Mathematical Biosciences, Vol. 176 (1) (2002) pp. 59-69

PII: S0025-5564(01)00105-5

Multivariate approach for selecting sets of differentially expressed genes

A. Chilingaryan [a], N. Gevorgyan [a], A. Vardanyan [a], D. Jones [b] and A. Szabo [b] [*]
aniko.szabo@hci.utah.edu

[a] Cosmic Ray Division, Yerevan Physics Institute, 2 Alikhanian Brothers st., Yerevan 375036, Armenia
[b] Huntsman Cancer Institute and Department of Oncological Sciences, University of Utah, 2000 Circle of Hope, Salt Lake City, UT 84112-5550, USA

Received 21 May 2001; received in revised form 13 June 2001; accepted 20 November 2001

Abstract

An important problem addressed using cDNA microarray data is the detection of genes differentially expressed in two tissues of interest. Currently used approaches ignore the multidimensional structure of the data. However it is well known that correlation among covariates can enhance the ability to detect less pronounced differences. We use the Mahalanobis distance between vectors of gene expressions as a criterion for simultaneously comparing a set of genes and develop an algorithm for maximizing it. To overcome the problem of instability of covariance matrices we propose a new method of combining data from small-scale random search experiments. We show that by utilizing the correlation structure the multivariate method, in addition to the genes found by the one-dimensional criteria, finds genes whose differential expression is not detectable marginally.

Keywords: Microarray; Random search; Mahalanobis distance; Simulation study

[*]Corresponding author. Tel.: +1-801 585 5182; fax: +1-801 585 5357/5337