



eseach Documentation

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

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

Summary

This module assesses enrichment and its significance of gene set(s) in a molecular profiling dataset. The molecular profiling data can be transcriptome abundance, statistic measuring association with phenotype (e.g., t-statistic calculated from group comparison), etc. that allows rank-ordering of the molecular features in the dataset. 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 $-\log_{10}(\text{nominal p-value})$ 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).
marker data	Gene set database (.gmx)
Output name	Prefix for output file names. Default: "eseach".
norm method	Data normalization method. "row standardization" for expression data. "none" for rank or pre-calculated rank metric.
gene ordering	Gene ordering. "descending" for expression data. "ascending" for gct of rank (not for rank metric, e.g., t-statistic). 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.
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. eeach_ES.obs.gct
Matrix of enrichment score (ES).
2. eeach_NES.gct
Matrix of normalized enrichment score (NES).
3. eeach_nom.p.gct
nominal p-values for the ESs.
4. eeach_neg.log10.p.ESign.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)