Package ‘CelliD’

March 11, 2024

Type Package

Title Unbiased Extraction of Single Cell gene signatures using Multiple Correspondence Analysis

Version 1.10.1

Description CelliD is a clustering-free multivariate statistical method for the robust extraction of per-cell gene signatures from single-cell RNA-seq. CelliD allows unbiased cell identity recognition across different donors, tissues-of-origin, model organisms and single-cell omics protocols. The package can also be used to explore functional pathways enrichment in single cell data.

Depends R (>= 4.1), Seurat (>= 4.0.1), SingleCellExperiment

License GPL-3 + file LICENSE

Encoding UTF-8

LazyData true

Imports Rcpp, RcppArmadillo, stats, utils, Matrix, tictoc, scater, stringr, irlba, data.table, glue, pbapply, umap, Rtsne, reticulate, fastmatch, matrixStats, ggplot2, BiocParallel, SummarizedExperiment, fgsea

Suggests knitr, rmarkdown, BiocStyle, testthat, tidyverse, ggpubr, destiny, ggrepel

VignetteBuilder knitr

RoxygenNote 7.1.1

biocViews RNASeq, SingleCell, DimensionReduction, Clustering, GeneSetEnrichment, GeneExpression, ATACSeq

LinkingTo Rcpp, RcppArmadillo

git_url https://git.bioconductor.org/packages/CelliD

git_branch RELEASE_3_18

git_last_commit 5dc220a

git_last_commit_date 2023-11-30

Repository Bioconductor 3.18

Date/Publication 2024-03-11
Author Akira Cortal [aut, cre],
Antonio Rausell [aut, ctb]

Maintainer Akira Cortal <akira.cortal@institutimagine.org>

R topics documented:

CelliD-package ..................................................... 2
checkCelliDArg ........................................................ 4
DimPlotMC ............................................................... 5
DistSort ............................................................... 6
fgseaCelliD ............................................................. 6
GetCellGeneDistance .................................................. 7
GetCellGeneRanking ................................................... 8
GetCellGeneSet ......................................................... 9
GetGeneCellCoordinates ............................................... 10
GetGroupCoordinates .................................................. 11
GetGroupGeneDistance ................................................ 12
GetGroupGeneRanking .................................................. 13
GetGroupGeneSet ....................................................... 14
GetGSEAMatrix ......................................................... 15
Hallmark ................................................................. 15
HgProteinCodingGenes ................................................ 16
import ................................................................. 16
MgProteinCodingGenes ............................................... 17
pairDist ............................................................... 17
plotReducedDimMC .................................................... 18
RunCellGSEA ........................................................... 19
RunCellHGT ............................................................. 21
RunGroupGSEA .......................................................... 22
RunMCA ................................................................. 24
RunMCDMAP ............................................................. 25
RunMCTSNE .............................................................. 27
RunMCUMAP ............................................................. 28
setDimMCSlot ........................................................... 29
seuratPbmc ............................................................. 30

Index ............................................................................. 31

CelliD-package Multiple Correspondence Analysis on Single Cell for Joint Dimension-
ality Reduction of Gene and Cell, Cells Geneset Extraction and Geneset Enrichment Analysis
**Description**

CelliD is a clustering-free multivariate statistical method for the robust extraction of per-cell gene signatures from single-cell RNA-seq. CelliD allows unbiased cell identity recognition across different donors, tissues-of-origin, model organisms and single-cell omics protocols. The package can also be used to explore functional pathways enrichment in single cell data.

**Author(s)**

**Maintainer**: Akira Cortal <akira.cortal@institutimagine.org>

Authors:

- Akira Cortal
- Antonio Rausell

**References**


- Stuart and Butler et al. Comprehensive integration of single cell data. bioRxiv (2018). [https://doi.org/10.1101/460147](https://doi.org/10.1101/460147)


**See Also**


checkCelliDArg

Check for CelliD arguments

Description

Performs multiple check of consistency of the argument provided by the user for different CelliD functions. It notably check if the provided features or cells name are actually contained in the high level object.

Usage

checkCelliDArg(X, group.by, reduction, dims, features, cells)

## S3 method for class 'Seurat'
checkCelliDArg(
  X,
  group.by = NULL,
  reduction,
  dims,
  features = NULL,
  cells = NULL
)

## S3 method for class 'SingleCellExperiment'
checkCelliDArg(
  X,
  reduction,
  dims,
  features = NULL,
  cells = NULL,
  group.by = NULL
)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Seurat or SingleCell Experiment Object</td>
</tr>
<tr>
<td>group.by</td>
<td>Name of meta.data or ColData column.</td>
</tr>
<tr>
<td>reduction</td>
<td>Which dimensionality reduction to use, must be based on MCA.</td>
</tr>
<tr>
<td>dims</td>
<td>A vector of integers indicating which dimensions to use of specified reduction embeddings and loadings.</td>
</tr>
<tr>
<td>features</td>
<td>Character vector of feature names to subset feature coordinates. If not specified will take all features available from specified reduction loadings.</td>
</tr>
<tr>
<td>cells</td>
<td>Character vector of cell names to subset cell coordinates. If not specified will take all features available from specified reduction Embeddings.</td>
</tr>
</tbody>
</table>
**DimPlotMC**

Value

list of corrected arguments if no error is thrown.

