Chapter 1

DNA Microarrays: Background, Interactive Databases, and Hands-on Data Analysis

***A. Malcolm Campbell*[1] *and Laurie J. Heyer*[2]**

[1]Biology Department
Box 7118
Davidson College
Davidson, NC 28035
macampbell@davidson.edu

[2]Mathematics Department
Box 6959
Davidson College
Davidson, NC 28035
laheyer@davidson.edu

Malcolm Campbell obtained his Ph.D. from Johns Hopkins University in 1992. He was awarded a Pew Teacher-Scholar postdoctoral Fellowship to help make the transition from bench to classroom. Campbell is the founding Director of the Genome Consortium for Active Teaching (GCAT; <*www.bio.davidson.edu/GCAT*>). He is also co-Editor-in-Chief of Cell Biology Education, an online journal for the life sciences <*www.cellbioed.org*>. In the fall of 2002, Campbell and Laurie Heyer co-authored the first true genomics textbook *Discovering Genomics, Proteomics and Bioinformatics,* which was published jointly by Cold Spring Harbor Laboratory Press and Benjamin Cummings. Campbell enjoys teaching introductory biology, immunology, molecular biology and genomics.

Laurie Heyer received her Ph.D. in Applied Mathematics from the University of Colorado at Boulder in 1998. She was a postdoctoral fellow in the Center for Computational and Experimental Genomics at the University of Southern California. Her primary interests are in mathematical modeling (especially of genomes and cellular processes), probability theory and operations research. Heyer teaches a broad array of mathematics courses, including computational biology.

Reprinted From: Malcolm, A. and L. J. Heyer. 2004. DNA Microarrays: Background, interactive databases, and hands-on data analysis. Pages 1-39, *in* Tested studies for laboratory teaching, Volume 25 (M. A. O'Donnell, Editor). Proceedings of the 25th Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 414 pages.

Although the laboratory exercises in ABLE proceedings volumes have been tested and due consideration has been given to safety, individuals performing these exercises must assume all responsibility for risk. The Association for Biology Laboratory Education (ABLE) disclaims any liability with regards to safety in connection with the use of the exercises in its proceedings volumes.

Contents

Introduction

This paper describes a series of exercises that can be used in a modular fashion. The introduction to DNA microarrays could supplement any class that addresses the topic. The interactive questions can be used to emphasize the introduction, as homework, or as a way to illustrate the immense power of public domain databases containing DNA microarray datasets. Finally, the software MAGIC Tool permits faculty to develop hands-on activities for students in either a laboratory setting, or as homework assignments to be discussed in class. MAGIC Tool is free, as are the raw data files so cost is not an issue. These resources could be used to augment a seminar course where students read primary literature including papers that utilize DNA microarrays.

DNA microarrays are best saved for upper level students who have already successfully completed Genetics. They need to be very familiar with central dogma of DNA → RNA → Protein. Students will catch on faster if they are familiar with cDNA and molecular methods such as Southern and Northern blots.

The time needed to prepare depends on how much of these materials you want to use. The introduction takes no more time than preparing for any other lecture. The interactive questions take about 2 hours to fully explore. The instructor should go through the questions to make sure he or she is familiar with the databases and their interfaces. MAGIC Tool takes about 3-6 hours to learn well enough to teach it. With continued use, you will discover many of the nuances that make it rich enough for teaching and research applications.

MAGIC Tool was written in Java, so it works on Macintosh OSX, Windows 98 or later, Sun, Linux and Solaris. It is free so every student may download his or her own copy. It will be necessary for each user to begin with the User's Guide since it walks you through the steps to perform DNA microarray analysis. A "tutorial" is provided as a suggested set of gradually more difficult steps for becoming proficient with MAGIC Tool. The Instructor's Guide (file available on MAGIC Tool web site) provides additional background information, including Math Minutes to help you and your students understand how the data are analyzed mathematically. Calculus is not needed to understand the math behind DNA microarray analysis.

Materials

This major workshop was a part of a new experiment on the part of ABLE. Rather than requiring all workshops to be tried and tested, they wanted to see if some "cutting edge" labs could be introduced into the annual meeting. Furthermore, this particular laboratory was a computer lab experience. No wet lab materials are needed at all. This is a bioinformatics laboratory paper that cost nothing other than computers and Internet access. MAGIC Tool was written in Java, so it works on Macintosh OSX, Windows 98 or later, Sun, Linux and Solaris. RAM may be a limiting factor for some computers. For those with 256 MB of RAM, MAGIC Tool will work fine with the smaller raw data files. If you have at least 512 MB of RAM, you can work with any of the raw data files. The introduction and interactive database questions will work on any computer that has a web browser and Internet access. Modems may prove to be too frustrating for the database interactions.

Notes for the Instructor

Introduction

The introductory materials works best when combined with web a browser and Internet access. Full color figures and the microarray animation are available at this web site: <*www.bio.davidson.edu/people/macampbell/ABLE/ABLE.html*>. You may use this web site and all the figures but I request that you do not post them on your own web server. My administration takes note of the number of hits my web pages receive, so I benefit from your using my page but not mirroring it. Some of the workshop participants requested more introductory materials and practical examples of microarrays. Campbell and Heyer have written a textbook (*Discovering Genomics, Proteomics and Bioinformatics*) that devotes 56 pages to introducing DNA microarrays and providing practical examples where microarrays have been used. In addition, the book contains many interactive "Discover Questions" for further exploration. An Instructor's Guide is available on line and contains all the answers to the Discovery Questions. You may also use materials found on my genomics course web site <*www.bio.davidson.edu/genomics*>. The reader is urged to consult these materials for more in depth coverage.

You can download the DNA Microarray animation and store it on your hard drive for class use. All I ask is that you give me credit for creating the animation, you send people to my web site if they ask where they can find it, and you not post it on a web server of your own. I get recognition from my administration when I can show how many hits my pages get.

Here are the directions:

1) View the animation from the web.
2) Look in your browser's cache folder for the flash file. The name of the file will end in ".swf" and there will be an html file with a similar name.
3) Move the .swf and .html files to another location on your hard drive.
4) Your cache folder will be emptied when you turn off the browser so make sure you move the files you want before quitting the browser.
5) All I ask in return is that you cite me as the author and refer people to the original web site. It is beneficial for me to know how many people are accessing my animations.

Databases

Interactive databases can change and thus are double-edged swords. The good news is they are constantly updated which means you and your students have access to the same databases as researchers all over the world. The bad news is that your results may change with time which can lead to some awkward teaching moments when you don't get the same answer you got last month. Even more frustrating can be the changing interfaces each web site presents. Just when you got used to a particular look and feel of a database, some IT person is going to "upgrade" the web site and leave your teaching instructions out of date. However, these two inconveniences pail in comparison to the power of knowledge provided by public databases. So, it is worth double-checking the database the day you are teaching it, or sending an assignment home for students. Our textbook contains many interactive "Discover Questions" for further exploration. An Instructor's Guide to the Discovery Questions is available and contains all the answers to the Discovery Questions. You can access the Discover Questions at this URL <*www.aw.com/genomics*>.

MAGIC Tool

Because MAGIC Tool is written in Java, there are some limitations inherent in the code. First, Java only requests a limited amount of RAM from your computer if you are running a Unix-based machine such as Mac OSX. We have addressed this limitation by writing an applescript called "MAGIC_launch". MAGIC Tool application file ends in ".jar" because it is a Java file. You can launch MAGIC Tool by double clicking on MAGICTool.jar but you will be limited in the RAM it will utilize. If you double click on MAGIC_launch, then the extra RAM will be allocated to MAGIC Tool. We have set the ceiling at 800 MB of RAM, but if your computer does not have this much RAM, it will request as much as possible.

On a related note, Java programs do not purge old files, they hang on to them unless told to purge. We recognize this as a bug that needs to be fixed and will be soon. Until that time, you may find that opening many raw data files will cause MAGIC Tool to slow down. To solve this problem, you can quit and relaunch MAGIC Tool. The best time to quit is before loading a new pair of raw data files (tiff images).

Because Mac OSX is Unix based, we need to follow certain Unix naming conventions when working with MAGIC Tool. For example, "MAGIC_launch" does not have a space between the two words; it has an underscore. Unix files cannot have spaces, slashes "/" or dots ".". The same is true of folders containing the MAGIC Tool software. Another quirk of Unix is that you cannot use any shortcuts or Favorites to navigate to your folder containing MAGIC Tool. You must click directly from your hard drive until you reach the MAGIC Tool folder, which you may want to call MAGIC_Tool. Finally, you cannot use the multi-panel view available in Mac OSX. You can only have one window open at a time if you want to use the applescript to launch MAGIC Tool. The reason for this is the applescript instructs the computer to determine the name of the folder containing MAGIC Tool but if you have the multi-panel view option selected, the computer does not know which window is the correct one. These Unix-based issues are annoying, but the power provided by Unix on your Mac makes these annoyances trivial by comparison.

Finally, we STRONGLY recommend you and your students use the Quick Start portion of the User's Guide when first starting out with MAGIC Tool. The User's Guide provides step-by-step instructions for using MAGIC Tool. As you confidence and familiarity with MAGIC Tool increase, you can use different starting places or create your own instructions.

Student Outline

Introduction to DNA Microarrays

DNA microarrays are produced by placing small drops of liquid containing genes on a glass microscope slide, and allowing the spots to dry. Each spot of liquid contains many copies of a single gene and the identity of each spot's gene is known (Figure 1). Placing the drops of liquid in a precise grid pattern, or array, is the job of a one armed robot (to see a movie of the process, go to <*www.bio.davidson.edu/projects/GCAT/arrayer.mpg*>).

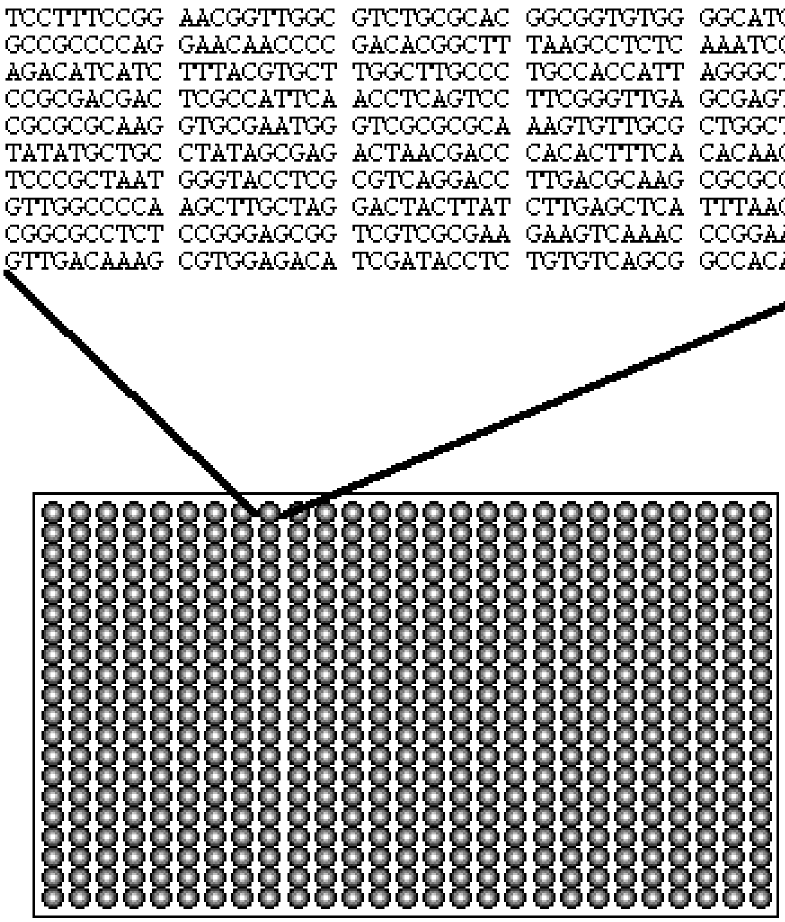

TCCTTTCCGG AACGGTTGGC GTCTGCGCAC GGCGGTGTGG GGCATGACAT
GCCGCCCCAG GAACAACCCC GACACGGCTT TAAGCCTCTC AAATCGCTGT
AGACATCATC TTTACGTGCT TGGCTTGCCC TGCCACCATT AGGGCTGTTC
CCGCGACGAC TCGCCATTCA ACCTCAGTCC TTCGGGTTGA GCGAGTGGGT
CGCGCGCAAG GTGCGAATGG GTCGCGCGCA AAGTGTTGCG CTGGCTGTAT
TATATGCTGC CTATAGCGAG ACTAACGACC CACACTTTCA CACAAGGATT
TCCCGCTAAT GGGTACCTCG CGTCAGGACC TTGACGCAAG CGCGCCTTCG
GTTGGCCCCA AGCTTGCTAG GACTACTTAT CTTGAGCTCA TTTAACATCC
CGGCGCCTCT CCGGGAGCGG TCGTCGCGAA GAAGTCAAAC CCGGAACGGC
GTTGACAAAG CGTGGAGACA TCGATACCTC TGTGTCAGCG GCCACAAATC

Figure 1. Cartoon of a DNA microarray where each circle represents a spot on the microarray and each spot contains a different DNA sequence. One such sequence is shown above the microarray.

There are many ways to used DNA microarrays, but the most common is shown in figures 2 – 4 below. In Figure 2, you see two populations of yeast grown under different conditions. The fundamental concept is that a genome must respond to the environment in which it is located and therefore regulate its genes accordingly. DNA microarrays allow us to measure the gene activity of every gene in an organism's genome. The mRNA is isolated from each population and each population of mRNA is converted into colored cDNA (typically red and green).

Once the two populations of cDNAs are produced, they are mixed and incubated with the DNA microarray and unbound cDNA is washed off (Figure 3). The spotted DNA on the microarray has been denatured so it is single stranded. The amount of colored cDNA that binds to its complementary single-stranded DNA is proportional to the activity of the gene. If a gene A produced 10 mRNAs and gene B produced 2 mRNAs, then we would expect spot A to be 5 times brighter than spot B. The DNA microarray is scanned to detect the two colors of cDNA and the green and red images are stored. Software merges the two colors and spots bound by both colors of cDNA appear yellow.

All of these steps can be seen in a Flash animation that is freely available on the web at this URL: <*www.bio.davidson.edu/courses/genomics/chip/chip.html*>. There are sound effects with this animation so turn on your speakers if you are not in the library. There is also a silent version if you prefer.

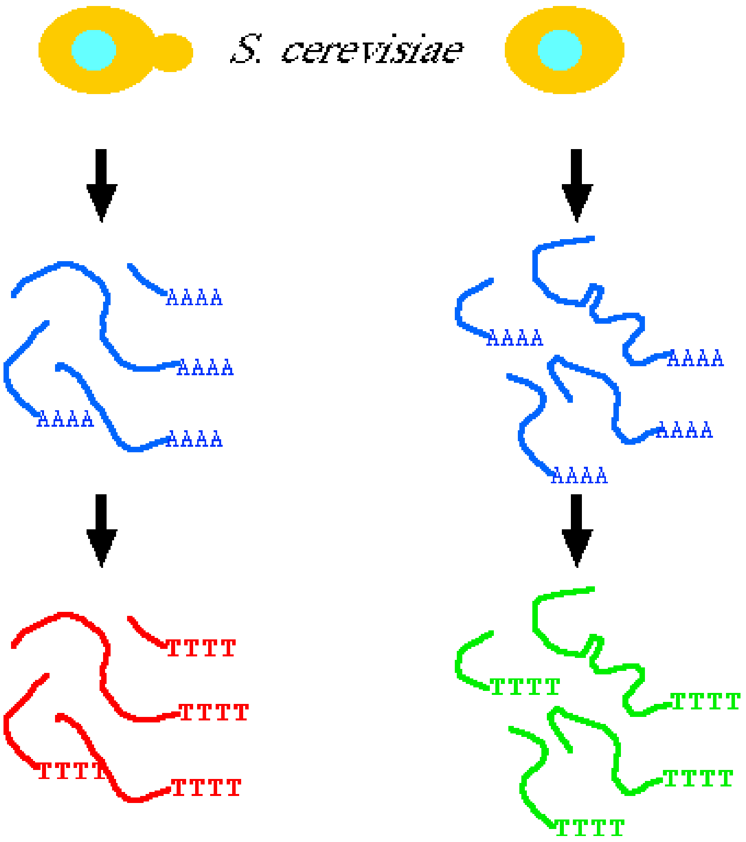


Figure 2. Method for producing labeled cDNA from two populations of cells. The colored cDNAs will be used to probe the DNA microarray to determine which genes were activated in each growth condition.

