Package 'cTRAP'

October 17, 2020

Title Identification of candidate causal perturbations from differential gene expression data

Version 1.6.1

Description Compare differential gene expression results with those from known cellular perturbations (such as gene knock-down, overexpression or small molecules) derived from the Connectivity Map. Such analyses allow not only to infer the molecular causes of the observed difference in gene expression but also to identify small molecules that could drive or revert specific transcriptomic alterations.

Depends R (>= 3.6.0)

License MIT + file LICENSE

Encoding UTF-8

LazyData true

biocViews DifferentialExpression, GeneExpression, RNASeq, Transcriptomics, Pathways, ImmunoOncology, GeneSetEnrichment

URL https://nuno-agostinho.github.io/cTRAP, https://github.com/nuno-agostinho/cTRAP

BugReports https://github.com/nuno-agostinho/cTRAP/issues

Suggests testthat, knitr, covr, rmarkdown, spelling

RoxygenNote 7.1.1

Imports biomaRt, cowplot, data.table, dplyr, fgsea, ggplot2, ggrepel, graphics, httr, limma, methods, pbapply, R.utils, readxl, reshape2, rhdf5, scales, stats, tools, utils

VignetteBuilder knitr

Language en-GB

git_url https://git.bioconductor.org/packages/cTRAP

git_branch RELEASE_3_11

git_last_commit 3570b03

git_last_commit_date 2020-08-17

Date/Publication 2020-10-16

Author Bernardo P. de Almeida [aut], Nuno Saraiva-Agostinho [aut, cre], Nuno L. Barbosa-Morais [aut, led]

Maintainer Nuno Saraiva-Agostinho <nunodanielagostinho@gmail.com>

R topics documented:

analyseDrugSetEnrichment . . . 2
as.table.referenceComparison . . . 4
convertENSEMBLtoGeneSymbols . . . 4
cTRAP . . . 5
downloadENCODEknockdownMetadata . . . 6
filterCMapMetadata . . . 6
getCMapConditions . . . 7
getCMapPerturbationTypes . . . 8
listExpressionDrugSensitivityAssociation . . . 9
loadCMapData . . . 10
loadCMapZscores . . . 11
loadDrugDescriptors . . . 11
loadENCODEsamples . . . 12
loadExpressionDrugSensitivityAssociation . . . 13
parseCMapID . . . 14
performDifferentialExpression . . . 14
plot.perturbationChanges . . . 15
plot.referenceComparison . . . 17
plotDrugSetEnrichment . . . 19
plotTargetingDrugsVSsimilarPerturbations . . . 20
predictTargetingDrugs . . . 21
prepareCMapPerturbations . . . 22
prepareDrugSets . . . 24
prepareENCODEgeneExpression . . . 24
print.similarPerturbations . . . 25
rankSimilarPerturbations . . . 26

Index **28**

`analyseDrugSetEnrichment` *Analyse drug set enrichment*

Description

Analyse drug set enrichment

Usage

```
analyseDrugSetEnrichment(
  sets,
  stats,
  col = NULL,
  nperm = 10000,
  maxSize = 500,
  ...
)
```

Arguments

sets	Named list of characters: named sets containing compound identifiers (obtain drug sets by running prepareDrugSets())
stats	Named numeric vector or either a similarPerturbations or a targetingDrugs object (obtained after running rankSimilarPerturbations or predictTargetingDrugs, respectively)
col	Character: name of the column to use for statistics (only required if class of stats is either similarPerturbations or targetingDrugs)
nperm	Number of permutations to do. Minimial possible nominal p-value is about 1/nperm
maxSize	Maximal size of a gene set to test. All pathways above the threshold are excluded.
...	Arguments passed on to fgsea::fgsea minSize Minimal size of a gene set to test. All pathways below 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

Enrichment analysis based on GSEA

See Also

Other functions for drug set enrichment analysis: loadDrugDescriptors(), plotDrugSetEnrichment(), prepareDrugSets()

Examples

```
descriptors <- loadDrugDescriptors()
drugSets <- prepareDrugSets(descriptors)

# Analyse drug set enrichment in ranked targeting drugs for a differential
# expression profile
data("diffExprStat")
gdsc      <- loadExpressionDrugSensitivityAssociation("GDSC")
predicted <- predictTargetingDrugs(diffExprStat, gdsc)

analyseDrugSetEnrichment(drugSets, predicted)
```

`as.table.referenceComparison`

Cross Tabulation and Table Creation

Description

Cross Tabulation and Table Creation

Usage

```
## S3 method for class 'referenceComparison'
as.table(x, ..., clean = TRUE)
```

Arguments

`x`	`referenceComparison` object
`...`	Extra parameters not currently used
`clean`	Boolean: only show certain columns (to avoid redundancy)?

Value

Complete table with metadata based on a `targetingDrugs` object

See Also

Other functions related with the ranking of CMap perturbations: `filterCMapMetadata()`, `getCMapConditions()`, `getCMapPerturbationTypes()`, `loadCMapData()`, `loadCMapZscores()`, `parseCMapID()`, `plot.perturbationChang`
`plot.referenceComparison()`, `plotTargetingDrugsVSsimilarPerturbations()`, `prepareCMapPerturbations()`
`print.similarPerturbations()`, `rankSimilarPerturbations()`

Other functions related with the prediction of targeting drugs: `listExpressionDrugSensitivityAssociation()`, `loadExpressionDrugSensitivityAssociation()`, `plot.referenceComparison()`, `plotTargetingDrugsVSsimila`
`predictTargetingDrugs()`

`convertENSEMBLtoGeneSymbols`

Convert ENSEMBL gene identifiers to gene symbols

Description

Convert ENSEMBL gene identifiers to gene symbols

Usage

```
convertENSEMBLtoGeneSymbols(
  genes,
  dataset = "hsapiens_gene_ensembl",
  mart = "ensembl"
)
```

Arguments

genes	Character: ENSEMBL gene identifiers
dataset	Character: biomaRt dataset name
mart	Character: biomaRt database name

Value

Named character vector where names are the input ENSEMBL gene identifiers and the values are the matching gene symbols

---

cTRAP *cTRAP package*

Description

Compare differential gene expression results with those from big datasets (e.g. CMap), allowing to infer which types of perturbations may explain the observed difference in gene expression.