<table>
<thead>
<tr>
<th>DimPlotMC</th>
<th>Seurat DimPlot for MCA like Dimensionality Reduction</th>
</tr>
</thead>
</table>

**Description**

Small modification of the regular Seurat DimPlot function to enable plotting features for mca like dimensionality reduction. Allows to represent a set of genes of interest on top of the regular cell scatter plot. The label of the genes can be overlayed also but it is recommended to plot less than 50 genes label as it can overcrowd the plot severely.

**Usage**

```r
DimPlotMC(
  X,
  reduction = "mca",
  dims = c(1, 2),
  features = NULL,
  size.feature = 2,
  size.feature.text = 5,
  as.text = FALSE,
  ...
)
```

**Arguments**

- **X** a Seurat object
- **reduction** Which dimensionality reduction to use. If not specified, searches for mca.
- **dims** Dimensions to plot, must be a two-length numeric vector specifying x- and y-dimensions
- **features** character vector of features to plot, must be present in the specified dimension loadings
- **size.feature** integer indicating size of geom_point for features
- **size.feature.text** integer indicating size of geom_text for features
- **as.text** logical indicating as to include text label for feature plotting, will produce warning if TRUE and length(features) > 50
- **...** Other arguments passed to DimPlot

**Value**

A ggplot object
Examples

```r
seuratPbmc <- RunMCA(seuratPbmc, nmc = 5)
seuratPbmc <- DimPlotMC(seuratPbmc, features = Seurat::VariableFeatures(seuratPbmc))
```

DistSort

Sort Gene Cell Distance Matrix

**Description**

Sort Gene Cell Distance Matrix

**Usage**

```r
DistSort(distance)
```

**Arguments**

- `distance`: distance matrix with features at rows and cell at columns

**Value**

list of ranking of genes by cells

fgseaCelliD

Slight change in fgsea for ram and speed efficiency in CelliD

**Description**

Slight change in fgsea for ram and speed efficiency in CelliD

**Usage**

```r
fgseaCelliD(
  pathways,
  stats,
  nperm = 1000,
  minSize = 10,
  maxSize = 500,
  gseaParam = 0
)
```
Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pathways</td>
<td>List of gene sets to check</td>
</tr>
<tr>
<td>stats</td>
<td>Named vector of gene-level stats. Names should be the same as in 'pathways'</td>
</tr>
<tr>
<td>nperm</td>
<td>Number of permutations to do. Minimal possible nominal p-value is about 1/nperm</td>
</tr>
<tr>
<td>minSize</td>
<td>Minimal size of a gene set to test. All pathways below the threshold are excluded.</td>
</tr>
<tr>
<td>maxSize</td>
<td>Maximal size of a gene set to test. All pathways above the threshold are excluded.</td>
</tr>
<tr>
<td>gseaParam</td>
<td>GSEA parameter value, all gene-level stats are raised to the power of 'gseaParam' before calculation of GSEA enrichment scores</td>
</tr>
</tbody>
</table>

Value

A table with GSEA results. Each row corresponds to a tested pathway. The columns are the following:

- pathway – name of the pathway as in ‘names(pathway)’;
- pval – an enrichment p-value;
- padj – a BH-adjusted p-value;
- ES – enrichment score, same as in Broad GSEA implementation;
- NES – enrichment score normalized to mean enrichment of random samples of the same size;
- nMoreExtreme – a number of times a random gene set had a more extreme enrichment score value;
- size – size of the pathway after removing genes not present in ‘names(stats)’;

Examples

```r
seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
ranking <- GetCellGeneRanking(seuratPbmc, reduction = "mca", dims = 1:5)
fgseaCelliD(pathways = Hallmark, stats = ranking[[1]])
```

GetCellGeneDistance  Distance Calculation

Description

Small intermediate function for euclidean distance calculation between MCA feature coordinates and cell coordinates. Due to MCA pseudo barycentric relationship, the closer a gene g is to a cell c, the more specific to such a cell it can be considered.
GetCellGeneRanking

Usage

GetCellGeneDistance(X, reduction, dims, features, cells)

## S3 method for class 'Seurat'
GetCellGeneDistance(X, reduction = "mca", dims, features = NULL, cells = NULL)

## S3 method for class 'SingleCellExperiment'
GetCellGeneDistance(X, reduction = "MCA", dims, features = NULL, cells = NULL)

Arguments

- **X**: Seurat or SingleCell Experiment Object
- **reduction**: Which dimensionality reduction to use, must be based on MCA.
- **dims**: A vector of integers indicating which dimensions to use with reduction embedding and loading for distance calculation.
- **features**: Character vector of feature names to subset feature coordinates. If not specified will take all features available from specified reduction loading.
- **cells**: Character vector of cell names to subset cell coordinates. If not specified will take all cells available from specified reduction embedding.

