

SEL.TS.AREA

This program selects differentially expressed genes in time series experiments based on the area of the region bounded by the time series expression profiles to be compared, and considers the gene differentially expressed if the area exceeds a threshold based on a model of the experimental error.

This is a preliminary version of the software. Only comparison between time series (e.g. control vs treated) has been implemented. Comparison of a time series vs its baseline has not yet been implemented.

To use the program source the R script from R workspace (you can download R at <http://www.r-project.org/>).

Usage

`SEL.TS.AREA(replicates,data1,data2,sampling.grid=NULL,takelog=FALSE,binsize=0.1,B=10000,NAcontrol=1)`

Arguments

replicates:

a matrix of two columns containing two vectors of available replicated measurements.

For example, if arrays $C(t_{1bis})$, $C(t_{2bis})$ represent experimental replicates of $C(t_1), C(t_2)$ the first column of "replicates" must contain data from $C(t_1), C(t_2)$ (aligned in a single vector) and the second column of "replicates" must contain data from C_{t1bis} , C_{t2bis} (aligned in a single vector)

data1,data2:

data to be compared. data1, data2 are matrix containing genes in rows and samples (time samples) in columns. data2 contains samples to be compared with those in data1.

sampling.grid:

time grid at which the data are sampled. Sampling grid must have length equal to the number of columns of data1 and data2. If time samples are equally spaced sampling grid can be set to NULL (default).

takelog:

should the log be taken or data are already inserted as log?

binsize:

size of intensity intervals to be used for error standardization

B:

number of time series generated by resampling to calculate the null hypothesis distribution

NAcontrol:

minimum number of available samples (not NA) in a gene time series that must be available in order the gene to be considered for selection.

Details

replicates, data1, data2 must contain already pre-processed data (e.g normalized, filtered and not containing NA).

The algorithm works iteratively asking to the user:

1. whether using a constant or an intensity dependent model of the error (a plot of the variance of replicates1-Replicates2 vs average intensity is shown to help the decision).
2. to choose the best model among the ones fitting the null hypothesis distribution based on goodness of fit. Parameter precision and plots are given to help the choice. It's also possible to use quantiles instead of distribution models

Value

No value returned. The file "Result.txt" is saved in the workspace

Author:

Barbara Di Camillo

References:

Barbara Di Camillo, Gianna Toffolo, Sreekumaran K. Nair, Laura J. Greenlund, Claudio Cobelli.

Significance analysis of microarray transcript levels in time series experiments.

BMC Bioinformatics – Accepted.