

# Package ‘tinyarray’

January 22, 2024

**Type** Package

**Title** Expression Data Analysis and Visualization

**Version** 2.3.2

**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, stringr, tibble, pheatmap, ggplot2, survival, survminer, patchwork

**Suggests** testthat, AnnoProbe, GEOquery, Biobase, VennDiagram, FactoMineR, factoextra, knitr, rmarkdown, cowplot, ggpubr, ggplotify, tidyr, labeling, Rtsne, scatterplot3d, ComplexHeatmap, circlize, org.Hs.eg.db, org.Rn.eg.db, org.Mm.eg.db

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

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

**Depends** R (>= 3.5.0)

**RoxygenNote** 7.2.3

**NeedsCompilation** no

**Author** Xiaojie Sun [aut, cre]

**Repository** CRAN

**Date/Publication** 2024-01-22 09:00:02 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 . . . . .	17
draw_volcano2 . . . . .	18
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_count_txt . . . . .	26
get_deg . . . . .	27
get_deg_all . . . . .	28
get_gpl_txt . . . . .	30
ggheat . . . . .	30
hypertest . . . . .	32
interaction_to_edges . . . . .	33
intersect_all . . . . .	34
lnc_anno . . . . .	34
lnc_annot23 . . . . .	35
make_tcga_group . . . . .	35
match_exp_cl . . . . .	36
metal . . . . .	37
mRNA_anno . . . . .	37
mRNA_annot23 . . . . .	38
multi_deg . . . . .	38
multi_deg_all . . . . .	40
pkg_all . . . . .	41
plcortest . . . . .	42
plot_deg . . . . .	42
point_cut . . . . .	44
quick_enrich . . . . .	45
risk_plot . . . . .	46

<i>box_surv</i>	3
<i>sam_filter</i> . . . . .	47
<i>surv_cox</i> . . . . .	48
<i>surv_KM</i> . . . . .	49
<i>trans_array</i> . . . . .	50
<i>trans_entrezexp</i> . . . . .	51
<i>trans_exp</i> . . . . .	52
<i>trans_exp_new</i> . . . . .	53
<i>t_choose</i> . . . . .	54
<i>union_all</i> . . . . .	55
<b>Index</b>	<b>56</b>

---

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

---

### Description

draw box plot for a hub gene expression matrix

### Usage

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

### Arguments

<i>exp_hub</i>	an expression matrix for hubgenes
<i>exprSet_hub</i>	a tumor expression set for hubgenes
<i>meta</i>	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]]
```

---

cod	<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
```

---

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

---

**Description**

cor.test for all variables(each two columns)

**Usage**

```
cor.full(x, drop = min(x) - 0.001, min.obs = 10)
```

**Arguments**

x	A numeric matrix or data.frame
drop	drop values
min.obs	minimum number of observations after dropping

**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	<i>cor.test for one variable with all variables</i>
---------	---

---

**Description**

cor.test for all variables(each two columns)

**Usage**

```
cor.one(
  x,
  var,
  drop.var = min(x[, var]) - 0.001,
  drop.other = min(x[, -which(colnames(x) == var)]) - 0.001,
  min.obs = 10
)
```

**Arguments**

x	A numeric matrix or data.frame
var	your chosen variable,only one.
drop.var	drop values in var
drop.other	drop values in other columns
min.obs	minimum number of observations after dropping

**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

*deg*

---

**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

*deseq\_data*

---

**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)
```

---

double_enrich	<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**

deg	a data.frame contains at least two columns:"ENTREZID" and "change"
n	how many terms will you perform for up and down genes respectively
color	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)
```

---

draw_boxplot	<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**

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

**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 paramter, exp will be scaled.

**Usage**

```
draw_heatmap(
  n,
  group_list,
  scale_before = FALSE,
  n_cutoff = 3,
  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**

n                    A numeric matrix  
group\_list          A factor with duplicated character or factor

scale_before	deprecated parameter
n_cutoff	3 by default, scale before plot and set a cutoff, usually 2 or 1.6
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
...	other parameters from pheatmap

**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, my_genes = NULL, heat_union = TRUE, ...)
```

**Arguments**

exp	A numeric matrix
group_list	A factor with duplicated character or factor
deg	a data.frame created by Differential analysis
my_genes	genes for pheatmap
heat_union	logical ,use union or intersect DEGs for heatmap
...	other parameters from draw_heatmap

**Value**

a heatmap plot according to exp and grouped by group.

