

Package ‘bt2m’

January 12, 2023

Type Package

Title A iteratively bifurcated clustering strategy for single-cell sequencing data

Version 0.4.9

Description A iteratively bifurcated clustering strategy for single-cell sequencing data.

License GPL-3 + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Depends R (>= 3.6)

Imports stats,
utils,
Seurat,
dplyr,
clustree,
ggplot2,
ComplexHeatmap,
RColorBrewer,
circlize,
fastcluster,
parallelDist,
HiClimR,
clusterProfiler,
org.Hs.eg.db,
org.Mm.eg.db,
Hmisc,
stringr,
cowplot,
plyr

Suggests testthat

URL <https://github.com/leezx/bt2m>

BugReports <https://github.com/leezx/bt2m/issues>

R topics documented:

add.missing.genes 2

AddMarkerExpressionPct	3
Bt2mBifucation.graph	3
Bt2mBifucation.hclust	4
Bt2mBifucation.kmeans	4
Bt2mColors	5
Bt2mEnrichGO	5
clusterProfilerORA	6
DataframeToVector	6
DEG.cluster.list	7
DrawBarplotGO	7
DrawBt2mClusterTree	8
DrawGOchain	8
DrawMarkerChainDotplot	9
DrawMarkerChainHeatmap	9
FindBifurcationResolution	10
gene.anno.GRCh38.v3	11
gene.anno.mm10.v3	11
GetClusterChain	12
GetInitialBifurcation	12
GetTestData	13
HGNC.gene.pairs	13
IdentifyBinaryMarkers	13
JASPAR.motif.rmDup	14
lr.pair.all.uniq.hs	14
lr.pair.all.uniq.mm	15
MGI.gene.pairs	15
OrderCluster	16
PrepareExpressionMatrix	16
RemoveDuplicatedGO	17
RemoveDuplicatedMarker	17
RenameBt2m	18
RunBt2m	18
subsetSeuratObj	19
subset_cells	19
WriteBt2mIntoSeurat	20
Index	21

add.missing.genes	<i>Add missing genes to the gene expression matrix (for multiple datasets integratoin)</i>
-------------------	--

Description

Add missing genes to the gene expression matrix (for multiple datasets integratoin)

Usage

```
add.missing.genes(epxrM, all.genes)
```

Arguments

<code>exprM</code>	expression matrix in data.frame, matrix or dgCMatrx format
<code>all.genes</code>	full list of genes

Value

A gene expression matrix in dgCMatrx format

AddMarkerExpressionPct

filter uniquely expressed markers by expression percentage

Description

filter uniquely expressed markers by expression percentage

Usage

```
AddMarkerExpressionPct(
  seuratObj,
  bt2m.cellMeta,
  bt2m.marker.chain,
  assay = "RNA",
  slot = "data"
)
```

Arguments

<code>seuratObj</code>	A Seurat object
<code>bt2m.cellMeta</code>	bt2m.cellMeta dataframe from bt2m
<code>bt2m.marker.chain</code>	bt2m.marker.chain dataframe contains all the markers
<code>assay</code>	Assay used for prediction
<code>slot</code>	slot used for prediction

Value

A new dataframe with two additional columns: `cluster_pct` (expression percentage in the cluster) and `bkg_pct` (expression percentage in the background cells)

`Bt2mBifucation.graph` *Bifurcation based on graph-based clustering*

Description

Bifurcation based on graph-based clustering

Usage

```
Bt2mBifucation.graph(seuratObj, resolution.sets = 50)
```

Arguments

<code>seuratObj</code>	A Seurat object
<code>resolution.sets</code>	The number of resolution for searching

Value

A bifurcated Seurat object (see `active.ident`).

`Bt2mBifucation.hclust` *Bifurcation based on hierarchical clustering*

Description

Bifurcation based on hierarchical clustering

Usage

```
Bt2mBifucation.hclust(seuratObj, method = "euclidean")
```

Arguments

<code>seuratObj</code>	A Seurat object
<code>method</code>	The distance measurement method "euclidean or correlation"

Value

A bifurcated Seurat object (see `active.ident`).

