Gene expression

MAGIC Tool: integrated microarray data analysis

L. J. Heyer[1,*], D. Z. Moskowitz[1], J. A. Abele[1], P. Karnik[1], D. Choi[2], A. M. Campbell[2], E. E. Oldham[1] and B. K. Akin[1]

[1]Department of Mathematics and [2]Department of Biology, Davidson College, Davidson, NC, USA

ABSTRACT

Summary: Several programs are now available for analyzing the large datasets arising from cDNA microarray experiments. Most programs are expensive commercial packages or require expensive third party software. Some are freely available to academic researchers, but are limited to one operating system. MicroArray Genome Imaging and Clustering Tool (MAGIC Tool) is an open source program that works on all major platforms, and takes users 'from tiff to gif'. Several unique features of MAGIC Tool are particularly useful for research and teaching.

Availability: http://www.bio.davidson.edu/MAGIC

Contact: laheyer@davidson.edu

1 INTRODUCTION

MicroArray Genome Imaging and Clustering Tool (MAGIC Tool) enables investigators to explore and analyze all types of gene-expression data on all major operating systems (Windows, Mac OS X, Linux and Solaris). Flexible entry and exit points, ranging from reading and quantifying tiff image files to clustering and other data mining operations, allow users to run MAGIC Tool exclusively, or in conjunction with other programs. Novel features of MAGIC Tool contribute to its power and simplicity, including a 3-click gridding procedure, interactive segmentation and several dynamic display formats.

MAGIC Tool is particularly well suited for academic labs, having several advantages over other microarray analysis packages:

(1) The user interface is designed to illuminate the analysis procedures and algorithms, encouraging the user to understand each step rather than running data through a 'black box.'

(2) The Java source code is freely available and open source (under the Gnu General Public License), allowing faculty and students to test their own specialized algorithms within the existing interface.

(3) The program runs on multiple platforms, without need for compilation or specialized configuration.

(4) Using a single program for all facets of microarray data analysis greatly simplifies software management and user training.

---

*To whom correspondence should be addressed.

2 METHODS

2.1 Image analysis

MAGIC Tool reads from a tab-delimited text file the names of genes (including replicates) represented on the array, listed in spot order. The first spot may be at any of the four corners of any grid; the order is determined by the horizontal and vertical directions of increasing spot numbers, and the placement of Spot 2 relative to Spot 1 (either horizontally or vertically). The user enters these three parameters (horizontal order, vertical order and Spot 2 orientation) after being shown an overlayed image of two tiff files (corresponding to the two dyes used to label the hybridized samples).

Gridding in MAGIC Tool does not require grid and feature dimensions or spacing. Zooming in on the first grid, the user clicks on three points: the top left spot, the top right spot and any point on the bottom row. Entering the number of rows and columns completes the geometric description of the first grid, placing each spot in its own grid square. Remaining grids may be copied from existing grids, or entered independently with the 3-click method.

Segmentation, the process of separating spot signal from background, is performed with one of three algorithms: fixed circle, adaptive circle or seeded region growing. The fixed circle is centered in the grid square, with a user-specified radius. The adaptive circle algorithm examines the signal in each grid square to determine the most appropriate center and radius (within a user-specified range) for each circle. The red and green images for the current grid square are considered individually, and all pixels with intensity above a user-specified threshold (percentile) are marked as 'on.' The resulting binary images are combined with a logical OR, creating a single binary indication of 'on' pixels in either image. Finally, the adaptive circle's center and radius are set to be those containing the largest number of 'on' pixels. The Seeded Region Growing algorithm (Adams and Bischof, 1994) connects each pixel to a background or foreground region, continuing until all pixels are assigned. A user-specified threshold and geometric considerations (i.e. foreground near the center, background near the corners) determine which pixels are used to 'seed' the regions. Screen shots at the web page illustrate the gridding and segmentation processes.

A unique feature of MAGIC Tool is the interactive segmentation browser, a gray-scale view of red and green channel images for individual grid squares. The browser displays all pixels in the current grid square, foreground and background intensities, and the result of the chosen ratio computation method (pixel average or total, with or without background subtraction). The user can jump to a particular spot by number or gene name, or navigate through a region of a slide, one spot at a time. A screen shot of the segmentation browser is available at the web page. During segmentation, the gridding window can be reopened for an overview of the slide, assisting the user in finding spots of particular interest. When satisfied with segmentation performance, the user saves the resulting column of expression ratios into a new file, or appends it to an existing file with the same gene list.

2.2 Exploring and clustering

An expression file, consisting of one or more columns of gene expression ratios or levels, can be imported and analyzed in many different ways. MAGIC Tool has a novel intensity display that shows the expression level of each gene numerically and on a gray or red–green color scale. By adjusting the center, minimum and maximum color values, expression values in a certain range can be distinguished on a finer scale. Another unique display is Circle Plot, which places a user-selected group of gene names around the circumference of a circle, and connects genes whose correlation is above a user-specified threshold. Clicking on a gene name highlights all connections to that gene, providing a convenient way to explore similarity among genes selected for function, chromosomal location, etc. Gene expression can also be displayed in graphs of level versus experimental condition (e.g. time), or in a scatter plot of one experimental condition versus another.

One of the most useful exploratory functions in MAGIC Tool is a flexible filtering tool that allows the user to build groups of genes based on biological and expression profile characteristics (e.g. all genes with 'ribosome' in the cellular component description field whose expression ratio is > 2.0 at any point in a time course). Expression data for these groups can be displayed in any of the ways described above, and group members saved for further exploration and reference. The web page includes a screen shot of the characteristics that can be used to form queries.

MAGIC Tool can cluster gene-expression patterns using hierarchical (e.g Eisen *et al.*, 1998), *k*-means, or QT-Clust (Heyer *et al.*, 1999) algorithms. Expression profiles can be compared using correlation coefficients and other norms. Cluster visualization tools include hierarchical trees, intensity displays and graphs of expression level versus condition.

3 CONCLUSION

MAGIC Tool fills a need for an open, extensible, multipurpose microarray data analysis platform. All graphical displays in MAGIC Tool can be saved as jpeg or gif images, enabling inclusion of its results in publications. The software supports research, with its flexible exploratory approach, and teaching, with its interactive and clearly observable functions. Further details, including a User's Guide, sample data files and other resources, are available at the website.

REFERENCES

Adams,R. and Bischof,L. (1994) Seeded region growing. *IEEE Trans. Pattern Anal. Mach. Intell.*, **16**, 641–647.

Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.

Heyer,L.J., Kruglyak,S. and Yooseph,S. (1999) Exploring gene expression: identification and analysis of co-expressed genes. *Genome Res.*, **9**, 1106–1115.