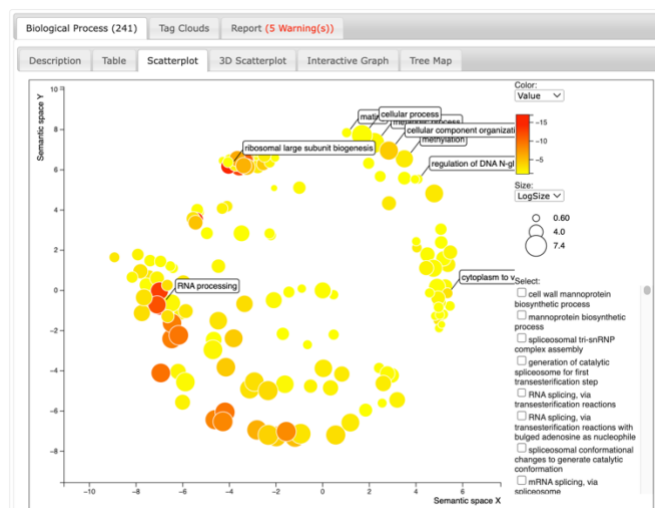
Note: you can sort genes in your results using the sort options within a column:

Genes in your result with this term ?	Percent of bkgd genes in your result ?
202	7.6
184	4.3
181	4.5

6. 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 user-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/>)

Creating queries across FungiDB and SGD (optional exercise)

During a genetic screen in *Lomentospora prolificans*, you identified several interesting genes, including jhhlp_004726, which is a hypothetical protein. Take advantage of FungiDB and SGD records to learn more about this gene.

1. Navigate to jhhlp_004726 in FungiDB and examine available records.

https://fungidb.org/fungidb/app/record/gene/jhhlp_004726

- Run an InterPro search and a GPI anchor prediction tool. What did you learn about this protein?

Hint: InterPro and GPI search tools can be found in the Protein features and properties section of the gene record page.

- Export orthologs of this gene.
Click on the Download gene link and select to export orthologs in VEuPathDB option

Download Gene: jhhlp_004726

Choose a Report: ☒ Text - choose from columns and/or tables ☐ FASTA - sequence retrieval, configurable

Choose Attributes

Choose Tables

Get Genes

- Navigate to the SGD gene lists search and copy and paste *S. cerevisiae* orthologs for jhhlp_004726: <https://www.yeastgenome.org/locus/YDR144C>



Create a new list

Select the type of list to create and either enter in a list of identifiers or upload identifiers from a file. A search will be performed for all the identifiers in your list.

- Separate identifiers by a **comma, space, tab, new line** or **semi-colon**.
- Qualify any identifiers that contain whitespace with double quotes like so: "even skipped".

Select Type:

for Organism:

Type/Paste in identifiers [\(click to see an example\)](#)

YDR144C
YGL259W
YIL015W
YIR039C
YLR120C
YLR121C

or Upload identifiers from a txt No file chosen

- Give your list a name such as 'Yeast orthologs 1'.
- Click on the GeneIDs to examine *S. cerevisiae* genes. What is the function of MKC7 (YDR144C) in *S. cerevisiae*? Does it encode a protein with enzymatic activity? Where in the cell does the protein execute its function? What biological process?
Hint: see the **Gene Ontology** section on the locus page or click on the Gene Ontology tab at the top of the page.

Functional relationships between genes and pathways can sometimes be revealed by examining genetic interactions between two or more genes. Genes are described as having a genetic interaction if the simultaneous mutation of both genes produces a phenotype that is unexpected, given the phenotypes of the single mutants.

- **Find known genetic interactions for MKC7.**
 - In SGD, find the MKC7 locus page and navigate to the **Interactions** tab, which is listed in the Quick Links panel near the top. The interactions are divided into separate physical interactions and genetic interactions tables below the summary.
 - Filter the **Genetic Interactions** table on “synthetic”. This table will show only the genetic interactions where some sort of synthetic growth defect, haploinsufficiency, or lethality is produced.

Summary Sequence Protein Gene Ontology Phenotype **Interactions** Regulation Expression Literature Homology

MKC7 / YDR144C Interactions [Interaction Help](#)

Summary: The mkc7 null mutant is viable; the null mutant of paralog yps1 is viable; the mkc7 yps1 double mutant has osmoremedial heat sensitivity, increased sensitivity to caffeine, congo red, caspofungin, calcofluor white, growth at low pH and a secretion defect; a mkc7 yps1 yps3 triple mutant has severe osmoremedial heat sensitivity and decreased tolerance to high salt.

Source: All physical and genetic interaction annotations listed in SGD are curated by [BioGRID](#).

