Supplementary Materials to

Computational Enhancement of a Shrinkage-Based ANOVA F-test Proposed for Differential Gene Expression Analysis

Stanley B. Pounds, Ph.D.
Assistant Member
Department of Biostatistics, MS 768
St. Jude Children's Research Hospital
Memphis, TN 38105

stanley.pounds@stjude.org
Phone: 901-495-5052
Fax: 901-544-8843

This document describes how to obtain and use the R routines described in the article.

Terms of Use

The routines are free to use provided that Pounds (2007) and Cui et al. (2005) are cited in all resulting presentations and publications.

Obtaining and Using the Routines

First, download the R code file from www.stjuderesearch.org/depts/biostats/shrinkage and use the `source` command in R to make the routines available for use.

The library defines the functions `array.cui.Fstat`, `cui.mse`, `cui.B`, and `cui.V`. The `array.cui.Fstat` function will be of primary interest to most users. It accepts a vector `grps` and a matrix `Y` of (log-transformed) expression values as arguments and returns a vector of F-statistics. The other functions are used by `array.cui.Fstat`. The `cui.mse` routine computes the shrinkage-based mean squared error term described by Cui et al. (2005). The constants *B* and *V* described by Cui et al (2005) are computed by `cui.B` and `cui.V`, respectively.

Next, the data should be prepared by storing the expression data in a matrix `Y` and the group labels in a vector `grps`. Each row of `Y` should have the (log-transformed) expression values for a specific gene (or micrroary probe set) and each column corresponds to one subject. Each entry of `grps` should give the group label for the corresponding column of `Y`.

Once the routines have been made available using the source command and the data have been stored in this fashion, the calculations can be performed by issuing the command below.

```
res <- array.cui.Fstat(grps,Y)
```

This command passes `grps` and `Y` as arguments to the routine `array.cui.Fstat`. The `array.cui.Fstat` computes the F-statistic for each gene and returns the result as a vector. The result vector is assigned to the object `res`. Each entry of `res` gives the F-statistic for the gene of the corresponding row of `Y`.

Works Cited

Cui, X. et al. (2005) Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics* 6: 59-75.