

Coexpression Module-guide

Introduction

Coexpression refers to the phenomenon when two or more genes are expressed or turned on at the same time. Coexpression is calculated by looking at the correlation in RNA abundance for pairs of genes across multiple samples. For example, if across many different samples you see that whenever gene_1 is highly expressed, gene_2 tends to be also highly expressed, and when gene_1 is lowly expressed, gene_2 is also lowly expressed, we would then say that gene_1 and gene_2 are coexpressed.

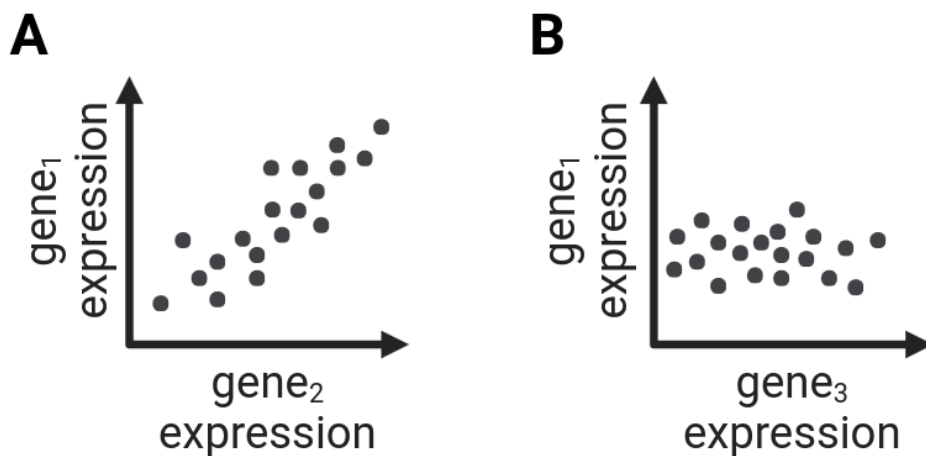


Figure 1: RNA abundance plots. **A)** Each dot represents a sample and shows the RNA expression level for gene_1 and gene_2 . Gene_1 and gene_2 have correlated expression such that when gene_1 is highly expressed, so is gene_2 ; therefore, gene_1 and gene_2 are coexpressed. **B)** The RNA expression level for gene_1 and gene_3 in each sample. Gene_1 and gene_3 do not have correlated expression; therefore, gene_1 and gene_3 are not coexpressed. Figure made with BioRender.com

Coexpression can be used to identify potential biological roles of genes; this is because research has shown that genes that are involved in similar biological processes tend to show similar expression patterns. Therefore, in this module, you will investigate which genes and other proto-genes have similar expression patterns (coexpression) to your proto-gene and what biological processes these genes are involved in. The dataset you will be analyzing looked at the correlation in expression for 5,803 annotated genes (denoted canonical ORFs) and 5,827 proto-genes (denoted noncanonical ORFs) across ~4,000 samples that had a wide variety of experimental conditions.

We can use coexpression to see if other genes that are also differentially expressed under the same conditions tend to have high coexpression, or expression similarity, across many different conditions. We will also explore whether these genes are known to have similar transcriptional regulators.

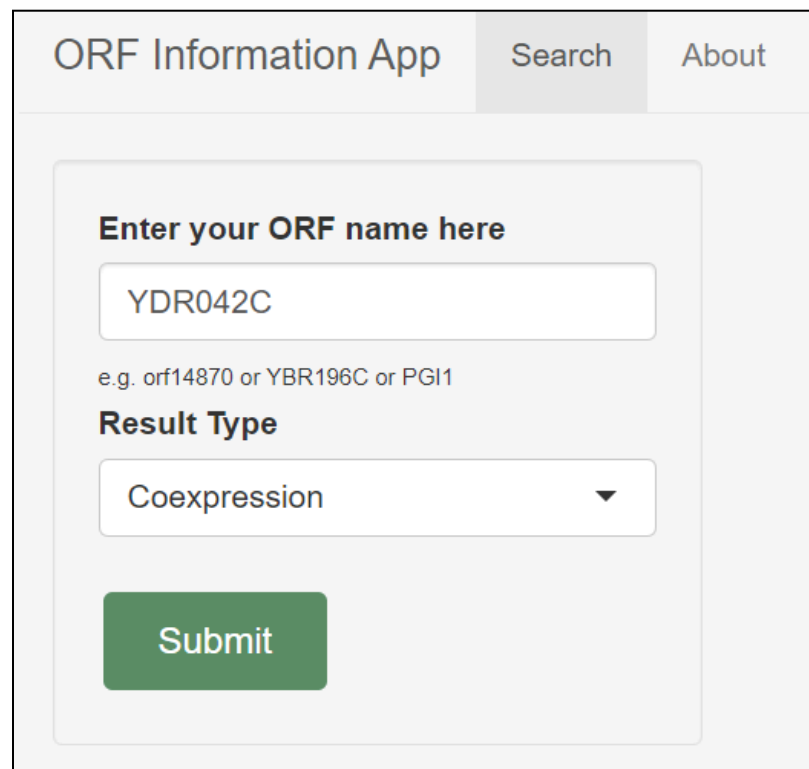
Goals: In this module, participants will use an online tool to investigate coexpression patterns among genes and the biological processes that these genes are involved in to see if these genes are known to share transcription regulators.

