Package 'multipletR'

October 30, 2026

Type Package

Title Adaptive Detection of Human-Mouse Multiplets in PDX Single-Cell Data

Version 0.99.9

Description Detects human-mouse multiplets in patient-derived xenograft (PDX) single-cell RNA-seq data using an adaptive threshold method that does not assume a fixed species proportion. Takes a 10x CellRanger GEM classification file and returns the data with added multiplet classifications, with optional diagnostic plots and helpers to annotate a Seurat or SingleCellExperiment object with multiplet classifications or remove multiplets from it. The adaptive approach starts from a conservative central region of balanced droplets and expands its thresholds until the selected cells no longer resemble true multiplets, which handles the imbalanced species mixtures common in real PDX samples.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

Imports diptest, ggplot2, methods, patchwork, scales, SingleCellExperiment, stats, utils

Suggests BiocStyle, Seurat, SummarizedExperiment, knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

biocViews SingleCell, RNASeq, Preprocessing, QualityControl, Transcriptomics, Classification

URL https://github.com/Alex05a/multipletR

BugReports https://github.com/Alex05a/multipletR/issues

Config/testthat/edition 3

git_url https://git.bioconductor.org/packages/multipletR

git_branch devel

git_last_commit a3af58a

git_last_commit_date 2026-09-21

Repository Bioconductor 3.24

Date/Publication 2026-10-29

Author Alexandra Gerveni [aut, cre] (ORCID:
<https://orcid.org/0009-0004-5428-5889>),
Amy Olex [aut],
Mikhail Dozmorov [aut, fnd] (P50AA022537),
Gamze Bulut [aut],
J. Chuck Harrell [fnd] (U54CA283762),
Jose Trevino [fnd] (U54CA283762)

Maintainer Alexandra Gerveni <alexandra.gerveni@hotmail.com>

Contents

multipletR-package 2
.plot_percent 3
.plot_total_reads 4
calculate_adaptive_metrics 4
detect_multiplets 5
expand_threshold 6
find_adaptive_thresholds_3t 7
find_mode 8
initialize_thresholds 8
remove_multiplets 9
select_cells_in_rectangle 10
should_stop_expanding 10
test_bimodality 11

Index **12**

`multipletR-package` *multipletR: Adaptive Detection of Human-Mouse Multiplets in PDX Single-Cell Data*

Description

Detects human-mouse multiplets in patient-derived xenograft (PDX) single-cell RNA-seq data using an adaptive threshold method that does not assume a fixed species proportion. It reads a 10x CellRanger GEM classification file and returns the data with added multiplet classifications, with optional diagnostic plots and helpers to annotate or filter a Seurat or SingleCellExperiment object.

Main functions

- `detect_multiplets`: read a 10x CellRanger GEM classification file, detect human-mouse multiplets with the adaptive threshold method, and write out the classified data, with optional diagnostic plots.
- `remove_multiplets`: add multipletR's per-cell classification to a Seurat or SingleCellExperiment object and, optionally, remove the detected multiplets so the object is ready for downstream analysis.

Author(s)

Maintainer: Alexandra Gerveni <alexandra.gerveni@hotmail.com> (ORCID)

Authors:

- Amy Olex
- Mikhail Dozmorov (P50AA022537) [funder]
- Gamze Bulut

Other contributors:

- J. Chuck Harrell (U54CA283762) [funder]
- Jose Trevino (U54CA283762) [funder]

See Also

Useful links:

- https://github.com/Alex05a/multipletR
- Report bugs at https://github.com/Alex05a/multipletR/issues

---

.plot_percent *Percent plots: total reads vs percent mouse, 10x vs multipletR classification*

Description

Draws two side-by-side scatter plots of total reads (x) against percent mouse content (y): one colored by the original 10x classification, one by the multipletR classification.

Usage

```
.plot_percent(cc)
```

Arguments

cc	Data frame with columns total_reads, pct_mouse_10x, call_10x, and our_classification.

Value

A patchwork object (two panels side by side).

