

Interpreting microarray data: Cluster analysis and statistical detection of differential gene expression

David Bickel, Ph.D.
Medical College of Georgia

Abstract:

Different methods of analyzing microarray data achieve different goals. Cluster analysis can classify genes by their expression patterns, whereas hypothesis testing, Bayesian probability computation, and decision-theoretic optimization can determine which genes appear to have differences in expression between groups of patients.

A robust graphical method of summarizing the results of cluster analysis and a biological method of determining the number of clusters are presented. These methods are applied to the data of DeRisi et al. (1997), showing that rank-based methods perform better than log-based methods.

In addition, a decision-theoretic method of detecting differential gene expression is presented. As a measure of error in testing multiple hypotheses, the decisive false discovery rate (dFDR), the ratio of the expected number of false discoveries to the expected total number of discoveries, has advantages over the false discovery rate (FDR) and positive FDR (pFDR). The dFDR can be optimized and often controlled using decision theory, and some previous estimators of the FDR can estimate the dFDR without assuming weak dependence or the randomness of hypothesis truth values. While it is suitable in frequentist analyses, the dFDR is also exactly equal to a posterior probability under the assumption of random truth values, even without independence.