Details

Input: To use this package, a named vector of differentially expressed gene metric is needed, where its values represent the significance and magnitude of the differentially expressed genes (e.g. t-statistic) and its names are gene symbols.

Workflow: The differentially expressed genes will be compared against selected perturbation conditions by:

- Spearman or Pearson correlation with z-scores of differentially expressed genes after perturbations from CMap. Use function `rankSimilarPerturbations` with `method = "spearman"` or `method = "pearson"`
- Gene set enrichment analysis (GSEA) using the (around) 12 000 genes from CMap. Use function `rankSimilarPerturbations` with `method = gsea`.

Available perturbation conditions for CMap include:

- Cell line(s).
- Perturbation type (gene knockdown, gene upregulation or drug intake).
- Drug concentration.
- Time points.

Values for each perturbation type can be listed with `getCMapPerturbationTypes()`

Output: The output includes a data frame of ranked perturbations based on the associated statistical values and respective p-values.

downloadENCODEknockdownMetadata *Download metadata for ENCODE knockdown experiments*

Description

Download metadata for ENCODE knockdown experiments

Usage

```
downloadENCODEknockdownMetadata(cellLine = NULL, gene = NULL)
```

Arguments

cellLine	Character: cell line
gene	Character: target gene

Value

Data frame containing ENCODE knockdown experiment metadata

See Also

Other functions related with using ENCODE expression data: loadENCODEsamples(), performDifferentialExpressi
prepareENCODEgeneExpression()

Examples

```
downloadENCODEknockdownMetadata("HepG2", "EIF4G1")
```

filterCMapMetadata *Filter CMap metadata*

Description

Filter CMap metadata

Usage

```
filterCMapMetadata(
  metadata,
  cellLine = NULL,
  timepoint = NULL,
  dosage = NULL,
  perturbationType = NULL
)
```

Arguments

metadata	Data frame (CMap metadata) or character (respective filepath)
cellLine	Character: cell line (if NULL, all values are loaded)
timepoint	Character: timepoint (if NULL, all values are loaded)
dosage	Character: dosage (if NULL, all values are loaded)
perturbationType	Character: type of perturbation (if NULL, all perturbation types are loaded)

Value

Filtered CMap metadata

See Also

Other functions related with the ranking of CMap perturbations: as.table.referenceComparison(), getCMapConditions(), getCMapPerturbationTypes(), loadCMapData(), loadCMapZscores(), parseCMapID(), plot.perturbationChanges(), plot.referenceComparison(), plotTargetingDrugsVSsimilarP prepareCMapPerturbations(), print.similarPerturbations(), rankSimilarPerturbations()

Examples

```
## Not run:
cmapMetadata <- loadCMapData("cmapMetadata.txt", "metadata")

## End(Not run)
filterCMapMetadata(cmapMetadata, cellLine="HEPG2", timepoint="2 h",
                   dosage="25 ng/mL")
```

---

getCMapConditions *List available conditions in CMap datasets*

Description

Downloads metadata if not available

Usage

```
getCMapConditions(
  metadata,
  cellLine = NULL,
  timepoint = NULL,
  dosage = NULL,
  perturbationType = NULL,
  control = FALSE
)
```

Arguments

metadata	Data frame (CMap metadata) or character (respective filepath)
cellLine	Character: cell line (if NULL, all values are loaded)
timepoint	Character: timepoint (if NULL, all values are loaded)
dosage	Character: dosage (if NULL, all values are loaded)
perturbationType	Character: type of perturbation (if NULL, all perturbation types are loaded)
control	Boolean: show controls for perturbation types?

Value

List of conditions in CMap datasets

See Also

Other functions related with the ranking of CMap perturbations: as.table.referenceComparison(), filterCMapMetadata(), getCMapPerturbationTypes(), loadCMapData(), loadCMapZscores(), parseCMapID(), plot.perturbationChanges(), plot.referenceComparison(), plotTargetingDrugsVSsimilarP prepareCMapPerturbations(), print.similarPerturbations(), rankSimilarPerturbations()

Examples

```
## Not run:
cmapMetadata <- loadCMapData("cmapMetadata.txt", "metadata")

## End(Not run)
getCMapConditions(cmapMetadata)
```

---

getCMapPerturbationTypes

Get CMap perturbation types

Description

Get CMap perturbation types

Usage

```
getCMapPerturbationTypes(control = FALSE)
```

Arguments

control	Boolean: return perturbation types used as control?

Value

Perturbation types and respective codes as used by CMap datasets

See Also

Other functions related with the ranking of CMap perturbations: as.table.referenceComparison(), filterCMapMetadata(), getCMapConditions(), loadCMapData(), loadCMapZscores(), parseCMapID(), plot.perturbationChanges(), plot.referenceComparison(), plotTargetingDrugsVSsimilarPerturbations() prepareCMapPerturbations(), print.similarPerturbations(), rankSimilarPerturbations()

Examples

```
getCMapPerturbationTypes()
```

---

listExpressionDrugSensitivityAssociation

List available gene expression and drug sensitivity correlation matrices

---

Description

List available gene expression and drug sensitivity correlation matrices

Usage

```
listExpressionDrugSensitivityAssociation(url = FALSE)
```

Arguments

url	Boolean: return download link?

