Yeast ORFan Gene Project: Module 8 Guide

Genetic and Physical Interactors and Expression Data

In the last module we investigated phenotypes that occur in strains where your gene has been deleted. These studies included looking for other gene mutation strains that demonstrated the same phenotypes as your mutated strain – for instance in response to treatment with various drugs. These data could help us predict the function of your gene, as it may be similar to a pathway that these other genes are involved in. In this module we will again explore your gene in the context of other genes, starting by searching for genetic interactors, then examining physical protein interactors and finally studying gene expression data in response to various stimuli.

GeneMania

GeneMANIA helps you predict the function of your favorite genes and gene sets. GeneMANIA finds other genes that are related to a set of input genes, using a very large set of functional association data. Association data include protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity. You can use GeneMANIA to find new members of a pathway or complex, find additional genes you may have missed in your screen or find new genes with a specific function, such as protein kinases. Your question is defined by the set of genes you input. If members of your gene list make up a protein complex, GeneMANIA will return more potential members of the protein complex. If you enter a gene list, GeneMANIA will return connections between your genes, within the selected datasets.*

Navigate to GeneMania at http://www.genemania.org/. On the search bar in the upper left corner click on the picture of the *Homo sapiens*, a drop down menu should appear, change the selection to "**Saccharomyces cerevisiae**" and enter your gene name into the search box and start the search.

You will be routed to a new page that displays the known interactors, co-expression, co-localization and protein domain similarity data for your gene of interest. Before we discuss how to navigate and mine for data let's first look at what all of these categories mean. GeneMANIA searches many large, publicly available biological datasets to find related genes. These include protein-protein, protein-DNA and genetic interactions, pathways, reactions, gene and protein expression data, protein domains and phenotypic screening profiles. Data is regularly updated. Networks names describe the data source and are either generated from the PubMed entry associated with the data source (first author-last author-year), or simply the name of the data source (BioGRID, PathwayCommons-(original data source), Pfam)*

- **Co-expression:** Gene expression data. Two genes are linked if their expression levels are similar across conditions in a gene expression study. Most of these data are collected from the Gene Expression Omnibus (GEO); GeneMania only collects data associated with a publication.*
- **Physical Interaction:** Protein-protein interaction data. Two gene products are linked if they were found to interact in a protein-protein interaction study. These data are collected from primary studies found in protein interaction databases, including BioGRID and PathwayCommons.*
- **Genetic Interaction:** Genetic interaction data. Two genes are functionally associated if the effects of perturbing one gene were found to be modified by perturbations to a second gene. These data are collected from primary studies and BioGRID.*
- **Shared protein domains:** Protein domain data. Two gene products are linked if they have the same protein domain. These data are collected from domain databases, such as InterPro, SMART and Pfam.*
- **Co-localization:** Genes expressed in the same tissue, or proteins found in the same location. Two genes are linked if they are both expressed in the same tissue or if their gene products are both identified in the same cellular location.*

* Information regarding the functionality of each software/algorithm and operational information was taken directly from the websites and included here for instructional purposes only.

Yeast ORFan Gene Project: Module 8 Guide

- **Pathway:** Pathway data. <u>Two gene products are linked if they participate in the same reaction within a pathway.</u> These data are collected from various source databases, such as Reactome and BioCyc, via PathwayCommons.*
- **Predicted:** <u>Predicted functional relationships between genes, often protein interactions. A major source of predicted data is mapping known functional relationships from another organism via orthology. For instance, two proteins are predicted to interact if their orthologs are known to interact in another organism.</u> Also, GeneMania includes predicted functional associations from other groups that combine multiple data sources for a given organism, e.g., the entire YeastNet predicted network: Lee-Marcotte-2007 YeastNet; the genetic interaction data used to generate YeastNet: Lee-Marcotte-2007 Genetic interactions. In these cases, the network name indicates the original publication detailing the predicted network, and (in some cases) lists the individual network that was used to generate the entire predicted network (for latter example above). *
- **Other:** <u>Networks that do not fit into any of the above categories.</u> Examples include phenotype correlations from Ensembl, disease information from OMIM and chemical genomics data.*

Networks link interacting or functionally associated genes
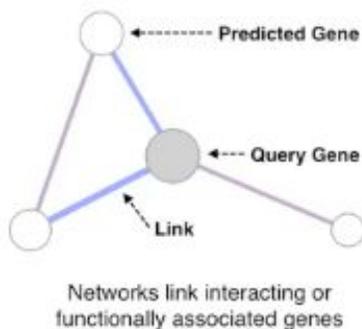
You can change the network visualization in many ways, which will ease the ability to focus on specific types or networks and create a less complicated picture. This network is presented in a basic node and edge design – where each node (or circle) represents a different gene and the edges (or connector lines) and their color represent the type of network or data that is being represented. For instance, co-expression data is represented by pink lines and co-localization is represented by blue lines, thus in the diagram to the left your query gene is linked to the predicted gene by co-localization data.*

These data networks are most easily viewed one at a time. To change this setting simply move to the right-hand portion of the screen and select ONLY ONE (for example, Co-expression) data box at a time. Once you have selected one and only one type of data you can also manipulate the network by clicking any given node and dragging it to a new location in the left-side box.

