

# Package ‘CEDA’

May 7, 2026

**Type** Package

**Title** CRISPR Screen and Gene Expression Differential Analysis

**Version** 1.1.1

**Description** Provides analytical methods for analyzing CRISPR screen data at different levels of gene expression. Multi-component normal mixture models and EM algorithms are used for modeling.

**Depends** R(>= 3.5.0), limma

**Imports** stats, mixtools, ggplot2, dplyr, ggsci, ggridges, ggprism

**Suggests** knitr, rmarkdown

**License** Apache License (== 2.0)

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.2.1

**VignetteBuilder** knitr, rmarkdown

**NeedsCompilation** no

**Author** Lianbo Yu [aut, cre],  
Yue Zhao [aut],  
Kevin R. Coombes [aut],  
Lang Li [aut]

**Maintainer** Lianbo Yu <Lianbo.Yu@osumc.edu>

**Repository** CRAN

**Date/Publication** 2024-02-27 06:10:02 UTC

## Contents

alphaBeta . . . . .	2
calculateGeneLFC . . . . .	2
calculateGenePval . . . . .	3
densityPlot . . . . .	3
EMFit . . . . .	4

makeRhoNull . . . . .	4
mda231 . . . . .	5
medianNormalization . . . . .	5
normalMM . . . . .	6
permuteLimma . . . . .	7
preparePlotData . . . . .	7
ridgePlot . . . . .	8
runLimma . . . . .	9
scatterPlot . . . . .	9

## Index 11

---

alphaBeta *Calculating a significance score of a gene based on the corresponding sgRNAs' p-values of the gene.*

---

### Description

Code was adapted from R package gscreend.

### Usage

```
alphaBeta(pvec)
```

### Arguments

pvec            A numeric vector of p-values.

### Value

A min value of the kth smallest value based on the beta distribution  $B(k, n-k+1)$ , where the n is the number of probabilities in the vector. This min value is the significance score of the gene.

---

calculateGeneLFC *Calculating gene-level log fold ratios*

---

### Description

Log fold ratios of all sgRNAs of a gene are averaged to obtain the gene level log fold ratio.

### Usage

```
calculateGeneLFC(lfcs, genes)
```

### Arguments

lfcs            A numeric vector containing log fold change of sgRNAs.  
genes           A character string containing gene names corresponding to sgRNAs.

**Value**

A numeric vector containing log fold ratio of genes.

---

calculateGenePval	<i>Calculating gene level p-values using modified robust rank aggregation (alpha-RRA method) on sgRNAs' p-values</i>
-------------------	--

---

**Description**

Code was adapted from R package gscreend. The alpha-RRA method is adapted from MAGeCK.

**Usage**

```
calculateGenePval(pvec, genes, alpha, nperm = 20)
```

**Arguments**

pvec	A numeric vector containing p-values of sgRNAs.
genes	A character string containing gene names corresponding to sgRNAs.
alpha	A numeric number denoting the alpha cutoff (i.e. 0.05).
nperm	Number of permutations, default is 20

**Value**

A list with four elements: 1) a list of genes with their p-values; 2) a numeric matrix of rho null, each column corresponding to a different number of sgRNAs per gene; 3) a numeric vector of rho; 4) a numeric vector of number of sgRNAs per gene.

---

densityPlot	<i>2D density contour plot of gene log2 fold ratios against gene expression levels</i>
-------------	--

---

**Description**

This function generates a scatter plot with 2D density contour of log2 fold ratios of sgRNAs against the corresponding gene expression levels.

**Usage**

```
densityPlot(data, ...)
```

**Arguments**

data	A data frame from the output of preparePlotData function
...	Other graphical parameters

**Value**

No return value

---

EMFit	<i>Fitting multi-component normal mixture models by R package mixtools</i>
-------	--

---

**Description**

The function normalmixEM in R package mixtools is employed for fitting multi-component normal mixture models.

**Usage**

```
EMFit(x, k0, mean_constr, sd_constr, npara, d0)
```

**Arguments**

x	A numeric vector
k0	Number of components in the normal mixture model
mean_constr	A constrain on means of components
sd_constr	A constrain on standard deviations of components
npara	Number of parameters
d0	Number of times for fitting mixture model using different starting values

**Value**

Normal mixture model fit and BIC value of the log-likelihood

---

makeRhoNull	<i>Generating the null distribution of the significance score of a gene.</i>
-------------	--

---

**Description**

Code was adapted from R package gscreend.

**Usage**

```
makeRhoNull(n, p, nperm)
```

**Arguments**

n	An integer representing sgRNA number of a gene.
p	A numeric vector which contains the percentiles of the p-values that meet the cut-off (alpha).
nperm	Number of permutation runs.

**Value**

A numeric vector which contains all the significance scores (rho) of genes generated by a permutation test where the sgRNAs are randomly assigned to genes.

---

mda231	<i>CRISPR screen data of cell line MDA-MB-231.</i>
--------	--

---

**Description**

A dataset containing the expression data of sgRNAs in a CRISPR screen experiment of cell line MDA-MB-231.

**Usage**

```
mda231
```

**Format**

A data frame with a list of two elements:

**sgRNA** Raw Read counts of sgRNAs

**negene** A list of non-essential genes

---

medianNormalization	<i>Median normalization of sgRNA counts</i>
---------------------	---

---

**Description**

This function adjusts sgRNA counts by the median ratio method. The normalized sgRNA read counts are calculated as the raw read counts divided by a size factor. The size factor is calculated as the median of all size factors calculated from negative control sgRNAs (eg., sgRNAs corresponding to non-targeting or non-essential genes).

