# Package 'scMAGeCK'

April 12, 2022

**Type** Package

**Title** Identify genes associated with multiple expression phenotypes in single-cell CRISPR screening data

**Version** 1.6.0

**Date** 2021-10-22

**Author** Wei Li, Xiaolong Cheng, Lin Yang

**Maintainer** Xiaolong Cheng <xiaolongcheng1120@gmail.com>

**Description**
scMAGeCK is a computational model to identify genes associated with multiple expression phenotypes from CRISPR screening coupled with single-cell RNA sequencing data (CROP-seq)

**License** BSD_2_clause

**biocViews** CRISPR, SingleCell, RNASeq, PooledScreens, Transcriptomics, GeneExpression, Regression

**NeedsCompilation** yes

**Imports** Seurat, ggplot2, stats, utils

**Suggests** knitr, rmarkdown

**VignetteBuilder** knitr

**git_url** https://git.bioconductor.org/packages/scMAGeCK

**git_branch** RELEASE_3_14

**git_last_commit** 194b6e1

**git_last_commit_date** 2021-10-26

**Date/Publication** 2022-04-12

## R topics documented:

scMAGeCK-package . . . . . 2
featurePlot . . . . . 3
pre_processRDS . . . . . 4
scmageck_lr . . . . . 5
scmageck_rra . . . . . 6
selectPlot . . . . . 7

**Index** 9

scMAGeCK-package *Identify genes associated with multiple expression phenotypes in single-cell CRISPR screening data*

## Description

scMAGeCK is a computational model to identify genes associated with multiple expression phenotypes from CRISPR screening coupled with single-cell RNA sequencing data (CROP-seq)

## Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.
scMAGeCK is a computational model to identify genes associated with multiple expression phenotypes from CRISPR screening coupled with single-cell RNA sequencing data (CROP-seq).scMAGeCK is based on our previous MAGeCK and MAGeCK-VISPR models for pooled CRISPR screens.

The scMAGeCK manuscript can be found at bioRxiv(https://www.biorxiv.org/content/10.1101/658146v1/).

## Author(s)

Wei Li, Xiaolong Cheng, Lin Yang

Maintainer: Xiaolong Cheng <xiaolongcheng1120@gmail.com>

## Examples

```
### BARCODE file contains cell identity information, generated from
### the cell identity collection step
BARCODE <- system.file("extdata","barcode_rec.txt",package = "scMAGeCK")
### RDS can be a Seurat object or local RDS file path that contains
### the scRNA-seq dataset
RDS <- system.file("extdata","singles_dox_mki67_v3.RDS",package = "scMAGeCK")

target_gene <- "MKI67"

### Set RRA executable file path or activate scmageck env if needed (see https://bitbucket.org/weililab/scmageck
RRAPATH <- NULL

rra_result <- scmageck_rra(BARCODE=BARCODE, RDS=RDS, GENE=target_gene,
                           RRAPATH=RRAPATH, LABEL='dox_mki67',
                           NEGCTRL=NULL, KEEPTMP=FALSE,
                           PATHWAY=FALSE, SAVEPATH=NULL)
head(rra_result)

lr_result <- scmageck_lr(BARCODE=BARCODE, RDS=RDS, LABEL='dox_scmageck_lr',
        NEGCTRL = 'NonTargetingControlGuideForHuman', PERMUTATION = 1000,
        SAVEPATH=NULL, LAMBDA=0.01)
```

```
lr_score <- lr_result[1]
lr_score_pval <- lr_result[2]
head(lr_score_pval)
```

---

`featurePlot` *Detect the sgRNA distribution and generate Vlnplot to identity gene regulation between different cells.*

---

## Description

identify how many sgRNAs entered the cells, visualize the counts of sgRNAs distribution and gene regulation.

## Usage

```
featurePlot(RDS, TYPE = plot.type, BARCODE = NULL, sgRNA = NULL, GENE = NULL, CONTROL = NULL, palette = N

plot.type
# c("Dis", "Vln", "Den")
```

## Arguments

| | |
|---|---|
| RDS | RDS object from the pre-processRDS step |
| TYPE | Type of the plot. |
| BARCODE | A txt file to include cell identity information, generated from the cell identity collection step. |
| sgRNA | Generate whole sgRNAs distribution when sgRNA = NULL, add sgRNAs to see the specific sgRNA distribution. Mutiple sgRNAs can be provided, separated by ",". For example, "APC,TP53". |
| GENE | Genes whose expressions are to be compared under different cell coditions. Mutiple genes can be provided, separated by ",". For example, "APC,TP53". when provide mutiple genes, it would show the average gene expression. |
| CONTROL | Set up the sepecifc clusters to compare the gene expression. it would compare gene expression across the dataset when CONTROL = NULL. |
| palette | The color palette to change the color of VlnPlot. |
| label.size | Text size of label. |
| axis.size | Text size of axis.text. |
| title.size | Text size of axis/pics' title. |
| legend.text | Text size of figure legend. |
| fill | Fill colour. |