Bt2mBifucation.kmeans	<i>Bifurcation based on K-means clustering</i>
-----------------------	--

Description

Bifurcation based on K-means clustering

Usage

```
Bt2mBifucation.kmeans(seuratObj, method = "euclidean")
```

Arguments

seuratObj	A Seurat object
method	The distance measurement method "euclidean or correlation"

Value

A bifurcated Seurat object (see active.ident).

Bt2mColors	<i>Colors for bt2m</i>
------------	------------------------

Description

Colors for bt2m

Usage

```
Bt2mColors()
```

Value

A color vector

Bt2mEnrichGO	<i>Perform GO annotation of bt2m.marker.chain dataframe</i>
--------------	---

Description

Perform GO annotation of bt2m.marker.chain dataframe

Usage

```
Bt2mEnrichGO(
  bt2m.marker.chain,
  organism = "hs",
  pvalueCutoff = 0.05,
  min_count = 3
)
```

Arguments

bt2m.marker.chain	terbi.marker.chain dataframe contains all the markers
organism	"hs" for Homo sapiens, or "mm" for Mus musculus
pvalueCutoff	cut off P-value for GO annotations
min_count	Minimal count of genes for GO annotations

Value

GO annotation dataframe labeled with cluster

clusterProfilerORA	<i>GO KEGG ORA analysis by clusterProfiler</i>
--------------------	--

Description

GO KEGG ORA analysis by clusterProfiler

Usage

```
clusterProfilerORA(geneList = markerList, organism = "hs")
```

Arguments

geneList	a gene list
organism	"hs" for Homo sapiens, or "mm" for Mus musculus

Value

a list with GO and KEGG annotation result (clusterProfiler format)

DataframeToVector	<i>transform dataframe to vector</i>
-------------------	--------------------------------------

Description

transform dataframe to vector

Usage

```
DataframeToVector(df)
```

Arguments

df A dataframe from Seuratobj[["attr"]]

Value

A vector with names

DEG.cluster.list	<i>A general function to identify DEGs between case and control</i>
------------------	---

Description

A general function to identify DEGs between case and control

Usage

```
DEG.cluster.list(  
  seuratObj,  
  cluster.list,  
  ident.1 = "Vcl cKO",  
  ident.2 = "Control",  
  assay = "RNA"  
)
```

DrawBarplotGO	<i>Draw barplot for GO annotation</i>
---------------	---------------------------------------

Description

Draw barplot for GO annotation

Usage

```
DrawBarplotGO(barplot_df, cluster.chain)
```

Arguments

barplot_df	A GO annotation dataframe from clusterProfiler
cluster.chain	cluster.chain from bt2m

Value

A ggplot barplot

DrawBt2mClusterTree	<i>Draw cluster tree/chain by clustree</i>
---------------------	--

Description

Draw cluster tree/chain by clustree

Usage

```
DrawBt2mClusterTree(
  seuratObj,
  bt2m.cellMeta,
  node_text_size = 7,
  node_size = 8
)
```

Arguments

seuratObj	A Seurat object
bt2m.cellMeta	bt2m.cellMeta from bt2m
node_text_size	see node_text_size in clustree function
node_size	see node_size in clustree function

Value

clustree object

DrawGOchain

Get GO chain corresponding to a cluster chain

Description

Get GO chain corresponding to a cluster chain

Usage

```
DrawGOchain(bt2m.GO.anno, cluster.chain, top_n = 3)
```

Arguments

bt2m.GO.anno	A GO annotation dataframe from clusterProfiler
cluster.chain	cluster.chain from bt2m
top_n	top n GO terms for barplot (sort by P-value)

Value

A ggplot barplot showed GO chain

DrawMarkerChainDotplot

Draw cluster tree/chain by clustree

Description

Draw cluster tree/chain by clustree

Usage

```
DrawMarkerChainDotplot(
  seuratObj,
  bt2m.marker.chain,
  rmDup = T,
  top_n = 3,
  rel_heights = c(0.5, 9, 1.5),
  rel_widths = c(1, 4)
)
```

