Copyright © Orly Alter 2017

Databases

Problem 1:
Access the data from the publication –
http://www.alterlab.org/teaching/BIOEN6770/papers/Spellman_1998.pdf
at the Princeton University Microarray Database (PUMAdb).
Record the steps that lead you from – http://puma.princeton.edu/
to – http://puma.princeton.edu/cgi-bin/publication/viewPublication.pl?pub_no=90

Click on the "search by data set," pick a results type of "publication," then chose the organism "Saccharomyces cerevisiae" and data identifier "Spellman PT et al. (1998) Mol Biol Cell 9:3272-97" before clicking on the "display data" button.

Problem 2:
At the Gene Ontology (GO) website or at the Saccharomyces Genome Database (SGD), find the GO IDs for the biological processes "cell cycle" and "response to pheromone." Download and prepare a list of the yeast open reading frames (YORFs) that correspond to the GO term "response to pheromone."

ID for "cell cycle" is GO:0007049
http://amigo.geneontology.org/cgi-bin/amigo/term_details?term=GO:0007049
ID for "response to pheromone" is GO:0019236
http://www.yeastgenome.org/cgi-bin/GO/goTerm.pl?goid=0019236

http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/2/ResponseToPheromone.txt

Problem 3:
For the list of "response to pheromone" YORFs and, separately, for the cell cycle genes selected by Spellman et al. –
http://genome-www.stanford.edu/cellcycle/data/rawdata/ or
http://www.alterlab.org/teaching/BIOEN6770/labs/Spellman_Cell_Cycle.txt –
download the Spellman et al. alpha-factor block-release data as follows:

a) Gene Selection and Annotation:
Enter the list of YORFs alphabetically sorted. Use experiment name. Include "Biological Process," "Cellular Component," and "Molecular Function," in the biological annotations for the genes. Where are these annotations taken from?

b) Data Filtering Options:
Download default "Log(base2) of R/G Normalized Ratio (Mean)." Select only features with no flag. Select default filters 2 AND 3 with the cutoff >1.2. Retrieve spot coordinates.

c) Gene Filtering Options:

Center data for each gene by mean. Do not filter genes on the basis of data values. Use only genes with >99% valid data.

d) Clustering and Image Generation:
Use default clustering. Choose a contrast of 1. Choose a color scheme.

http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/3/CellCycle_Report.html
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/3/ResponseToPheromone_Report.html

e) Download or print into a file both the red/green (or yellow/blue) raster and the spot image displays. Also download the clustered data ".cdt" file.

http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/3/CellCycle_Raster.png
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/3/CellCycle_Spot.gif
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/3/CellCycle.cdt

http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/3/ResponseToPheromone_Raster.png
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/3/ResponseToPheromone_Spot.gif
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/3/ResponseToPheromone.cdt

f) Compare the results for the Spellman cell cycle genes (or YORFs) with the results for the "response to pheromone" genes (or YORFs).

It appears that the red to green shift of the most prominent gene cluster in the pheromone response raster is followed by a green to red shift in the uppermost cell cycle cluster and a red to green shift in the midway cell cycle cluster. Interestingly, both of these cell cycle gene clusters contain genes for cyclin-dependent protein kinase regulator activity.

g) Compare the raster and spot image displays. Can you detect similar expression patterns in both displays?

Yes, there are similar expression patterns present in the raster and spot image display for a given gene set.

Problem 4 (Extra Credit):
Compare your results to figures 1A and 1B in Spellman et al., *MBC* (1998), pp. 3280 and 3281.

Note that ordering the cell cycle genes by cell cycle phase, as in figure 1A, rather than clustering by gene expression pattern, as in figure 1B, should reveal a traveling wave with a distinctly diagonal slant from the upper left to lower right of the display. The raster pattern would then seem to reflect the order of expression in the cell cycle as a whole. Also note that the cell cycle phase was determined by using Fourier analysis of the data, as in equations 1–4.

http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/4/CellCycle_Report.html
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/4/CellCycle_Raster.png
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/4/CellCycle_Spot.gif
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/4/CellCycle.cdt

However, currently the database output the genes in an order different than the input order. Therefore, repeating problem 3 with the list of YORFs sorted according to their cell cycle phases as determined by Spellman et al., and selecting "no gene clustering" in addition to the default "no experiment clustering" under Clustering and Image Generation, gives the genes ordered by the internal database order rather than by the input cell cycle phase order.

Problem 5 (Extra Credit):
Repeat problem 3(a–e), making your own choices in each step. For example, select different gene lists based on specific GO terms, different cutoffs, or even different data to download. Describe and explain your choices. What do you learn from the resulting raster and spot displays?

http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/5/Genome_Report.html
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/5/Genome.cdt

Out of the 6,194 yeast genes which passed the initial data channel filters, only 4,397 passed the gene filter for >99% good data. Disabling the gene filter allowed the full 6,194 yeast genes to pass.

Problem 6 (Extra Credit):
Repeat problem 3(a–e) for all yeast genes. How many genes pass your filter? What happens when all gene filters are disabled?

Filtering the data by using the GO term for DNA replication (ID GO:0006260)
http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0006260
we find a red-green raster pattern similar to those in the clustered cell cycle raster, confirming my preexisting knowledge that DNA replication occurs during the cell cycle.

http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/6/DNAReplication_Report.html
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/6/DNAReplication_Raster.png
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/6/DNAReplication_Spot.gif

http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/6/DNAReplication.cdt

Problem 7 (Extra Credit):
Separately sort each of the samples in the .cdt dataset you downloaded in problem 3(e), once in increasing and once in decreasing order. Upload the sorted list of genes into GOrilla at – http://cbl-gorilla.cs.technion.ac.il/
Tabulate the key results for each sample. What do you learn from this analysis?

Using the data from problem 5, for the 4,397 genes in the 6th time point, we find that sorted from the most to the least overexpressed, the ordered gene list is enriched in "nucleosome assembly" and related annotations. This suggests that this time point is during the cell cycle S-phase, when DNA is being replicated.

http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/7/OverexpressedProcess_Tree.png
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/7/OverexpressedProcess_pValues.png
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/7/OverexpressedFunction_Tree.png
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/7/OverexpressedFunction_pValues.png
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/7/OverexpressedComponent_Tree.png
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/7/OverexpressedComponent_pValues.png

When the genes are sorted from the most to the least underexpressed, the ordered gene list is enriched in "cytokinetic process" and related annotations. This suggests that this time point is during the cell cycle S-phase, when the M phase-exclusive cytokinetic process is being suppressed.

http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/7/UnderexpressedProcess_Tree.png
http://www.alterlab.org/teaching/BIOEN6770/labs/Lab_1/7/UnderexpressedProcess_pValues.png