

Package ‘phantom’

March 29, 2018

Title Gene Set Pareto Heterogeneity Analysis of Time-Course Gene Expression Data

Version 0.1.3

Author Jinghua Gu [aut, cre],
Xuan Wang [aut],
Jinayn Chan [aut, cre],
Nicole Baldwin [aut],
Jacob Turner [aut]

Maintainer Jinyan Chan <Jinyan.Chan@bswhealth.org>

Description Pareto front based statistical tool for detecting heterogeneity in gene sets and biological modules from time-course data. Details about this method are described in Gu, J., et al. (2017) <doi: 10.1093/bioinformatics/btx348>: Phantom: investigating heterogeneous gene sets in time-course data.

License GPL-2

Encoding UTF-8

LazyData true

Depends R (>= 3.3)

biocViews

Imports Rcpp (>= 0.12.3), NMF (>= 0.20.6), MASS (>= 7.3-35), RColorBrewer (>= 1.1-2), gplots (>= 2.16.0), qusage (>= 2.12.0)

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 6.0.1

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation yes

Repository CRAN

Date/Publication 2018-03-29 21:59:44 UTC

R topics documented:

geneset.names	2
load.data	3
load.geneset	3
run.phantom	4
run.phantom.batch	5
time.course.data	6
Index	7

geneset.names	<i>Print all geneset names in geneset list</i>
---------------	--

Description

This function allows user to print all the geneset names in the loaded geneset list such that user can find the specific query_geneset name for run.phantom function.

Usage

```
geneset.names(geneset_list = NULL)
```

Arguments

`geneset_list` User provided genesets list loaded by `load.geneset()`. Phantom package provides four geneset lists from different resources: kegg, reactome, emory geneset and baylor modules. These genesets can be obtained with `data()`, e.g. `data(kegg.geneset)`

Examples

```
## store all the geneset names in one vector
## Not run:
g.names = geneset.names(reactome.geneset)

## End(Not run)
```

load.data	<i>Load data function</i>
-----------	---------------------------

Description

This function allows user to load in the time-course gene differential expression data in a format that can be used by run.phantom and run.phantom.batch functions.

Usage

```
load.data(filename)
```

Arguments

filename	the time-course gene differential expression data file path and name on user's local machine
----------	--

load.geneset	<i>Load geneset in gmt file format</i>
--------------	--

Description

This function allows user to load in the genesets that are downloaded from GSEA (gmt format) into a format that can be used by run.phantom and run.phantom.batch functions.

Usage

```
load.geneset(geneset_file)
```

Arguments

geneset_file	the genesets gmt file path and name on user's local machine
--------------	---

run.phantom	<i>Run phantom analysis</i>
-------------	-----------------------------

Description

This function allows user to run individual geneset heterogeneity analysis with phantom

Usage

```
run.phantom(data = NULL, geneset_list = NULL, query_geneset = NULL,  
            ncluster = 2, nsample = 1000)
```

Arguments

data	User provided time-course data loaded by load.data()
geneset_list	User provided genesets list loaded by load.geneset(). Phantom package provides four geneset lists from different resources: kegg, reactome, emory geneset and baylor modules. These genesets can be obtained with data(), e.g. data(kegg.geneset)
query_geneset	The name of a geneset user wants to analysis. This geneset should be from the geneset_list designated by geneset_list parameter
ncluster	The number of clusters within a geneset user wants to use to identify the heterogeneity of this geneset.
nsample	The times of random sampling that is used to build the NULL distribution for parato front analysis.

Examples

```
## load in the demo data in phantom package  
data("time.course.data")  
  
## store the analysis result in an object  
## Not run:  
obj = run.phantom(data = time.course.data, geneset_list = reactome.geneset,  
                 query_geneset = 'REACTOME_ANTIVIRAL_MECHANISM_BY_IFN_STIMULATED_GENES',  
                 ncluster = 2, nsample = 1000)  
  
## End(Not run)
```

run.phantom.batch	<i>Run phantom batch analysis</i>
-------------------	-----------------------------------

Description

This function allows user to run batch analysis of a full geneset list with phantom, and download the identified heterogeneous genesets

Usage

```
run.phantom.batch(data = NULL, geneset_list = NULL, maxncluster = 5,  
  nsample = 1000, report_pval = 0.05, report_nmin = 5,  
  output_dir = file.path("./phantom_result"))
```

Arguments

data	User provided time-course data loaded by load.data()
geneset_list	User provided genesets list loaded by load.geneset(). Phantom package provides four geneset lists from different resources: kegg, reactome, emory geneset and baylor modules. These genesets can be obtained with data(), e.g. data(kegg.geneset)
maxncluster	The maximum number of clusters within a geneset user wants to test with. All numbers from 1 to ncluster will be tested and an optimal cluster number will be selected to identify the heterogeneity of this geneset.
nsample	The times of random sampling that is used to build the NULL distribution for parato front analysis.
report_pval	The maximum p value of a geneset that will be reported as a significant heterogeneous geneset. Genesets with p value larger than report_pval wil not be reported
report_nmin	The minmum size of subcluster in a geneset that will be reported as a significant heterogeneous geneset.
output_dir	The directory where user wants to put the phantom batch analysis results

Examples

```
## load in the demo data and geneset in phantom package  
data("time.course.data")  
data("kegg.geneset")  
  
## store the analysis result in an object  
## Not run: obj = run.phantom.batch(data = time.course.data, geneset_list = kegg.geneset,  
  maxncluster = 5, nsample = 1000, report_pval = 0.05, report_nmin = 5,  
  output_dir = file.path(getwd(), '/phantom_result'))  
## End(Not run)
```

time.course.data	<i>A sample time.course.data with 15 time points</i>
------------------	--

Description

This data set gives a sample of time-course gene differential expression data, designated by t statistics of symptomatic samples versus asymptomatic samples gene expression across 15 time points.

Usage

```
time.course.data
```

Format

A large matrix containing the t statistics of 11745 genes across 15 time points

Source

publications

Index

- *Topic **batch**
 - run.phantom.batch, 5
 - *Topic **datasets**
 - time.course.data, 6
 - *Topic **data**
 - load.data, 3
 - *Topic **geneset**
 - geneset.names, 2
 - load.geneset, 3
 - *Topic **load**
 - load.data, 3
 - load.geneset, 3
 - *Topic **names**
 - geneset.names, 2
 - *Topic **phantom**
 - run.phantom, 4
 - run.phantom.batch, 5
 - *Topic **print**
 - geneset.names, 2
 - *Topic **run**
 - run.phantom, 4
 - run.phantom.batch, 5
- geneset.names, 2
- load.data, 3
- load.geneset, 3
- run.phantom, 4
- run.phantom.batch, 5
- time.course.data, 6