

Ratio Statistics of Gene Expression Levels and Applications to Microarray Data Analysis

Yidong Chen¹, Vishnu Kamat², Edward R. Dougherty^{2*}, Michael L. Bittner¹, Paul S. Meltzer¹, and Jeffery M. Trent¹

¹Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Building 50, Room 5154, 50 South Drive, MSC 8000, Bethesda, MD 20892

²Department of Electrical Engineering, Texas A&M University, 3128 TAMU, College Station, Texas 77843

Abstract

Motivation: Expression-based analysis for large families of genes has recently become possible owing to the development of cDNA microarrays, which allow simultaneous measurement of transcript levels for thousands of genes. For each spot on a microarray, signals in two channels must be extracted from their backgrounds. This requires algorithms to extract signals arising from tagged mRNA hybridized to arrayed cDNA locations and algorithms to determine the significance of signal ratios.

Results: This paper focuses on estimation of signal ratios from the two channels, and the significance of those ratios. The key issue is the determination of whether a ratio is significantly high or low in order to conclude whether the gene is up-regulated or down-regulated. The paper builds on an earlier study that involved a hypothesis test based on a ratio statistic under the supposition that the measured fluorescent intensities subsequent to image processing can be assumed to reflect the signal intensities. Here, a refined hypothesis test is considered in which the measured intensities forming the ratio are assumed to be combinations of signal and background. The new method involves a signal-to-noise ratio, and for a high signal-to-noise ratio the new test reduces (with close approximation) to the original test. The effect of low signal-to-noise ratio on the ratio statistics constitutes the main theme of the paper. Finally, and in this vein, a quality metric is formulated for spots. This measure can be used to decide whether or not a spot ratio should be deleted, or to adjust various measurements to reflect confidence in the quality of the measurement.

Key words: genomics, image processing, microarray, ratio statistics