Package ‘gateR’

February 2, 2023

Type Package

Title Flow/Mass Cytometry Gating via Spatial Kernel Density Estimation

Version 0.1.13

Date 2023-02-01

Maintainer Ian D. Buller <ian.buller@alumni.emory.edu>

Description Estimates statistically significant marker combination values within which one immunologically distinctive group (i.e., disease case) is more associated than another group (i.e., healthy control), successively, using various combinations (i.e., “gates”) of markers to examine features of cells that may be different between groups. For a two-group comparison, the ‘gateR’ package uses the spatial relative risk function estimated using the ‘sparr’ package. Details about the ‘sparr’ package methods can be found in the tutorial: Davies et al. (2018) <doi:10.1002/sim.7577>. Details about kernel density estimation can be found in J. F. Bithell (1990) <doi:10.1002/sim.4780090616>. More information about relative risk functions using kernel density estimation can be found in J. F. Bithell (1991) <doi:10.1002/sim.4780101112>.

License Apache License (>= 2.0)

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Depends R (>= 3.5.0)

Imports fields, graphics, grDevices, lifecycle, rlang, sparr, SpatialPack, spatstat.geom, stats, terra, tibble

Suggests dplyr, R.rsp, spelling, testthat, usethis, utils

VignetteBuilder R.rsp

Language en-US

URL https://github.com/lance-waller-lab/gateR

BugReports https://github.com/lance-waller-lab/gateR/issues

NeedsCompilation no
The gateR Package: Flow/Mass Cytometry Gating via Spatial Kernel Density Estimation

Description

Estimates statistically significant fluorescent marker combination values within which one immunologically distinctive group (i.e., disease case) is more associated than another group (i.e., healthy control), successively, using various combinations (i.e., "gates") of fluorescent markers to examine features of cells that may be different between groups.

Details


This package provides a function to perform a gating strategy for flow cytometry data. The 'gateR' package also provides basic visualization for each gate.

Key content of the 'gateR' package include:

Gating Strategy

gating Extracts cells within statistically significant combinations of fluorescent markers, successively, for a set of markers. Statistically significant combinations are identified using two-tailed p-values of a relative risk surface assuming asymptotic normality. This function is currently available for two-level comparisons of a single condition (e.g., case/control) or two conditions (e.g.,
gating

Gating strategy for mass cytometry data using spatial relative risk functions

description

Extracts cells within statistically significant combinations of fluorescent markers, successively, for a set of markers. Statistically significant combinations are identified using two-tailed p-values of a relative risk surface assuming asymptotic normality. This function is currently available for two-level comparisons of a single condition (e.g., case/control) or two conditions (e.g., case/control at time 1 and time 2). Provides functionality for basic visualization and multiple testing correction.

rrs

Estimates a relative risk surface and computes the asymptotic p-value surface for a single gate with a single condition, including features for basic visualization. This function is used internally within the gating function to extract the points within the significant areas. This function can also be used as a standalone function.

lotrrs

Estimates a ratio of relative risk surfaces and computes the asymptotic p-value surface for a single gate with two conditions, including features for basic visualization. This function is used internally within the gating function to extract the points within the significant areas. This function can also be used as a standalone function.

Flow Cytometry Data

randCyto

A sample dataset containing information about flow cytometry data with two binary categorical variables. The data are a random subset of the 'extdata' data in the 'flowWorkspaceData' package found on Bioconductor http://bioconductor.org/packages/release/data/experiment/html/flowWorkspaceData.html and formatted for 'gateR' input.

Dependencies

The 'gateR' package relies heavily upon sparr, spatstat.geom, and terra. For a two-level comparison, the spatial relative risk function uses the risk function. The calculation of a Bonferroni correction for multiple testing accounting for the spatial correlation of the estimated surface uses the modified.ttest function. Basic visualizations rely on the image.plot function.

