

A novel method using edge detection for signal extraction from cDNA microarray image analysis

Jin Hyuk Kim^{1,3}, Hye Young Kim² and Yong Sung Lee¹

¹Institute of Mental Health, Hanyang University, 17 Hengdang-dong, Sungdong-ku, Seoul 133-791, Korea

²Department of Physiology, College of Medicine, Hanyang University, 17, Hengdang-dong, Sungdong-ku, Seoul 133-791, Korea

³Corresponding author: Tel, +82-2-2290-0614; Fax, +82-2-2281-3603; E-mail, jhkim1@hanyang.ac.kr

Accepted 15 June, 2001

Abbreviations: SAGE, serial analysis of gene expression; IMAGE, integrated molecular analysis of genomes and their expression; TIFF, tag-based image file format; SNR, signal to noise ratio; SD, standard deviation

Abstract

Gene expression analyses by probes of hybridization from mRNA to cDNA targets arrayed on membranes or activated glass surfaces have revolutionized the way of profiling mega level gene expression. The main remaining problems however are sensitivity of detection, reproducibility and data processing. During processing of microarray images, especially irregularities of spot position and shape could generate significant errors: small regions of signal spots can be mis-included into background area and vice versa. Here we report a novel method to eliminate such obstacles by sensing their edges. Application of edge detection technology on separating spots from the background decreases the probability of the errors and gives more accurate information about the states of spots such as the pixel number, degree of fragmentation, width and height of spot, and circumference of spot. Such information can be used for the quality control of cDNA microarray experiments and filtering of low quality spots. We analyzed the cDNA microarray image that contains 10,368 genes using edge detection and compared the result with that of conventional method which draws circle around the spot.

Keywords: cDNA microarray, image analysis, edge detection, segmentation of spot

Introduction

As the human genome project has nearly been completed, the attention is being focused on understanding the gene expression profiling of disease state of cells and tissues as well as the development of platform technology or methodology for detecting and quantitating gene expression levels utilizing Northern blots, S1 nuclease protection, differential display, sequencing of cDNA libraries, and SAGE analysis. These methods are augmented by two microarray-based technologies – *i.e.* cDNA and oligonucleotide arrays (Duggan *et al.*, 1999). In the case of cDNA microarray, cDNA amplified from IMAGE, cDNA clone set, or a custom cDNA library are usually spotted on the glass at high density. Arrays of cDNA spots in small areas (more than 10,000 cDNAs on a 2 × 2 glass slide) are now commonly referred to as microarrays (Ermolaeva *et al.*, 1998).

Glass slide has many advantages over other materials (Cheung *et al.*, 1998; Cheung *et al.*, 1999). Especially, as a consequence of its low fluorescence, it does not significantly contribute to background noise. And also it is a durable material that sustains high temperatures and washes of high ionic strength. It is important to choose a weak background noise as a support of microarray chip, because a strong background noise increases the SNR and consequently it is hard to isolate the spots from the noise.

To generate high-density gridded microarrays, spotting with precisely controlled robot is essential. During the spotting, rotation of pin and vibrations cause the perturbation of positioning spots and the irregularities of spot shape. Several factors such as the hydrophobicity of glass surface, humidity during drying, and speed of drying induce unequal distribution of cDNA in the spot. All of the above and unknown causes contribute to the displacement of spot position and irregular shape. Commonly used method to detect the spot makes a circle around it and calculates various statistical parameters from the intensity of digital image pixels inside the circle. However, the deviation of spot position and distorted shape of spots often give rise to not only an inclusion of background pixels in circle but also an exclusion of spot pixels.

In this report we describe the development and application of a spot isolation technique for microarray image analysis. This technique is able to isolate the spot from the background and all statistical parameters generated with it are more accurate and sophisticated than the conventional circle technique.