Arguments

seuratObj	A Seurat object
bt2m.marker.chain	bt2m.marker.chain from bt2m
rmDup	remove duplicated genes or not
top_n	top n genes for dotplot (sort by P-value)
rel_heights	relative heights for plot_grid(), the first and last control the height of top and bottom margin, the middle one controls the height of tree
rel_widths	relative widths for plot_grid(), the first one controls the widths of tree, the second one controls the widths of dotplot

Value

A plot_grid integrated tree and dotplot

DrawMarkerChainHeatmap

Draw heatmap by ComplexHeatmap

Description

Draw heatmap by ComplexHeatmap

Usage

```
DrawMarkerChainHeatmap(
  seuratObj,
  bt2m.cellMeta,
  bt2m.marker.chain,
  compare_anno = "",
  known_markers = c()
)
```

Arguments

seuratObj	A Seurat object
bt2m.cellMeta	bt2m.cellMeta from bt2m
bt2m.marker.chain	bt2m.marker.chain from bt2m
compare_anno	select a annotation in seurat object for comparision (e.g. previous annotation)
known_markers	the genes to include in the heatmap (like known markers)

Value

ComplexHeatmap object

FindBifurcationResolution

Find the best resolution for bifurcation in graph-based clustering

Description

Find the best resolution for bifurcation in graph-based clustering

Usage

```
FindBifurcationResolution(seuratObj, resolution.sets = 50)
```

Arguments

seuratObj A Seurat object
resolution.sets The number of resolution for searching

Value

The best resolution which can bifurcate all cells

gene.anno.GRCh38.v3 *Gene annotation for human genome*

Description

All human genes and annotation

Usage

gene.anno.GRCh38.v3

Format

A vector

Source

<https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest>

gene.anno.mm10.v3 *Gene annotation for mouse genome*

Description

All mouse genes and annotation

Usage

gene.anno.mm10.v3

Format

A vector

Source

<https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest>

GetClusterChain	<i>Find the cluster chain of any given cluster</i>
-----------------	--

Description

Find the cluster chain of any given cluster

Usage

```
GetClusterChain(bt2m.cellMeta, bt2m.bifucation, target.cluster)
```

Arguments

bt2m.cellMeta	bt2m.cellMeta from bt2m
bt2m.bifucation	bt2m.bifucation from bt2m
target.cluster	target cluster

Value

A vector, names are cluster name, element shows the states of clusters

GetInitialBifurcation	<i>Find initial bifurcation event of any two clusters in bt2m.cellMeta</i>
-----------------------	--

Description

Find initial bifurcation event of any two clusters in bt2m.cellMeta

Usage

```
GetInitialBifurcation(
  seuratObj,
  bt2m.cellMeta,
  cluster_1,
  cluster_2,
  balance_cells = T
)
```

Arguments

seuratObj	A Seurat object
bt2m.cellMeta	bt2m.cellMeta from bt2m
cluster_1	cluster 1
cluster_2	cluster 2
balance_cells	output balanced cell number (T or F)

Value

A list. subset_seuratObj is the subset cells in cluster 1 and cluster 2 diff_marker_1 is the marker of cluster 1, diff_marker_2 is the marker of cluster 2

GetTestData	<i>Download raw single-cell matrix for testing</i>
-------------	--

Description

Download raw single-cell matrix for testing

Usage

```
GetTestData()
```

Value

downloaded single-cell dataset

HGNC.gene.pairs	<i>Mouse transcription factors</i>
-----------------	------------------------------------

Description

A vector of genes used in marker annotation

Usage

```
HGNC.gene.pairs
```

Format

A vector

Source

<http://bioinfo.life.hust.edu.cn/AnimalTFDB/#!/download>

IdentifyBinaryMarkers	<i>Identify binary markers</i>
-----------------------	--------------------------------

Description

Identify binary markers

Usage

```
IdentifyBinaryMarkers(
  seuratObj,
  p_value = 0.01,
  min_avg_logFC = 0.1,
  min_correlation = 0.3
)
```

Arguments

seuratObj	A Seurat object
p_value	The number of resolution for searching
min_avg_logFC	Minimal threshold for logFC to pick markers
min_correlation	Minimal threshold for correlation to pick markers

Value

A list. b1 is the positive marker of cluster 0, b2 is the positive marker of cluster 1.

