

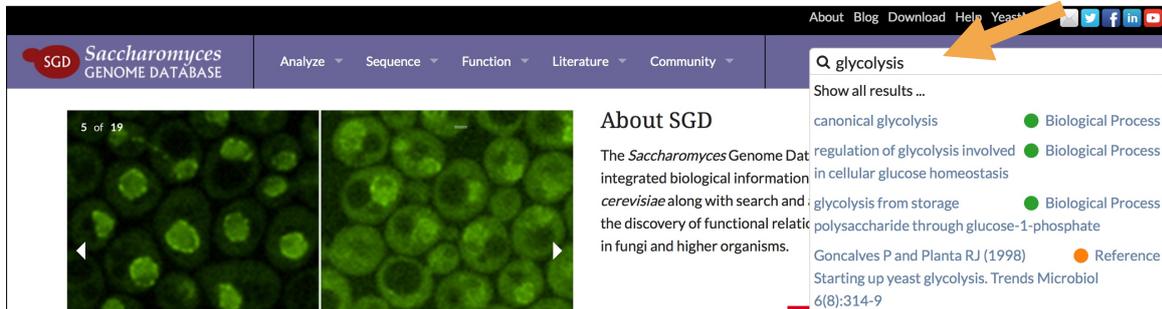
Using SPELL to Analyze Expression Datasets & Coexpressed Genes at SGD

SPELL (Serial Pattern of Expression Levels Locator) is a query-driven search engine for large gene expression microarray compendia. Given a small set of query genes, SPELL identifies which datasets are most informative for these genes, then within those datasets additional genes are identified with expression profiles most similar to the query set.

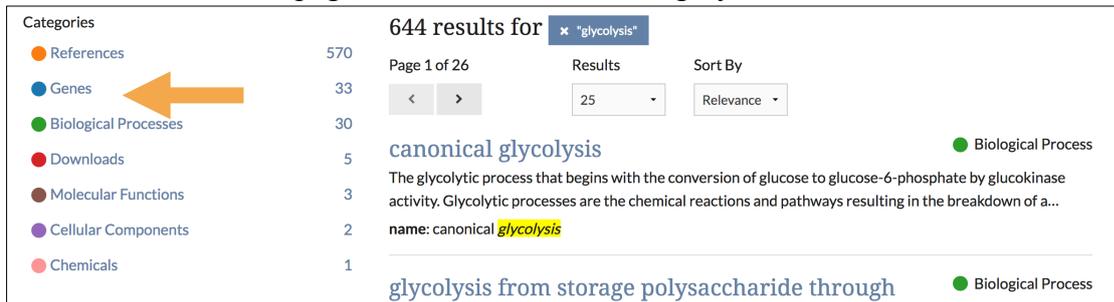
Use SPELL to find out which genes are coexpressed with genes involved in glycolysis.

Compile a list of genes involved in glycolysis.

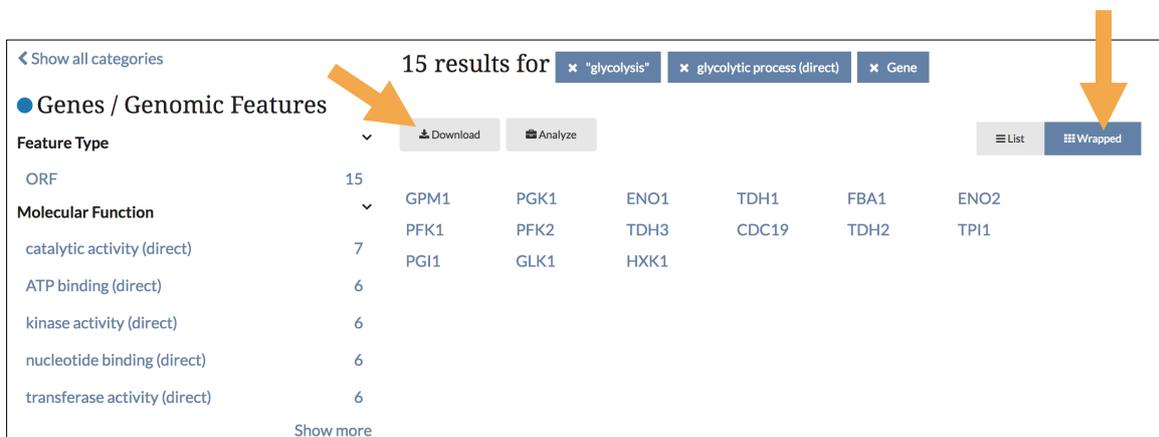
- On the SGD home page (www.yeastgenome.org), enter glycolysis into the search box and hit Enter.



