



SurvivalGene Documentation

Description: Select genes correlated with time-to-event clinical outcome

Author: Yujin Hoshida, hoshida@broadinstitute.org

Summary

This module evaluates correlation of each gene's expression level with time-to-event clinical outcome such as time to tumor recurrence, disease progression, or death using Cox score [1,2] or Cox regression coefficient. Significance of the correlation is assessed based on standard random permutation test.

In contrast to gene selection based on class comparison (e.g., comparison between two sample groups using t-test), the correlation with time-to-event outcome is more affected by measurement in each single sample. That is, an outlier measurement in a single poor quality sample could falsely yield a high correlation. To avoid selecting such false positives, it is recommended to further refine the gene list by running LoocvSurvival module (this module takes an output file of SurvivalGene module, "SurvivalGene_emp.stat.txt", as an input).

References

[1] N Engl J Med. 2008 Nov 6;359(19):1995-2004

[2] PLoS Med. 2006 Jan;3(1):e13

Parameters

Name	Description
input filename gct	Gene expression dataset (.gct)
input filename clinical	Clinical data (tab-delimited .txt, sample order should be the same with .gct)
output file	Prefix for output file name, Default: SurvivalGene
time field	Field name for time variable, Default: time



sensor field	Field name for censoring information (0: censor, 1: event) , Default: status
statistic selection	Statistic to use, Choice: Cox.score, Cox.regression.coefficient, Default: Cox.score
trim percent 2 side	Proportion of samples with outlier outcome to trim (2-side, e.g., "0.05"), Default: 0
nperm	Number of sample name permutations to compute empirical distribution of statistic (must be >2 for Shapiro-Wilk normality test), Default: 1000
rnd seed	Random seed, Default: 56438219
emp stat dist	Output mean & SD of empirical distribution of statistic used for LoocvSurvival module? Perform Shapiro-Wilk normality test? (for Cox score only), Default: No

Input Files

1. input filename gct
Normalized gene expression dataset in GenePattern GCT file format.
2. input filename clinical
Clinical dataset including time-to-clinical event variable and censoring information.
Tab-delimited text. First column should be sample identified in the same order with
the gene expression dataset (.gct).

Output Files

1. SurvivalGene.txt
Gene list with statistic (Cox score or Cox regression coefficient), p-value, Benjamini-Hochberg false discovery rate (BH.FDR), and Bonferroni-corrected p-value. Tab-delimited text.
2. SurvivalGene_emp.stat.txt (optional)
Gene list with mean and sample standard deviation for Cox score. Last column is p-Shapiro-Wilk test p-value (test of normality for Cox score).



Example Data

"SurvivalGene_example_gene_expression_data.gct": Gene expression data

"SurvivalGene_example_clinical_data.txt": Clinical data

Platform Dependencies

Module type: Gene List Selection

CPU type: Any

OS: Any

Language: R