# Package 'EGSEA'

October 12, 2016

**Title** Ensemble of Gene Set Enrichment Analyses

**Version** 1.0.3

**Date** 10-05-2016

**Author** Monther Alhamdoosh, Milica Ng and Matthew Ritchie

**Maintainer** Monther Alhamdoosh <m.hamdoosh@gmail.com>

**Description** This package implements the Ensemble of Gene Set Enrichment Analyses (EGSEA) method for gene set testing.

**biocViews** DifferentialExpression, GO, GeneExpression, GeneSetEnrichment, Genetics, Microarray, MultipleComparison, OneChannel, Pathways, RNASeq, Sequencing, Software, SystemsBiology, TwoChannel,Metabolomics, Proteomics, KEGG, GraphAndNetwork

**Depends** R (>= 3.3), Biobase, gage (>= 2.14.4), AnnotationDbi, topGO (>= 2.16.0), pathview (>= 1.4.2)

**Imports** PADOG (>= 1.6.0), GSVA (>= 1.12.0), globaltest (>= 5.18.0), limma (>= 3.20.9), edgeR (>= 3.6.8), HTMLUtils (>= 0.1.5), hwriter (>= 1.2.2), gplots (>= 2.14.2), ggplot2 (>= 1.0.0), safe (>= 3.4.0), stringi (>= 0.5.0), parallel, stats, grDevices, graphics, utils, org.Hs.eg.db, org.Mm.eg.db, org.Rn.eg.db, EGSEAdata

**License** GPL-2

**LazyLoad** yes

**NeedsCompilation** no

**Suggests** BiocStyle, knitr, testthat

**VignetteBuilder** knitr

**RoxygenNote** 5.0.1

## R topics documented:

---

`EGSEA-package` *Ensemble of Gene Enrichment Analysis (EGSEA)*

---

## Description

This packages provides the implementatino of the EGSEA algorithm and addition functions to help perform GSE analysis

## Author(s)

Monther Alhamdoosh, Milica Ng and Matthew Ritchie

---

`buildCustomIdxEZID` *Custom Gene Set Collection Index*

---

## Description

It creates gene set collections from a given list of gene sets to be used for the EGSEA analysis.

## Usage

```
buildCustomIdxEZID(entrezIDs, gsets, anno = NULL, label = "custom",
  name = "Custom", species = "Human", min.size = 1)
```

## Arguments

| | |
|---|---|
| `entrezIDs` | character, a vector that stores the Entrez Gene IDs tagged in your dataset. The order of the Entrez Gene IDs should match those of the count/expression matrix row names. |
| `gsets` | list, list of gene sets. Each gene set is character vector of Enterz IDs. The names of the list should match the GeneSet column in the `anno` argument (if it is provided). |

| | |
|---|---|
| anno | list, dataframe that stores a detailed annotation for each gene set. Some of its fields can be ID, GeneSet, PubMed, URLs, etc. The GeneSet field is mandatory and should have the same names as the gsets' names. |
| label | character,a unique id that identifies the collection of gene sets |
| name | character,the collection name to be used in the EGSEA report |
| species | character, determine the organism of selected gene sets: "human", "mouse" or "rat". |
| min.size | integer, the minium number of genes required in a testing gene set |

## Details

It indexes newly created gene sets and attach gene set annotation if provided.

## Value

indexed gene set annotation that can be used with other functions in the package. Each annotation is a list of seven elements: original stores the original gene sets, idx stores the indexed gene sets, anno that stores detailed annotation for each gene set, label a unique id that identifies the collection of gene sets, featureIDs stores the entrezIDs used in building the annotation, species stores that organism name of gene sets and name stores the collection name to be used in the EGSEA report.