**Author(s)**

Xiaojie Sun

**See Also**

[draw\\_pca](#); [draw\\_volcano](#); [draw\\_venn](#)

**Examples**

```
## Not run:
gse = "GSE474"
geo = geo_download(gse, destdir=tempdir())
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)

## End(Not run)
```

---

draw_KM	<i>draw_KM</i>
---------	----------------

---

**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
...	other parameters from ggsurvplot

**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	<i>draw PCA plots</i>
----------	-----------------------

---

**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",  
  title = "",  
  ...  
)
```

**Arguments**

exp	A numeric matrix
group_list	A factor with duplicated character or factor
color	color vector
addEllipses	logical,add ellipses or not
style	plot style,"default","ggplot2"and "3D"
color.label	color legend label
title	plot title
...	other paramters from fviz_pca_ind

**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$Species)
draw_pca(t(iris[,1:4]),iris$Species,style = "ggplot2")
draw_pca(t(iris[,1:4]),iris$Species,style = "3D")
#change color
draw_pca(t(iris[,1:4]),iris$Species,color = c("#E78AC3", "#A6D854", "#FFD92F"))
```

---

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	<i>draw a venn plot</i>
-----------	-------------------------

---

### Description

print a venn plot for deg result created by three packages

### Usage

```
draw_venn(
  x,
  main,
  color = c("#2874C5", "#f87669", "#e6b707", "#868686", "#66C2A5", "#FC8D62", "#8DA0CB",
    "#E78AC3", "#A6D854", "#FFD92F", "#E5C494", "#B3B3B3"),
  imagetype = "png",
  filename = NULL,
  lwd = 1,
  lty = 1,
  col = color[1:length(x)],
  fill = color[1:length(x)],
  cat.col = color[1:length(x)],
  cat.cex = 1,
  cat.dist = -0.15,
  rotation.degree = 0,
  main.cex = 1,
  cex = 1,
  alpha = 0.1,
  reverse = TRUE,
  ...
)
```

### Arguments

x	a list for plot
main	Character giving the main title of the diagram
color	color vector
imagetype	Specification of the image format (e.g. tiff, png or svg)

filename	Filename for image output, or if NULL returns the grid object itself
lwd	width of the circle's circumference
lty	dash pattern of the circle's circumference
col	Colour of the circle's circumference
fill	Colour of the circle's area
cat.col	Colour of the category name
cat.cex	size of the category name
cat.dist	The distance (in npc units) of the category name from the edge of the circle (can be negative)
rotation.degree	Number of degrees to rotate the entire diagram
main.cex	Number giving the cex (font size) of the main title
cex	size of the area label
alpha	Alpha transparency of the circle's area
reverse	logical,reflect the three-set Venn diagram along its central vertical axis of symmetry. Use in combination with rotation to generate all possible set orders
...	other parameters from venn.diagram

**Value**

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

**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	<i>draw a volcano plot</i>
--------------	----------------------------

---

### Description

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

### Usage

```
draw_volcano(
  deg,
  lab = NA,
  xlab.package = TRUE,
  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, if NA , x axis names by package
xlab.package	whether to use the package name as the x axis name
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	<i>draw_volcano2</i>
---------------	----------------------

---

**Description**

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

**Usage**

```
draw_volcano2(deg, pkg = 4, lab, ...)
```

**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, if NA , x axis names by package
...	other parameters from draw_volcano

**Value**

one or more volcano plot

**Author(s)**

Xiaojie Sun

**See Also**

[geo\\_download](#);[draw\\_volcano](#);[draw\\_venn](#)

**Examples**

```
## Not run:
#two group
gse = "GSE42872"
geo = geo_download(gse,destdir=tempdir())
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())
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_volcano2(deg)
draw_volcano2(deg, color = c("darkgreen", "grey", "darkred"))

## End(Not run)
```

---

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	<i>edges_to_nodes</i>
----------------	-----------------------

---

**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",
  "c, e, g",
  "a, b, d"))
edges = interaction_to_edges(df)
nodes = edges_to_nodes(edges)
```

---

exists_anno_list	<i>exists_anno_list</i>
------------------	-------------------------

---

**Description**

AnnoProbe supported GPLs

**Usage**

```
exists_anno_list
```

**Format**

An object of class character of length 175.