Objectives: After completing this module, participants will be able to:

- 1) Identify genes coexpressed with a given query
- 2) Identify known regulators for genes
- 3) Identify what cellular processes are enriched in a set of coexpressed genes

Activity

1. Go to the ORF information app website:
<https://carvunislab.csb.pitt.edu/shiny/coexpression/>
2. Enter the name of your proto-gene of interest in the search box, in this example we are using YDR042C which is a differentially expressed proto-gene from module 1. In the 'Result Type' menu select '**Coexpression**'



The screenshot shows the 'ORF Information App' interface. At the top, there are three tabs: 'ORF Information App' (selected), 'Search', and 'About'. Below the tabs, there is a search form. The form has a title 'Enter your ORF name here' and a text input field containing 'YDR042C'. Below the input field, there is a hint text 'e.g. orf14870 or YBR196C or PGI1'. Below the hint text, there is a 'Result Type' label and a dropdown menu showing 'Coexpression'. At the bottom of the form, there is a green 'Submit' button.

3. On the next page click on the tab that says **Coexpression Network**. This returns a table showing what genes or proto-genes your proto-gene is coexpressed with, listed in

descending order. The coexpression values range from -1 to 1, where a value greater than 0.85 represents strong coexpression. The table also contains information about whether or not the ORFs are canonical or noncanonical, where canonical ORFs are genes with annotations (Verified or Uncharacterized in SGD) and noncanonical ORFs are unannotated ORFs (ie proto-genes).

Coexpression Info for ORF YDR042C

[Gene set enrichment](#)
[Coexpression Network](#)

Coexpression

Filter by ORF classification:

all

Show 10 entries

Search:

ORF ID	Systematic name	Gene name	ORF classification	Coexpression (rho)
orf103139	YGR273C		canonical	0.942
orf41858	YDL186W		canonical	0.94
orf214476	YNL130C-A	DGR1	canonical	0.939
orf106038	YGL138C		canonical	0.937

You can filter to display only canonical genes (i.e. genes with annotations as Verified or Uncharacterized in SGD) by selecting '**canonical**' in the '*Filter by ORF classification*' drop down menu.

Filter by ORF classification:

canonical

To search for a specific gene, you can type the name of the gene into the search box above the table. For example let's take a look at a gene that was also highly upregulated in the treatment condition (from Module 1) to see if it coexpressed with our proto-gene of interest.

Show 10 entries		Search: TKL2		
ORF ID	Systematic name	Gene name	ORF classification	Coexpression (rho)
orf15388	YBR117C	TKL2	canonical	0.345

This is below the 0.85 threshold, so TKL2 is not strongly co-expressed with the protogene YDR042C.

Let's find out more about TKL2, including if it is known to also be regulated by Mot3, which was the predicted regulator from module 2.

On SGD's website: <https://www.yeastgenome.org/> search for the gene and go to the **Regulation** tab at the top.