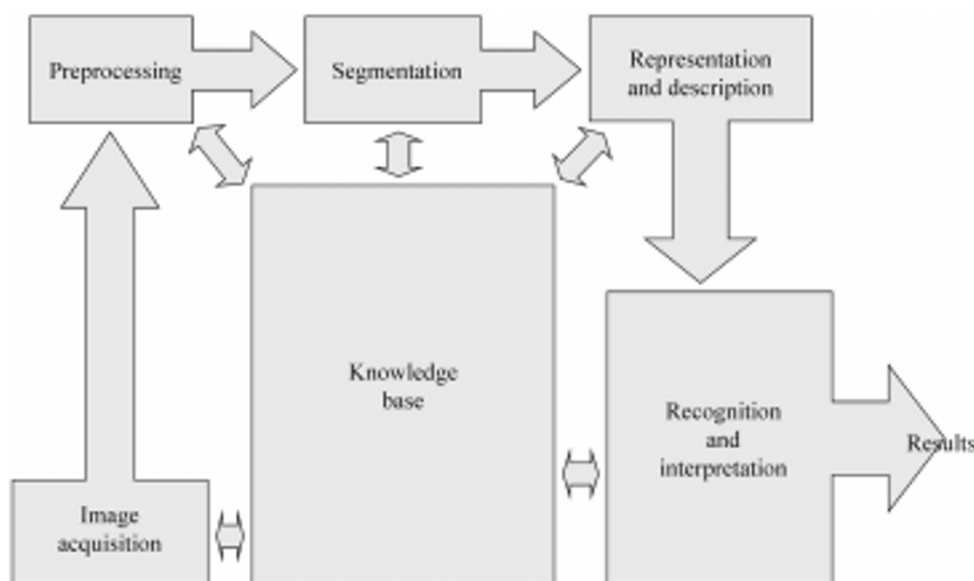


Figure 1. Fundamental steps in digital image processing.

Materials and Methods

Image analysis

Microarray images are typically captured as unsigned 16-bits TIFF formatted files with scanner (GMS418[®], Genetic Microsystems). Usually all images captured were stored in fixed disk. All image analysis processes were performed on the Windows 2000[®] platforms. Fundamental steps in digital image processing consist of several steps that include image acquisition, preprocessing of image, segmentation, representation and description, and recognition and interpretation (Figure 1) (Gonzalez and Woods 1993). Every step in general image processing is closely related to the information of the image to be analyzed and knowledge of general digital image. In the cDNA microarray image analysis, the concept of image process is based on the above-mentioned algorithm. Therefore, our package consists of four modules, which are (i) image file retrieve and preprocessing, (ii) segmentation of image by the position of spot, (iii) detection of spot and calculation of statistical parameters and (iv) normalization.

Image file retrieve and preprocessing

Hybridized microarray chips were scanned twice for Cy3 and Cy5 signals. It produced two digital image files that represent the control and experimental group. To compare the gene expression profile between control and experimental group, the image analyzer should retrieve the Cy3 and Cy5 unsigned 16 bits TIFF formatted image files. It is impossible to scan Cy3 and Cy5 signals simultaneously with any kind of scanners. Successive scanning of two signals frequently causes a little differ-

ence of position between two images. Therefore, preprocessing of image must include the conversion of 16 bits Cy3 and Cy5 images to overlapped RGB or 8 bits image after repositioning of Cy3 or Cy5 image.

Segmentation of image by the positions of spots

In order to generate the useful statistical value about gene expression from the cDNA microarray images, it is essential to segment the image around spots as centers. Since embossed cDNA spots were printed automatically to the predefined positions, it was assumed that the arrangement of probe signals forms a lattice. Therefore, it seems reasonable that the predefined grid was overlaid on the image. However, during spotting and hybridization, various causes deviate the center of spot from the predefined position. It must be corrected manually by tilting the predefined grid in horizontal and/or vertical directions.

Detection of spot and calculation of statistical parameters

Signal representing gene expression level should be isolated from each segmented region. Usually, to spot the cDNA on a slide glass coated with various materials, a pen filled with viscous cDNA solution is used. As the end of pen is tapered, the diameter of round tip is less than 100 μm . Therefore, the diameter of spot is about 100 μm and its shape is round but the viscosity of cDNA solution, vibration of spotting machine, drying out of spot from the edge to the center etc. practically deform the spot. In order to produce accurate results, it is important to detect the spot precisely. Generally, an edge means a significant change in the grayscale values between

pixels in an image. It is necessary to increase the contrast for detecting the edge of vague spot in microarray image. Passing through the proper kernel increases it, and it must be filtered with Gaussian smoothing filter to estimate the location of edges even under conditions of poor SNR images. After detecting an edge of the spot from the modified image, inner and outer edges are stored as the spot and background template in the memory respectively. Now there are several images from original 16 bits TIFF microarray image: 8 bits image for segmentation and edge detection, RGB image for presentation to users, and black and white spot and background template images. It is easy to get the data of gene expression levels from 16 bits image with information about segmentation, and spot and background templates.