## Examples

```
library(EGSEAdata)
data(il13.data)
v = il13.data$voom
kegg = buildIdxEZID(entrezIDs=rownames(v$E), species="human",
msigdb.gsets="none",
         kegg.updated=FALSE, kegg.exclude = c("Metabolism"))
gsets = kegg$kegg$original[1:50]
gs.annots = buildCustomIdxEZID(entrezIDs=rownames(v$E), gsets= gsets,
species="human")
names(gs.annots)
```

---

# buildGeneSetDBIdxEZID *Gene Set Collection Indexes from the GeneSetDB Database*

---

## Description

It prepares the GeneSetDB gene set collections to be used for the EGSEA analysis.

## Usage

```
buildGeneSetDBIdxEZID(entrezIDs, species, by.category = TRUE, min.size = 1,
  rdata.dir = NULL)
```

### Arguments

| | |
|---|---|
| entrezIDs | character, a vector that stores the Entrez Gene IDs tagged in your dataset. The order of the Entrez Gene IDs should match those of the count/expression matrix row names. |
| species | character, determine the organism of selected gene sets: "human", "mouse" or "rat". |
| by.category | logical, whether to group the gene sets into collections or not. Default TRUE. |
| min.size | integer, the minium number of genes required in a testing gene set |
| rdata.dir | character, directory from which the GeneSetDB collections are loaded. If NULL, EGSEA tries to load the data from **EGSEAdata**. |

### Details

It indexes the GeneSetDB gene sets and loads gene set annotation.

### Value

indexed gene set annotation that can be used with other functions in the package. Each annotation is a list of seven elements: `original` stores the original gene sets, `idx` stores the indexed gene sets, `anno` that stores detailed annotation for each gene set, `label` a unique id that identifies the collection of gene sets, `featureIDs` stores the entrezIDs used in building the annotation, `species` stores that organism name of gene sets and `name` stores the collection name to be used in the EGSEA report.

### Examples

```
library(EGSEAdata)
data(il13.data)
v = il13.data$voom
gs.annots = buildGeneSetDBIdxEZID(entrezIDs=rownames(v$E), species="human")
names(gs.annots)
```

---

## buildIdxEZID *Generate Gene Set Collection Indexes from the MSigDB and KEGG Databases*

---

### Description

It prepares the MSigDB and KEGG gene set collections to be used for the EGSEA analysis.

### Usage

```
buildIdxEZID(entrezIDs, species = "human", msigdb.gsets = NULL,
  kegg.updated = FALSE, kegg.exclude = c(), min.size = 1,
  rdata.dir = NULL)
```

### Arguments

| | |
|---|---|
| entrezIDs | character, a vector that stores the Entrez Gene IDs tagged in your dataset. The order of the Entrez Gene IDs should match those of the count/expression matrix row names. |
| species | character, determine the organism of selected gene sets: "human", "mouse" or "rat". |
| msigdb.gsets | character, a vector determines which gene set collections should be used. It can take values from this list: "h", "c1", "c2", "c3", "c4", "c5", "c6","c7". "h" and "c1" are human specific. If NULL, all available gene set collections are loaded. If "none", MSigDB collections are excluded. |
| kegg.updated | logical, set to TRUE if you want to download the most recent KEGG pathways. |
| kegg.exclude | character, vector used to exclude KEGG pathways of specific type(s): Disease, Metabolism, Signaling. If "all", none fo the KEGG collections is included. |
| min.size | integer, the minium number of genes required in a testing gene set |
| rdata.dir | character, directory from which the MSigDB collections are loaded. If NULL, EGSEA tries to load the data from **EGSEAdata**. |

### Details

It indexes the MSigDB and KEGG gene sets and loads gene set annotation.

### Value

indexed gene set annotation that can be used with other functions in the package. Each annotation is a list of seven elements: `original` stores the original gene sets, `idx` stores the indexed gene sets, `anno` that stores detailed annotation for each gene set, `label` a unique id that identifies the collection of gene sets, `featureIDs` stores the entrezIDs used in building the annotation, `species` stores that organism name of gene sets and `name` stores the collection name to be used in the EGSEA report.

