

Microarray imaging data pre-processing

R.Rizzi, M.Falchi, L.Milanesi

CNR-ITBA, Via Fratelli Cervi, 93, Segrate (Milano)

The new technique of microarray gene expression assay in the field of molecular biology and genetics, is a highly parallel method for investigating the expression behavior of a great number of genes in a living organism.

Gene sequences are arranged onto small membranes of different materials according to a grid scheme. The gene panel is then submitted to hybridization reaction with messenger RNA which is extracted from one or more cellular populations to be studied, and labeled with fluorescent substances. The produced signal is finally read by means of a scanning device.

The resulting digital image presents a pattern composed of spots with signal more or less intense. The spot band intensity is directly proportional to the expression level of the gene in the cellular population attached to the band. Perfect conditions should provide images containing only the expression signal. Some signal disturb actually distorts the pattern.

The peculiar signal pattern drives to read the image through regular grids of interest areas arranged as rows and columns. If the grid is correctly positioned over the image, one pattern spot corresponds to one grid element. The grid can be approximately positioned on the microarray image. An automatic adjustment of the preliminary positioning can then apply an iterative affine transformation in order to lead the grid to its optimal position on the signal. To accomplish this task, a program 'SpotView' has been developed (entirely in JAVA for platform portability), that produces data files in standard formats suitable to be used with relational databases.

Each grid element is composed by an area of interest for the expression spot and an area and for the surrounding background, so that an integration value and a value of local noise for each signal band are obtained. The calculation of the local background is given by multiplying the mean intensity value in the background area, by the interest area for the expression spot. Disturbs of neighboring light spots can be avoided by using the median of the neighborhood's background values. The correct spot value and the background subtraction are very important for the final quantification of the microarray image. Furthermore a quality spot measure can be given on the basis of some spot statistics: standard deviation of the intensity and saturation percentage and percentage of pixels with intensity greater than or equal to one standard deviation above the median intensity in the interest area of expression.