**Examples**

```
exists_anno_list
```

---

<code>exprSet_hub1</code>	<i>exprSet_hub1</i>
---------------------------	---------------------

---

**Description**

An cpm expression matrix from TCGA,tumor samples only

**Usage**

```
exprSet_hub1
```

**Format**

An object of class matrix (inherits from array) with 8 rows and 177 columns.

**Examples**

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

---

<code>exp_boxplot</code>	<i>exp_boxplot</i>
--------------------------	--------------------

---

**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	<i>exp_hub1</i>
----------	-----------------

---

**Description**

An expression matrix from TCGA and Gtex

**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	<i>exp_surv</i>
----------	-----------------

---

**Description**

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

**Usage**

```
exp_surv(exprSet_hub, meta, cut.point = FALSE, color = c("#2874C5", "#f87669"))
```

**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 default
color	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)+patchwork::plot_layout(guides = "collect")
tmp2 = exp_surv(exprSet_hub1,meta1,cut.point = TRUE)
patchwork::wrap_plots(tmp2)+patchwork::plot_layout(guides = "collect")
```

---

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(),
  n = 1
)
```

**Arguments**

gse	gse assession number
by_annoprobe	download data by geoquery or annoprobe
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.
n	For data with more than one ExpressionSet, specify which one to analyze



**Value**

a list with exp,pd and gpl

**Author(s)**

Xiaojie Sun

**See Also**

[find\\_anno](#)

**Examples**

```
## Not run:  
gse = "GSE42872"  
a = geo_download(gse,destdir=tempdir())  
  
## End(Not run)
```

---

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**

```
## Not run:
#two group
gse = "GSE42872"
geo = geo_download(gse, destdir=tempdir())
group_list = rep(c("A", "B"), each = 3)
ids = AnnoProbe::idmap('GPL6244', destdir=tempdir())
deg = get_deg(geo$exp, group_list, ids)
cgs = get_cgs(deg)
#mutigroup
gse = "GSE474"
geo = geo_download(gse, destdir=tempdir())
geo$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)

## End(Not run)
```

---

get\_count\_txt

*get count from GEO*


---

**Description**

get RNA-seq count file from GEO database

**Usage**

```
get_count_txt(gse, destdir = getwd(), download = FALSE)
```

**Arguments**

gse	gse assession number
destdir	The destination directory for data downloads.
download	download the txt file or not

**Value**

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

**Author(s)**

Xiaojie Sun

**See Also**[geo\\_download](#)**Examples**

```
get_count_txt("GSE162550", destdir = tempdir())
```

---

`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,  
  species = "human"  
)
```

**Arguments**

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

**Value**

a deg data.frame with 10 columns

**Author(s)**

Xiaojie Sun

**See Also**

[multi\\_deg](#); [get\\_deg\\_all](#)

**Examples**

```
## Not run: gse = "GSE42872"
geo = geo_download(gse, destdir=tempdir())
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)

## End(Not run)
```

---

get\_deg\_all

*get\_deg\_all*

---

**Description**

do diffiencial analysis according to exprission set and group information

**Usage**

```
get_deg_all(
  exp,
  group_list,
  ids,
  symmetry = TRUE,
  my_genes = NULL,
  show_rownames = FALSE,
  cluster_cols = TRUE,
  color_volcano = c("#2874C5", "grey", "#f87669"),
  logFC_cutoff = 1,
  pvalue_cutoff = 0.05,
  adjust = FALSE,
  entriz = TRUE,
  n_cutoff = 2,
  annotation_legend = FALSE,
  lab = NA,
  species = "human"
)
```

**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
symmetry	a logical value ,would you like to get your plot symmetrical
my_genes	genes for pheatmap
show_rownames	logical,show rownames or not
cluster_cols	boolean values determining if columns should be clustered or hclust object.
color_volcano	color for volcano
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.
n_cutoff	3 by defalut , scale before plot and set a cutoff,usually 2 or 1.6
annotation_legend	logical,show annotation legend or not
lab	label for x axis in volcano plot, if NA , x axis names by package
species	choose human or mouse, or rat, default: human

**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**

```
## Not run:
gse = "GSE42872"
geo = geo_download(gse,destdir=tempdir())
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

## End(Not run)
```

---

`get_gpl_txt`                    *get gpl txt from GEO*

---

**Description**

get gpl annotation txt file from GEO database

**Usage**