.plot_total_reads *Total-reads plots: mouse reads vs human reads, 10x vs multipletR classification*

Description

Draws two side-by-side scatter plots of mouse reads (x) against human reads (y): one colored by the original 10x classification, one by the multipletR classification.

Usage

```
.plot_total_reads(cc)
```

Arguments

cc	Data frame with columns human_10x, mouse_10x, call_10x, and our_classification.

Value

A patchwork object (two panels side by side).

calculate_adaptive_metrics *Distribution-quality metrics for a set of selected cells*

Description

Computes the number of cells, the percent overlap between the human and mouse log-read distributions, the area overlap, the difference between the distribution modes, and bimodality of each distribution.

Usage

```
calculate_adaptive_metrics(data, min_cells_for_bimodality = 30)
```

Arguments

data	Data frame with human_10x and mouse_10x columns.
min_cells_for_bimodality	Minimum cells to run the bimodality test.

Value

A named list of metrics.

`detect_multiplets` *Detect human-mouse multiplets in PDX single-cell data*

Description

Reads a 10x CellRanger GEM classification file, applies the adaptive threshold method to identify human-mouse multiplets without assuming a fixed species proportion, and writes the input data back out with added classification columns. Optionally produces diagnostic plots.

Usage

```
detect_multiplets(
  fileIn,
  fileOut,
  T1 = 70,
  T2 = 30,
  T3 = 25,
  plotPercent = TRUE,
  plotTotalReads = FALSE,
  overlapDrop = 10,
  modeDiff = 0.9,
  verbose = FALSE
)
```

Arguments

`fileIn`	Path to the input GEM classification CSV (from 10x CellRanger). Expected columns: a barcode column, a human read-count column (GRCh38), a mouse read-count column (e.g. GRCm39 / mm10 / mm39 / mouse_reads), and a 10x call column.
`fileOut`	Path to write the output CSV. The output is the input data with added columns: our_classification (Multiplet / Singlet), pct_human, and pct_mouse.
`T1`	Upper percent-mouse threshold: the starting upper bound on percent mouse content. Cells above this are too mouse-heavy. Default 70.
`T2`	Lower percent-mouse threshold: the starting lower bound on percent mouse content (equivalently, the percent-human side). Cells below this are too human-heavy. Default 30.
`T3`	Total-reads threshold: the starting lower bound on total reads, given as a percentile of the total-reads range (e.g. 25 = 25th percentile). Default 25.
`plotPercent`	Logical, whether to draw the two percent plots (one for the original 10x classification, one for our classification), total reads vs percent mouse. Default TRUE.
`plotTotalReads`	Logical, whether to draw the total-reads plots (mouse reads vs human reads, colored by 10x and our classification). Default FALSE.
`overlapDrop`	Stop expanding a threshold when the overlap between the human and mouse read distributions drops by more than this percentage from its running peak. Lower values are more conservative. Default 10.
`modeDiff`	Stop expanding when the difference between the human and mouse distribution modes increases by more than this amount from the previous step. Default 0.9.
`verbose`	Logical, print progress messages including the final thresholds, multiplet count, and the step-by-step algorithm trace. Set FALSE to run silently. Default FALSE.

Details

The three starting thresholds T1, T2 and T3 follow the notation used in the manuscript. They define the conservative starting region that is then expanded adaptively; the defaults are the values we recommend and they rarely need to be changed.

Value

Invisibly, a data frame: the input data with added columns our_classification, pct_human, and pct_mouse. Also written to `fileOut`.

Examples

```
# Detect multiplets in the bundled example PDX dataset (sample PC65)
gem_file <- system.file("extdata", "PC65_gem_classification.csv",
  package = "multipletR"
)
res <- detect_multiplets(gem_file, tempfile(fileext = ".csv"),
  plotPercent = FALSE, plotTotalReads = FALSE
)
table(res$our_classification)
```

`expand_threshold` *Expand one threshold by a single step*

Description

Expand one threshold by a single step

Usage

```
expand_threshold(thresholds, which_threshold, step_size, min_reads, max_reads)
```

Arguments

`thresholds`	Current thresholds list.
`which_threshold`	1 = T1 (mouse% up), 2 = T2 (human% down), 3 = T3 (reads down).
`step_size`	Step magnitude.
`min_reads`	Minimum total reads (floor for T3).
`max_reads`	Maximum total reads (unused; upper boundary is fixed).

