

Package ‘tinyarray’

November 8, 2021

Type Package

Title Expression Data Analysis and Visualization

Version 2.2.7

Maintainer Xiaojie Sun <18763899370@163.com>

Description Gene Expression Omnibus(GEO) and The Cancer Genome Atlas(TCGA) are common bioinformatics public databases. We integrate the regular analysis and charts for expression data, to analyze and display the data concisely and intuitively.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports BiocManager, clusterProfiler, dplyr, limma, org.Hs.eg.db,
stringr, tibble, pheatmap, ggplot2, survival, survminer,
patchwork

Suggests testthat, AnnoProbe, GEOquery, Biobase, VennDiagram,
FactoMineR, factoextra, knitr, rmarkdown, cowplot, ggpunbr,
ggplotify, tidyverse, labeling, Rtsne, scatterplot3d,
ComplexHeatmap, circlize

URL <https://github.com/xjsun1221/tinyarray>

BugReports <https://github.com/xjsun1221/tinyarray/issues>

Depends R (>= 3.5.0)

RoxygenNote 7.1.2

NeedsCompilation no

Author Xiaojie Sun [aut, cre]

Repository CRAN

Date/Publication 2021-11-08 10:00:02 UTC

R topics documented:

box_surv	3
cod	4
cor.full	4
cor.one	5
deg	6
deseq_data	6
double_enrich	7
draw_boxplot	7
draw_heatmap	9
draw_heatmap2	11
draw_KM	12
draw_pca	13
draw_tsne	14
draw_venn	15
draw_volcano	16
draw_volcano2	17
dumd	19
edges_to_nodes	20
exists_anno_list	20
exprSet_hub1	21
exp_boxplot	21
exp_hub1	22
exp_surv	22
find_anno	23
genes	24
geo_download	24
get_cgs	25
get_deg	26
get_deg_all	27
ggheat	29
hypertest	30
interaction_to_edges	31
intersect_all	32
lnc_anno	33
lnc_annov23	33
make_tcga_group	34
match_exp_cl	34
meta1	35
mRNA_anno	36
mRNA_annov23	36
multi_deg	37
multi_deg_all	38
pkg_all	40
plcortest	40
point_cut	41
quick_enrich	42

<i>box_surv</i>	3
-----------------	---

sam_filter	43
surv_cox	44
surv_KM	45
trans_array	46
trans_exp	47
t_choose	48
union_all	49

Index	50
--------------	----

box_surv *box_surv*

Description

draw box plot for a hub gene expression matrix

Usage

```
box_surv(exp_hub, exprSet_hub, meta)
```

Arguments

exp_hub	an expression matrix for hubgenes
exprSet_hub	a tumor expression set for hubgenes
meta	meta data corresponds to expression set

Value

patchwork result for hub genes boxplot and survival plot

Author(s)

Xiaojie Sun

See Also

[exp_boxplot](#); [exp_surv](#)

Examples

```
k = box_surv(log2(exp_hub1+1),exprSet_hub1,meta1);k[[1]]
```

<code>cod</code>	<i>cod</i>
------------------	------------

Description

An expression matrix form TCGA

Usage

`cod`

Format

An object of class `matrix` (inherits from `array`) with 100 rows and 512 columns.

Examples

`cod`

<code>cor.full</code>	<i>cor: test for all variables</i>
-----------------------	------------------------------------

Description

`cor.test` for all variables(each two columns)

Usage

`cor.full(x)`

Arguments

`x` A numeric matrix or data.frame

Value

a data.frame with `cor.test` p.value and estimate

Author(s)

Xiaojie Sun

See Also

[cor.one](#)

Examples

```
x = iris[,-5]  
cor.full(x)
```

cor.one

cor.test for one variable with all variables

Description

cor.test for all variables(each two columns)

Usage

```
cor.one(x, var)
```

Arguments

x	A numeric matrix or data.frame
var	your chosen variable,only one.

Value

A data.frame with cor.test p.value and estimate

Author(s)

Xiaojie Sun

See Also

[cor.full](#)

Examples

```
x = iris[,-5]  
cor.one(x,"Sepal.Width")
```

deg	<i>deg</i>
-----	------------

Description

limma differential analysis result for GSE42872

Usage

deg

Format

An object of class `data.frame` with 18591 rows and 10 columns.

Examples

`head(deg)`

deseq_data	<i>deseq_data</i>
------------	-------------------

Description

DESeq2 differential analysis result

Usage

deseq_data

Format

An object of class `data.frame` with 552 rows and 6 columns.

Examples

`head(deseq_data)`

<code>double_enrich</code>	<i>draw enrichment bar plots for both up and down genes</i>
----------------------------	---

Description

draw enrichment bar plots for both up and down genes,for human only.

Usage

```
double_enrich(deg, n = 10, color = c("#2874C5", "#f87669"))
```

Arguments

<code>deg</code>	a data.frame contains at least two columns:"ENTREZID" and "change"
<code>n</code>	how many terms will you perform for up and down genes respectively
<code>color</code>	color for bar plot

Value

a list with kegg and go bar plot according to up and down genes enrichment result.

Author(s)

Xiaojie Sun

See Also

[quick_enrich](#)

Examples

```
double_enrich(deg)
```

<code>draw_boxplot</code>	<i>draw boxplot for expression</i>
---------------------------	------------------------------------

Description

draw boxplot for expression

Usage

```
draw_boxplot(
  exp,
  group_list,
  method = "kruskal.test",
  sort = TRUE,
  drop = FALSE,
  width = 0.5,
  pvalue_cutoff = 0.05,
  xlab = "Gene",
  ylab = "Expression",
  grouplab = "Group",
  p.label = FALSE,
  add_error_bar = FALSE,
  color = c("#2874C5", "#f87669", "#e6b707", "#868686", "#66C2A5", "#FC8D62",
  "#8DA0CB", "#E78AC3", "#A6D854", "#FFD92F", "#E5C494", "#B3B3B3")
)
```

Arguments

<code>exp</code>	A numeric matrix
<code>group_list</code>	A factor with duplicated character or factor
<code>method</code>	one of kruskal.test,aov,t.test and wilcox.test
<code>sort</code>	whether the boxplot will be sorted
<code>drop</code>	whether to discard insignificant values
<code>width</code>	width of boxplot and error bar
<code>pvalue_cutoff</code>	if drop = TRUE,genes with p-values below the threshold will be drawn
<code>xlab</code>	title of the x axis
<code>ylab</code>	title of the y axis
<code>grouplab</code>	title of group legend
<code>p.label</code>	whether to show p value in the plot
<code>add_error_bar</code>	whether to add error bar
<code>color</code>	color vector

Value

a boxplot according to `exp` and grouped by `group`.