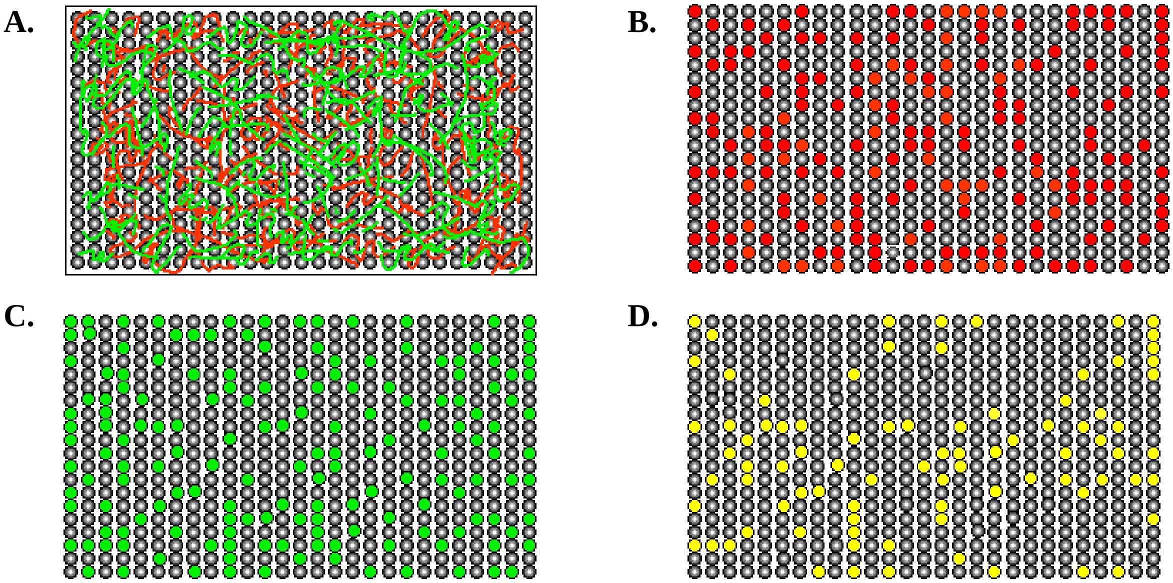
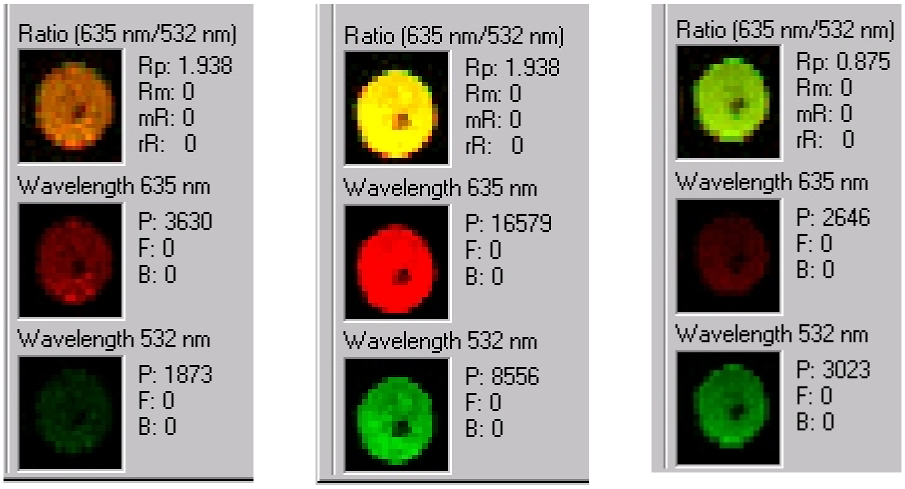


Figure 3. Colored cDNA is incubated with the microarray (panel A) and complementary sequences are allowed to bind. Unbound cDNA is washed off and then the microarray is scanned for red cDNA (panel B) and green cDNA (panel C). Spots bound by both colors of cDNA appear yellow on the computer screen used to visualize the microarray data.

In Figure 4, you can see some real data. The first gene appears orange because it was transcribed under both growth conditions but more strongly so in the red growth condition. The middle gene appears pure yellow because it was equally transcribed in both growth conditions. The third gene appears lime yellow because it too was transcribed in both growth conditions, but more strongly so in the green growth condition. This type of comparison allows us to make relative measurements of gene activity between two growth conditions. If the red condition is experimental and the green is control, then gene one was induced by the experimental conditions, gene two was unaffected but gene three was repressed by the experimental growth condition.



Ratio (635 nm/532 nm)
Rp: 1.938
Rm: 0
mR: 0
rR: 0
Wavelength 635 nm
P: 3630
F: 0
B: 0
Wavelength 532 nm
P: 1873
F: 0
B: 0

Ratio (635 nm/532 nm)
Rp: 1.938
Rm: 0
mR: 0
rR: 0
Wavelength 635 nm
P: 16579
F: 0
B: 0
Wavelength 532 nm
P: 8556
F: 0
B: 0

Ratio (635 nm/532 nm)
Rp: 0.875
Rm: 0
mR: 0
rR: 0
Wavelength 635 nm
P: 2646
F: 0
B: 0
Wavelength 532 nm
P: 3023
F: 0
B: 0

Figure 4. Real microarray data for three genes. In the text, the genes are called one, two and three from left to right. The top panel for each gene shows the merged image from the red and green channels.

Conversion of Raw Data to Numerical Data

The computer measures the amount of red and green light emitted from each spot and generates a number for each spot (Table 1). Also included in the table are the spots location (first three columns), the light intensity for each channel (red and green). The ratio is the final product from the process with which the remaining effort of analysis. However, most experiments are not conducted on static processes. Typically, we want to see how cells adopt over time so a time course experiment is performed where mRNA samples are collected every two hours from zero to ten hours (Table 2). Table 2 shows all the data but the genes are listed in alphabetical order. It is very hard to interpret these data, so investigators tend to cluster genes which means genes with similar patterns of ratios are placed next to each other in a table as shown in Table 3. However, large tables such as these (a yeast has about 6000 genes) are difficult to read and interpret. Therefore, a new color-coded figure is typically displayed (Figure 5A).

Table 1. Subset of numerical data from a DNA microarray experiment.

Grid	Column	Row	Gene Name	Red Signal	Green Signal	Ratio Red ÷ Green
1	1	1	tub1	2345	2467	0.95
1	1	2	tub2	3589	2158	1.66
1	1	3	sec1	4109	1469	2.80
1	1	4	sec2	1500	3589	0.42
1	1	5	sec3	1246	1259	0.99
1	1	6	act1	1937	2104	0.92
1	1	7	act2	2561	1562	1.64
1	1	8	fus1	2962	3012	0.98
1	1	9	idp2	3585	1209	2.97
1	1	10	idp1	2796	1005	2.78

Table 2. Hypothetical data for 10 genes (D – M) in a time course experiment.

Gene Name	0 Hours	2 Hours	4 Hours	6 Hours	8 Hours	10 Hours
Gene C	1.00	8.00	12.00	16.00	12.00	8.00
Gene D	1.00	3.00	4.00	4.00	3.00	2.00
Gene E	1.00	4.00	8.00	8.00	8.00	8.00
Gene F	1.00	1.00	1.00	0.25	0.25	0.10
Gene G	1.00	2.00	3.00	4.00	3.00	2.00
Gene H	1.00	0.50	0.33	0.25	0.33	0.50
Gene I	1.00	4.00	8.00	4.00	1.00	0.50
Gene J	1.00	2.00	1.00	2.00	1.00	2.00
Gene K	1.00	1.00	1.00	1.00	3.00	3.00
Gene L	1.00	2.00	3.00	4.00	3.00	2.00
Gene M	1.00	0.33	0.25	0.25	0.33	0.50
Gene N	1.00	0.12	0.083	0.62	0.83	0.12

Table 3. Clustered ratio data putting similar genes next to each other.

Gene Name	0 Hours	2 Hours	4 Hours	6 Hours	8 Hours	10 Hours
Gene M	1.00	0.33	0.25	0.25	0.33	0.50
Gene N	1.00	0.12	0.083	0.62	0.83	0.12
Gene H	1.00	0.50	0.33	0.25	0.33	0.50
Gene K	1.00	1.00	1.00	1.00	3.00	3.00
Gene J	1.00	2.00	1.00	2.00	1.00	2.00
Gene E	1.00	4.00	8.00	8.00	8.00	8.00
Gene C	1.00	8.00	12.00	16.00	12.00	8.00
Gene L	1.00	2.00	3.00	4.00	3.00	2.00
Gene G	1.00	2.00	3.00	4.00	3.00	2.00
Gene D	1.00	3.00	4.00	4.00	3.00	2.00
Gene I	1.00	4.00	8.00	4.00	1.00	0.50
Gene F	1.00	1.00	1.00	0.25	0.25	0.10

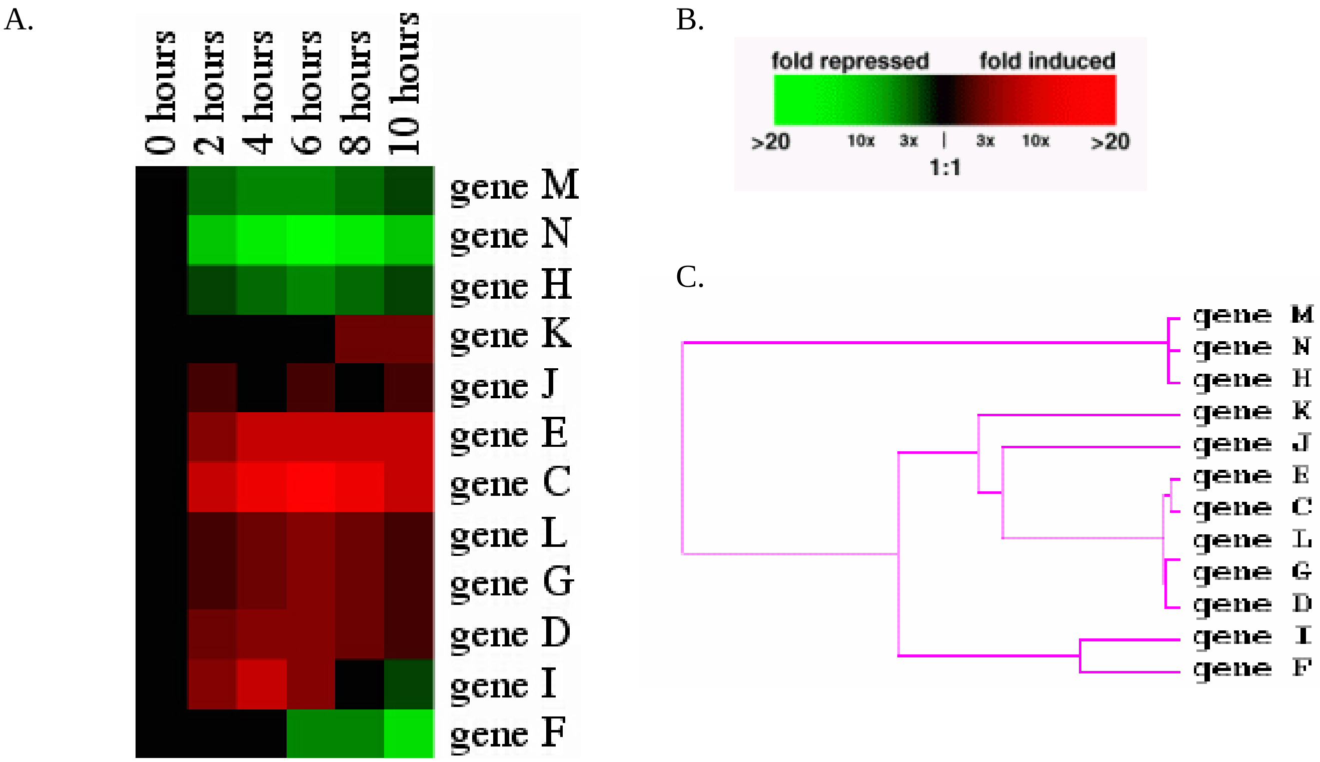


Figure 5. Color-coded version of Table 3 (panel A). Genes that were induced (more red experimental cDNA bound than green) are shaded red according to the color scale (panel B). The greater the induction (e.g., >20-fold induction), the brighter the shade of red. Conversely, repressed genes (more green control cDNA bound) are shaded green with bright green indicating the greatest repression. Genes with no induction or repression (spots that appeared yellow) are shaded black to indicate at 1:1 ratio of cDNA bound to a spot. Color-coded tables can also be displayed as dendrograms (panel C) that utilize the lateral length of lines to indicate the relatedness of gene expression patterns.

With the color-coded table, each row is a gene and each column an experimental time point. These data are much easier to read and detect genes with similar expression patterns and overall trends in the data. Tables such as this one can be in any color scheme, but green-black-red is a common one. The algorithms that produce color-coded tables can also generate "trees" called dendrograms (Figure 5C). Dendrograms illustrate the similarity of gene expression patterns with the most similar genes connected by short horizontal lines and less similar genes connected by long horizontal lines.

To illustrate the power of good data visualization, lets look at a table of 6200 yeast genes where you can see about two-thirds of the genes were repressed and one-third induced (Figure 6). In addition, you can see that genes affected early in the experiment are near the top of their portion of the colors while those affected late in the experiment are clustered near the bottom of their color sections.

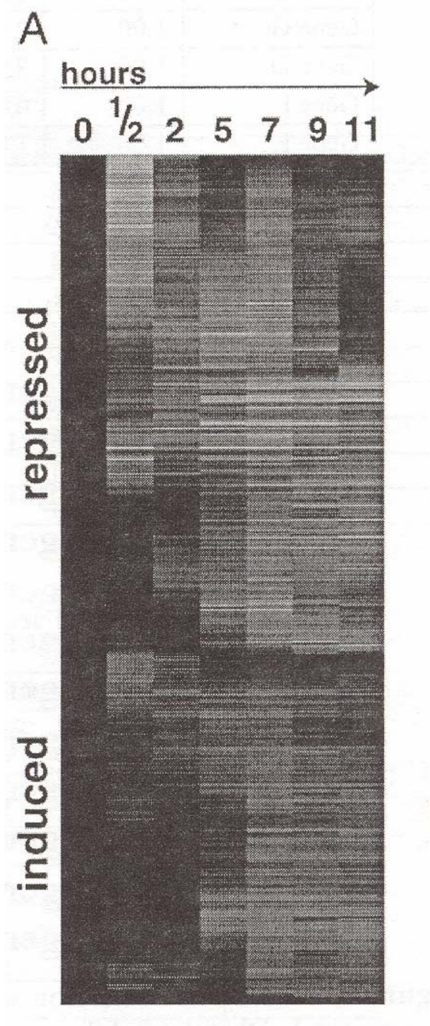


Figure 6. Whole-genome data from a yeast experiment. Each gene has been compressed vertically to be a small dash of color. Each column is an experimental time point. The colors used are the same as in Figure 5. This figure was provided by Dr. Pat Brown of Stanford University.

Conclusions

Now you have learned the essence of DNA microarrays. Investigators around the world are producing more and more data each day. Some people want to understand how medications work. Others want to determine which drug would be appropriate for a particular patient's form of cancer. Luckily for us, much of these data are being deposited in public domain databases. This means that anyone with a web browser and Internet connection can analyze their data. The next section will provide you with some challenging questions to teach you how to mine microarray databases.

If you are interested, you can see two examples of "lessons learned" at the ABLE presentation web site <*www.bio.davidson.edu/people/macampbell/ABLE/introchips.html*>.
The links are called "Yeast Lesson" and "Lymphoma Lesson." These two case studies are taken from Gasch et al. (2000) and Alzadeh et al. (2000), respectively.

Interactive Database Questions

This portion can be read only on an internet-ready computer – it works equally well on Mac and PC. You will need a web browser, and ideally a high-speed connection. Modems can work, but will frustrate you with the slow pace. To access this set of questions, go to <*www.bio.davidson.edu/people/macampbell/ABLE/ChipDatabases.html*>.

The first link is to the consortium known collectively as Gene Ontology, or GO. Gene Onotology is a much-needed effort to unify terminology associated with genomics. The term "function has become outdated and this consortium created three new terms to be more specific. **Molecular Function** is the task performed by individual gene products; examples are *carbohydrate binding* and *ATPase activity*. **Biological Process** defines broad biological goals, such as *mitosis* or *purine metabolism,* that are accomplished by ordered assemblies of molecular functions. **Cellular Component** describes subcellular structures, locations, and macromolecular complexes; examples include *nucleus, telomere,* and *origin recognition complex.* The GO database includes a large assortment of genomic organisms including yeast, fly, human, mouse, *E. coli*, *Arabidopsis*, as well as many microbes and other model systems. This GO web site provides a searchable interface so you can learn more about the three new terms and examples of each.