Value

Distance Matrix with genes at row and cells at column

GetCellGeneRanking Raking Extraction

Description

Intermediate function for ranking extraction from Cell Gene Distance Matrix. Genes are ordered from the most specific to the least specific to the cell according to their euclidean distances. Value indicates the euclidean distances between the cell and the genes in the MCA coordinates.

Usage

GetCellGeneRanking(X, reduction, dims, features, cells)

## S3 method for class 'Seurat'
GetCellGeneRanking(
  X,
  reduction = "mca",
  dims = seq(50),
  features = NULL,
  cells = NULL
)
GetCellGeneRanking

## S3 method for class 'SingleCellExperiment'
GetCellGeneRanking(
  X,
  reduction = "MCA",
  dims = seq(50),
  features = NULL,
  cells = NULL
)

### Arguments

- **X**: Seurat or SingleCellExperiment Object
- **reduction**: Which dimensionality reduction to use, must be based on MCA.
- **dims**: A vector of integers indicating which dimensions to use with reduction embedding and loading for distance calculation.
- **features**: Character vector of feature names to subset feature coordinates. If not specified will take all features available from specified reduction. Loading
- **cells**: Character vector of cell names to subset cell coordinates. If not specified will take all features available from specified reduction. Embedding.

### Value

A cell named list of gene rankings ordered by distances from shortest (most specific) to farthest (less specific)

### Examples

```r
seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
ranking <- GetCellGeneRanking(seuratPbmc, reduction = "mca", dims = 1:5)
```

GetCellGeneSet

### Description

Calculate cells and genes distances, rank them per cell and extract top n features. The obtained top n features represents features that are highly specific to that cell.

### Usage

```r
GetCellGeneSet(X, reduction = "mca", dims, features, cells, n.features)
```

### Examples

```r
seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
ranking <- GetCellGeneRanking(seuratPbmc, reduction = "mca", dims = 1:5)
```
GetGeneCellCoordinates

dims = seq(50),
features = NULL,
cells = NULL,
n.features = 200
)

## S3 method for class 'SingleCellExperiment'
GetCellGeneSet(
  X,
  reduction = "MCA",
dims = seq(50),
features = NULL,
cells = NULL,
n.features = 200
)

Arguments

X Seurat or SingleCell Experiment Object
reduction Which dimensionality reduction to use, must be based on MCA.
dims A vector of integers indicating which dimensions to use with reduction embeddings and loadings for distance calculation.
features Character vector of feature names to subset feature coordinates. If not specified will take all features available from specified reduction Loadings
cells Character vector of cell names to subset cell coordinates. If not specified will take all features available from specified reduction Embeddigns.
n.features single integer specifying how many top features should be extracted from the ranking

Value

A cell named list of gene rankings ordered by distances from shortest (most specific) to farthest (less specific)

Examples

seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
GroupGeneRanking <- GetGroupGeneRanking(seuratPbmc, group.by = "seurat_clusters", dims = 1:5)

GetGeneCellCoordinates

Description

Get coordinates of both cells and features in a matrix
GetGroupCoordinates

Usage

GetGeneCellCoordinates(X, reduction, dims, features)

Arguments

X Seurat or SingleCellExperiment Object
reduction Which dimensionality reduction to use, must be based on MCA.
dims A vector of integers indicating which dimensions to use with reduction embeddings and loadings for distance calculation.
features Character vector of feature names to subset feature coordinates. If not specified will take all features available from specified reduction Loadings.

Value

A matrix with gene and cell coordinates of MCA

GetGroupCoordinates

Description

Centroids calculation for a given group of cells defined for instance by cell type/condition.

Usage

GetGroupCoordinates(X, group.by, reduction, dims, ...)