### Examples

```
library(EGSEAdata)
data(il13.data)
v = il13.data$voom
gs.annots = buildIdxEZID(entrezIDs=rownames(v$E), species="human",
         msigdb.gsets = c("h", "c2"),
         kegg.exclude = c("Metabolism"))
names(gs.annots)
```

---

## buildKEGGIdxEZID *Gene Set Collection Index from the KEGG Database*

---

### Description

It prepares the KEGG pathway collection to be used for the EGSEA analysis.

### Usage

```
buildKEGGIdxEZID(entrezIDs, species = "human", min.size = 1,
  updated = FALSE, rdata.dir = NULL)
```

### Arguments

| | |
|---|---|
| entrezIDs | character, a vector that stores the Entrez Gene IDs tagged in your dataset. The order of the Entrez Gene IDs should match those of the count/expression matrix row names. |
| species | character, determine the organism of selected gene sets: "human", "mouse" or "rat". |
| min.size | integer, the minium number of genes required in a testing gene set |
| updated | logical, set to TRUE if you want to download the most recent KEGG pathways. |
| rdata.dir | character, directory from which the KEGG pathway collections are loaded. If NULL, EGSEA tries to load the data from **EGSEAdata**. |

### Details

It indexes the KEGG pathway gene sets and loads gene set annotation.

### Value

indexed gene set annotation that can be used with other functions in the package. Each annotation is a list of seven elements: `original` stores the original gene sets, `idx` stores the indexed gene sets, `anno` that stores detailed annotation for each gene set, `label` a unique id that identifies the collection of gene sets, `featureIDs` stores the entrezIDs used in building the annotation, `species` stores that organism name of gene sets and `name` stores the collection name to be used in the EGSEA report.

### Examples

```
library(EGSEAdata)
data(il13.data)
v = il13.data$voom
gs.annots = buildKEGGIdxEZID(entrezIDs=rownames(v$E), species="human")
```

---

## buildMSigDBIdxEZID *Gene Set Collection Indexes from the MSigDB Database*

---

### Description

It prepares the MSigDB gene set collections to be used for the EGSEA analysis.

### Usage

```
buildMSigDBIdxEZID(entrezIDs, geneSets = NULL, species = "Homo \nsapiens",
  min.size = 1, rdata.dir = NULL)
```

### Arguments

| | |
|---|---|
| entrezIDs | character, a vector that stores the Entrez Gene IDs tagged in your dataset. The order of the Entrez Gene IDs should match those of the count/expression matrix row names. |
| geneSets | character, a vector determines which gene set collections should be used. It can take values from this list: "h", "c1", "c2", "c3", "c4", "c5", "c6","c7". "h" and "c1" are human specific. If NULL, all available gene set collections are loaded. |
| species | character, determine the organism of selected gene sets: "human", "mouse" or "rat". |
| min.size | integer, the minium number of genes required in a testing gene set |
| rdata.dir | character, directory from which the MSigDB collections are loaded. If NULL, EGSEA tries to load the data from **EGSEAdata**. |

### Details

It indexes the MSigDB gene sets and loads gene set annotation.

### Value

indexed gene set annotation that can be used with other functions in the package. Each annotation is a list of seven elements: `original` stores the original gene sets, `idx` stores the indexed gene sets, `anno` that stores detailed annotation for each gene set, `label` a unique id that identifies the collection of gene sets, `featureIDs` stores the entrezIDs used in building the annotation, `species` stores that organism name of gene sets and `name` stores the collection name to be used in the EGSEA report.

### Examples

```
library(EGSEAdata)
data(il13.data)
v = il13.data$voom
gs.annots = buildMSigDBIdxEZID(entrezIDs=rownames(v$E), geneSets=c("h",
"c2"), species="human")
names(gs.annots)
```

---

## egsea *Ensemble of Gene Set Enrichment Analyses Function*

---

### Description

This is the main function to carry out gene set enrichment analysis using the EGSEA algorithm. This function is aimed to extend the limma-voom pipeline of RNA-seq analysis.