The second link is to GEO, which is NCBI's repository for microarray data. The early DNA microarray experiments were performed in a "wild west" environment where each investigator created his or her own rules and conventions for methodology and reporting results. Within a couple years, it became clear that universal standard were needed in order for results from different labs to be compared. A series of international meetings produced commonly accepted practices and requirements for publishing DNA microarray data. GEO is a searchable public database that allows anyone with a web browser to download data and all the experimental details associated with the data.

Expression Connection is another searchable database but it is much more user-friendly and simpler to extract meaningful information. You will see a Discovery Question (DQ) link to the right of Expression Connection. When you click on this DQ, you will get a new browser window with three frames. The left frame contains three separate sets of directions with questions that lead you to particular findings. The right frame begins as a placeholder, but as you click on the links in the left frame, the new web sites will appear in the right frame. Work your way through all three methods of querying the database and you will discover some new information and learn how to use Expression Connection at the same time.

Stanford Microarray Database (SMD) is a public repository for DNA microarray papers and their associated data sets. This too is a searchable database and the associated Discovery Question will guide you to a new insight and teach you how to use SMD. The SMD Discovery Question also uses frames with five steps provided in the left frame and the web sites appear in the right frame.

Conclusions

The purpose of this section was to illustrate some of the public domain datasets that are available for you to utilize. There are many more Discovery Questions available at the textbook companion web site <*www.awl.com/genomics*> and in the textbook. When a laboratory group publishes a paper based on DNA microarray data, the information is so abundant that no one group could possibly extract all the interesting conclusions from their data. This is why you will find recent publications based on older datasets because the new authors have uncovered new lessons by mining the public domain data. This new trend in basic science publication means that you too can mine the public databases and make new discoveries even if you cannot perform DNA microarray experiments on your own. Making new discoveries with old data is one reason so many leading scientists are emphasizing the importance of free and timely access to data and published papers.

Hands-on Learning with MAGIC Tool Tutorial

The purpose of this tutorial is to provide the new MAGIC Tool user with a few exercises to help you learn how to discover what MAGIC Tool can do. In order to make the most of this tutorial, you must have Internet access in order to download the MAGIC Tool software. The MAGIC Tool User's Guide is printed below. All MAGIC Tool resources can be found at the MAGIC Tool web site <*www.bio.davidson.edu/MAGIC/*>. The resources include:

1) MAGIC Tool User's Guide (most recent edition)
2) MAGIC Tool Instructor's Guide
2) MAGIC Tool Math Background
3) Raw Data files of varying complexity
4) Expression files of varying complexity
5) Original data from the modern classic paper by DeRisi et al., 1997.

The first section of the User's Guide called "Getting Started" has the step by step directions for a beginner. However, the User's Guide assumes the user has data files already available. This tutorial will show you how to obtain some simple data files that can be used as a way learn about MAGIC Tool.

Tutorial

1. Go to the **MAGIC Tool** web site: <*www.bio.davidson.edu/MAGIC/*>.
2. Scroll down to the Download link. Click on the text and you will find the GNU license that is required of all open source software. If you do not intend to alter the software or sell it, you can scroll to the bottom and click on "I agree". From the resulting page, you will find a few options for downloading the software. Choose the format that is appropriate for you.
3. Go back to the MAGIC Tool main page and scroll down to find the **User's Guide**. This PDF file is designed to show you how to get started and go deeper into MAGIC Tool. The instructor's

Guide provides (online) more details about the algorithms used in MAGIC Tool. The PDF link called “Additional Math Background for Understanding Data Analysis” provides Math Minutes to explain the basic math behind DNA microarray data analysis.

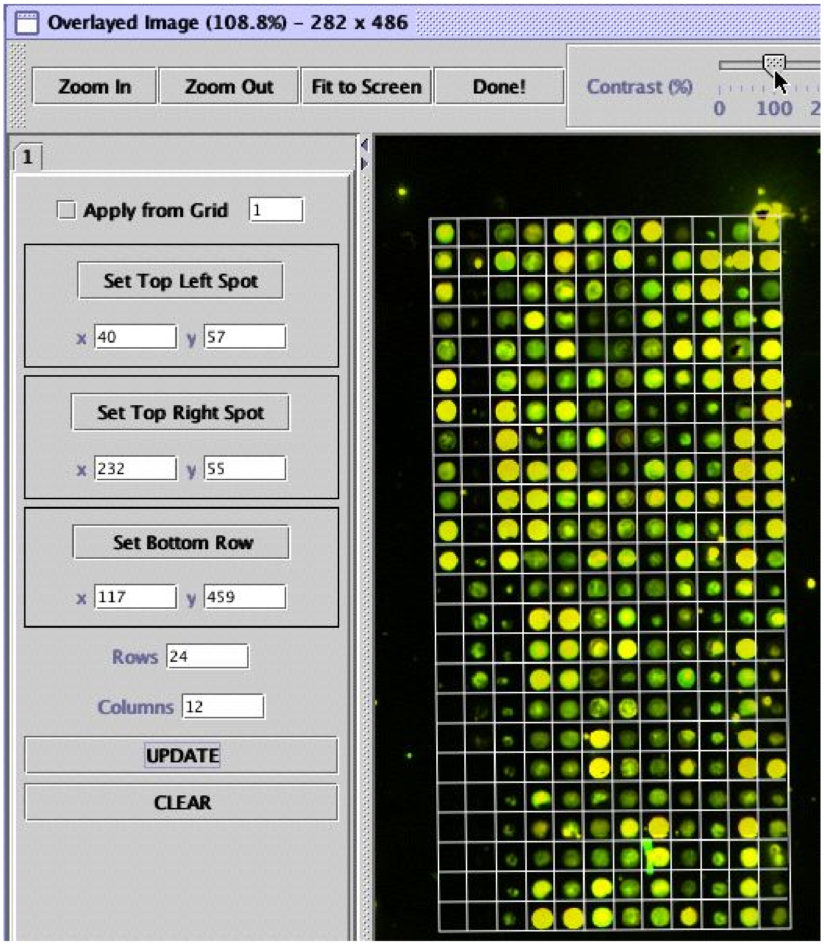
4. Now download the **tiff files** that are the raw data. There are four levels of options. The simplest is the one grid set which is only one small section from a much larger microarray. The next level up is a four-grid set of spots. The next level up has two similar types of data: Whole yeast genome (Y01 Version from Stanford) and Whole yeast Genome (ISB Version). These are much larger sets of spots from two different DNA microarrays. These are single experiments from a larger set of experiments. The final and most complex set of tiff files are the ones used by Joe DeRisi in the landmark paper examining the diauxic shift of yeast metabolism (*Science*. 1997. 278: 680-686).
5. The last two sets of files you will need are the **gene lists** (some times referred to as GodLists) and the **gene info files**. At this time, GCAT has compiled only gene info files for yeast, but other model organism files are under development. Gene lists identify each spot on the microarray so you will know which genes you are looking at on the chip. Gene info files contain the chromosomal location and functional information for each annotated gene in the genome. This information can be imported to enhance your interpretation and searching of the data.
6. Read the User’s Guide to see the best way to launch MAGIC Tool. Note that software written in Java automatically requests a limited amount of RAM from your computer. This is not a problem to launch the program but since DNA microarrays generate large data sets, you should boost your RAM if possible. Therefore, we have created scripts to boost the allocation of RAM to MAGIC Tool.
7. Although it is not critical for the practice data since you are not interested in the biology behind the data, it may be a good exercise to download the excel files and the jpeg file to understand how the chips were printed. Understanding the chip’s layout is critical to analyzing real data (either your own or the DeRisi set of data). Knowledge of the chip’s layout is important for addressing the chip (telling MAGIC Tool how the chip was printed) so the software will know how the spots relate to the gene lists.
8. A good starting place is to work with the one grid set of data. Keep in mind that MAGIC Tool running on Mac OSX or Linux uses Unix and so you will want to eliminate any folders or file names that contain spaces. Check your file names and folders at this time to ensure no surprises later.
9. Follow the Getting Started directions using the one grid set of tiff files. In this project, you will grid, segment and explore the data. You cannot cluster single time points.
10. Perform similar steps for the four-grid set of data. The only difference is the ability to apply the grid patter from one group of spots to additional groups of spots.
11. The next level of complexity is a whole-genome chip. You can choose either the Y01 set where genes were printed once or the ISB chip where each gene was spotted in duplicate (side by side).

The duplicate printing allows you to examine consistency and to utilize the MAGIC Tool averaging capacity.

12. Finally, you can use the DeRisi tiff files to create a data set that has multiple time points. This is a way to generate an expression file that can be clustered by MAGIC Tool in several ways.
13. We also have pre-fab/hypothetical data sets for you if you don’t want to endure the tedious steps necessary to generate your own time course data. You can use the 1-grid, four-grid, or DeRisi expression files for use with clustering. Clustering is explained at increasingly detailed information in the User’s Guide, Instructor’s Guide and the “Additional Math Background for Understanding Data Analysis” file (both are available on the MAGIC Tool web site).
14. Instructors can create expression files using a spreadsheet program and then saving the file as a tab delimited text file (one of the “save as” options). These tab-delimited files can be imported and used for further practice by students.
15. MAGIC Tool is public domain and open source which means you and your students can download the software as many times as you want and load on as many computers as you want. We hope that MAGIC Tool will continue to improve over time, so you may want check the download page periodically to see if any new versions have been posted.

MAGIC Tool User's Guide

MicroArray Genome Imaging and Clustering Tool



Overlayed Image (108.8%) - 282 x 486

☐ Apply from Grid 1

Set Top Left Spot: x 40 y 57

Set Top Right Spot: x 232 y 55

Set Bottom Row: x 117 y 459

Rows 24

Columns 12

MAGIC Tool Version number: 1.0; June 16, 2003

MAGIC Tool is distributed freely by Davidson College for non-commercial, academic use.

The Goal for MAGIC Tool

The purpose of MAGIC Tool is to allow the user to begin with DNA microarray tiff files and end with biologically meaningful information. You can start with tiff files or expression files (spreadsheet of ratios). MAGIC Tool was created with the novice in mind but it is not a dumbed down program. In fact, MAGIC is also designed to illuminate all the black boxes inherent in software programs. MAGIC allows the user to change parameters for clustering, data quantification etc. The Instructor's Guide (available online at MAGIC Tool web site) explains the math behind all these different options. This User's Guide will teach you how to use the software but leave the theoretical explanations to the Instructor's Guide.

Comparative hybridization data (glass chips) and Affymetrix data are compatible with MAGIC Tool. You are also encouraged to visit related sites:

GCAT: <www.bio.davidson.edu/GCAT>
Tutorial for Clustering: <www.bio.davidson.edu/courses/compbio/jas/home.htm>
MAGIC web site: <www.bio.davidson.edu/MAGIC>
Genomics Course: <www.bio.davidson.edu/genomics>

System Requirements

- Windows 2000 or later
- Mac OSX 10.2 or later
- Linux 7.x or later
- 256 MB RAM minimum; 500 MB to 1 GB of RAM recommended.
- Several hundred MB of hard drive space available, depending on the files you work with and what type of analyses you perform

Vocabulary

Addressing is the short process of telling MAGIC Tool the layout of the spots and grids in the tiff file as viewed within MAGIC.

Chip is a synonym for a microarray.

Feature is a synonym for a single spot on a microarray.

Flag is a verb that means you mark a particular spot to indicate its data are not reliable. This may be due to high background in the area, a dust bunny sitting on the spot, etc.

Grid is a compact arrangement of spots with even spacing.

Gridding is the process that MAGIC uses to find the spots on your tiff files

Metagrid is a higher order level of organization. A set of grids are organized into groups called metagrids. For a more complete description, see this web page <www.bio.davidson.edu/projects/GCAT/Griding.html>.

Segmentation is the process of finding the signal and distinguishing it from the background. There are three methods in MAGIC: Seeded Region Growing, Adaptive Circle and Fixed Circle.

Tiff files (e.g. file_name.tif) are the raw data that are produced when a DNA microarray is scanned. One tiff file is produced for each color on each chip scanned.

WARNING! Java programs, including MAGIC Tool, do not like files or folders with spaces in the names. Therefore, when you put MAGIC Tool on your computer, make sure its folder, and all upper level folders, have underscores " _ " instead of spaces.

Getting Started

Overview of Steps

If you start with two tiff files, you will need to perform the following steps in order to produce clusters or explore your data.

1) Start MAGIC Tool
2) Start a project
3) Load tiff files
4) Load gene list
5) Locate spots
6) Distinguish signal from background
7) Generate expression file
8) Manipulate Data – transform (maybe normalize)
9) Calculate correlation coefficients
10) Cluster genes
11) Graphically display data – many options
12) Explore data (can be done immediately after transforming)

Start MAGIC Tool

1) Double click on the MAGIC_launch icon. On some computer systems, you can start MAGIC Tool by clicking on the MAGICTool.jar file, but the JAVA software may restrict the amount of RAM dedicated to MAGIC Tool, which is not good for big projects. Therefore, we have created scripts for Windows and Mac OSX called MAGIC_Launch that ensure MAGIC Tool will access more of your available RAM.
In order for MAGIC Tool to work properly, you have two follow three rules.

A) MAGICTool.jar and MAGIC_launch must be located in the same folder.

B) The name of the folder containing MAGIC Tool and MAGIC_launch cannot contain any spaces. Likewise, all parent folders cannot have spaces in their names. You can use underscore “ _ ” but not spaces or dashes “ - ” or slashes “ / ” or “ \ ”.

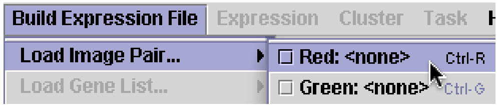
C) On OSX, you must navigate through the hard drive directly. You cannot use any aliases or Favorites. Furthermore, you must have the folder containing MAGIC Tool open with no parent folders open. You cannot use the multi-column view.

Start a Project

2) Under the Project menu, create a new Project. You can save this project in a convenient location on your hard drive. Project files end with the suffix “.gprj”.

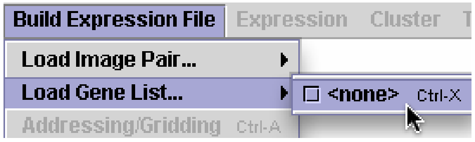
Load Tiff Files (Control R and Control G)

3) Under the Build Expression File menu, load the red and green tiff image pairs. Remember that red is a longer wavelength than green, so if your files are identified by the wavelengths, you should still be able to determine which color is which.



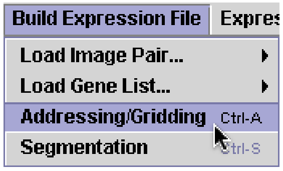
Load Gene List (Control X)

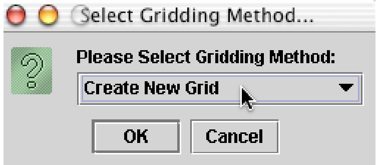
4) Load the gene list, also under the Build Expression File menu. This should be a text file with suffix of ".txt". However, many other file formats will also work. To see non-txt files, you may need to alter the dropdown menu for file types to read "All Files".



Locate Spots (Control A)

5) Under the Build Expression File, select Addressing/Gridding option. Note that you can type Control and A to accomplish this step.

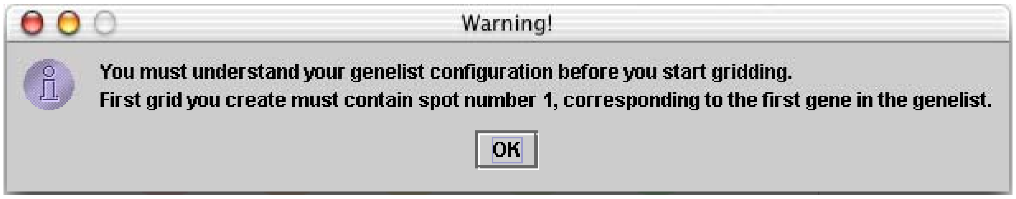




Please Select Gridding Method: Create New Grid

You will be asked if you want to create a new grid or load a saved grid. Unless you have done this before, you will need to create a new grid.

