PRSG_Nirvana — A parallel tool to assess differential gene expression

David Henderson

PRSG_Nirvana is a statistical package to assess statistical significance in differential gene expression data sets. The basis of the method is a two stage mixed linear model approach popularized in a fairly recent publication by R. Wolfinger. The package takes advantage of the SNOW and Rmpi packages to create a parallel computing environment on common commodity computing clusters. Without the added computing power of cluster computing, bootstrapping of residuals to utilize the empirical distribution of a test statistic would not be practical. One would then have to rely on asymptotic thresholds for less than large data sets as inference is based upon gene specific observations. Additional features of the package are a custom version of the now ubiquitous shrinkage estimators for individual gene variances and a dependent test false discovery rate adjustment of bootstrap p-values based upon work by Benjamini and Yekutieli. The performance of the package for computational speed is compared to other single processor compiled packages. The resulting lists of genes from each method are also compared using a small example data set where results are assumed known.