Author(s)

Xiaojie Sun

See Also

[draw_heatmap](#); [draw_volcano](#); [draw_venn](#)

Examples

```
draw_boxplot(t(iris[,1:4]),iris$Species)
exp <- matrix(rnorm(60),nrow = 10)
colnames(exp) <- paste0("sample",1:6)
rownames(exp) <- paste0("gene",1:10)
exp[,4:6] = exp[,4:6] +10
exp[1:4,1:4]
group_list <- factor(rep(c("A","B"),each = 3))
draw_boxplot(exp,group_list)
draw_boxplot(exp,group_list,color = c("grey","red"))
```

`draw_heatmap`

draw a heatmap plot

Description

print a heatmap plot for expression matrix and group by group_list parameter, exp will be scaled.

Usage

```
draw_heatmap(
  n,
  group_list,
  scale_before = FALSE,
  n_cutoff = 3,
  cluster_cols = TRUE,
  legend = FALSE,
  show_rownames = FALSE,
  annotation_legend = FALSE,
  split_column = FALSE,
  show_column_title = FALSE,
  color = (grDevices::colorRampPalette(c("#2fa1dd", "white", "#f87669")))(100),
  color_an = c("#2fa1dd", "#f87669", "#e6b707", "#868686", "#92C5DE", "#F4A582",
    "#66C2A5", "#FC8D62", "#8DA0CB", "#E78AC3", "#A6D854", "#FFD92F", "#E5C494",
    "#B3B3B3"),
  scale = TRUE,
  main = NA
)
```

Arguments

<code>n</code>	A numeric matrix
<code>group_list</code>	A factor with duplicated character or factor
<code>scale_before</code>	deprecated parameter
<code>n_cutoff</code>	3 by default , scale before plot and set a cutoff,usually 2 or 1.6
<code>cluster_cols</code>	if F,heatmap will nor cluster in column

```

legend           logical,show legend or not
show_rownames   logical,show rownames or not
annotation_legend
  logical,show annotation legend or not
split_column    split column by group_list
show_column_title
  show column title or not
color            color for heatmap
color_an         color for column annotation
scale            logical,scale the matrix or not
main             the title of the plot

```

Value

a heatmap plot according to exp and grouped by group.

Author(s)

Xiaojie Sun

See Also

[draw_pca](#); [draw_volcano](#); [draw_venn](#)

Examples

```

#example data
exp = matrix(abs(rnorm(60, sd = 16)), nrow = 10)
exp[,4:6] <- exp[,4:6]+20
colnames(exp) <- paste0("sample",1:6)
rownames(exp) <- paste0("gene",1:10)
exp[1:4,1:4]
group_list = factor(rep(c("A","B"),each = 3))
draw_heatmap(exp,group_list)
#use iris
n = t(iris[,1:4]); colnames(n) = 1:150
group_list = iris$Species
draw_heatmap(n,group_list)
draw_heatmap(n,group_list,color = colorRampPalette(c("green","black","red"))(100),
  color_an = c("red","blue","pink") )

```

draw_heatmap2	<i>draw heatmap plots</i>
---------------	---------------------------

Description

print heatmap plots for expression matrix and group by group_list paramter

Usage

```
draw_heatmap2(
  exp,
  group_list,
  deg,
  heat_union = TRUE,
  heat_id = 1,
  gene_number = 200,
  show_rownames = FALSE,
  scale_before = FALSE,
  n_cutoff = 3,
  cluster_cols = TRUE,
  annotation_legend = FALSE,
  legend = FALSE,
  color = (grDevices::colorRampPalette(c("#2fa1dd", "white", "#f87669")))(100),
  color_an = c("#2fa1dd", "#f87669", "#e6b707", "#868686", "#92C5DE", "#F4A582",
    "#66C2A5", "#FC8D62", "#8DA0CB", "#E78AC3", "#A6D854", "#FFD92F", "#E5C494",
    "#B3B3B3"))
)
```

Arguments

exp	A numeric matrix
group_list	A factor with duplicated character or factor
deg	a data.frame created by Differential analysis
heat_union	logical ,use union or intersect DEGs for heatmap
heat_id	id of heatmap,1 for all DEGs,2 for head and tail,3 for top n DEGs
gene_number	how many DEGs will heatmap show .
show_rownames	logical,show rownames or not
scale_before	deprecated parameter
n_cutoff	3 by defalut , scale before plot and set a cutoff,usually 2 or 1.6
cluster_cols	if F,heatmap will nor cluster in column
annotation_legend	logical,show annotation legend or not
legend	logical,show legend or not
color	color vector
color_an	color for column annotation

Value

a heatmap plot according to exp and grouped by group.

Author(s)

Xiaojie Sun

See Also

[draw_pca](#); [draw_volcano](#); [draw_venn](#)

Examples

```
gse = "GSE474"
geo = geo_download(gse, destdir=tempdir(), by_annoprobe = FALSE)
geo$exp[1:4,1:4]
geo$exp=log2(geo$exp+1)
group_list=ifelse(stringr::str_detect(geo$pd$title,"MObese"),"MObese",
ifelse(stringr::str_detect(geo$pd$title,"NonObese"),"NonObese","Obese"))
group_list=factor(group_list,levels = c("NonObese","Obese","MObese"))
find_anno(geo/gpl)
ids <- AnnoProbe::idmap(geo/gpl,destdir = tempdir())
deg = multi_deg(geo$exp,group_list,ids,adjust = FALSE)
draw_heatmap2(geo$exp,group_list,deg)
```

draw_KM

draw_KM

Description

draw KM-plot with two or more group

Usage

```
draw_KM(
  meta,
  group_list,
  time_col = "time",
  event_col = "event",
  legend.title = "Group",
  legend.labs = levels(group_list),
  color = c("#2874C5", "#f87669", "#e6b707", "#868686", "#92C5DE", "#F4A582",
  "#66C2A5", "#FC8D62", "#8DA0CB", "#E78AC3", "#A6D854", "#FFD92F", "#E5C494",
  "#B3B3B3")
```