Value

Character vector of available gene expression and drug sensitivity correlation matrices

See Also

Other functions related with the prediction of targeting drugs: as.table.referenceComparison(), loadExpressionDrugSensitivityAssociation(), plot.referenceComparison(), plotTargetingDrugsVSsimila predictTargetingDrugs()

Examples

```
listExpressionDrugSensitivityAssociation()
```

`loadCMapData` *Load CMap data*

Description

Load CMap data (if not found, `file` will be automatically downloaded)

Usage

```
loadCMapData(
  file,
  type = c("metadata", "geneInfo", "zscores", "compoundInfo"),
  zscoresID = NULL
)
```

Arguments

`file`	Character: path to file
`type`	Character: type of data to load (`metadata`, `geneInfo`, `zscores` or `compoundInfo`)
`zscoresID`	Character: identifiers to partially load z-scores file (for performance reasons; if `NULL`, all identifiers will be loaded)

Value

Metadata as a data table

Note

If `type = "compoundInfo"`, two files from **The Drug Repurposing Hub** will be downloaded containing information about drugs and perturbations. The files will be named `file` with `_drugs` and `_samples` before their extension, respectively.

See Also

Other functions related with the ranking of CMap perturbations: `as.table.referenceComparison()`, `filterCMapMetadata()`, `getCMapConditions()`, `getCMapPerturbationTypes()`, `loadCMapZscores()`, `parseCMapID()`, `plot.perturbationChanges()`, `plot.referenceComparison()`, `plotTargetingDrugsVSsimilarP`
`prepareCMapPerturbations()`, `print.similarPerturbations()`, `rankSimilarPerturbations()`

Examples

```
# Load CMap metadata (data is automatically downloaded if not available)
cmapMetadata <- loadCMapData("cmapMetadata.txt", "metadata")

# Load CMap gene info
loadCMapData("cmapGeneInfo.txt", "geneInfo")

# Load CMap zscores based on filtered metadata
cmapMetadataKnockdown <- filterCMapMetadata(
  cmapMetadata, cellLine="HepG2",
  perturbationType="Consensus signature from shRNAs targeting the same gene")
loadCMapData("cmapZscores.gctx.gz", "zscores", cmapMetadataKnockdown$sig_id)
```

loadCMapZscores — *Load matrix of CMap perturbation's differential expression z-scores (optional)*

Description

Load matrix of CMap perturbation's differential expression z-scores (optional)

Usage

```
loadCMapZscores(data, inheritAttrs = FALSE, verbose = TRUE)
```

Arguments

data	perturbationChanges object
inheritAttrs	Boolean: convert to perturbationChanges object and inherit attributes from data?
verbose	Boolean: print messages?

Value

Matrix containing CMap perturbation z-scores (genes as rows, perturbations as columns)

See Also

Other functions related with the ranking of CMap perturbations: as.table.referenceComparison(), filterCMapMetadata(), getCMapConditions(), getCMapPerturbationTypes(), loadCMapData(), parseCMapID(), plot.perturbationChanges(), plot.referenceComparison(), plotTargetingDrugsVSsimilarP prepareCMapPerturbations(), print.similarPerturbations(), rankSimilarPerturbations()

Examples

```
metadata <- loadCMapData("cmapMetadata.txt", "metadata")
metadata <- filterCMapMetadata(metadata, cellLine="HepG2")
perts <- prepareCMapPerturbations(metadata, "cmapZscores.gctx",
                                  "cmapGeneInfo.txt")
zscores <- loadCMapZscores(perts[ , 1:10])
```

loadDrugDescriptors — *Load table with drug descriptors*

Description

Load table with drug descriptors

Usage

```
loadDrugDescriptors(
  source = c("NCI60", "CMap"),
  type = c("2D", "3D"),
  file = NULL
)
```

Arguments

source	Character: molecular descriptors for compounds in NCI60 or CMap
type	Character: load 2D or 3D molecular descriptors
file	Character: filepath to drug descriptors (automatically downloaded if file does not exist)

Value

Data table with drug descriptors

See Also

Other functions for drug set enrichment analysis: analyseDrugSetEnrichment(), plotDrugSetEnrichment(), prepareDrugSets()

Examples

```
loadDrugDescriptors()
```

---

loadENCODEsamples *Load ENCODE samples*

Description

Samples are automatically downloaded if they are not found in the current working directory.

Usage

```
loadENCODEsamples(metadata)
```

Arguments

metadata	Character: ENCODE metadata

Value

List of loaded ENCODE samples

See Also

Other functions related with using ENCODE expression data: downloadENCODEknockdownMetadata(), performDifferentialExpression(), prepareENCODEgeneExpression()

Examples

```
if (interactive()) {
  # Load ENCODE metadata for a specific cell line and gene
  cellLine <- "HepG2"
  gene <- c("EIF4G1", "U2AF2")
  ENCODEmetadata <- downloadENCODEknockdownMetadata(cellLine, gene)

  # Load samples based on filtered ENCODE metadata
  loadENCODEsamples(ENCODEmetadata)
}
```

`loadExpressionDrugSensitivityAssociation`
Load gene expression and drug sensitivity correlation matrix

Description

Load gene expression and drug sensitivity correlation matrix

Usage

```
loadExpressionDrugSensitivityAssociation(source, file = NULL)
```

Arguments

`source`	Character: source of matrix to load; see `listExpressionDrugSensitivityAssociation`
`file`	Character: filepath to gene expression and drug sensitivity association dataset (automatically downloaded if file does not exist)

