

The Funnel Algorithm: Toward an Epistemologically Rigorous and Statistically Powerful Procedure for the Analysis of Microarray Data

Jacob Pieter Laurens Brand, Ph.D.
University of Alabama, Birmingham

Abstract:

Microarray technology is an exciting new technology enabling scientists to measure gene expression levels for thousands of genes at a time. However, analysis of the large volumes of data generated in microarray experiments is challenging and accounting for interactions and associations among genes is even more difficult. A common situation in microarray data analysis is the testing of k null-hypotheses where k is the number of genes (k is often more than 10,000) and each null-hypothesis is related to one particular gene only. A standard approach to this situation is to do an independent statistical test on each gene. Two major limitations of the standard approach are i.) interactions and dependencies between genes are not taken into account and ii) the large number of tests, often more than 10,000, can potentially lead to a large number of false positives, for which avoidance of which require prohibitive multiple comparison corrections. We propose a new algorithm, called the Funnel Algorithm, as a (partial) solution to the two problems mentioned above. The Funnel algorithm is an iterative algorithm, where each iteration is a two step procedure. Step 1 is to partition genes into meaningful groups based on a clustering algorithm or prior biological information. Step 2 is to do a statistical test of significance on each group to detect whether a group of genes contains biologically interesting genes. The algorithm stops when all remaining groups consist of one gene and the interesting genes are found. By choosing an appropriate method to partition the genes into groups, relationships and dependencies between genes can be taken into account. By mapping the Funnel algorithm to the framework of a closed testing procedure (Marcus et al 1976) it is shown that far less stringent multiple comparison correction are required for controlling the FWER (Family Wise Error Rate) at a confidence level of α .