Arguments

meta	survival data with time and event column
group_list	A factor with duplicated character or factor
time_col	colname of time
event_col	colname of event
legend.title	legend title
legend.labs	character vector specifying legend labels
color	color vector

Value

a KM-plot

Author(s)

Xiaojie Sun

Examples

```
require("survival")
x = survival::lung
draw_KM(meta = x,
        group_list = x$sex, event_col = "status")
```

draw_pca

draw PCA plots

Description

do PCA analysis and print a PCA plot

Usage

```
draw_pca(
  exp,
  group_list,
  color = c("#2874C5", "#f87669", "#e6b707", "#868686", "#92C5DE", "#F4A582",
  "#66C2A5", "#FC8D62", "#8DA0CB", "#E78AC3", "#A6D854", "#FFD92F", "#E5C494",
  "#B3B3B3"),
  addEllipses = TRUE,
  style = "default",
  color.label = "Group"
)
```

Arguments

<code>exp</code>	A numeric matrix
<code>group_list</code>	A factor with duplicated character or factor
<code>color</code>	color vector
<code>addEllipses</code>	logical,add ellipses or not
<code>style</code>	plot style,"default","ggplot2"and "3D"
<code>color.label</code>	color legend label

Value

a pca plot according to `exp` and grouped by `group`.

Author(s)

Xiaojie Sun

See Also

[draw_heatmap](#); [draw_volcano](#); [draw_venn](#)

Examples

```
draw_pca(t(iris[,1:4]),iris$Species)
exp <- matrix(rnorm(60),nrow = 10)
colnames(exp) <- paste0("sample",1:6)
rownames(exp) <- paste0("gene",1:10)
exp[1:4,1:4]
group_list <- factor(rep(c("A","B"),each = 3))
draw_pca(exp,group_list)
draw_pca(exp,group_list,color = c("blue","red"))
```

`draw_tsne`

draw_tsne

Description

draw tsne plot with annotation by ggplot2

Usage

```
draw_tsne(
  exp,
  group_list,
  perplexity = 30,
  color = c("#2874C5", "#f87669", "#e6b707", "#868686", "#92C5DE", "#F4A582",
  "#66C2A5", "#FC8D62", "#8DA0CB", "#E78AC3", "#A6D854", "#FFD92F", "#E5C494",
```

```

    "#B3B3B3"),
color.label = "group",
addEllipses = TRUE
)

```

Arguments

exp	A numeric matrix
group_list	A factor with duplicated character or factor
perplexity	numeric; perplexity parameter for Rtsne
color	color vector
color.label	color legend label
addEllipses	logical, add ellipses or not

Value

a ggplot object

Author(s)

Xiaojie Sun

Examples

```

exp <- matrix(rnorm(10000), nrow = 50)
colnames(exp) <- paste0("sample", 1:200)
rownames(exp) <- paste0("gene", 1:50)
exp[1:4, 1:4]
exp[, 1:100] = exp[, 1:100]+10
group_list <- factor(rep(c("A", "B"), each = 100))
draw_tsne(exp, group_list)

```

draw_venn

draw a venn plot

Description

print a venn plot for deg result created by three packages

Usage

```

draw_venn(
  x,
  name,
  color = c("#2874C5", "#f87669", "#e6b707", "#868686", "#66C2A5", "#FC8D62",
  "#8DA0CB", "#E78AC3", "#A6D854", "#FFD92F", "#E5C494", "#B3B3B3")
)

```

Arguments

x	a list for plot
name	main of the plot
color	color vector

Value

a venn plot according to x, y and.z named "name" parameter

Author(s)

Xiaojie Sun

See Also

[draw_pca](#); [draw_volcano](#); [draw_heatmap](#)

Examples

```
x = list(Deseq2=sample(1:100,30),edgeR = sample(1:100,30),limma = sample(1:100,30))
draw_venn(x,"test")
draw_venn(x,"test",color = c("darkgreen", "darkblue", "#B2182B"))
```

draw_volcano

draw a volcano plot

Description

print a volcano plot for Differential analysis result in data.frame format.

Usage

```
draw_volcano(
  deg,
  lab = NA,
  pvalue_cutoff = 0.05,
  logFC_cutoff = 1,
  pkg = 1,
  adjust = FALSE,
  symmetry = FALSE,
  color = c("#2874C5", "grey", "#f87669")
)
```

Arguments

deg	a data.frame created by Differential analysis
lab	label for x axis in volcano plot
pvalue_cutoff	Cutoff value of pvalue,0.05 by default.
logFC_cutoff	Cutoff value of logFC,1 by default.
pkg	a integer ,means which Differential analysis packages you used,we support three packages by now, 1,2,3,4 respectively means "DESeq2","edgeR","limma(voom)","limma"
adjust	a logical value, would you like to use adjusted pvalue to draw this plot,FAISE by default.
symmetry	a logical value ,would you like to get your plot symmetrical
color	color vector

Value

a volcano plot according to logFC and P.value(or adjust P.value)

Author(s)

Xiaojie Sun

See Also

[draw_heatmap](#); [draw_pca](#); [draw_venn](#)

Examples

```
head(deseq_data)
draw_volcano(deseq_data)
draw_volcano(deseq_data,pvalue_cutoff = 0.01,logFC_cutoff = 2)
draw_volcano(deseq_data,color = c("darkgreen", "darkgrey", "#B2182B"))
```

draw_volcano2

draw_volcano2

Description

print one or more volcano plot for Differential analysis result in data.frame fomat.

