LoocvSurvival Documentation

Description:	Leave-one-out cross-validation (LOOCV)-type prediction based on time-to-event clinical outcome. Select genes robustly correlated with time-to-event clinical outcome throughout the LOO steps.
Author:	Yujin Hoshida, hoshida@broadinstitute.org

Summary

This module performs leave-one-out cross validation (LOOCV)-type prediction with genes correlated with time-to-event clinical outcome. In each LOO step, the outcome-correlated genes are selected using Cox score at a certain significance threshold [1,2]. In the assessment of significance, each Cox score is standardized using mean and sample standard deviation calculated by SurvivalGene module (the module outputs "SurvivalGene_emp.stat.txt" file, which can be used as input to LoocvSurvival module). Outcome prediction is made by nearest template prediction method [3]. The module also outputs which genes are recurrently selected as part of predictive signature in each LOO step, thereby user can know which specific genes are robustly correlated with outcome and worth further evaluation, assay development, etc.

References

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

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

[3] PLoS One. 2010 Nov 23;5(11):e15543

Parameters

Name	Description
input dataset	Gene expression dataset (.gct)
input clinical data	Clinical dataset (tab-delimited .txt). Order of samples should be the same with .gct.
time field	Column name for time variable, Default: time
censor field	Column name for censoring information (0: censored, 1: event), Default: status

trim percent 2 side	Proportion of samples with outlier outcome to trim (2-side, e.g., "0.05"), Default: 0
cox score sig	Empirical distibution of Cox score (output of SurvivalGene module, "SurvivalGene_emp.stat.txt") to standardize each Cox score before computing a p-value. 1st column: probe/gene ID, 2nd column: mean of Cox score, 3rd column: SD of Cox score. If this step is not needed, use 0 for mean (2nd column) and 1 for SD (3rd column), Default: 0.05
temp nn wt	Weight gene by Cox score for prediction?, Default: Yes
dist selection	Distance metric for prediction, Default: Cosine.distance
nresample	# of resampling to compute prediction confidence, Default: 1000
rnd seed	Random seed, Default: 4675921
output file	Prefix for output file name, Default: LOOCV_Survival

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).
3. emp cox file
   Gene list with mean and sample standard deviation for Cox score. Tab-delimited text. Output of SurvivalGene module.

Output Files

1. LOOCV_Survival_poor.features.txt

   Gene list indicating which genes are selected as predictive signature for each LOO step. Genes with "poor" are selected as signature genes correlated with poor outcome when the sample is left out. "percent.poor" column indicates proportion of patients for which each gene is selected as a signature gene. For example,

"percent.poor" of 1 means that the gene is robustly correlated with outcome because it is chosen as part of predictive signature when each one of all patients is left out.

2. LOOCV_Survival_good.features.txt
   Gene list indicating which genes are selected as predictive signature for each LOO step. Genes with "good" are selected as signature genes correlated with good outcome when the sample is left out. "percent.good" column indicates proportion of patients for which each gene is selected as a signature gene. For example, "percent.good" of 1 means that the gene is robustly correlated with outcome because it is chosen as part of predictive signature when each one of all patients is left out.
3. LOOCV_Survival_prediction.result.txt
   Summary of the LOOCV-type prediction with prediction confidence.

Example Data

"SurvivalGene_example_gene_expression_data.gct": Gene expression data

"SurvivalGene_example_clinical_data.txt": Clinical data

"SurvivalGene_emp.stat.txt": Mean and sample standard deviation for each Cox score (output of SurvivalGene module)

Platform Dependencies

Module type:	Prediction, Gene List Selection, Survival Analysis
CPU type:	Any
OS:	Any
Language:	R