Normalization

Usually signal intensities of Cy3 and Cy5 images are not equal because of the difference between the Cy3 and

Cy5 sensitivity to fluorescence, variation of PMT tube gain, and laser strength. The mean of ratios of Cy5 to Cy3 signals is deviated from 1.0. Therefore, it is necessary to normalize the signal intensities. When the normalization is performed, one of Cy5 or Cy3 intensity is fixed and the other is normalized. For an example, a normalized intensity of Cy5 is calculated as

$$I_{Cy5\text{-normalized}}^k = I_{Cy5}^k \cdot 10^{n/\sum \log r_i}$$

where $I_{Cy5\text{-normalized}}^k$ and I_{Cy5}^k is the Cy5 intensity of k th spot after and before the normalization, and n and r_i are the number of data and the ratio of Cy5 to Cy3 intensity at each spot respectively.

Results

To test the cDNA microarray image analyzer, human microarray chips and human gastric carcinoma and

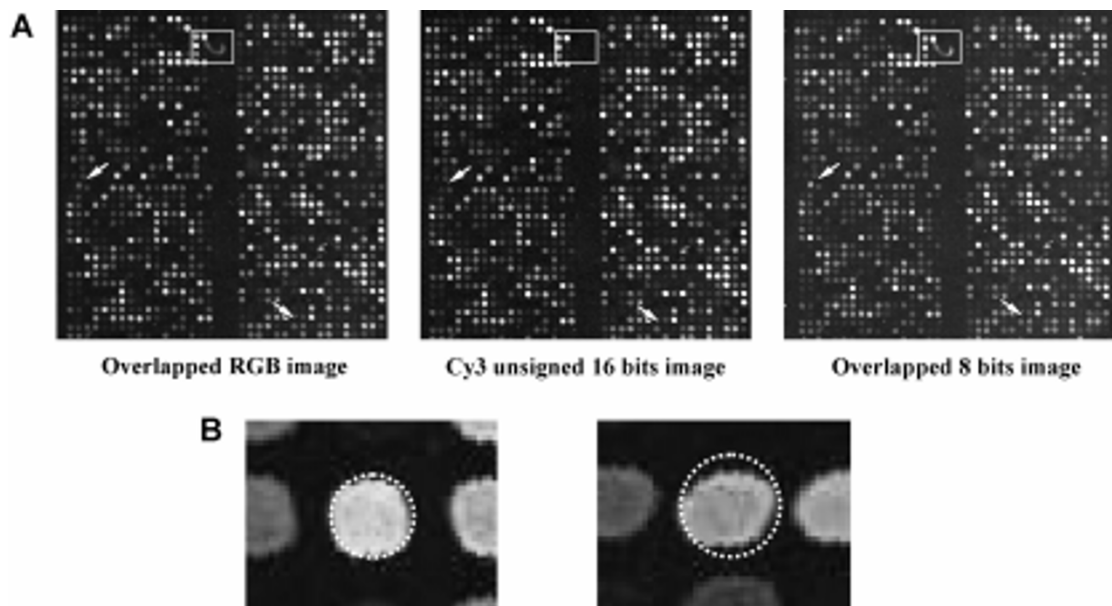


Figure 2. Various images of cDNA microarray and spots. (A) During image analysis, several images are generated by image processing package. Red spots and particles in RGB image indicated by arrows and box cannot be seen in Cy3 16 bits image. In the overlapped 8 bits image, the intensities of pixels are the averages of Cy3 and Cy5 pixel intensities. (B) Various shapes of spot were shown. In the right, there is a little discrepancy between elliptical spot and circle.

