Introduction to Microarray Image Processing

1/11/2011

Copyright © 2011 Dan Nettleton

1

A microarray scanner creates a digital image of a microarray.

- A *digital image* is a rectangular array of intensity values.

- Each intensity value corresponds to a *pixel*.

- The *color depth* of an image is the number of bits used to store the intensity value of one pixel.

- A color depth of 16 bits/pixel (common for microarray scanners) means that the intensity value of each pixel will be an integer between 0 and $65{,}535 = 2^{16} - 1$.

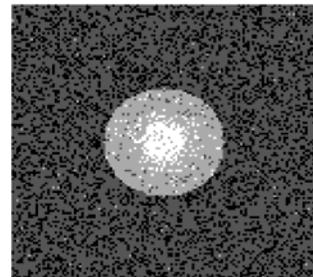- The number of pixels contained in the digital image is called the *resolution*.

2

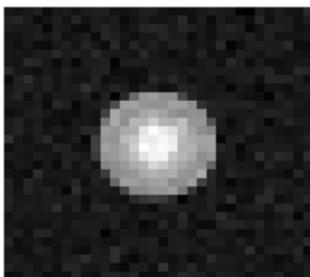Color Depth=6   Resolution=128 x 128
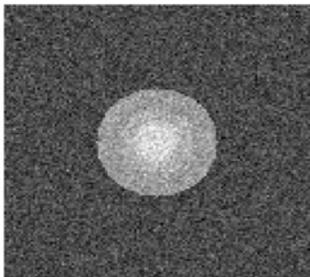
3

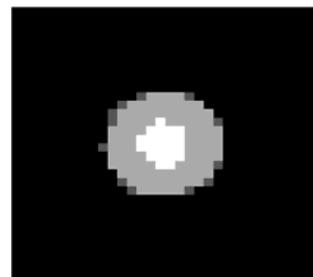Color Depth=2   Resolution=128 x 128

4

Color Depth=6   Resolution=32 x 32

5

Color Depth=2   Resolution=32 x 32

6

Image Processing for Spotted Arrays

Image processing for spotted arrays (cDNA and spotted oligo) can be divided into four basic steps:

1. Array localization – find the spots
2. Image segmentation – categorize each pixel as spot signal, background, or other
3. Quantification – assign signal and background values to each spot
4. Spot quality assessment – compute measures of spot quality for each spot

These steps are typically carried out with the aid of specialized software and can involve varying degrees of human input.

7

1. Array Localization

- Users may aid software by outlining grids and providing information about spot size and the number of rows and columns spotted on the slide.
- Using such information, software draws circles around each spot.
- Users may be able to make manual adjustments to improve upon automated spot identifications.

8

Software locates spots using info about grid.
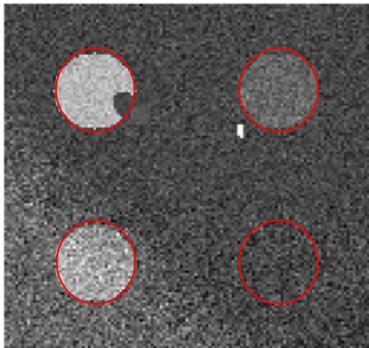
9

2. Image Segmentation

- There are a variety of proprietary commercial approaches for identifying each pixel as spot signal, background, or other.
- *Spot signal* or simply *signal* is fluorescence intensity due to target molecules hybridized to probe sequences contained in a spot (what we would like to measure) plus background fluorescence (what we would rather not measure).
- *Background* is fluorescence that may contribute to spot pixel intensities but is not due to fluorescence from target molecules hybridized to spot probe sequences.
- Background may be due to dust particles, stray fluorescent molecules, fluorescence in the slide itself, etc.
- Background will vary across the slide so most software packages attempt to measure local background by quantifying pixel intensities around each spot.
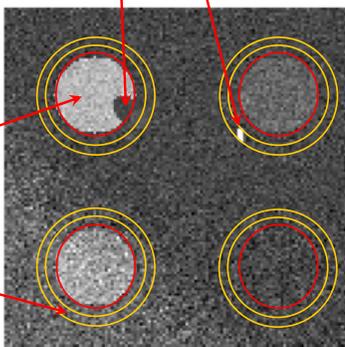
10

Artifacts can make signal and background segmentation challenging.

