Chapter XII
Uncovering Fine Structure in Gene Expression Profile by Maximum Entropy Modeling of cDNA Microarray Images and Kernel Density Methods

George Sakellaropoulos
University of Patras, Greece

Antonis Daskalakis
University of Patras, Greece

George Nikiforidis
University of Patras, Greece

Christos Argyropoulos
University of Pittsburgh Medical Center, USA

ABSTRACT

The presentation and interpretation of microarray-based genome-wide gene expression profiles as complex biological entities are considered to be problematic due to their featureless, dense nature. Furthermore microarray images are characterized by significant background noise, but the effects of the latter on the holistic interpretation of gene expression profiles remains under-explored. We hypothesize that a framework combining (a) Bayesian methodology for background adjustment in microarray images with (b) model-free modeling tools, may serve the dual purpose of data and model reduction, exposing hitherto hidden features of gene expression profiles. Within the proposed framework, microarray image restoration and noise adjustment is facilitated by a class of prior Maximum Entropy distributions. The resulting gene expression profiles are non-parametrically modeled by kernel density methods, which not only normalize the data, but facilitate the generation of reduced mathematical descriptions of biological variability as mixture models.

INTRODUCTION

The advent of complementary DNA (cDNA) microarray technologies enabled the simultaneous and specific assessment of the expression levels of thousands of genes (Southern, Mir, & Shchepinov, 1999). The conventional approach to analyze such datasets is to explore quantitative co-expression relations across a variety of experimental conditions prior to invoking putative similarities in gene regulation or function (DeRisi, Iyer, & Brown, 1997; Eisen, Spellman, Brown, & Botstein, 1998). The alternative viewpoint considers gene expression profiles from specific conditions to be informative of distinct molecular signatures that characterize cellular states. Such genome wide, transcriptional signatures have been used to distinguish normal from abnormal samples in benign developmental conditions (Barnes et al., 2005), solid tumors and hematologic malignancies (Febbo et al., 2005; Valentini, 2002) and differentiate distinct disease states of renal allografts (Sarwal et al., 2003). It has been suggested that the thousands of expression values in a microarray experiment are too dense and irregular to be directly interpreted in a holistic manner and that alternative transformations of the normalized gene profiles should be sought after (Guo, Eichler, Feng, Ingber, & Huang, 2006). Nevertheless one could justifiably argue that the irregularity of the gene profiles is due to incomplete modeling and adjustment for the presence of measurement noise. However this alternative hypothesis has not been adequately addressed in the current literature. These considerations underline the impetus for the present work, which aims to:

1. Establish the role of microarray image background in the irregularity and "featureless" appearance of gene expression profiles (GEP) from individual experimental states.
2. Propose a data and model reduction framework for the analysis of GEP consisting of:
    a. A probabilistic Bayesian algorithm for background adjustment of microarray images based on Maximum Entropy distributions.
    b. Non-parametric kernel density estimation methods for the mathematical representation and exploration of the resultant gene expression profiles.

BACKGROUND

The basic microarray procedure involves hybridization of complementary nucleic acid molecules, one of which (target) has been immobilized in a solid substrate (e.g. glass) using a robotically controlled device (arrayer). Such targets form spots at the vertices of a rectangular lattice on the solid substrate surface; each spot then serves as a highly specific and sensitive detector of the corresponding gene. Technical factors operating at different stages of the microarray pipeline impart the final microarray image with uneven and non-negligible background. Background correction has been considered a necessary step in the microarray pipeline (L. Qin, Rueda, Ali, & Ngom, 2005), which may in fact have a significantly higher bearing on the final results than normalization. This viewpoint is supported by a number of empirical studies showing that conventional approaches to background correction may lead to nonsensical results (e.g. negative gene expression measures) and hinder the ability to detect differential gene expression (L. X. Qin & Kerr, 2004). In spite of the nodal role of background adjustment methods upon the quantification of single gene expression, the effects of de-noising upon the totality of the gene profile remains unexplored. A notable exception is the study by (Kooperberg, Fazzio, Delrow, & Tsukiyama, 2002); the authors argued that incomplete background adjustment is expected to predominantly affect