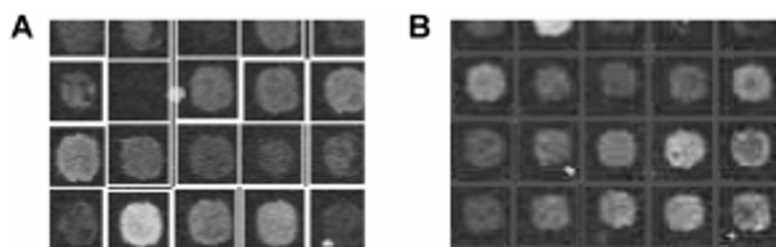


Figure 3. Results of segmentation. (A) Automated segmentation is a convenient tool but bright blob deviates frame of segment. (B) Manual segmentation shows the regular arrangement of segments.

placenta RNA were used. Both parts of 16 bits unsigned TIFF Cy3 microarray image and Cy3-Cy5 overlapped and converted RGB image are shown in Figure 2. Red colored spot in the RGB image is vague in the Cy3 16 bits image. Various conditions during spotting and hybridization caused the shape and position of spots irregular. Even in the same chip, the shape of spot varied (Figure 2) so that circle around the spot sometimes included the background pixels inside the circle.

To isolate the spot from surrounding ones, the auto-

matic and manual methods for segmentation were applied and their results were shown in Figure 3. Automated segmentation has many advantages but bright blobs in the background pull the frame of segment toward themselves. It is important to remove the blobs from the image. Manual segmentation needs more efforts than the other and it is hard to locate the spot on the center of the segment when the manual segmentation is applied. In a segment, intensities of the pixels were shown in the 3D graph of Figure 4. A large number of spots in the

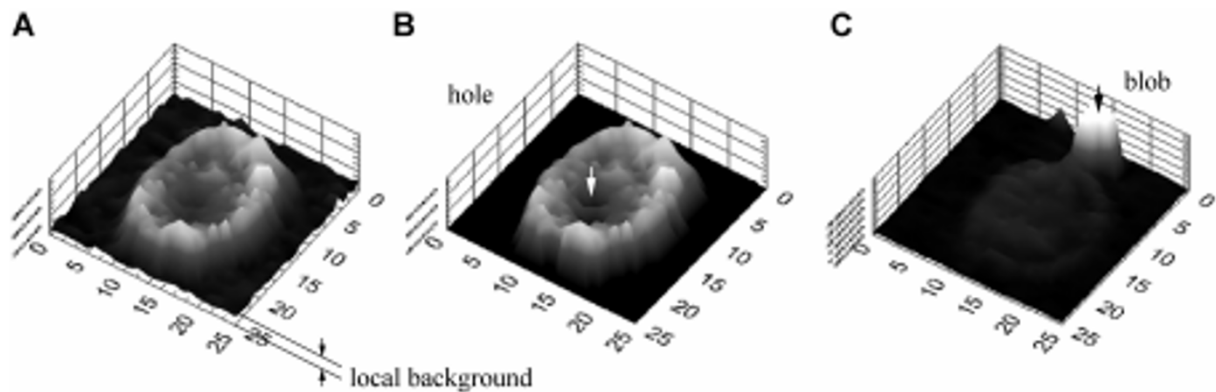


Figure 4. Three dimensional graph of a segment. (A) Typical annular shape spot and its local background is shown. (B) 3D graph of the same segment that was subtracted by local background mean shows two holes (indicated by white arrow) in the center of the spot. (C) Saturated blob (indicated by black arrow) is shown in the 3D graph of another segment.

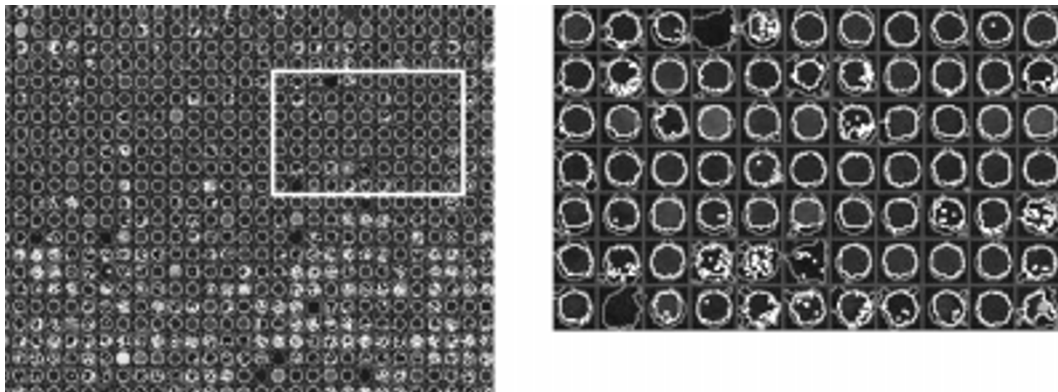


Figure 5. Result of edge detection. Blue line is the frame of segment and white and red lines are edges of spots and local backgrounds. Shapes and areas of spot are various and some spots have several holes or are fragmented.

