Package 'directPA'

November 16, 2020

Type Package

Title Direction Analysis for Pathways and Kinases

Version 1.5

Date 2020-11-16

Author Pengyi Yang & Ellis Patrick

Maintainer Pengyi Yang <yangpy7@gmail.com>

Description Direction analysis is a set of tools designed to identify
combinatorial effects of multiple treatments/conditions on pathways
and kinases profiled by microarray, RNA-seq, proteomics, or phosphoproteomics
data. See Yang P et al (2014) <doi:10.1093/bioinformatics/btt616>; and Yang P et al. (2016) <doi:10.1002/pmic.201600068>.

License GPL-3

Depends R (>= 3.10.0),

Imports grDevices, graphics, stats, plotly, calibrate

NeedsCompilation no

Repository CRAN

Date/Publication 2020-11-16 05:10:02 UTC

RoxygenNote 7.1.1

R topics documented:

directPA-package   *Direction Pathway Analysis Package*

Description

The directPA-package is designed for analysing pathways in experiments with multiple perturbations. The combination effects of different treatments are tested by rotating polar coordinates in two-dimentional space when the experiment contains two perturbations and corresponding controls, or spherical coordinates in three-dimensional space when the experiment contains three perturbations and corresponding controls.

Details

Package:	directPA
Type:	Package
Version:	1.0
Date:	2013-10-21
License:	GPL-3

Author(s)

Pengyi Yang <yangpy7@gmail.com> & Ellis Patrick <epatrick87@gmail.com>

References

Pengyi Yang, Ellis Patrick, Shi-Xiong Tan, Daniel J. Fazakerley, James Burchfield, Christopher Gribben, Matthew J. Prior, David E. James, Yee Hwa Yang, Direction pathway analysis of large-scale proteomics data reveals novel features of the insulin action pathway, submitted.

---

directExplorer2d   *Batch Direction Analysis in 2-dimentional space*

Description

Rotate to the direction of interest in polar coordinates by degree (e.g. pi/4).

Usage

```
directExplorer2d(Tc, annotation=NULL, gene.method="OSP",
path.method="Stouffer", top=10, nd=8, ...)
```

Arguments

Tc	a numeric matrix with 2 columns. The rows are genes or phosphorylation sites and the columns are treatments vs control statistics.
annotation	a list with names correspond to pathways or kinases and elements correspond to genes or substrates belong to each pathway or kinase, respectively.
gene.method	the method to be used for integrating statistics across treatments for each gene or phosphorylation site. Available methods are Stouffer, OSP, Fisher, and maxP. Default method is OSP.
path.method	the method to be used for integrating statistics of all genes or phosphorylation sites that belongs to a pathway or kinase. Available methods are Stouffer, OSP, Fisher, and maxP. Default method is Stouffer.
top	the number of entries to be highlighted in the plot.
nd	the number of directions to plot (4 or 8)
...	parameters for controlling the plot.

Value

The the list of enrichment analysis in tables.

Examples

```
# load the phosphoproteomics dataset
data(HEK)

# load the kinase-substrate annoations
data(PhosphoSite)

# test enrichment on 8 directions in polar coordinate system.
bda <- directExplorer2d(Tc=HEK, annotation=PhosphoSite.mouse)

# the direction are denoted as follow for the two treatments vs control:
# ++: up-regulated in both treatments
# +*: up-regulated in the first treatment and unchanged in the second treatment
```

```
# +-: up-regulated in the first treatment and down-regulated in the second treatment
# *-: unchanged in the first treatment and down-regulated in the second treatment
# --: down-regulated in both treatments
# -*: down-regulated in the first treatment and unchanged in the second treatment
# -+: down-regulated in the first treatment and up-regulated in the second treatment
# *+: unchanged in the first treatment and up-regulated in the second treatment

# sort the most enriched phosphorylation sites and kinases on down-regulaiton from both
# treatments (i.e. "--") and displa the top-10 entries
bda$gene.tab[order(bda$gene.tab[,"--"]),][1:10,]
bda$path.tab[order(bda$path.tab[,"--"]),][1:10,]
```

---

directPA   *Direction Analysis for Pathways*

Description

The main function of direction Analysis. This function takes in a matrix of test statistics with two
(2-dimensional space) or three (3-dimensional space) columns, the direction of interests, and the
annotation list such as pathway annotation, and test for enrichment of pathways on the specified
direction.