Author(s)

Ian D. Buller

Social & Scientific Systems, Inc., a division of DLH Corporation, Silver Spring, Maryland, USA (current); Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, Maryland, USA (former); Environmental Health Sciences, James T. Laney School of Graduate Studies, Emory University, Atlanta, Georgia, USA. (original)

Maintainer: I.D.B. <ian.buller@alumni.emory.edu>
gating

Usage

gating(
  dat,
  vars,
  n_condition = c(1, 2),
  numerator = TRUE,
  bandw = NULL,
  alpha = 0.05,
  p_correct = "none",
  nbc = NULL,
  plot_gate = FALSE,
  save_gate = FALSE,
  name_gate = NULL,
  path_gate = NULL,
  rcols = c("FF0000", "CCCCCC", "0000FF"),
  lower_lrr = NULL,
  upper_lrr = NULL,
  c1n = NULL,
  c2n = NULL,
  win = NULL,
  ...
)

Arguments

dat Input data frame flow cytometry data with the following features (columns): 1) ID, 2) Condition A ID, 3) Condition B ID (optional), and a set of markers.

vars A vector of characters with the name of features (columns) within dat to use as markers for each gate. See details below.

n_condition A numeric value of either 1 or 2 designating if the gating is performed with one condition or two conditions.

numerator Logical. If TRUE (the default), cells will be extracted within all statistically significant numerator (i.e., case) clusters. If FALSE, cells will be extracted within all statistically significant denominator (i.e., control) clusters.

bandw Optional, numeric. Fixed bandwidth for the kernel density estimation. Default is based on the internal [sparr]{OS} function.

alpha Numeric. The two-tailed alpha level for significance threshold (default is 0.05).

p_correct Optional. Character string specifying whether to apply a correction for multiple comparisons including a False Discovery Rate p_correct = "FDR", a spatially dependent Sidak correction p_correct = "correlated Sidak", a spatially dependent Bonferroni correction p_correct = "correlated Bonferroni", an independent Sidak correction p_correct = "uncorrelated Sidak", an independent Bonferroni correction p_correct = "uncorrelated Bonferroni", and a correction based on Random Field Theory using an equation by Adler and Hasofer p_correct = "Adler and Hasofer" or an equation by Friston et al. p_correct
gating

= "Friston". If \( p_{\text{correct}} = \text{"none"} \) (the default), then no correction is applied.

nbc

Optional. An integer for the number of bins when \( p_{\text{correct}} = \text{"correlated"} \).
Similar to \text{nbccl} argument in \text{modified.ttest}. The default is 30.

plot_gate

Logical. If \text{TRUE}, the output includes basic data visualizations.

save_gate

Logical. If \text{TRUE}, the output saves each visualization as a separate PNG file.

name_gate

Optional, character. The filename of the visualization(s). The default is "gate_k" where "k" is the gate number.

path_gate

Optional, character. The path of the visualization(s). The default is the current working directory.

rcols

Character string of length three (3) specifying the colors for: 1) group A (numerator), 2) neither, and 3) group B (denominator) designations. The defaults are \text{c("FF0000", "cccccc", ")0000FF")} or \text{c("red", "grey80", "blue")}.

lower_lrr

Optional, numeric. Lower cut-off value for the log relative risk value in the color key (typically a negative value). The default is no limit, and the color key will include the minimum value of the log relative risk surface.

upper_lrr

Optional, numeric. Upper cut-off value for the log relative risk value in the color key (typically a positive value). The default is no limit, and the color key will include the maximum value of the log relative risk surface.

c1n

Optional, character. The name of the level for the numerator of condition A. The default is NULL, and the first level is treated as the numerator.

c2n

Optional, character. The name of the level for the numerator of condition B. The default is NULL, and the first level is treated as the numerator.

win

Optional. Object of class \text{owin} for a custom two-dimensional window within which to estimate the surfaces. The default is NULL and calculates a convex hull around the data.

... Arguments passed to \text{risk} to select resolution.

doplot

[Deprecated] \text{doplot} is no longer supported and has been renamed \text{plot_gate}.

verbose

[Deprecated] \text{verbose} is no longer supported; this function will not display verbose output from internal \text{risk} function.

