

# Recombinant DNA Techniques

---

We will be using yeast microarray data for our microarray analyses.

## Using MagicSpot

---

### Tutorial

1. In the Genomics CD, double click on the MagicTool directory.
2. Double click on the MAGIC\_Users\_Guide.pdf file.
3. Go to page 6, Getting Started.
4. To start the MagicSpot program, select MagicTool.bat.
5. Go through the tutorial. The following files are the ones mentioned in the tutorial:
  - a. When you start a new project, enter any name that you can remember.
  - b. The files listed below are located on the CD in the **RecomMicroarrayData** directory.
  - c. The control R and G files are green1.tif and red1.tif (for the single grid).
  - d. The gene list file is Y01godlist1grid.txt. (for the single grid).
  - e. Go through step 25. This will give you a basic understanding of how to compile the data for the microarray.

### On your own

1. Repeat the steps in the tutorial for the 4 grid microarray. Use the same file names as in the tutorial, except that the 1's are changed to 4's (e.g. red4.tif instead of red1.tif).

## Using MagicSpot to analyze the yeast data

---

### Tutorial

**Note:** This is modified from the one at [http://gcat.davidson.edu/GCAT/workshop2/derisi\\_lab.html](http://gcat.davidson.edu/GCAT/workshop2/derisi_lab.html)

1. Open MAGIC Tool using the launcher, which gives you access to additional RAM.
2. Create a project called derisi\_lab. This creates a folder called derisi\_lab, containing a file called **derisi\_lab.gprj**. (See the [MAGIC Tool project tutorial](#) for a reminder of how projects are organized.)
3. Copy the files in the “derisi lab project” folder from the CD to the derisi\_lab folder on the computer.
4. Go back to MAGIC Tool and select “Update Project” under the Project menu. This copies all the files from the top-level project folder into their appropriate subfolders. No will probably get an error message about DeRisiGodList.xls. Ignore it.

## Generating expression ratios

5. In MAGIC Tool, load the Red and Green image files for OD 6.9 [the ones with 690 in the name]. Use `derisi_genelist.txt` as the gene list. When you begin the addressing step, you can either practice creating a new grid, or open the saved grid `1309.grid`.
6. During segmentation, create two different expression files:
  - a. Using fixed circle with a radius of 3 pixels, and total signal (without background subtraction) create a new file named **my3\_10**, labeling the column 10 (the number of hours that have passed between OD 0.14 and OD 6.9).
  - b. Using fixed circle with a radius of 5 pixels, and total signal (without background subtraction) create a new file named **my5\_10**, once again labeling the column 10.
7. Repeat steps 5 and 6 for the OD 7.3 array [730], with the same settings as above. You must exit the program and restart it to load new image files.

For the OD 7.3 array, the alternative to creating your own grid for addressing is to use `1313.grid`. During segmentation, append the 7.3 data to the 6.9 data in files called **my5\_last2**, and **my3\_last2**. In each file, label the current column 12.

8. Now we will see how your data compares to the published DeRisi data.
  - a. Use the command “Merge Expression Files” to combine the two expression files that you have just created, `my5_last2.exp` and `my3_last2.exp`, calling the result **my\_last2.exp** (override the default name by simply typing over it). Accept the default nicknames for the two files, which will be appended to the column names. The merge will take a few minutes; you will not be able to open any menus until it is done.
  - b. Use the command “Merge Expression Files” to combine the existing expression file `derisi_last2.exp` with the merged expression file you just created, calling the result **all\_last2.exp**.

