Chapter XIX

Image Processing and Machine Learning Techniques for the Segmentation of cDNA Microarray Images

Nikolaos Giannakeas
University of Ioannina, Greece

Dimitrios I. Fotiadis
University of Ioannina, Greece

ABSTRACT

Microarray technology allows the comprehensive measurement of the expression level of many genes simultaneously on a common substrate. Typical applications of microarrays include the quantification of expression profiles of a system under different experimental conditions, or expression profile comparisons of two systems for one or more conditions. Microarray image analysis is a crucial step in the analysis of microarray data. In this chapter an extensive overview of the segmentation of the microarray image is presented. Methods already presented in the literature are classified into two main categories: methods which are based on image processing techniques and those which are based on Machine learning techniques. A novel classification-based application for the segmentation is also presented to demonstrate efficiency.

INTRODUCTION

Several types of microarrays have been developed to address different biological processes: (i) cDNA microarrays (Eisen, 1999) are used for the monitoring of the gene expression levels to study the effects of certain treatments, diseases, and developmental stages on gene expression. As a result, microarray gene expression profiling can be used to identify disease genes by com-
paring gene expression in diseased and normal cells. (ii) Comparative genomic hybridization application assesses genome content in different cells or closely related organisms (Pollack et al., 1999). (iii) SNP detection arrays identify single nucleotide polymorphism among alleles within or between populations (Moran & Whitney, 2004). (iv) Finally, Chromatin immunoprecipitation (ChIP) technologies determine protein binding site occupancy throughout the genome, employing ChIP-on-chip technology (Buck & Lieb, 2004).

The experiment of cDNA microarrays typically starts by taking two biological tissues and extracting their mRNA. The mRNA samples are reverse transcribed into complementary DNA (cDNA) and labelled with fluorescent dyes resulting in a fluorescence-tagged cDNA. The most common dyes for tagging cDNA are the red fluorescent dye Cy5 (emission from 630-660 nm) and the green-fluorescent dye Cy3 (emission from 510-550 nm). Next, the tagged cDNA copy, called the sample probe, is hybridized on a slide containing a grid or array of single-stranded cDNAs called probes. Probes are usually known genes of interest which were printed on a glass microscope slide by a robotic arrayer. According to the hybridization principles, a sample probe will only hybridize with its complementary probe. The probe-sample hybridization process on a microarray typically occurs after several hours. All unhybridized sample probes are then washed off and the microarray is scanned twice, at different wavelengths corresponding to the different dyes used in the assay. The digital image scanner records the intensity level at each grid location producing two grayscale images. The intensity level is correlated with the absolute amount of RNA in the original sample, and thus, the expression level of the gene associated with this RNA.

Automated quantification of gene expression levels is realized analyzing the microarray images. Microarray images contain several blocks (or subgrids) which consist of a number of spots, placed in rows and columns (Fig. 1). The level of intensity of each spot represents the amount of sample which is hybridized with the corresponding gene. The processing of microarray images (Schena et al., 1995) includes three stages: initially, spots and blocks are preliminarily located from the images (gridding). Second, using the available gridding information, each microarray spot is individually segmented into foreground and background. Finally, intensity extraction,
Related Content

Assistive Technology for Individuals with Disabilities
www.igi-global.com/chapter/assistive-technology-individuals-disabilities/53664?camid=4v1a

Nurses and Telehealth: Current Practice and Future Trends
www.igi-global.com/chapter/nurses-telehealth-current-practice-future/27325?camid=4v1a

Extracting Data from the National Inpatient Sample
Patricia Cerrito and John Cerrito (2010). Clinical Data Mining for Physician Decision Making and Investigating Health Outcomes: Methods for Prediction and Analysis (pp. 69-93).
www.igi-global.com/chapter/extracting-data-national-inpatient-sample/44267?camid=4v1a

An Introduction to Women's Health and Informatics
Peter Stone (2009). Medical Informatics in Obstetrics and Gynecology (pp. 1-12).
www.igi-global.com/chapter/introduction-women-health-informatics/26181?camid=4v1a