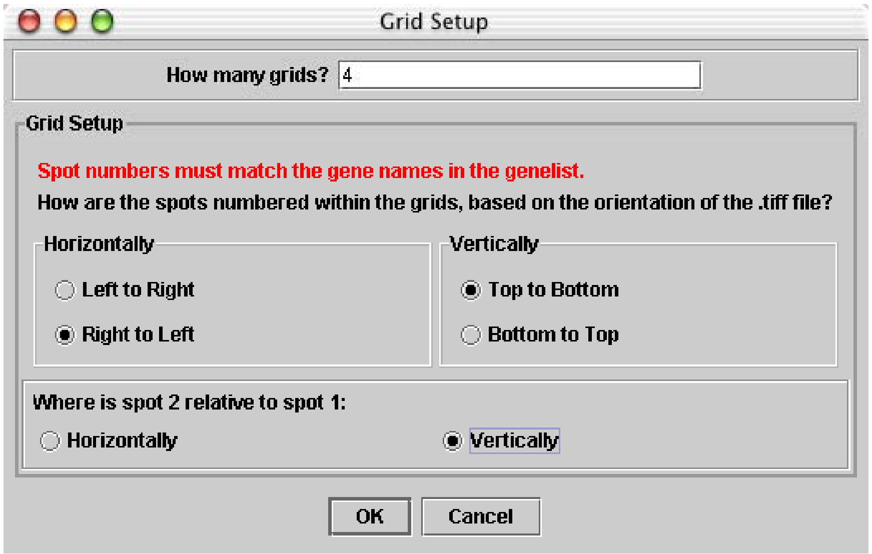
When you begin, you will get a warning window that is normal and intentional.



You must understand your genelist configuration before you start gridding.
First grid you create must contain spot number 1, corresponding to the first gene in the genelist.

The warning is a reminder that you MUST understand how your spots are arranged on your microarray. For more information about this process, consult the Instructor's Guide under "Gridding". Do not proceed any further if you do not understand the organization of your microarray. Failure to do this will result in features being incorrectly identified.

6) Two windows will appear. One will show your merged tiff files and the other will permit you to address the tiff file. The smaller (moveable) window will ask you information about how your microarray is organized; this is called addressing. First, enter the total number of grids on the tiff file.



How many grids? 4

Grid Setup

Spot numbers must match the gene names in the genelist.

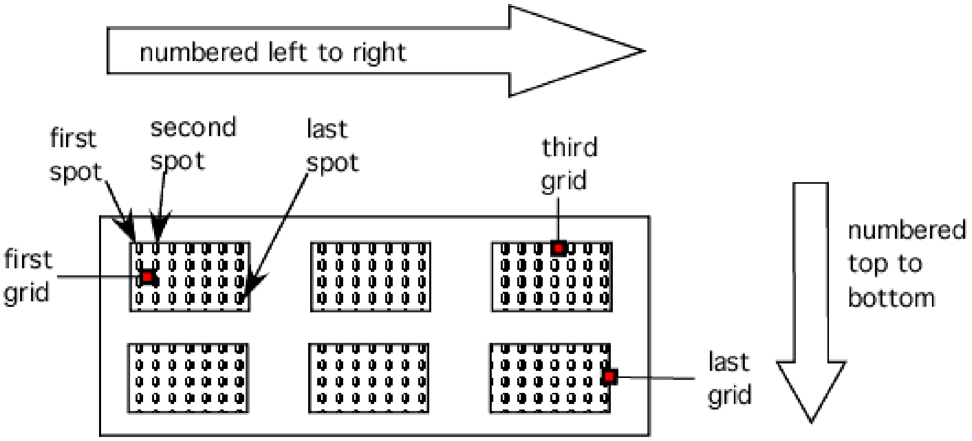
How are the spots numbered within the grids, based on the orientation of the .tiff file?

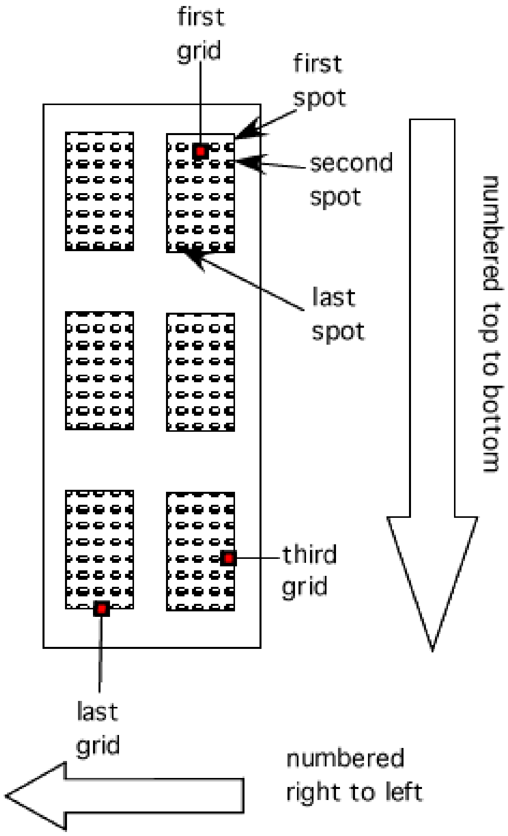
Horizontally
○ Left to Right
● Right to Left

Vertically
● Top to Bottom
○ Bottom to Top

Where is spot 2 relative to spot 1:
○ Horizontally ● Vertically

This is the easiest step to make a disastrous mistake. Answer the four questions based on **the way you are seeing your microarray at this time**. Here are examples to illustrate the point. If the image has been rotated 90 degrees clockwise compared to the way you normally think about your chip but your gene list is not altered. The way you are seeing your tiff file will not match what you think of as your microarray organization. Notice in the image above, the spots are described as being numbered from top to bottom and from right to left. Also, spot number 2 is below (vertical) spot number 1. This addressing is due to the chip's image being rotated 90 degrees clockwise, as shown below. If you study the before and after rotation images, you will see how the spots have moved and why the new orientation resulted in the addressing provided in the figure above. You can change your answers to these addressing problems by selecting "Grid properties…" under the file menu of the gridding window.





Here, the chip image has been rotated 90° clockwise but the numbering of the spots has not changed.

It is important that you keep track of the spots if the chip is rotated. Use the pattern of missing spots and the comments in your gene list to help you become reoriented if necessary. The layout and number of grids is an easy way to orient yourself as well.

If you find you have made a mistake in your answers to the four addressing questions, you can correct yourself at any time by re-setting the grid properties.

7) At this time, you should see a merged image of your tiff files. Where red and green are superimposed, you should see a shade of yellow. The number one tab should be selected as the default. Tab number 1 indicates you should work with grid number one (based on the gene list order) first. The grid numbers on the microarray should correspond to the tab numbers. Again, if you do not follow this procedure of matching grid numbers with tab

numbers, you will cause the features to be incorrectly identified.

8) Navigate the image until you can see the first grid as defined by the gene list. Remember that spots and genes do not change their numbers with image rotation. In the example above where the image is rotated 90 degrees clockwise, the first grid would be the grid in the top right corner.

9) If you want, you can adjust the contrast to help illuminate faint spots. To do this, slide the indicator that is currently pointing to 100% contrast near the top of this window. Adjusting contrast does NOT affect the raw data, it only allows you to see spots better for this step.

10) The goal of gridding is to tell MAGIC where the spots within each grid are located. This feature is one of the best innovations in MAGIC Tool. You may find it helpful to zoom in on the first grid of spots. To zoom in, click on the "Zoom In" button and then click on the grid where you want the zoom to center.

11) Click on the button that says "Set Top Left Spot" and then click on the center of the top left spot.

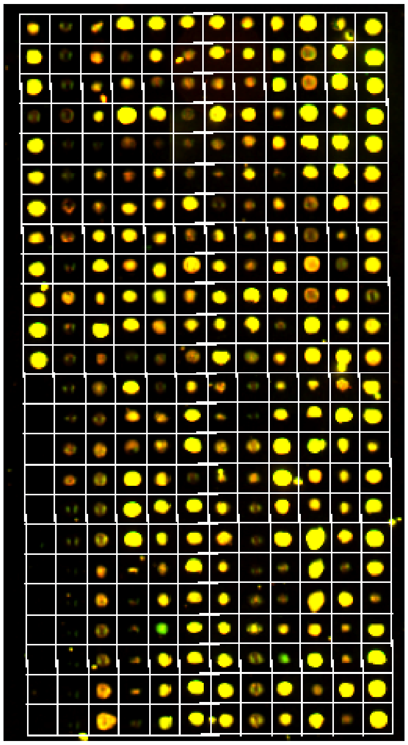
12) Click on the button that says "Set Top Right Spot" and then click on the center of the top right spot.

13) Click on the button that says "Set Bottom Row" and then click on the center of any spot in the bottom row. Choose a big round spot to make this step easier.

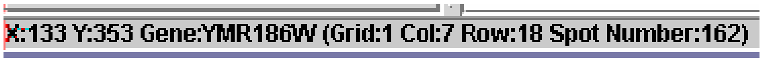
14) Enter the information for the number of rows and columns.

This is to be answered based on the way you are currently viewing the tiff file. In this example, there are 24 rows and 12 columns.

15) Click the "Update" button. At this time, you should see all the spots in the first grid surrounded by boxes as shown: (You may need to zoom out to see the full grid.)



At any time in the gridding process, you can mouse over a spot and identify its location as well as its identity. This information is displayed in the bottom left corner and is especially useful for navigating during segmentation.



X:133 Y:353 Gene:YMR186W (Grid:1 Col:7 Row:18 Spot Number:162)

16) At this time, see if the spots look centered in the boxes. If not, then adjust the position of the boxes either by clicking on the appropriate button and then the correct spot, or by manually typing in numbers to adjust the boxes. For the one grid file, the optimum numbers are:
Top Left: x = 41 and y = 57
Top Right; x = 235 and y = 55
Bottom: x = many values and y = 460

With practice, entering the numbers manually is much easier. Note that the position of the mouse is displayed in the bottom left corner of the window so you can determine if the numbers should be bigger or smaller to shift the boxes in the correct direction. This step takes a bit of practice, but it is MUCH easier than most other methods for gridding.

17) If you only have one grid, skip to step 20. If you have more than one grid, continue. Once the first grid is properly gridded (surrounded with boxes with the spots in the centers), it is time to repeat this process for grid that is number two in your gene list. Click on tab 2 at the top of the window. You may need to zoom in again to see clearly the spots in grid number 2. Click on the button that says "Set Top Left Spot" and then click on the center of the top left spot.

18) At this time, you should see a box next to the phrase (apply from grid 1). Click this box and then click on the “Update” button. This should apply a set of boxes around the spots that nearly centers all spots in the boxes. Again, you can adjust to size and location of the boxes as needed by typing in X and Y values to resize and move the grid of boxes.

19) Continue the gridding process until all the grids on the microarray are boxed with the spots in the center of the boxes. At any time, you can change your answers to the four addressing problems by selecting “Grid properties…” under the file menu of the gridding window.

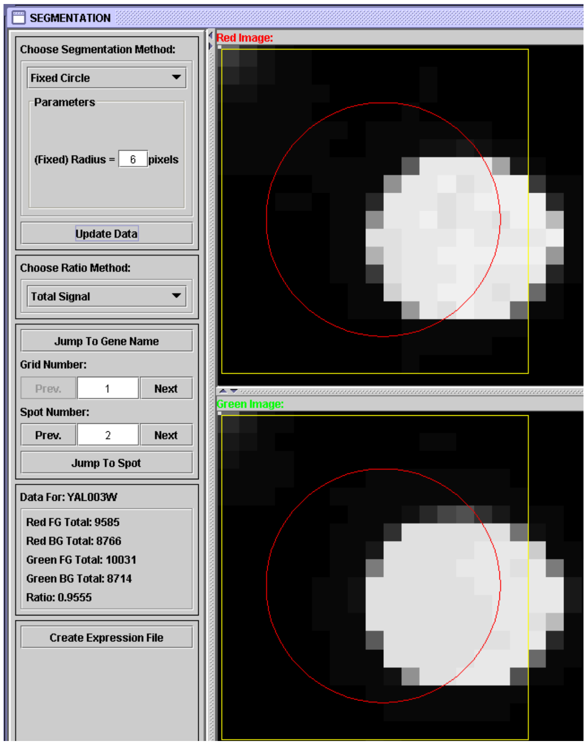
20) When you have finished gridding all your grids, click on the “Done!” button. A grid file should be saved in your project folder and have a suffix of “.grid”. If the default file name ends in .txt (a temporary bug), then delete the .txt and then save. You may also save a partially gridded project using the “Save Current Grid As…” under the file menu of the gridding window.

If your gene file and the number of spots do not match, you will get an error message. Your gene list should match the number of spots. If not, you probably will make an error identifying the spots later so you are urged to fix this problem now. If your gene list and the number of spots you gridded match, then you will be informed of the total number of spots and allowed to save the grid file for further use.

21) You can also save an image of the combined tiff files, with or without grids. You can save as tiff, jpg or gif. Tiff format works on all drawing and word processing programs so it is a universal format. Jpeg is good for images such as this that have many shades, like a photograph. Gif is the simplest format but may lose some of the subtlety of your original file. This saved merged image is useful if you want to take a picture of the overall grid and can be used for publishing or teaching.

Distinguish signal from background - segmentation

22) From this point on, there are choices the user must make that will influence the outcome of the data interpretation. The first choice you have to make is which method for distinguishing signal from background.



Fixed Circle

(Fixed) Radius = 6 pixels

Total Signal

Grid Number: 1

Spot Number: 2

Data For: YAL003W

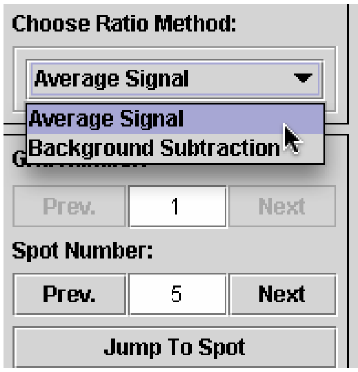
Red FG Total: 9585
Red BG Total: 8766
Green FG Total: 10031
Green BG Total: 8714
Ratio: 0.9555

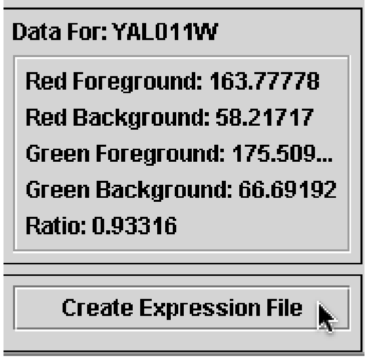
a) The most common way is to simply place a circle in the middle of the squares you drew for gridding. This is called *fixed circle*, though you can adjust the radius of this circle as shown below. Note that even if the circle is bigger than the box, only signal inside the box is used for measuring signal.

b) The second method is the *adaptive circle*. The size and location of the circle changes depending of the size on the feature. However, the shape is always a circle.

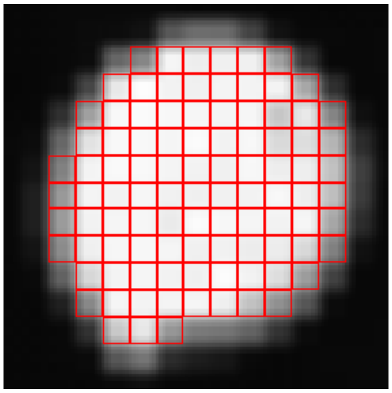
c) This method for segmentation is designed to find the signal for each spot based on the distribution of the signal. Seeded region growing looks for the brightest pixel and then connects all pixels adjacent to this pixel into one shape.

You can visually inspect the features to verify the gridding and segmentation were performed adequately. This inspection gives you a chance to note any features you think should not be considered during subsequent data analysis. You can choose to subtract background or not (Ratio Method) and you can navigate around the spots, with a summary of each spot's data below.



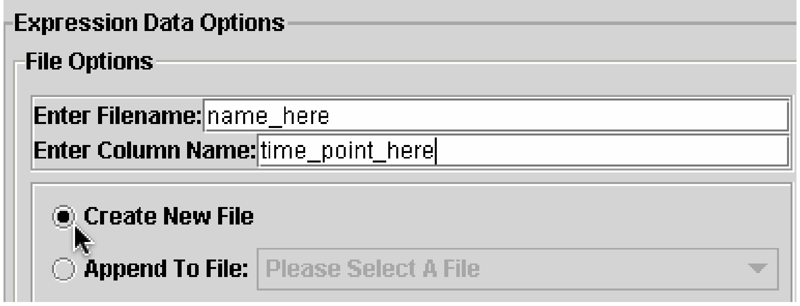

Data For: YAL011W

Red Foreground: 163.77778
Red Background: 58.21717
Green Foreground: 175.509...
Green Background: 66.69192
Ratio: 0.93316



Generate expression file

23) Click on "Create Expression File" when you are satisfied with the segmentation process. This will generate an expression file, which was the goal of the previous 22 steps. An expression file contains the numerical values for each color for each spot as well as the ratios for each spot (red ÷ green). The ratios will be used for all subsequent data analysis. You do not need the tiff files any more.


