

ClusTCR2

This package is for performing the single cell RNA-seq T cell receptor (TCR) sequence similarity of the hypervariable CDR3 sequence in the TCR. This package was based on the python software ClusTCR. This uses an edit distance (hamming distance) of one and of the same v_gene.

```
library(ClusTCR2)
```

Step 1. Example of how to use ClusTCR = hamming distance and creating the matrix.

```
# 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)
#> creating empty matrices
#> Performing edit distance
#> keeping edit distance of 1
#> Creating target and source object
#> Creating matrix for MCL
#> Matrix complete
# Print the result
step1[1:6,1:6]
#>
#>
#> CAAFNQAGTALIF_TRAV2 CAAGGSARQLTF_TRAV23DV6 CAAGPQGGSEKLVF_TRAV29DV5 CAAGY
#> CAAFNQAGTALIF_TRAV2 0 0 0
#> CAAGGSARQLTF_TRAV23DV6 0 0 0
#> CAAGPQGGSEKLVF_TRAV29DV5 0 0 0
#> CAAGYNFNKFYF_TRAV22 0 0 0
#> CAAHNNNDMRF_TRAV13-1 0 0 0
#> CAAHSGAGSYQLTF_TRAV29DV5 0 0 0
```

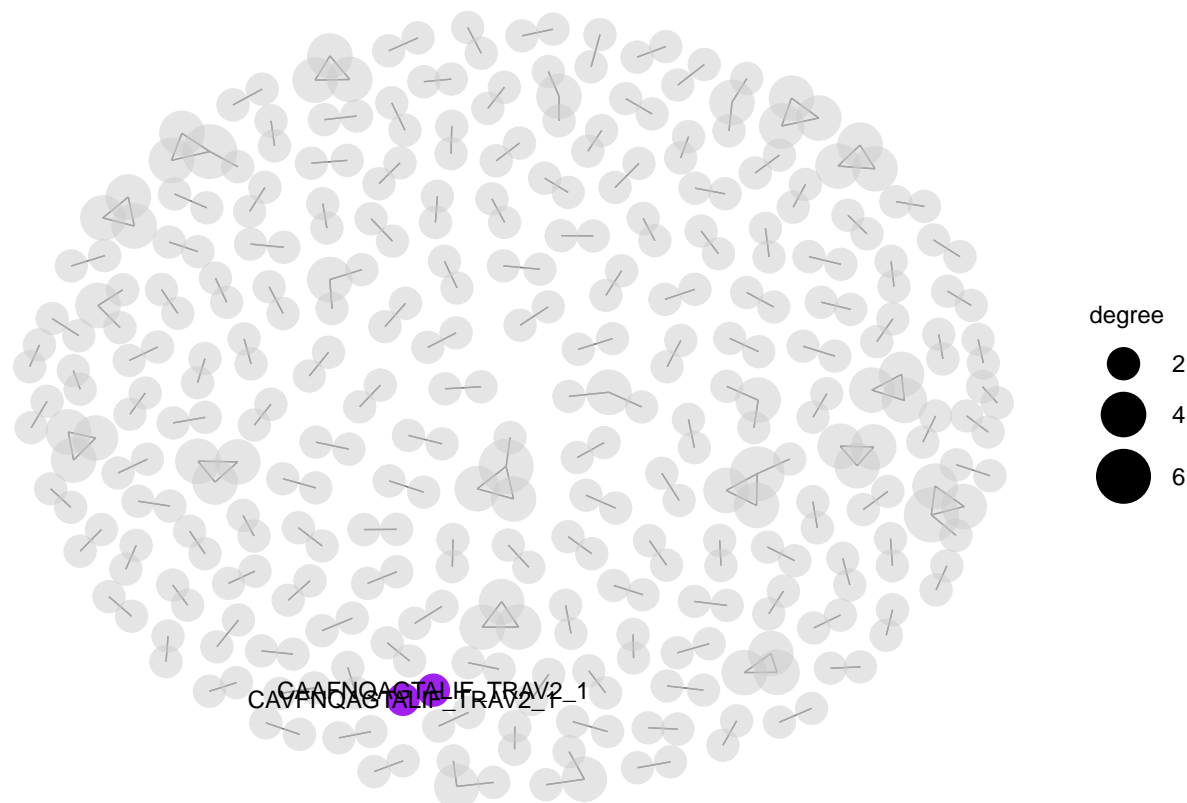
Step 2. Creating both the file need for visulising the motif and network from step 1.

```
# Example usage of mcl_cluster function with a stored file
step2 <- mcl_cluster(step1)
#> Iteration complete: 2
#> Inflation complete
#> MCL complete
#> Finished correcting matrix to binary
#> relabelling nodes for MCL start.
#> iteration 1 completed in: 0.01715087890625
#> iteration 2 completed in: 0.0105979442596436
#> iteration 3 completed in: 0.0159671306610107
#> iteration 4 completed in: 0.0104598999023438
#> iteration 5 completed in: 0.0155000686645508
#> iteration 6 completed in: 0.0103540420532227
#> this process took 0.37 seconds
#> Completed process, can now display data as either motif or netplot.
step2[[1]][1:3,1:3]
#> Original_cluster nodes #_of_connections
```

```
#> 1          1 1 240          2
#> 129        2 2 76          2
#> 236        3 3 7          2
head(step2[[2]][1:6,1:6])
#>
#>          CAAFNQAGTALIF_TRAV2 CAAGGSARQLTF_TRAV23DV6 CAAGPQGGSEKLVF_TRAV29DV5 CAAGY
#> CAAFNQAGTALIF_TRAV2          0.5          0.0          0.0
#> CAAGGSARQLTF_TRAV23DV6        0.0          0.5          0.0
#> CAAGPQGGSEKLVF_TRAV29DV5      0.0          0.0          0.5
#> CAAGYNFNKFYF_TRAV22          0.0          0.0          0.0
#> CAAHNNNDMRF_TRAV13-1        0.0          0.0          0.0
#> CAAHSGAGSYQLTF_TRAV29DV5    0.0          0.0          0.0
```

Visualization: Network plot

```
# Visualization of the network plot
netplot_ClusTCR2(step2, label = "Name_selected")
```



Visualization: Motif plot

```
# Visualization of the network plot
# step2[[1]]
subset(step2[[1]], step2[[1]]$Clust_size_order == 1)
#>   Original_cluster   nodes #_of_connections order cluster count Clust_size_order CD
#> 59          14      14 67          2      14      12      4      1 CAAPSGNTPLVF_T
#> 58          14     36 67 71          3     36      12      4      1 CAASEGNTPLVF_T
#> 60          14 14 36 67 71          4     67      12      4      1 CAASSGNTPLVF_T
#> 61          14     36 67 71          3     71      12      4      1 CAASTGNTPLVF_T
motif_plot(step2, Clust_selected = 1)
```

