Package 'clustermole'

January 26, 2021

Type Package

Title Unbiased Single-Cell Transcriptomic Data Cell Type Identification

Version 1.1.0

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://igordot.github.io/clustermole/, https://github.com/igordot/clustermole

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

Depends R (>= 3.6)

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

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

biocViews

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

VignetteBuilder knitr

NeedsCompilation no

Author Igor Dolgalev [aut, cre]

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

Repository CRAN

Date/Publication 2021-01-26 06:40:02 UTC

R topics documented:

`clustermole_enrichment`

Cell types based on the expression of all genes

Description

Perform enrichment of cell type signatures based on the full gene expression matrix.

Usage

```
clustermole_enrichment(expr_mat, species, method = "gsva")
```

Arguments

`expr_mat`	Expression matrix (logCPMs, logFPKMs, or logTPMs) with genes as rows and clusters/populations/samples as columns.
`species`	Species: `hs` for human or `mm` for mouse.
`method`	Enrichment method: `ssgsea`, `gsva`, `singscore`, or `all`. The method to use for the estimation of gene set enrichment scores. The options are ssGSEA (Barbie et al, 2009), GSVA (Hänzelmann et al, 2013), singscore (Foroutan et al, 2018), or a combination of all three methods.

Value

A data frame of enrichment results.

References

Barbie, D., Tamayo, P., Boehm, J. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462, 108–112 (2009). doi: 10.1038/nature08460

Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* 14, 7 (2013). doi: 10.1186/14712105147

Foroutan, M., Bhuva, D.D., Lyu, R. et al. Single sample scoring of molecular phenotypes. *BMC Bioinformatics* 19, 404 (2018). doi: 10.1186/s1285901824354

Examples

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

clustermole_markers *Available cell type markers*

Description

Retrieve the full list of cell type markers in the `clustermole` database.

Usage

```
clustermole_markers(species = c("hs", "mm"))
```

Arguments

`species`	Species: `hs` for human or `mm` for mouse.

Value

A data frame of cell type markers (one gene per row).

Examples

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

clustermole_overlaps *Cell types based on overlap of marker genes*

Description

Perform overrepresentation analysis for a set of genes compared to all cell type signatures.

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