```
get_gpl_txt(gpl, destdir = getwd(), download = F)
```

**Arguments**

<code>gpl</code>	gpl accession from GEO database
<code>destdir</code>	The destination directory for data downloads.
<code>download</code>	download the txt file or not

**Value**

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

**Author(s)**

Xiaojie Sun

**See Also**

[geo\\_download](#)

**Examples**

```
get_gpl_txt("GPL23270", destdir = tempdir())
```

---

`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,  
  cluster_rows = FALSE,  
  cluster_cols = FALSE,  
  groupname = "group",  
  expname = "exp",  
  fill_mid = TRUE  
)
```

**Arguments**

dat	expression matrix for plot
group	group for expression colnames
cluster	logical, cluster in both rows and column or not, default F, now replaced by cluster_rows and cluster_cols.
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
cluster_rows	logical, if rows (on the plot) should be clustered, default F
cluster_cols	logical, if column (on the plot) should be clustered, default F
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_rows = TRUE)
ggheat(exp_dat,group,cluster_rows = TRUE,show_rownames = FALSE,
       show_colnames = FALSE,groupname = "risk",expname = "expression")
```

---

hypertest

*hypertest*


---

**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	<i>intersect_all</i>
---------------	----------------------

---

**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	<i>lnc_anno</i>
----------	-----------------

---

**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(Inc_anno)
```

---

Inc_annov23	<i>Inc_annov23</i>
-------------	--------------------

---

**Description**

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

**Usage**

```
Inc_annov23
```

**Format**

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

**Examples**

```
head(Inc_annov23)
```

---

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

---

**Description**

make tcga group for given tcga expression matrix

**Usage**

```
make_tcga_group(exp)
```

**Arguments**

exp                    TCGA or TCGA\_Gtex expression set from gdc 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)
```

---

match_exp_cl	<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 column 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_annov23	<i>mRNA_annov23</i>
--------------	---------------------

---

**Description**

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

**Usage**

```
mRNA_annov23
```

**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 diffiential analysis according to expression set and group information

**Usage**

```
multi_deg(  
  exp,  
  group_list,  
  ids,  
  logFC_cutoff = 1,  
  pvalue_cutoff = 0.05,  
  adjust = FALSE,  
  species = "human",  
  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,FAISE by default.
species	choose human or mouse, or rat, default: human
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**

```
## Not run:
gse = "GSE474"
geo = geo_download(gse, destdir=tempdir())
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]])

## End(Not run)
```

---

multi_deg_all	<i>multi_deg_all</i>
---------------	----------------------

---

**Description**

do diffiencial analysis according to exprission set and group information

**Usage**

```
multi_deg_all(
  exp,
  group_list,
  ids,
  symmetry = TRUE,
  my_genes = NULL,
  show_rownames = FALSE,
  cluster_cols = TRUE,
  color_volcano = c("#2874C5", "grey", "#f87669"),
  pvalue_cutoff = 0.05,
  logFC_cutoff = 1,
  adjust = FALSE,
  entriz = TRUE,
  annotation_legend = FALSE,
  lab = NA,
  species = "human"
)
```

**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
symmetry	a logical value ,would you like to get your plot symmetrical
my_genes	genes for pheatmap
show_rownames	boolean specifying if column names are be shown.
cluster_cols	boolean values determining if columns should be clustered or hclust object.
color_volcano	color for volcano
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.
entriz	whether convert symbols to entriz ids
annotation_legend	boolean value showing if the legend for annotation tracks should be drawn.



lab label for x axis in volcano plot, if NA , x axis names by package  
species choose human or mouse, or rat, default: human

**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**

```
## Not run:
gse = "GSE474"
geo = geo_download(gse, destdir=tempdir())
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]]

## End(Not run)
```

---

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 nrow 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
```

---

plot\_deg

*plot\_deg*

---

### Description

plot pca plot, volcano plot, heatmap, and venn plot for Differential analysis result

**Usage**

```
plot_deg(
  exp,
  group_list,
  deg,
  symmetry = TRUE,
  my_genes = NULL,
  show_rownames = FALSE,
  cluster_cols = TRUE,
  color_volcano = c("#2874C5", "grey", "#f87669"),
  pvalue_cutoff = 0.05,
  logFC_cutoff = 1,
  adjust = FALSE,
  annotation_legend = FALSE,
  lab = NA,
  species = "human"
)
```

**Arguments**

exp	A numeric matrix
group_list	A factor with duplicated character or factor
deg	result of multi_deg or get_deg function
symmetry	a logical value ,would you like to get your plot symmetrical
my_genes	genes for pheatmap
show_rownames	boolean specifying if column names are be shown.
cluster_cols	boolean values determining if columns should be clustered or hclust object.
color_volcano	color for volcano
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.
annotation_legend	boolean value showing if the legend for annotation tracks should be drawn.
lab	label for x axis in volcano plot, if NA , x axis names by package
species	choose human or mouse, or rat, default: human

**Value**

plots

**Author(s)**

Xiaojie Sun

## Examples