Usage

```
draw_volcano2(
  deg,
  pkg = 4,
  lab,
  pvalue_cutoff = 0.05,
  logFC_cutoff = 1,
  adjust = FALSE,
  symmetry = TRUE,
  color = c("#2874C5", "grey", "#f87669")
)
```

Arguments

deg	a data.frame created by Differential analysis
pkg	a integer ,means which Differential analysis packages you used,we support three packages by now, 1,2,3,4 respectively means "DESeq2","edgeR","limma(voom)","limma"
lab	label for x axis in volcano plot
pvalue_cutoff	Cutoff value of pvalue,0.05 by default.
logFC_cutoff	Cutoff value of logFC,1 by default.
adjust	a logical value, would you like to use adjusted pvalue to draw this plot,FAISE by default.
symmetry	a logical value ,would you like to get your plot symmetrical
color	color vector

Value

one or more volcano plot

Author(s)

Xiaojie Sun

See Also

[geo_download](#); [draw_volcano](#); [draw_venn](#)

Examples

```
#two group
gse = "GSE42872"
geo = geo_download(gse, destdir=tempdir(), by_annoprobe = FALSE)
group_list = rep(c("A", "B"), each = 3)
ids = AnnoProbe::idmap('GPL6244', destdir = tempdir())
deg = get_deg(geo$exp, group_list, ids)
draw_volcano2(deg)
#multigroup
```

```

gse = "GSE474"
geo = geo_download(gse, destdir=tempdir(), by_annoprobe = FALSE)
geo$exp[1:4, 1:4]
geo$exp=log2(geo$exp+1)
group_list=ifelse(stringr::str_detect(geo$pd$title,"M0bese"),"M0bese",
ifelse(stringr::str_detect(geo$pd$title,"NonObese"),"NonObese","Obese"))
group_list=factor(group_list,levels = c("NonObese","Obese","M0bese"))
find_anno(geo$gpl)
ids <- AnnoProbe::idmap(geo$gpl,destdir = tempdir())
deg = multi_deg(geo$exp,group_list,ids,adjust = FALSE)
draw_volcano2(deg)
draw_volcano2(deg,color = c("darkgreen","grey","darkred"))

```

dumd

count unique values in every columns for data.frame

Description

in geo analysis, this function can help you simplify pdata, delete columns with unique values, which can't be used as group vector

Usage

```
dumd(x)
```

Arguments

x	A data.frame.
---	---------------

Value

The simple data.frame of columns unique values count in x

Examples

```

dumd(iris)
data(ToothGrowth)
x = ToothGrowth
dumd(ToothGrowth)

```

`edges_to_nodes` *edges_to_nodes*

Description

get nodes from edges

Usage

`edges_to_nodes(edges)`

Arguments

`edges` data.frame

Value

nodes data.frame

Author(s)

Xiaojie Sun

See Also

[interaction_to_edges](#)

Examples

```
df = data.frame(a = c("gene1", "gene2", "gene3"),
b = c("d,f,a,b",
"e,c,g",
"a,b,d"))
edges = interaction_to_edges(df)
nodes = edges_to_nodes(edges)
```

`exists_anno_list` *exists_anno_list*

Description

AnnoProbe supported GPLs

Usage

`exists_anno_list`

Format

An object of class `character` of length 175.

Examples

```
exists_anno_list
```

*exprSet_hub1**exprSet_hub1***Description**

An expression matrix from TCGA,tumor samples only

Usage

```
exprSet_hub1
```

Format

An object of class `data.frame` with 8 rows and 177 columns.

Examples

```
exprSet_hub1[1:4,1:4]
```

*exp_boxplot**exp_boxplot***Description**

draw box plot for a hub gene expression matrix

Usage

```
exp_boxplot(exp_hub, color = c("grey", "red"))
```

Arguments

<code>exp_hub</code>	an expression matrix for hubgenes
<code>color</code>	color for boxplot

Value

box plots list for all genes in the matrix

Author(s)

Xiaojie Sun

See Also

[exp_surv](#); [box_surv](#)

Examples

```
k = exp_boxplot(log2(exp_hub1+1));k[[1]]
```

`exp_hub1`

exp_hub1

Description

An expression matrix from TCGA

Usage

`exp_hub1`

Format

An object of class `matrix` (inherits from `array`) with 8 rows and 350 columns.

Examples

```
exp_hub1[1:4,1:4]
```

`exp_surv`

exp_surv

Description

draw surv plot for a hub gene expression matrix for tumor samples

Usage

```
exp_surv(exprSet_hub, meta, color = c("grey", "red"))
```

Arguments

- | | |
|--------------------------|---|
| <code>exprSet_hub</code> | a tumor expression set for hubgenes |
| <code>meta</code> | meta data corresponds to expression set |
| <code>color</code> | color for boxplot |

Value

survival plots list for all genes

Author(s)

Xiaojie Sun

See Also

[exp_boxplot](#); [box_surv](#); [draw_venn](#)

Examples

```
tmp = exp_surv(exprSet_hub1,meta1)
patchwork::wrap_plots(tmp)
```

find_anno

find annotation package or files

Description

find gpl annotation package or files

Usage

```
find_anno(gpl, install = FALSE, update = FALSE)
```

Arguments

gpl	a gpl accession
install	whether to install and library the package
update	whether to update the package

Value

a list with deg data.frame, volcano plot and a list with DEGs.

Author(s)

Xiaojie Sun

See Also

[geo_download](#)

Examples

```
find_anno("GPL570")
```

genes	<i>genes</i>
-------	--------------

Description

some gene entriz ids

Usage

```
genes
```

Format

An object of class character of length 511.

Examples

```
genes
```

geo_download	<i>geo_download</i>
--------------	---------------------

Description

download gse data and get informations

Usage

```
geo_download(
  gse,
  by_annoprobe = TRUE,
  simpd = TRUE,
  colon_remove = FALSE,
  destdir = getwd()
)
```

Arguments

gse	gse assessment number
by_annoprobe	getGEO or geoChina
simpd	get simplified pdata,drop out columns with all same values
colon_remove	whether to remove duplicated columns with colons
destdir	The destination directory for data downloads.

Value

a list with exp,pd and gpl

Author(s)

Xiaojie Sun

See Also

[find_anno](#)

Examples

```
gse = "GSE42872"
a = geo_download(gse,by_annoprobe = FALSE,destdir=tempdir())
```

get_cgs

get_cgs

Description

extract DEGs from deg data.frame

Usage

`get_cgs(deg)`

Arguments

`deg` a data.frame created by Differential analysis

Value

a list with upgenes,downgenes,diffgenes.

