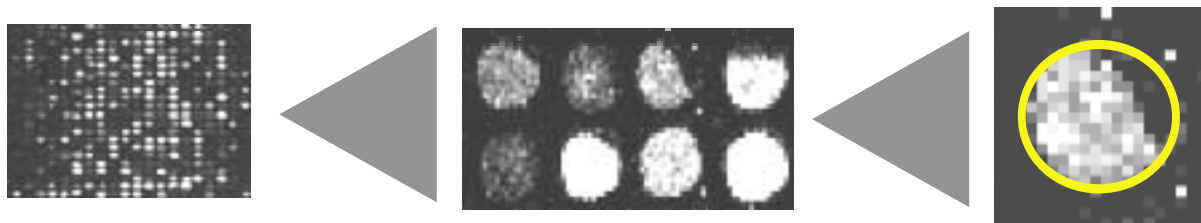




Image Analysis and Quantitation of cDNA Microarray Images

Chad Shaw and Jeff Tollett
January 19, 2001



Overview

In a typical cDNA microarray experiment laser scanning produces two 16 bit TIFF image files. One file corresponds to the reference sample dye-channel and one file to the experimental sample dye-channel. The goal of image analysis is to reduce the large number of pixel values in the image files to a small set of summary values representing each printed spot on the array. The total data reduction accomplished by image quantitation reduces 40MB of image file data to about 150KB of summary quantitation data.

The data set which is generated by image quantitation represents the expression information in the microarray experiment, and this quantitation data set is the object used in all downstream analysis of a microarray study. One should keep in mind, however, that the raw data are the image files themselves, not the quantitation files. If the raw image files are archived, this raw data may be re-processed any number of times with any number of image analysis techniques.

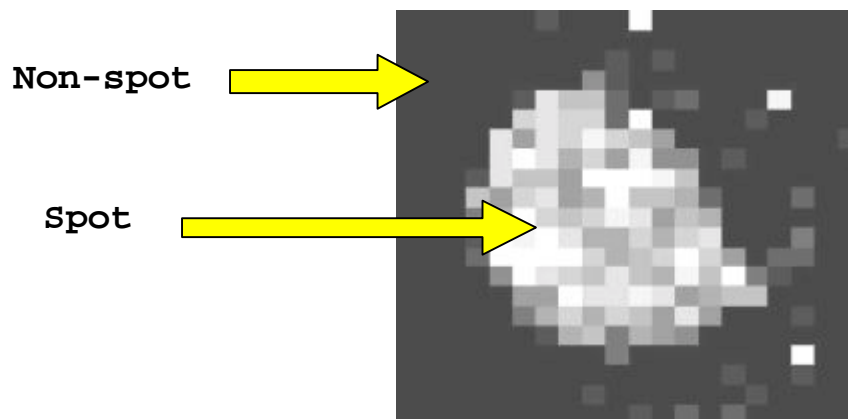
The fundamental operations needed to accomplish quantitation of microarray image files are:

1. **Spot Discrimination:** localizing the small region occupied by each printed cDNA.
2. **Spot Summarization:** a statistical procedure which condenses the pixel values within each spot to a single number or a small set of numbers representing the intensity of the spot.

The quantitation of cDNA microarray images can be accomplished by a variety of currently available commercial software. Software available for use at the Baylor Microarray Core Facility includes QuantArray, Axon, and Gleams.

Spot Discrimination

A spot is a region of the array occupied by printed cDNA. Non-spot regions are the empty space between printed material. Much of the array surface consists of intervening non-spot regions.



A first step in the quantitation of array data is to discriminate spot and non-spot regions in the microarray image file. The premise of spot discrimination is that positions where cDNA is printed have different brightness characteristics when compared to non-spot regions without printed cDNA.

For spots where sufficient labeled probe material has hybridized the spot is much brighter than the surrounding non-spot area. The spot is brighter because laser scanning detects fluorescence of bound labeled probe material and records the fluorescence intensities in the pixel values written into the TIFF file. For positions where material is printed but no labeled probe has hybridized, pixel values are equal to or lower than surrounding background regions. Values lower than

background can occur because the printed cDNA material excludes the material which generates background pixel values.