For each of the data sets, individually create the network, copy and paste it into your **Module 8 Worksheet,** by data type. [Note – exporting the file creates a vector image file (.svg) that is best opened with Adobe Illustrator – if you do not have access to this program, using the Snipping Tool (Window) or capturing the image with Grab (Mac) is the best way to save and insert these network graphics.]

To learn more about the genes in any given network you can do several things: 1) Click on the relevant node and gene information will pop up 2) Change to the Genes Tab in the right side panel and gene information is displayed (again clicking a gene will display additional information 3) Look up the gene name in SGD 4) By clicking the arrow next to the type of data displayed in the right side box the papers or program from which the data were mined or generated is displayed, you can find the article or program specific for the interaction of interest and read/run through to get more information. Include summary information of the relevant genes and their functions in your **Module 8 Worksheet.**

* Information regarding the functionality of each software/algorithm and operational information was taken directly from the websites and included here for instructional purposes only.

Yeast ORFan Gene Project: Module 8 Guide

Saccharomyces cerevisiae Genome Database: Interactions Tab

Part of the curation and compilation of the *Saccharomyces cerevisiae* Genome Database (SGD) involves reading all of the pertinent literature and incorporating this data into the website. Part of this curation involves searching for information that has been published from recent work that looks to characterize physical and genetic interactions. This data is collected and stored in the Interactions Tab of your gene's SGD page.

Navigate to SGD at http://www.yeastgenome.org/ and enter your standard gene name in the Search box and click go. When your gene/protein page loads, link onto the Interactions Tab. Scroll down to the Interaction Network section and copy interaction maps into your **Module 8 Worksheet**, taking time to switch between Physical and Genetic Interactions maps as you go. Compare results on this map to the one from GeneMania above, note genes found in both data sets and those that are novel. [Note – Snipping Tool (Window) or capturing the image with Grab (Mac) may be the best way to save and insert the table.]

SPELL

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 (even just one), 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. *

Many of the experiments that are mined here are based on DNA microarray technology. This type of microarray is composed of a collection of DNA spots attached to a solid surface, usually some sort of glass slide. The microarrays (small arrays of DNA spots) can then be utilized to simultaneously measure the expression level of all of the genes in any given organisms. This is achieved because each DNA spot (also known as a probe) corresponds to a gene or DNA element and can hybridize to cDNA or anti-sense RNA (also known as the target) that is expressed at any given moment in the cell. Cells can therefore be treated in a certain way, for example in the presence of a stimulus or chemical and then the cDNA and/or anti-sense RNA is harvested and hybridized to the probes on the chip. Probe-target interactions are then visualized by fluorophores or chemiluminescence such that any cDNA that is expressed and hybridizes to the microarray causes a signal to be released, the signal is detected and we can measure that this cDNA is expressed in response to the given stimuli. Furthermore relative abundance can be measured and in this manner tell us if the cDNA (corresponding to a gene) is expressed at high, low or normal levels under the given conditions.*

Navigate to SPELL – *S. cerevisiae* at http://spell.yeastgenome.org/ and enter your standard gene name in the Search box and submit the Search. (Leave the # Results at the default of 20)

You will be directed to a new page where microarray data is displayed for 10 experiments at a time. In this display the published **Dataset** is hyperlinked at the top of each column, underneath which the type of experiments is briefly displayed (**Tags**) and below this there is a description of the type of genes that were found or the type of treatment that was applied (i.e ploidy regulation of gene expression or mitochondrial dysfunction). You will then see a red, green and black graphic chart with gene names listed down the left side. Your gene of interest should be listed in the top, left position.   As described above microarrays can be used to measure increases, decreases and no change results of expression of a given gene – these are represented as green (decrease), red (increase) or black (no change) colorimetric data in the microarray chart. If you click on any given colored bar – the numerical data and changes to stimuli will appear in a new window. Using the expression data shown for you gene across the top row, SPELL has searched and found genes that show similar expression profiles to those seen for your gene. As these data are collected across

---

Yeast ORFan Gene Project: Module 8 Guide

various experiments, and the expression patterns consistently correlate this is strong data that your gene could be functioning in the same, similar or parallel ways as the other genes listed. Clicking on a gene name in the list will link you to that gene's SGD page to find out more description data on its function. Furthermore, below the microarray chart, SPELL has curated and is displaying the GO (gene ontology) data for this list of gene. These will once again be helpful in predicting the type of processes that the other genes, and therefore possibly your gene, are involved in.

[*Note: If no datasets are found to contain a significant signal for a given query then we are unable to assign per-dataset weights for the search. In this case a warning message is displayed, and all datasets are equally weighted for that query. If no genes are related to the query set at a reliable confidence level, then a warning message is displayed and the confidence level is weakened until results can be obtained. Both of these cases typically only occur when the query genes are either largely unrelated, or highly unique. Neither of these cases occurs very often.*]

In your **Module 8 Worksheet** copy and paste the page with microarray data including the Dataset, Tags, Description data and the chart for all 20 genes displayed. Also copy and paste the GO Term Enrichment Data that is located at the bottom of this page into your **Module 8 Worksheet**. [Note – Snipping Tool (Window) or capturing the image with Grab (Mac) may be the best way to save and insert the table.] Investigate the expression data sets to comment on how your gene's expression is changed due to these treatments. Investigate the functionality of the most similarly expressed genes from the list down the left-hand column and comment on their roles in the cell. Investigate the GO Term Enrichment and comment on how this relates to your other data collected to date.