```
## Not run:
gse = "GSE474"
geo = geo_download(gse, destdir=tempdir())
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$gpl, destdir = tempdir())
deg = get_deg(geo$exp,group_list,ids,adjust = FALSE)
plot_deg(geo$exp,group_list,deg)

## End(Not run)
```

---

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	<i>quick_enrich</i>
--------------	---------------------

---

**Description**

do diffiencial analysis according to exprission set and group information,for human only

**Usage**

```
quick_enrich(  
  genes,  
  kkggo_file = "kkggo_file.Rdata",  
  destdir = getwd(),  
  species = "human"  
)
```

**Arguments**

genes	a gene symbol or entrizid vector
kkggo_file	Rdata filename for kegg and go result
destdir	destdir to save kkgofile
species	choose human or mouse, or rat, default: human

**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]]
```

---

`risk_plot`*risk\_plot*

---

**Description**

draw risk plot

**Usage**

```
risk_plot(  
  exprSet_hub,  
  meta,  
  riskscore,  
  cut.point = FALSE,  
  color = c("#2fa1dd", "#f87669")  
)
```

**Arguments**

<code>exprSet_hub</code>	a tumor expression set for hubgenes
<code>meta</code>	meta data corresponds to expression set
<code>riskscore</code>	a numeric vector of riskscore
<code>cut.point</code>	logical , use <code>cut_point</code> or not, if FALSE,use median by default
<code>color</code>	color for boxplot

**Value**

risk plot

**Author(s)**

Xiaojie Sun

**See Also**

[exp\\_boxplot](#); [box\\_surv](#); [draw\\_venn](#)

**Examples**

```
risk_plot(exprSet_hub1,meta1,riskscore = rnorm(nrow(meta1)))
```

---

sam_filter	<i>sam_filter</i>
------------	-------------------

---

**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**

<code>exprSet_hub</code>	a tumor expression set for hubgenes
<code>meta</code>	meta data corresponds to expression set
<code>cut.point</code>	logical , use <code>cut_point</code> or not, if FALSE,use median by default
<code>pvalue_cutoff</code>	p value cut off ,0.05 by default
<code>HRkeep</code>	one of "all", "protect" or "risk"
<code>continuous</code>	logical, gene expression or gene expression group
<code>min_gn</code>	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)
surv_cox(exprSet_hub1,meta1,cut.point = TRUE,continuous = TRUE)
surv_cox(exprSet_hub1,meta1,cut.point = TRUE,continuous = TRUE,pvalue_cutoff = 1)
```



---

surv_KM	<i>surv_KM</i>
---------	----------------

---

## 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 default
pvalue_cutoff	p value cut off ,0.05 by default
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,pvalue_cutoff = 1)  
surv_KM(exprSet_hub1,meta1,cut.point = TRUE)
```

---

trans_array	<i>trans_array</i>
-------------	--------------------

---

**Description**

transform rownames for microarray or rnaseq expression matrix

**Usage**

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

**Arguments**

exp	microarray expression matrix with probe_id as rownames
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_entrezexp	<i>trans_entrezexp</i>
-----------------	------------------------

---

## Description

transform rownames of expression set from "entrez" to "symbol", according to the bitr function.

## Usage

```
trans_entrezexp(entrezexp, species = "human")
```

## Arguments

entrezexp	expression set with entrezid as rownames
species	choose human or mouse, or rat, default: human

## Value

a transformed expression set with symbol

## Author(s)

Xiaojie Sun

## See Also

[trans\\_exp](#)

## Examples

```
exp = matrix(rnorm(200), ncol = 10)
rownames(exp) = c("79691", "56271", "8662", "10394", "55630", "159162", "23541",
                 "79723", "54413", "22927", "92342", "23787", "5550", "8924",
                 "55274", "866", "8844", "353299", "587", "1473")
colnames(exp) = paste0("s", 1:10)
exp2 = trans_entrezexp(exp)
```