Author(s)

Xiaojie Sun

See Also

[geo_download](#); [draw_volcano](#); [draw_venn](#)

Examples

```
#two group
gse = "GSE42872"
geo = geo_download(gse, destdir=tempdir(), by_annoprobe = FALSE)
group_list = rep(c("A","B"), each = 3)
ids = AnnoProbe::idmap('GPL6244', destdir=tempdir())
deg = get_deg(geo$exp, group_list, ids)
cgs = get_cgs(deg)
#mutigroup
gse = "GSE474"
geo = geo_download(gse, destdir=tempdir(), by_annoprobe = FALSE)
geo$exp[1:4,1:4]
geo$exp=log2(geo$exp+1)
group_list=ifelse(stringr::str_detect(geo$pd$title,"MObese"),"MObese",
ifelse(stringr::str_detect(geo$pd$title,"NonObese"),"NonObese","Obese"))
group_list=factor(group_list, levels = c("NonObese","Obese","MObese"))
find_anno(geo$gpl)
ids = AnnoProbe::idmap(geo$gpl, destdir = tempdir())
deg = multi_deg(geo$exp, group_list, ids, adjust = FALSE)
cgs = get_cgs(deg)
```

get_deg

get_deg

Description

do differential analysis according to expression set and group information

Usage

```
get_deg(
  exp,
  group_list,
  ids,
  logFC_cutoff = 1,
  pvalue_cutoff = 0.05,
  adjust = FALSE,
  entriz = TRUE
)
```

Arguments

exp	A numeric matrix
group_list	A factor with duplicated character or factor
ids	a data.frame with 2 columns,including probe_id and symbol
logFC_cutoff	Cutoff value of logFC,1 by default.

pvalue_cutoff Cutoff value of pvalue,0.05 by default.
 adjust a logical value, would you like to use adjusted pvalue to draw this plot, FALSE by default.
 entriz whether convert symbols to entriz ids

Value

a deg data.frame with 10 columns

Author(s)

Xiaojie Sun

See Also

[multi_deg](#); [get_deg_all](#)

Examples

```
gse = "GSE42872"
geo = geo_download(gse, destdir=tempdir(), by_annoprobe = FALSE)
Group = rep(c("control", "treat"), each = 3)
Group = factor(Group)
find_anno(geo$gpl)
ids <- AnnoProbe::idmap(geo$gpl, destdir = tempdir())
deg = get_deg(geo$exp, Group, ids)
head(deg)
```

get_deg_all

get_deg_all

Description

do differential analysis according to expression set and group information

Usage

```
get_deg_all(
  exp,
  group_list,
  ids,
  logFC_cutoff = 1,
  pvalue_cutoff = 0.05,
  adjust = FALSE,
  entriz = TRUE,
  scale_before = FALSE,
  n_cutoff = 3,
```

```

cluster_cols = TRUE,
annotation_legend = FALSE,
show_rownames = FALSE,
legend = FALSE,
lab = NA,
pkg = 4,
symmetry = FALSE,
heat_union = TRUE,
heat_id = 1,
gene_number = 200,
color_volcano = c("#2874C5", "grey", "#f87669")
)

```

Arguments

exp	A numeric matrix
group_list	A factor with duplicated character or factor
ids	a data.frame with 2 columns,including probe_id and symbol
logFC_cutoff	Cutoff value of logFC,1 by default.
pvalue_cutoff	Cutoff value of pvalue,0.05 by default.
adjust	a logical value, would you like to use adjusted pvalue to draw this plot,FAISE by default.
entriz	logical , if TRUE ,convert symbol to entriz id.
scale_before	deprecated parameter
n_cutoff	3 by defalut , scale before plot and set a cutoff,usually 2 or 1.6
cluster_cols	if F,heatmap will nor cluster in column
annotation_legend	logical,show annotation legend or not
show_rownames	logical,show rownames or not
legend	logical,show legend or not
lab	label for x axis in volcano plot
pkg	a integer ,means which Differential analysis packages you used,we support three packages by now, 1,2,3,4 respectively means "DESeq2","edgeR","limma(voom)","limma"
symmetry	a logical value ,would you like to get your plot symmetrical
heat_union	logical ,use union or intersect DEGs for heatmap
heat_id	id of heatmap,1 for all DEGs,2 for head and tail,3 for top n DEGs
gene_number	how many DEGs will heatmap show .
color_volcano	color for volcano plot

Value

a list with deg data.frame, volcano plot ,pca plot ,heatmap and a list with DEGs.

Author(s)

Xiaojie Sun

See Also

[get_deg](#); [multi_deg_all](#)

Examples

```
gse = "GSE42872"
geo = geo_download(gse, destdir=tempdir(), by_annoprobe = FALSE)
group_list = rep(c("A", "B"), each = 3)
group_list = factor(group_list)
find_anno(geo$gpl)
ids <- AnnoProbe::idmap(geo$gpl, destdir = tempdir())
dcp = get_deg_all(geo$exp, group_list, ids)
head(dcp$deg)
dcp$plots
```

ggheat

ggheat

Description

draw heatmap plot with annotation by ggplot2

Usage

```
ggheat(
  dat,
  group,
  cluster = FALSE,
  color = c("#2874C5", "white", "#f87669"),
  legend_color = c("#2874C5", "#f87669", "#e6b707", "#868686", "#66C2A5", "#FC8D62",
    "#8DA0CB", "#E78AC3", "#A6D854", "#FFD92F", "#E5C494", "#B3B3B3"),
  show_rownames = TRUE,
  show_colnames = TRUE,
  groupname = "group",
  expname = "exp",
  fill_mid = TRUE
)
```

Arguments

dat	expression matrix for plot
group	group for expression colnames
cluster	logical,cluster or not, default F
color	color for heatmap
legend_color	color for legend
show_rownames	logical,show rownames in plot or not,default T
show_colnames	logical,show colnames in plot or not,default T
groupname	name of group legend
expname	name of exp legend
fill_mid	use median value as geom_tile fill midpoint

Value

a ggplot object

Author(s)

Xiaojie Sun

Examples

```
exp_dat = matrix(sample(100:1000,40),ncol = 4)
exp_dat[1:(nrow(exp_dat)/2),] = exp_dat[1:(nrow(exp_dat)/2),]-1000
rownames(exp_dat) = paste0("sample",1:nrow(exp_dat))
colnames(exp_dat) = paste0("gene",1:ncol(exp_dat))
group = rep(c("A","B"),each = nrow(exp_dat)/2)
group = factor(group,levels = c("A","B"))
ggheat(exp_dat,group)
ggheat(exp_dat,group,cluster = TRUE)
ggheat(exp_dat,group,cluster = TRUE,show_rownames = FALSE,
       show_colnames = FALSE,groupname = "risk",expname = "expression")
```

