Package ‘scds’

April 9, 2024

Type Package
Title In-Silico Annotation of Doublets for Single Cell RNA Sequencing Data
Version 1.18.0
Description In single cell RNA sequencing (scRNA-seq) data combinations of cells are sometimes considered a single cell (doublets). The scds package provides methods to annotate doublets in scRNA-seq data computationally.
License MIT + file LICENSE
Encoding UTF-8
biocViews SingleCell, RNASeq, QualityControl, Preprocessing, Transcriptomics, GeneExpression, Sequencing, Software, Classification
RoxygenNote 6.1.1
Depends R (>= 3.6.0)
Imports Matrix, S4Vectors, SingleCellExperiment, SummarizedExperiment, xgboost, methods, stats, dplyr, pROC
Suggests BiocStyle, knitr, rsvd, Rtsne, scater, cowplot, rmarkdown
VignetteBuilder knitr
git_url https://git.bioconductor.org/packages/scds
git_branch RELEASE_3_18
git_last_commit dff0c08
git_last_commit_date 2023-10-24
Repository Bioconductor 3.18
Date/Publication 2024-04-08
Author Dennis Kostka [aut, cre], Bais Abha [aut]
Maintainer Dennis Kostka <kostka@pitt.edu>
R topics documented:

bcds ................................................................. 2
cxds ................................................................. 3
cxds_bcds_hybrid .................................................. 4
cxds_getTopPairs .................................................... 4
get_dblCalls_ALL ..................................................... 5
get_dblCalls_dist ..................................................... 6
get_dblCalls_ROC ..................................................... 6
sce_chcl ............................................................... 7

Index 8

bcds  

Find doublets/multiplets in UMI scRNA-seq data;

Description

Annotates doublets/multiplets using a binary classification approach to discriminate artificial doublets from original data.

Usage

bcds(sce, ntop = 500, srat = 1, verb = FALSE, retRes = FALSE,
nmax = "tune", varImp = FALSE, estNdbl = FALSE)

Arguments

sce single cell experiment (SingleCellExperiment) object to analyze; needs counts in assays slot.
ntop integer, indicating number of top variance genes to consider. Default: 500
srat numeric, indicating ratio between original number of "cells" and simulated doublets; Default: 1
verb progress messages. Default: FALSE
retRes logical, should the trained classifier be returned? Default: FALSE
nmax maximum number of training rounds; integer or "tune". Default: "tune"
varImp logical, should variable (i.e., gene) importance be returned? Default: FALSE
estNdbl logical, should the number of doublets be estimated from the data. Enables doublet calls. Default: FALSE. Use with caution.

Value

sce input sce object SingleCellExperiment with doublet scores added to colData as "bcds_score" column, and possibly more (details)
cxds

Examples

data("sce_chcl")
## create small data set using only 100 cells
sce_chcl_small = sce_chcl[, 1:100]
sce_chcl_small = bxds(sce_chcl_small)

cxds

Find doublets/multiplets in UMI scRNA-seq data;

Description

Annotates doublets/multiplets using co-expression based approach

Usage

cxds(sce, ntop = 500, binThresh = 0, verb = FALSE, retRes = FALSE, 
estNdbl = FALSE)

Arguments

sce single cell experiment (SingleCellExperiment) object to analyze; needs counts
in assays slot.
ntop integer, dimessageing number of top variance genes to consider. Default: 500
binThresh integer, minimum counts to consider a gene "present" in a cell. Default: 0
verb progress messages. Default: FALSE
retRes logical, whether to return gene pair scores & top-scoring gene pairs? Default: FALSE.
estNdbl logical, should the numer of doublets be estimated from the data. Enables dou-
blet calls. Default: FALSE. Use with caution.

Value

sce input sce object SingleCellExperiment with doublet scores added to colData as "cxds_score"
column.