## S3 method for class 'matrix'
GetGroupCoordinates(X, group.by, reduction = NULL, dims, ...)

## S3 method for class 'Seurat'
GetGroupCoordinates(X, group.by = NULL, reduction = "mca", dims = seq(50), ...)

## S3 method for class 'SingleCellExperiment'
GetGroupCoordinates(X, group.by = NULL, reduction = "MCA", dims, ...)

Arguments

X Seurat or SingleCellExperiment object, alternatively a matrix.
group.by column name of meta.data (Seurat) or ColData (SingleCellExperiment). For Seurat object if NULL active.ident slot will be taken.
reduction Which dimensionality reduction to use, must be based on MCA.
dims A vector of integers indicating which dimensions to use with reduction embeddings and loadings for distance calculation.
... Other arguments passed to methods
GetGroupGeneDistance

Value
A data.table with coordinates of the group centroids for the specified dims.

Description
Distance calculation between genes and group of cells centroids.

Usage
GetGroupGeneDistance(X, group.by, reduction, dims, features)

## S3 method for class 'Seurat'
GetGroupGeneDistance(
  X,
  group.by = NULL,
  reduction = "mca",
  dims = seq(50),
  features = NULL
)

## S3 method for class 'SingleCellExperiment'
GetGroupGeneDistance(
  X,
  group.by,
  reduction = "MCA",
  dims = seq(50),
  features = NULL
)

Arguments
- **X**: Seurat or SingleCellExperiment object, alternatively a matrix.
- **group.by**: column name of meta.data (Seurat) or ColData (SingleCellExperiment)
- **reduction**: Which dimensionality reduction to use, must be based on MCA.
- **dims**: A vector of integers indicating which dimensions to use with reduction embeddings and loadings for distance calculation.
- **features**: A character vector of features name to subset feature coordinates for distance calculation.

Value
Distance Matrix between groups (column) and genes (row)
**GetGroupGeneRanking**

**Gene Specificity Ranking Calculation**

**Description**

Gene Specificity Ranking Calculation

**Usage**

```r
GetGroupGeneRanking(X, group.by, reduction, dims, features)
```

```r
## S3 method for class 'Seurat'
GetGroupGeneRanking(
  X,
  group.by = NULL,
  reduction = "mca",
  dims = seq(50),
  features = NULL
)
```

```r
## S3 method for class 'SingleCellExperiment'
GetGroupGeneRanking(
  X,
  group.by,
  reduction = "MCA",
  dims = seq(50),
  features = NULL
)
```

**Arguments**

- **X**: Seurat or SingleCellExperiment object, alternatively a matrix.
- **group.by**: column name of meta.data (Seurat) or ColData (SingleCellExperiment)
- **reduction**: Which dimensionality reduction to use, must be based on MCA.
- **dims**: A vector of integers indicating which dimensions to use with reduction embeddings and loadings for distance calculation.
- **features**: A character vector of features name to subset feature coordinates for distance calculation.

**Value**

List of genes ranking for each groups

**Examples**

```r
seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
GroupGeneRanking <- GetGroupGeneRanking(seuratPbmc, group.by = "seurat_clusters", dims = 1:5)
```
GetGroupGeneSet

Extract cluster/group gene sets from MCA

Description

Extract cluster/group gene sets from MCA

Usage

GetGroupGeneSet(X, group.by, reduction, dims, features, n.features)

## S3 method for class 'Seurat'
GetGroupGeneSet(
  X,
  group.by = NULL,
  reduction = "mca",
  dims = seq(50),
  features = NULL,
  n.features = 200
)

## S3 method for class 'SingleCellExperiment'
GetGroupGeneSet(
  X,
  group.by = NULL,
  reduction = "MCA",
  dims = seq(50),
  features = NULL,
  n.features = 200
)

Arguments

X Seurat or SingleCellExperiment object, alternatively a matrix.
group.by column name of meta.data (Seurat) or ColData (SingleCellExperiment).
reduction Which dimensionality reduction to use, must be based on MCA.
dims A vector of integers indicating which dimensions to use with reduction for distance calculation.
features A character vector of features name to subset feature coordinates for distance calculation.
n.features A single integer specifying how many top features will be extracted from ranking.

Value

Distance Matrix between groups (column) and genes (row)
GetGSEAMatrix

Examples

```r
seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
GroupGeneSet <- GetGroupGeneSet(seuratPbmc, dims = 1:5, group.by = "seurat_clusters")
```

---

GetGSEAMatrix

Get Matrix from Enrichment Results

Description

Extract enrichment score Matrix from RunGSEA functions.

Usage

```
GetGSEAMatrix(X, metric = "ES")
```

Arguments

- **X**: an enrichment results obtained by RunGroupGSEA or RunCellGSEA
- **metric**: a character indicating which metric to use as value of matrix (ES, NES, padj, pval)

Value

A matrix of geneset enrichment metric with cell/group at columns and pathways/genesets at rows

Examples

```r
seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
GSEAResults <- RunGroupGSEA(seuratPbmc, Hallmark, group.by = "seurat_clusters", dims = 1:5)
GSEAMatrix <- GetGSEAMatrix(GSEAResults)
```

---

Hallmark

Hallmark Pathways from MSigDB

Description

A dataset containing the Hallmark gene sets from MSigDB.