## Examples

```
### Loading required package
require(Seurat)

### BARCODE file contains cell identity information, generated from the cell identity collection step
BARCODE <- system.file("extdata","barcode_rec.txt",package = "scMAGeCK")

### RDS can be a Seurat object or local RDS file path that contains the scRNA-seq dataset
RDS <- system.file("extdata","singles_dox_mki67_v3.RDS",package = "scMAGeCK")

### For using the featurePlot function, it needs to do the preprocessRDS first
Demo <- pre_processRDS(BARCODE = BARCODE, RDS = RDS)

### For the sgRNA distribution
featurePlot(BARCODE = BARCODE, RDS = Demo, TYPE = "Dis")

### For the density of sgRNA, clustering needed to be done first.
Demo <- RunUMAP(Demo, dims = 1:10)
featurePlot(RDS = Demo, sgRNA = NULL, TYPE = "Den")

###For the Vlnplot to display the gene regulation, take MKI67 for examples
featurePlot(RDS = Demo, GENE = "MKI67", sgRNA = "TP53", TYPE = "Vln")
```

---

# pre_processRDS *Integrate the imformation of sgRNA into RDS file for the further analysis.*

---

## Description

Pre-process the sgRNA count from previous step, and generate the sgRNA expression matrix.

## Usage

```
pre_processRDS(BARCODE, RDS)
```

## Arguments

| | |
|---|---|
| BARCODE | A txt file to include cell identity information, generated from the cell identity collection step. |
| RDS | A Seurat object or local RDS file path that contains the scRNA-seq dataset. Note that the dataset has to be normalized and scaled. |

## Examples

```
### BARCODE file contains cell identity information, generated from the cell identity collection step
BARCODE <- system.file("extdata","barcode_rec.txt",package = "scMAGeCK")

### RDS can be a Seurat object or local RDS file path that contains the scRNA-seq dataset
RDS <- system.file("extdata","singles_dox_mki67_v3.RDS",package = "scMAGeCK")
```

```
Demo <- pre_processRDS(BARCODE = BARCODE, RDS = RDS)
```

---

scmageck_lr *Use linear regression to test the association of gene knockout with all possible genes*

---

## Description

echo "Use linear regression to test the association of gene knockout with all possible genes"

## Usage

```
scmageck_lr(BARCODE, RDS, NEGCTRL, SELECT_GENE=NULL, LABEL = NULL,
PERMUTATION = NULL, SIGNATURE = NULL, SAVEPATH = "./",LAMBDA=0.01,GENE_FRAC=0.01)
```

## Arguments

| | |
|---|---|
| BARCODE | A txt file to include cell identity information, generated from the cell identity collection step. |
| RDS | A Seurat object or local RDS file path that contains the scRNA-seq dataset. Note that the dataset has to be normalized and scaled. |
| NEGCTRL | The name of the genes (separated by ",") served as negative controls. |
| SELECT_GENE | The list of genes for regression. By default, all genes in the table are subject to regression. |
| LABEL | The label of the output file. |
| PERMUTATION | The number of permutations for p value calculation. |
| SAVEPATH | The save path of result. Default save path is the current working directory. If you don't need save the result, set SAVEPATH as NULL. |
| LAMBDA | A paramter for the LR model for ridge regression. Default: 0.01. |
| GENE_FRAC | A paramter for filtering low expressed genes. By default, only genes that have expressions in at least that fractions of cells (in raw count table) are kept. If raw count table is not found, will look into scaled expression, and only keep genes whose expression is greater than zero in at least that fraction of cells. Default: 0.01. |
| SIGNATURE | A GMT text file, the format must be as follows:(1)Column 1: name of gene set; (2)Colum 2: Empty, or the information about gene set e.g. the source of the gene set; (3)Column 3 and onwards: ids of genes beloging to a particular gene set. Note that this parameter for applying LR model for the gene set analysis. Reference:http://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats |

## Value

The result for object RDS

## Examples

```
### BARCODE file contains cell identity information, generated from the cell identity collection step
 BARCODE <- system.file("extdata","barcode_rec.txt",package = "scMAGeCK")

### RDS can be a Seurat object or local RDS file path that contains the scRNA-seq dataset
 RDS <- system.file("extdata","singles_dox_mki67_v3.RDS",package = "scMAGeCK")

lr_result <- scmageck_lr(BARCODE=BARCODE, RDS=RDS, LABEL='dox_scmageck_lr', SIGNATURE = NULL,
     NEGCTRL = 'NonTargetingControlGuideForHuman', PERMUTATION = 1000, SAVEPATH=NULL, LAMBDA=0.01)
 lr_score <- lr_result[1]
 lr_score_pval <- lr_result[2]
 head(lr_score_pval)
```

---

# scmageck_rra *Use RRA to test the association of gene knockout with certain marker expression*

---

## Description

echo "Use RRA to test the association of gene knockout with certain marker expression"

## Usage

```
scmageck_rra(BARCODE, RDS, GENE, RRAPATH = NULL, LABEL = NULL, NEGCTRL = NULL,
SIGNATURE = NULL, KEEPTMP = FALSE, PATHWAY = FALSE, SAVEPATH = "./")
```