Description

make hypertest for given lncRNA and mRNA common miRNAs

Usage

```
hypertest(lnc, pc, deMIR = NULL, lnctarget, pctarget)
```

Arguments

lnc	lncRNA names
pc	mRNA names
deMIR	miRNA names , default NULL
lnctarget	a data.frame with two column,lncRNA in the first column ,miRNA in the second column
pctarget	a data.frame with two column,mRNA in the first column ,miRNA in the second column

Value

a data.frame with hypertest result

Author(s)

Xiaojie Sun

See Also

[plcortest](#)

Examples

```
# to update
```

interaction_to_edges *interaction_to_edges*

Description

split interactions by sep paramter,return edges data.frame

Usage

```
interaction_to_edges(df, a = 1, b = 2, sep = ",")
```

Arguments

df	interactions data.frame
a	column to replicate
b	column to split
sep	a character string to separate b column

Value

a new data.frame with two column ,one interaction by one rows

Author(s)

Xiaojie Sun

See Also

[edges_to_nodes](#)

Examples

```
df = data.frame(a = c("gene1", "gene2", "gene3"),
b = c("d,f,a,b",
      "c,e,g",
      "a,b,d"))
interaction_to_edges(df)
```

`intersect_all`

intersect_all

Description

calculate intersect set for two or more elements

Usage

`intersect_all(...)`

Arguments

`...` some vectors or a list with some vectors

Value

vector

Author(s)

Xiaojie Sun

See Also

[union_all](#)

Examples

```
x1 = letters[1:4]
x2 = letters[3:6]
x3 = letters[3:4]
re = intersect_all(x1,x2,x3)
re2 = intersect_all(list(x1,x2,x3))
re3 = union_all(x1,x2,x3)
```

`lnc_anno`*lnc_anno*

Description

annotation for TCGA expression matrix(lncRNA),form genecode v22 gtf file.

Usage`lnc_anno`**Format**

An object of class `data.frame` with 14826 rows and 3 columns.

Examples`head(lnc_anno)`

`lnc_anno23`*lnc_anno23*

Description

annotation for TCGA and gtex expression matrix(lncRNA),form genecode v23 gtf file.

Usage`lnc_anno23`**Format**

An object of class `data.frame` with 14852 rows and 3 columns.

Examples`head(lnc_anno23)`

<code>make_tcga_group</code>	<i>make_tcga_group</i>
------------------------------	------------------------

Description

make tcga group for given tcga expression matrix

Usage

```
make_tcga_group(exp)
```

Arguments

<code>exp</code>	TCGA or TCGA_Gtex expression set from gcd or xena
------------------	---

Value

a group factor with normal and tumor ,correspond to colnames for expression matrix

Author(s)

Xiaojie Sun

See Also

[sam_filter](#); [match_exp_cl](#)

Examples

```
k = make_tcga_group(exp_hub1); table(k)
```

<code>match_exp_cl</code>	<i>match_exp_cl</i>
---------------------------	---------------------

Description

match exp and clinical data from TCGA

Usage

```
match_exp_cl(exp, cl, id_column = "id", sample_centric = TRUE)
```

Arguments

exp	TCGA expression set
cl	TCGA clinical data.frame
id_column	which column contains patient ids, column number or colnmn name.
sample_centric	logical,default T,keep all samples from the same patients.if FALSE,keep only one tumor sample for one patient.

Value

a transformed clinical data.frame with sample ids.

Author(s)

Xiaojie Sun

See Also

[make_tcga_group](#); [sam_filter](#)

Examples

```
a = match_exp_cl(exp_hub1,meta1[,2:4],"X_PATIENT")
exp_matched = a[[1]]
cl_matched = a[[2]]
b = match_exp_cl(exp_hub1,meta1[,2:4],"X_PATIENT",sample_centric = FALSE)
exp_matched = b[[1]]
cl_matched = b[[2]]
```

meta1

meta1

Description

clinical messages for some TCGA patients,correspond to exprSet_hub1

Usage

meta1

Format

An object of class `data.frame` with 177 rows and 4 columns.

Examples

`head(meta1)`

*mRNA_anno**mRNA_anno*

Description

annotation for TCGA and gtex expression matrix(mRNA),form genecode v22 gtf file.

Usage

mRNA_anno

Format

An object of class `data.frame` with 19814 rows and 3 columns.

Examples

```
head(mRNA_anno)
```

*mRNA_anno*v23*mRNA_anno*v23

Description

annotation for TCGA and gtex expression matrix(mRNA),form genecode v23 gtf file.

Usage

*mRNA_anno*v23

Format

An object of class `data.frame` with 19797 rows and 3 columns.

Examples

```
head(mRNA_annov23)
```

multi_deg	<i>multi_deg</i>
-----------	------------------

Description

do differential analysis according to expression set and group information

Usage

```
multi_deg(  
  exp,  
  group_list,  
  ids,  
  logFC_cutoff = 1,  
  pvalue_cutoff = 0.05,  
  adjust = FALSE,  
  entriz = TRUE  
)
```

Arguments

exp	A numeric matrix
group_list	A factor with duplicated character or factor
ids	a data.frame with 2 columns,including probe_id and symbol
logFC_cutoff	Cutoff value of logFC,1 by default.
pvalue_cutoff	Cutoff value of pvalue,0.05 by default.
adjust	a logical value, would you like to use adjusted pvalue to draw this plot, FALSE by default.
entriz	whether convert symbols to entriz ids

Value

a deg data.frame with 10 columns

Author(s)

Xiaojie Sun

See Also

[get_deg](#); [multi_deg_all](#)

Examples

```

gse = "GSE474"
geo = geo_download(gse, destdir=tempdir(), by_annoprobe = FALSE)
geo$exp[1:4,1:4]
geo$exp=log2(geo$exp+1)
group_list=ifelse(stringr::str_detect(geo$pd$title,"MObese"),
"MObese", ifelse(stringr::str_detect(geo$pd$title,"NonObese"),
"NonObese","Obese"))
group_list=factor(group_list,levels = c("NonObese","Obese","MObese"))
find_anno(geo/gpl)
ids <- AnnoProbe::idmap(geo/gpl,destdir = tempdir())
deg = multi_deg(geo$exp,group_list,ids,adjust = FALSE)
names(deg)
head(deg[[1]])
head(deg[[2]])
head(deg[[3]])