Expression Data Options

File Options

Enter Filename: name_here
Enter Column Name: time_point_here

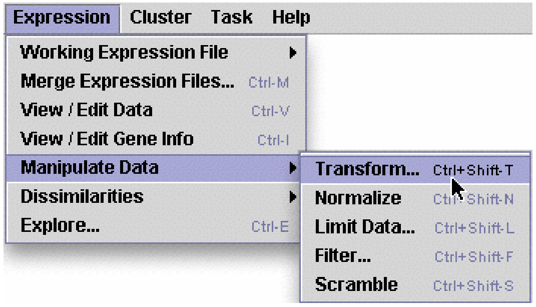
● Create New File
○ Append To File: Please Select A File

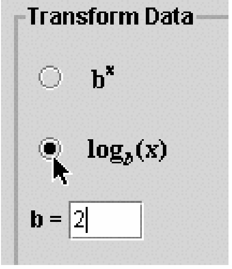
You will need to name the expression file and the column (e.g. time point, treatment, etc.). You can append this to an existing file or create a new one.

Fixed circle is the fastest method and seeded region growing is the slowest. Keep this in mind when you are learning MAGIC Tool.

Manipulate Data

Although this step sounds like a point and click way to conduct scientific fraud, it is actually a beneficial step to consider (see Instructor's Guide online). You can: transform your data; normalize the ratios; temporarily restrict your data analysis to a subset of experimental conditions (e.g. certain time points, or dye reversals); filter out some features that don't meet certain criteria; or generate a random set of data to use as a comparison.



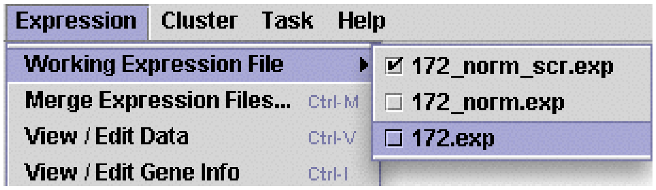

Transform Data

○ b^x

● $\log_b(x)$

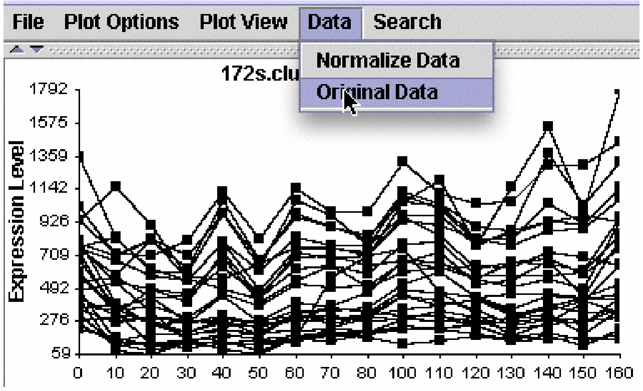
b = 2

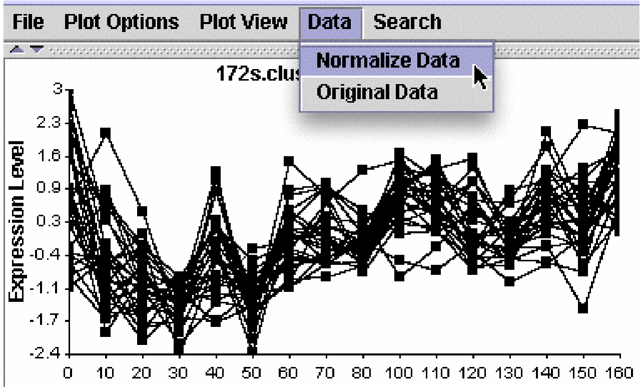
If you manipulate your data, you will generate a collection of new expression files with names that match the manipulation. Be sure to verify which expression file you are working with in subsequent steps. It is easy to get confused.



At this time, you should transform your data. Typically, this is done using a $\log_2$ transformation to indicate the number of two fold changes in gene expression (thus 4 fold changes resulted in numerical values of 2). This will convert your ratios into values that are on the same numerical scale so that a gene that is 4 fold induced (+2) has the same numerical value as a gene that is 4 fold repressed (-2 instead of 0.25).

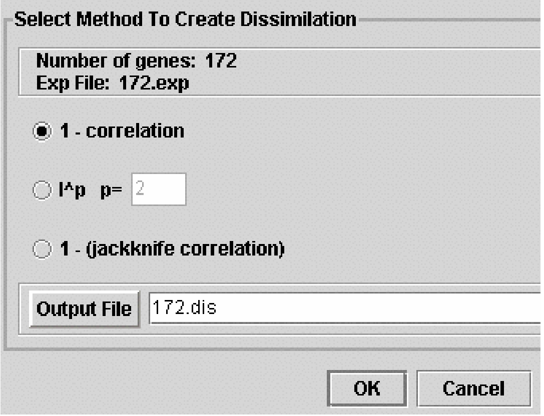
You may also want to normalize your ratios, but this is not necessary. Normalization does not affect the correlation between two genes, but since normalized gene expression patterns all have a variance of 1, the graphs of normalized genes are on the same scale. When you plot the various groups or clusters of genes, you can view the data as normalized or original ratio values.





Calculate correlation coefficients

24) From this point on, you are comparing different genes to one another. The first step in this process is to generate correlation coefficients (see Instructor's Guide online for a detailed explanation). Under the Expression menu, choose "Dissimilarities" and then "compute". When you do this, a window will appear where you have to choose from three choices. This is another decision that will affect the data analysis.



Select Method To Create Dissimilation

Number of genes: 172
Exp File: 172.exp

● 1 - correlation

○ l^p p= 2

○ 1 - (jackknife correlation)

Output File 172.dis

25) The most common method is the default "1 – correlation." The other two methods are described in the Instructor's Guide. When this step is complete, MAGIC generates a dissimilarity file which you can name in the output file box, but be sure to retain the suffix ".dis". Click on OK to begin this process. The progress is monitored in a scale bar below the OK button (not shown here). You can calculate dissimilarities on any expression file (.exp) but you should use your transformed ratios rather than non-transformed ratios. You can also use transformed and normalized expression files containing ratios.

Determine Biological Meanings

At this point, you can generate a series of clusters using four different methods. Clustering is a very popular process for DNA microarrays, so we will describe this first, but remember that exploration is equally valid (see below). Exploring your data can be performed any time after segmentation. All you need to explore are expression files (*.exp).

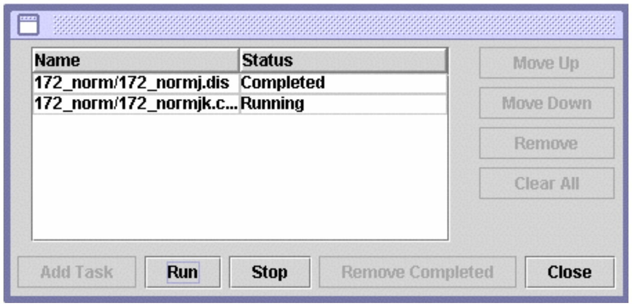
Cluster genes

With MAGIC Tool, there are four ways to cluster genes. You can cluster from any dissimilarity file. First you have to calculate the clusters and then you can display them in a variety of ways. The most common way to cluster is called hierarchical clustering, which you can do with MAGIC. However, we prefer Q-T clustering (see Instructor's Guide online for details). You can also cluster by k-means or supervised clustering.

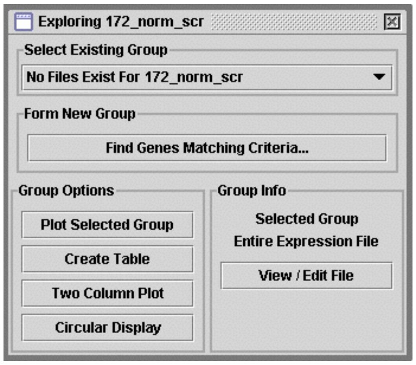
Once you have clustered the genes, you can display the results in several ways. MAGIC allows you to view these clusters in a variety of dynamic displays. Each display can be saved as a image file for publishing or teaching. Display options are addressed in more detail later in this manual.

Automating Tasks

As your datasets get bigger, the time it will take to make all the necessary calculations will increase rapidly. Therefore, MAGIC allows you to establish a list of tasks to be performed in sequence. You can tell MAGIC to begin a series of steps and then walk away from your computer. MAGIC will perform this sequence of tasks while you do other things. For example, you can establish a list of tasks to perform and go home for the night. When you return the next morning, MAGIC will have completed the series of tasks.



Name	Status
172_norm/172_normj.dis	Completed
172_norm/172_normjk.c...	Running



Exploring 172_norm_scr

Select Existing Group: No Files Exist For 172_norm_scr

Explore data

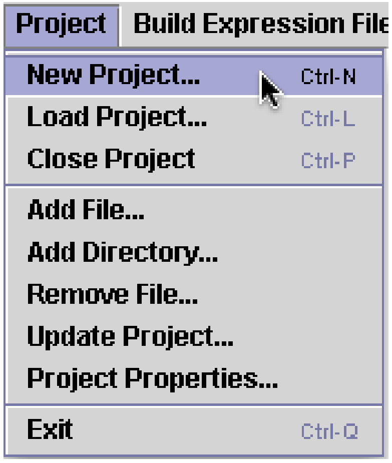
Data exploration is a way to find relationships that were not apparent after simple clustering. For example, you can find all genes that were up-regulated after a certain time point, or all genes that increased their fold repression four times or greater at any time point. Once you have identified such genes, you can display them in a number of dynamic ways and also save these images for publishing or teaching.

Closing Comments

This section was intended as a way to get you launched into the MAGIC Tool way of working with DNA microarrays. MAGIC allows you to compare the consequences of different choices for quantifying, comparing and clustering the same raw dataset. This capacity to compare methods is a powerful way to understand better the assumptions and implications inherent in data analysis as published each week. MAGIC allows you to explore data and data analysis during the early days of DNA microarrays when the research community has not settled upon standards for comparing results. MAGIC was designed to empower the user and make DNA microarrays more approachable for a wider audience. In the following section, every option available in MAGIC Tool will be spelled out so you can utilize the full potential of MAGIC Tool.

Complete List of MAGIC Tool Options

Project Menu



New Project (Control N)
This begins a new project. All work done within MAGIC Tool must be associated with a project. The name you give to the project should be unique from other projects. The file name will automatically terminate with the suffix ".gprj" and a folder will be created. All subsequent steps and files will be stored automatically in this project folder.

Load Project (Control L)
This allows you to reopen a previous project.

Close Project (Control P)
Allows you to stop project without quitting MAGIC Tool completely.

Add File….
This allows you to add files (e.g. expression files) from other projects to your current project. You will be directed to a window from which you can click your way through the hard drive in search of the files you want to add. You can hold down the control key and click on multiple files to select them.

Add Directory…..
This allows you to add entire folders to your current project.

Remove File….
This lets you remove unwanted files from your current project folder. Or, if you want to delete a dissimilarity or cluster file so you can create a new one, this can be accomplished by writing over the older version (You will be prompted to verify you want to write over the existing file with the same name.) You can hold down the control key and click on multiple files to select them.

Update Project….
Allows you to drag files into existing folders and then update the currently active project. This allows the user to quickly move tiff, grid, expression, dissimilarity, and cluster files around and then utilize them in different projects.

Project Properties
This allows you to *remove* or *ignore* any genes in your current project that do not have complete data. When a DNA microarray is printed, some features will be missing and therefore you cannot collect data for this gene.

If you choose to *ignore*, you will be prompted to determine what cutoff of possible data (in percent) must be available for a gene to be included in your data analysis. This allows you to work with genes that are occasionally missing data from a series of DNA microarrays. Ignoring will retain the gene in subsequent analysis, but ignore those columns that lack the data (as long as the number of columns with data is at least the percent designated). If the gene contains less than the designated percent of data columns, then no dissimilarities are computed for ignored genes.

If you choose to remove all genes missing data, then genes missing any data from one or more columns will not be used for calculating dissimilarities. This is a more drastic cutoff than ignoring genes with high percentages of available data but missing ratios.

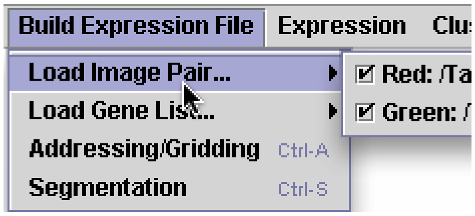
Exit (Control Q)

This quits MAGIC Tool. All completed steps will be saved in your project folder. Steps only partially completed will be lost.

Build Expression File

Load Image Pair…. (Control R and Control G)

This allows you to browse your hard drive to find the tiff files for the two colors. You can load the two tiff files in either order. Just be sure to match the colors and the files. Remember that red is a longer wavelength than green.



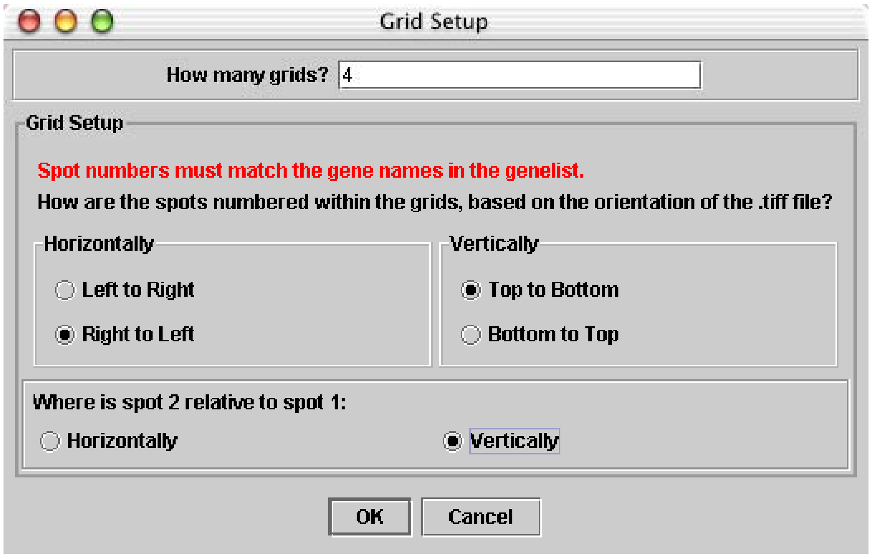
Load Gene List… (Control X)

You can browse to find your gene list that associates each feature on the microarray with a gene name. Often, gene lists have additional information such as which features did not print, alternative names for the gene, etc. You can open your gene list to see what information it contains. If it contains information about the plates and wells for each gene, this is not useful information for MAGIC but was used to help the people who printed the chips to keep track of what they were doing during the manufacturing of the chips.

Because every microarray manufacturer has a different layout for their gene files, MAGIC Tool requires you to create a new gene list that contains the ORF names in the first column. If you open the gene list that came with your microarrays using a spreadsheet program, you can quickly create the gene list MAGIC needs. To do this, find the column that contains ORF names such as YBL023c or YAR002W, etc. Copy this ORF column and paste it in the first column (you may have to create a new column to hold this information). Remove column labels, so that each row in your file corresponds to a gene. Save this as a new file that ends with the suffix ".txt" and use it for the gene list. Although it takes a bit of manual labor to create this MAGIC gene list, it allows the user to quickly adapt to different microarray production styles. Later, you will learn how to import additional information about genes from commonly studied organisms.

Addressing/Gridding (Control A)

Addressing is telling MAGIC Tool how the spots are numbered. This step is the easiest one to make a mistake on, so be very careful when answering the four questions as they appear in the window. It is vital you understand how your spots are organized on the microarray and in the gene list. All questions should be answered according to the way you see the merged image of your microarray in the viewing window. Are the genes printed in duplicate? If so, are the duplicate spots horizontal or vertical? You will need to know how many grids there are as well as the order of the spots in your gene list compared to the image in MAGIC Tool. It cannot be overemphasized how critical this step is. If you get this part wrong, you will not know the correct identity of any of the spots.


Grid Setup

How many grids? 4

Grid Setup

Spot numbers must match the gene names in the genelist.

How are the spots numbered within the grids, based on the orientation of the .tiff file?