## Usage

```
egsea(voom.results, contrasts, logFC = NULL, gs.annots, symbolsMap = NULL,
  baseGSEAs = egsea.base(), minSize = 2, display.top = 20,
  combineMethod = "fisher", combineWeights = NULL, sort.by = "p.adj",
  egsea.dir = "./", kegg.dir = NULL, logFC.cutoff = 0,
  sum.plot.axis = "p.adj", sum.plot.cutoff = NULL, vote.bin.width = 5,
  num.threads = 4, report = TRUE, print.base = FALSE, verbose = FALSE)
```

## Arguments

| | |
|---|---|
| voom.results | list, an EList object generated using the voom function. Entrez Gene IDs should be used as row names. |
| contrasts | double, an N x L matrix indicates the contrast of the linear model coefficients for which the test is required. N is number of experimental conditions and L is number of contrasts. |
| logFC | double, an K x L matrix indicates the log2 fold change of each gene for each contrast. K is the number of genes included in the analysis. If logFC=NULL, the logFC values are estimated using the ebayes for each contrast. |
| gs.annots | list, indexed collections of gene sets. It is generated using one of these functions: buildIdxEZID, buildMSigDBIdxEZID, buildKEGGIdxEZID, buildGeneSetDBIdxEZID, and buildCustomIdxEZID. |
| symbolsMap | dataframe, an K x 2 matrix stores the gene symbol of each Entrez Gene ID. It is used for the heatmap visualization. The order of rows should match that of the **voom.results**. Default symbolsMap=NULL. |
| baseGSEAs | character, a vector of the gene set tests that should be included in the ensemble. Type egsea.base to see the supported GSE methods. By default, all supported methods are used. |
| minSize | integer, the minimum size of a gene set to be included in the analysis. Default minSize= 2. |
| display.top | integer, the number of top gene sets to be displayed in the EGSEA report. You can always access the list of all tested gene sets using the returned gsa list. Default is 20. |
| combineMethod | character, determines how to combine p-values from different GSEA method. Type egsea.combine() to see supported methods. |
| combineWeights | double, a vector determines how different GSEA methods will be weighted. Its values should range between 0 and 1. This option is not supported currently. |
| sort.by | character, determines how to order the analysis results in the stats table. Type egsea.sort() to see all available options. |
| egsea.dir | character, directory into which the analysis results are written out. |
| kegg.dir | character, the directory of KEGG pathway data file (.xml) and image file (.png). Default kegg.dir=paste0(egsea.dir, "/kegg-dir/"). |
| logFC.cutoff | numeric, cut-off threshold of logFC and is used for Sginificance Score and Regulation Direction Calculations. Default logFC.cutoff=0. |

| | |
|---|---|
| `sum.plot.axis` | character, the x-axis of the summary plot. All the values accepted by the **sort.by** parameter can be used. Default sum.plot.axis="p.value". |
| `sum.plot.cutoff` | numeric, cut-off threshold to filter the gene sets of the summary plots based on the values of the **sum.plot.axis**. Default sum.plot.cutoff=NULL. |
| `vote.bin.width` | numeric, the bin width of the vote ranking. Default vote.bin.width=5. |
| `num.threads` | numeric, number of CPU threads to be used. Default num.threads=2. |
| `report` | logical, whether to generate the EGSEA interactive report. It takes longer time to run. Default is True. |
| `print.base` | logical, whether to write out the results of the individual GSE methods. Default FALSE. |
| `verbose` | logical, whether to print out progress messages and warnings. |

## Details

EGSEA, an acronym for *Ensemble of Gene Set Enrichment Analyses*, utilizes the analysis results of eleven prominent GSE algorithms from the literature to calculate collective significance scores for gene sets. These methods include: **ora**, **globaltest**, **plage**, **safe**, **zscore**, **gage**, **ssgsea**, **roast**, **padog**, **camera** and **gsva**. The ora, gage, camera and gsva methods depend on a competitive null hypothesis while the remaining seven methods are based on a self-contained hypothesis. Conveniently, the algorithm proposed here is not limited to these eleven GSE methods and new GSE tests can be easily integrated into the framework. This function takes the voom object and the contrast matrix as parameters.

