

Improvements to the RMA Algorithm for Gene Expression Microarray Analysis

Monnie McGee & Zhongxue Chen

Department of Statistical Science

Southern Methodist University

Abstract

Gene expression microarrays (or "gene chips") are devices which allow the response of thousands of genes to various stimulants to be assessed all at once. There are typically two reasons why microarrays are done as part of an experiment: to determine which genes are highly differentially expressed in response to a stimulus, and to determine which genes behave similarly. However, before one can do statistical inference or clustering, one must have clean, reliable data. The data coming from microarray chips are extremely noisy. In addition, the number of replicates of a microarray is typically an order of magnitude (or two) smaller than the number of genes that reside on the microarray. Both of these realities present huge statistical challenges.

In this talk, I describe the composition of both two-channel and single-channel chips, and I describe Robust Multichip Average (RMA), a very popular method for background correction, normalization, and summarization of probe-level gene chip data to obtain a measure of gene expression. Then, I show simulation results indicating deficiencies in RMA. I propose several simple ways to correct for these problems, and give strengths and weaknesses of each. The new algorithms perform better than RMA under most circumstances.