## Arguments

| | |
|---|---|
| BARCODE | A txt file to include cell identity information, generated from the cell identity collection step. |
| RDS | A Seurat object or local RDS file path that contains the scRNA-seq dataset. Note that the dataset has to be normalized and scaled. |
| GENE | Genes whose expressions are to be tested. Multiple genes can be provided, separated by ",". For example, "MKI67,TP53" |
| RRAPATH | The path to the RRA program, if RRA cannot be found in the PATH environment variable. |
| LABEL | The label of the output file. |
| NEGCTRL | The name of the negative control gene. For example, "NonTargetingControlGuideForHuman". Default is NULL (do not use any negative controls). |
| KEEPTMP | Keep temporary files. |
| PATHWAY | Treat genes in –GENE option as a pathway. In other words, the averaged expression of these genes will be used for testing. |
| SAVEPATH | The save path of result. Default save path is the current working directory. If you don't need save the result, set SAVEPATH as NULL. |

SIGNATURE A GMT text file, the format must be as follows:(1)Column 1: name of gene set; (2)Colum 2: Empty, or the information about gene set e.g. the source of the gene set; (3)Column 3 and onwards: ids of genes beloging to a particular gene set. Note that if you don't set the parameter "SAVEPATH", this parameter would create a folder called "GENE_SET"" in the current working directory to store the results from applying RRA program to do gene set analysis. Reference:http://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats

## Value

The result for object RDS

## Examples

```
### BARCODE file contains cell identity information, generated from the cell identity collection step
 BARCODE <- system.file("extdata","barcode_rec.txt",package = "scMAGeCK")

### RDS can be a Seurat object or local RDS file path that contains the scRNA-seq dataset
 RDS <- system.file("extdata","singles_dox_mki67_v3.RDS",package = "scMAGeCK")

 target_gene <- "MKI67"
### Set RRA executable file path or activate scmageck env if needed (see https://bitbucket.org/weililab/scmageck
 RRAPATH <- NULL

 rra_result <- scmageck_rra(BARCODE=BARCODE, RDS=RDS, GENE=target_gene,
                            RRAPATH=RRAPATH, LABEL='dox_mki67',
                            NEGCTRL=NULL, KEEPTMP=FALSE, SIGNATURE = NULL,
                            PATHWAY=FALSE, SAVEPATH=NULL)
 head(rra_result)
```

---

# selectPlot *generate the selection plot*

---

## Description

detect the gene regulation relationship between genes and perturbation by using RRA or LR test.

## Usage

```
selectPlot(GENE = NULL, lr_result = NULL, CUTOFF = 0.05, ADJ = "fdr",
RRA_re1 = NULL, RRA_re2 = NULL, TYPE = select.type, QUALITY = 10)

select.type
#c("lr", "rra")
```

## Arguments

| | |
|---|---|
| GENE | Genes whose expressions are to be tested under the LR test. Multiple genes can be provided, separated by ",". For example, "MKI67,TP53" |
| lr_result | The result from the scmageck-lr step. |
| CUTOFF | Determine the significant pvalue. |
| ADJ | P.adjust.methods. Choose one of correction method. c("holm","hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none") |
| RRA_re1 | RRA result from the scmageck-rra step. |
| RRA_re2 | Optional input. The second RRA result from the scmageck-rra step. Add this input to visualize gene selection under two different cell condition. |
| TYPE | The type of the scMAGeCK results. Can be either "rra" or "lr". |
| QUALITY | The number of single-cells that passes the threshold when use the RRA test, default is 10. Lower quality could improve the sensitivity but reduce accuracy. |

## Examples

```
 ### by using RRA test, take MKI67 for example
 ### only works if you have RRA installed
#selectPlot(RRA_re1 = rra_result, CUTOFF = 0.05, QUALITY = 10, ADJ = "fdr", TYPE = "rra")

 ### by using LR test, take MKI67 for example

### BARCODE file contains cell identity information, generated from the cell identity collection step
 BARCODE <- system.file("extdata","barcode_rec.txt",package = "scMAGeCK")

### RDS can be a Seurat object or local RDS file path that contains the scRNA-seq dataset
 RDS <- system.file("extdata","singles_dox_mki67_v3.RDS",package = "scMAGeCK")

lr_result <- scmageck_lr(BARCODE=BARCODE, RDS=RDS, LABEL='dox_scmageck_lr', SIGNATURE = NULL,
     NEGCTRL = 'NonTargetingControlGuideForHuman', PERMUTATION = 1000, SAVEPATH=NULL, LAMBDA=0.01)
selectPlot(GENE = "MKI67", lr_result = lr_result, CUTOFF = 0.05, ADJ = "fdr", TYPE = "lr")
```

# Index

∗ **package**
scMAGeCK-package, 2

featurePlot, 3

pre_processRDS, 4

scmageck *(*scMAGeCK-package*)*, 2
scMAGeCK-package, 2
scmageck_lr, 5
scmageck_rra, 6
selectPlot, 7