Usage

```
directPA(Tc, direction, annotation, minSize=5, gene.method="OSP",
path.method="Stouffer", visualize=TRUE, ...)
```

Arguments

Tc	a numeric matrix. Rows are genes and columns are treatments vs control statistics.
direction	the direction to be tested for enrichment. Either specified as a degree for two-dimensional analysis or as contrast (in a triplet) for three-dimensional analysis.
annotation	a list with names correspond to pathways and elements correspond to genes belong to each pathway, respectively.
minSize	the size of annotation groups to be considered for calculating enrichment. Groups that are smaller than the minSize will be removed from the analysis.
gene.method	the method to be used for integrating statistics across treatments for each gene. Available methods are Stouffer, OSP, Fisher, and maxP. Default method is OSP.
path.method	the method to be used for integrating statistics of all genes that belongs to a pathway. Available methods are Stouffer, OSP, Fisher, and maxP. Default method is Stouffer.
visualize	whether to visualize the plot.
...	other visualization parameters to pass on.

Value

a list that contains directional p-values for each gene and directional enrichment for each pathway.

Examples

```
# load the proteomics dataset
data(PM)

# load pathway annotations
data(Pathways)

# display reactome pathways. Could be replaced by any other pathway databases
Pathways.reactome[1:5]

# direction pathway analysis in 3-dimensional space. Implemnted as rotating by contrast
# (1) test combined effect of all 3 treatments (stimulation and inhibitions) vs control (basal)
# on the original direction.
dPA <- directPA(Tc=PM, direction=c(1,1,1), annotation=Pathways.reactome)
dPA$gst[order(unlist(dPA$gst[,1])),][1:20,]
# rank substrates on the direciton of interest
sort(dPA$gene.pvalues)[1:20]

# (2) test combined effect of all 3 treatments vs controls on direction c(1,-1, 0)
# this rotates Ins by 0 degree, Wmn by 90 degree, and MK by 45 degree.
dPA <- directPA(Tc=PM, direction=c(1,-1,0), annotation=Pathways.reactome)
dPA$gst[order(unlist(dPA$gst[,1])),][1:20,]

# (3) test combined effect of all 3 perturbations vs controls on direction c(1,-1, 1)
# this rotates Ins by 0 degree, Wmn by 90 degree, and MK by 0 degree.
dPA <- directPA(Tc=PM, direction=c(1,-1,1), annotation=Pathways.reactome)
dPA$gst[order(unlist(dPA$gst[,1])),][1:20,]
```

---

geneStats	*Molecule Level Statistics*

Description

Takes a vector of statistics with each element corresponds to a treatment vs control comparison, and calculates a combined statistics accross multiple treatments.

Usage

```
geneStats(T, method="OSP")
```

Arguments

T	a vector of statistics (z-scores converted) with each element correspond to a treatment vs control comparison.
method	the p-value integration method for combining accross multiple treatments. Available methods are Stouffer, OSP, Fisher, and maxP. The default method is OSP.

Value

a p-value after integration across treatments.

Examples

```
# load the example data
data(PM)

# convert statistics into z-scores
PM.zscores <- apply(PM, 2, function(x){qnorm(rank(x)/(nrow(PM)+1))})

# Rotate the matrix by contrast 1, -1, -1 (i.e. up-regulation, down-regulation, dow-regulation).
PM.rotated <- rotate3d(PM.zscores, contrast = c(1, -1, -1))

# combine rotated statistics across treatments
gene.pvalues <- apply(PM.rotated, 1, geneStats)
```

HEK   *Phosphoproteomics of HEK-293E*

Description

The data object contains all quantified phosphoryaltion sites in HEK-293E cells with Rapamycin or Torin1 treatment vs insulin stimulation. See Hsu et al. Science, 332(6035), 1317-1322, 2011, for more details.

---

kinasePA   *Direction Analysis for Kinases*

Description

This is a wrapper for runing directPA for kinase perturbation analysis (kinasePA)

Usage

```
kinasePA(Tc, direction, annotation, minSize=5, substrate.method="OSP",
kinase.method="Stouffer", visualize=TRUE, ...)
```

Arguments

Tc	a numeric matrix. The columns are phosphorylation sites and the columns are treatments vs control statistics.
direction	the direction to be tested for enrichment. Either specified as a degree for two-dimensional analysis or as contrast (in a triplet) for three-dimensional analysis.
annotation	a list with names correspond to kinases and elements correspond to substrates belong to each kinase, respectively.
minSize	the size of annotation groups to be considered for calculating enrichment. Groups that are smaller than the minSize will be removed from the analysis.
substrate.method	
	the method to be used for integrating statistics across treatments for each substrate (phosphorylation site). Available methods are Stouffer, OSP, Fisher, and maxP. Default method is OSP.
kinase.method	the method to be used for integrating statistics of all phosphorylation sites that belongs to a kinase. Available methods are Stouffer, OSP, Fisher, and maxP. Default method is Stouffer.
visualize	whether to visualize the plot.
...	other visualization parameters to pass on.