Value

Correlation matrix between gene expression (rows) and drug sensitivity (columns)

See Also

Other functions related with the prediction of targeting drugs: `as.table.referenceComparison()`, `listExpressionDrugSensitivityAssociation()`, `plot.referenceComparison()`, `plotTargetingDrugsVSsimila`
`predictTargetingDrugs()`

Examples

```
gdsc <- listExpressionDrugSensitivityAssociation()[[1]]
loadExpressionDrugSensitivityAssociation(gdsc)
```

parseCMapID *Parse CMap identifier*

Description

Parse CMap identifier

Usage

```
parseCMapID(id, cellLine = FALSE)
```

Arguments

id	Character: CMap identifier
cellLine	Boolean: if TRUE, return cell line information from CMap identifier; else, return the CMap identifier without the cell line

Value

Character vector with information from CMap identifiers

See Also

Other functions related with the ranking of CMap perturbations: as.table.referenceComparison(), filterCMapMetadata(), getCMapConditions(), getCMapPerturbationTypes(), loadCMapData(), loadCMapZscores(), plot.perturbationChanges(), plot.referenceComparison(), plotTargetingDrugsVSsimi prepareCMapPerturbations(), print.similarPerturbations(), rankSimilarPerturbations()

Examples

```
id <- c("CVD001_HEPG2_24H:BRD-K94818765-001-01-0:4.8",
        "CVD001_HEPG2_24H:BRD-K96188950-001-04-5:4.3967",
        "CVD001_HUH7_24H:BRD-A14014306-001-01-1:4.1")
parseCMapID(id, cellLine=TRUE)
parseCMapID(id, cellLine=FALSE)
```

performDifferentialExpression *Perform differential gene expression based on ENCODE data*

Description

Perform differential gene expression based on ENCODE data

Usage

```
performDifferentialExpression(counts)
```

Arguments

counts	Data frame: gene expression

Value

Data frame with differential gene expression results between knockdown and control

See Also

Other functions related with using ENCODE expression data: `downloadENCODEknockdownMetadata()`, `loadENCODEsamples()`, `prepareENCODEgeneExpression()`

Examples

```
if (interactive()) {
  # Download ENCODE metadata for a specific cell line and gene
  cellLine <- "HepG2"
  gene <- "EIF4G1"
  ENCODEmetadata <- downloadENCODEknockdownMetadata(cellLine, gene)

  # Download samples based on filtered ENCODE metadata
  ENCODEsamples <- loadENCODEsamples(ENCODEmetadata)[[1]]

  counts <- prepareENCODEgeneExpression(ENCODEsamples)

  # Remove low coverage (at least 10 counts shared across two samples)
  minReads   <- 10
  minSamples <- 2
  filter <- rowSums(counts[ , -c(1, 2)] >= minReads) >= minSamples
  counts <- counts[filter, ]

  # Convert ENSEMBL identifier to gene symbol
  counts$gene_id <- convertENSEMBLtoGeneSymbols(counts$gene_id)

  # Perform differential gene expression analysis
  diffExpr <- performDifferentialExpression(counts)
}
```

---

plot.perturbationChanges

Operations on a `perturbationChanges` *object*

Description

Operations on a `perturbationChanges` object

Usage

```
## S3 method for class 'perturbationChanges'
plot(
  x,
  perturbation,
  input,
  method = c("spearman", "pearson", "gsea"),
  geneSize = 150,
  genes = c("both", "top", "bottom"),
  ...,
```

```
  title = NULL
)

## S3 method for class 'perturbationChanges'
x[i, j, drop = FALSE, ...]

## S3 method for class 'perturbationChanges'
dim(x)

## S3 method for class 'perturbationChanges'
dimnames(x)
```

Arguments

x	perturbationChanges object
perturbation	Character (perturbation identifier) or a similarPerturbations table (from which the respective perturbation identifiers are retrieved)
input	Named numeric vector of differentially expressed genes whose names are gene identifiers and respective values are a statistic that represents significance and magnitude of differentially expressed genes (e.g. t-statistics); or character of gene symbols composing a gene set that is tested for enrichment in reference data (only used if method includes gsea)
method	Character: one or more methods to compare data (spearman, pearson or gsea)
geneSize	Numeric: number of top up-/down-regulated genes to use as gene sets to test for enrichment in reference data; if a 2-length numeric vector, the first index is the number of top up-regulated genes and the second index is the number of down-regulated genes used to create gene sets; only used if method includes gsea and if input is not a gene set
genes	Character: when plotting gene set enrichment analysis (GSEA), plot most up-regulated genes (genes = "top"), most down-regulated genes (genes = "bottom") or both (genes = "both"); only used if method = "gsea" and geneset = NULL
...	Extra arguments
title	Character: plot title (if NULL, the default title depends on the context; ignored when plotting multiple perturbations)
i, j	Character or numeric indexes specifying elements to extract
drop	Boolean: coerce result to the lowest possible dimension?

