MulCom: a multiple comparison statistical test for microarray data in Bioconductor

Isella C[1], Renzulli T[1], Corà D[2] ,Medico E[1]

Motivation

Together with the development of -omics platforms and in particular of microarray analysis of gene expression, the bioinformatics community has developed several statistical analysis tools to handle data and detection of differentially regulated genes through weighing of systematic and random errors coming from the highly parallel -but poorly replicated- microarray expression data. Unfortunately, statistical analysis suffers from the limited number of replicates usually performed in experimental design. Typically the algorithms are based on pair-wise t-test (SAM. Limma) to identify gene lists discriminating two experimental subgroups. However several gene expression experiments encompass multiple experimental points to be compared. This is the case of a time-course or multiple independent points; although standard approaches can be applied to this type of data, they work suboptimaly.

Methods

MulCom is an R package designed to identify differentially expressed genes in DNA microarray data obtained on multiple experimental points. The package implements the Dunnett's t-test for multiple comparisons against a common reference, with an optional tunable fold-change threshold. Test parameters are automatically optimized by systematic estimate of the false discovery rate through sample permutations.

Results

A time-course microarray experiment conducted in parallel on two microarray platforms was successfully used to cross-validate the lists of genes generated by MulCom and other algorithms, such as Limma and SAM. Running the analyses on the two platforms, the MulCom gene lists showed greater overlap in respect to the lists of gene generated either by Limma or SAM, supporting the higher precision of the MulCom test in the identification of differentially expressed genes in a time-course experiment. Furthermore Mulcom-based analysis generated gene lists more robustly enriched in functional annotations related to the biological effects of the

[1] Department of Oncological Sciences and Laboratory of Oncogenomics, Institute for Cancer Research and Treatment (IRCC), University of Turin Medical School, Candiolo, Torino, Italy [2] Systems Biology Lab Institute for Cancer Research and Treatment (IRCC), School of Medicine, University of Torino, Str. Prov. 142 Km. 3.95, I-10060 Candiolo, Torino, Italy

time-course experiment. The MulCom package can also easily be adapted to handle any -omics datasets such as miRNAomic, proteomic and metabolomic studies where a Dunnetts type of statistical approach is required.

Contact e-mail
claudio.isella@ircc.it