

“Statistical Methods for Identification of Differentially Expressed Gene Groups and Pathways”

David M. Rocke,
Distinguished Professor
Division of Biostatistics
University of California, Davis

Abstract

There are a number of reasons why it may be important to search for differential expression of groups of genes and pathways rather than for differential expression of individual genes. If biological samples are taken at a fixed time following an intervention, a transcriptional cascade may occur at different speeds in different individuals. At a fixed time, the differential transcription will then lie in different genes within the pathway for different individuals, thus resulting in a signal that occurs in one gene in the pathway for some individuals and other genes for other individuals. Polymorphisms and differences in physiology can result in differential expression of one gene from a class (e.g., MAP Kinases) in one individual and other genes from the same class in other individuals. Up- or down-regulation may be broad across a class of genes but with a signal that is too diffuse and weak to be detected in the results from individual genes. The greater power from aggregation of results may increase the sensitivity.

The earliest methods for handling this general problem class involved computing whether the gene group at issue is over-represented in a set of, for example, significantly differentially expressed genes. This is sometimes called Gene Set Enrichment Analysis (GSEA). In Rocke et al. (2005), a new method of analyzing gene groups and pathways was introduced, called the Test of Test Statistics (ToTS) method, in which a set of test statistics, such as in the case of the example a t-test for the dose-response slope, is tested for a positive or negative bias using the Wilcoxon one-sample test. This proved to be much more powerful than tests of individual genes. In order to make the procedure robust to correlations in the tests, a re-sampling based method is used to determine significance, rather than the usual asymptotic p-values for the Wilcoxon test. We have subsequently investigated a number of alternative approaches for evaluating the statistical significance of such results, and found methods that are effective under a wide variety of assumptions.