Value

The updated thresholds list.

`find_adaptive_thresholds_3t` *Adaptive three-threshold multiplet detection (internal engine)*

Description

Round-robin expansion of three thresholds (upper percent-mouse, lower percent-mouse, lower total reads) from a conservative starting region until the human and mouse read distributions begin to separate. The upper reads boundary is fixed at the maximum. This is the internal engine called by `detect_multiplets`.

Usage

```
find_adaptive_thresholds_3t(
  data,
  percent_lower = 30,
  percent_upper = 70,
  reads_percentile = 25,
  step_size_reads = 500,
  step_size_pct = 1,
  overlap_drop_threshold = 10,
  mode_diff_increase_threshold = 0.9,
  overlap_patience = 3,
  min_cells_for_bimodality = 30,
  verbose = FALSE
)
```

Arguments

`data`	Data frame with columns total_reads, pct_mouse_10x, human_10x, mouse_10x.
`percent_lower`	Lower percent-mouse starting threshold (T2). Default 30.
`percent_upper`	Upper percent-mouse starting threshold (T1). Default 70.
`reads_percentile`	Lower total-reads starting threshold as a percentile. Default 25.
`step_size_reads`	Expansion step for the reads threshold. Default 500.
`step_size_pct`	Expansion step for the percent thresholds. Default 1.
`overlap_drop_threshold`	Stop when overlap drops more than this percent below the running peak. Default 10.
`mode_diff_increase_threshold`	Stop when the mode difference increases more than this. Default 0.9.
`overlap_patience`	Consecutive below-peak steps tolerated before stopping. Default 3.
`min_cells_for_bimodality`	Minimum cells to run the bimodality test. Default 30.
`verbose`	Print progress. Default FALSE.

Value

A list with the final and initial thresholds, stop reasons, iteration counts, initial and final metrics, the detected multiplets, the expansion history, and the total number of rounds.

`find_mode` *Find the primary mode of a distribution via kernel density*

Description

Find the primary mode of a distribution via kernel density

Usage

```
find_mode(x)
```

Arguments

`x`	Numeric vector.

Value

The x-position of the density peak, or NA.

`initialize_thresholds` *Initialize the conservative starting thresholds*

Description

Initialize the conservative starting thresholds

Usage

```
initialize_thresholds(
  data,
  percent_lower = 30,
  percent_upper = 70,
  reads_percentile = 25
)
```

Arguments

`data`	Data frame with a total_reads column.
`percent_lower`	Lower percent-mouse threshold (T2). Default 30.
`percent_upper`	Upper percent-mouse threshold (T1). Default 70.
`reads_percentile`	Lower total-reads threshold as a percentile. Default 25.

Value

A list of the initial thresholds.

`remove_multiplets` *Annotate (and optionally remove) multiplets in a Seurat or SingleCellExperiment object*

Description

Adds multipletR's per-cell classification to a single-cell object's cell metadata. By default all cells are kept and only annotated (useful for visualizing where the multiplets fall, e.g. on a UMAP colored by `multipletR_class` before deciding whether to filter). Set `remove = TRUE` to drop the detected multiplets so the object is ready for downstream analysis. The object type (Seurat or `SingleCellExperiment`) is detected automatically.

