Package 'clustermole'

January 20, 2020

Type Package

Title Unbiased Cell Type Identification of Single-Cell Transcriptomic Data

Version 1.0.0

Description A typical computational pipeline to process single-cell RNA sequencing (scRNA-seq) data involves clustering of cells. Assignment of cell type labels to those 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 if you are not familiar with all the captured subpopulations or have unexpected contaminants. 'clustermole' provides a comprehensive meta collection of cell identity markers for thousands of human and mouse cell types sourced from a variety of databases as well as methods to query them.

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, roxygen2, testthat (>= 2.1.0), knitr, rmarkdown

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 2020-01-20 10:00:02 UTC

R topics documented:

`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.

`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 *Read a GMT file into a data frame*

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