Horizontally: ○ Left to Right ● Right to Left

Vertically: ● Top to Bottom ○ Bottom to Top

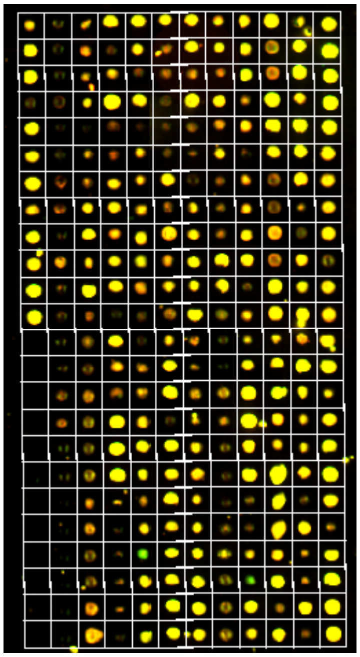
Where is spot 2 relative to spot 1: ○ Horizontally ● Vertically

Gridding is much easier. The purpose of gridding is to draw little boxes around each feature so the spots are in the center of the boxes. To perform gridding, you must know where the first grid is on the image you are viewing. Perform gridding on the number one grid of spots first. **Each subsequent grid must be gridded in the same order as they are in the gene list.**

At this time, you should see a merged image of your tiff files. You may find it helpful to zoom in on the first grid of spots. To zoom in, click on the "Zoom In" button and then click where you want the zoom to center. Where red and green are superimposed, you should see a shade of yellow. The number one tab should be selected as the default.

Navigate the image until you can see the first grid as the one you know to be the first grid in the original layout of your microarray. If you want, you can adjust the contrast to help illuminate faint spots. To do this, slide the indicator that is currently pointing to 100% contrast near the top of this window. Adjusting contrast does NOT affect the raw data, it only allows you to see spots better for this step.

To grid, you simply click on three spots. First, click on the button that says "Set Top Left Spot" and then click on the center of the top left spot. Second, click on the button that says "Set Top Right Spot" and then click on the center of the top right spot. Third, click on the button that says "Set Bottom Row" and then click on the center of any spot in the bottom row. Choose a good spot to make this step easier. Enter the information for the number of rows and columns. Rows and columns are defined based on the way you are currently viewing the tiff file. To finish this grid, click on "Update" button. At this time, you should see all the spots in the first grid surrounded by boxes as shown to the right. (You may need to zoom out to see the full grid.)



At this time, see if the spots look centered in the boxes. If not, then adjust the position of the boxes either by clicking on the appropriate button and then the correct spot, or by manually typing in numbers to adjust the boxes. Note that the position of the mouse is displayed in the bottom left corner of the window so you can determine if the numbers should be bigger or smaller to shift the boxes in the correct direction. This step take a bit of practice, but it is WAY easier than most other methods for gridding.

Once the first grid is properly gridded, it is time to repeat this process for grid number two. Click on tab 2 at the top of the window. You may need to zoom in again. Click on the button that says "Set Top Left Spot" and then click on the center of the top left spot. At this time, you should see a box next to the phrase (apply from grid 1). Select this box and then click on the "Update" button. This should apply a set of boxes around the spots that is close to what you wanted to do. Again, you can adjust this grid of boxes as needed. Continue this process until all the grids are surrounded with the boxes.

When you have finished gridding all the grids on the microarray, click on the "Done!" button. A grid file should be saved in your project folder and have a suffix of ".grid". If the default file name ends in .txt (a temporary bug), then delete the .txt and then save. You may also save a partially gridded project using the "Save Current Grid As…" under the file menu of the gridding window.

If your gene file and the number of spots do not match, you will get an error message. Your gene list should match the number of spots. If not, you probably will make an error identifying the spots later so you are urged to fix this problem now. If your gene list and the number of gridded spots match, then you will be informed of the total number of spots and allowed to save the grid file for further use.

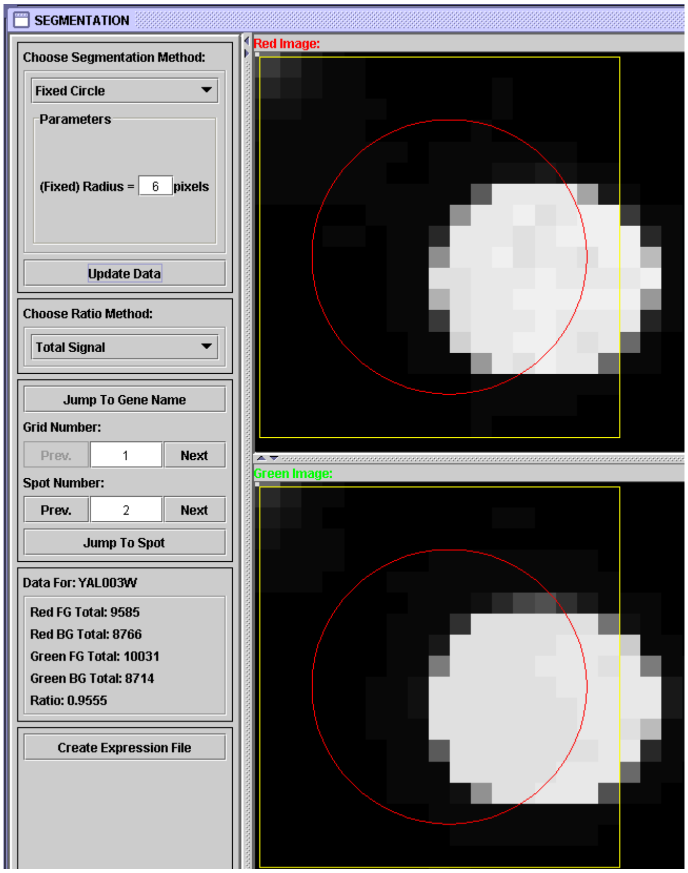
You can also save a file of the combined tiff images. You can save as tiff, jpg or gif. Tiff format works on all drawing and word processing programs so it is a universal format. Jpeg is good for images such as this that have many shades, like a photograph. Gif is the simplest format but may lose some of the subtlety of your original file. This saved merged image is useful if you want to take a picture of the overall grid and can be used for publishing or teaching.

Segmentation (Control S)

Segmentation is the process of distinguishing signal from background. There are three methods available for this process. During segmentation, you will have the opportunity to view each feature on the entire microarray but in this step, the two tiff files are separated again so the red image is on top and the green image on bottom. In this example, you can see the features are in the box, but they are not centered. This is an important component of segmentation.

Fixed Circle

Fixed Circle simply places a circle in the middle of the box. This indicates all pixels inside the circle (and inside the box) will be considered signal and pixels outside the circle will be background. You can set the radius of the circle in pixel units. Because each feature is of variable size, there may be disadvantages for this method. However, fixed circle is the most common method for segmentation. This is the fastest of the three segmentation methods.



Choose Segmentation Method: Fixed Circle

(Fixed) Radius = 6 pixels

Choose Ratio Method: Total Signal

Grid Number: 1

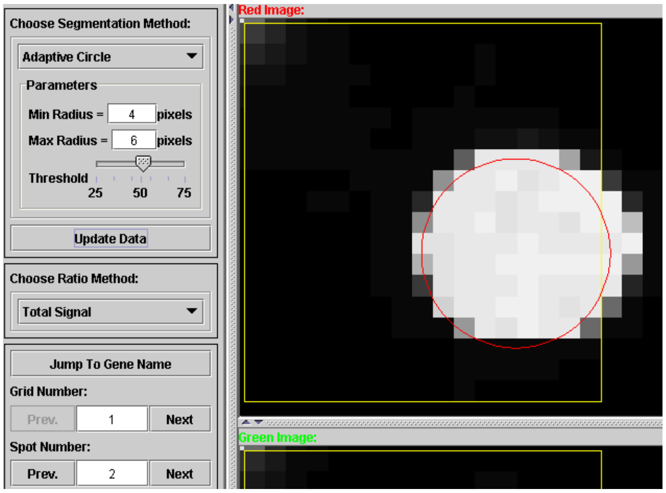
Spot Number: 2

Data For: YAL003W

Red FG Total: 9585
Red BG Total: 8766
Green FG Total: 10031
Green BG Total: 8714
Ratio: 0.9555

Adaptive Circle

This method changes the center and radius of the circle to fit the size and location of each feature. The algorithm considers all pixels above a user-specified threshold to be "on," and finds the circle with the highest percentage of pixels that are on. The radius can range between a user-specified lower and upper bound; the center can be anywhere inside the grid box. This method is slightly slower than Fixed Circle, but generally covers the actual spot better.



Choose Segmentation Method: Adaptive Circle

Min Radius = 4 pixels

Max Radius = 6 pixels

Choose Ratio Method: Total Signal

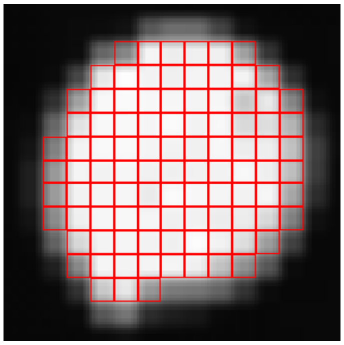
Grid Number: 1

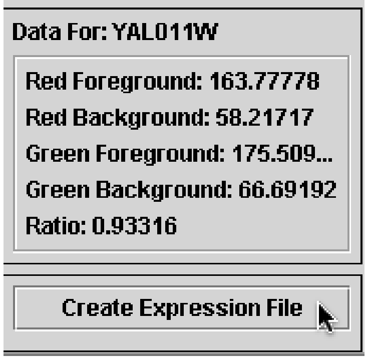
Spot Number: 2

Seeded Region Growing

This method for segmentation is designed to find the signal for each spot based on the distribution of the signal. Seeded region growing looks for the brightest pixel and then connects all pixels adjacent to this pixel into one shape. The algorithm simultaneously connects pixels to background and foreground regions, continuing until all pixels are in one of the regions. A user-specified threshold determines which pixels can be used to "seed" the regions. This is the slowest method since each pixel is processed individually.

Regardless of which method you choose, you can visually inspect the features to verify the gridding and segmentation were performed adequately. This inspection gives you a chance to flag any features you think should not be considered during subsequent data analysis.



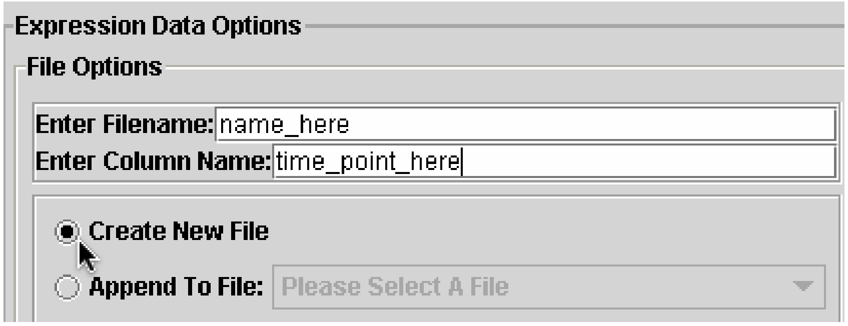


Data For: YAL011W

Red Foreground: 163.77778
Red Background: 58.21717
Green Foreground: 175.509...
Green Background: 66.69192
Ratio: 0.93316

When you complete segmentation, you will produce an expression file. Click on "Create Expression File" when you are satisfied with the segmentation process. This will generate an expression file, which was the goal of the first half of MAGIC Tool. An expression file contains the numerical values for each color for each spot as well as the ratios for each spot (red ÷ green). The ratios will be used for all subsequent data analysis. You do not need the tiff files any more.

You will need to name the expression file and the column (e.g. time point, treatment, etc.). You can append this to an existing file or create a new one. MAGIC will ignore certain entries in the gene name column ("blank", "empty", "missing" and "none"; case insensitive). Genes with non-unique names will be assigned new names corresponding to how many times the name appears. For example, if YBL023c appears in four different locations on the array, the first location (in grid and spot order) will be designated YBL023c_rep1, the second YBL023c_rep2, and so on.



Expression Data Options

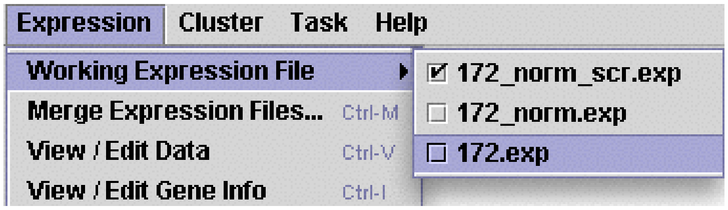
File Options

Enter Filename: name_here
Enter Column Name: time_point_here

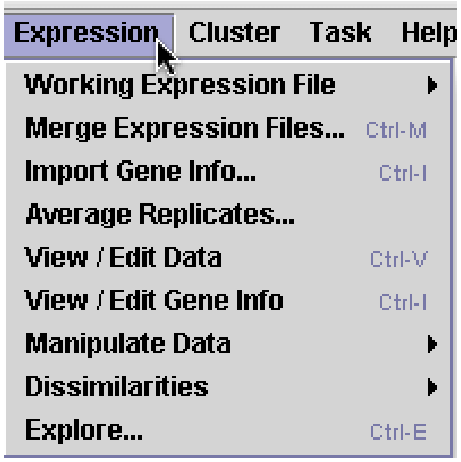
● Create New File
○ Append To File: Please Select A File

Expression

Working Expression File

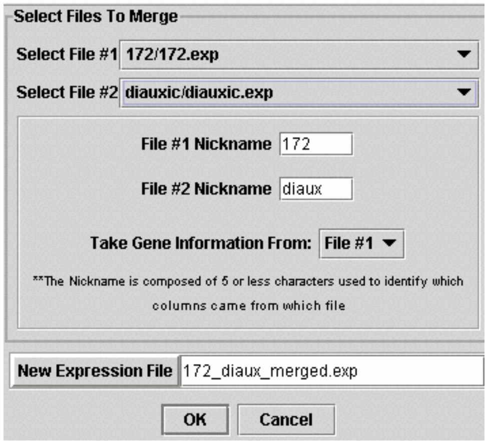


This option allows you to choose from a range of expression files within a single project. As you can see from the image on the left, you can choose which one is active simply by clicking on it.



Merge Expression Files… (Control M)

Merging expression files allows you to combine data from multiple chips so you can evaluate time course data, or other related data sets. You merge files one at a time and provide nicknames to assist MAGIC in keeping track of the soon to be combined data. Also, you can select one gene information list as the one that is retained with the merged data set. A new file will be created, so your two original files are not lost.


Select Files To Merge

Select File #1 172/172.exp

Select File #2 diauxic/diauxic.exp

File #1 Nickname 172

File #2 Nickname diaux

Take Gene Information From: File #1

**The Nickname is composed of 5 or less characters used to identify which columns came from which file

New Expression File 172_diaux_merged.exp

Import Gene Info… (Control I)

This allows you to compile more complete information about your ORFs. For example, we have created a text file that describes the chromosomal location, the three categories of gene ontology annotation, and synonym for all yeast genes. This permits you to search by each of these fields to help detect trends and meaningful information.

Average Replicates

MAGIC Tool treats every spot as a unique feature and does not average for replicate genes automatically. This preserves your raw data. After you have created expression files, you may choose to average duplicate spots as defined by ORF name. Expression data will be averaged over all genes with the same name up to the unique “_rep#” tag.

View/Edit Data (Control V)

After an expression file is created or merged, you can view and edit the data. This option should not be used often, but we did want you to have access to the ratio data if you deem it necessary. It is also helpful if you want to verify steps or pick up a project after an extended period of time.

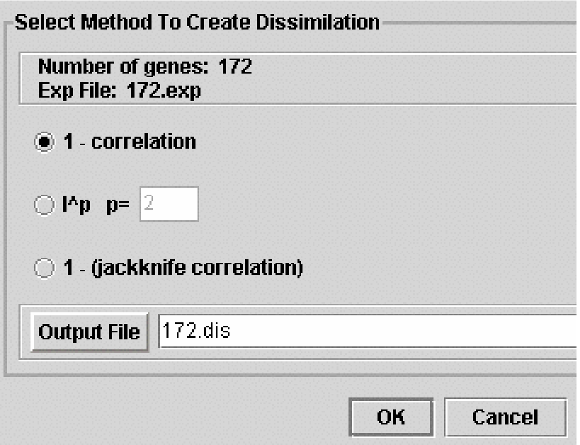
View/Edit Gene Info (Control I)

Similar to above, this option allows you to view and modify the gene list. Of course, you can view and edit the gene list outside MAGIC Tool, but this option provides you an opportunity to do so within MAGIC. Perhaps you will want to perform a search on the gene function. Viewing the list can allow you to select appropriate terms for searching.

Dissimilarities (Control D)

Calculating dissimilarities allows you to compare different genes to one another. The first step in this process is to generate correlation coefficients (see Instructor's Guide online for a detailed explanation). When you do this, a window will appear where you have to choose from three options.

