

Package ‘clustermole’

January 27, 2020

Type Package

Title Unbiased Single-Cell Transcriptomic Data Cell Type Identification

Version 1.0.1

Description Assignment of cell type labels to single-cell RNA sequencing (scRNA-seq) clusters is often a time-consuming process that involves manual inspection of the cluster marker genes complemented with a detailed literature search. This is especially challenging when unexpected or poorly described populations are present. The clustermole R package provides methods to query thousands of human and mouse cell identity markers sourced from a variety of databases.

License MIT + file LICENSE

URL <https://github.com/igordot/clustermole>

BugReports <https://github.com/igordot/clustermole/issues>

Depends R (>= 3.4)

Imports dplyr, GSVA (>= 1.26.0), magrittr, methods, rlang (>= 0.1.2), tibble, tidyr, utils

Suggests covr, knitr, prettydoc, rmarkdown, roxygen2, testthat (>= 2.1.0)

biocViews

Encoding UTF-8

LazyData true

RoxygenNote 7.0.2

VignetteBuilder knitr

NeedsCompilation no

Author Igor Dolgalev [aut, cre]

Maintainer Igor Dolgalev <igor.dolgalev@nyumc.org>

Repository CRAN

Date/Publication 2020-01-27 10:00:02 UTC

R topics documented:

clustermole_enrichment	2
clustermole_markers	2
clustermole_overlaps	3
read_gmt	4

Index	5
--------------	----------

clustermole_enrichment

Perform cell type enrichment for a given gene expression matrix

Description

Perform cell type enrichment for a given gene expression matrix

Usage

```
clustermole_enrichment(expr_mat, species)
```

Arguments

expr_mat	Expression matrix (logCPMs, logFPKMs, or logTPMs) with genes as rows.
species	Species: "hs" for human or "mm" for mouse.

Value

A data frame of enrichment results.

Examples

```
# my_enrichment <- clustermole_enrichment(expr_mat = my_expr_mat, species = "hs")
```

clustermole_markers

Retrieve the available cell type markers

Description

Retrieve the available cell type markers

Usage

```
clustermole_markers(species = "hs")
```

Arguments

species Species for the appropriate gene symbol format: "hs" for human or "mm" for mouse.

Value

A data frame of markers with one gene per row.

Examples

```
markers <- clustermole_markers()
head(markers)
```

clustermole_overlaps *Perform cell type overrepresentation analysis for a set of genes*

Description

Perform cell type overrepresentation analysis for a set of genes

Usage

```
clustermole_overlaps(genes, species)
```

Arguments

genes A vector of genes.
species Species: "hs" for human or "mm" for mouse.

Value

A data frame of enrichment results with hypergeometric test p-values.

Examples

```
my_genes <- c("CD2", "CD3D", "CD3E", "CD3G", "TRAC", "TRBC2", "LTB")
my_overlaps <- clustermole_overlaps(genes = my_genes, species = "hs")
head(my_overlaps)
```

read_gmt	<i>Read a GMT file into a data frame</i>
----------	--

Description

Read a GMT file into a data frame

Usage

```
read_gmt(file, geneset_label = "celltype", gene_label = "gene")
```

Arguments

file	A connection object or a character string (can be a URL).
geneset_label	Column name for gene sets (first column of the GMT file) in the output data frame.
gene_label	Column name for genes (variable columns of the GMT file) in the output data frame.

Value

A data frame with gene sets as the first column and genes as the second column (one gene per row).

Examples

```
gmt <- "http://software.broadinstitute.org/gsea/msigdb/supplemental/scsig.all.v1.0.symbols.gmt"
gmt_tbl <- read_gmt(gmt)
head(gmt_tbl)
```

Index

`clustermole_enrichment`, [2](#)

`clustermole_markers`, [2](#)

`clustermole_overlaps`, [3](#)

`read_gmt`, [4](#)