```

multi_deg_all

multi_deg_all

Description

do differential analysis according to expression set and group information

Usage

```

multi_deg_all(
  exp,
  group_list,
  ids,
  logFC_cutoff = 1,
  pvalue_cutoff = 0.05,
  adjust = FALSE,
  entriz = TRUE,
  scale_before = FALSE,
  n_cutoff = 3,
  cluster_cols = TRUE,
  annotation_legend = FALSE,
  show_rownames = FALSE,
  legend = FALSE,
  lab = NA,
  pkg = 4,
  symmetry = FALSE,
  heat_union = TRUE,
  heat_id = 1,
  gene_number = 200,

```

```
color_volcano = c("#2874C5", "grey", "#f87669")
)
```

Arguments

exp	A numeric matrix
group_list	A factor with duplicated character or factor
ids	a data.frame with 2 columns,including probe_id and symbol
logFC_cutoff	Cutoff value of logFC,1 by default.
pvalue_cutoff	Cutoff value of pvalue,0.05 by default.
adjust	a logical value, would you like to use adjusted pvalue to draw this plot, FALSE by default.
entriz	whether convert symbols to entriz ids
scale_before	deprecated parameter
n_cutoff	3 by defalut , scale before plot and set a cutoff,usually 2 or 1.6
cluster_cols	if F,heatmap will nor cluster in column
annotation_legend	logical,show annotation legend or not
show_rownames	logical,show rownames or not
legend	logical,show legend or not
lab	label for x axis in volcano plot
pkg	a integer ,means which Differential analysis packages you used,we support three packages by now, 1,2,3,4 respectively means "DESeq2","edgeR","limma(voom)","limma"
symmetry	a logical value ,would you like to get your plot symmetrical
heat_union	logical ,use union or intersect DEGs for heatmap
heat_id	id of heatmap,1 for all DEGs,2 for head and tail,3 for top n DEGs
gene_number	how many DEGs will heatmap show .
color_volcano	color for volcano

Value

a list with deg data.frame, volcano plot and a list with DEGs.

Author(s)

Xiaojie Sun

See Also

[geo_download](#); [draw_volcano](#); [draw_venn](#)

Examples

```

gse = "GSE474"
geo = geo_download(gse, destdir=tempdir(), by_annoprobe = FALSE)
geo$exp[1:4,1:4]
geo$exp=log2(geo$exp+1)
group_list=ifelse(stringr::str_detect(geo$pd$title,"MObese"),"MObese",
ifelse(stringr::str_detect(geo$pd$title,"NonObese"),"NonObese","Obese"))
group_list=factor(group_list,levels = c("NonObese","Obese","MObese"))
find_anno(geo/gpl)
ids = AnnoProbe::idmap(geo/gpl,destdir = tempdir())
dcp = multi_deg_all(geo$exp,
group_list,ids,adjust = FALSE)
dcp[[3]]

```

pkg_all

pkg_all

Description

bioconductor annotation packages for GPLs

Usage

```
pkg_all
```

Format

An object of class `data.frame` with 85 rows and 3 columns.

Examples

```
head(pkg_all)
```

plcortest

plcortest

Description

make cor.test for given lncRNA and mRNA

Usage

```
plcortest(lnc_exp, mRNA_exp, cor_cutoff = 0)
```

Arguments

lnc_exp	lncRNA expression set
mRNA_exp	mRNA expression set which now equal to lncRNA_exp
cor_cutoff	cor estimate cut_off, default 0

Value

a list with cor.test result, names are lncRNAs, element are mRNAs

Author(s)

Xiaojie Sun

See Also

[hypertest](#)

Examples

```
# to update
```

point_cut *point_cut*

Description

calculate cut point for multiple genes

Usage

```
point_cut(exprSet_hub, meta)
```

Arguments

exprSet_hub	a tumor expression set for hubgenes
meta	meta data corresponds to expression set

Value

a vector with cutpoint for genes

Author(s)

Xiaojie Sun

See Also

[surv_KM;surv_cox](#)

Examples

```
point_cut(exprSet_hub1,meta1)
```

quick_enrich

quick_enrich

Description

do differential analysis according to expression set and group information, for human only

Usage

```
quick_enrich(genes, kkgo_file = "kkgo_file.Rdata", destdir = getwd())
```

Arguments

genes	a gene symbol or entrezid vector
kkgo_file	Rdata filename for kegg and go result
destdir	destdir to save kkgofile

Value

enrichment results and dotplots

Author(s)

Xiaojie Sun

See Also

[double_enrich](#)

Examples

```
head(genes)
g = quick_enrich(genes, destdir = tempdir())
names(g)
g[[1]][1:4,1:4]
g[[3]]
g[[4]]
```

sam_filter

sam_filter

Description

drop duplicated samples from the same patients

Usage

```
sam_filter(exp)
```

Arguments

exp	TCGA or TCGA_Gtex expression set from gdc or xena
-----	---

Value

a transformed expression set without duplicated samples

Author(s)

Xiaojie Sun

See Also

[make_tcga_group](#); [match_exp_cl](#)

Examples

```
cod[1:4,1:4]
dim(cod)
cod2 = sam_filter(cod)
dim(cod2)
g = make_tcga_group(cod);table(g)
library(stringr)
table(!duplicated(str_sub(colnames(cod[,g=="tumor"]),1,12)))
```

surv_cox*surv_cox*

Description

calculate cox p values and HR for genes

Usage

```
surv_cox(
  exprSet_hub,
  meta,
  cut.point = FALSE,
  pvalue_cutoff = 0.05,
  HRkeep = "all",
  continuous = FALSE,
  min_gn = 0.1
)
```

Arguments

exprSet_hub	a tumor expression set for hubgenes
meta	meta data corresponds to expression set
cut.point	logical , use cut_point or not, if FALSE,use median by defult
pvalue_cutoff	p value cut off ,0.05 by defult
HRkeep	one of "all","protect"or"risk"
continuous	logical, gene expression or gene expression group
min_gn	Depending on the expression of a gene, there may be a large difference in the number of samples between the two groups, and if a smaller group of samples is less than 10 percent (default) of all, the gene will be discarded

Value

a matrix with gene names ,cox p value and HR

Author(s)

Xiaojie Sun

See Also

[point_cut](#); [surv_KM](#)