The most common method is the default '1 – correlation.' The other two methods are described in the Instructor's Guide. When this step is complete, MAGIC generates a dissimilarity file which you can name in the output file box, but be sure to retain the suffix ".dis". Click on OK to begin this process. The progress is monitored in a scale bar below the OK button (not shown here).



Select Method To Create Dissimilation

Number of genes: 172
Exp File: 172.exp

● 1 - correlation

○ l^p p= 2

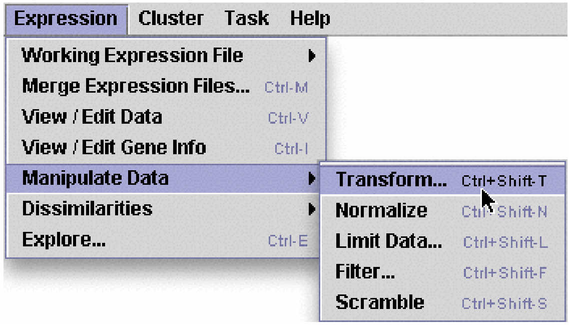
○ 1 - (jackknife correlation)

Output File: 172.dis

'1 – jackknife' correlation is worth mentioning briefly. This method is beneficial because it adjusts the dissimilarity calculation to take into account single columns of data that may be outliers from the rest of the data. For example, if all ratios are near 1 but a single time point has a ratio of 6, this would skew the dissimilarity calculation to place excessive emphasis on the single ratio of 6. Although '1-jackknife' requires a much longer time to compute, the results may produce more meaningful clusters.

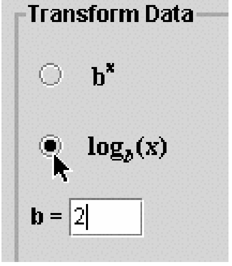
Manipulate Data

Manipulating data is not as bad as it sounds. This option allows you to choose from five options. These options do NOT alter your original data; they simply allow you to process the data further prior to clustering or exploring your data.



Transform (Control Shift T)

A standard process you should perform is transforming your data before performing any analysis (exploring or calculating dissimilarities and clustering). You want to log-transform your ratios so you eliminate any fractions. It is important to get all ratios on the same scale of magnitude. For example, if a gene is repressed 16 fold, the ratio will be 0.0625 while a gene that is induced 16 fold will have a ratio of 16.0. Before analyzing your data, you should log-transform your data. After transformation (typically $\log_2$), the two genes would be altered (-4 vs. +4) with equal magnitude but in opposite directions. See Instructor's Guide online for more information. You should explore after transforming, but may or may not want to normalize before exploring (see below). If you want to "un-transform" your transformed data, you can use the exponent function b^x.



Transform Data

○ b^x

● $\log_b(x)$

b = 2

Normalize (Control Shift N)

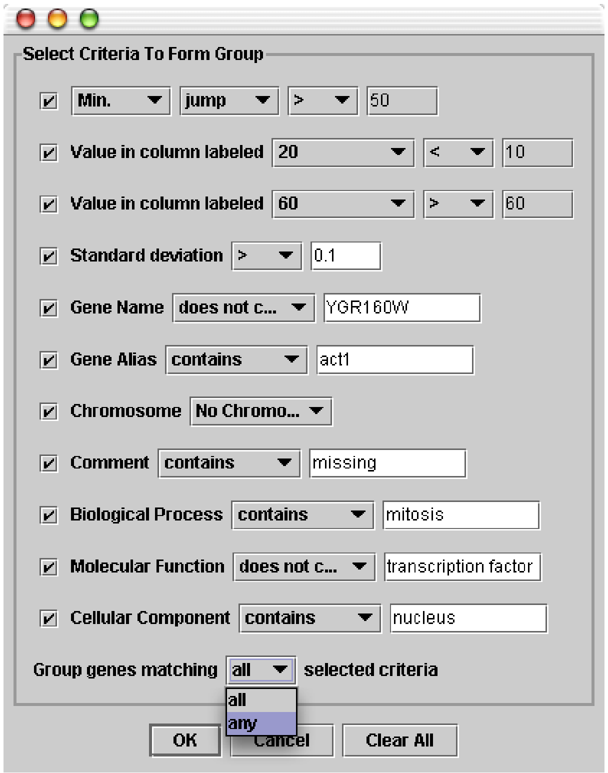
This process takes your (transformed) ratios and corrects for the magnitude of a gene's ratios and the variation among each gene's ratios. Normalization is not necessary before calculating the correlation coefficient. Although you can perform normalization on raw ratio values (expression files), we strongly recommend you log-transform your data first (see above). The best sequence of steps is to transform your data, normalize the transformed data, calculate the dissimilarities, and then cluster. You should explore after transforming, but may or may not want to normalize before exploring. See Instructor's Guide online for more details.

Limit Data (Control Shift L)

If you have merged data from many microarrays (e.g. a time course experiment), you may want to study only certain portions of your merged data independently. Limiting data allows you to select column headings and retain these selected data for analysis in a "limited data set." Your original merged file is left unaltered and a new file is created. The new expression file will terminate with the name "x_limted.exp" where x would be the original expression file name.

Filter (Control Shift F)

Filtering allows you to remove from further consideration genes that do or do not meet user-defined criteria. The image below shows you the types of filters that can be applied.



Select Criteria To Form Group

☑ Min. ▼ jump ▼ > ▼ 50

☑ Value in column labeled 20 ▼ < ▼ 10

☑ Value in column labeled 60 ▼ > ▼ 60

☑ Standard deviation > ▼ 0.1

☑ Gene Name does not c... ▼ YGR160W

☑ Gene Alias contains ▼ act1

☑ Chromosome No Chromo... ▼

☑ Comment contains ▼ missing

☑ Biological Process contains ▼ mitosis

☑ Molecular Function does not c... ▼ transcription factor

☑ Cellular Component contains ▼ nucleus

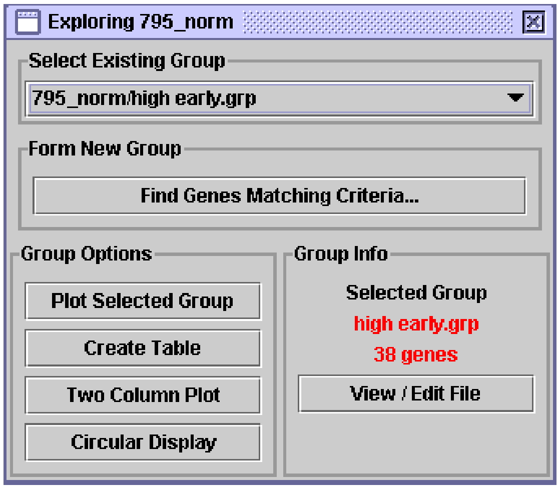
Group genes matching all ▼ selected criteria

Scramble (Control Shift S)

Scrambling your expression data and repeating certain analysis steps can help you verify that patterns in your data are biologically significant, rather than due to chance. There are three different ways to scramble your data: (1) scramble each column (experimental condition) independently, (2) scramble each row (gene) independently, and (3) scramble all data.

Explore (Control E)

After you have transformed your data, you can explore it in a number of ways. The default group of genes is the full set. You can select a subset of genes via the Form New Group button called "Find Genes Matching Criteria…" You can search for criteria similar to those shown for the filter set on the previous page. When you have identified genes of interest, the window changes as shown to the right in red text. To save this new group of genes, click on the "View/Edit file" button just below the red text.



Exploring 795_norm

Select Existing Group

795_norm/high early.grp ▼

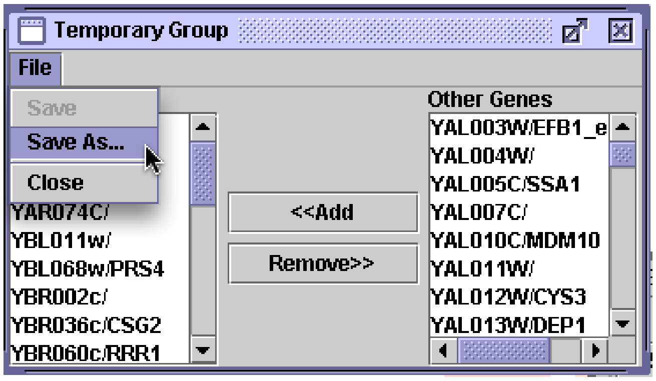
Group Info

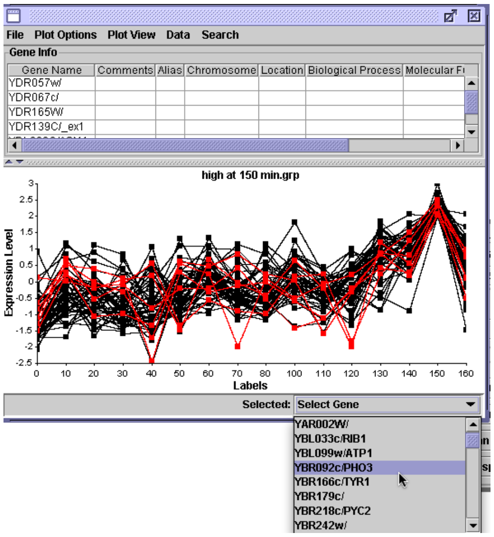
Selected Group

high early.grp

38 genes

A new window will appear that lets you view the list of genes in your newly formed group. You can modify this group if you want, or you can “save as” under the file menu. You can create many subgroups of genes and explore them individually using the “select Existing Group” pull down menu. Once you have subsets of genes to explore, you can visualize them in a number of ways:

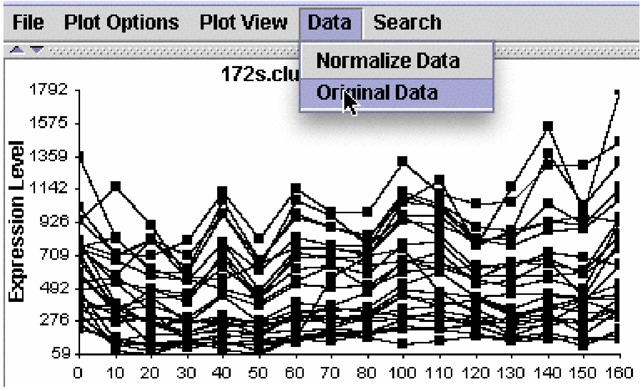


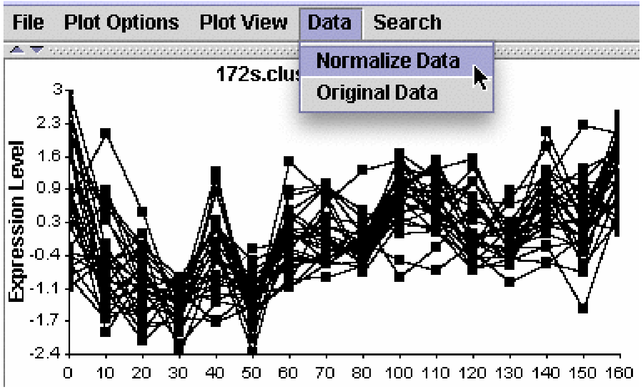


Gene Name	Comments	Alias	Chromosome	Location	Biological Process	Molecular F
YDR057w/						
YDR067c/						
YDR165W/						
YDR139C/_ex1						

Plot Selected Group

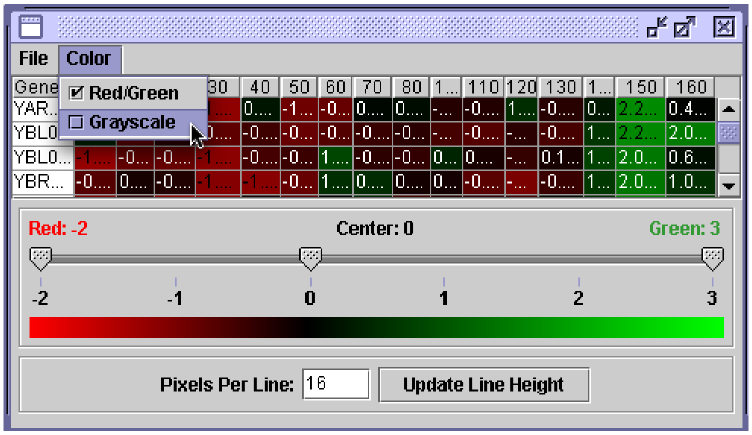
You can have the ratios plotted graphically. You can select one gene using the pull down menu in the bottom right corner. Or, as shown here, you can click on one node at a time and hold down the shift key to select multiple genes (in this case, those with the lowest ratios in the group). These selected genes are listed in the top window (which you can pull down to see) as well as any other information about these genes in your gene list. You can adjust the size of the plot, as well as zoom in on a section. For example, this group of genes was selected by having a ratio of 2 or more at 150 minutes. To untangle the crowded lines, you can zoom in on any region of interest. To do this, hold down the control button then click and drag a box around the crowded area to zoom in. You can unzoom using the Plot View menu at the top of the window. In addition, you can label the axes, save this as a file, print this plot, normalize the data (if you have not already done so), change the size and shape of the points, and search for certain terms for the genes based on the gene list from which these genes are derived.





Create Table

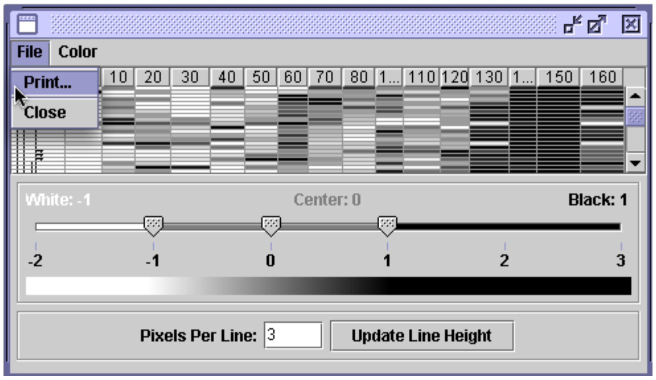
This feature is unique to MAGIC Tool and creates a dynamic table. The default is a grayscale table, but you can change this to a red-green scale if you prefer. The most interesting feature of this interactive table is the scale bar and the three sliding tabs. Imagine a gene set that has one gene with a very high ratio (e.g. +16) and one gene with a very low ratio (-16) but with most genes having ratios between +3 and –3. Because of these two extreme genes, the color differences in the remaining genes would be lost. However, if you adjust the tabs, you can compress the color scale on the extreme ends and bring more color variation to the middle of the range of ratios, where most of your genes are located. In the image below, the scale has been adjusted and converted to grayscale to illustrate the potential views in the Create Table option. In this view, the gene lines have been reduced from 16 pixels high to 3 pixels high, the color scale changed to grayscale and the range reduced to –1 to +1. This reduction makes all high and low values either white or black, but allows the intermediate values to be on the grayscale.



Gene				30	40	50	60	70	80	1...	110	120	130	1...	150	160
YAR...				....	0....	-1...	-0...	0....	0....	-...	-0....	1....	-0....	0...	2.2...	0.4...
YBL0				0....	-0....	-0...	-0...	-0....	-0...	-...	-0....	-...	-0....	1...	2.2...	2.0...
YBL0...	-1....	-0...	-0....	-1....	-0....	-0...	1....	-0....	-0...	0...	0....	-...	0.1...	1...	2.0...	0.6...
YBR...	-0....	0....	-0....	-1....	-1....	-0...	1....	0....	0....	0...	-0....	-...	-0....	1...	2.0...	1.0...

Red: -2 Center: 0 Green: 3

Pixels Per Line: 16

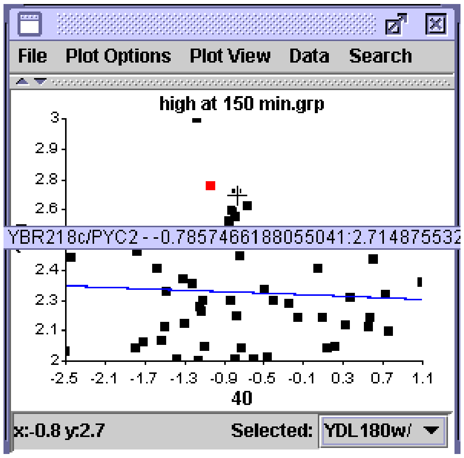


White: -1 Center: 0 Black: 1

Pixels Per Line: 3

Two Column Plot

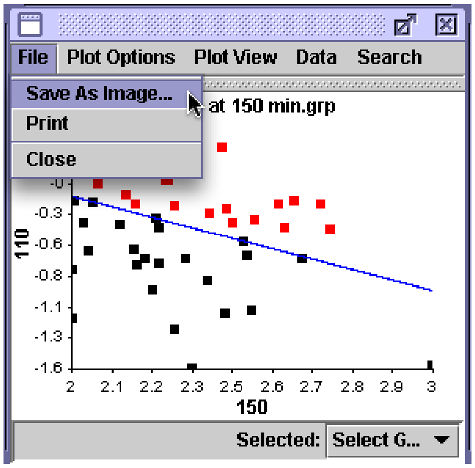
This plot allows you to select two columns of data and compare their ratios. As you can see, some comparisons are more similar than others. In this plot, you can select a single gene (left) or many genes (right; hold down the shift key while clicking). If you mouse over a gene, the display will tell you the two ratios for the two time points. You can also see an approximation in the bottom left corner.



