

Package ‘scTGIF’

April 15, 2020

Type Package

Title Cell type annotation for unannotated single-cell RNA-Seq data

Version 1.0.0

Depends R (>= 3.6.0)

Imports GSEABase, Biobase, SingleCellExperiment, BiocStyle, plotly, tagcloud, rmarkdown, Rcpp, grDevices, graphics, utils, knitr, S4Vectors, SummarizedExperiment, RColorBrewer, nnTensor, methods, scales, msigdb

Suggests testthat

Description scTGIF connects the cells and the related gene functions without cell type label.

License Artistic-2.0

biocViews DimensionReduction, QualityControl, SingleCell, Software, GeneExpression

VignetteBuilder knitr

git_url <https://git.bioconductor.org/packages/scTGIF>

git_branch RELEASE_3_10

git_last_commit 5e82bd9

git_last_commit_date 2019-10-29

Date/Publication 2020-04-14

Author Koki Tsuyuzaki [aut, cre]

Maintainer Koki Tsuyuzaki <k.t.the-answer@hotmail.co.jp>

R topics documented:

| | |
|--------------------------------------|----------|
| scTGIF-package | 2 |
| calcTGIF | 2 |
| DistalLungEpithelium | 3 |
| label.DistalLungEpithelium | 3 |
| pca.DistalLungEpithelium | 4 |
| reportTGIF | 4 |
| settingTGIF | 6 |
| Index | 8 |

 scTGIF-package

Cell type annotation for unannotated single-cell RNA-Seq data

Description

scTGIF connects the cells and the related gene functions without cell type label.

Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.

[calcTGIF](#) function calculates what kind of cellular patterns and functional patterns are contained in single-cell RNA-seq data and [reportTGIF](#) function generates report of analytic result. The algorithm is based on joint NMF, which is implemented in nnTensor package.

Author(s)

NA

Maintainer: NA

References

Dominic Grun, Anna Lyubimova, Lennart Kester, Kay Wiebrands, Onur Basak, Nobuo Sasaki, Hans Clevers, Alexander van Oudenaarden (2015) Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature*, **525**: 251-255

 calcTGIF

Function for connecting cellular patterns and functional patterns using jNMF

Description

[calcTGIF](#) function calculates what kind of cellular patterns and functional patterns are contained in single-cell RNA-seq data and [reportTGIF](#) function generates report of analytic result.

Usage

```
calcTGIF(sce, rank)
```

Arguments

sce A object generated by instantizaton of SingleCellExperiment-class.

rank Rank parameter of joint NMF algorithm.

Value

The result is saved to metadata slot of SingleCellExperiment object.

Author(s)

NA

Examples

```
showMethods("calcTGIF")
```

DistalLungEpithelium *Gene expression matrix of DistalLungEpithelium dataset containing five cluster.*

Description

A data frame with 3397 rows (genes) with following 80 columns (cells).

The data is downloaded as supplementary information of the distal lung epithelium paper (<https://www.nature.com/article>)

Low-expressed genes are filtered.

All Gene ID is converted to Human Entrez Gene ID for applying the data to scTGIF.

Usage

```
data("DistalLungEpithelium")
```

Source

<http://www.nature.com/nbt/journal/v33/n2/full/nbt.3102.html>

References

Treutlein, B. et al. (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371-375

Examples

```
data("DistalLungEpithelium")
```

label.DistalLungEpithelium
Cellular label of DistalLungEpithelium dataset containing five cluster.

Description

A vector containing 80 elements (cells).

Usage

```
data("label.DistalLungEpithelium")
```

References

Treutlein, B. et al. (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371-375

Examples

```
data("label.DistalLungEpithelium")
```

```
pca.DistalLungEpithelium
```

The result of PCA of the DistalLungEpithelium dataset.

Description

A matrix having 80 (cells) * 2 (PCs) elements.

Usage

```
data("pca.DistalLungEpithelium")
```

References

Treutlein, B. et al. (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371-375

Examples

```
data("pca.DistalLungEpithelium")
```

```
reportTGIF
```

Function for reporting the result of [calcTGIF](#) function

Description

[calcTGIF](#) function calculates what kind of cellular patterns and functional patterns are contained in single-cell RNA-seq data and [reportTGIF](#) function generates report of analytic result.

Usage

```
reportTGIF(sce, out.dir=tempdir(), html.open=TRUE,  
  title="The result of scTGIF",  
  author="The person who runs this script",  
  assayNames="counts")
```

Arguments

| | |
|------------|--|
| sce | A object generated by instantiztion of SingleCellExperiment-class. |
| out.dir | Output directory user want to save the report (Default: tempdir()). |
| html.open | Whether html is opened when <code>reportTGIF</code> is finished (Default: TRUE) |
| title | Title of report (Default: "The result of scTGIF") |
| author | The name of user name (Default: "The person who runs this script") |
| assayNames | The unit of gene expression for using scTGIF (e.g. normcounts, cpm...etc) (Default: "counts"). |

Value

Some file is generated to output directory user specified.

