Package 'scCATCH'

January 7, 2022

Type Package

Title Single Cell Cluster-Based Annotation Toolkit for Cellular Heterogeneity

Version 3.0

Depends R (>= 4.0.0)

Author Xin Shao

Maintainer Xin Shao<xin_shao@zju.edu.cn>

Description An automatic cluster-based annotation pipeline based on evidence-based score by matching the marker genes with known cell markers in tissue-specific cell taxonomy reference database for single-cell RNA-seq data. See Shao X, et al (2020) <doi:10.1016/j.isci.2020.100882> for more details.

URL https://github.com/ZJUFanLab/scCATCH

License GPL (>= 3)

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

Suggests rmarkdown, knitr, testthat

VignetteBuilder knitr

Imports Matrix, methods, progress, stats, crayon, reshape2

NeedsCompilation no

Repository CRAN

Date/Publication 2022-01-07 13:32:42 UTC

R topics documented:

`cellmatch` *cellmatch*

Description

Marker genes of `'Human'` and `'Mouse'`.

Usage

```
cellmatch
```

Format

An object of class `data.frame` with 49625 rows and 11 columns.

Source

https://github.com/ZJUFanLab/scCATCH/tree/master/data

`createscCATCH` *scCATCH object*

Description

create scCATCH object using single-cell count data and cluster information.

Usage

```
createscCATCH(data, cluster)
```

Arguments

`data`	A matrix or dgCMatrix containing normalized single-cell RNA-seq data, each column representing a cell, each row representing a gene. See `demo_data`.
`cluster`	A character containing the cluster information for each cell. The length of it must be equal to the ncol of the data.

Value

scCATCH object

`demo_data` *Demo data of single-cell RNA-seq data*

Description

Demo data of single-cell RNA-seq data

Usage

```
demo_data()
```

Details

`data` used in `createscCATCH` must be a `matrix` object, each column representing a cell, each row representing a gene.

Value

A demo data matrix.

Examples

```
data_demo <- demo_data()
```

`demo_geneinfo` *Demo data of geneinfo*

Description

Demo data of geneinfo

Usage

```
demo_geneinfo()
```

Details

`geneinfo` used in `rev_gene` must be a `data.frame` object with three columns, namely `'symbol'`, `'synonyms'`, `'species'`.

Value

A demo geneinfo data.frame.

Examples

```
geneinfo_demo <- demo_geneinfo()
```

`demo_marker` *Demo data of markers*

Description

Demo data of markers

Usage

```
demo_marker()
```

Details

`markers` used in `findmarkergene` must be a `data.frame` object with eleven columns.

Value

A demo marker data.frame.

Examples

```
markers_demo <- demo_marker()
```

`findcelltype` *Evidence-based score and annotation for each cluster*

Description

Evidence-based score and annotation for each cluster.

Usage

```
findcelltype(object, verbose = TRUE)
```

Arguments

`object`	scCATCH object generated from `findmarkergene`.
`verbose`	Show progress messages.

Value

scCATCH object containing the results of predicted cell types for each cluster.

findcelltype,scCATCH-method *Evidence-based score and annotation for each cluster*

Description

Evidence-based score and annotation for each cluster.

Usage

```
## S4 method for signature 'scCATCH'
findcelltype(object, verbose = TRUE)
```

Arguments

object	scCATCH object generated from findmarkergene.
verbose	Show progress messages.

Value

scCATCH object containing the results of predicted cell types for each cluster.

findmarkergene *Find potential marker genes for each cluster*

Description

Identify potential marker genes for each cluster.

Usage

```
findmarkergene(
  object,
  species = NULL,
  cluster = "All",
  if_use_custom_marker = FALSE,
  marker = NULL,
  cancer = "Normal",
  tissue = NULL,
  use_method = "1",
  comp_cluster = NULL,
  cell_min_pct = 0.25,
  logfc = 0.25,
  pvalue = 0.05,
  verbose = TRUE
)
```