Analyze

☒ Physical ☒ Genetic ☒ Intersection ☒ All

Genetic Interactions

Genetic Interactions 121 entries for 102 genes

Interactor	Allele	Assay	Annotation	Action	Phenotype	SGA score	P-value	Reference
ACT1		Synthetic Haploinsufficiency	high-throughput	Hit				Haarer B, et al. (2007) PMID:17167106
GIM5		Synthetic Growth Defect	high-throughput	Hit	vegetative growth: decreased Mutant Type: unspecified			Tong AH, et al. (2004) PMID:14764870

- Click on the **Download** button, which is located under the results table, and save this gene list. *Rename the file to **synthetic.txt**.*

*Note: Rename the file to **synthetic.txt** so that we can find it easily later.*

- Click on the **Analyze** button, then on **GO Term Finder**.
- Run a **process** enrichment for the MKC7 genetic interaction genes.

Hint: GO Term Finder finds common Gene Ontology (GO) annotations between genes. To run a Biological Process enrichment, select the Process button as shown below, then submit the form. More ways to customize your GO Term Finder query can be found in the GO Term Finder exercise.

Step 2. Choose Ontology

Pick an ontology aspect:

☒ Process
☐ Function
☐ Component

Search using default settings or use Step 3 and/or Step 4 below to customize your options.

- Scroll down the results page to see the table of enriched biological processes. What kind of processes are associated with the genes we analyzed? What do these results suggest about MKC7’s functional relationships in the cell?
- Click on any of the genes shown for a biological process of interest to visit the gene’s page on SGD. Use the gene page to uncover how the respective gene is involved in the biological process you were interested in.

Result Table

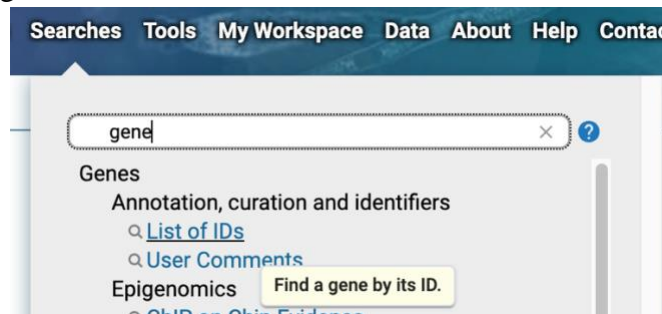
Terms from the Process Ontology of gene_association.sgd with p-value <= 0.01						
Gene Ontology term	Cluster frequency	Genome frequency	Corrected P-value	FDR	False Positives	Genes annotated to the term
tubulin complex assembly	3 of 9 genes, 33.3%	10 of 7166 genes, 0.1%	1.96e-05	0.00%	0.00	YML094W, YLR200W, YGR078C
protein folding	4 of 9 genes, 44.4%	121 of 7166 genes, 1.7%	0.00109	0.00%	0.00	YML094W, YLR200W, YKL117W, YGR078C
peptide pheromone maturation	2 of 9 genes, 22.2%	9 of 7166 genes, 0.1%	0.00603	0.67%	0.02	YNL238W, YLR120C
chaperone-mediated protein complex assembly	2 of 9 genes, 22.2%	9 of 7166 genes, 0.1%	0.00603	0.50%	0.02	YKL117W, YLR200W
fungal-type cell wall organization	4 of 9 genes, 44.4%	205 of 7166 genes, 2.9%	0.00878	0.40%	0.02	YHR079C, YLR120C, YLR121C, YFL039C

Now, let’s go back to the file of MKC7 “synthetic” genetic interactors we downloaded earlier and find the orthologs of these genes in *Lomentospora prolificans*.

- Open this file in Excel and copy the Gene IDs in the **Interactor Systematic Name** column (not including the header)

Interactor	Interactor Sy	Interactor	Interactor Systematic Name	Type	Assay	Annotation
MKC7	YDR144C	ACT1	YFL039C	Genetic	Synthetic Ha	high-through
MKC7	YDR144C	GIM5	YML094W	Genetic	Synthetic Gri	high-through
MKC7	YDR144C	IRE1	YHR079C	Genetic	Synthetic Gri	manually cur
MKC7	YDR144C	KEX2	YNL238W	Genetic	Synthetic Let	manually cur
MKC7	YDR144C	PAC10	YGR078C	Genetic	Synthetic Let	high-through
MKC7	YDR144C	SBA1	YKL117W	Genetic	Synthetic Let	high-through
MKC7	YDR144C	YKE2	YLR200W	Genetic	Synthetic Gri	high-through
MKC7	YDR144C	YPS1	YLR120C	Genetic	Synthetic Let	manually cur
MKC7	YDR144C	YPS1	YLR120C	Genetic	Synthetic Let	manually cur
MKC7	YDR144C	YPS1	YLR120C	Genetic	Synthetic Gri	manually cur
MKC7	YDR144C	YPS1	YLR120C	Genetic	Synthetic Let	manually cur
MKC7	YDR144C	YPS3	YLR121C	Genetic	Synthetic Let	manually cur

- Visit FungiDB again and initiate the List of IDs search query
- The query can be deployed from the “Searches” menu at the top or the “Search for Genes” section on the main page.



