Package 'fgsea'

October 17, 2017

Title Fast Gene Set Enrichment Analysis

Version 1.2.1

Date 2016-11-21

Description The package implements an algorithm for fast gene set enrichment analysis. Using the fast algorithm allows to make more permutations and get more fine grained p-values, which allows to use accurate stantard approaches to multiple hypothesis correction.

biocViews GeneExpression, DifferentialExpression, GeneSetEnrichment, Pathways

SystemRequirements C++11

Depends R (>= 3.3), Rcpp

Imports data.table, BiocParallel, stats, ggplot2 (>= 2.2.0), gridExtra, grid, fastmatch

Suggests testthat, knitr, rmarkdown, reactome.db, AnnotationDbi, parallel

License MIT + file LICENCE

LazyData true

LinkingTo Rcpp

RoxygenNote 5.0.1

VignetteBuilder knitr

URL https://github.com/ctlab/fgsea/

BugReports https://github.com/ctlab/fgsea/issues

NeedsCompilation yes

Author Alexey Sergushichev [aut, cre]

Maintainer Alexey Sergushichev <alsergbox@gmail.com>

R topics documented:

`calcGseaStat` *Calculates GSEA statistics for a given query gene set*

Description

Takes *O(k log k)* time, where *k* is a size of 'selectedSize'.

Usage

```
calcGseaStat(stats, selectedStats, gseaParam = 1, returnAllExtremes = FALSE,
  returnLeadingEdge = FALSE)
```

Arguments

`stats`	Named numeric vector with gene-level statistics sorted in decreasing order (order is not checked).
`selectedStats`	Indexes of selected genes in the 'stats' array.
`gseaParam`	GSEA weight parameter (0 is unweighted, suggested value is 1).
`returnAllExtremes`	If TRUE return not only the most extreme point, but all of them. Can be used for enrichment plot
`returnLeadingEdge`	If TRUE return also leading edge genes.

Value

Value of GSEA statistic if both returnAllExtremes and returnLeadingEdge are FALSE. Otherwise returns list with the folowing elements:

- res – value of GSEA statistic
- tops – vector of top peak values of cumulative enrichment statistic for each gene;
- bottoms – vector of bottom peak values of cumulative enrichment statistic for each gene;
- leadingGene – vector with indexes of leading edge genes that drive the enrichment, see http://software.broadinstitute.org/gsea/doc/GSEAUserGuideTEXT.htm#_Running_a_Leading.

Examples

```
data(exampleRanks)
data(examplePathways)
ranks <- sort(exampleRanks, decreasing=TRUE)
es <- calcGseaStat(ranks, na.omit(match(examplePathways[[1]], names(ranks))))
```

examplePathways *Example list of mouse Reactome pathways.*

Description

The list was obtained by selecting all the pathways from ‘reactome.db‘ package that contain mouse genes. The exact script is available as system.file("gen_reactome_pathways.R", package="fgsea")

exampleRanks *Example vector of gene-level statistics obtained for Th1 polarization.*

Description

The data were obtained by doing differential expression between Naive and Th1-activated states for GEO dataset GSE14308. The exact script is available as system.file("gen_gene_ranks.R", package="fgsea")

fgsea *Runs preranked gene set enrichment analysis.*

Description

The function takes about *O(nk^{3/2})* time, where *n* is number of permutations and *k* is a maximal size of the pathways. That means that setting ‘maxSize‘ parameter with a value of ~500 is strongly recommended.

Usage

```
fgsea(pathways, stats, nperm, minSize = 1, maxSize = Inf, nproc = 0,
  gseaParam = 1, BPPARAM = NULL)
```

Arguments

pathways	List of gene sets to check.
stats	Named vector of gene-level stats. Names should be the same as in ’pathways’
nperm	Number of permutations to do. Minimial possible nominal p-value is about 1/nperm
minSize	Minimal size of a gene set to test. All pathways below the threshold are excluded.
maxSize	Maximal size of a gene set to test. All pathways above the threshold are excluded.
nproc	If not equal to zero sets BPPARAM to use nproc workers (default = 0).
gseaParam	GSEA parameter value, all gene-level statis are raised to the power of ‘gseaParam‘ before calculation of GSEA enrichment scores.
BPPARAM	Parallelization parameter used in bplapply. Can be used to specify cluster to run. If not initialized explicitly or by setting ‘nproc‘ default value ‘bpparam()‘ is used.