Summary

Sequence

Protein

Gene Ontology

Phenotype

Interactions

Regulation

Expression

Literature

Homology

TKL2 / YBR117C

Locus Overview

Sequence

Protein

Alleles

Gene Ontology

Pathways

Phenotype

Interaction

Regulation

Expression

Summary Paragraph

Literature

History

References

Resources

TKL2 / YBR117C Overview

Standard Name:

TKL2¹

Systematic Name:

YBR117C

SGD ID:

SGD:S0000000321

Feature Type:

ORF, Verified

Description:

Transketolase; catalyzes conversion of xylulose-5-phosphate and ribose-5-phosphate to sedoheptulose-7-phosphate and glyceraldehyde-3-phosphate in the pentose phosphate pathway; needed for synthesis of aromatic amino acids; TKL2 has a paralog, TKL1, that arose from the whole genome duplication^{1 2}

Name Description:

TransKetoLase¹

Paralog:

TKL1²

Comparative Info:

Sequence

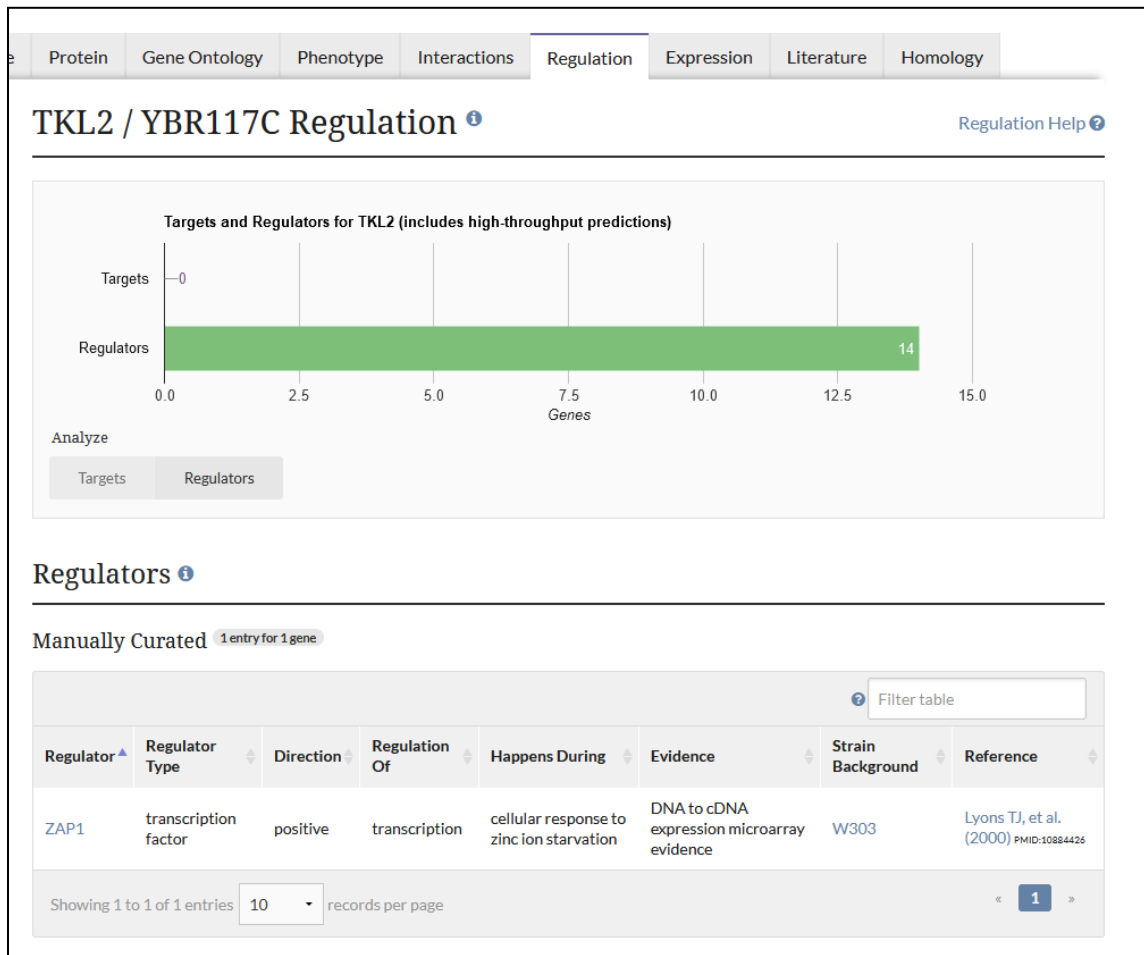
Summary:

TKL2 has a paralog, TKL1, that arose from the whole genome duplication

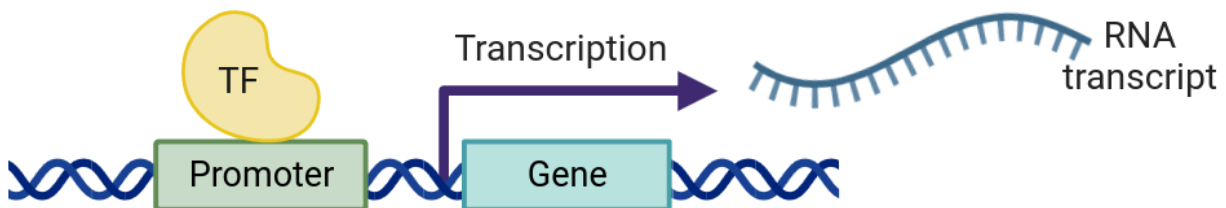
Download (.fsa)

View in: JBrowse

TKL2 Location: Chromosome II 474392..476437



Here it will show the number of targets and regulators for this gene, which targets are the genes that your query gene regulates and regulators is the list of proteins that regulate your query gene. Below are tables you can search for your regulator of interest. There will be two tables listing known regulators, a 'manually curated' table and a 'high-throughput' table. The manually curated table lists regulators that have been experimentally tested on a small scale and can be seen as high confidence/ more thoroughly investigated, whereas the 'high-throughput' table lists regulators that have been experimentally tested during an experiment where many other things were tested. These regulators may include transcription factors (TF) that regulate gene transcription.



Searching for our example regulator, Mot3 (TF predicted to bind upstream of proto-genes that are differentially expressed in myriocin treatment) didn't return any results for *TKL2*. This doesn't mean that this gene is not regulated by Mot3, just that there isn't any evidence suggesting it is, which could be because it hasn't been tested for example.

High-throughput 15 entries for 14 genes

Search: Mot3

Regulator	Regulator Type	Direction	Regulation Of	Happens During	Evidence	Strain Background	Reference
No matching records found							

Showing 0 to 0 of 0 entries (filtered from 15 total entries) 10 records per page

Download (.txt) Analyze

Repeat this analysis for your potential regulator.

Answer question 1 in the worksheet: What is your proto-gene's coexpression level with the most upregulated gene (module 1 question 5)? Is your proto-gene strongly coexpressed with this gene?

Answer question 2 in the worksheet: Is the most upregulated gene known to be regulated by one of the predicted regulators from the regulatory motif module (module 2 question 3)? Check the *Regulation* tab on SGD.

Answer question 3 in the worksheet: What is your proto-genes coexpression level with the most downregulated gene (module 1 question 6)? Is your proto-gene strongly coexpressed with this gene?

Answer question 4 in the worksheet: Is the most downregulated gene known to be regulated by one of the predicted regulators from the regulatory motif module (module 2 question 3)?

- Go back to the ORF information app: <https://carvunislabs.cs.pitt.edu/shiny/coexpression/>
Type your proto-gene name in the search box and select the '**Coexpression**' in the '*Result Type*' drop down menu.
Click on the **Gene set enrichment** tab to learn about what cellular processes the genes your proto-gene is coexpressed with are involved in. The table displays the cellular pathway, the padj, and the normalized enrichment score. The padj represents how likely

the result is due to random chance, i.e. a smaller padj number means the result is more likely to be biologically meaningful and not due to random chance. The NES is the normalized enrichment score which reflects the degree to which the genes coexpressed with your proto-gene are involved in this biological process, where a higher number means a larger proportion of the genes are involved (values > 1), and a smaller number means a smaller proportion (values < 1). Click on the '**NES**' column to sort from highest to lowest enrichment score.

Coexpression Info for ORF
YDR042C

Gene set enrichment

Coexpression Network

Gene Set Enrichment Analysis

Download GSEA Results

Show

10

entries

Search:

GO ID	TERM	padj	NES
GO:0048646	anatomical structure formation involved in morphogenesis	2.7e-23	3.77
GO:0006259	DNA metabolic process	1.1e-18	2.75
GO:0051276	chromosome organization	1.0e-11	2.64
GO:0051726	regulation of cell cycle	1.0e-11	2.58
GO:0006974	cellular response to DNA damage stimulus	1.0e-10	2.44

Answer question 5 in the worksheet: What are the top biological processes that genes coexpressed with your proto-gene are involved in?

Sources

- van Dam, S. et al. (2017). Gene co-expression analysis for functional classification and gene–disease predictions, *Briefings in Bioinformatics*. <https://doi.org/10.1093/bib/bbw139>
- Rich, A. et al. (2023). Exploring the noncanonical translome using massively integrated coexpression analysis, *bioRxiv*. <https://doi.org/10.1101/2023.03.16.533058>