- Paste the list of Gene IDs that had the “synthetic” genetic interactions with MKC7 into FungiDB query and click on the **Get Answer** button.

Identify Genes based on List of IDs

Configure Search | Learn More | View Data Sets Used

Reset values to default

Gene ID input set

Enter a list of IDs or text:

☐ Upload a text file: No file chosen
Maximum size 10MB. The file should contain the list of IDs.

☐ Upload from a URL:
The URL should resolve to a list of IDs.

☐ Copy from My Basket: 3 records will be copied from your basket.

☐ Copy from My Strategy: NRPS (766 records)

IDs List
9 Genes

Add a step

Step 1

9 Genes (8 ortholog groups) [Revise this search](#)

Gene Results

Genome View

Analyze Results

Rows per page: 1000

Download

Send to...

Add Columns

	Gene ID	Transcript ID	Gene Name or Symbol	Organism	Genomic Location (Gene)	Product Description
	YFL039C	YFL039C-126_1	ACT1	<i>Saccharomyces cerevisiae</i> S288C	BK006940:53,260..54,696(-)	actin
	YML094W	YML094W-126_1	GIM5	<i>Saccharomyces cerevisiae</i> S288C	BK006946:82,275..82,849(+)	Gim5p
	YHR079C	YHR079C-126_1	IRE1	<i>Saccharomyces cerevisiae</i> S288C	BK006934:258,244..261,591(-)	bifunctional endoribonuclease/protein

- Find orthologs in *Lomentospora prolificans*. Click on the “Add a step” button to **Transform** the list **into related records**. Select the option to transform into **orthologs**, then use the search bar to filter on *Lomentospora prolificans* and **Run Step**.

Gene ID(s)
9 Genes

Add a step

Step 1

Add a step to your search strategy

Combine with other Genes

Transform into related records

Use Genomic Colocation to combine with other features

Transform 9 Genes into...

Orthologs

Add a step to your search strategy

Your Genes from Step 1 will be converted into Orthologs

Organism

Note: You must select at least 1 values for this parameter.
1 selected, out of 163

add these | clear these | select only these

Lom

Fungi

Ascomycota

Sordariomycetes

Microascales

☒ Lomentospora prolificans JH4-5317

select all | clear all

Syntenic Orthologs Only?

no

Run Step

Gene ID	Transcript ID	Organism	Genomic Location (Gene)	Product Description	Input Ortholog(s)	Ortholog Group	Paralog count	Ortholog count
jhhlp_002587	jhhlp_002587-t41_1	Lomentospora prolificans JHH-5317	NLAX01000008:3,258,120..3,260,362(-)	hypothetical protein	YFL039C	OG6_100127	0	239
jhhlp_004481	jhhlp_004481-t41_1	Lomentospora prolificans JHH-5317	NLAX01000010:4,766,898..4,769,585(+)	hypothetical protein	YNL238W	OG6_100362	0	167
jhhlp_004364	jhhlp_004364-t41_1	Lomentospora prolificans JHH-5317	NLAX01000010:4,180,492..4,181,475(-)	hypothetical protein	YKL117W	OG6_101574	0	157

How many of the interacting *S. cerevisiae* genes have a hypothetical protein ortholog in *Lomentospora prolificans*? Can you find jhhlp_004726 amongst these genes?

Strategy URL: <https://fungidb.org/fungidb/app/workspace/strategies/import/c0978bdb48a8392d>

Glycosylphosphatidylinositol (GPI)-anchored proteins are involved in cell wall integrity and cell-cell interactions and perturbations in GPI biosynthesis lead to hypersensitivity to host defenses. Given the accumulated biological information we uncovered at SGD and FungiDB, summarize your predictions about the hypothetical *L. prolificans* protein jhhlp_004726.

- What is the likely jhhlp_004726 ortholog in *S. cerevisiae*?
 - Is this gene a GPI-protein in yeast?
- Do you have sufficient information to think the hypothetical gene in *L. prolificans* may be a putative GPI-anchor protein?
- How many “synthetic” genetic interactors exist in SGD for MKC7 in yeast?
 - What GO terms were enriched in biological processes associated with MKC7 interactors in *S. cerevisiae*?
 - How many orthologs of these genes are found in *L. prolificans*?
 - Why do you think the number of genes vary between *S. cerevisiae* and *L. prolificans*?

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/>