Value

Subset, plot or return dimensions or names of a perturbationChanges object

See Also

Other functions related with the ranking of CMap perturbations: as.table.referenceComparison(), filterCMapMetadata(), getCMapConditions(), getCMapPerturbationTypes(), loadCMapData(), loadCMapZscores(), parseCMapID(), plot.referenceComparison(), plotTargetingDrugsVSsimilarPerturbati prepareCMapPerturbations(), print.similarPerturbations(), rankSimilarPerturbations()

Examples

```
data("diffExprStat")
data("cmapPerturbationsKD")

compareKD <- rankSimilarPerturbations(diffExprStat, cmapPerturbationsKD)
EIF4G1knockdown <- grep("EIF4G1", compareKD[[1]], value=TRUE)
plot(cmapPerturbationsKD, EIF4G1knockdown, diffExprStat, method="spearman")
plot(cmapPerturbationsKD, EIF4G1knockdown, diffExprStat, method="pearson")
plot(cmapPerturbationsKD, EIF4G1knockdown, diffExprStat, method="gsea")

data("cmapPerturbationsCompounds")
pert <- "CVD001_HEPG2_24H:BRD-A14014306-001-01-1:4.1"
plot(cmapPerturbationsCompounds, pert, diffExprStat, method="spearman")
plot(cmapPerturbationsCompounds, pert, diffExprStat, method="pearson")
plot(cmapPerturbationsCompounds, pert, diffExprStat, method="gsea")

# Multiple cell line perturbations
pert <- "CVD001_24H:BRD-A14014306-001-01-1:4.1"
plot(cmapPerturbationsCompounds, pert, diffExprStat, method="spearman")
plot(cmapPerturbationsCompounds, pert, diffExprStat, method="pearson")
plot(cmapPerturbationsCompounds, pert, diffExprStat, method="gsea")
```

`plot.referenceComparison` *Plot data comparison*

Description

If `element = NULL`, comparison is plotted based on all elements. Otherwise, show scatter or GSEA plots for a single element compared with previously given differential expression results.

Usage

```
## S3 method for class 'referenceComparison'
plot(
  x,
  element = NULL,
  method = c("spearman", "pearson", "gsea", "rankProduct"),
  n = c(3, 3),
  showMetadata = TRUE,
  plotNonRankedPerturbations = FALSE,
  alpha = 0.3,
  genes = c("both", "top", "bottom"),
  ...,
  zscores = NULL,
  title = NULL
)
```

Arguments

x	`referenceComparison` object: obtained after running `rankSimilarPerturbations()` or `predictTargetingDrugs()`

element	Character: identifier in the first column of x
method	Character: method to plot results; choose between spearman, pearson, gsea or rankProduct (the last one is only available if element = NULL)
n	Numeric: number of top and bottom genes to label (if a vector of two numbers is given, the first and second numbers will be used as the number of top and bottom genes to label, respectively); only used if element = NULL
showMetadata	Boolean: show available metadata information instead of identifiers (if available)? Only used if element = NULL
plotNonRankedPerturbations	Boolean: plot non-ranked data in grey? Only used if element = NULL
alpha	Numeric: transparency; only used if element = NULL
genes	Character: when plotting gene set enrichment analysis (GSEA), plot most up-regulated genes (genes = "top"), most down-regulated genes (genes = "bottom") or both (genes = "both"); only used if method = "gsea" and geneset = NULL
...	Extra arguments currently not used
zscores	Data frame (GCTX z-scores) or character (respective filepath to load data from file)
title	Character: plot title (if NULL, the default title depends on the context; ignored when plotting multiple perturbations)

Value

Plot illustrating the reference comparison

See Also

Other functions related with the ranking of CMap perturbations: as.table.referenceComparison(), filterCMapMetadata(), getCMapConditions(), getCMapPerturbationTypes(), loadCMapData(), loadCMapZscores(), parseCMapID(), plot.perturbationChanges(), plotTargetingDrugsVSsimilarPerturbati prepareCMapPerturbations(), print.similarPerturbations(), rankSimilarPerturbations()

Other functions related with the prediction of targeting drugs: as.table.referenceComparison(), listExpressionDrugSensitivityAssociation(), loadExpressionDrugSensitivityAssociation(), plotTargetingDrugsVSsimilarPerturbations(), predictTargetingDrugs()

Examples

```
# Example of a differential expression profile
data("diffExprStat")

## Not run:
# Download and load CMap perturbations to compare with
cellLine <- "HepG2"
cmapMetadataKD <- filterCMapMetadata(
  "cmapMetadata.txt", cellLine=cellLine,
  perturbationType="Consensus signature from shRNAs targeting the same gene")

cmapPerturbationsKD <- prepareCMapPerturbations(
  cmapMetadataKD, "cmapZscores.gctx", "cmapGeneInfo.txt", loadZscores=TRUE)

## End(Not run)
```

```
# Rank similar CMap perturbations
compareKD <- rankSimilarPerturbations(diffExprStat, cmapPerturbationsKD)

# Plot ranked list of CMap perturbations
plot(compareKD, method="spearman")
plot(compareKD, method="spearman", n=c(7, 3))
plot(compareKD, method="pearson")
plot(compareKD, method="gsea")

# Plot results for a single perturbation
pert <- compareKD[[1, 1]]
plot(compareKD, pert, method="spearman", zscores=cmapPerturbationsKD)
plot(compareKD, pert, method="pearson", zscores=cmapPerturbationsKD)
plot(compareKD, pert, method="gsea", zscores=cmapPerturbationsKD)

# Predict targeting drugs based on a given differential expression profile
gdsc <- loadExpressionDrugSensitivityAssociation("GDSC 7")
predicted <- predictTargetingDrugs(diffExprStat, gdsc)

# Plot ranked list of targeting drugs
plot(predicted, method="spearman")
plot(predicted, method="spearman", n=c(7, 3))
plot(predicted, method="pearson")
plot(predicted, method="gsea")

# Plot results for a single targeting drug
drug <- predicted$compound[[4]]
plot(predicted, drug, method="spearman")
plot(predicted, drug, method="pearson")
plot(predicted, drug, method="gsea")
```

`plotDrugSetEnrichment` *Plot drug set enrichment*

Description

Plot drug set enrichment

Usage

```
plotDrugSetEnrichment(
  sets,
  stats,
  col = "rankProduct_rank",
  selectedSets = NULL
)
```

