

Package ‘rPanglaoDB’

May 13, 2021

Type Package

Title Download and Merge Single-Cell RNA-Seq Data from the PanglaoDB Database

Version 0.2.1

Description Download and merge labeled single-cell RNA-seq data from the PanglaoDB <<https://panglaoDB.se/>> into a Seurat object.

URL <https://github.com/dosorio/rPanglaoDB/>

License GPL-3

Encoding UTF-8

LazyData false

RoxygenNote 7.1.1

Imports methods, XML, xml2, Seurat, pbapply, Matrix

NeedsCompilation no

Author Daniel Osorio [aut, cre] (<<https://orcid.org/0000-0003-4424-8422>>),
Marieke Kuijjer [aut] (<<https://orcid.org/0000-0001-6280-3130>>),
James J. Cai [aut] (<<https://orcid.org/0000-0002-8081-6725>>)

Maintainer Daniel Osorio <daniecos@uio.no>

Repository CRAN

Date/Publication 2021-05-12 22:00:02 UTC

R topics documented:

getMarkers	2
getSampleComposition	3
getSampleList	4
getSamples	5

Index	7
--------------	----------

getMarkers	<i>Get the list of samples from the panglaoDB database with a pattern of expression for a set of molecular markers.</i>
------------	---

Description

Get the list of samples from the panglaoDB database with a pattern of expression for a set of molecular markers.

Usage

```
getMarkers(include, exclude = NULL)
```

Arguments

include	A set of molecular markers to query the database. This set of genes needs to be expressed in the sample.
exclude	A set of molecular markers to query the database. This set of genes needs to be absent in the sample.

Value

The rows in the data frame are the samples matching the requested pattern. The returned data frame contain 7 columns as follows:

- SRA: The SRA identifier of the biological sample in the SRA database
- SRS: The SRS identifier of the biological sample in the SRA database
- Specie: The specie from which the biological samples originated from
- Tissue: The tissue from which the biological samples originated from
- Cluster: The cluster-id assigned by the panglaoDB database to the cells matching the requested pattern
- Cell-Type: The cell-type from which the counts originates from
- Markers: The recovered pattern for the marker genes requested

Examples

```
## Not run:
Fibrocytes <- getMarkers(include = c('ACTA2', 'CD34', 'FAP'))
Fibrocytes
## End(Not run)

#      SRA      SRS      Specie Tissue Cluster  Cell-Type      Markers
# SRA681285 SRS3121028 Mus musculus Dermis      4 Fibroblasts +ACTA2+CD34+FAP
```

getSampleComposition *Get the cell-type content for each sample from the panglaoDB database.*

Description

Get the cell-type content for each sample from the panglaoDB database.

Usage

```
getSampleComposition(  
  sra = "All",  
  srs = "All",  
  tissue = "All",  
  protocol = "All",  
  specie = "All",  
  verbose = TRUE  
)
```

Arguments

sra	Filter based on the SRA identifier of the biological sample in the SRA database
srs	Filter based on the SRS identifier of the biological sample in the SRA database
tissue	Filter based on the tissue from which the biological samples originates from
protocol	Filter based on the single-cell library preparation protocol used to generate the data
specie	Filter based on the specie from which the biological samples originates from
verbose	A boolean value TRUE or FALSE to activate the verbose mode

Value

This function returns the cell-type composition of the samples included in the PanglaoDB database in a data frame with 8 columns as follows:

- SRA: The SRA identifier of the biological sample in the SRA database
- SRS: The SRS identifier of the biological sample in the SRA database
- Tissue: The tissue from which the biological samples originated from
- Protocol: The single-cell library preparation protocol used to generate the data
- Species: The species from which the biological samples originated from
- Cluster: The cluster-id assigned by the panglaoDB database to the cells in the sample
- Cells: The number of cells included in the cluster
- Cell Type: The cell-type from which the counts originates from

Examples

```
# From PanglaoDB
# https://panglaodb.se/list_clusters_and_cell_types.html?sra=SRA689041&srs=SRS3166675

SRS3166675 <- getSampleComposition(srs = 'SRS3166675')
head(SRS3166675)
```