Author(s)

NA

Examples

```

if(interactive()){
  # Package loading
  library("SingleCellExperiment")
  library("GSEABase")
  library("msigdb")

  # Test data
  data("DistalLungEpithelium")
  data("pca.DistalLungEpithelium")
  data("label.DistalLungEpithelium")

  # Test data
  par(ask=FALSE)
  plot(pca.DistalLungEpithelium, col=label.DistalLungEpithelium, pch=16,
       main="Distal lung epithelium dataset", xlab="PCA1",
       ylab="PCA2", bty="n")
  text(0.1, 0.05, "AT1", col="#FF7F00", cex=2)
  text(0.07, -0.15, "AT2", col="#E41A1C", cex=2)
  text(0.13, -0.04, "BP", col="#A65628", cex=2)
  text(0.125, -0.15, "Clara", col="#377EB8", cex=2)
  text(0.09, -0.2, "Cilliated", col="#4DAF4A", cex=2)

  # Load the gmt file from MSigDB
  # Only "Entrez Gene ID" can be used in scTGIF
  # e.g. gmt <- GSEABase::getGmt(
  #   "/PATH/YOU/SAVED/THE/GMTFILES/h.all.v6.0.entrez.gmt")
  # Here we use msigdb to retrieve mouse gene sets

  # Mouse gene set (NCBI Gene ID)
  m_df <- msigdb(species = "Mus musculus", category = "H")[,
    c("gs_name", "entrez_gene")]

  # Convert to GeneSetCollection
  hallmark = unique(m_df$gs_name)
  gsc <- lapply(hallmark, function(h){

```

```

        target = which(m_df$gs_name == h)
        geneIds = unique(as.character(m_df$entrez_gene[target]))
        GeneSet(setName=h, geneIds)
    })
gmt <- GeneSetCollection(gsc)

# SingleCellExperiment-class
sce <- SingleCellExperiment(
  assays = list(counts = DistalLungEpithelium)
reducedDims(sce) <- SimpleList(PCA=pca.DistalLungEpithelium)

# User's Original Normalization Function
CPMED <- function(input){
  libsize <- colSums(input)
  median(libsize) * t(t(input) / libsize)
}
# Normalization
normcounts(sce) <- log10(CPMED(counts(sce)) + 1)

# Registration of required information into metadata(sce)
sce2 <- settingTGIF(sce, gmt, reducedDimNames="PCA",
  assayNames="normcounts")

# Functional Annotation based on jNMF
sce2 <- calcTGIF(sce2, rank=7)

# HTML Reprt
reportTGIF(sce2,
  html.open=TRUE,
  title="scTGIF Report for DistalLungEpithelium dataset",
  author="Koki Tsuyuzaki")
}

```

settingTGIF

Parameter setting for scTGIF

Description

All parameters is saved to metadata slot of SingleCellExperiment object.

Usage

```
settingTGIF(sce, gmt, reducedDimNames, assayNames="counts", grid.size=50)
```

Arguments

| | |
|-----------------|--|
| sce | A object generated by instantiation of SingleCellExperiment-class. |
| gmt | Object generated from GSEABase::getGmt function. GMT file can be downloaded from MSigDB web (site http://software.broadinstitute.org/gsea/login.jsp#msigdb). Please confirm that the gmt file contains Human Entrez Gene ID, not gene symbol. Also confirm that the DataMatrix has Human Entrez Gene ID. |
| reducedDimNames | The names of reducedDim(sce) that user want use in scTGIF. |

| | |
|-------------------------|---|
| <code>assayNames</code> | The unit of gene expression for using scTGIF (e.g. normcounts, cpm...etc) (Default: "counts"). |
| <code>grid.size</code> | The grid size for segmentation of the two dimensional plot of reducedDim(sce) (Default: 50, which means 50*50 grids). |

Value

The result is saved to metadata slot of SingleCellExperiment object.

Author(s)

NA

Examples

```
showMethods("settingTGIF")
```

Index

*Topic **datasets**

DistalLungEpithelium, 3
label.DistalLungEpithelium, 3
pca.DistalLungEpithelium, 4

*Topic **methods**

calcTGIF, 2
reportTGIF, 4
settingTGIF, 6

*Topic **package**

scTGIF-package, 2

calcTGIF, 2, 2, 4
calcTGIF, SingleCellExperiment-method
(calcTGIF), 2

DistalLungEpithelium, 3

label.DistalLungEpithelium, 3

pca.DistalLungEpithelium, 4

reportTGIF, 2, 4, 4, 5
reportTGIF, SingleCellExperiment-method
(reportTGIF), 4

scTGIF (scTGIF-package), 2
scTGIF-package, 2
settingTGIF, 6
settingTGIF, SingleCellExperiment-method
(settingTGIF), 6