Arguments

object	scCATCH object generated from createscCATCH.
species	The specie of cells. The species must be defined. 'Human' or 'Mouse'. When if_use_custom_marker is set TRUE, no need to define the species.
cluster	Select which clusters for potential marker genes identification. e.g. '1', '2', etc. The default is 'All' to find potential makrer genes for each cluster.
if_use_custom_marker	Whether to use custom markers data.frame.
marker	A data.frame containing marker genes. See demo_marker. Default is to use the system cellmatch data.frame.
cancer	If the sample is from cancer tissue, then the cancer type may be defined. When if_use_custom_marker is set TRUE, no need to define the species.
tissue	Tissue origin of cells must be defined. Select one or more related tissue types. When if_use_custom_marker is set TRUE, no need to define the species.
use_method	'1' is to compare with other every cluster. '2' is to compare with other clusters together.
comp_cluster	Number of clusters to compare. Default is to compare all other cluster for each cluster. Set it between 1 and length of unique clusters. More marker genes will be obtained for smaller comp_cluster.
cell_min_pct	Include the gene detected in at least this many cells in each cluster.
logfc	Include the gene with at least this fold change of average gene expression compared to every other clusters.
pvalue	Include the significantly highly expressed gene with this cutoff of p value from wilcox test compared to every other clusters.
verbose	Show progress messages.

Details

Details of available tissues see https://github.com/ZJUFanLab/scCATCH/wiki

Value

scCATCH object

---

findmarkergene,scCATCH-method

Find potential marker genes for each cluster

Description

Identify potential marker genes for each cluster.

Usage

```
## S4 method for signature 'scCATCH'
findmarkergene(
  object,
  species = NULL,
  cluster = "All",
  if_use_custom_marker = FALSE,
  marker = NULL,
  cancer = "Normal",
  tissue = NULL,
  use_method = "1",
  comp_cluster = NULL,
  cell_min_pct = 0.25,
  logfc = 0.25,
  pvalue = 0.05,
  verbose = TRUE
)
```

Arguments

object	scCATCH object generated from createscCATCH.
species	The specie of cells. The species must be defined. 'Human' or 'Mouse'. When if_use_custom_marker is set TRUE, no need to define the species.
cluster	Select which clusters for potential marker genes identification. e.g. '1', '2', etc. The default is 'All' to find potential makrer genes for each cluster.
if_use_custom_marker	Whether to use custom markers data.frame.
marker	A data.frame containing marker genes. See demo_marker. Default is to use the system cellmatch data.frame.
cancer	If the sample is from cancer tissue, then the cancer type may be defined. When if_use_custom_marker is set TRUE, no need to define the cancer.
tissue	Tissue origin of cells must be defined. Select one or more related tissue types. When if_use_custom_marker is set TRUE, no need to define the tissue.
use_method	'1' is to compare with other every cluster. '2' is to compare with other clusters together.
comp_cluster	Number of clusters to compare. Default is to compare all other cluster for each cluster. Set it between 1 and length of unique clusters. More marker genes will be obtained for smaller comp_cluster.
cell_min_pct	Include the gene detected in at least this many cells in each cluster.
logfc	Include the gene with at least this fold change of average gene expression compared to every other clusters.
pvalue	Include the significantly highly expressed gene with this cutoff of p value from wilcox test compared to every other clusters.
verbose	Show progress messages.

Details

Details of available tissues see https://github.com/ZJUFanLab/scCATCH/wiki

Value

scCATCH object

---

geneinfo *geneinfo*

Description

Gene symbols of 'Human' and 'Mouse' updated on Jan. 2, 2022 for revising genes.

Usage

```
geneinfo
```

Format

An object of class data.frame with 227791 rows and 3 columns.

Source

https://www.ncbi.nlm.nih.gov/gene

---

rev_gene *Pre-processing step: revising gene symbols*

Description

Revise genes according to NCBI Gene symbols updated in Jan. 2, 2022 for count matrix, user-custom cell marker data.frame.

Usage

```
rev_gene(data = NULL, data_type = NULL, species = NULL, geneinfo = NULL)
```

Arguments

data	A matrix or dgCMatrix containing count or normalized data, each column representing a spot or a cell, each row representing a gene; Or a data.frame containing cell markers, use demo_marker.
data_type	A character to define the type of data, select 'data' for the data matrix, 'marker' for the data.frame containing cell markers.
species	Species of the data.'Human' or 'Mouse'.
geneinfo	A data.frame of the system data containing gene symbols of 'Human' and 'Mouse' updated on Jan. 1, 2022.

Value

A new matrix or data.frame.

---

scCATCH *Definition of 'scCATCH' class*

Description

An S4 class containing the data, meta, and results of inferred cell types.

Slots

`data` A list containing normalized data. See `demo_data`.

`meta` A data frame containing the meta data.

`para` A list containing the parameters.

`markergene` A data frame containing the identified markers for each cluster.

`celltype` A data frame containing the cell types for each cluster.

`marker` A data frame containing the known markers. See `demo_marker`.

Index