MAGeCK.Pathway.Analysis

Test if a pathway is enriched in one particular pathway ranking using RRA.

Author
Clarence Mah;Mesirov Lab

Contact
ckmah@ucsd.edu

Algorithm version
0.5.5

Task Type
CRISPR

CPU Type
Any

Operating System
Any

Language
Python, C

References

Li, W., Xu, H., Xiao, T., Cong, L., Love, M. I., Zhang, F., ... Mesirov, J. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology 2014 15:12*, *15*(12), 819–823. https://doi.org/10.1126/SCIENCE.1231143

Input Files

1. Gene ranking

The contents of each column are as follows:

Column	Content
id	Gene ID
num	The number of targeting sgRNAs for each gene
neg|score	The RRA lo value of this gene in negative selection

neg|p-value	The raw p-value (using permutation) of this gene in negative selection
neg|fdr	The false discovery rate of this gene in negative selection
neg|rank	The ranking of this gene in negative selection
neg|goodsgrna	The number of "good" sgRNAs, i.e., sgRNAs whose ranking is below the alpha cutoff (determined by the gene test FDR threshold option), in negative selection.
neg|lfc	The log fold change of this gene in negative selection
pos|score	The number of targeting sgRNAs for each gene in positive selection (usually the same as num.neg)
pos|score	The RRA lo value of this gene in negative selection
pos|p-value	The raw p-value of this gene in positive selection
pos|fdr	The false discovery rate of this gene in positive selection
pos|rank	The ranking of this gene in positive selection
pos|goodsgrna	The number of "good" sgRNAs, i.e., sgRNAs whose ranking is below the alpha cutoff (determined by the gene test FDR threshold option), in positive selection.
pos|lfc	The log fold change of this gene in positive selection

Genes are ranked by the `neg|p-value` (by default). If you need a ranking by the `pos|p-value`, you can use the "sort criteria" option.

2. Pathways file

The GMT file format stores the pathway information and is consistent with the GMT file in Gene Set Enrichment Analysis (GSEA). The details of the GMT format can be found at GSEA website. You can also download different pathway files directly from GSEA MSigDB database. They can be used directly by MAGeCK.

Output Files

1. Count Table
The output of the pathway summary is similar to the gene summary:

Column	Content
id	The pathway ID
num	The number of genes in the pathway

neg|score	The RRA lo value of this pathway in negative selection
neg|p-value	The raw p-value (using permutation) of this pathway in negative selection
neg|fdr	The false discovery rate of this pathway in negative selection
neg|rank	The ranking of this pathway in negative selection
neg|goodsgrna	The number of "good" sgRNAs, i.e., sgRNAs whose ranking is below the alpha cutoff (determined by the pathway test FDR threshold option), in negative selection.
pos|num	The number of targeting sgRNAs for each pathway in positive selection (usually the same as num.neg)
pos|score	The RRA lo value of this pathway in negative selection
pos|p-value	The raw p-value of this pathway in positive selection
pos|fdr	The false discovery rate of this pathway in positive selection
pos|rank	The ranking of this pathway in positive selection
pos|goodsgrna	The number of "good" sgRNAs, i.e., sgRNAs whose ranking is below the alpha cutoff (determined by the pathway test FDR threshold option), in positive selection.

2. Log file
This file includes the logging information during the execution.

3. Intermediate files
These files will be automatically deleted after the completion of each command. To keep these files, use the "keep intermediate files" option during the execution.

Example Data

http:/github.com/ckmah/mageck_pathway_analysis/blob/master/data/demo.count.txt
http:/github.com/ckmah/mageck_pathway_analysis/blob/master/data/c6.all.v5.2.symbols.gmt

Requirements

MAGeCK can be run on either Mac or Linux system. Since MAGeCK is written in Python and C, Python 2.7 (>2.7) and a C compiler is needed.

Module Parameters

*required

Column	Content
Gene ranking*	The gene ranking file generated by the gene test step.
Gmt file*	The pathway file in GMT format.
Single ranking	The provided file is a (single) gene ranking file, either positive or negative selection. Only one enrichment comparison will be performed.
Output prefix	The prefix of the output file(s). Default sample1.
Sort criteria	Sorting criteria, either by negative selection (neg) or positive selection (pos). Default negative selection.
Keep intermediate files	Keep intermediate files.
Ranking column	Column number or label in gene summary file for gene ranking; can be either an integer of column number, or a string of column label. Default "2" (the 3rd column).
Ranking column 2	Column number or label in gene summary file for gene ranking; can be either an integer of column number, or a string of column label. This option is used to determine the column for positive selections and is disabled if --single-ranking is specified. Default "5" (the 6th column).
Column	Content
Gene ranking	The gene ranking file generated by the gene test step.
Gmt file	The pathway file in GMT format.