Software locates spots using info about grid.

Pixels in red circle may be segmented as signal.

Pixels between gold lines may be segmented as background.

Segmentation algorithms vary in complexity and effectiveness.

11

3. Quantification

Microarray imaging software will compute the following statistics for both signal and background using the segmented pixels for each spot:

1. mean – mean of pixel intensities
2. median – median of pixel intensities
3. mode – location of peak in histogram of intensities
4. area – number of pixels
5. total – sum of pixel intensities

12

Signal Histogram for One Spot

mean   median   mode
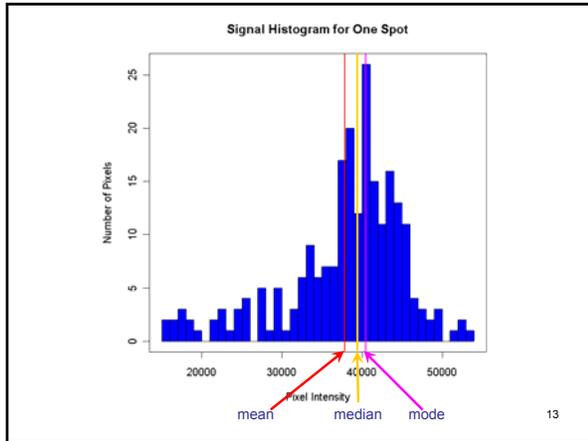
13

Spot Quality Assessment

The following are some spot quality statistics computed by some microarray imaging software.

- standard deviation – standard deviation of pixel intensities computed for both signal and background
- shape regularity – First signal area of a spot is inscribed into a circle. Then the number of non-signal pixels that fall within this circle is computed and divided by the circle's area. This ratio subtracted from 1 is defined as "shape regularity".
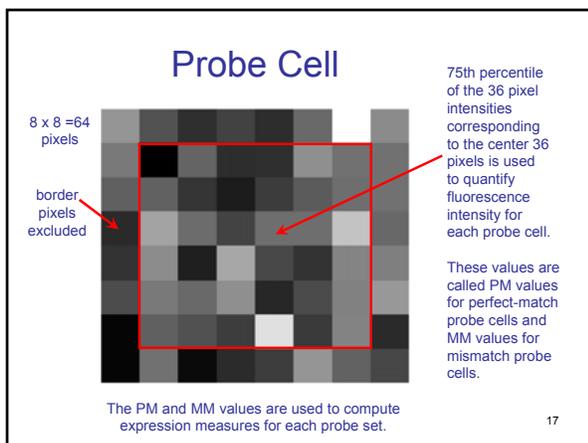
14

Spot Quality Measures (ctd.)

- area to perimeter = $\dfrac{(\text{spot area})*4\pi}{\text{perimeter}^2}$

  Ranges from 0 (highly non-circular shape) to 1 (a perfect circle).
- diameter – diameter of spot's grid circle in pixels
- saturation – indicates whether some pixels were censored at $2^{16}-1$
- signal contamination – indicates whether signal pixels were "contaminated" (contained outliers)
- background contamination – indicates whether background pixels were "contaminated"
- other measures involving spot location

15

Image Processing for Affymetrix GeneChips

- Image processing for Affymetrix GeneChips is typically done using proprietary Affymetrix software.
- The entire surface of a GeneChip is covered with square-shaped cells containing probes.
- Probes are synthesized on the chip in precise locations.
- Thus spot finding and image segmentation are not major issues.

16

Probe Cell

8 x 8 =64 pixels

border pixels excluded

75th percentile of the 36 pixel intensities corresponding to the center 36 pixels is used to quantify fluorescence intensity for each probe cell.

These values are called PM values for perfect-match probe cells and MM values for mismatch probe cells.

The PM and MM values are used to compute expression measures for each probe set.

17