Arguments

`sets`	Named list of characters: named sets containing compound identifiers (obtain drug sets by running `prepareDrugSets()`)
`stats`	Named numeric vector or either a `similarPerturbations` or a `targetingDrugs` object (obtained after running `rankSimilarPerturbations` or `predictTargetingDrugs`, respectively)

col	Character: name of the column to use for statistics (only required if class of stats is either similarPerturbations or targetingDrugs)
selectedSets	Character: drug sets to plot (if NULL, plot all)

Value

List of GSEA plots per drug set

See Also

Other functions for drug set enrichment analysis: analyseDrugSetEnrichment(), loadDrugDescriptors(), prepareDrugSets()

Examples

```
descriptors <- loadDrugDescriptors()
drugSets <- prepareDrugSets(descriptors)

# Analyse drug set enrichment in ranked targeting drugs for a differential
# expression profile
data("diffExprStat")
gdsc      <- loadExpressionDrugSensitivityAssociation("GDSC")
predicted <- predictTargetingDrugs(diffExprStat, gdsc)

plotDrugSetEnrichment(drugSets, predicted)
```

---

plotTargetingDrugsVSsimilarPerturbations

Plot similar perturbations against predicted targeting drugs

Description

Plot similar perturbations against predicted targeting drugs

Usage

```
plotTargetingDrugsVSsimilarPerturbations(
  targetingDrugs,
  similarPerturbations,
  column,
  labelBy = "pert_iname",
  quantileThreshold = 0.25,
  showAllScores = FALSE
)
```

Arguments

targetingDrugs	targetingDrugs object
similarPerturbations	similarPerturbations object
column	Character: column to plot (must be available in both databases)

labelBy — Character: column in similarPerturbations, its metadata or compound information to be used for labelling

quantileThreshold — Numeric: quantile to use for highlight values within [0, 1]

showAllScores — Boolean: show all scores? If FALSE, only the best score per compound will be plotted

Value

ggplot2 plot

See Also

Other functions related with the ranking of CMap perturbations: as.table.referenceComparison(), filterCMapMetadata(), getCMapConditions(), getCMapPerturbationTypes(), loadCMapData(), loadCMapZscores(), parseCMapID(), plot.perturbationChanges(), plot.referenceComparison(), prepareCMapPerturbations(), print.similarPerturbations(), rankSimilarPerturbations()

Other functions related with the prediction of targeting drugs: as.table.referenceComparison(), listExpressionDrugSensitivityAssociation(), loadExpressionDrugSensitivityAssociation(), plot.referenceComparison(), predictTargetingDrugs()

---

predictTargetingDrugs *Predict targeting drugs*

Description

Identify compounds that may target the phenotype associated with a user-provided differential expression profile by comparing such against a correlation matrix of gene expression and drug sensitivity.

Usage

```
predictTargetingDrugs(
  input,
  expressionDrugSensitivityCor,
  method = c("spearman", "pearson", "gsea"),
  geneSize = 150,
  isDrugActivityDirectlyProportionalToSensitivity = NULL
)
```

Arguments

input — Named numeric vector of differentially expressed genes whose names are gene identifiers and respective values are a statistic that represents significance and magnitude of differentially expressed genes (e.g. t-statistics); or character of gene symbols composing a gene set that is tested for enrichment in reference data (only used if method includes gsea)

expressionDrugSensitivityCor — Matrix: correlation matrix of gene expression (rows) and drug sensitivity (columns) across cell lines. Pre-prepared gene expression and drug sensitivity associations are available to download using loadExpressionDrugSensitivityAssociation().

`method` Character: one or more methods to compare data (`spearman`, `pearson` or `gsea`)

`geneSize` Numeric: number of top up-/down-regulated genes to use as gene sets to test for enrichment in reference data; if a 2-length numeric vector, the first index is the number of top up-regulated genes and the second index is the number of down-regulated genes used to create gene sets; only used if `method` includes `gsea` and if `input` is not a gene set

`isDrugActivityDirectlyProportionalToSensitivity` Boolean: are the values used for drug activity directly proportional to drug sensitivity? See details.

Details

If `isDrugActivityDirectlyProportionalToSensitivity = NULL`, the attribute `isDrugMetricDirectlyProportiona` of `expressionDrugSensitivityCor` is used (see `loadExpressionDrugSensitivityAssociation()`).

Value

Data table with correlation or GSEA results comparing differential expression values against gene expression and drug sensitivity associations

GSEA score

Weighted connectivity scores (WTCS) are calculated when `method = "gsea"` (https://clue.io/connectopedia/cmap_algorithms).

See Also

Other functions related with the prediction of targeting drugs: `as.table.referenceComparison()`, `listExpressionDrugSensitivityAssociation()`, `loadExpressionDrugSensitivityAssociation()`, `plot.referenceComparison()`, `plotTargetingDrugsVSsimilarPerturbations()`

Examples

```
# Example of a differential expression profile
data("diffExprStat")

# Load expression and drug sensitivity association derived from GDSC data
gdsc <- loadExpressionDrugSensitivityAssociation("GDSC 7")

# Predict targeting drugs on a differential expression profile
predictTargetingDrugs(diffExprStat, gdsc)
```

---

prepareCMapPerturbations

Prepare CMap perturbation data

Description

Prepare CMap perturbation data

Usage

```
prepareCMapPerturbations(
  metadata,
  zscores,
  geneInfo,
  compoundInfo = NULL,
  loadZscores = FALSE
)
```

