PairedEseach Documentation

Description: Assess enrichment of gene sets in each individual paired samples in a dataset.

Author: Yujin Hoshida, Yujin.Hoshida@UTSouthwestern.edu

Summary

This module assesses enrichment and its significance of gene set(s) in each individual pair of samples (e.g., samples obtained before and after treatment in a patient) a molecular profiling dataset. The molecular profiling data can be abundance of transcripts, proteins, etc. that allows rank-ordering of the molecular features in the dataset by log2 ratio. Enrichment is calculated based on the Kolmogorov–Smirnov statistic-based gene set enrichment score (PNAS 102;15545, 2005). Significance of the enrichment is calculated by random-permutation-based nominal p-value, which is also converted to Gene Set Enrichment Index (GSEI) as -log10(nominal p-value) with sign of enrichment score.

References

Cancer Cell. 2016 Dec 12;30(6):879-890. PMID: 27960085

Parameters

Name	Description
input data	Input data file (.gct). Samples should be ordered like "A-1, B-1, C-1,...,A-2, B-2, C-2,..", where 1 and 2 indicate paired measurement for samples (e.g., patients), A, B, and C.
marker data	Gene set database (.gmx)
Output name	Prefix for output file names. Default: "peseach".
norm method	Data normalization method. "none" is default. "row standardization" can also be chosen.

dir change	Ratio (fold change) calculation by paired "measurement 1" over "measurement 2" (i.e., "measurement 1"/"measurement 2") or "measurement 2" over "measurement 1" (i.e., "measurement 2"/"measurement 1").
rank metric	Gene ranking metric. Currently only log2-ratio.
gene ordering	Direction of gene ordering. Default: "descending".
nperm	# of gene permutation. Default: 1000.
min size	Size threshold for gene set. Default: 10.
weighted score type	Weight enrichment statistic with rank metric? Default:yes.
fraction ES perm w same sign	Required fraction of permutation-based enrichment scores to calculate normalized enrichment score. Default: 0.1.
ces	Use combined enrichment score? Only for paired 2 gene sets (e.g., up- and down-regulated genes, poor- and good-prognosis genes). Default: "no".
plot	Plot running gene set enrichment? Default: yes.
line width	Line width for gene set enrichment plot. Default: 2.
line col	Line color for gene set enrichment plot. Default: green.
rnd seed	Random seed. Default: 4736859.

Input Files

1. input data
   - Molecular profiling data in .gct format.
2. Marker data

- Gene set database in .gmx format.

Output Files

1. peseach_ES.obs.gct
   Matrix of enrichment score (ES).

2. peseach_NES.gct
   Matrix of normalized enrichment score (NES).

3. peseach_nom.p.gct
   nominal p-values for the ESs.

4. peseach_neg.log10.p.ESsign.gct
   Matrix of Gene Set Enrichment Index (GSEI).

Example Data

Any .gct and .gmx format files can be used.

Platform Dependencies

Module type:	Projection
CPU type:	Any
OS:	Any
Language:	R (ver. 2.0 or higher)