- On the Results page, click on the **Genes** category.



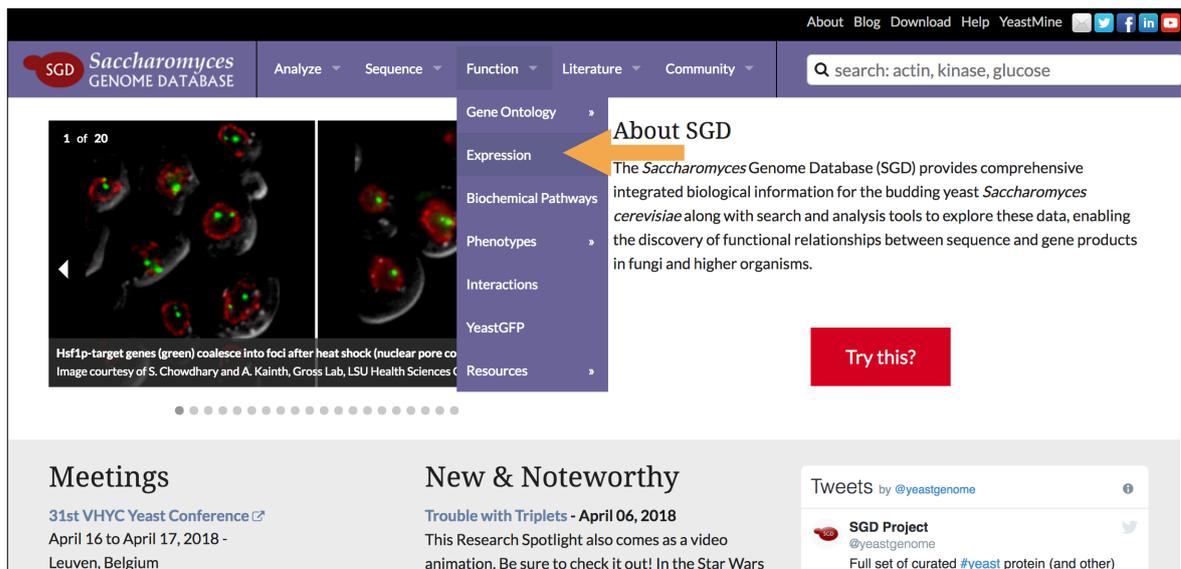
- Scroll down the page and find the **Biological Process** category on the left hand menu. Hit Show more and select **glycolytic process (direct)**.
- To download the list of genes, click on **Wrapped** and then on **Download**.



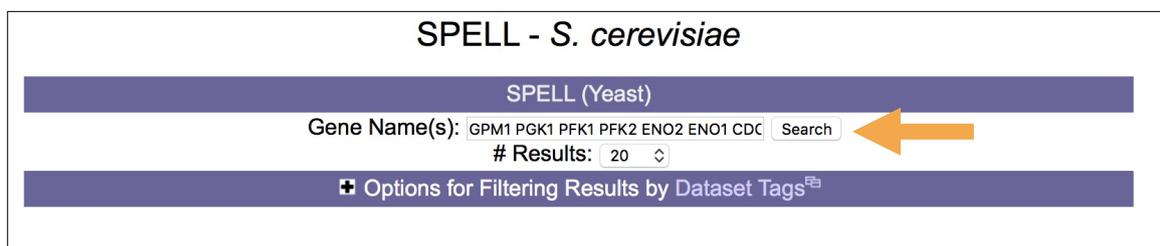
- The **Analyze** button, directly to the right of Download, enables you to import your search results directly into SPELL (among other tools at SGD). However, for the sake of demonstration, in this exercise we are instead going to enter our gene list into SPELL manually.