**Important:** you must select `derisi_last2.exp` as File #1, because all genes in File #1 need to be in File #2 for the merge to work properly.
  - c. Log base 2 transform the expression file (expression, manipulate data, transform).
  - d. From the Explore window, perform **two-column plots** comparing your 3 pixel segmentation to the published DeRisi data for the OD 7.3 array (12 hours into experiment), and your 5 pixel segmentation to the published DeRisi data for that same array (expression, explore, two column plot). Each plot will take a minute or so to appear, so be patient.
    - i. Click on an outlier point in one of the plots, turning the point red and causing the ORF name to appear in the bottom right corner.
    - ii. Go to the other plot, and select the **same gene** from the drop-down menu in the bottom right corner. Is the ratio in the second plot closer to the published data, or even more different?
    - iii. Go back to Segmentation in the Build Expression File Menu, which should still contain the OD 7.3 array. Jump to the gene you identified in step (i), and try to explain why the ratio at this particular spot was difficult to determine. Experiment with different segmentation methods to see what you think the

- best answer is for the ratio at this spot. [An answer is at [http://gcat.davidson.edu/GCAT/workshop2/ans\\_2col.html](http://gcat.davidson.edu/GCAT/workshop2/ans_2col.html)]
- iv. As time permits, explore more outliers in the first set of plots, and/or repeat the analysis with the OD 7.3 array.
  - v. Explain why it was important to log transform the data before looking for outliers in the two-column plots. [An answer is at [http://gcat.davidson.edu/GCAT/workshop2/ans\\_2col\\_log.html](http://gcat.davidson.edu/GCAT/workshop2/ans_2col_log.html)]

### Analyzing expression ratios

9. 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 to list the files in this order, and change the nicknames for both files to `t`. Call the merged file `derisi.exp`. After the merge is complete, examine `derisi.exp` using “View / Edit Data”, to be sure the column labels are in order.
10. Add the gene information in `yeastgenes.info` to `derisi.exp`, forming `derisi_i.exp`. Use this merged and annotated file, which is the complete time course published by DeRisi, to answer the remaining questions.
11. How many genes’ expression change by at least a factor of 2 in the first two hours? [An answer is at [http://gcat.davidson.edu/GCAT/workshop2/ans\\_early\\_factor2.html](http://gcat.davidson.edu/GCAT/workshop2/ans_early_factor2.html)]
12. 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 #3? [An answer is at [http://gcat.davidson.edu/GCAT/workshop2/ans\\_zero\\_factor2.html](http://gcat.davidson.edu/GCAT/workshop2/ans_zero_factor2.html)]
13. How many genes' expression increases by a factor of at least 4 sometime during the time course? How many genes' expression diminishes by a factor of at least 4 sometime during the time course? [An answer is at [http://gcat.davidson.edu/GCAT/workshop2/ans\\_factor4.html](http://gcat.davidson.edu/GCAT/workshop2/ans_factor4.html)]
14. 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. [An answer is at [http://gcat.davidson.edu/GCAT/workshop2/ans\\_mrp\\_rp.html](http://gcat.davidson.edu/GCAT/workshop2/ans_mrp_rp.html)]
15. Find genes with the "late induction profile" in which levels increased by more than ninefold at the last timepoint, but less than threefold at the preceding timepoint. [An answer is at [http://gcat.davidson.edu/GCAT/workshop2/ans\\_late\\_induct.html](http://gcat.davidson.edu/GCAT/workshop2/ans_late_induct.html)]

### On your own

For your microarray project, each person will examine a different set of genes. For this set of genes, you will examine whether the changes in expression when comparing the anaerobic vs. aerobic over time makes sense.

The following projects can be done. If you wish, you can look at some other set of genes, but it must be approved by the instructor. In general, you want to look at approximately 10 genes. If the process has 20 or 30, examine a subset of them.

1. Ten genes with the largest increase [aerobic vs. anaerobic]

2. Ten genes with the largest decrease [aerobic vs. anaerobic]
3. Genes in glycolysis
4. Genes in citric acid cycle (tricarboxylic acid cycle)
5. Genes involved in DNA replication
6. Genes binding/modifying chromatin
7. Genes involved in protein biosynthesis
8. Genes involved in transport
9. Genes in ATP synthase
10. Genes in electron transport
11. Genes involved in fermentation
12. Mitochondrial proteins

Look at whether the patterns of expression are the same, whether the relative expression is the same, etc. Should the group of genes go up or down when comparing the two conditions? Did they? You will need to determine the known function(s) of each of the genes and use the function to determine if the up/down regulation makes sense.