

Improved Statistical Tests for Differential Gene Expression by Shrinking the Estimates of Variance Components

Xiangqin Cui, Ph.D.
The Jackson Laboratory, Bar Harbor, Maine

Abstract:

The large amount of data generated by microarray technology is due mainly to the large number of genes represented on the array. For each gene the number of RNA samples assayed is typically small. Therefore, testing for differential expression one gene at a time often has low power and combining information across genes is desirable. Here we develop an estimator of the error variance that can borrow information across genes using the James-Stein shrinkage concept. A new test statistic FS is constructed using this estimator. The new statistic is compared with some other statistics used to test for differential expression, namely the gene-specific F test (F1), the pooled-variance F statistic (F3), and a hybrid statistic (F2) that uses the average of the individual and pooled variances. The FS test shows best or nearly best power for detecting differentially expressed genes over a wide range of simulated data in which the variance components associated with individual genes are either homogeneous or heterogeneous. Comparisons between FS and other types of information-borrowing statistics will also be discussed.