Value

a list that contains directional p-values for each substrate and directional enrichment for each kinase.

Examples

```
# load the phosphoproteomics dataset
data(HEK)

# load the kinase-substrate annoations
data(PhosphoSite)

# direction pathway analysis in 2-dimensional space. Implemented as rotating by degree
# (1) test combined effect of Torin1 and Rapamycin vs insul both on "down-regulation"
# (180 degree to original direction)
kPA <- kinasePA(Tc=HEK, direction=pi, annotation=PhosphoSite.mouse)
kPA$kinase[order(unlist(kPA$kinase[,1])),][1:20,]
# rank substrates on the direciton of interest
sort(kPA$substrate.pvalues)[1:20]

# (2) test combined effect of Torin1 and Rapamycin vs insul on "no change and down-regulation"
# (135 degree to the original direction)
kPA <- kinasePA(Tc=HEK, direction=pi*3/4, annotation=PhosphoSite.mouse)
kPA$kinase[order(unlist(kPA$kinase[,1])),][1:20,]

# (3) test combined effect of Torin1 and Rapamycin vs insul on "down-regulation and no change"
# (225 degree to the original direction)
kPA <- kinasePA(Tc=HEK, direction=pi*5/4, annotation=PhosphoSite.mouse)
```

```
kPA$kinase[order(unlist(kPA$kinase[,1])),][1:20,]
```

Pathways.KEGG   *KEGG pathway annotations*

Description

The data object contains the annotations of KEGG pathways.

Pathways.reactome   *Reactome pathway annotations*

Description

The data object contains the annotations of reactome pathways.

---

pathwayStats   *Pathway Level Statistics*

Description

Takes a vector of statistics with each element corresponds to a gene or phosphorylation site, and calculates a combined statistics for those that belong to the same pathway or kinase.

Usage

```
pathwayStats(PGs, T, minSize=5, method="Stouffer")
```

Arguments

PGs	an array of names indicating genes or substrates that belong to a given pathway or kinase.
T	a vector of statistics (z-scores converted) with each element correspond to a gene or phosphorylation site that belong to the same pathway or kinase.
minSize	the size of annotation groups to be considered for calculating enrichment. Groups that are smaller than the minSize will be removed from the analysis.
method	the p-value integration method for combining accross multiple treatments. Available methods are Stouffer, OSP, Fisher, and maxP. The default method is Stouffer.

Value

a doublet corresponding to the enrichment after integration across all genes or substrates that belong to the same pathway or kinase, and the size of the mapped genes or substrates to that pathway or kinase.

Examples

```
# load the example data
data(PM)

# load pathway annotations
data(Pathways)

# convert statistics into z-scores
PM.zscores <- apply(PM, 2, function(x){qnorm(rank(x)/(nrow(PM)+1))})

# Rotate the matrix by contrast 1, -1, -1 (i.e. up-regulation, down-regulation, dow-regulation).
PM.rotated <- rotate3d(PM.zscores, contrast = c(1, -1, -1))

# combine rotated statistics across treatments
gene.pvalues <- apply(PM.rotated, 1, geneStats)

# compute statistics for all reactome pathways
gene.zscores <- qnorm(gene.pvalues, lower.tail = FALSE)
gst <- t(sapply(Pathways.reactome, pathwayStats, gene.zscores))
```

---

perturbPlot2d   *Perturbation Plot*

Description

This function takes in a matrix of test statistics with two columns (2-dimensional space) and the annotation list such as pathway annotation or kinase-substrate annotation, and visualize the enrichment of pathways or kinases in direction specific manner.

Usage

```
perturbPlot2d(Tc, annotation, minSize=5, ...)
```

Arguments

Tc	a numeric matrix. The columns are genes or phosphorylation sites and the columns are treatments vs control statistics.
annotation	a list with names correspond to pathways or kinases and elements correspond to genes or substrates belong to each pathway or kinase, respectively.
minSize	the size of annotation groups to be considered for calculating enrichment. Groups that are smaller than the minSize will be removed from the analysis.
...	parameters for controling the plot.

Value

a list of coordinates for pathways or kinases

Examples

```
# load the phosphoproteomics dataset
data(HEK)

# load the kinase-substrate annoations
data(PhosphoSite)

perturbPlot2d(Tc=HEK, annotation=PhosphoSite.mouse, cex=3)
```

---

perturbPlot3d   *Perturbation Plot 3D*

Description

This function takes in a matrix of test statistics with two columns (3-dimensional space) and the
annotation list such as pathway annotation or kinase-substrate annotation, and visualize the enrich-
ment of pathways or kinases in direction specific manner.

