Interaction Networks \2017

Practical 8 - Comp Gen 2017

Assistants: Daniel Morgan and Mateusz Kaduk

OBJECTIVES

1. Characterize the topology of functional association networks
2. Compare networks of differentially expressed genesets found in a yeast chromosome via state of the art networks analysis tools (FunCoup, STRING).
3. Analyze for significant enrichment in differentially expressed genes (PathwAX, DAVID).

ACTIVITY

Perform the following steps in this order

Comparative network analysis using STRING

1. Extract the networks for all your bacterial and eukaryotic genomes from the STRING (protein.links.v10.txt.gz).
   Tip1: The file is very large don't decompress it but use python gzip.open() to extract the links for your species. Proteins in the links file are prefixed with the NCBI taxonomy ID of the species.
   Tip2: Python ETE3 package contains NCBITaxa class which can be used to translate NCBI taxa identifiers to species names.
2. Write a Python script to calculate the average connectivity (nr of links/nr of proteins) for each interactome.
3. Plot the degree distribution as a log-log scale scatter plot for each interactome. On

one axis of the plot should be the node degree the other axis the frequency. Do you observe a power-law distribution?

Experimental gene sets

Imagine an experiment has been performed and e.g. differentially expressed genes have been detected. Obviously, these would be a subset of your yeast chromosome.

1. Download the *S.cerevisiae* S228C proteome from Uniprot (http://www.uniprot.org/proteomes/) and use BLAST to match your predicted genes against UniProt proteome.
2. Parse blast results to extract gene names (GN=) for genes present on chromosome.
3. Find two experimental gene sets (experiments.txt) that overlap most with genes on your chromosome. The goal is to find out if any of these are present on the chromosome.
4. Report how many genes overlap and save two experimental gene sets for further analysis.

Comparative network analysis using FunCoup and STRING

1. Using your two experimental gene sets query FunCoup and STRING for sub-networks containing these genes.
    a. FunCoup works with space delimited list of gene names, STRING expects each gene name in new line for a query.
    b. Use the same expansion depth or max number of interactors for searching both databases, so the results are comparable.
2. Compare results
    a. How do these networks differ in terms of nodes, links, and hubs (the three nodes with highest degree)?
    b. What is the most common evidence type with high confidence (>0.9)?
    c. Can you explain the differences in terms of underlying data sources in the databases?

Enrichment analysis using PathwAX and DAVID

1. Analyze the same two experimental sets for enrichment of pathways via PathwAX and DAVID, using KEGG.
    - Which pathways are enriched ?
    - Disparity between results from PathwAX and DAVID ?
    - Does the number of (input) genes matter for results ?
    - If so explain why