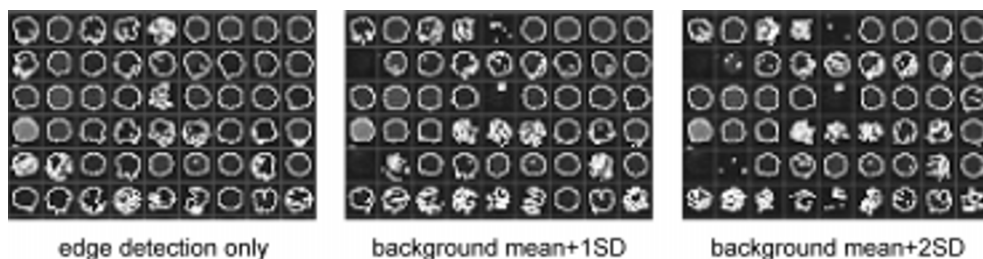


Figure 6. Changes of shape, size, and fragmentation of spots are shown as the thresholds increase. Some spots are disappeared.

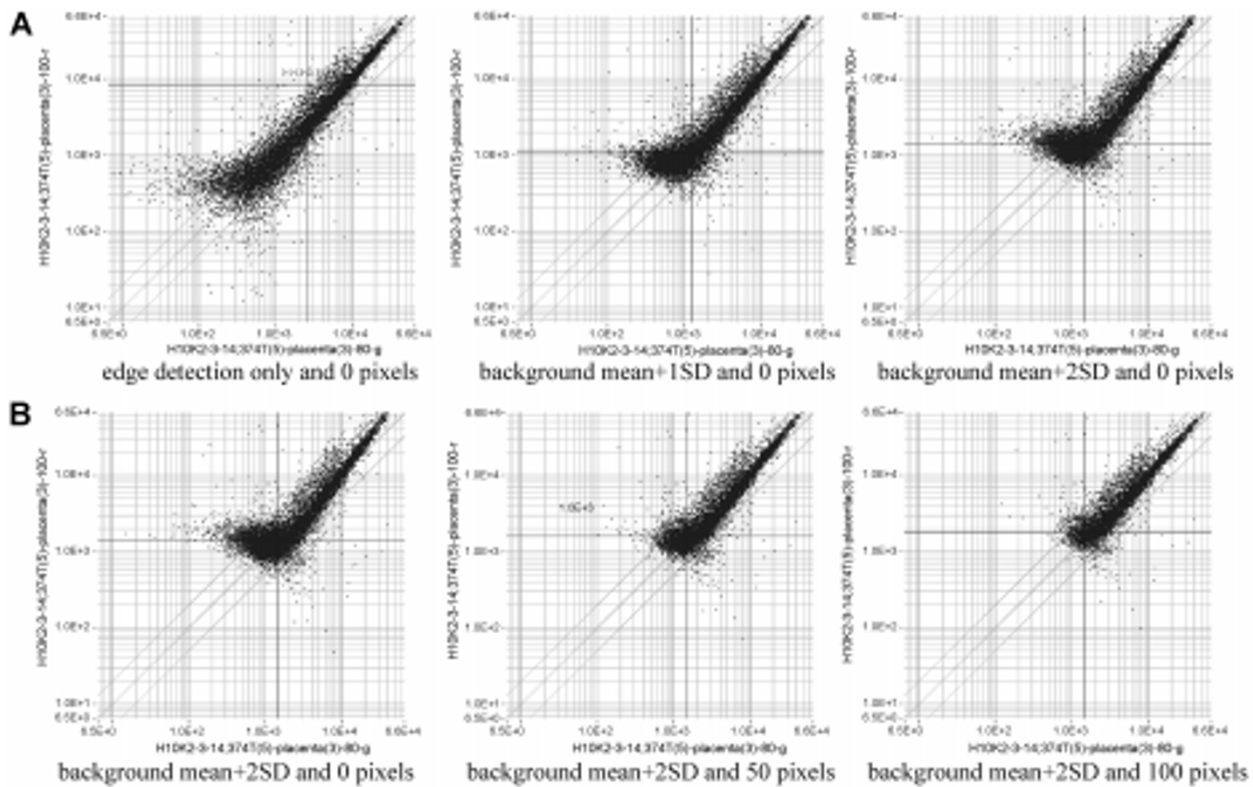


Figure 7. Normalized scatter plot of cDNA microarray. (A) Increments of threshold for the pixel intensities move the points in the lower and left region of scatter plot upward and rightward. (B) Application of spot area filter decreases the number of significant points in the scatter plot.