Details

This function performs a sequential gating strategy for mass cytometry data comparing two levels with one or two conditions. Gates are typically two-dimensional space comprised of two fluorescent markers. The two-level comparison allows for the estimation of a spatial relative risk function and the computation of \( p \)-value based on an assumption of asymptotic normality. Cells within statistically significant areas are extracted and used in the next gate. This function relies heavily upon the \text{risk} function. Basic visualization is available if \text{plot_gate} = \text{TRUE}.

The \text{vars} argument must be a vector with an even-numbered length where the odd-numbered elements are the markers used on the x-axis of a gate, and the even-numbered elements are the markers used on the y-axis of a gate. For example, if \text{vars} = \text{c("V1", "V2", "V3", and "V4")} then the first gate is "V1" on the x-axis and "V2" on the y-axis and then the second gate is "V3" on the x-axis and "V4" on the y-axis. Makers can be repeated in successive gates.
The n_condition argument specifies if the gating strategy is performed for one condition or two conditions. If n_condition = 1, then the function performs a one condition gating strategy using the internal rrs function, which computes the statistically significant areas (clusters) of a relative risk surface at each gate and selects the cells within the clusters specified by the numerator argument. If n_condition = 2, then the function performs a two conditions gating strategy using the internal lotrrs function, which computes the statistically significant areas (clusters) of a ratio of relative risk surfaces at each gate and selects the cells within the clusters specified by the numerator argument. The condition variable(s) within dat must be of class 'factor' with two levels. The first level is considered the numerator (i.e., "case") value, and the second level is considered the denominator (i.e., "control") value. The levels can also be specified using the c1n and c2n parameters. See the documentation for the internal rrs and lotrrs functions for more details.

The p-value surface of the ratio of relative risk surfaces is estimated assuming asymptotic normality of the ratio value at each gridded knot. The bandwidth is fixed across all layers.

Provides functionality for a correction for multiple testing. If p_correct = "FDR", calculates a False Discovery Rate by Benjamini and Hochberg. If p_correct = "uncorrelated Sidak", calculates an independent Sidak correction. If p_correct = "uncorrelated Bonferroni", calculates an independent Bonferroni correction. If p_correct = "correlated Sidak" or if p_correct = "correlated Bonferroni", then the corrections take into account the spatial correlation of the surface. (NOTE: If p_correct = "correlated Sidak" or if p_correct = "correlated Bonferroni", it may take a considerable amount of computation resources and time to calculate). If p_correct = "Adler and Hasofer" or if p_correct = "Friston", then calculates a correction based on Random Field Theory. If p_correct = "none" (the default), then the function does not account for multiple testing and uses the uncorrected alpha level. See the internal pval_correct function documentation for more details.

Value

An object of class list. This is a named list with the following components:

obs An object of class 'tibble' of the same features as dat that includes the information for the cells extracted with significant clusters in the final gate.

n An object of class 'list' of the sample size of cells at each gate. The length is equal to the number of successful gates plus the final result.

gate An object of class 'list' of 'rrs' objects from each gate. The length is equal to the number of successful gates.

note An object of class 'character' of the gating diagnostic message.

The objects of class 'rrs' is similar to the output of the risk function with two additional components:

rr An object of class 'im' with the relative risk surface.

f An object of class 'im' with the spatial density of the numerator.

g An object of class 'im' with the spatial density of the denominator.

P An object of class 'im' with the asymptotic p-value surface.

lrr An object of class 'im' with the log relative risk surface.

alpha A numeric value for the alpha level used within the gate.
Examples

```r
if (interactive()) {
  ## Single condition, no multiple testing correction
  test_gate <- gating(dat = randCyto,
                      vars = c("arcsinh_CD4", "arcsinh_CD38",
                               "arcsinh_CD8", "arcsinh_CD3"),
                      n_condition = 1)
}
```

**Description**

Estimates a ratio of relative risk surfaces and computes the asymptotic p-value surface for a single gate with two conditions. Includes features for basic visualization. This function is used internally within the `gating` function to extract the points within the significant areas. This function can also be used as a standalone function.