Usage

```
remove_multiplets(
  x,
  multiplets,
  remove = FALSE,
  barcode_col = "barcode",
  verbose = TRUE
)
```

Arguments

`x`	A single-cell object whose cell names (`colnames`) are barcodes: either a Seurat object or a `SingleCellExperiment`. The object type is detected automatically.
`multiplets`	The data frame returned by `detect_multiplets` (must contain `barcode`, `our_classification`, `pct_human`, and `pct_mouse`).
`remove`	Logical. If `TRUE`, the detected multiplets are removed from the returned object. If `FALSE` (the default), all cells are kept and only the metadata is added.
`barcode_col`	Name of the barcode column in `multiplets`. Default "barcode".
`verbose`	Logical. If `TRUE` (the default), print a summary of the annotation and removal. Set `FALSE` to suppress these messages.

Details

The classification follows the same idea as doublet-detection tools such as DoubletFinder: each cell is labeled Human, Mouse, or Multiplet, and the percent human / percent mouse are added per cell.

Value

The input object with three added cell-metadata columns (`multipletR_class`, `multipletR_pct_human`, `multipletR_pct_mouse`). If `remove = TRUE`, the multiplet cells are also dropped.

Examples

```
# Detect multiplets in the bundled example dataset
gem_file <- system.file("extdata", "PC65_gem_classification.csv",
  package = "multipletR"
)
res <- detect_multiplets(gem_file, tempfile(fileext = ".csv"),
```

```
  plotPercent = FALSE, plotTotalReads = FALSE
)

# Build a minimal Seurat object whose cell names are the same barcodes
counts <- matrix(rpois(20 * nrow(res), 5),
  nrow = 20,
  dimnames = list(paste0("gene", 1:20), res$barcode)
)
seu <- Seurat::CreateSeuratObject(counts)

# Annotate only (keep all cells), e.g. to visualize before filtering
seu <- remove_multiplets(seu, res)
table(seu$multipletR_class)

# Or annotate and remove the multiplets in one step
seu_clean <- remove_multiplets(seu, res, remove = TRUE)
```

`select_cells_in_rectangle` *Select cells within the current threshold rectangle*

Description

Select cells within the current threshold rectangle

Usage

```
select_cells_in_rectangle(data, thresholds)
```

Arguments

`data`	Data frame with total_reads and pct_mouse_10x columns.
`thresholds`	A thresholds list from initialize_thresholds().

Value

The subset of rows inside the rectangle.

`should_stop_expanding` *Decide whether a threshold should stop expanding*

Description

Decide whether a threshold should stop expanding

Usage

```
should_stop_expanding(
  current_metrics,
  previous_metrics,
  overlap_drop_threshold = 10,
  mode_diff_increase_threshold = 0.9
)
```

Arguments

current_metrics	Metrics for the candidate selection.
previous_metrics	Metrics for the last accepted selection.
overlap_drop_threshold	Stop if overlap drops more than this from the previous step.
mode_diff_increase_threshold	Stop if the mode difference increases more than this.

Value

A list: stop (logical) and reason (character).

---

test_bimodality *Test a distribution for bimodality (Hartigan dip test)*

Description

Test a distribution for bimodality (Hartigan dip test)

Usage

```
test_bimodality(x, alpha = 0.05, min_cells_for_test = 30)
```

Arguments

x	Numeric vector.
alpha	Significance level. Default 0.05.
min_cells_for_test	Minimum values needed to run the test. Default 30.

Value

A list: is_bimodal, p_value, dip_statistic, reason.

Index

∗ **internal**
.plot_percent, 3
.plot_total_reads, 4
calculate_adaptive_metrics, 4
expand_threshold, 6
find_adaptive_thresholds_3t, 7
find_mode, 8
initialize_thresholds, 8
multipletR-package, 2
select_cells_in_rectangle, 10
should_stop_expanding, 10
test_bimodality, 11
.plot_percent, 3
.plot_total_reads, 4

calculate_adaptive_metrics, 4

detect_multiplets, *2*, 5, *7*, *9*

expand_threshold, 6

find_adaptive_thresholds_3t, 7
find_mode, 8

initialize_thresholds, 8

multipletR *(*multipletR-package*)*, 2
multipletR-package, 2

remove_multiplets, *2*, 9

select_cells_in_rectangle, 10
should_stop_expanding, 10

test_bimodality, 11