

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 A is highly expressed, gene B tends to be also highly expressed, and when gene A is lowly expressed, gene B is also lowly expressed, we would then say that gene A and gene B are coexpressed.

Coexpression can be used for identifying 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.

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.

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

- 1) Identify canonical and noncanonical genes coexpressed with a given ORF
- 2) Visualize coexpression data as network
- 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 ORF ID for your proto-gene in the search box and in the 'Result Type' menu select 'Coexpression'

ORF Information App

Enter your ORF name here

Result Type

3. The first table shows 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).

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.

Coexpression Info for ORF YBR196C-A

Coexpression Network

[Gene set enrichment](#)

Coexpression

Filter by ORF classification:

all

Show 10 entries

Search:

ORF ID	Systematic name	Gene name	ORF classification	Coexpression (rho)
orf17727			noncanonical	0.885
orf227059	YOR002W	ALG6	canonical	0.875
orf69823			noncanonical	0.864
orf143867			noncanonical	0.86
orf20744	YBR196C-B		canonical	0.859
orf237351			noncanonical	0.858
orf101037			noncanonical	0.857
orf121569			noncanonical	0.856
orf254537	YPL028W	ERG10	canonical	0.856
orf69196			noncanonical	0.852

Showing 1 to 10 of 11,629 entries

Previous

1

2

3

4

5

...

1,163

Next

For the top ten ORFs coexpressed with your proto-gene, how many are canonical ORFs (ie annotated genes)?

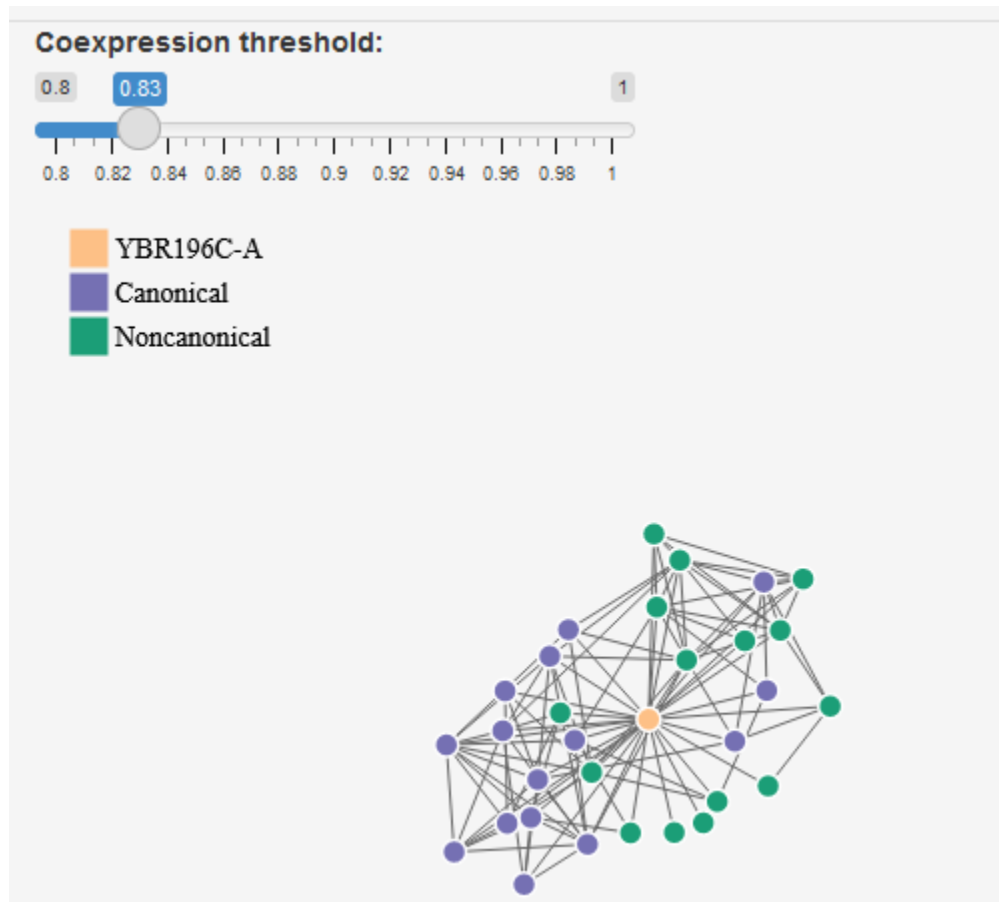
What is the canonical gene most coexpressed with your proto-gene? Use SGD to find out where this ORF is located and what its function is.

What is the noncanonical ORF most coexpressed with your proto-gene? Search for this ORF on the ORF information app to see where in the genome it is located.

Answer questions 1-3 in the worksheet.

4. Coexpression data is often visualized as a network, where each node represents a gene or protogene and lines between nodes represent coexpression. By constructing a coexpression network, you can see the overall structure and organization of the transcriptional co-regulation of your proto-gene and see how many other genes have similar expression patterns.

Beside the coexpression table is a visualization of your proto-gene's coexpression network. The yellow node represents your proto-gene and all other nodes represent other ORFs your proto-gene has similar expression patterns with. As you increase the coexpression threshold, only ORFs that are more highly coexpressed will remain.



Answer question 4 in the worksheet: What do you notice about the network as you increase the threshold? Do you primarily see canonical (purple nodes) or noncanonical ORFs (green nodes) in the network?

5. Next click on the 'Gene set enrichment analysis' 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 Network Gene set enrichment

Gene Set Enrichment Analysis

Show 10 entries Search:

pathway	TERM	padj	NES
GO:0006629	lipid metabolic process	1.0e-20	4.02
GO:0006486	protein glycosylation	3.1e-13	3.89
GO:0006497	protein lipidation	7.4e-09	3.27
GO:0051603	proteolysis involved in cellular protein catabolic process	9.4e-11	3.19
GO:0006873	cellular ion homeostasis	6.0e-09	3.16
GO:0042592	homeostatic process	6.7e-09	2.96
GO:0006464	cellular protein modification process	1.1e-10	2.64
GO:0016192	vesicle-mediated transport	1.8e-07	2.53
GO:0006811	ion transport	1.5e-05	2.45
GO:0051604	protein maturation	1.8e-04	2.45

Showing 1 to 10 of 35 entries Previous 1 2 3 4 Next

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>