Usage

```
Hallmark
```

Format

A named list of length 50 containing Hallmark gene sets.
Source

http://software.broadinstitute.org/gsea/msigdb/download_file.jsp?filePath=/resources/msigdb/6.2/h.all.v6.2.symbols.gmt

References


HgProteinCodingGenes  Homo Sapiens Protein Coding Genes

Description

A gene list of human protein coding genes extracted from biomaRt.

Usage

HgProteinCodingGenes

Format

A list of 19308 gene ontology terms with the corresponding genes.

Source

http://software.broadinstitute.org/gsea/msigdb/collections.jsp#C5

References


import  Import

Description

Import

Usage

import()

Value

updates NAMESPACE import
MgProteinCodingGenes

MgProteinCodingGenes  Mus Musculus Protein Coding Genes

Description
A gene list of mouse protein coding genes extracted from biomaRt.

Usage
MgProteinCodingGenes

Format
A list of 3857 gene ontology terms with the corresponding genes.

Source
http://software.broadinstitute.org/gsea/msigdb/collections.jsp#C5

References

pairDist  Distance Calculation

Description
Small function to calculate quickly the distance between rows of two matrix.

Usage
pairDist(x, y)

Arguments
  x  a matrix
  y  a matrix

Value
A Distance Matrix
plotReducedDimMC

Scater plotReducedDim for MCA like dimensionality Reduction

Description

Small modification of the Scater plotReducedDim function to enable plotting features for mca like dimensionality reduction. Allows to represent a set of genes of interest on top of the regular cell scatter plot. The label of the genes can be overlaid also but it is recommended to plot less than 50 genes label as it can overcrowd the plot severely.

Usage

plotReducedDimMC(
  X,
  reduction = "MCA",
  dims = c(1, 2),
  features = NULL,
  size.feature = 3,
  size.feature.text = 5,
  as.text = FALSE,
  ...
)

Arguments

X a Single Cell Experiment Object
reduction Which dimensionality reduction to use. If not specified, searches for mca.
dims Dimensions to plot, must be a two-length numeric vector specifying x- and y-dimensions
features character vector of features to plot, must be present in the specified dimension loadings
size.feature integer indicating size of geom_point for features
size.feature.text integer indicating size of geom_text for features
as.text logical indicating as to include text label for feature plotting, will produce warning if TRUE and length(features) > 50.
... Other arguments passed to plotReducedDim

Value

A ggplot object

Examples

scePBMC <- as.SingleCellExperiment(seuratPbmc)
scePBMC <- RunMCA(scePBMC, nmcs = 5)
plotReducedDimMC(scePBMC)
RunCellGSEA

Run Gene Set Enrichment Analysis on cells

Description
Calculate cells gene specificity ranking and then perform geneset enrichment analysis (fgsea) on it. However, due to the very long running time of gene set enrichment analysis, we recommend the usage of RunCellHGT.

Usage

RunCellGSEA(
  X, 
  pathways, 
  reduction, 
  dims, 
  features, 
  cells, 
  nperm, 
  minSize, 
  maxSize, 
  gseaParam, 
  n.core
)

## S3 method for class 'Seurat'
RunCellGSEA(
  X, 
  pathways, 
  reduction = "mca", 
  dims = seq(50), 
  features = NULL, 
  cells = NULL, 
  nperm = 1000, 
  minSize = 10, 
  maxSize = 500, 
  gseaParam = 0, 
  n.core = 1
)

## S3 method for class 'SingleCellExperiment'
RunCellGSEA(
  X, 
  pathways, 
  reduction = "mca", 
  dims = seq(50), 
  features = NULL,
RunCellGSEA

cells = NULL,
nperm = 1000,
minSize = 10,
maxSize = 500,
gseaParam = 0,
n.core = 1
)

Arguments

X Seurat or SingleCellExperiment object
pathways List of gene sets to check
reduction Which dimensionality reduction to use, must be based on MCA.
dims A vector of integers indicating which dimensions to use with reduction embeddings and loadings for distance calculation.
features Character vector of feature names to subset feature coordinates. If not specified will take all features available from specified reduction Loadings.
cells Character vector of cell names to subset cell coordinates. If not specified will take all features available from specified reduction Embeddings
nperm Number of permutations to do. Minimal possible nominal p-value is about 1/nperm
minSize Minimal size of a gene set to test. All pathways below the threshold are excluded.
maxSize Maximal size of a gene set to test. All pathways above the threshold are excluded.
gseaParam GSEA parameter value, all gene-level stats are raised to the power of ‘gseaParam’ before calculation of GSEA enrichment scores
n.core A single integer to specify the number of core for parallelisation.