cDNA microarray image show annular shapes and have one or more holes in the center of them. Nonspecific bright blobs in the microarray image were shown in a few segments. Their brightness is saturated in most cases and therefore their mean pixel values are almost 65,535. Even a small bright blob in the segment can increase the mean value of spot or background.

In our package, the separation of signal from the background could be achieved with an edge detection technology. The results were shown in Figure 5. In the case of vague spot, there are no or a few clusters of pixels inside the edges and sometimes two or more fragments appear in the spot region. However, a clear spot has a large fragment in the central region of the segment and its line of edge is smooth and round. Edge detection module used in our cDNA microarray image analyzer has a function that obscures pixels in the neutral area. Such pixels were shown between the white and red lines in Figure 5.

To determine the validity of pixels inside the edge of the spot, an image analyzer verifies whether the intensity value of pixel is larger than the background mean + $x \cdot SD$, where x is an input by user. As x increases from 0 to 2, increments of both fragmented spot and empty segment were demonstrated in Figure 6. In the scatter plot, an increased threshold (x value) pushes the point

to the upper and right area of plot (Figure 7A) and also the pixel number filter, which determines whether the effective spot area is larger than threshold or not, removes small spots (Figure 7B).

Discussion

We have developed an efficient image-processing package for the use with cDNA microarrays, and in this report we demonstrated the new technology of microarray image analysis. Our package has several distinctive features. One of the most representative features is an edge detection technology and the others are automatic segmentation and spot area (pixel number) filter.

Generally, there are many chances to make errors during the manufacture of cDNA microarray chip, hybridization of mRNA extracted from the sample, and scanning of chips. These facts demand the compensation of defects to the image processing procedures. However, a small inaccuracy of image processor has been ignored because of the confusion of the error sources. In the future, the advance of chip technology should contribute to the precise analysis of the microarray image.

The application of circle method on the elliptical or irregular spot tends to include the background pixel

inside the circle. It can distort gene expression profiles to some degree. Many packages using circle method have processes that prevent mis-inclusion of spots or background pixels. It can be achieved by discarding pixels whose intensities are in the upper or lower 15%. Usually the user can change a percentage but it is impossible to determine an accurate percentage for removing background pixels from the circle and vice versa. Therefore, either a removal of normal pixels or an inclusion of background pixels can generate mistakes. Especially, bright blobs certainly induce the mis-inclusion of pixels in a large extent. However, an edge detection technology enables to recognize the accurate margin of spot (Kuklin *et al.*, 2000).

Another advantage of our package is the generation of data that cannot be produced by others. They include the number of fragmentation, number of holes in the spot, and various parameters of each fragment or hole. However, a further study is needed to define the meaning of them in the cDNA microarray image analysis.

Acknowledgement

This work was financially supported by '21C Frontier, The Center for Functional Analysis of Human Genome'

project.

References

- Cheung VG, Gregg JP, Gogolin-Ewens KJ, Bandong J, Stanley CA, Baker L, Higgins MJ, Nowak NJ, Shows TB, Ewens WJ, Nelson SF, Spielman RS. Linkage-disequilibrium mapping without genotyping. *Nat Genet* 1998;18:225-30
- Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G. Making and reading microarrays. *Nat Genet* 1999; 21(1 Suppl):15-19
- Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nat Genet* 1999; 21(1 Suppl):10-14
- Ermolaeva O, Rastogi M, Pruitt KD, Schuler GD, Bittner ML, Chen Y, Simon R, Meltzer P, Trent JM, Boguski MS. Data management and analysis for gene expression arrays. *Nat Genet* 1998;20:19-23
- Gonzalez RC, Woods RE. Fundamental steps in image processing. In *Digital image* (Gonzalez, R.C. and Woods, R.E., eds.), 1993, pp. 7-10, Addison-Wesley, New York
- Kuklin A, Shams S, Shah S. High throughput screening of gene expression signatures. *Genetica* 2000;108:41-46