YBR218c/PYC2 - -0.7857466188055041:2.714875532

x:-0.8 y:2.7

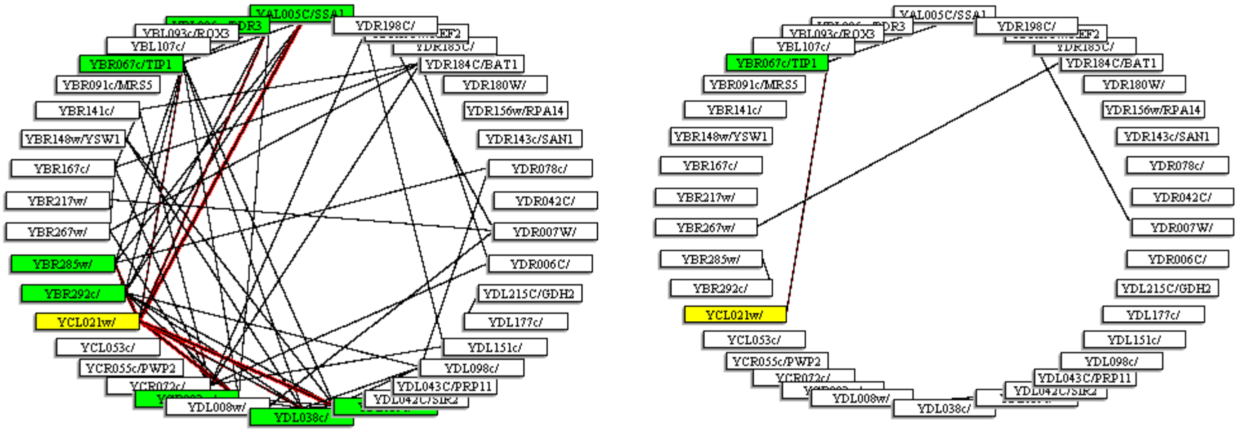
Selected: YDL180w/



Selected: Select G...

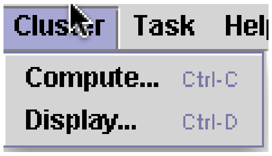
Circular Display

Another unique MAGIC Tool display is the circular one. Let's imagine you have created a group of genes and you want to know the correlation coefficients among them. The default setting is correlation coefficient of 0.8 which is shown on the left. Using the display menu, you can change the radius of the circle and the threshold for reporting correlations. Change the threshold to 0.1 (correlation of 0.9) and you see fewer lines connecting the genes (right). In this case, the same gene was clicked on (yellow) and the genes that met the threshold are colored green with the lines colored red.



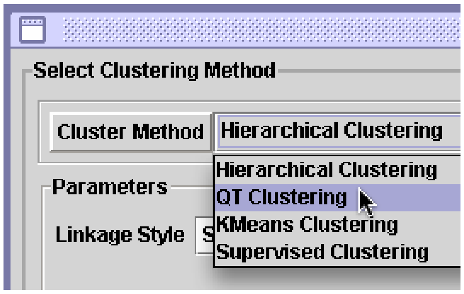
Cluster

Compute… (Control C)
Once you have created dissimilarity file, you may cluster your data. To do this you must computer the cluster using one of four methods. Details for these four methods can be found in the Instructor's Guide online.



Hierarchical Clustering

Hierarchical clustering produces a tree-like structure (a *dendrogram*) by connecting genes according to the similarity of their expression data. When a gene joins with another gene or group of genes in the tree, the entire collection of genes is represented as a single pseudo-gene. The similarity between a given gene and the gene (or pseudo-gene) to which it is connected is indicated by the horizontal length of the branches joining them. At each stage in the algorithm, the two most similar genes or pseudo-genes are joined together. The process continues until all genes have joined the tree.



QT Clustering

QT Cluster takes every gene under consideration and one at a time, builds a temporary cluster for each gene with a user-defined cutoff value for similarity. Whichever gene garnered the most genes in its cluster is used to create permanent cluster and all the genes associated in this cluster are removed from the list of genes for the next round of creating permanent clusters. QT Cluster repeats the process of creating temporary clusters, one gene at a time, and then forms the second permanent cluster using the largest temporary cluster. This process is repeated until all the genes are in clusters, or the remaining genes form clusters smaller than a user-defined size. These remaining genes (called *singletons*) are not presented in the clustering displays unless the user defined 1 as the minimal size for a permanent cluster.

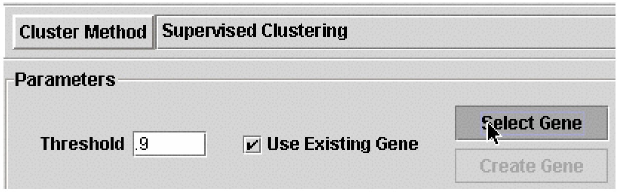
When you use QT Cluster, you should adjust the threshold value. The default of 0.9 means correlation coefficients of +0.1 through +1.0. If you change the threshold setting to 0.2, you will cluster genes only if their correlation coefficients are +0.8 through +1.0. The range of settings for threshold is from 0 (correlation of +1.0) through 1 (correlation of 0, i.e. not similar at all) to 2 (correlation of –1.0; track opposite each other). Therefore, by setting the threshold at 2, you would get every single gene placed in one cluster.

K-Means Clustering

In this method, you determine *a priori* how many clusters there will be (K = the number of clusters) and MAGIC tool will make sure all genes fit into this number of clusters. This is the first step in Self Organized Maps but both methods begin with the investigator determining how many clusters to generate.

Supervised Clustering

This method performs a QT cluster but you can define the threshold and choose one gene around which you want your cluster built. This allows you to focus your research on your favorite gene. On the left, you see that "Use Existing Gene" is selected. Click on the "Select Gene" button and then choose form the genes in your gene list of the currently active expression file.



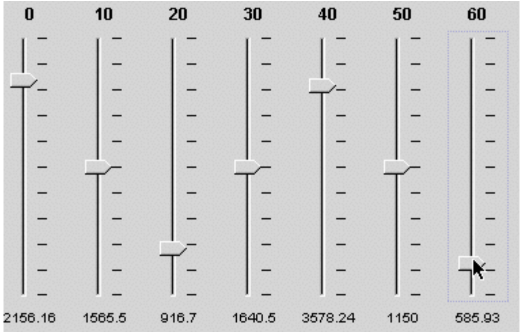
Cluster Method: Supervised Clustering

Parameters

Threshold .9

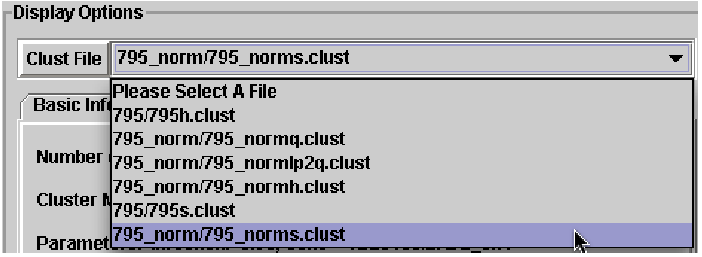
☑ Use Existing Gene

Alternatively, you can deselect the "Use Existing Gene" option and then click on "Create Gene". This produces a window that allows you to manipulate the sliders to create an expression profile for which you want to find genes with similar profiles (based on the threshold you choose). This is a quick way to find complex patters of interest to you.



Display…

Once you have create a cluster or two, you can display them. First, choose the cluster file you want to display. Each type of cluster has its own display options.



Display Options

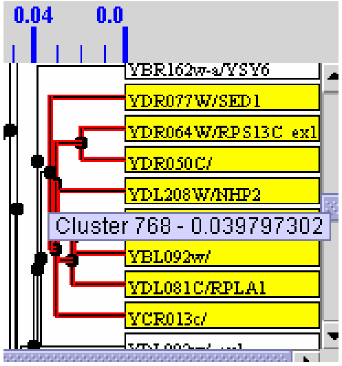
Clust File: 795_norm/795_norms.clust

Please Select A File
795/795h.clust
795_norm/795_normq.clust
795_norm/795_normlp2q.clust
795_norm/795_normh.clust
795/795s.clust
795_norm/795_norms.clust

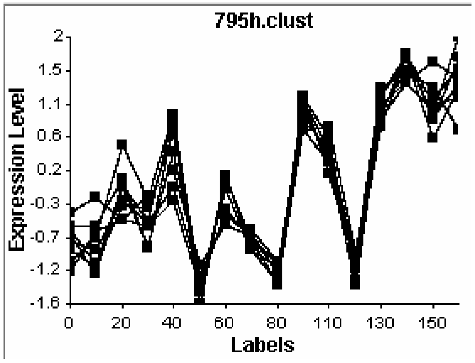
Hierarchical Cluster Display

You have three options for display, each of which has its own options. Metric Tree is unique to hierarchical clustering. It produces a dendogram with nodes plotted at indicated thresholds. The smaller the threshold number, the higher the correlation coefficient.

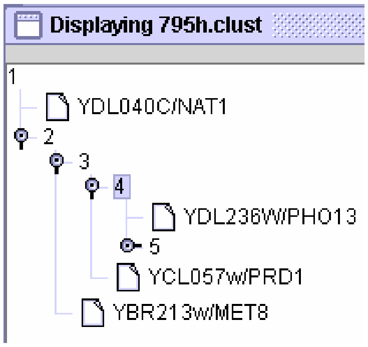
You can click on a branch point and highlight all the genes within this cluster as shown. If you mouse over the branch point, you can see the exact threshold, which is 1 minus the correlation coefficient (~0.96). You can plot this cluster and as you would image with this high a correlation coefficient, the normalized data plot as a very tight group.



YBR162w-a/YSY6
YDR077W/SED1
YDR064W/RPS13C exl
YDR050C/
YDL208W/NHP2
Cluster 768 - 0.039797302
YBL092w/
YDL081C/RPLA1
YCR013c/



Exploding Tree is an efficient way to show clusters and gradually expand the contents of each node. In this example, there is one gene and then all other genes are within node number 2. As you click on the nodes, they expand and if you click a second time, they collapse. You can explode the node completely by highlighting the number and clicking on the explode button, or explode it one at a time by clicking on the node directly. You can also plot any cluster within a node by clicking on the "Plot Node As Group" button.

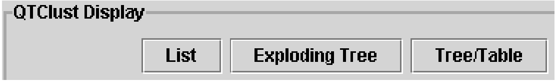

Displaying 795h.clust

1
YDL040C/NAT1
2
3
4
YDL236W/PHO13
5
YCL057w/PRD1
YBR213W/MET8

Tree/Table is a way to combine the Table view and the dendrogram. The dendrogram is on the far left and the colored table (the majority of the window) is displayed on the right (view not shown).

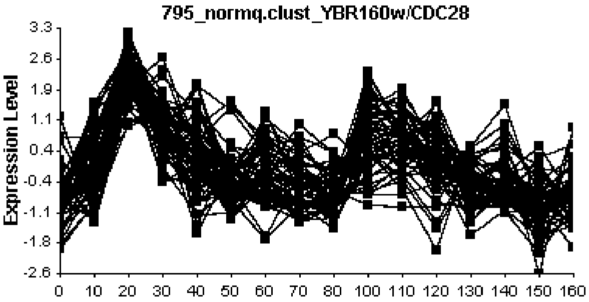
QT Cluster Display

QT cluster also allows Exploding tree and Tree/Table, but it has replaced the metric tree with List. List allows you to see the name of the root gene for each cluster. If you click on the root gene, then all the genes within this cluster are displayed. You can plot this cluster as shown here.



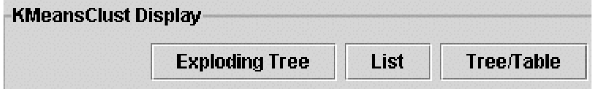
Supervised (QT) Cluster Display

Supervised Cluster hast the same display options as regular QT Cluster. However, when you are choosing your display, you should note the box that indicates what threshold was used and which gene was used as the root. In this case, ERD2, the KDEL receptor exon 1 was used as the root for this cluster with a correlation coefficient of 0.95 (plot not shown).



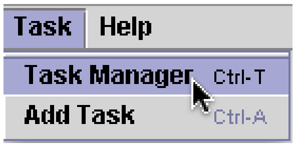
K-means Cluster Display

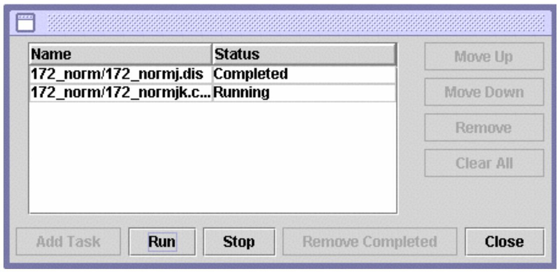
The three displays possible for K-means cluster display are described above.



Task

As your datasets get bigger, the time it will take to make all the necessary calculations will increase rapidly. Therefore, MAGIC allows you to establish a list of tasks to be performed in sequence. You can tell MAGIC to begin a series of steps and then walk away from your computer. MAGIC will perform this sequence of tasks while you do other things. For example, you can establish a list of tasks to perform and go home for the night. When you return the next morning, MAGIC will have completed the series of tasks. At this time, the only tasks that can be performed are calculating dissimilarities and clusters.





Name	Status
172_norm/172_normj.dis	Completed
172_norm/172_normjk.c...	Running

Task Manager (Control Shift M)

The window above is the task manager. It allows you to add or remove a task, change the order of a task as well as various housekeeping chores.

Add Task (Control T)

This option allows you to add a task without going through the task manager.

Help (Control H)

Currently, electronic help is under development. We hope to get this working soon. Its content will be very similar to this User's Guide.

Credits

MAGIC Tool version 1.0 was written in JAVA by Adam Abele, Brian Akin, Danielle Choi, Parul Karnik, and David Moskowitz. Laurie J. Heyer and A. Malcolm Campbell were advisors to the code-writing team.
MAGIC Tool was developed at Davidson College and supported by the NSF, Duke Endowment, and Davidson College.
We would like to thank Wolfgang Christian and Mario Belloni for sharing their knowledge and resources with us.
The ImageJ package integrated into MAGIC Tool is an excellent general image analysis package by Wayne Rasband (wayne@codon.nih.gov). ImageJ was inspired by NIH Image, and is freely available at <http://rsbweb.nih.gov/ij/>. We are grateful for this generous contribution to the public domain.

Acknowledgements

We would like to thank all the members of GCAT, especially our co-PIs Laura Hoopes and Todd Eckdahl. We are particularly indepted to the Davidson College undergraduates who wrote the code for MAGIC Tool: Adam Abele, Brian Akin, Danielle Choi, Parul Karnik, and David Moskowitz. This work was supported by Davidson College Faculty Study and Research grants, the NSF DBI-0099720. the Duke Endowment, and the Associated Colleges of the South.

Literature Cited

Alzadeh, Ash, et al. 2000. Distinctive types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503-511.

Campbell, A. Malcolm and Laurie J. Heyer. 2003. Discovering Genomics, Proteomics and Bioinformatics. First Edition. Cold Spring Harbor Laboratory Press and Benjamin Cummings. San Francisco, CA. 352 pages.

Campbell, A. Malcolm. 2003. Instructor's Guide to Discovering Genomics, Proteomics, and Bioinformatics. Benjamin Cummings. San Francisco, CA. 136 pages.

Campbell, A. Malcolm and Laurie J. Heyer. 2003. Instructor's Guide to MAGIC Tool. <http://www.bio.davidson.edu/projects/magic/MAGIC_Instrs_Guide.pdf> Accessed 22 August, 2003.

Campbell, A. Malcolm and Laurie J. Heyer. 2003. Additional Math Background for Understanding Data Analysis. <http://www.bio.davidson.edu/projects/magic/Math_Analysis.pdf> Accessed 22 August, 2003.

DeRisi, et al. 1997. Exploring metabolic and genetic control of gene expression on a genomic scale. Science 278:680 – 686.

Gasch, Audrey et al. 2000. Genomic expression programs in the response of yeast cells to environmental changes. Molecular Biology of the Cell 11:4241-4257.