Import your gene list into SPELL and run a query:

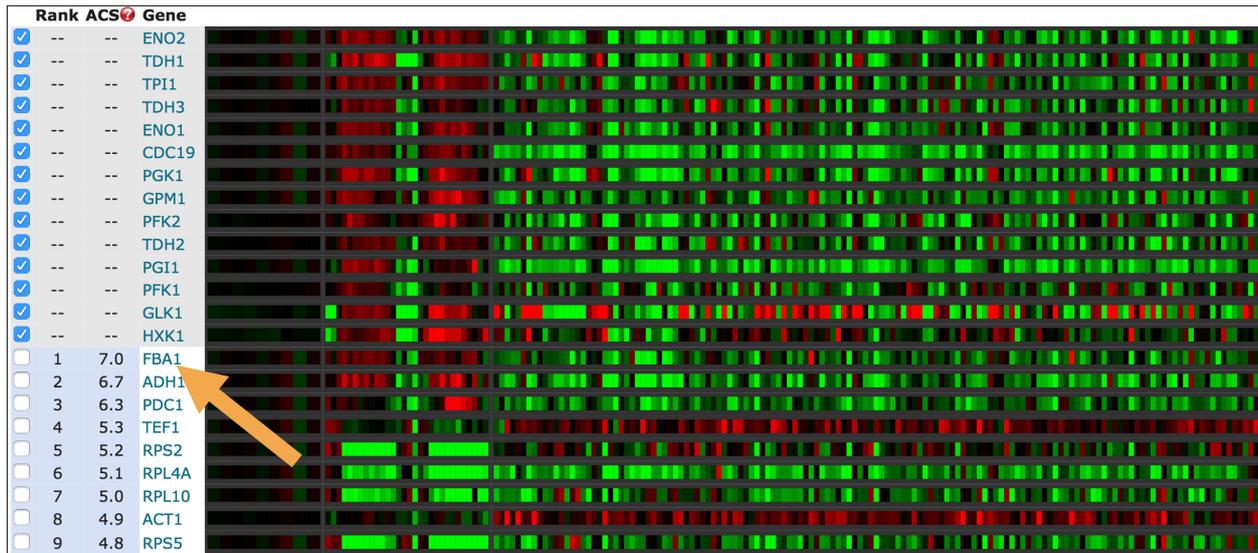
- To access SPELL, go to the SGD home page at www.yeastgenome.org, open the **Function** tab on top of the page and click on **Expression**. Or, if you are already on a Locus Summary page, open the Expression tab and click on the SPELL link under the histogram.



- On the SPELL page, copy and paste the list of glycolysis genes you downloaded in step 1 into the Gene Name(s) box. For the sake of demonstration, remove **FBA1** from your list before hitting Search. This is to test if SPELL can properly identify missing members of glycolysis based on coexpression.



- Scroll down the list of genes on the left. Genes with checked boxes are from our query; the remaining genes are "hits", ordered from top to bottom according to their ranks. The rank reflects the correlation of expression of that gene with the query gene(s), given the relevance weight of that expression dataset. Thus, genes that show the highest degree of coexpression with the query genes in the most relevant datasets receive the highest rank.



○ Notice that the glycolysis gene we deleted earlier, FBA1, is indeed the highest-ranking gene!

- Examine other genes enriched for this query set. You can click on their names to be taken to their respective summary pages at SGD. Does it make sense for any of these genes to be highly coexpressed with members of glycolysis?
- Click on + **Additional Display Options** to change the default mapping method and color scheme to blue/yellow. Directly above this section are options to change the number of genes and datasets shown in your results.

of Result Genes to Show: Datasets to view:

+ Additional Display Options

| | Mapping method | Color scheme |
|---|--|--|
| For single channel data: <input type="checkbox"/> | <input type="text" value="Per-gene log2 fold change"/> | <input type="text" value="Red/Green"/> |
| For dual channel data: <input type="checkbox"/> | <input type="text" value="Reported log2 fold change"/> | <input type="text" value="Red/Green"/> |