Value

A data.table with geneset enrichment analysis statistics.

Examples

seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
GSEAResults <- RunCellGSEA(seuratPbmc, Hallmark, dims = 1:5)
RunCellHGT

**Run HyperGeometric Test on cells**

**Description**

RunCellHGT calculates the gene signatures for each cells and performs hypergeometric test against a user defined gene signatures/pathways (named list of genes). It returns a score of enrichment in the form of \(-\log_{10} p\text{value}\) (see `log.trans` argument). The obtained matrix can then be integrated in Seurat or SingleCellExperiment object. It can notably be used with cell type signatures to predict cell types or with functional pathways.

**Usage**

```r
RunCellHGT(
  X,
  pathways,
  reduction,
  n.features,
  features,
  dims,
  minSize,
  log.trans,
  p.adjust
)
```

```r
## S3 method for class 'SingleCellExperiment'
RunCellHGT(
  X,
  pathways,
  reduction = "MCA",
  n.features = 200,
  features = NULL,
  dims = seq(50),
  minSize = 10,
  log.trans = TRUE,
  p.adjust = TRUE
)
```

```r
## S3 method for class 'Seurat'
RunCellHGT(
  X,
  pathways,
  reduction = "mca",
  n.features = 200,
  features = NULL,
  dims = seq(50),
  minSize = 10,
  log.trans = TRUE,
  p.adjust = TRUE
)
```
Arguments

X Seurat or SingleCellExperiment object with mca performed
pathways geneset to perform hypergeometric test on (named list of genes)
reduction name of the MCA reduction
n.features integer of top n features to consider for hypergeometric test
features vector of features to calculate the gene ranking by default will take everything in the selected mca reduction.
dims MCA dimensions to use to compute n.features top genes.
minSize minimum number of overlapping genes in geneset and
log.trans if TRUE transform the pvalue matrix with -log10 and convert it to sparse matrix
p.adjust if TRUE apply Benjamini Hochberg correction to p-value

Value
a matrix of benjamini hochberg adjusted pvalue pvalue or a sparse matrix of (-log10) benjamini hochberg adjusted pvalue

Examples
seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
Enrichment <- RunCellHGT(X = seuratPbmc, pathways = Hallmark, dims = 1:5)

Description
Calculate group gene specificity ranking and then perform geneset enrichment analysis on it.

Usage
RunGroupGSEA(
  X,  
  pathways,  
  group.by,  
  reduction,  
  dims,  
  features,  
  nperm,  
  minSize,
RunGroupGSEA

\[
\text{maxSize,}
\text{gseaParam}
\]

## S3 method for class 'Seurat'
RunGroupGSEA(
  \text{X, pathways, group.by = NULL, reduction = "mca",}
  \text{dims = seq(50), features = NULL, nperm = 1000,}
  \text{minSize = 10, maxSize = 500, gseaParam = 0}
)

## S3 method for class 'SingleCellExperiment'
RunGroupGSEA(
  \text{X, pathways, group.by, reduction = "MCA",}
  \text{dims = seq(50), features = NULL, nperm = 1000,}
  \text{minSize = 10, maxSize = 500, gseaParam = 0}
)

**Arguments**

- \text{X}: pathways List of gene sets to check
- \text{pathways}: reduction Which dimensionality reduction to use, must be based on MCA.
- \text{group.by}: dims A vector of integers indicating which dimensions to use with reduction embeddings and loadings for distance calculation.
- \text{reduction}: features Character vector of feature names to subset feature coordinates. If not specified will take all features available from specified reduction Loadings.
- \text{dims}: cells Character vector of cell names to subset cell coordinates. If not specified will take all features available from specified reduction Embeddings
- \text{features}: cells Character vector of cell names to subset cell coordinates. If not specified will take all features available from specified reduction Embeddings
- \text{nperm}: nperm Number of permutations to do. Minimal possible nominal p-value is about 1/nperm
RunMCA

Run Multiple Correspondence Analysis

Description

RunMCA allows to compute the Multiple Correspondence Analysis on the single cell data contained in Seurat or SingleCellExperiment. MCA is a statistical technique close to PCA that provides a simultaneous representation of observations (e.g. cells) and variables (e.g. genes) in low-dimensional space. The barycentric relation among cells and genes is a distinctive feature of MCA biplots and represents a major advantage as compared to other types of biplots such as those produced by Principal Component Analysis as well as over alternative low-dimensional transformations providing only cell projections. Thus, in the MCA biplot, analytical distances can be calculated not only between cells and between genes, but also between each cell and each gene in order to estimate its association. Thus, the closer a gene g is to a cell c, the more specific to such a cell it can be considered. Gene-to-cell distances can then be ranked for each individual cell, and the top-ranked genes may be regarded as a unique gene signature representing the identity card of the cell.

