Package 'ClusTCR2'

January 20, 2025

Title Identifying Similar T Cell Receptor Hyper-Variable Sequences with 'ClusTCR2'

Version 1.7.3.01

Date 2024-05-15

Author Kerry A. Mullan [aut, cre],
Sebastiaan Valkiers [aut, ctb],
Kris Laukens [aut, ctb],
Pieter Meysman [aut, ctb]

Description Enhancing T cell receptor (TCR) sequence analysis, 'ClusTCR2', based on 'ClusTCR' python program, leverages Hamming distance to compare the complement-determining region three (CDR3) sequences for sequence similarity, variable gene (V gene) and length. The second step employs the Markov Cluster Algorithm to identify clusters within an undirected graph, providing a summary of amino acid motifs and matrix for generating network plots. Tailored for single-cell RNA-seq data with integrated TCR-seq information, 'ClusTCR2' is integrated into the Single Cell TCR and Expression Grouped Ontologies (STEGO) R application or 'STEGO.R'. See the two publications for more details. Sebastiaan Valkiers, Max Van Houcke, Kris Laukens, Pieter Meysman (2021) <doi:10.1093/bioinformatics/btab446>, Kerry A. Mullan, My Ha, Sebastiaan Valkiers, Nicky de Vrij, Benson Ogunjimi, Kris Laukens, Pieter Meysman (2023) <doi:10.1101/2023.09.27.559702>.

Maintainer Kerry A. Mullan <Kerry.Mullan@uantwerpen.be>

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.1

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

Imports DescTools, ggplot2, ggseqlogo, network, plyr, RColorBrewer, stringr, scales, sna, VLF

biocViews GeneTarget, SingleCell

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2024-05-16 15:00:19 UTC

Contents

`ClusTCR` *Creates ClusTCR matrix This function identifies similar CDR3 amino acid sequences based on the same length and V_gene*

Description

Creates ClusTCR matrix This function identifies similar CDR3 amino acid sequences based on the same length and V_gene

Usage

```
ClusTCR(my_file, allele = NULL, v_gene = "v_call")
```

Arguments

`my_file`	uploaded file with junction_aa (CD3 sequences), variable gene.
`allele`	The allele, if present as *00 will be removed if the user requires it.
`v_gene`	Variable gene column name

Value

X by Y matrix of structurally related CDR3 sequences.

Examples

```
# Example usage of ClusTCR function with a stored file
example_file <- read.csv(system.file("extdata", "my_data.csv", package = "ClusTCR2"))
# Perform clustering using ClusTCR function
step1 <- ClusTCR(example_file, allele = FALSE)
# Print the result
print(step1)
```

`ClusTCR_Large` *Creates ClusTCR matrix This function identifies similar CDR3 amino acid sequences based on the same length and V_gene*

Description

Creates ClusTCR matrix This function identifies similar CDR3 amino acid sequences based on the same length and V_gene

Usage

```
ClusTCR_Large(my_file, allele = NULL, v_gene = "v_call")
```

Arguments

`my_file`	uploaded file with junction_aa (CD3 sequences), variable gene.
`allele`	The allele, if present as *00 will be removed if the user requires it.
`v_gene`	Variable gene column name

Value

X by Y matrix of structurally related CDR3 sequences.

`ggnet2` *Copied code from ggnet's ggnet2 function*

Description

Copied code from ggnet's ggnet2 function

Usage

```
ggnet2(
  net,
  mode = "fruchtermanreingold",
  layout.par = NULL,
  layout.exp = 0,
  alpha = 1,
  color = "grey75",
  shape = 19,
  size = 9,
  max_size = 9,
  na.rm = NA,
  palette = NULL,
  alpha.palette = NULL,
```

```
  alpha.legend = NA,
  color.palette = palette,
  color.legend = NA,
  shape.palette = NULL,
  shape.legend = NA,
  size.palette = NULL,
  size.legend = NA,
  size.zero = FALSE,
  size.cut = FALSE,
  size.min = NA,
  size.max = NA,
  label = FALSE,
  label.alpha = 1,
  label.color = "black",
  label.size = max_size/2,
  label.trim = FALSE,
  node.alpha = alpha,
  node.color = color,
  node.label = label,
  node.shape = shape,
  node.size = size,
  edge.alpha = 1,
  edge.color = "grey50",
  edge.lty = "solid",
  edge.size = 0.25,
  edge.label = NULL,
  edge.label.alpha = 1,
  edge.label.color = label.color,
  edge.label.fill = "white",
  edge.label.size = max_size/2,
  arrow.size = 0,
  arrow.gap = 0,
  arrow.type = "closed",
  legend.size = 9,
  legend.position = "right",
  ...
)
```