Value

A table with GSEA results. Each row corresponds to a tested pathway. The columns are the following:

- pathway – name of the pathway as in ‘names(pathway)‘;
- pval – an enrichment p-value;
- padj – a BH-adjusted p-value;
- ES – enrichment score, same as in Broad GSEA implementation;
- NES – enrichment score normalized to mean enrichment of random samples of the same size;
- nMoreExtreme‘ – a number of times a random gene set had a more extreme enrichment score value;
- size – size of the pathway after removing genes not present in ‘names(stats)‘.
- leadingEdge – vector with indexes of leading edge genes that drive the enrichment, see http://software.broadinstitute.org/gsea/doc/GSEAUserGuideTEXT.htm#_Running_a_Leading.

Examples

```
data(examplePathways)
data(exampleRanks)
fgseaRes <- fgsea(examplePathways, exampleRanks, nperm=10000, maxSize=500)
# Testing only one pathway is implemented in a more efficient manner
fgseaRes1 <- fgsea(examplePathways[1], exampleRanks, nperm=10000)
```

---

gmtPathways *Returns a list of pathways from a GMT file.*

Description

Returns a list of pathways from a GMT file.

Usage

```
gmtPathways(gmt.file)
```

Arguments

gmt.file	Path to a GMT file.

Value

A list of vectors with gene sets.

Examples

```
pathways <- gmtPathways(system.file(
     "extdata", "mouse.reactome.gmt", package="fgsea"))
```

plotEnrichment *Plots GSEA enrichment plot.*

Description

Plots GSEA enrichment plot.

Usage

```
plotEnrichment(pathway, stats, gseaParam = 1)
```

Arguments

pathway	Gene set to plot.
stats	Gene-level statistics.
gseaParam	GSEA parameter.

Value

ggplot object with the enrichment plot.

Examples

```
data(examplePathways)
data(exampleRanks)
## Not run:
plotEnrichment(examplePathways[["5991130_Programmed_Cell_Death"]],
               exampleRanks)

## End(Not run)
```

plotGseaTable *Plots table of enrichment graphs using ggplot and gridExtra.*

Description

Plots table of enrichment graphs using ggplot and gridExtra.

Usage

```
plotGseaTable(pathways, stats, fgseaRes, gseaParam = 1, colwidths = c(5, 3,
  0.8, 1.2, 1.2))
```

Arguments

pathways	Pathways to plot table, as in ‘fgsea‘ function.
stats	Gene-level stats, as in ‘fgsea‘ function.
fgseaRes	Table with fgsea results.
gseaParam	GSEA-like parameter. Adjusts displayed statistic values, values closer to 0 flatten plots. Default = 1, value of 0.5 is a good choice too.
colwidths	Vector of five elements corresponding to column width for grid.arrange.

Value

TableGrob object returned by grid.arrange.

Examples

```
data(examplePathways)
data(exampleRanks)
fgseaRes <- fgsea(examplePathways, exampleRanks, nperm=1000,
                  minSize=15, maxSize=100)
topPathways <- fgseaRes[head(order(pval), n=15)][order(NES), pathway]
## Not run:
plotGseaTable(examplePathways[topPathways], exampleRanks,
              fgseaRes, gseaParam=0.5)

## End(Not run)
```

---

reactomePathways *Returns a list of Reactome pathways for given Entrez gene IDs*

Description

Returns a list of Reactome pathways for given Entrez gene IDs

Usage

```
reactomePathways(genes)
```

Arguments

genes	Entrez IDs of query genes.

Value

A list of vectors with gene sets.

Examples

```
data(exampleRanks)
pathways <- reactomePathways(names(exampleRanks))
```

Index