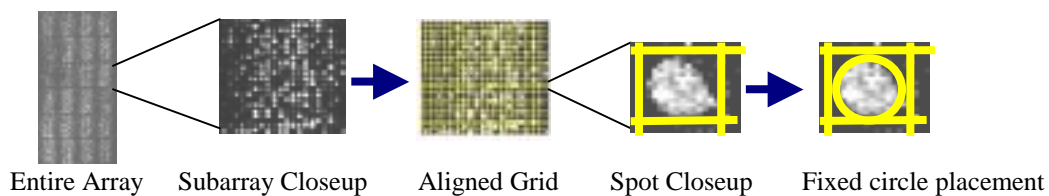
The pixels for non-spot regions with no printed cDNA are called background pixels, and the non-zero intensity values observed for these pixels result from unwashed fluorescent material adhering to the slide surface and from the optical properties of the blank slide, which can be highly non-uniform across the slide.

Grid Placement

The known geometry or approximate geometry of the cDNA printing procedure enhances the spot-finding procedure. In the first step of spot finding, a grid is overlaid and adapted to the geometric pattern of the printed cDNA. Placement of the grid can be manual, semi-automatic, or automatic. Grid placement procedures are implemented in all softwares in use at Baylor. Almost all of these procedures are semi-automated.

The center of each small square in the grid is an idealized spot center, and the region around each spot center are used to identify the boundaries of the spot in the grid. The pixel values in each grid box are the values ultimately used to summarize the expression values for each particular cDNA.

Once the grid is aligned and spot centers are identified, several methods can be used to identify the precise pixels within the spot. An important method is fixed circle, which inscribes a circle of fixed radius in each grid region. The pixel values within the inscribed circle are summarized to create the quantitation values for the spot.



Besides the fixed circle method, adaptive methods can also be used to identify spot regions within an aligned grid square. The adaptive method allows for variable spot morphology beyond the idealized circle shape. The adaptive methods still require grid placement, and these methods attempt to find spot regions within each

grid square. Both Gleams and QuantArray softwares implement adaptive methods in addition to fixed circle quantitation.

Adaptive spot finding within a grid region is difficult because of the irregular shapes of spots, the variation in background across the chip, and the presence of large defects, such as specks of dust. Fixed circle methods which floor out low pixel values may be more reliable for reporting spot intensities (as of 1/19/01).

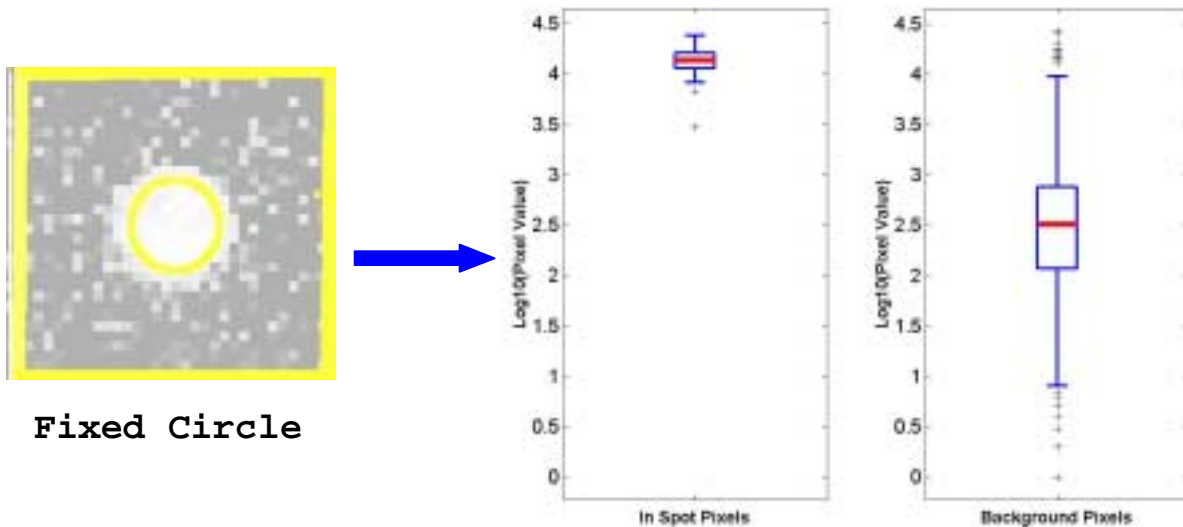
Spot Summarization

Once spot regions are discriminated on the array, statistical analysis condenses the pixel values within each spot region and summarizes these values into a single value or a small set of values. Spot summarization results in the final quantitation information used for further analysis.

Univariate Pixels

If spot discrimination is done separately for each of the two image files, then each spot is associated with a univariate set of pixel values. The goal of pixel summarization is to find the center and spread of the pixel values for each spot for each channel.