Usage

RunMCA(X, nmcs, features, reduction.name, slot, ...)

## S3 method for class 'matrix'
RunMCA(X, nmcs = 50, features = NULL, reduction.name = "MCA", ...)

## S3 method for class 'Seurat'
RunMCA(
  X,
  nmcs = 50,
  features = NULL,
  reduction.name = "mca",
  slot = "data",

minSize

minSize Minimal size of a gene set to test. All pathways below the threshold are excluded.

maxSize

maxSize Maximal size of a gene set to test. All pathways above the threshold are excluded.

gseaParam

gseaParam GSEA parameter value, all gene-level stats are raised to the power of 'gseaParam' before calculation of GSEA enrichment scores

Value

A data.table with geneset enrichment analysis statistics.

Examples

seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
GSEAResults <- RunGroupGSEA(seuratPbmc, Hallmark, group.by = "seurat_clusters", dims = 1:5)
## RunMCDMAP

```r
assay = DefaultAssay(X),
...
)

## S3 method for class 'SingleCellExperiment'
RunMCA(
  X,
  nmcs = 50,
  features = NULL,
  reduction.name = "MCA",
  slot = "logcounts",
  ...
)
```

### Arguments

- **X**: Seurat, SingleCellExperiment or matrix object
- **nmcs**: number of components to compute and store, default set to 30
- **features**: character vector of feature names. If not specified all features will be taken.
- **reduction.name**: name of the reduction default set to 'MCA' for SingleCellExperiment and mca
- **slot**: Which slot to pull expression data from? Default to logcounts for SingleCellExperiment and data for Seurat.
- **...**: other arguments passed to methods
- **assay**: Name of Assay MCA is being run on

### Value

Seurat or SCE object with MCA calculation stored in the reductions slot.

### Examples

```r
seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
```

---

**RunMCDMAP**

**Diffusion Map on MCA coordinates**

---

### Description

(!EXPERIMENTAL) Run DiffusionMap on MCA cell and feature coordinates. This will allow to draw the trajectory of both cells and the genes at the same time.
Usage

RunMCDMAP(X, reduction, features, dims, reduction.name, ...)

## S3 method for class 'Seurat'
RunMCDMAP(
  X,
  reduction = "mca",
  features = NULL,
  dims = seq(50),
  reduction.name = "mcdmap",
  assay = DefaultAssay(X),
  ...
)

## S3 method for class 'SingleCellExperiment'
RunMCDMAP(
  X,
  reduction = "MCA",
  features = NULL,
  dims = seq(50),
  reduction.name = "MCDMAP",
  ...
)

Arguments

X Seurat or SingleCellExperiment object
reduction Which dimensionality reduction to use, must be based on MCA.
features Character vector of feature names to subset feature coordinates. If not specified
  will take all features available from specified reduction Loadings.
dims A vector of integers indicating which dimensions to use with reduction embeddings
  and loadings for distance calculation.
reduction.name name of the created dimensionality reduction, default set to "mca" for Seurat
  and "MCA" for SCE.
... other arguments passed to methods or DiffusionMap
assay Seurat Asssay slot name.

Value

Seurat or SingleCellExperiment object with MCDMAP stored in the reduction slot

Examples

seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
seuratPbmc <- RunMCDMAP(seuratPbmc, dims = seq(5), k = 5)
RunMCTSNE

tSNE on MCA coordinates

Description

(EXPERIMENTAL) Run TSNE on MCA features and cells coordinates. This will allow to embed in 2D both cells and the genes at the same time.

Usage

RunMCTSNE(X, reduction, dims, features, reduction.name, ...)

## S3 method for class 'Seurat'
RunMCTSNE(
  X,
  reduction = "mca",
  dims = seq(50),
  features = NULL,
  reduction.name = "mctsne",
  assay = DefaultAssay(X),
  ...
)

## S3 method for class 'SingleCellExperiment'
RunMCTSNE(
  X,
  reduction = "MCA",
  dims = seq(50),
  features = NULL,
  reduction.name = "MCTSNE",
  ...
)

Arguments

X Seurat or SingleCellExperiment object
reduction Which dimensionality reduction to use, must be based on MCA.
dims A vector of integers indicating which dimensions to use with reduction embeddings and loadings for distance calculation.
features Character vector of feature names to subset feature coordinates. If not specified will take all features available from specified reduction Loadings.
reduction.name name of the created dimensionality reduction, default set to "mca" for Seurat and "MCA" for SCE.
... other arguments passed to methods or Rtsne::Rtsne
assay Seurat assay slot. When not specified set with DefaultAssay(X)
RunMCUMAP

Value
Seurat or SingleCellExperiment object with MCTSNE stored in the reduction slot

Examples
seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
seuratPbmc <- RunMCTSNE(seuratPbmc, dims = seq(5))

RunMCUMAP  UMAP on MCA coordinates

Description
(!EXPERIMENTAL) Run UMAP on MCA features and cells coordinates. This will allow to embed
in 2D both cells and the genes at the same time.

