Package 'Polytect'

June 25, 2024

Title An R package for digital data clustering

Version 0.99.0

Description An automatic clustering and labeling method for multi-color digital PCR data to classify partitions into groups based on subpopulations generated by flowPeaks.

biocViews ddPCR, Clustering

License MIT

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Depends R (>= 3.5.0)

Imports mvtnorm,
sn,
dplyr,
flowPeaks,
ggplot2,
tidyverse,
cowplot,
mlrMBO,
DiceKriging

Suggests testthat (>= 3.0.0),
knitr,
rmarkdown

VignetteBuilder knitr

R topics documented:

BPV *BPV data*

Description

A 3-color dPCR data of bovine papilloma virus assay

Usage

```
BPV
```

Format

A data frame of fluorescence intensities in each channel. Each row represents each partitions, and each column each channel.

channel1 fluorescence intensities of color 1

channel2 fluorescence intensities of color 2

channel3 fluorescence intensities of color 3

Examples

```
data(BPV)
```

CA *CA data*

Description

2-color competitive assay of competition BRAF V600E assay with 1% mutant

Usage

```
CA
```

Format

A data frame of fluorescence intensities in each channel. Each row represents each partitions, and each column each channel. data is not orthogonal.

channel1 fluorescence intensities of color 1

channel2 fluorescence intensities of color 2

Examples

```
data(CA)
```

conc_cal *concentration calculation function*

Description

This function takes a data frame of fluorescence intensities and partition clusters as input. It can be results from polytect_clust or polytect_merge. It will give the target concentration as output.

Usage

```
conc_cal(df_data, sampvol = 0.91, volmix = 20, voltemp = 20, type = "2color")
```

Arguments

df_data	A data frame containing partition fluorescence intensities and corresponding cluster label. This can be the output of polytect_merge or any data frame containing the above information.
sampvol	The sample volume in microliters (μL)
volmix	The volume of the mixture
voltemp	The volume of the template
type	The assay design, including the number of channels and targets. type=c("2color", "2colorHO","3color","4color"). "2color" is chosen when there are 2 colors and 2 targets. "2colorHO" means higher-order 2-color data (2 color and 3 targets). "3color" means 3-color and 3-target. "4-color" is chosen when there are 4 colors.

Value

a data frame of target concentration.

Examples

```
data(HR)
df_data<-polytect_clust(HR,4)
conc_cal(df_data)
```

HIV *HIV data*

Description

A 4-color dPCR data of intact HIV-1 proviruses

Usage

```
HIV
```

Format

A data frame of fluorescence intensities in each channel. Each row represents each partitions, and each column each channel.

channel1 fluorescence intensities of color 1

channel2 fluorescence intensities of color 2

channel3 fluorescence intensities of color 3

channel4 fluorescence intensities of color 4

Source

https://www.biorxiv.org/content/10.1101/2023.08.18.553846v1

Examples

```
data(HIV)
```

HR *HR data*

Description

A high-resolution 2-color dPCR data of RPP30 genomic DNA assay

Usage

```
HR
```

Format

A data frame of fluorescence intensities in each channel. Each row represents each partitions, and each column each channel. good separation but some crosstalk.

channel1 fluorescence intensities of color 1

channel2 fluorescence intensities of color 2

Source

https://pubmed.ncbi.nlm.nih.gov/33992770/

Examples

```
data(HR)
```

LR *LR data*

Description

A low-resolution 2-color dPCR data of development of genotyping assays for plants various

Usage

```
LR
```

Format

A data frame of fluorescence intensities in each channel. Each row represents each partitions, and each column each channel. barely separable on x-axis.

channel1 fluorescence intensities of color 1

channel2 fluorescence intensities of color 2

Examples

```
data(LR)
```

MM *MM data*

Description

A multi-mode 2-color dPCR data of HIV gBlock sequences

Usage

```
MM
```

Format

A data frame of fluorescence intensities in each channel. Each row represents each partitions, and each column each channel. obvious multimodality.

channel1 fluorescence intensities of color 1

channel2 fluorescence intensities of color 2

Source

https://pubmed.ncbi.nlm.nih.gov/37827643/

Examples

```
data(MM)
```

polytect_clust *Main function for clustering*

Description

This is the main function for clustering. The function will start with flowPeaks, then merge the excess clusters. It will return a data frame of fluorescence intensities and partition labels.