**Usage**

```
medianNormalization(data, control)
```

**Arguments**

data	A numeric matrix containing raw read counts of sgRNAs with rows corresponding to sgRNAs and columns corresponding to samples.
control	A numeric matrix containing raw read counts of negative control sgRNAs with rows corresponding to sgRNAs and columns corresponding to samples. Sample ordering is the same as in data.

**Value**

A list with two elements: 1) size factors of all samples; 2) normalized counts of sgRNAs.

**Examples**

```
count <- matrix(rnbinom(5000 * 6, mu=500, size=3), ncol = 6)
colnames(count) = paste0("sample", 1:6)
rownames(count) = paste0("sgRNA", 1:5000)
control <- count[1:100,]
normalizedcount <- medianNormalization(count, control)
```

---

normalMM

*Performing empirical Bayes modeling on limma results*


---

**Description**

This function perform an empirical Bayes modeling on log fold ratios and return the posterior log fold ratios.

**Usage**

```
normalMM(data, theta0, n.b = 5, d = 10)
```

**Arguments**

data	A numeric matrix containing limma results and log <sub>2</sub> gene expression levels that has a column named 'lfc' and a column named 'exp.level.log2'
theta0	Standard deviation of log <sub>2</sub> fold changes under permutations
n.b	Number of bins, default is 5 bins
d	Number of times for fitting mixture model using different starting values, default is 10

**Value**

A numeric matrix containing limma results, RNA expression levels, posterior log<sub>2</sub> fold ratio, log p-values, and estimates of mixture model

---

permutelimma	<i>Modeling CRISPR data with a permutation test between conditions by R package limma</i>
--------------	---

---

**Description**

The lmFit function in R package limma is employed for group comparisons under permutations.

**Usage**

```
permutelimma(data, design, contrast.matrix, nperm)
```

**Arguments**

data	A numeric matrix containing log2 expression level of sgRNAs with rows corresponding to sgRNAs and columns to samples.
design	A design matrix with rows corresponding to samples and columns to coefficients to be estimated.
contrast.matrix	A matrix with columns corresponding to contrasts.
nperm	Number of permutations

**Value**

A numeric matrix containing log2 fold changes with permutations

**Examples**

```
y <- matrix(rnorm(1000*6),1000,6)
condition <- gl(2,3,labels=c("Control","Baseline"))
design <- model.matrix(~ 0 + condition)
contrast.matrix <- makeContrasts("conditionControl-conditionBaseline",levels=design)
fit <- permutelimma(y,design,contrast.matrix,20)
```

---

preparePlotData	<i>Prepare data for density plot and ridge plot</i>
-----------------	---

---

**Description**

Input a data frame with each gene one row, and geneID, geneLFC, geneFDR as columns. This function will stratify genes into five groups based on their FDR levels:  $\leq 0.001$ , (0.001,0.01], (0.01,0.05], (0.05,0.5], (0.5,1]

**Usage**

```
preparePlotData(data, gene.fdr)
```

**Arguments**

data	A data frame containing each gene in one row, and at least three columns with geneID, geneLFC, and geneFDR.
gene.fdr	A numeric variable (column) in the data frame, corresponding to the gene level FDR

**Value**

A data frame based on the original data frame, with an additional column "group" indicating which FDR group this gene belongs to.

---

ridgePlot	<i>Density ridgeline plot of gene expression levels for different FDR groups.</i>
-----------	---

---

**Description**

This function generates a density ridgeline plot of gene expression levels for different FDR groups.

**Usage**

```
ridgePlot(data, ...)
```

**Arguments**

data	A data frame from the output of preparePlotData function
...	Other graphical parameters

**Value**

No return value

---

`runLimma`*Modeling CRISPR screen data by R package limma*

---

**Description**

The `lmFit` function in R package `limma` is employed for group comparisons.

**Usage**

```
runLimma(data, design, contrast.matrix)
```

**Arguments**

`data` A numeric matrix containing log2 expression levels of sgRNAs with rows corresponding to sgRNAs and columns corresponding to samples.

`design` A design matrix with rows corresponding to samples and columns corresponding to coefficients to be estimated.

`contrast.matrix` A matrix with columns corresponding to contrasts.

**Value**

A data frame with rows corresponding to sgRNAs and columns corresponding to limma results

**Examples**

```
y <- matrix(rnorm(1000*6),1000,6)
condition <- gl(2,3,labels=c("Treatment","Baseline"))
design <- model.matrix(~ 0 + condition)
contrast.matrix <- makeContrasts("conditionTreatment-conditionBaseline",levels=design)
limma.fit <- runLimma(y,design,contrast.matrix)
```

---

`scatterPlot`*Scatter plot of log2 fold ratios against gene expression levels*

---

**Description**

This function generates a scatter plot of log2 fold ratios of sgRNAs against the corresponding gene expression levels.

**Usage**

```
scatterPlot(data, fdr, ...)
```

**Arguments**

<code>data</code>	A numeric matrix from the output of <code>normalMM</code> function
<code>fdr</code>	A level of false discovery rate
<code>...</code>	Other graphical parameters

**Value**

No return value

# Index

## \* datasets

mda231, 5

alphaBeta, 2

calculateGeneLFC, 2

calculateGenePval, 3

densityPlot, 3

EMFit, 4

makeRhoNull, 4

mda231, 5

medianNormalization, 5

normalMM, 6

permutelimma, 7

preparePlotData, 7

ridgePlot, 8

runLimma, 9

scatterPlot, 9