Arguments

metadata	Data frame (CMap metadata) or character (respective filepath to load data from file)
zscores	Data frame (GCTX z-scores) or character (respective filepath to load data from file)
geneInfo	Data frame (CMap gene info) or character (respective filepath to load data from file)
compoundInfo	Data frame (CMap compound info) or character (respective filepath to load data from file)
loadZscores	Boolean: load matrix of perturbation z-scores? Not recommended in systems with less than 30GB of RAM; if FALSE, downstream functions will read the file chunk by chunk (this strategy impacts performance at the expense of a much lower memory footprint)

Value

CMap perturbation data attributes and filename

See Also

Other functions related with the ranking of CMap perturbations: as.table.referenceComparison(), filterCMapMetadata(), getCMapConditions(), getCMapPerturbationTypes(), loadCMapData(), loadCMapZscores(), parseCMapID(), plot.perturbationChanges(), plot.referenceComparison(), plotTargetingDrugsVSsimilarPerturbations(), print.similarPerturbations(), rankSimilarPerturbations

Examples

```
## Not run:
metadata <- loadCMapData("cmapMetadata.txt", "metadata")
metadata <- filterCMapMetadata(metadata, cellLine="HepG2")
prepareCMapPerturbations(metadata, "cmapZscores.gctx", "cmapGeneInfo.txt")

## End(Not run)
```

prepareDrugSets *Prepare drug sets from a table with compound descriptors*

Description

Prepare drug sets from a table with compound descriptors

Usage

```
prepareDrugSets(table, id = 1, maxUniqueElems = 15)
```

Arguments

table	Data frame: drug descriptors
id	Integer or character: index or name of the column containing identifiers
maxUniqueElems	Numeric: maximum number of unique elements in a descriptor to consider when creating discrete drug sets

Value

Named list of characters: named drug sets with respective compound identifiers as list elements

See Also

Other functions for drug set enrichment analysis: analyseDrugSetEnrichment(), loadDrugDescriptors(), plotDrugSetEnrichment()

Examples

```
descriptors <- loadDrugDescriptors("NCI60")
prepareDrugSets(descriptors)
```

prepareENCODEgeneExpression *Load an ENCODE gene expression data*

Description

Load an ENCODE gene expression data

Usage

```
prepareENCODEgeneExpression(samples)
```

Arguments

samples	List of loaded ENCODE samples

Value

Data frame containing gene read counts

See Also

convertENSEMBLtoGeneSymbols()

Other functions related with using ENCODE expression data: downloadENCODEknockdownMetadata(), loadENCODEsamples(), performDifferentialExpression()

Examples

```
if (interactive()) {
  # Load ENCODE metadata for a specific cell line and gene
  cellLine <- "HepG2"
  gene <- "EIF4G1"
  ENCODEmetadata <- downloadENCODEknockdownMetadata(cellLine, gene)

  # Load samples based on filtered ENCODE metadata
  ENCODEsamples <- loadENCODEsamples(ENCODEmetadata)[[1]]

  prepareENCODEgeneExpression(ENCODEsamples)
}
```

---

print.similarPerturbations

Print a similarPerturbations *object*

Description

Print a similarPerturbations object

Usage

```
## S3 method for class 'similarPerturbations'
print(x, perturbation = NULL, ...)
```

Arguments

x	similarPerturbations object
perturbation	Character (perturbation identifier) or numeric (perturbation index)
...	Extra parameters passed to print

Value

Information on perturbationChanges object or on specific perturbations

See Also

Other functions related with the ranking of CMap perturbations: as.table.referenceComparison(), filterCMapMetadata(), getCMapConditions(), getCMapPerturbationTypes(), loadCMapData(), loadCMapZscores(), parseCMapID(), plot.perturbationChanges(), plot.referenceComparison(), plotTargetingDrugsVSsimilarPerturbations(), prepareCMapPerturbations(), rankSimilarPerturbations()

`rankSimilarPerturbations` *Rank CMap perturbations' similarity to a differential expression profile*

Description

Compare differential expression results against CMap perturbations.

Usage

```
rankSimilarPerturbations(
  input,
  perturbations,
  method = c("spearman", "pearson", "gsea"),
  geneSize = 150,
  cellLineMean = "auto",
  rankPerCellLine = FALSE
)
```

Arguments

`input`	Named `numeric` `vector` of differentially expressed genes whose names are gene identifiers and respective values are a statistic that represents significance and magnitude of differentially expressed genes (e.g. t-statistics); or `character` of gene symbols composing a gene set that is tested for enrichment in reference data (only used if `method` includes `gsea`)
`perturbations`	`perturbationChanges` object: CMap perturbations (check `prepareCMapPerturbations`)
`method`	Character: comparison method (`spearman`, `pearson` or `gsea`; multiple methods may be selected at once)
`geneSize`	Numeric: number of top up-/down-regulated genes to use as gene sets to test for enrichment in reference data; if a 2-length numeric vector, the first index is the number of top up-regulated genes and the second index is the number of down-regulated genes used to create gene sets; only used if `method` includes `gsea` and if `input` is not a gene set
`cellLineMean`	Boolean: add a column with the mean score across cell lines? If `cellLineMean = "auto"` (default), the mean score will be added when data for more than one cell line is available.
`rankPerCellLine`	Boolean: rank results based on both individual cell lines and mean scores across cell lines (`TRUE`) or based on mean scores alone (`FALSE`)? If `cellLineMean = FALSE`, individual cell line conditions are always ranked.

Value

Data table with correlation or GSEA results comparing differential expression values with those associated with CMap perturbations

GSEA score

Weighted connectivity scores (WTCS) are calculated when `method = "gsea"` (https://clue.io/connectopedia/cmap_algorithms).