## Value

A list of elements, each with two/three elements that store the top gene sets and the detailed analysis results for each contrast and the comparative analysis results.

## References

Monther Alhamdoosh, Milica Ng, Nicholas J. Wilson, Julie M. Sheridan, Huy Huynh, Michael J. Wilson and Matthew E. Ritchie. Combining multiple tools outperforms individual methods in gene set enrichment analyses.

## See Also

`egsea.base`, `egsea.sort`, `buildIdxEZID`, `buildMSigDBIdxEZID`, `buildKEGGIdxEZID`, `buildGeneSetDBIdxEZID`, and `buildCustomIdxEZID`

## Examples

```
library(EGSEAdata)
data(il13.data)
v = il13.data$voom
contrasts = il13.data$contra
gs.annots = buildIdxEZID(entrezIDs=rownames(v$E), species="human",
msigdb.gsets="none",
```

```
         kegg.updated=FALSE, kegg.exclude = c("Metabolism"))
# set report = TRUE to generate the EGSEA interactive report
gsa = egsea(voom.results=v, contrasts=contrasts,  gs.annots=gs.annots,
         symbolsMap=v$genes,
baseGSEAs=egsea.base()[-c(2,5,6,9)], display.top = 5,
          sort.by="avg.rank", egsea.dir="./il13-egsea-report",
         num.threads = 2, report = FALSE)
```

---

`egsea.base` *EGSEA Base GSE Methods*

---

## Description

It lists the supported GSEA methods

## Usage

```
egsea.base()
```

## Value

It returns a character vector of supported GSE methods.

## Examples

```
egsea.base()
```

---

`egsea.cnt` *Ensemble of Gene Set Enrichment Analyses Function*

---

## Description

This is the main function to carry out gene set enrichment analysis using the EGSEA algorithm. This function is aimed to use the raw count matrix to perform the EGSEA analysis.

## Usage

```
egsea.cnt(counts, group, design = NULL, contrasts, logFC = NULL, gs.annots,
  symbolsMap = NULL, baseGSEAs = egsea.base(), minSize = 2,
  display.top = 20, combineMethod = "fisher", combineWeights = NULL,
  sort.by = "p.adj", egsea.dir = "./", kegg.dir = NULL,
  logFC.cutoff = 0, sum.plot.axis = "p.adj", sum.plot.cutoff = NULL,
  vote.bin.width = 5, num.threads = 4, report = TRUE,
  print.base = FALSE, verbose = FALSE)
```