#	SRA	SRS	Tissue	Protocol	Species	Cluster	Cells	Cell Type
#	SRA689041	SRS3166675	Colon	10x chromium	Mus musculus	0	735	Fibroblasts
#	SRA689041	SRS3166675	Colon	10x chromium	Mus musculus	1	526	Smooth muscle cells
#	SRA689041	SRS3166675	Colon	10x chromium	Mus musculus	2	465	Unknown
#	SRA689041	SRS3166675	Colon	10x chromium	Mus musculus	3	157	Unknown
#	SRA689041	SRS3166675	Colon	10x chromium	Mus musculus	4	140	Goblet cells
#	SRA689041	SRS3166675	Colon	10x chromium	Mus musculus	5	100	Fibroblasts

```
getSampleList
```

Get the sample list from the panglaoDB database.

Description

Get the sample list from the panglaoDB database.

Usage

```
getSampleList()
```

Value

This function returns a data frame with 6 columns as follows:

- SRA: The SRA identifier of the biological sample in the SRA database
- SRS: The SRS identifier of the biological sample in the SRA database
- Tissue: The tissue from which the biological samples originated from
- Protocol: The single-cell library preparation protocol used to generate the data
- Species: The specie from which the biological samples originated from
- Cells: The number of cells included in the sample

Examples

```
# From the PanglaoDB database
# https://panglaodb.se/samples.html

sampleList <- getSampleList()
head(sampleList)
```

#	SRA	SRS	Tissue	Protocol	Species	Cells
#	SRA553822	SRS2119548	Cultured embryonic stem cells	10x chromium	Homo sapiens	6501

```

# SRA570744 SRS2253536 Lung mesenchyme 10x chromium Mus musculus 4611
# SRA598936 SRS2428405 Kidney cortex 10x chromium Homo sapiens 3759
# SRA644036 SRS2808714 Cervical and lumbar spinal cord 10x chromium Mus musculus 1025
# SRA670243 SRS3078084 Ventral midbrain 10x chromium Mus musculus 5603
# SRA689041 SRS3166675 Colon 10x chromium Mus musculus 2878

```

getSamples	<i>Download the expression matrix and annotations from the panglaoDB database.</i>
------------	--

Description

Download the expression matrix and annotations from the panglaoDB database.

Usage

```

getSamples(
  sra = "All",
  srs = "All",
  tissue = "All",
  protocol = "All",
  specie = "All",
  celltype = "All",
  include = NA,
  exclude = NA,
  merge = TRUE
)

```

Arguments

sra	Filter based on the SRA identifier of the biological sample in the SRA database
srs	Filter based on the SRS identifier of the biological sample in the SRA database
tissue	Filter based on the tissue from which the biological samples originates from
protocol	Filter based on the single-cell library preparation protocol used to generate the data
specie	Filter based on the specie from which the biological samples originates from
celltype	Filter based on the cell-type from which the counts originates from
include	A set of molecular markers to filter the dataset. This set of genes needs to be expressed in each cell.
exclude	A set of molecular markers to filter the dataset. This set of genes needs to be absent in each cell.
merge	A boolean value TRUE or FALSE defining if the samples should be returned as a list or as a unique Seurat object

Value

A Seurat object, as described in `?SeuratObject::`Seurat-class``

Examples

```
# From PanglaoDB SRS3805255
# https://panglaodb.se/view_data.php?sra=SR705190&srs=SRS4139632

## Not run:
SRS4139632 <- getSamples(srs = 'SRS4139632')
SRS4139632
## End(Not run)

# An object of class Seurat
# 19859 features across 102 samples within 1 assay
# Active assay: RNA (19859 features, 0 variable features)

# Metadata from the PanglaoDB database can be accessed as follows:
# head(SRS4139632[[[]]])
```

Index

getMarkers, 2
getSampleComposition, 3
getSampleList, 4
getSamples, 5