JASPAR.motif.rmDup	<i>Human ligand and receptor</i>
--------------------	----------------------------------

Description

A list of genes used in marker annotation

Usage

```
JASPAR.motif.rmDup
```

Format

A list of two vectors

human.ligand ligand genes in human

human.receptor receptor genes in human

Source

<http://www.cellchat.org/>

lr.pair.all.uniq.hs	<i>Mouse ligand and receptor</i>
---------------------	----------------------------------

Description

A list of genes used in marker annotation

Usage

```
lr.pair.all.uniq.hs
```

Format

A list of two vectors

mouse.ligand ligand genes in mouse

mouse.receptor receptor genes in mouse

Source

<http://www.cellchat.org/>

lr.pair.all.uniq.mm	<i>Mouse ligand and receptor</i>
---------------------	----------------------------------

Description

A list of genes used in marker annotation

Usage

lr.pair.all.uniq.mm

Format

A list of two vectors

mouse.ligand ligand genes in mouse

mouse.receptor receptor genes in mouse

Source

<http://www.cellchat.org/>

MGI.gene.pairs	<i>Human transcription factors</i>
----------------	------------------------------------

Description

A vector of genes used in marker annotation

Usage

MGI.gene.pairs

Format

A vector

Source

<http://bioinfo.life.hust.edu.cn/HumanTFDB/#!/download>

OrderCluster	<i>order the clusters according to similarity inside the bt2m result</i>
--------------	--

Description

order the clusters according to similarity inside the bt2m result

Usage

```
OrderCluster(seuratObj, bt2m.result)
```

Arguments

seuratObj	A Seurat object
bt2m.result	A result file from RunBt2m() function

Value

A re-ordered list. cellMeta contains the final bifurcation for each level marker_chain contains all the significant markers for each cluster bifucation contains the bifurcation details (parent, child1, child2)

PrepareExpressionMatrix

Prepare expression matrix for heatmap visualization, estimate the overall expression of marker modules

Description

Prepare expression matrix for heatmap visualization, estimate the overall expression of marker modules

Usage

```
PrepareExpressionMatrix(
  seuratObj,
  bt2m.marker.chain,
  assay = "RNA",
  slot = "scale.data",
  known_markers = c()
)
```

Arguments

seuratObj	A Seurat object
bt2m.marker.chain	bt2m.marker.chain from bt2m
assay	Assay used for prediction
slot	slot used for prediction
known_markers	the genes to include in the heatmap (like known markers)

Value

A expression matrix contain the average expression of marker modules and known markers

RemoveDuplicatedGO	<i>remove duplicated GO terms based on overlaps of genes</i>
--------------------	--

Description

remove duplicated GO terms based on overlaps of genes

Usage

```
RemoveDuplicatedGO(tmp.GO.df, max.overlap = 0.6)
```

Arguments

tmp.GO.df	GO annotation dataframe from clusterProfiler package
max.overlap	Maximum overlapped percentage of genes

Value

GO annotation dataframe without duplications

RemoveDuplicatedMarker	<i>Remove duplicated markers (just for plotting)</i>
------------------------	--

Description

Remove duplicated markers (just for plotting)

Usage

```
RemoveDuplicatedMarker(bt2m.marker.chain, method = "correlation")
```

Arguments

bt2m.marker.chain	bt2m.marker.chain from bt2m
method	select a method to sort markers

Value

bt2m.marker.chain without duplicated genes

RenameBt2m	<i>rename the clusters inside the bt2m result</i>
------------	---

Description

rename the clusters inside the bt2m result

Usage

```
RenameBt2m(bt2m.result)
```

Arguments

bt2m.result A result file from RunBt2m() function

Value

A renamed list. cellMeta contains the final bifurcation for each level marker_chain contains all the significant markers for each cluster bifurcation contains the bifurcation details (parent, child1, child2)

RunBt2m	<i>The main function to perform iteratively bifurcation clustering</i>
---------	--