## Arguments

| | |
|---|---|
| counts | double, numeric matrix of read counts where genes are the rows and samples are the columns. |
| group | character, vector or factor giving the experimental group/condition for each sample/library |
| design | double, numeric matrix giving the design matrix of the linear model fitting. |
| contrasts | double, an N x L matrix indicates the contrast of the linear model coefficients for which the test is required. N is number of experimental conditions and L is number of contrasts. |
| logFC | double, an K x L matrix indicates the log2 fold change of each gene for each contrast. K is the number of genes included in the analysis. If logFC=NULL, the logFC values are estimated using the eBayes for each contrast. |
| gs.annots | list, indexed collections of gene sets. It is generated using one of these functions: buildIdxEZID, buildMSigDBIdxEZID, buildKEGGIdxEZID, buildGeneSetDBIdxEZID, and buildCustomIdxEZID. |
| symbolsMap | dataframe, an K x 2 matrix stores the gene symbol of each Entrez Gene ID. It is used for the heatmap visualization. The order of rows should match that of the **counts**. Default symbolsMap=NULL. |
| baseGSEAs | character, a vector of the gene set tests that should be included in the ensemble. Type egsea.base to see the supported GSE methods. By default, all supported methods are used. |
| minSize | integer, the minimum size of a gene set to be included in the analysis. Default minSize= 2. |
| display.top | integer, the number of top gene sets to be displayed in the EGSEA report. You can always access the list of all tested gene sets using the returned gsa list. Default is 20. |
| combineMethod | character, determines how to combine p-values from different GSEA method. Type egsea.combine() to see supported methods. |
| combineWeights | double, a vector determines how different GSEA methods will be weighted. Its values should range between 0 and 1. This option is not supported currently. |
| sort.by | character, determines how to order the analysis results in the stats table. Type egsea.sort() to see all available options. |
| egsea.dir | character, directory into which the analysis results are written out. |
| kegg.dir | character, the directory of KEGG pathway data file (.xml) and image file (.png). Default kegg.dir=paste0(egsea.dir, "/kegg-dir/"). |
| logFC.cutoff | numeric, cut-off threshold of logFC and is used for Sginificance Score and Regulation Direction Calculations. Default logFC.cutoff=0. |
| sum.plot.axis | character, the x-axis of the summary plot. All the values accepted by the **sort.by** parameter can be used. Default sum.plot.axis="p.value". |
| sum.plot.cutoff | numeric, cut-off threshold to filter the gene sets of the summary plots based on the values of the **sum.plot.axis**. Default sum.plot.cutoff=NULL. |
| vote.bin.width | numeric, the bin width of the vote ranking. Default vote.bin.width=5. |

| | |
|---|---|
| num.threads | numeric, number of CPU threads to be used. Default num.threads=2. |
| report | logical, whether to generate the EGSEA interactive report. It takes longer time to run. Default is True. |
| print.base | logical, whether to write out the results of the individual GSE methods. Default FALSE. |
| verbose | logical, whether to print out progress messages and warnings. |

## Details

EGSEA, an acronym for *Ensemble of Gene Set Enrichment Analyses*, utilizes the analysis results of eleven prominent GSE algorithms from the literature to calculate collective significance scores for gene sets. These methods include: **ora**, **globaltest**, **plage**, **safe**, **zscore**, **gage**, **ssgsea**, **roast**, **padog**, **camera** and **gsva**. The ora, gage, camera and gsva methods depend on a competitive null hypothesis while the remaining seven methods are based on a self-contained hypothesis. Conveniently, the algorithm proposed here is not limited to these eleven GSE methods and new GSE tests can be easily integrated into the framework. This function takes the raw count matrix, the experimental group of each sample, the design matrix and the contrast matrix as parameters. It performs TMM normalization and then applies voom to calculate the logCPM and weighting factors.

## Value

A list of elements, each with two/three elements that store the top gene sets and the detailed analysis results for each contrast and the comparative analysis results.

## References

Monther Alhamdoosh, Milica Ng, Nicholas J. Wilson, Julie M. Sheridan, Huy Huynh, Michael J. Wilson and Matthew E. Ritchie. Combining multiple tools outperforms individual methods in gene set enrichment analyses.

## See Also

egsea.base, egsea.sort, buildIdxEZID, buildMSigDBIdxEZID, buildKEGGIdxEZID, buildGeneSetDBIdxEZID, and buildCustomIdxEZID

## Examples

```
library(EGSEAdata)
data(il13.data.cnt)
cnt = il13.data.cnt$counts
group = il13.data.cnt$group
design = il13.data.cnt$design
contrasts = il13.data.cnt$contra
genes = il13.data.cnt$genes
gs.annots = buildIdxEZID(entrezIDs=rownames(cnt), species="human",
msigdb.gsets="none",
         kegg.updated=FALSE, kegg.exclude = c("Metabolism"))
# set report = TRUE to generate the EGSEA interactive report
gsa = egsea.cnt(counts=cnt, group=group, design=design, contrasts=contrasts,
         gs.annots=gs.annots,
```

```
         symbolsMap=genes, baseGSEAs=egsea.base()[-c(2,5,6,9)],
display.top = 5,
          sort.by="avg.rank",
egsea.dir="./il13-egsea-cnt-report",
         num.threads = 2, report = FALSE)
```