**Usage**

```r
lotrrs(
  dat,
  bandw = NULL,
  alpha = 0.05,
  p_correct = "none",
  nbc = NULL,
  plot_gate = FALSE,
  save_gate = FALSE,
  name_gate = NULL,
  path_gate = NULL,
  rcols = c("#FF0000", "#CCCCCC", "#0000FF"),
  lower_lrr = NULL,
  upper_lrr = NULL,
  c1n = NULL,
  c2n = NULL,
  win = NULL,
  ...
)
```

**Arguments**

bandw  Optional, numeric. Fixed bandwidth for the kernel density estimation. Default is based on the internal [sparr] function.

alpha    Numeric. The two-tailed alpha level for significance threshold (default is 0.05).

p_correct Optional. Character string specifying whether to apply a correction for multiple comparisons including a False Discovery Rate p_correct = "FDR", a spatially dependent Sidak correction p_correct = "correlated Sidak", a spatially dependent Bonferroni correction p_correct = "correlated Bonferroni", an independent Sidak correction p_correct = "uncorrelated Sidak", an independent Bonferroni correction p_correct = "uncorrelated Bonferroni", and a correction based on Random Field Theory using an equation by Adler and Hasofer p_correct = "Adler and Hasofer" or an equation by Friston et al. p_correct = "Friston". If p_correct = "none" (the default), then no correction is applied.

nbc       Optional. An integer for the number of bins when p_correct = "correlated". Similar to nbclass argument in modified.ttest. The default is 30.

plot_gate Logical. If TRUE, the output includes basic data visualization.

save_gate Logical. If TRUE, the output saves the visualization as a separate PNG file.

name_gate Optional, character. The filename of the visualization. The default is "gate".

path_gate Optional, character. The path of the visualization. The default is the current working directory.

cols      Character string of length three (3) specifying the colors for: 1) group A (numerator), 2) neither, and 3) group B (denominator) designations. The defaults are c("#FF0000", "#cccccc", "#0000FF") or c("red", "grey80", "blue").

lower_lrr Optional, numeric. Lower cut-off value for the log relative risk value in the color key (typically a negative value). The default is no limit, and the color key will include the minimum value of the log relative risk surface.

upper_lrr Optional, numeric. Upper cut-off value for the log relative risk value in the color key (typically a positive value). The default is no limit, and the color key will include the maximum value of the log relative risk surface.

c1n        Optional, character. The name of the level for the numerator of condition A. The default is NULL, and the first level is treated as the numerator.

c2n        Optional, character. The name of the level for the numerator of condition B. The default is NULL, and the first level is treated as the numerator.

win        Optional. Object of class owin for a custom two-dimensional window within which to estimate the surfaces. The default is NULL and calculates a convex hull around the data.

...        Arguments passed to risk to select resolution.

doplot     [Deprecated] doplot is no longer supported and has been renamed plot_gate.

verbose    [Deprecated] verbose is no longer supported; this function will not display verbose output from internal risk function.
Details

This function estimates a ratio of relative risk surfaces and computes the asymptotic p-value surface for a single gate with two conditions using three successive `risk` functions. A relative risk surface is estimated for Condition A at each level of Condition B, and then a ratio of the two relative risk surfaces is computed.

\[
RR_{\text{Condition B}1} = \frac{\text{Condition A2 of B1}}{\text{Condition A1 of B1}}
\]

\[
RR_{\text{Condition B}2} = \frac{\text{Condition A2 of B2}}{\text{Condition A1 of B2}}
\]

\[
\ln(rRR) = \ln \left( \frac{RR_{\text{Condition B2}}}{CRR_{\text{Condition B2}}} \right)
\]

The p-value surface of the ratio of relative risk surfaces is estimated assuming asymptotic normality of the ratio value at each gridded knot. The bandwidth is fixed across all layers. Basic visualization is available if `plot_gate = TRUE`.