Description

The main function to perform iteratively bifurcation clustering

Usage

```
RunBt2m(
  seuratObj,
  method = "graph",
  min.marker.num = 100,
  max.level.num = 20,
  min.cell.count = 50,
  resolution.sets = 30,
  verbose = T
)
```

Arguments

seuratObj	A Seurat object
method	The method to perform bifurcation clustering "graph (default), hclust or kmeans"
min.marker.num	Minimal number of markers to confirm a bifurcation
max.level.num	Maximum number of level for bifurcation
min.cell.count	Minimal number of cells to perform bifurcation (must bigger than PC number: 50)
resolution.sets	The number of resolution for searching
verbose	Print detail processing messages

Value

A list. cellMeta contains the preliminary bifurcation for each level marker_chain contains all the significant markers for each cluster bifurcation contains the bifurcation details (parent, child1, child2)

subsetSeuratObj	<i>Subset seurat object by cell names</i>
-----------------	---

Description

Subset seurat object by cell names

Usage

```
subsetSeuratObj(seuratObj, tmp.cells)
```

Arguments

seuratObj	A Seurat object
tmp.cells	cells for subsetting

Value

A subset of seurat object

subset_cells	<i>a subset function for seurat obj, original cannot work sometimes</i>
--------------	---

Description

a subset function for seurat obj, original cannot work sometimes

Usage

```
subset_cells(seuratObj, condition)
```

WriteBt2mIntoSeurat	<i>Write bt2m result to Seurat object</i>
---------------------	---

Description

Write bt2m result to Seurat object

Usage

```
WriteBt2mIntoSeurat(seuratObj, bt2m.result)
```

Arguments

seuratObj	A Seurat object
bt2m.result	bt2m.result

Value

A Seurat object contains bt2m.result. bt2m will be stored in assay data region of Seurat object ("seuratObj@assays\$bt2m")

Index

* datasets

- gene.anno.GRCh38.v3, [11](#)
- gene.anno.mm10.v3, [11](#)
- HGNC.gene.pairs, [13](#)
- JASPAR.motif.rmDup, [14](#)
- lr.pair.all.uniq.hs, [14](#)
- lr.pair.all.uniq.mm, [15](#)
- MGI.gene.pairs, [15](#)

* data

- gene.anno.GRCh38.v3, [11](#)
- gene.anno.mm10.v3, [11](#)
- HGNC.gene.pairs, [13](#)
- JASPAR.motif.rmDup, [14](#)
- lr.pair.all.uniq.hs, [14](#)
- lr.pair.all.uniq.mm, [15](#)
- MGI.gene.pairs, [15](#)

add.missing.genes, [2](#)

AddMarkerExpressionPct, [3](#)

Bt2mBifucation.graph, [3](#)

Bt2mBifucation.hclust, [4](#)

Bt2mBifucation.kmeans, [4](#)

Bt2mColors, [5](#)

Bt2mEnrichGO, [5](#)

clusterProfilerORA, [6](#)

DataframeToVector, [6](#)

DEG.cluster.list, [7](#)

DrawBarplotGO, [7](#)

DrawBt2mClusterTree, [8](#)

DrawGOchain, [8](#)

DrawMarkerChainDotplot, [9](#)

DrawMarkerChainHeatmap, [9](#)

FindBifurcationResolution, [10](#)

gene.anno.GRCh38.v3, [11](#)

gene.anno.mm10.v3, [11](#)

GetClusterChain, [12](#)

GetInitialBifurcation, [12](#)

GetTestData, [13](#)

HGNC.gene.pairs, [13](#)

IdentifyBinaryMarkers, [13](#)

JASPAR.motif.rmDup, [14](#)

lr.pair.all.uniq.hs, [14](#)

lr.pair.all.uniq.mm, [15](#)

MGI.gene.pairs, [15](#)

OrderCluster, [16](#)

PrepareExpressionMatrix, [16](#)

RemoveDuplicatedGO, [17](#)

RemoveDuplicatedMarker, [17](#)

RenameBt2m, [18](#)

RunBt2m, [18](#)

subset_cells, [19](#)

subsetSeuratObj, [19](#)

WriteBt2mIntoSeurat, [20](#)