---

`egsea.combine` *EGSEA P-value Combining Options*

---

## Description

It lists the p-value combining methods

## Usage

```
egsea.combine()
```

## Value

It returns a character vector of available methods for the combineMethod argument in egsea

## Examples

```
egsea.combine()
```

---

`egsea.ora` *Over-representation Analysis with EGSEA Reporting Capabilities*

---

## Description

This is the main function to carry out gene set enrichment analysis using the over-representation analysis (ORA) only.

## Usage

```
egsea.ora(entrezIDs, universe = NULL, logFC = NULL, title = NULL,
  gs.annots, symbolsMap = NULL, minSize = 2, display.top = 20,
  sort.by = "p.adj", egsea.dir = "./", kegg.dir = NULL,
  logFC.cutoff = 0, sum.plot.axis = "p.adj", sum.plot.cutoff = NULL,
  vote.bin.width = 5, num.threads = 4, report = TRUE,
  print.base = FALSE, verbose = FALSE)
```

## Arguments

| | |
|---|---|
| entrezIDs | character, a vector of Entrez Gene IDs to be tested for ORA. |
| universe | character, a vector of Enterz IDs to be used as a background list. If universe=NULL, the background list is created from the **AnnotationDbi** package. |
| logFC | double, is a matrix or vector of log fold changes of the same length of entrezIDs. If logFC=NULL, 1 is used as a default value. Then, the regulation direction in heatmaps and pathway maps is not indicative to the gene regulation direction. |
| title | character, a short description of the experimental contrast. |
| gs.annots | list, indexed collections of gene sets. It is generated using one of these functions: buildIdxEZID, buildMSigDBIdxEZID, buildKEGGIdxEZID, buildGeneSetDBIdxEZID, and buildCustomIdxEZID. |
| symbolsMap | dataframe, an K x 2 matrix stores the gene symbol of each Entrez Gene ID. It is used for the heatmap visualization. The order of rows should match that of the **entrezIDs**. Default symbolsMap=NULL. |
| minSize | integer, the minimum size of a gene set to be included in the analysis. Default minSize= 2. |
| display.top | integer, the number of top gene sets to be displayed in the EGSEA report. You can always access the list of all tested gene sets using the returned gsa list. Default is 20. |
| sort.by | character, determines how to order the analysis results in the stats table. It takes "p.value", "p.adj" or "Significance". |
| egsea.dir | character, directory into which the analysis results are written out. |
| kegg.dir | character, the directory of KEGG pathway data file (.xml) and image file (.png). Default kegg.dir=paste0(egsea.dir, "/kegg-dir/"). |
| logFC.cutoff | numeric, cut-off threshold of logFC and is used for Sginificance Score and Regulation Direction Calculations. Default logFC.cutoff=0. |
| sum.plot.axis | character, the x-axis of the summary plot. All the values accepted by the **sort.by** parameter can be used. Default sum.plot.axis="p.adj". |
| sum.plot.cutoff | numeric, cut-off threshold to filter the gene sets of the summary plots based on the values of the **sum.plot.axis**. Default sum.plot.cutoff=NULL. |
| vote.bin.width | numeric, the bin width of the vote ranking. Default vote.bin.width=5. |
| num.threads | numeric, number of CPU threads to be used. Default num.threads=2. |
| report | logical, whether to generate the EGSEA interactive report. It takes longer time to run. Default is True. |
| print.base | logical, whether to write out the results of the individual GSE methods. Default FALSE. |
| verbose | logical, whether to print out progress messages and warnings. |

## Details

This function takes a list of Entrez gene IDs and uses the gene set collections from **EGSEAdata** or a custom-built collection to find over-represented gene sets in this list. It takes the advantage of the existing EGSEA reporting capabilities and generate an interative report for the ORA analysis.