Arguments

net	net plot from step 2.
mode	= "fruchtermanreingold"
layout.par	= NULL,
layout.exp	= 0
alpha	= 1
color	= "grey75"
shape	= 19

`size` = 9

`max_size` = 9

`na.rm` = NA

`palette` = NULL

`alpha.palette` = NULL

`alpha.legend` = NA

`color.palette` = palette

`color.legend` = NA

`shape.palette` = NULL

`shape.legend` = NA

`size.palette` = NULL

`size.legend` = NA

`size.zero` = FALSE

`size.cut` = FALSE

`size.min` = NA

`size.max` = NA

`label` = FALSE

`label.alpha` = 1

`label.color` = "black"

`label.size` = max_size/2

`label.trim` = FALSE

`node.alpha` see `alpha`

`node.color` see `color`

`node.label` see `label`

`node.shape` see `shape`

`node.size` see `size`

`edge.alpha` = 1

`edge.color` the color of the edges, as a color value, a vector of color values, or as an edge attribute containing color values. Defaults to `"grey50"`.

`edge.lty` = "solid"

`edge.size` = 0.25

`edge.label` = NULL

`edge.label.alpha` = 1

`edge.label.color` = label.color

`edge.label.fill` = "white"

edge.label.size
= max_size/2
arrow.size = 0
arrow.gap = 0
arrow.type = "closed"
legend.size = 9
legend.position
= "right"
... Other functions in ggplot2

Value

A ggplot object displaying the network plot.

`mcl_cluster` *Create the files for labeling the linked clusters from ClusTCR_list_to_matrix function*

Description

Create the files for labeling the linked clusters from ClusTCR_list_to_matrix function

Usage

```
mcl_cluster(my_file, max.iter = 10, inflation = 1, expansion = 1)
```

Arguments

my_file	Matrix file produce from ClusTCR
max.iter	Number of iterations to find the steady state of MCL.
inflation	numeric value
expansion	numeric value

Value

A list containing two elements:

- 'Cluster_lab': Data frame containing information about the clusters
- 'Normalised_tabel': Normalized table used in the clustering process

Examples

```
# Example usage of mcl_cluster function with a stored file
example_file <- read.csv(system.file("extdata", "my_data.csv",package = "ClusTCR2"))
# Perform clustering using mcl_cluster function
step1 <- ClusTCR(example_file,allele = FALSE)
# perform mcl
step2 <- mcl_cluster(step1)
```

`mcl_cluster_large` *Create the files for labeling the linked clusters from ClusTCR_list_to_matrix function*

Description

Create the files for labeling the linked clusters from ClusTCR_list_to_matrix function

Usage

```
mcl_cluster_large(my_file, max.iter = 10, inflation = 1, expansion = 1)
```

Arguments

`my_file`	Matrix file produce from ClusTCR
`max.iter`	Number of iterations to find the steady state of MCL.
`inflation`	numeric value
`expansion`	numeric value

Value

A list containing two elements:

- 'Cluster_lab': Data frame containing information about the clusters
- 'Normalised_tabel': Normalized table used in the clustering process

`Motif_from_cluster_file` *Code for plotting the Motif based on a specific CDR3 length and V gene (see netplot_ClusTCR2 for details).*

Description

Code for plotting the Motif based on a specific CDR3 length and V gene (see netplot_ClusTCR2 for details).

Usage

```
Motif_from_cluster_file(
  ClusTCR,
  Clust_selected = NULL,
  selected_cluster_column = "Clust_size_order"
)
```

Arguments

ClusTCR	Cluster file produced from mcl_cluster.
Clust_selected	Select which cluster to review.
selected_cluster_column	Select the column "Clust_size_order" of the cluster ordered.

Value

A ggplot object representing the motif.