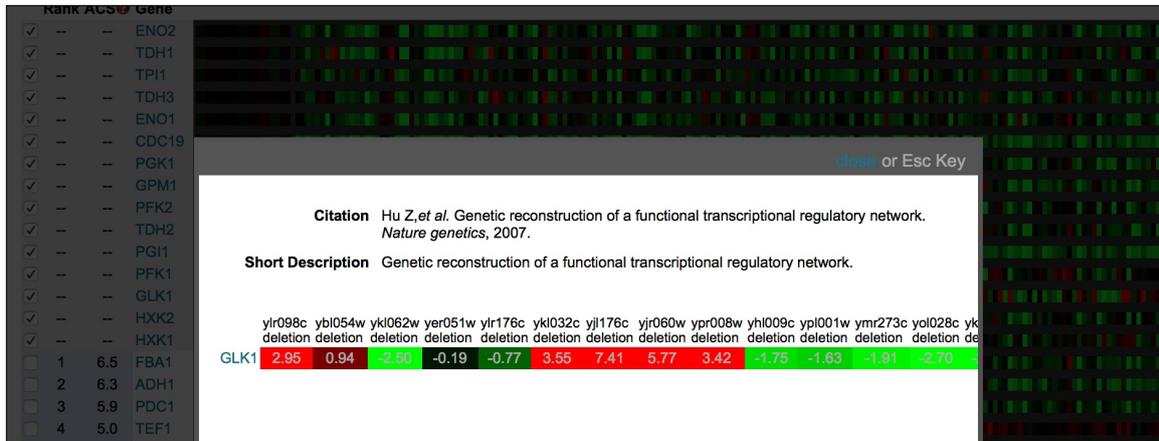
- To select only datasets with particular tags, click on + **Options** for Filtering Results.

Dataset Tags

Select:

| | | | |
|--|--|--|--|
| <input type="checkbox"/> amino acid metabolism | <input type="checkbox"/> evolution | <input type="checkbox"/> organelles, biogenesis, structure, and function | <input type="checkbox"/> RNA catabolism |
| <input type="checkbox"/> amino acid utilization | <input type="checkbox"/> fermentation | <input type="checkbox"/> osmotic stress | <input type="checkbox"/> signaling |
| <input type="checkbox"/> carbon utilization | <input type="checkbox"/> filamentous growth | <input type="checkbox"/> oxidative stress | <input type="checkbox"/> sporulation |
| <input type="checkbox"/> cell aging | <input type="checkbox"/> flocculation | <input type="checkbox"/> oxygen level alteration | <input type="checkbox"/> starvation |
| <input type="checkbox"/> cell cycle regulation | <input type="checkbox"/> genetic interaction | <input type="checkbox"/> phosphorus utilization | <input type="checkbox"/> stationary phase entry |
| <input type="checkbox"/> cell morphogenesis | <input type="checkbox"/> genome variation | <input type="checkbox"/> ploidy | <input type="checkbox"/> stationary phase maintenance |
| <input type="checkbox"/> cell wall organization | <input type="checkbox"/> heat shock | <input type="checkbox"/> protein dephosphorylation | <input type="checkbox"/> stress |
| <input type="checkbox"/> cellular ion homeostasis | <input type="checkbox"/> histone modification | <input type="checkbox"/> protein glycosylation | <input type="checkbox"/> sulfur utilization |
| <input type="checkbox"/> chemical stimulus | <input type="checkbox"/> lipid metabolism | <input type="checkbox"/> protein modification | <input type="checkbox"/> synthetic biology |
| <input type="checkbox"/> chromatin organization | <input type="checkbox"/> mating | <input type="checkbox"/> protein phosphorylation | <input type="checkbox"/> transcription |
| <input type="checkbox"/> cofactor metabolism | <input type="checkbox"/> metabolism | <input type="checkbox"/> protein trafficking, localization and degradation | <input type="checkbox"/> transcriptional regulation |
| <input type="checkbox"/> diauxic shift | <input type="checkbox"/> metal or metalloid ion stress | <input type="checkbox"/> proteolysis | <input type="checkbox"/> translational regulation |
| <input type="checkbox"/> disease | <input type="checkbox"/> mitotic cell cycle | <input type="checkbox"/> QTLs | <input type="checkbox"/> ubiquitin or ULP modification |
| <input type="checkbox"/> DNA damage stimulus | <input type="checkbox"/> mRNA processing | <input type="checkbox"/> radiation | |
| <input type="checkbox"/> DNA replication, recombination and repair | <input type="checkbox"/> nitrogen utilization | <input type="checkbox"/> respiration | |
| <input type="checkbox"/> environmental-sensing | <input type="checkbox"/> nutrient utilization | <input type="checkbox"/> response to unfolded protein | |

- Click on any patch in the heat map to open a page with information about its parent dataset.



