Statistical Framework for Integrative Analysis of Multiple Gene Expression Profiles

Jia Li, Sunghee Oh, and George C. Tseng
Department of Biostatistics, University of Pittsburgh, Pittsburgh, USA

With the availability of tons of gene expression profiles on the web, the needs of different forms of meta-analyses to enhance different types of microarray data analyses, including detection of differentially expressed genes, classification analysis and generation of gene networks, are obvious. For detection of differentially expressed genes, most of the current efforts are focused on comparing and evaluating gene lists obtained from each individual data set and a real sense of information integration is rarely performed. In this paper, we propose a general framework of statistical integration by a weighted averaged statistics when multiple biologically relevant data sets are available and a permutation analysis is applied to control the false discovery rate. A subset of differentially expressed genes identified in the integrative analysis are otherwise ignored in each individual analysis. We will show the advantage of combining information from multiple data sets through simulated data sets and real data sets of lung cancer and prostate cancer.

[George C. Tseng, Department of Biostatistics, 130 Desoto Street, Pittsburgh, PA 15261, USA; ctseng@pitt.edu]