---

trans_exp	<i>trans_exp</i>
-----------	------------------

---

### Description

transform rownames of TCGA or TCGA\_Gtex expression set from gdc 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 gdc 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_annot23$gene_id, 100)
colnames(exp) = c(paste0("TCGA", 1:5), paste0("GTEx", 1:5))
k = trans_exp(exp)
```

---

trans_exp_new	<i>trans_exp_new</i>
---------------	----------------------

---

### Description

transform rownames of expression set from "ensembl" to "symbol", according to the new information from ensembl database.

### Usage

```
trans_exp_new(exp, mrna_only = FALSE, lncrna_only = FALSE, species = "human")
```

### Arguments

exp	expression set with ensembl as rownames
mrna_only	only keep mrna rows in result
lncrna_only	only keep lncrna rows in result
species	choose human or mouse, or rat, default: human

### Value

a transformed expression set with symbol

### Author(s)

Xiaojie Sun

### See Also

[trans\\_exp](#)

### Examples

```
exp = matrix(rnorm(1000), ncol = 10)
rownames(exp) = sample(mRNA_annot23$gene_id, 100)
colnames(exp) = c(paste0("TCGA", 1:5), paste0("GTEx", 1:5))
k = trans_exp_new(exp)
```

---

t_choose	<i>t_choose</i>
----------	-----------------

---

### 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

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

### 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	<i>union_all</i>
-----------	------------------

---

**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, [34](#)
- lnc\_annov23, [35](#)
- meta1, [37](#)
- mRNA\_anno, [37](#)
- mRNA\_annov23, [38](#)
- pkg\_all, [41](#)

box\_surv, [3](#), [22](#), [23](#), [46](#)

cod, [4](#)

cor.full, [4](#), [5](#)

cor.one, [4](#), [5](#)

deg, [6](#)

deseq\_data, [6](#)

double\_enrich, [7](#), [45](#)

draw\_boxplot, [7](#)

draw\_heatmap, [9](#), [9](#), [13](#), [16](#), [17](#)

draw\_heatmap2, [11](#)

draw\_KM, [12](#)

draw\_pca, [10](#), [11](#), [13](#), [16](#), [17](#)

draw\_tsne, [14](#)

draw\_venn, [9–11](#), [13](#), [15](#), [17](#), [18](#), [23](#), [25](#), [41](#), [46](#)

draw\_volcano, [9–11](#), [13](#), [16](#), [17](#), [18](#), [25](#), [41](#)

draw\_volcano2, [18](#)

dumd, [19](#)

edges\_to\_nodes, [20](#), [33](#)

exists\_anno\_list, [20](#)

exp\_boxplot, [3](#), [21](#), [23](#), [46](#)

exp\_hub1, [22](#)

exp\_surv, [3](#), [22](#), [22](#)

exprSet\_hub1, [21](#)

find\_anno, [23](#), [25](#)

genes, [24](#)

geo\_download, [18](#), [23](#), [24](#), [25](#), [27](#), [30](#), [41](#)

get\_cgs, [25](#)

get\_count\_txt, [26](#)

get\_deg, [27](#), [29](#), [39](#)

get\_deg\_all, [28](#), [28](#)

get\_gpl\_txt, [30](#)

ggheat, [30](#)

hypertest, [32](#), [42](#)

interaction\_to\_edges, [20](#), [33](#)

intersect\_all, [34](#), [55](#)

lnc\_anno, [34](#)

lnc\_annov23, [35](#)

make\_tcga\_group, [35](#), [36](#), [47](#)

match\_exp\_cl, [36](#), [36](#), [47](#)

meta1, [37](#)

mRNA\_anno, [37](#)

mRNA\_annov23, [38](#)

multi\_deg, [28](#), [38](#)

multi\_deg\_all, [29](#), [39](#), [40](#)

pkg\_all, [41](#)

plcortest, [32](#), [42](#)

plot\_deg, [42](#)

point\_cut, [44](#), [48](#), [49](#)

quick\_enrich, [7](#), [45](#)

risk\_plot, [46](#)

sam\_filter, [36](#), [47](#)

surv\_cox, [44](#), [48](#), [49](#)

surv\_KM, [44](#), [48](#), [49](#)



t\_choose, [54](#)  
trans\_array, [50](#), [52](#)  
trans\_entrezexp, [51](#)  
trans\_exp, [50](#), [51](#), [52](#), [53](#)  
trans\_exp\_new, [53](#)  
  
union\_all, [34](#), [55](#)