- SPELL also runs a **Gene Ontology (GO) enrichment** for the results of your query. GO enrichments can tell you which gene ontology terms (in this case, biological process terms) are significantly associated with your set of genes. You can scroll down to the bottom of the page to view it.

| GO Term | P-val | % query | % genome | Annotated Genes |
|---|----------|----------|-------------|---|
| glucose catabolic process (biological_process) | 1.33e-29 | 19 of 35 | 52 of 6381 | FBA1, TDH3, ENO1, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, ENO2, PDC1, ADH1, PDC5, CDC19, PGK1, TDH2 |
| hexose catabolic process (biological_process) | 2.39e-28 | 19 of 35 | 59 of 6381 | FBA1, TDH3, ENO1, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, ENO2, PDC1, ADH1, PDC5, CDC19, PGK1, TDH2 |
| monosaccharide catabolic process (biological_process) | 2.91e-27 | 19 of 35 | 66 of 6381 | FBA1, TDH3, ENO1, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, ENO2, PDC1, ADH1, PDC5, CDC19, PGK1, TDH2 |
| glycolysis (biological_process) | 4.79e-27 | 16 of 35 | 32 of 6381 | FBA1, TDH3, ENO1, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, ENO2, CDC19, PGK1, TDH2 |
| glucose metabolic process (biological_process) | 1.66e-23 | 19 of 35 | 99 of 6381 | FBA1, TDH3, ENO1, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, ENO2, PDC1, ADH1, PDC5, CDC19, PGK1, TDH2 |
| single-organism carbohydrate catabolic process (biological_process) | 3.62e-22 | 19 of 35 | 115 of 6381 | FBA1, TDH3, ENO1, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, ENO2, PDC1, ADH1, PDC5, CDC19, PGK1, TDH2 |
| hexose metabolic process (biological_process) | 4.32e-22 | 19 of 35 | 116 of 6381 | FBA1, TDH3, ENO1, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, ENO2, PDC1, ADH1, PDC5, CDC19, PGK1, TDH2 |
| monosaccharide metabolic process (biological_process) | 1.42e-21 | 19 of 35 | 123 of 6381 | FBA1, TDH3, ENO1, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, ENO2, PDC1, ADH1, PDC5, CDC19, PGK1, TDH2 |
| carbohydrate catabolic process (biological_process) | 1.97e-21 | 19 of 35 | 125 of 6381 | FBA1, TDH3, ENO1, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, ENO2, PDC1, ADH1, PDC5, CDC19, PGK1, TDH2 |
| generation of precursor metabolites and energy (biological_process) | 7.97e-18 | 19 of 35 | 190 of 6381 | FBA1, TDH3, ENO1, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, ENO2, PDC1, ADH1, PDC5, CDC19, PGK1, TDH2 |
| single-organism carbohydrate metabolic process (biological_process) | 1.60e-13 | 19 of 35 | 319 of 6381 | FBA1, TDH3, ENO1, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, ENO2, PDC1, ADH1, PDC5, CDC19, PGK1, TDH2 |
| gluconeogenesis (biological_process) | 3.72e-13 | 10 of 35 | 33 of 6381 | FBA1, TDH3, ENO1, GPM1, TPI1, TDH1, PGI1, ENO2, PGK1, TDH2 |
| hexose biosynthetic process (biological_process) | 5.25e-13 | 10 of 35 | 34 of 6381 | FBA1, TDH3, ENO1, GPM1, TPI1, TDH1, PGI1, ENO2, PGK1, TDH2 |
| monosaccharide biosynthetic process (biological_process) | 7.33e-13 | 10 of 35 | 35 of 6381 | FBA1, TDH3, ENO1, GPM1, TPI1, TDH1, PGI1, ENO2, PGK1, TDH2 |