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 (<u>www.yeastgenome.org</u>), enter glycolysis into the search box and hit Enter.



• On the Results page, click on the Genes category.

Categories		644 results fo	r × "glycolysis"			
References	570	Page 1 of 26	Results	Sort By		
Genes	33	< >	25 •	Relevance -		
Biological Processes	30					
Downloads	5	canonical glycolysis Biological Process				
Molecular Functions	3	The glycolytic process activity. Glycolytic pro	s that begins with the c ocesses are the chemic	conversion of glucose to glucose-6-phosphate by glucokinase cal reactions and pathways resulting in the breakdown of a		
Cellular Components	2	name: canonical glyco	olysis			
Chemicals	1	glycolysis fro	m storage pol	ysaccharide through Biological Process		

• 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.

Show all categories		15 result	s for 💌	glycolysis" 🗙 g	glycolytic process (dir	ect) 🗙 Gene		
• Genes / Genomic Features								
Feature Type	~	🛓 Download	📾 Analyze				≡List	III Wrapped
ORF	15							
Molecular Function	~	GPM1	PGK1	ENO1	TDH1	FBA1	ENO2	
catalytic activity (direct)	7	PFK1	PFK2		CDC19	TDH2	I PI1	
ATP binding (direct)	6	FGII	GLNI	HARI				
kinase activity (direct)	6							
nucleotide binding (direct)	6							
transferase activity (direct)	6							
Show	/ more							

• 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 <u>www.yeastgenome.org</u>, 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.

SPELL - S. cerevisiae
SPELL (Yeast)
Gene Name(s): GPM1 PGK1 PFK1 PFK2 ENO2 ENO1 CDC Search # Results: 20 0
Options for Filtering Results by Dataset Tags ^{Eb}

• 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.

Γ	F	lank	ACS	Gene	
	 Image: A second s			ENO2	
	 Image: A second s			TDH1	
	 Image: A second s			TPI1	
	v			TDH3	
	v			ENO1	
	v			CDC19	
	~			PGK1	
	~			GPM1	
	~			PFK2	
	~			TDH2	
	~			PGI1	
				PFK1	
	7			GLK1	
	7			HXK1	
C		1	7.0	FBA1	
C		2	6.7	ADH1	
0		3	6.3	PDC1	
0		4	5.3	TEE1	
C		5	5.2	RPS2	
ſ		6	5 1	RPI 4A	
C		7	5.0	RPI 10	
C		8	4.9	ACT1	
Ì	ň	0	1.9	DDC5	
		9	7.0	KF35	

- 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:	20 \$	Datasets to view:	Fr	om 1 to 10	٥
Additional Display Options	\$				
Mapping method				Color sche	eme
For single channel data:	Per-gene	ne log2 fold change 💲		Red/Green	٢
For dual channel data:	Reported	log2 fold change	٢	Red/Green	٥

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

Dataset Tags 🥹						
Select: all none previous query toggle						
amino acid metabolism	evolution	 organelles, biogenesis, structure, and function 	RNA catabolism			
amino acid utilization	fermentation	osmotic stress	signaling			
carbon utilization	filamentous growth	oxidative stress	sporulation			
cell aging	flocculation	oxygen level alteration	starvation			
cell cycle regulation	genetic interaction	phosphorus utilization	stationary phase entry			
cell morphogenesis	genome variation	ploidy	stationary phase maintenance			
cell wall organization	heat shock	protein dephosphorylation	□ stress			
cellular ion homeostasis	histone modification	protein glycosylation	sulfur utilization			
chemical stimulus	lipid metabolism	protein modification	synthetic biology			
chromatin organization	mating	protein phosphorylation	transcription			
Cofactor metabolism	metabolism	protein trafficking, localization and degradation	transcriptional regulation			
☐ diauxic shift	 metal or metalloid ion stress 	proteolysis	translational regulation			
disease	mitotic cell cycle	QTLs	 ubiquitin or ULP modification 			
DNA damage stimulus	mRNA processing	radiation				
DNA replication, recombination an repair	^d initrogen utilization	respiration				
environmental-sensing	nutrient utilization	response to unfolded protein				

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

Rank	ACS	Gene		
		TDH1		
		TPI1		
		TDH3		
		PGK1	close or Esc Key	
		PFK2	Citation Hu 7 at al Constin reconstruction of a functional transcriptional regulatory patwork	
		TDH2	Nature genetics. 2007.	
		PFK1	Short Description Genetic reconstruction of a functional transcriptional regulatory network.	
		GLK1		
			vin98c, vbi054w vki062w ver051w vir176c, vki032c, vii176c, vin660w von008w vbi009c vol001w vmr273c vol028c vk	
			deletion del	
1	6.5	FBA1	GLK1 2.95 0.94 <mark>-2.50</mark> -0.19 -0.77 3.55 7.41 5.77 3.42 -1.75 -1.63 -1.91 -2.70 -	
2	6.3	ADH1		
3	5.9	PDC1		
4	5.0	TEF1		

• 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 Enrichment@	GO Term Enrichment€							
GOTerm	P-val % quer	/ % genome	Annotated Genes					
<u>glucose catabolic process</u> (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, EN01, HXK1, HXK2, PFK2, GLK1, GPM1, PFK1, TPI1, TDH1, PGI1, EN02, 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	i 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, GDC19, 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, GDC19, 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, GDC19, 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, GDC19, 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					