

A Robust Unified Approach to Analyzing Methylation and Gene Expression Data

Shili Lin¹ and Abbas Khalili¹

¹ Department of Statistics, The Ohio State University, USA

Microarray technology has made it possible to investigate expression levels, and more recently methylation signatures, of thousands of genes simultaneously, in a biological sample. Since more and more data from different biological systems or technological platforms are being generated at an incredible rate, there is an increasing need to develop statistical methods that are applicable to multiple data types and platforms. Motivated by such a need, we propose a flexible finite mixture model that is applicable to methylation, gene expression, and potentially data from other biological systems. Two major thrusts of this approach are to allow for a variable number of components in the mixture to capture non-biological variation and small biases, and to use a robust procedure for parameter estimation and probe classification. We applied the method to the analysis of methylation signatures of three breast cancer cell lines. The method was also tested on three sets of expression microarray data to study its power and type I error rates. Comparison with a number of existing methods in the literature yielded very encouraging results; lower type I error rates and comparable/better power were achieved based on our limited study. Furthermore, our method also leads to more biologically interpretable results for the three breast cancer cell lines.