International Biometric Society

Is there an alternative to increasing the sample size in microarray studies?

Xing Qiu[1], Lev Klebanov[2], Andrei Yakovlev[1],

[1] Department of Biostatistics and Computational Biology, University of Rochester, U.S.A.
[2] Department of Probability and Statistics, Charles University, Czech Republic

All methods of significance testing in gene expression profiling are highly unstable and their power tends to be very low. These undesirable properties are due to the nature of multiple testing procedures, as well as extremely strong and long-ranged correlations between gene expression levels. The commonly used normalization procedures do not provide a satisfactory solution because of their distorting effects on the true expression signals. Such effects are especially pronounced in larger samples where control of type 1 errors may be entirely lost. Many authors resort to pooling test-statistics across genes to overcome the notorious sample size limitations. However, the correlation structure remains to be the main obstacle standing in the way of this approach. We propose a method overcoming this obstacle and demonstrate its usefulness in conjunction with the nonparametric empirical Bayes method (NEBM). While the proposed modification of the NEBM leads to dramatic improvements in its performance, the currently practiced small sample sizes still remain a major concern.