Usage

```
perturbPlot3d(Tc, annotation, minSize=5, ...)
```

Arguments

Tc	a numeric matrix. The columns are genes or phosphorylation sites and the columns are treatments vs control statistics.
annotation	a list with names correspond to pathways or kinases and elements correspond to genes or substrates belong to each pathway or kinase, respectively.
minSize	the size of annotation groups to be considered for calculating enrichment. Groups that are smaller than the minSize will be removed from the analysis.
...	parameters for controling the plot.

Value

a list of coordinates for pathways or kinases

PhosphoELM.human   *PhosphoELM annotations for human*

Description

The data object contains the annotations of kinases and their conrresponding substrates as phosphorylation sites in human. It is extracted from the PhosphoELM database. For details of PhosphoELM, please refer to the article: Phospho.ELM: a database of phosphorylation sites–update 2008. Diella, F., Gould, C.M., Chica, C., Via, A. & Gibson, T.J. Nucleic Acids Res. 2008 Jan;36(Database issue):D240-4.

PhosphoELM.mouse   *PhosphoELM annotations for mouse*

Description

The data object contains the annotations of kinases and their conrresponding substrates as phosphorylation sites in mouse. It is extracted from the PhosphoELM database. For details of PhosphoELM, please refer to the article: Phospho.ELM: a database of phosphorylation sites–update 2008. Diella, F., Gould, C.M., Chica, C., Via, A. & Gibson, T.J. Nucleic Acids Res. 2008 Jan;36(Database issue):D240-4.

PhosphoSite.human   *PhosphoSitePlus annotations for human*

Description

The data object contains the annotations of kinases and their conrresponding substrates as phosphorylation sites in human. It is extracted from the PhosphoSitePlus database. For details of PhosphoSitePlus, please refer to the article: Hornbeck et al. Nucleic Acids Res., 40:D261-70, 2012.

PhosphoSite.mouse   *PhosphoSitePlus annotations for mouse*

Description

The data object contains the annotations of kinases and their conrresponding substrates as phosphorylation sites in mouse. It is extracted from the PhosphoSitePlus database. For details of PhosphoSitePlus, please refer to the article: Hornbeck et al. Nucleic Acids Res., 40:D261-70, 2012.

PM   *Plasma Membrame Protoemics Data*

Description

The data object contains all quantified plasma membrame proteins with insulin simulation with and without prior treatments of MK or Wmn vs basal condition in mouse 3T3L1 cell lines. Please refer to the article: Yang et al. Bioinformatics, 30(6):808-14, 2014.

---

rotate2d   *Polar Coordinates Rotation*

Description

Rotate to the direction of interest in polar coordinates by degree (e.g. pi/4).

Usage

```
rotate2d(T, degree = 0)
```

Arguments

T	a numeric matrix (normally z-score converted) with 2 columns. The rows are genes or phosphorylation sites and the columns are treatments vs control statistics.
degree	the direction to be tested for enrichment. Specified as degree from to the original direction.

Value

A rotated matrix with respect to the direction of interest.

Examples

```
# load the phosphoproteomics dataset
data(HEK)

# convert statistics into z-scores
HEK.zscores <- apply(HEK, 2, function(x){qnorm(rank(x)/(nrow(HEK)+1))})

# Rotate the matrix by 1/2 pi (i.e. down-regulation, dow-regulation).
HEK.rotated <- rotate2d(HEK.zscores, degree = pi/2)
```

---

rotate3d   *Spherical Coordinates Rotation*

Description

Rotate to the direction of interest in spherical coordinates by contrasts (e.g. 1, -1, -1).

Usage

```
rotate3d(T, contrast)
```

Arguments

T	a numeric matrix (normally z-score converted) with 3 columns. The rows are genes or phosphorylation sites and the columns are treatments vs control statistics.
contrast	the direction to be tested for enrichment. Specified as contrast in a triplet (see example).

Value

A rotated matrix with respect to the direction of interest.

Examples

```
# load the example data
data(PM)

# convert statistics into z-scores
PM.zscores <- apply(PM, 2, function(x){qnorm(rank(x)/(nrow(PM)+1))})

# Rotate the matrix by contrast 1, -1, -1 (i.e. up-regulation, down-regulation, dow-regulation).
PM.rotated <- rotate3d(PM.zscores, contrast = c(1, -1, -1))
```

Index