Examples

data("sce_chcl")
## create small data set using only 100 cells
sce_chcl_small = sce_chcl[, 1:100]
sce_chcl_small = cxds(sce_chcl_small)
cxds_bcds_hybrid

Find doublets/multiples in UMI scRNA-seq data:

Description

Annotates doublets/multiplets using the hybrid approach

Usage

cxds_bcds_hybrid(sce, cxdsArgs = NULL, bcdsArgs = NULL, verb = FALSE, estNdbl = FALSE, force = FALSE)

Arguments

sce single cell experiment (SingleCellExperiment) object to analyze; needs counts in assays slot.
cxdsArgs list, arguments for cxds function in list form. Default: NULL
bcdsArgs list, arguments for bcds function in list form. Default: NULL
verb logical, switch on/off progress messages
estNdbl logical, should the number of doublets be estimated from the data. Enables doublet calls. Default: FALSE. Use with caution.
force logical, force a (re)run of cxds and bcds. Default: FALSE

Value

sce input sce object SingleCellExperiment with doublet scores added to colData as "hybrid_score" column.

Examples

data("sce_chcl")
## create small data set using only 100 cells
sce_chcl_small = sce_chcl[, 1:100]
sce_chcl_small = cxds_bcds_hybrid(sce_chcl_small)

cxds_getTopPairs

Extract top-scoring gene pairs from an SingleCellExperiment where cxds has been run

Description

Extract top-scoring gene pairs from an SingleCellExperiment where cxds has been run
get_dblCalls_ALL

Usage

cxds_getTopPairs(sce, n = 100)

Arguments

sce single cell experiment to analyze; needs "counts" in assays slot.
n integer. The number of gene pairs to extract. Default: 100

Value

matrix Matrix with two columns, each containing gene indexes for gene pairs (rows).

get_dblCalls_ALL Wrapper for getting doublet calls

Description

Wrapper for getting doublet calls

Usage

get_dblCalls_ALL(scrs_real, scrs_sim, rel_loss = 1)

Arguments

scrs_real numeric vector, the scores for the real/original data
scrs_sim numeric vector, the scores for the artificial doublets
rel_loss numeric scalar, relative weight of a false positive classification compared with a false negative. Default: 1 (same loss for fp and fn).

Value

numeric, matrix containing the (estimated) number of doublets, the score threshold and the fraction of artificial doublets missed (false negative rate, of sorts) as columns and four types of estimating: "youden", "balanced" and a false negative rate of artificial doublets of 0.1 and 0.01, respectively.
**get_dblCalls_dist**  
*Derive doublet calls from doublet scores*

**Description**

Given score vectors for real data and artificial doubles, derive doublet calls based on determining doublet score cutoffs.

**Usage**

```r
get_dblCalls_dist(scrs_real, scrs_sim, type = "balanced")
```

**Arguments**

- `scrs_real` numeric vector, the scores for the real/original data
- `scrs_sim` numeric vector, the scores for the artificial doublets
- `type` character or numeric, describes how the score threshold for calling doublets is determined. Either "balanced" or a number between zero and one that indicates the fraction of artificial doublets missed when making calls. Default: "balanced".

**Value**

numeric, vector containing the (estimated) number of doublets, the score threshold and the fraction of artificial doublets missed (false negative rate, of sorts)

---

**get_dblCalls_ROC**  
*Derive doublet calls from classification probabilities*

**Description**

Given class probabilities (or scores) discriminating real data from artificial doublets, derive doublet calls. Based on selecting a ROC cutoff, see *The Inconsistency of “Optimal” Cutpoints Obtained using Two Criteria basedon the Receiver Operating Characteristic Curve*, (doi).

**Usage**

```r
get_dblCalls_ROC(scrs_real, scrs_sim, rel_loss = 1)
```

**Arguments**

- `scrs_real` numeric vector, the scores for the real/original data
- `scrs_sim` numeric vector, the scores for the artificial doublets
- `rel_loss` numeric scalar, relative weight of a false positive classification compared with a false negative. Default: 1 (same loss for fp and fn).
Value

numeric, vector containing the (estimated) number of doublets, the score threshold and the fraction of artificial doublets missed (false negative rate, of sorts)

Example single cell experiment (SingleCellExperiment) object

Description

Example data set, created by randomly sampling genes and cells from a real data set (ch_cl, i.e., the cell lines data from https://satijalab.org/seurat/hashing_vignette.html). Contains raw counts in the counts assay slot.

Usage

sce_chcl

Format

a single cell experiment object (SingleCellExperiment) with raw counts in the counts in assays, and colData with experimental annotations.
Index

bcds, 2

cxds, 3
cxds_bcds_hybrid, 4
cxds_getTopPairs, 4

get_dblCalls_ALL, 5
get_dblCalls_dist, 6
get_dblCalls_ROC, 6

sce_chcl, 7