Examples

```
surv_cox(exprSet_hub1,meta1,cut.point = TRUE,HRkeep = "all")
```

*surv_KM**surv_KM*

Description

calculate log_rank test p values for genes

Usage

```
surv_KM(  
  exprSet_hub,  
  meta,  
  cut.point = FALSE,  
  pvalue_cutoff = 0.05,  
  min_gn = 0.1  
)
```

Arguments

exprSet_hub	a tumor expression set for hubgenes
meta	meta data corresponds to expression set
cut.point	logical , use cut_point or not, if FALSE,use median by defult
pvalue_cutoff	p value cut off ,0.05 by defult
min_gn	Depending on the expression of a gene, there may be a large difference in the number of samples between the two groups, and if a smaller group of samples is less than 10 percent (default) of all, the gene will be discarded

Value

a vector with gene names and log_rank p value

Author(s)

Xiaojie Sun

See Also

[point_cut](#); [surv_cox](#)

Examples

```
surv_KM(exprSet_hub1,meta1)  
surv_KM(exprSet_hub1,meta1,cut.point = TRUE)
```

trans_array*trans_array***Description**

transform rownames for microarray expression matrix

Usage

```
trans_array(exp, ids, from = "probe_id", to = "symbol")
```

Arguments

exp	TCGA or TCGA_Gtex expression set from gcd or xena
ids	data.frame with original rownames and new rownames
from	colname for original rownames
to	colname for new rownames

Value

a transformed expression set with new rownames

Author(s)

Xiaojie Sun

See Also

[trans_exp](#)

Examples

```
exp = matrix(1:50,nrow = 10)
rownames(exp) = paste0("g",1:10)
ids = data.frame(probe_id = paste0("g",1:10),
                 symbol = paste0("G",c(1:9,9)))
trans_array(exp,ids)
```

trans_exp

trans_exp

Description

transform rownames of TCGA or TCGA_Gtex expression set from gcd or xena,from ensembl id to gene symbol

Usage

```
trans_exp(exp, mrna_only = FALSE, lncrna_only = FALSE, gtex = FALSE)
```

Arguments

exp	TCGA or TCGA_Gtex expression set from gcd or xena
mrna_only	only keep mrna rows in result
lncrna_only	only keep lncrna rows in result
gtex	logical,whether including Gtex data

Value

a transformed expression set with symbol

Author(s)

Xiaojie Sun

See Also

[trans_array](#)

Examples

```
exp = matrix(rnorm(1000),ncol = 10)
rownames(exp) = sample(mRNA_annov23$gene_id,100)
colnames(exp) = c(paste0("TCGA",1:5),paste0("GTEX",1:5))
k = trans_exp(exp)
```

*t_choose**t_choose***Description**

choose differential expressed genes by simple t.test

Usage

```
t_choose(
  genes,
  exp,
  group_list,
  up_only = FALSE,
  down_only = FALSE,
  pvalue_cutoff = 0.05
)
```

Arguments

<code>genes</code>	a vector with some genes
<code>exp</code>	A numeric matrix
<code>group_list</code>	A factor with duplicated character or factor
<code>up_only</code>	keep up genes in the result only
<code>down_only</code>	keep down genes in the result only
<code>pvalue_cutoff</code>	p value cut off ,0.05 by defult

Value

a vector with differential expressed genes

Author(s)

Xiaojie Sun

Examples

```
exp = matrix(rnorm(1000),ncol = 10)
rownames(exp) = sample(mRNA_annov23$gene_id,100)
colnames(exp) = c(paste0("TCGA",1:5),paste0("GTEX",1:5))
exp2 = trans_exp(exp)
exp2[,1:5] = exp2[,1:5]+10
group_list = rep(c("A","B"),each = 5)
genes = sample(rownames(exp2),3)
t_choose(genes,exp2,group_list)
```

*union_all**union_all*

Description

calculate union set for two or more elements

Usage

`union_all(...)`

Arguments

... some vectors or a list with some vectors

Value

vector

Author(s)

Xiaojie Sun

See Also

[intersect_all](#)

Examples

```
x1 = letters[1:4]
x2 = letters[3:6]
x3 = letters[3:4]
re = intersect_all(x1,x2,x3)
re2 = intersect_all(list(x1,x2,x3))
re3 = union_all(x1,x2,x3)
```

Index

* **datasets**

- cod, 4
- deg, 6
- deseq_data, 6
- exists_anno_list, 20
- exp_hub1, 22
- exprSet_hub1, 21
- genes, 24
- lnc_anno, 33
- lnc_annov23, 33
- meta1, 35
- mRNA_anno, 36
- mRNA_annov23, 36
- pkg_all, 40

box_surv, 3, 22, 23

cod, 4

cor.full, 4, 5

cor.one, 4, 5

deg, 6

deseq_data, 6

double_enrich, 7, 42

draw_boxplot, 7

draw_heatmap, 8, 9, 14, 16, 17

draw_heatmap2, 11

draw_KM, 12

draw_pca, 10, 12, 13, 16, 17

draw_tsne, 14

draw_venn, 8, 10, 12, 14, 15, 17, 18, 23, 25, 39

draw_volcano, 8, 10, 12, 14, 16, 16, 18, 25, 39

draw_volcano2, 17

dumd, 19

edges_to_nodes, 20, 32

exists_anno_list, 20

exp_boxplot, 3, 21, 23

exp_hub1, 22

exp_surv, 3, 22, 22

exprSet_hub1, 21

find_anno, 23, 25

genes, 24

geo_download, 18, 23, 24, 25, 39

get_cgs, 25

get_deg, 26, 29, 37

get_deg_all, 27, 27

ggheat, 29

hypertest, 30, 41

interaction_to_edges, 20, 31

intersect_all, 32, 49

lnc_anno, 33

lnc_annov23, 33

make_tcga_group, 34, 35, 43

match_exp_cl, 34, 34, 43

meta1, 35

mRNA_anno, 36

mRNA_annov23, 36

multi_deg, 27, 37

multi_deg_all, 29, 37, 38

pkg_all, 40

plcortest, 31, 40

point_cut, 41, 44, 45

quick_enrich, 7, 42

sam_filter, 34, 35, 43

surv_cox, 41, 44, 45

surv_KM, 41, 44, 45

t_choose, 48

trans_array, 46, 47

trans_exp, 46, 47

union_all, 32, 49