See Also

Other functions related with the ranking of CMap perturbations: `as.table.referenceComparison()`, `filterCMapMetadata()`, `getCMapConditions()`, `getCMapPerturbationTypes()`, `loadCMapData()`, `loadCMapZscores()`, `parseCMapID()`, `plot.perturbationChanges()`, `plot.referenceComparison()`, `plotTargetingDrugsVSsimilarPerturbations()`, `prepareCMapPerturbations()`, `print.similarPerturbations`

Examples

```
# Example of a differential expression profile
data("diffExprStat")

## Not run:
# Download and load CMap perturbations to compare with
cellLine <- c("HepG2", "HUH7")
cmapMetadataCompounds <- filterCMapMetadata(
    "cmapMetadata.txt", cellLine=cellLine, timepoint="24 h",
    dosage="5 \u00B5M", perturbationType="Compound")

cmapPerturbationsCompounds <- prepareCMapPerturbations(
    cmapMetadataCompounds, "cmapZscores.gctx", "cmapGeneInfo.txt",
    "cmapCompoundInfo_drugs.txt", loadZscores=TRUE)

## End(Not run)
perturbations <- cmapPerturbationsCompounds

# Rank similar CMap perturbations (by default, Spearman's and Pearson's
# correlation are used, as well as GSEA with the top and bottom 150 genes of
# the differential expression profile used as reference)
rankSimilarPerturbations(diffExprStat, perturbations)

# Rank similar CMap perturbations using only Spearman's correlation
rankSimilarPerturbations(diffExprStat, perturbations, method="spearman")
```

Index

∗ **functions for drug set enrichment analysis**
  analyseDrugSetEnrichment, 2
  loadDrugDescriptors, 11
  plotDrugSetEnrichment, 19
  prepareDrugSets, 24
∗ **functions related with the prediction of targeting drugs**
  as.table.referenceComparison, 4
  listExpressionDrugSensitivityAssociation, 9
  loadExpressionDrugSensitivityAssociation, 13
  plot.referenceComparison, 17
  plotTargetingDrugsVSsimilarPerturbations, 20
  predictTargetingDrugs, 21
∗ **functions related with the ranking of CMap perturbations**
  as.table.referenceComparison, 4
  filterCMapMetadata, 6
  getCMapConditions, 7
  getCMapPerturbationTypes, 8
  loadCMapData, 10
  loadCMapZscores, 11
  parseCMapID, 14
  plot.perturbationChanges, 15
  plot.referenceComparison, 17
  plotTargetingDrugsVSsimilarPerturbations, 20
  prepareCMapPerturbations, 22
  print.similarPerturbations, 25
  rankSimilarPerturbations, 26
∗ **functions related with using ENCODE expression data**
  downloadENCODEknockdownMetadata, 6
  loadENCODEsamples, 12
  performDifferentialExpression, 14
  prepareENCODEgeneExpression, 24
[.perturbationChanges (plot.perturbationChanges), 15

analyseDrugSetEnrichment, 2, *12*, *20*, *24*
as.table.referenceComparison, 4, *7–11*, *13, 14*, *16*, *18*, *21–23*, *25*, *27*

compareAgainstCMap (rankSimilarPerturbations), 26
convertENSEMBLtoGeneSymbols, 4, *25*
cTRAP, 5

dim.perturbationChanges (plot.perturbationChanges), 15
dimnames.perturbationChanges (plot.perturbationChanges), 15
downloadENCODEknockdownMetadata, 6, *12*, *15*, *25*

fgsea::fgsea, *3*
filterCMapMetadata, *4*, 6, *8–11*, *14*, *16*, *18*, *21*, *23*, *25*, *27*

getCMapConditions, *4*, *7*, 7, *9–11*, *14*, *16*, *18*, *21*, *23*, *25*, *27*
getCMapPerturbationTypes, *4*, *7*, *8*, 8, *10, 11*, *14*, *16*, *18*, *21*, *23*, *25*, *27*

listExpressionDrugSensitivityAssociation, *4*, 9, *13*, *18*, *21, 22*
loadCMapData, *4*, *7–9*, 10, *11*, *14*, *16*, *18*, *21*, *23*, *25*, *27*
loadCMapZscores, *4*, *7–10*, 11, *14*, *16*, *18*, *21*, *23*, *25*, *27*
loadDrugDescriptors, *3*, 11, *20*, *24*
loadENCODEsamples, *6*, 12, *15*, *25*
loadExpressionDrugSensitivityAssociation, *4*, *9*, 13, *18*, *21, 22*

parseCMapID, *4*, *7–11*, 14, *16*, *18*, *21*, *23*, *25*, *27*
performDifferentialExpression, *6*, *12*, 14, *25*
plot.perturbationChanges, *4*, *7–11*, *14*, 15, *18*, *21*, *23*, *25*, *27*
plot.referenceComparison, *4*, *7–11*, *13, 14*, *16*, 17, *21–23*, *25*, *27*
plotDrugSetEnrichment, *3*, *12*, 19, *24*
plotTargetingDrugsVSsimilarPerturbations, *4*, *7–11*, *13, 14*, *16*, *18*, 20, *22, 23*, *25*, *27*

predictTargetingDrugs, *3*, *4*, *9*, *13*, *17–19*, *21*, 21
prepareCMapPerturbations, *4*, *7–11*, *14*, *16*, *18*, *21*, 22, *25–27*
prepareDrugSets, *3*, *12*, *20*, 24
prepareENCODEgeneExpression, *6*, *12*, *15*, 24
print.similarPerturbations, *4*, *7–11*, *14*, *16*, *18*, *21*, *23*, 25, *27*

rankSimilarPerturbations, *3–5*, *7–11*, *14*, *16–19*, *21*, *23*, *25*, 26