A picture of the univariate summarization procedure appears below. Placement of a fixed circle within a grid region discriminates in-spot pixels from background pixels. The in-spot and background pixel distributions are summarized by taking medians (or means) and calculating ranges (or standard deviations). A picture of the univariate summarization procedure appears below.



Bivariate Pixels

Historically array softwares operated a single channel at a time, so that independent analysis had to be performed on each image file. Some softwares now do image registration so that the pair of image files is brought into alignment, and the resulting pair-image file is analyzed (e.g. Axon scanners and software, new Gleams software).

If the two images are registered, the pixels values within each spot region may be viewed as a bivariate data set. Each spot region must be summarized into a single value quantifying the relative fluorescence intensity of each channel. Representative images and figures of two-color microarray data are shown below.

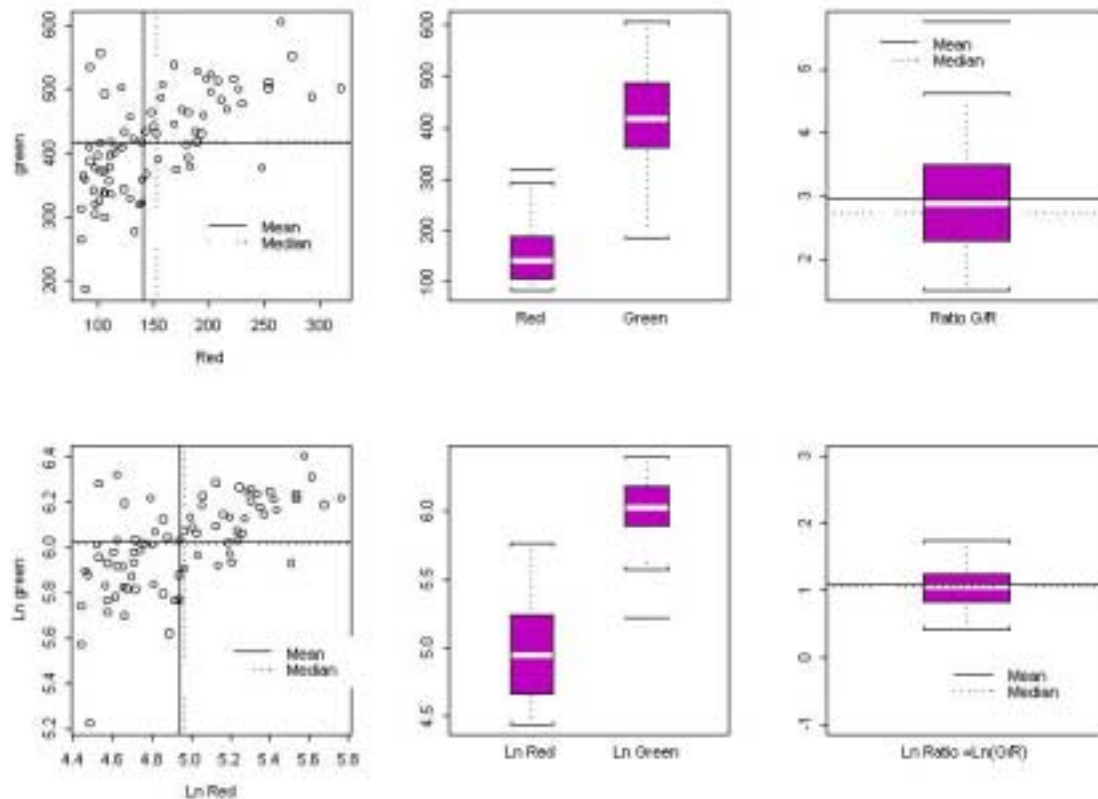


Image Analysis Issues

Image analysis continues to be an area of research for microarray data. The implementation of automated procedures for grid placement and spot finding offer hope to increase the ease of analysis of microarray data. However, the most reliable image analysis methods are still methods which require some human intervention to improve identification of spot regions.

Pixel values within spot regions show high dispersion. The source of variability and the statistical character of pixel distributions remain largely uninvestigated. In addition, systematic spatial effects must be accounted for during normalization of the quantitated data file. Data normalization is a distinct step from image processing.

References:

Y.H. Yang, M. J. Buckley, S. Dudoit and T.P.Speed. Comparison of methods for image analysis on cDNA microarray data. November 2000.

Y. H. Yang, S. Dudoit, P. Luu and T. P. Speed. Normalization for cDNA Microarray Data. December 2000.