Package: Cleanet (via r-universe)

December 20, 2024

Type Package

Title Automated doublet detection and classification for cytometry data

Version 1.0.0

Author Matei Ionita

Maintainer Matei Ionita <matei.ionita@gmail.com>

Description Automated method for doublet detection in flow or mass cytometry data, based on simulating doublets and finding events whose protein expression patterns are similar to the simulated doublets.

Imports Rcpp, RcppHNSW, ggplot2, dplyr, tidyr, readr, reshape2, tibble, magrittr, methods

LinkingTo Rcpp, RcppArmadillo

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.3.1

Depends R (>= 3.5)

Suggests knitr, rmarkdown

VignetteBuilder knitr

Config/pak/sysreqs libicu-dev libx11-dev

Repository https://matei-ionita.r-universe.dev

RemoteUrl https://github.com/matei-ionita/cleanet

RemoteRef HEAD

RemoteSha 734ce609ff83daf80adced07a3e093e63d5c2a9d

Contents

classify_doublets 2
cleanet 3
compare_doublets_exp_obs 4
filter_debris_cytof 5
filter_debris_flow 6

Index **7**

`classify_doublets` *Classify doublets (or multiplets) based on component singlets.*

Description

Extends a classification of singlets into a classification of doublets.

Usage

```
classify_doublets(cleanet_res, singlet_clas, max_multi = 4)
```

Arguments

`cleanet_res`	The output of a call to the cleanet function.
`singlet_clas`	An array giving a classification of the singlets, whose length must match the number of singlet events returned in cleanet_res.
`max_multi`	The highest cardinality of a multiplet to be considered.

Value

An array with the same length as the number of doublets found in cleanet_res, specifying the composition of each doublet.

Examples

```
path <- system.file("extdata", "df_mdipa.csv", package="Cleanet")
df_mdipa <- read.csv(path, check.names=FALSE)
cols <- c("CD45", "CD123", "CD19", "CD11c", "CD16",
          "CD56", "CD294", "CD14", "CD3", "CD20",
          "CD66b", "CD38", "HLA-DR", "CD45RA",
          "DNA1", "DNA2")
cleanet_res <- cleanet(df_mdipa, cols, cofactor=5)
singlet_clas <- df_mdipa$label[which(cleanet_res$status!="Doublet")]
doublet_clas <- classify_doublets(cleanet_res, singlet_clas)
```

`cleanet` *Detect doublets in a single cytometry sample*

Description

Augments data with simulated doublets, computes nearest neighbors for augmented dataset, identifies doublets as those events with a high share of simulated doublets among nearest neighbors.

Usage

```
cleanet(df, cols, cofactor, thresh = 5, is_debris = NULL)
```

Arguments

`df`	A data frame containing protein expression data.
`cols`	Columns to use in analysis.
`cofactor`	Parameter of arcsinh transformation, applied before computing nearest neighbors. Recommended values are 5 for mass cytometry and 500-1000 for flow cytometry.
`thresh`	Among the 15 nearest neighbors, how many should be simulated doublets in order for the event to be classified as doublet?
`is_debris`	Optional, binary array with length matching the number of rows in df. TRUE for debris events, FALSE for everything else. This package includes helper functions to compute this for flow or mass cytometry data.

Value

A list with multiple elements, among them the singlet/doublet status of each event.

Examples

```
path <- system.file("extdata", "df_mdipa.csv", package="Cleanet")
df_mdipa <- read.csv(path, check.names=FALSE)
cols <- c("CD45", "CD123", "CD19", "CD11c", "CD16",
          "CD56", "CD294", "CD14", "CD3", "CD20",
          "CD66b", "CD38", "HLA-DR", "CD45RA",
          "DNA1", "DNA2")
cleanet_res <- cleanet(df_mdipa, cols, cofactor=5)
```

compare_doublets_exp_obs *Tabulate expected and observed proportions of doublet types.*

Description

Given compatible classifications of singlets and doublets, this function computes expected proportions of doublets as the product of the proportions of their components.

Usage

```
compare_doublets_exp_obs(doublet_clas, singlet_clas, cleanet_res)
```

Arguments

doublet_clas	An array giving a classification of the doublets, whose length must match the number of doublet events returned in cleanet_res.
singlet_clas	An array giving a classification of the singlets, whose length must match the number of singlet events returned in cleanet_res.
cleanet_res	The output of a call to the cleanet function.

Value

A data frame tabulating expected and observed proportions for each unique doublet type.

Examples

```
path <- system.file("extdata", "df_mdipa.csv", package="Cleanet")
df_mdipa <- read.csv(path, check.names=FALSE)
cols <- c("CD45", "CD123", "CD19", "CD11c", "CD16",
          "CD56", "CD294", "CD14", "CD3", "CD20",
          "CD66b", "CD38", "HLA-DR", "CD45RA",
          "DNA1", "DNA2")
cleanet_res <- cleanet(df_mdipa, cols, cofactor=5)
singlet_clas <- df_mdipa$label[which(cleanet_res$status!="Doublet")]
doublet_clas <- classify_doublets(cleanet_res, singlet_clas)
df_exp_obs <- compare_doublets_exp_obs(doublet_clas, singlet_clas, cleanet_res)
```

`filter_debris_cytof` *Flag debris in mass cytometry data.*

Description

Detect events with low distance from 0 in protein space. This function aims for high specificity, but not high sensitivity: for Cleanet's purposes, it suffices to deplete debris, even if not all of it is eliminated.

Usage

```
filter_debris_cytof(
  df,
  cols,
  cols_plot = c("DNA1", "CD45"),
  cofactor = 5,
  threshold = 0.3
)
```

Arguments

`df`	A data frame containing protein expression data.
`cols`	Columns to use in analysis. It is recommended to use the same ones in the call to cleanet.
`cols_plot`	Two columns that are used for visual feedback.
`cofactor`	Parameter for arcsinh transformation used before computing distances. 5 is a good default for mass cytometry data.
`threshold`	Number between 0 and 1; distances are scaled between 0 and 1 and events whose distance to the origin is smaller than the threshold are flagged.

Value

A binary array with the same length as the number of rows in df. TRUE for debris, FALSE for everything else.

Examples

```
path <- system.file("extdata", "df_mdipa.csv", package="Cleanet")
df_mdipa <- read.csv(path, check.names=FALSE)
cols <- c("CD45", "CD123", "CD19", "CD11c", "CD16",
          "CD56", "CD294", "CD14", "CD3", "CD20",
          "CD66b", "CD38", "HLA-DR", "CD45RA",
          "DNA1", "DNA2")
is_debris <- filter_debris_cytof(df_mdipa, cols)
```

`filter_debris_flow` *Flag debris in flow cytometry data.*

Description

Detect events in the lower left corner of FSC-A/SSC-A plots. This function aims for high specificity, but not high sensitivity: for Cleanet's purposes, it suffices to deplete debris, even if not all of it is eliminated.

Usage

```
filter_debris_flow(df, sum_max = 50000, cols = c("FSC-A", "SSC-A"))
```

Arguments

`df`	A data frame containing scattering channels and protein expression data.
`sum_max`	Numeric; events whose sum of FSC-A and SSC-A is smaller than this value are flagged.
`cols`	Names of columns to use. This function is intended for use with the area channel of forward and side scatter.

Value

A binary array with the same length as the number of rows in df. TRUE for debris, FALSE for everything else.

Index

classify_doublets, 2
cleanet, 3
compare_doublets_exp_obs, 4

filter_debris_cytof, 5
filter_debris_flow, 6