

Gene Set Analysis in Expression Profiling Studies

James J. Chen

National Center for Toxicological Research, U.S. FDA

Jefferson, AR 72079, U.S.A.

Abstract

Biological phenomena often occur through the interactions of multiple genes, via signaling pathways, networks, or other functional relationships. Gene set analysis (GSA) is a statistical approach to determine whether some functionally predefined classes of genes express differently (enrichment and/or deletion) in different phenotypes. Most GSA methods use a univariate gene-by-gene analysis or assume independence without accounting for functional relationships among genes in the gene set. We propose using global and MANOVA (multivariate analysis of variance) tests for the one-sided and two-sided hypothesis, respectively. The one-sided test means that the changes of gene expressions in the gene class are in one direction: either up or down. The two-sided test means that changes of gene expressions in the gene class can be both up- and down-regulation. The global test and MANOVA tests are compared to several existing GSA methods. The proposed tests are shown to perform well in terms of the control of Type I error and power. Several publicly available microarray datasets under two and three experimental conditions are analyzed.