Usage
RunMCUMAP(X, reduction, dims, features, reduction.name, ...)

## S3 method for class 'Seurat'
RunMCUMAP(
  X,
  reduction = "mca",
  dims = seq(50),
  features = NULL,
  reduction.name = "mcumap",
  assay = DefaultAssay(X),
  ...
)

## S3 method for class 'SingleCellExperiment'
RunMCUMAP(
  X,
  reduction = "MCA",
  dims = seq(50),
  features = NULL,
  reduction.name = "MCUMAP",
  ...
)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Seurat or SingleCellExperiment object</td>
</tr>
<tr>
<td>reduction</td>
<td>Which dimensionality reduction to use, must be based on MCA.</td>
</tr>
<tr>
<td>dims</td>
<td>A vector of integers indicating which dimensions to use with reduction embeddings and loadings for distance calculation.</td>
</tr>
</tbody>
</table>
setDimMCSlot

features Character vector of feature names to subset feature coordinates. If not specified will take all features available from specified reduction Loadings.

reduction.name name of the created dimensionality reduction, default set to "mca" for Seurat and "MCA" for SCE.

... other arguments passed to methods or Rtsne::Rtsne

assay Seurat assay slot to assign MCUMAP. When not specified set to DefaultAssay(X)

Value
Seurat or SingleCellExperiment object with MCUMAP stored in the reduction slot

Examples

seuratPbmc <- RunMCA(seuratPbmc, nmcs = 5)
seuratPbmc <- RunMCUMAP(seuratPbmc, dims = seq(5))

Description
Integrate MCA in Seurat and SingleCellExperiment Dimensionality reduction Slot. It will set also a small parameter inside the dimensionality reduction object to signal if it is a MCA or not.

Usage

setDimMCSlot(X, cellEmb, geneEmb, stdev, reduction.name, ...)

## S3 method for class 'Seurat'
setDimMCSlot(
  X,
  cellEmb, geneEmb, stdev = NULL, reduction.name = "mca", assay = DefaultAssay(X), ...
)

## S3 method for class 'SingleCellExperiment'
setDimMCSlot(X, cellEmb, geneEmb, stdev = NULL, reduction.name = "MCA", ...)
Arguments

- `X`: Seurat or SingleCellExperiment object
- `cellEmb`: cell coordinates returned by MCA
- `geneEmb`: feature coordinates returned by MCA
- `stdev`: eigen value returned by MCA
- `reduction.name`: name of the created dimensionality reduction, default set to 'mca' for Seurat and 'MCA' for SCE.
- `...`: other arguments passed to methods
- `assay`: Seurat assay slot

Value

Seurat or SingleCellExperiment object with MC stored in the reduction slot

Description

A subset of the PBMC3k data from Seurat vignette. Normalisation, VariableFeatures, ScaleData and PCA has already been computed with default Seurat parameter.

Usage

`seuratPbmc`

Format

A seurat object.

Source

https://s3-us-west-2.amazonaws.com/10x.files/samples/cell/pbmc3k/pbmc3k_filtered_gene_bc_matrices.tar.gz

References

Butler et al., Nature Biotechnology 2018.
Index

* SCRNASeq, MCA, GSEA, Gene Sets
  CelliD-package, 2
  RunMCDMAP, 25
  RunMCTSNE, 27
  RunMCUMAP, 28
  setDimMCSlot, 29
  seuratPbmc, 30

* datasets
  Hallmark, 15
  HgProteinCodingGenes, 16
  MgProteinCodingGenes, 17
  seuratPbmc, 30

CelliD (CelliD-package), 2
CelliD-package, 2
checkCelliDArg, 4
DimPlotMC, 5
DistSort, 6

fgseaCelliD, 6

GetCellGeneDistance, 7
GetCellGeneRanking, 8
GetCellGeneSet, 9
GetGeneCellCoordinates, 10
GetGroupCoordinates, 11
GetGroupGeneDistance, 12
GetGroupGeneRanking, 13
GetGroupGeneSet, 14
GetGSEAMatrix, 15

Hallmark, 15
HgProteinCodingGenes, 16

import, 16

MgProteinCodingGenes, 17

pairDist, 17
plotReducedDimMC, 18

RunCellGSEA, 19
RunCellHGT, 21
RunGroupGSEA, 22
RunMCA, 24