Usage

```
polytect_clust(
  data,
  cluster_num,
  type = "2color",
  fp_par = "default",
  fp_optim = c(0.1, 1, 1.5),
  lambdas = rep(2, 12),
  coefs = rep(1, 4)
)
```

Arguments

data	A matrix of fluorescence intensities in each channel. Each row represents each partitions, and each column each channel.
cluster_num	The expected maximum number of clusters.
type	The assay design, including the number of channels and targets. type=c("2color", "2colorHO","3color","4color"). "2color" is chosen when there are 2 colors and 2 targets. "2colorHO" means higher-order 2-color data (2 color and 3 targets). "3color" means 3-color and 3-target. "4-color" is chosen when there are 4 colors.
fp_par	The parameters for flowPeaks. fp_par=c("default","manual","auto"). When "default" is chosen, the default parameters of flowPeaks will be used. With "manual", you have to fill in fp_optim.
fp_optim	The paramters for flowPeaks that users have to fill in manually when fp_par is set at "manual".
lambdas	The penalty terms for the deviation from the expected cluster centers. Higher lambdas penalizes the deviation more.
coefs	The coefficients to adjust for the expected cluster centers. The default is 1 which can be used for common assay designs and has to be modified for special assays such as competing assays.

Value

A data frame containing the original fluorescence intensity and the cluster labels.

Examples

```
data(HR)
polytect_clust(HR, 4)
```

`polytect_merge` *Function for merging*

Description

This function takes the clustering result as input. Users can first perform any clustering algorithm, then use this function. It will return a data frame of fluorescence intensities and partition labels.

Usage

```
polytect_merge(
  data,
  cluster_num,
  base_clust,
  type = "2color",
  lambdas = rep(2, 12),
  coefs = rep(1, 4)
)
```

Arguments

`data`	A matrix of fluorescence intensities in each channel. Each row represents each partitions, and each column each channel.
`cluster_num`	The expected maximum number of clusters.
`base_clust`	A list that contains partition labels given by initial clustering.
`type`	The assay design, including the number of channels and targets. `type=c("2color", "2colorHO","3color","4color")`. "2color" is chosen when there are 2 colors and 2 targets. "2colorHO" means higher-order 2-color data (2 color and 3 targets). "3color" means 3-color and 3-target. "4-color" is chosen when there are 4 colors.

Value

A data frame containing the original fluorescence intensity and the cluster labels.

Examples

```
data(HR)
dist_matrix <- dist(HR)
hc <- hclust(dist_matrix, method = "ward.D2")
hc_clusters <- cutree(hc, k = 6)
polytect_merge(HR, 4, hc_clusters)
```

polytect_plot *Plotting function for clustering results*

Description

This function takes results from polytect_clust and polytect_merge, or a data frame containing flurescence intensities and partition labels. It will output all combination of 2-color plots.

Usage

```
polytect_plot(df_data)
```

Arguments

df_data	A data frame containing partition fluorescence intensities and corresponding cluster label. This can be the output of polytect_clust and polytect_merge or any data frame containing the above information.

Value

2-color plots.

Examples

```
data(HR)
df_data<-polytect_clust(HR,4)
polytect_plot(df_data)
```

polytect_summary *Summary function*

Description

This function takes results from polytect_clust and polytect_merge, or a data frame containing flurescence intensities and partition labels. It will summarise cluster centers, cluster sizes and cluster silhouette coefficients.

Usage

```
polytect_summary(df_data)
```

Arguments

df_data	A data frame containing partition fluorescence intensities and corresponding cluster label. This can be the output of polytect_clust and polytect_merge or any data frame containing the above information.

Value

a data frame of the summary of cluster centers, cluster sizes and cluster silhouette coefficients.

Examples

```
data(HR)
df_data<-polytect_clust(HR,4)
polytect_summary(df_data)
```

`sil_plot` *Plotting function for silhouette coefficients*

Description

This function takes results from polytect_clust and polytect_merge, or a data frame containing flurescence intensities and partition labels. It will output the silhouette coefficients of each cluster.

Usage

```
sil_plot(df_data)
```

Arguments

`df_data`	A data frame containing partition fluorescence intensities and corresponding cluster label. This can be the output of `polytect_clust` and `polytect_merge` or any data frame containing the above information.

Value

plot of silhouette coefficients for each cluster.

Examples

```
data(HR)
df_data<-polytect_clust(HR,4)
sil_plot(df_data)
```

Index