Provides functionality for a correction for multiple testing. If `p_correct = "FDR"`, calculates a False Discovery Rate by Benjamini and Hochberg. If `p_correct = "uncorrelated Sidak"`, calculates an independent Sidak correction. If `p_correct = "uncorrelated Bonferroni"`, calculates an independent Bonferroni correction. If `p_correct = "correlated Sidak"` or if `p_correct = "correlated Bonferroni"`, then the corrections take into account the spatial correlation of the surface. (NOTE: If `p_correct = "correlated Sidak"` or if `p_correct = "correlated Bonferroni"`, it may take a considerable amount of computation resources and time to calculate). If `p_correct = "Adler and Hasofer"` or if `p_correct = "Friston"`, then calculates a correction based on Random Field Theory. If `p_correct = "none"` (the default), then the function does not account for multiple testing and uses the uncorrected `alpha` level. See the internal `pval_correct` function documentation for more details.

The two condition variables (Condition A and Condition B) within `dat` must be of class 'factor' with two levels. The first level in each variable is considered the numerator (i.e., "case") value, and the second level is considered the denominator (i.e., "control") value. The levels can also be specified using the `c1n` and `c2n` parameters.

Value

An object of class 'list' where each element is an object of class 'rrs' created by the `risk` function with two additional components:

- `rr` An object of class 'im' with the relative risk surface.
- `f` An object of class 'im' with the spatial density of the numerator.
- `g` An object of class 'im' with the spatial density of the denominator.
- `P` An object of class 'im' with the asymptotic p-value surface.
- `lrr` An object of class 'im' with the log relative risk surface.
- `alpha` A numeric value for the alpha level used within the gate.
Examples

test_lotrrs <- lotrrs(dat = randCyto)

Description

A sample dataset containing information about flow cytometry data with two binary conditions and four markers. The data are a random subset of the 'extdata' data in the 'flowWorkspaceData' package found on Bioconductor [http://bioconductor.org/packages/release/data/experiment/html/flowWorkspaceData.html](http://bioconductor.org/packages/release/data/experiment/html/flowWorkspaceData.html) and formatted for 'gateR' input. The selected markers are arcsinh transformed.

Usage

randCyto

Format

A data frame with 11763 rows and 7 variables:

- id  cell ID number
- g1  binary condition #1
- g2  binary condition #2
- arcsinh_CD4  arcsinh-transformed CD4
- arcsinh_CD38  arcsinh-transformed CD38
- arcsinh_CD8  arcsinh-transformed CD8
- arcsinh_CD3  arcsinh-transformed CD3

Source


Examples

head(randCyto)
**Description**

Estimates a relative risk surface and computes the asymptotic p-value surface for a single gate with a single condition, including features for basic visualization. This function is used internally within the `gating` function to extract the points within the significant areas. This function can also be used as a standalone function.

**Usage**

```r
rrs(
  dat,
  bandw = NULL,
  alpha = 0.05,
  p_correct = "none",
  nbc = NULL,
  plot_gate = FALSE,
  save_gate = FALSE,
  name_gate = NULL,
  path_gate = NULL,
  rcols = c("#FF0000", "#CCCCCC", "#0000FF"),
  lower_lrr = NULL,
  upper_lrr = NULL,
  c1n = NULL,
  win = NULL,
  ...
)
```

**Arguments**

- `dat` Input data frame flow cytometry data with four (4) features (columns): 1) ID, 2) Condition A ID, 3) Marker A as x-coordinate, 4) Marker B as y-coordinate.
- `bandw` Optional, numeric. Fixed bandwidth for the kernel density estimation. Default is based on the internal `[sparr]{OS}` function.
- `alpha` Numeric. The two-tailed alpha level for significance threshold (default is 0.05).
- `p_correct` Optional. Character string specifying whether to apply a correction for multiple comparisons including a False Discovery Rate `p_correct = "FDR"`, a spatially dependent Sidak correction `p_correct = "correlated Sidak"`, a spatially dependent Bonferroni correction `p_correct = "correlated Bonferroni"`, an independent Sidak correction `p_correct = "uncorrelated Sidak"`, an independent Bonferroni correction `p_correct = "uncorrelated Bonferroni"`, and a
correction based on Random Field Theory using an equation by Adler and Hasofer. If p_correct = "Adler and Hasofer" or an equation by Friston et al. p_correct = "Friston". If p_correct = "none" (the default), then no correction is applied.

nbc Optional. An integer for the number of bins when p_correct = "correlated". Similar to nbclass argument in `modified.ttest`. The default is 30.

plot_gate Logical. If TRUE, the output includes basic data visualization.

save_gate Logical. If TRUE, the output saves the visualization as a separate PNG file.

name_gate Optional, character. The filename of the visualization. The default is "gate".

path_gate Optional, character. The path of the visualization. The default is the current working directory.

rcols Character string of length three (3) specifying the colors for: 1) group A (numerator), 2) neither, and 3) group B (denominator) designations. The defaults are c("#FF0000", "#cccccc", "#0000FF") or c("red", "grey80", "blue").

lower_lrr Optional, numeric. Lower cut-off value for the log relative risk value in the color key (typically a negative value). The default is no limit, and the color key will include the minimum value of the log relative risk surface.

upper_lrr Optional, numeric. Upper cut-off value for the log relative risk value in the color key (typically a positive value). The default is no limit, and the color key will include the maximum value of the log relative risk surface.

c1n Optional, character. The name of the level for the numerator of condition A. The default is NULL, and the first level is treated as the numerator.

win Optional. Object of class `owin` for a custom two-dimensional window within which to estimate the surfaces. The default is NULL and calculates a convex hull around the data.

... Arguments passed to `risk` to select resolution.

doplot [Deprecated] doplot is no longer supported and has been renamed `plot_gate`.

verbose [Deprecated] verbose is no longer supported; this function will not display verbose output from internal `risk` function.

Details

This function estimates a relative risk surface and computes the asymptotic p-value surface for a single gate and single condition using the `risk` function. Bandwidth is fixed across both layers (numerator and denominator spatial densities). Basic visualization is available if `plot_gate = TRUE`.

Provides functionality for a correction for multiple testing. If p_correct = "FDR", calculates a False Discovery Rate by Benjamini and Hochberg. If p_correct = "uncorrelated Sidak", calculates an independent Sidak correction. If p_correct = "uncorrelated Bonferroni", calculates an independent Bonferroni correction. If p_correct = "correlated Sidak" or if p_correct = "correlated Bonferroni", then the corrections take into account the spatial correlation of the surface. (NOTE: If p_correct = "correlated Sidak" or if p_correct = "correlated Bonferroni", it may take a considerable amount of computation resources and time to calculate). If p_correct = "Adler and Hasofer" or if p_correct = "Friston", then calculates
a correction based on Random Field Theory. If p_correct = "none" (the default), then the function does not account for multiple testing and uses the uncorrected alpha level. See the internal pval_correct function documentation for more details.

The condition variable (Condition A) within dat must be of class 'factor' with two levels. The first level is considered the numerator (i.e., "case") value, and the second level is considered the denominator (i.e., "control") value. The level can also be specified using the c1n parameter.

Value

An object of class 'list' where each element is a object of class 'rrs' created by the risk function with two additional components:

rr An object of class 'im' with the relative risk surface.
f An object of class 'im' with the spatial density of the numerator.
g An object of class 'im' with the spatial density of the denominator.
P An object of class 'im' with the asymptotic p-value surface.
lrr An object of class 'im' with the log relative risk surface.
alpha A numeric value for the alpha level used within the gate.

Examples

test_rrs <- rrs(dat = randCyto)
Index

* datasets  
  randCyto, 10  
* package  
  gateR-package, 2  
  gateR (gateR-package), 2  
  gateR-package, 2  
  gating, 2, 3, 3, 7, 11  
  image.plot, 3  
  lotrrs, 3, 7  
  modified.ttest, 3, 5, 8, 12  
  randCyto, 3, 10  
  risk, 3, 5, 6, 8, 9, 12, 13  
  rrs, 3, 11  
  sparr, 2, 3  
  spatstat.geom, 3  
  terra, 3