Module on Microarray Statistics for Biochemistry: Metabolomics & Regulation
Part 1: Spot Gridding, Fold-Change Selection, Clustering
By Johanna Hardin and Laura Hoopes
Instructions and worksheet to be handed in
NAME_____

Lecture/Discussion on microarrays and spot gridding
- What is a microarray? (refer to "Intro to lab" information)
- Spot calling (addressing)
- Filtering: intensity, flag
- Gene filtering vs simple filtering
- Segmenting

Reading assignments: DeRisi et al, 1997 (see first lab information)
Draghici, S (2003) *Data Analysis Tools for DNA Microarrays*
Chap 2-3.

Dry lab work:

A. Go to http://www.bio.davidson.edu/courses/genomics/chip/chip.html and enjoy the animation in order to review how the expression microarrays are performed (what you are doing in lab). (Animation by Malcolm Campbell)
Comments on animation compared to what we did in lab:

B. Go to: http://www.bio.davidson.edu/projects/MAGIC/MAGIC.html
Download the MagicTool software so that you can gain insight into spot gridding. Download the User's Guide for use in this part and the next part of the exercises. (MagicTool by Davidson College undergraduates; exercises by Dr. Laurie Heyer)
Using the User's Guide, pp 1-14, work through the exercises on gridding, using the **1 grid** of data you can download from the site. Experiment with the different methods of spot fitting to see how much they affect ratios of reddish and greenish spots. Fill in the largest difference in ratio you were able to obtain by altering the spot fitting method of a single spot: _____
Comments on addressing and segmenting:

C. Go to http://gcat.davidson.edu/GCAT/workshop2/derisi_lab.html
1. Download the Dry-Lab Instructions to Explore DeRisi Experiment on diauxie. Choose 'Creating the Project'.
2. Use the command **Merge Expression Files** to combine the existing expression file **derisi-first5.exp** and the existing expression file **derisi-Last2.exp**. Be sure the list the files in this order, and change the nicknames for both files to t. Call the merged file derisi.exp.

3. After the merge is complete, examine **derisi.exp** using **View/Edit Data** to make sure the column labels are in order. (You want them ordered by time, as the cells enter diauxie).
4. Add the gene information in **yeastgenes.info** (part f on the same menu) to **derisi.exp**, forming **derisi_1.exp**. Use this merged and annotated file, which is the compelte time-course published by DeRisi et al., to answer the questions below.

Question list for dry lab exercise:

1. How many genes' expression change by at least a factor of 2 in the first two hours? (see p 680 in the paper by DeRisi)

2. How many genes' expression are greater than 2.0 or less than 0.5 in the time 0 microarray? How does this affect your interpretation of the answer to question 1?

3. How many genes' expression increase by a factor of at least 4 some time during the time course? How many genes' expression decrease by a factor of at least 4 some time during the time course? Compare testing for these numbers using MagicTool with doing the comparison in Excel in terms of difficulty. (p 680)

4. Investigate the change in expression of ribosomal genes by forming a group of ribosomal genes, plotting the group, and highlighting the mitochondrial genes in the plot (relate to p 681) What did you find out?

5. Find genes with the 'late induction profile' described on p 681, and graphed in Figure 5B, in which ratios increased by more than 9 fold at the last time point, but less than 3 fold at the preceding time point. Compare your results to those in Figure 5B and use http://www.yeastgenome.org to investigate and try to explain any differences. Explain what you found and how it compares:

6. Add the file **derisi_lab_i_tlog2.dis** to the project to help you answer the following questions. This file was generated by transforming the ratios with log

base 2, then computing dissimilarities using (1 – correlation). The process of calculating dissimilarities takes several hours even on a fast computer, so we are skipping this step in this laboratory. Form a supervised cluster with *SAM1*(YLR180W) as the seed gene, and compare your results to Figure 5E

a. Using 0.2 as the threshold
b. Using 0.02 as the threshold

Describe your findings:

7. If you didn't know to expect ribosome genes or *SAM1* to change, you would need to cluster the genes without supervision. Try unsupervised QT clustering with a threshold of 0.3 and a maximum number of clusters of 30. Compare the clusters you got to see if any of them might really belong to the same cluster when you think biologically. Comment on your findings:

8. Look again at the **derisi_1.exp** using **View/Edit Data** to remind yourself of what the format for this file needs to be. Transform the ratios from your class data into the same format with the genes in the same order (the order need not be the same as in the original file, but if not, call the file **derisi_2_class.exp**). Now examine questions 3-5 again using this data file. Describe any differences: