Guide for Bioinformatics Project Module 5

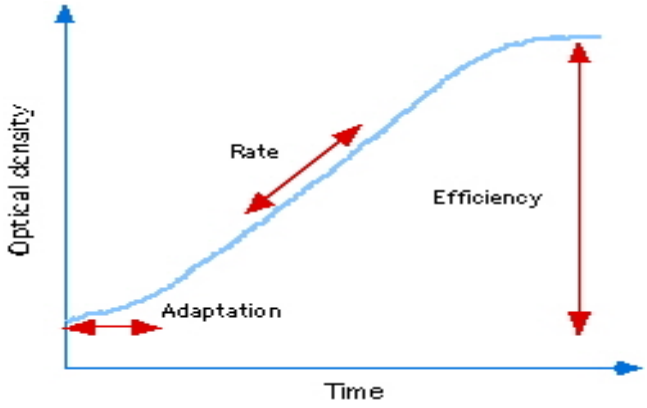
Gene Deletion Phenotypes

Strains have previously been created, as part of the deletion collection, that are either homozygously deleted or heterozygously deleted for your gene of interest. As we have discussed the deletion collection was created in part to allow large, high throughput experiments to be done on 6000+ strains each with a different gene mutation. The data generated from these experiments is publicly available (if previously published) and can be 'mined' to learn new information about what happens to strains if you gene of interest is removed and the cells are measured for some phenotype. We will now mine these data to determine if there is any information that might aid in our determination of the function of your protein.

[Note: The reason we are performing transformations and are making our own knockouts is to put your mutation in a yeast strain background that allows us to ask questions about the role of your gene in genome instability. We are not simply redoing what other scientists have already done.]

PROPHECY : PROfiling of PHEnotypic Characteristics in Yeast

PROPHECY provides quantitative information about phenotypes for the complete collection of deletion strains in yeast (Saccharomyces cerevisiae). PROPHECY evaluates the phenotype of a deletion strain on the basis of growth behavior during micro-cultivation. PROPHECY quantifies growth aberrations by estimating the rate of growth, the efficiency of growth and the adaptation time.



Navigate to PROPHECY at http://prophecy.lundberg.gu.se/Main.aspx and enter your standard gene name in the Quick Search box and submit the Search. [*Note: for some reason with this site you may need to click the Search button twice to get the website to proceed.*]

The data for cells harboring your gene mutation will appear on a new page with graphical information about the growth of these strains in environments of high salt (0.85M NaCl), oxidative stress [cells ability to tolerate production of destructive reactive oxygen species] (Diamide 1.4mM) and (Paraquat 200 µg/ml), reductive stress [ability of cells to maintain redox homeostasis] (DTT 1.6mM). As well, an overall Gene-by-environment phenotype (LPI) summary box is displayed which highlights any significant changes scientists have noted from the provided graphs. Click on each individual graph to open the data in a larger window and see how the growth of cells, mutant for your gene, are affected under these conditions relative to their wildtype (normal) counterparts.

Copy the large version of each graph and incorporate them into the given space in your Module 5 Worksheet. Also report the overall summary from the Gene-by-environment phenotype summary box on the first page and comment on any changes you noticed in the graphs that contributed to the summary provided. Make sure to comment on if the Rate, Adaptation or Efficiency is effected and how you take this into account when predicting what the effect could mean about the function of your protein.

Guide for Bioinformatics Project Module 5

Yeast Fitness Database

This site provides a graphical interface to the supplementary data presented in two publications: "The Chemical Genomic Portrait of Yeast: Uncovering a phenotype for all genes" (Hillenmeyer et al., Science 2008) and "Chemical genomic profiles predict gene function and mechanism of drug action" (Hillenmeyer et al., Genome Biol. 2010). This data should therefore provide us with information about the phenotype of cells with your gene mutated when exposed to various drugs and then based on these data the program will report what other gene mutations show similar sensitivities to the same drugs to which your mutated strains is sensitive.

Navigate to the Yeast Fitness Database at http://fitdb.stanford.edu/fitdb.cgi and enter your standard gene name in the Search box and submit the Search.

Sensitivity data can be informative based on how and what the drug is known to target. Within Section 1: Phenotypes scroll down to the Homozygous knockout section and identify the top 3 drugs with significant p-values (<0.05) and perform an Internet search to determine how this drug works in a cell.

[*Note: not all of the 'conditions' listed are drugs, you will have to determine by searching or by asking your professor – for instance ph8 is not a drug, it's a pH change to basic conditions and sorbitol is not a drug it's a sugar.*] (You will need to separately look up the mechanism of action for the drugs showing significant fitness defects because the list of tested drugs is too long to provide this information in this document).

It is important to research the CELLULAR role of your drug, for example if you only determine that doxorubicin is a chemotherapy drug this doesn't provide enough information to figure out what your protein might be doing. As well, if you find out that doxorubicin can cause heart damage this information will not be informative in our yeast module (since they don't have hearts). With further research you can elucidate the CELLULAR level mechanism of action and determine that the doxorubicin intercalates into DNA blocking topoisomerase II progression thereby inhibiting transcription. This level of detail is what you need to find in order to be useful in hypothesis generation or modification.

In your Module 5 Worksheet record the drug, the Fitness Defect (z-score), the p-value and the mechanism of action of the drug (include the citation of where you found this information).

In Section 2: Co-fitness interactions, YFD uses a new algorithm to identify cells with other mutated genes that show similar sensitivities to the drugs that your cells do; this can imply a similarity in function. **Scroll down to Homozygous co-fitness interactions section and look up the functions of the 10 given genes (listed under the Interactor column) and record that information in your Module 5 Worksheet, take time to think about how these data fit together and then generate a hypothesis.** [*To look up the genes and their known functions navigate to SGD and type the gene names into the search box, when the gene page loads copy the Description portion into your worksheet.*]

Not only are the names of these genes provided for you to investigate their known functions through SGD – but they have also been classified by Gene Ontology groupings (to the right of the homozygous co-fitness interaction diagram). Once again this data can be useful because it may indicate that your gene/protein will have a similar function and could be assigned to the groups/ontologies that are identified. **Record any Enriched GO terms of interactors that are indicated on your Yeast Fitness Database page in your Module 5 Worksheet.** If you do not know what any of the terms in the GO annotation mean look them up separately and together – and ask for help from your Professor if you are still confused.

Guide for Bioinformatics Project Module 5

Phenotype TAB in *Saccharomyces cerevisiae* Genome Database (SGD)

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 involved searching for information that has been published about resulting phenotypes due to mutation of your gene of interest. This data is collected and stored in the Phenotype Tab you can access from your gene homepage in SGD.

Navigate to the 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 Phenotype Tab. **Copy the Single Mutant Phenotypes Table into your Module 5 Worksheet, taking time to look up and research any terms, experiment types, phenotypes, chemicals or details presented here that you are unfamiliar with.** [*Note: It may be helpful to use the Snipping Tool to cut and paste the chart.*]