## Value

A list of elements, each with two/three elements that store the top gene sets and the detailed analysis results for each contrast and the comparative analysis results.

## References

Monther Alhamdoosh, Milica Ng, Nicholas J. Wilson, Julie M. Sheridan, Huy Huynh, Michael J. Wilson and Matthew E. Ritchie. Combining multiple tools outperforms individual methods in gene set enrichment analyses.

## See Also

buildIdxEZID, buildMSigDBIdxEZID, buildKEGGIdxEZID, buildGeneSetDBIdxEZID, and buildCustomIdxEZID

## Examples

```
library(EGSEAdata)
data(il13.data)
voom.results = il13.data$voom
contrast = il13.data$contra
library(limma)
vfit = lmFit(voom.results, voom.results$design)
vfit = contrasts.fit(vfit, contrast)
vfit = eBayes(vfit)
top.Table = topTable(vfit, coef=1, number=Inf, p.value=0.05, lfc=1)
deGenes = as.character(top.Table$FeatureID)
logFC =  top.Table$logFC
names(logFC) = deGenes
gs.annots = buildIdxEZID(entrezIDs=deGenes, species="human",
msigdb.gsets="none",
         kegg.updated=FALSE, kegg.exclude = c("Metabolism"))
# set report = TRUE to generate the EGSEA interactive report
gsa = egsea.ora(entrezIDs=deGenes, universe=
as.character(voom.results$genes[,1]),
             logFC =logFC, title="X24IL13-X24",
gs.annots=gs.annots,
             symbolsMap=top.Table[, c(1,2)], display.top = 5,
              egsea.dir="./il13-egsea-ora-report", num.threads = 2,
report = FALSE)
```

---

`egsea.sort` *EGSEA Result Sorting Options*

---

## Description

It lists the accepted sorting methods for analysis results

## Usage

```
egsea.sort()
```

## Value

It returns a character vector of the accepted values for the sort.by argument in egsea

## Examples

```
egsea.sort()
```

---

topSets *Table of Top Gene Sets from an EGSEA Analysis*

---

## Description

Extract a table of the top-ranked gene sets from an EGSEA analysis.

## Usage

```
topSets(gsa, contrast = 1, gs.label = 1, sort.by = NULL, number = 10,
  names.only = TRUE)
```

## Arguments

| | |
|---|---|
| gsa | list, the analysis result object from egsea, egsea.cnt or egsea.ora. |
| contrast | contrast column number or column name specifying which contrast is of interest. if contrast = 0 or "comparison" and the number of contrasts greater than 1, the comparative gene sets are retruned. |
| gs.label | the number or label of the gene set collection of interest. |
| sort.by | character, determines how to order the analysis results in the stats table. The accepted values depend on the function used to generate the EGSEA results. |
| number | integer, maximum number of gene sets to list |
| names.only | logical, whether to display the EGSEA statistics or not. |

## Value

A dataframe of top gene sets with the calculated statistics for each if names.only = FALSE.

## Examples

```
library(EGSEAdata)
data(il13.data)
v = il13.data$voom
contrasts = il13.data$contra
gs.annots = buildIdxEZID(entrezIDs=rownames(v$E), species="human",
msigdb.gsets="none",
         kegg.updated=FALSE, kegg.exclude = c("Metabolism"))
gsa = egsea(voom.results=v, contrasts=contrasts,  gs.annots=gs.annots,
         symbolsMap=v$genes,
baseGSEAs=egsea.base()[-c(2,5,6,9)], display.top = 5,
          sort.by="avg.rank", egsea.dir="./il13-egsea-report",
         num.threads = 2, report=FALSE)
topSets(gsa, contrast=1, gs.label="kegg", number = 10)
topSets(gsa, contrast=1, gs.label=1, sort.by="ora", number = 10,
names.only=FALSE)
topSets(gsa, contrast=0, gs.label="kegg", number = 10)
```

# Index