---

motif_plot *Code for plotting the Motif based on a specific CDR3 length and V gene (see netplot_ClusTCR2 for).*

Description

Code for plotting the Motif based on a specific CDR3 length and V gene (see netplot_ClusTCR2 for).

Usage

```
motif_plot(
  ClusTCR,
  Clust_column_name = "Clust_size_order",
  Clust_selected = NULL
)
```

Arguments

ClusTCR	Matrix file produce from mcl_cluster
Clust_column_name	Name of clustering column from mcl_cluster file e.g. cluster
Clust_selected	Select which cluster to display. Only one at a time.

Value

A ggplot object representing the motif.

Examples

```
# Example usage of mcl_cluster function with a stored file
example_file <- read.csv(system.file("extdata", "my_data.csv",package = "ClusTCR2"))
# Perform clustering using mcl_cluster function
step1 <- ClusTCR(example_file,allele = FALSE)
# perform mcl
step2 <- mcl_cluster(step1)
# print the motif plot for the simple clustering
print(motif_plot(step2,Clust_selected = 1))
```

`motif_plot_large` *Code for plotting the Motif based on a specific CDR3 length and V gene (see netplot_ClusTCR2 for details).*

Description

Code for plotting the Motif based on a specific CDR3 length and V gene (see netplot_ClusTCR2 for details).

Usage

```
motif_plot_large(
  ClusTCRFile_large,
  Clust_column_name = "Clust_size_order",
  Clust_selected = NULL
)
```

Arguments

`ClusTCRFile_large`	Matrix file produced from mcl_cluster_large.
`Clust_column_name`	Name of clustering column from mcl_cluster file e.g. cluster.
`Clust_selected`	Select which cluster to display. Only one at a time.

Value

A ggplot object representing the motif.

`netplot_ClusTCR2` *Code for displaying the network.*

Description

Code for displaying the network.

Usage

```
netplot_ClusTCR2(
  ClusTCR,
  filter_plot = 0,
  Clust_selected = 1,
  selected_col = "purple",
  selected_text_col = "black",
  selected_text_size = 3,
```

```
  non_selected_text_size = 2,
  Clust_column_name = "cluster",
  label = c("Name", "cluster", "CDR3", "V_gene", "Len"),
  non_selected_col = "grey80",
  non_selected_text_col = "grey40",
  alpha_selected = 1,
  alpha_non_selected = 0.5,
  colour = "color_test",
  all.colour = "default"
)
```

Arguments

ClusTCR	File produced from mcl_cluster
filter_plot	Filter's plot to remove connects grater than # e.g. 2 = 3 or more connections.
Clust_selected	Select which cluster to label.
selected_col	Color of selected cluster (Default = purple)
selected_text_col	Color of selected cluster text (Default = black)
selected_text_size	Text size of selected cluster (Default = 3)
non_selected_text_size	Text size of non-selected clusters (Default = 2)
Clust_column_name	Name of clustering column from mcl_cluster file e.g. cluster (Re-numbering the original_cluster), Original_cluster, Clust_size_order (Based on cluster size e.g. number of nodes)
label	Name to display on cluster: Name (CDR3_V_gene_Cluster), cluster, CDR3, V_gene, Len (length of CDR3 sequence), CDR3_selected, V_gene_selected, Name_selected,cluster_selected, (_selected only prints names of the chosen cluster), None
non_selected_col	Color of selected cluster (Default = grey80)
non_selected_text_col	Color of selected clusters text (Default = grey40)
alpha_selected	Transparency of selected cluster (default = 1)
alpha_non_selected	Transparency of non-selected clusters (default = 0.5)
colour	Colour selected = "color_test" or all = "color_all"
all.colour	Colours all points by: rainbow, random, heat.colors, terrain.colors, topo.colors, hcl.colors and default

Value

A ggplot object displaying the network plot.

Examples

```
# Example usage of mcl_cluster function with a stored file
example_file <- read.csv(system.file("extdata", "my_data.csv",package = "ClusTCR2"))
# Perform clustering using mcl_cluster function
step1 <- ClusTCR(example_file,allele = FALSE)
# perform mcl
step2 <- mcl_cluster(step1)
# print the clustering plot after performing step 1 and step 2
print(netplot_ClusTCR2(step2, label = "Name_selected"))
```

Index