Package 'scoper'

October 14, 2018

Type Package

Version 0.1.0

Date 2018-10-04

Title Spectral Clustering-Based Method for Identifying B Cell Clones

Description Provides a computational framework for unsupervised identification B cell clones from adaptive immune receptor repertoire sequencing (AIRR-Seq) datasets. This method is based on spectral clustering of the junction sequences of B cell receptors (BCRs, also referred to as Immunoglobulins, (Igs)) that share the same V gene, J gene and junction length.

License CC BY-SA 4.0

URL https://scoper.readthedocs.io

BugReports https://bitbucket.org/kleinstein/scoper/issues

LazyData true

BuildVignettes true

VignetteBuilder knitr

Encoding UTF-8

Depends R (>= 3.1.2), ggplot2 (>= 2.0.0)

Imports alakazam (>= 0.2.11), shazam (>= 0.1.10), doParallel, foreach, dplyr (>= 0.7.0), stringi, methods, stats, iterators, lazyeval

Suggests knitr, rmarkdown, testthat

RoxygenNote 6.1.0

Collate 'Data.R' 'Scoper.R' 'Functions.R'

NeedsCompilation no

Author Nima Nouri [aut, cre],
Jason Vander Heiden [ctb],
Steven Kleinstein [aut, cph]

Maintainer Nima Nouri <nima.nouri@yale.edu>

Repository CRAN

Date/Publication 2018-10-14 17:00:03 UTC

R topics documented:

`analyzeClones` *Clonal assignment analysis*

Description

The `analyzeClones` function performs a series of analysis to assess the performance of `defineClonesScoper` function.

Usage

```
analyzeClones(db, junction = "JUNCTION", v_call = "V_CALL",
  j_call = "J_CALL", clone = "CLONE", first = FALSE, cdr3 = FALSE,
  nproc = 1, progress = FALSE)
```

Arguments

`db`	data.frame with Change-O style columns containing sequence data.
`junction`	name of the column containing nucleotide sequences to compare. Also used to determine sequence length for grouping.
`v_call`	name of the column containing the V-segment allele calls.
`j_call`	name of the column containing the J-segment allele calls.
`clone`	name of the data column containing clone identifiers.
`first`	if `TRUE` only the first call of the gene assignments is used. if `FALSE` the union of ambiguous gene assignments is used to group all sequences with any overlapping gene calls.
`cdr3`	if `TRUE` remove 3 nts from both ends of `junction` (converts IMGT junction to CDR3 region).
`nproc`	number of cores to distribute the function over.
`progress`	if `TRUE` print a progress bar.

Value

Returns a ClonalAnalysis object.

Note

Arguments `first` and `cdr3` must match the corresponding arguments used in the defineClonesScoper function.

Examples

```
# Clonal assignment analysis
results <- analyzeClones(ClonedExampleDb, junction = "JUNCTION", v_call = "V_CALL",
                         j_call = "J_CALL", clone = "CLONE", first = TRUE)
# print threshold (a numeric)
results@threshold

# get inter and intra conal distances (a data.frame)
df <- results@inter_intra[[1]]

# density plot of inter versus intra clonal distances  (a ggplot).
results@plot_inter_intra

# get the neighborhoods used in spectral clustering (a numeric vector).
ngs <- results@neighborhoods

# plot histogram of neighborhoods (a ggplot).
results@plot_neighborhoods
```

`ClonalAnalysis-class` *Output of the analyzeClones function*

Description

`ClonalAnalysis` contains output from the analyzeClones function. It includes infromation to interpret clonal assignment performance.

Slots

`threshold` cut-off separating the inter (within) and intra (between) clonal distances.

`inter_intra` data.frame containing all inter and intra clonal distances.

`plot_inter_intra` density plot of inter versus intra clonal distances. The threshold is shown with a horizental dashed-line.

`neighborhoods` a numeric vector containing scale parameters used in spectral clustering process.

`plot_neighborhoods` histogram of neighborhoods. The threshold is shown with a vertical dashed-line.

See Also

analyzeClones

defineClonesScoper *Assigning Ig sequences into clonal groups*

Description

The defineClonesScoper function provides an unsupervised pipline for assigning Ig sequences into clonal groups sharing same V gene, J gene, and junction length.

Usage

```
defineClonesScoper(db, junction = "JUNCTION", v_call = "V_CALL",
  j_call = "J_CALL", first = FALSE, cdr3 = FALSE, mod3 = FALSE,
  iter_max = 1000, nstart = 25, nproc = 1, progress = FALSE,
  out_name = NULL, out_dir = ".")
```

Arguments

db	data.frame with Change-O style columns containing sequence data.
junction	name of the column containing nucleotide sequences to compare. Also used to determine sequence length for grouping.
v_call	name of the column containing the V-segment allele calls.
j_call	name of the column containing the J-segment allele calls.
first	if TRUE only the first call of the gene assignments is used. if FALSE the union of ambiguous gene assignments is used to group all sequences with any overlapping gene calls.
cdr3	if TRUE remove 3 nts from both ends of junction (converts IMGT junction to CDR3 region). if TRUE remove junction(s) with length less than 7 nts.
mod3	if TRUE remove junction(s) with number of nucleotides not modulus of 3.
iter_max	the maximum number of iterations allowed for kmean clustering step.
nstart	the number of random sets chosen for kmean clustering initialization.
nproc	number of cores to distribute the function over.
progress	if TRUE print a progress bar.
out_name	if not NULL save cloned data.frame and a summary of cloning performance. out_name string is used as the prefix of the successfully processed output files.
out_dir	specify to change the output directory. The input file directory is used if this is not specified while out_name is specified.

Details

An unsupervised pipeline to identify B cell clones from adaptive immune receptor repertoire sequencing (AIRR-Seq) datasets. This method is based on spectral clustering of the junction sequences of B cell receptors (BCRs, also referred to as Immunoglobulins, (Igs)) that share the same V gene, J gene and junction length. It uses an adaptive threshold that analyzes sequences in a local neighborhood.

Value

Returns a modified db data.frame with clone identifiers in the CLONE column. if out_name is not NULL, it will save the modified db and a summary of cloning performance in the current directory or the specified out_dir.

See Also

To assess the performance of clonal assignment process check analyzeClones.

Examples

```
# clone data using defineClonesScoper function
db <- defineClonesScoper(ExampleDb, junction = "JUNCTION", v_call = "V_CALL",
                         j_call = "J_CALL", first = TRUE)
```

ExampleDb *Example Change-O database*

Description

A small example database subset from Laserson and Vigneault et al, 2014.

Includes the CLONE column.

Usage

```
ExampleDb

ClonedExampleDb
```

Format

A data.frame with the following Change-O style columns:

- SEQUENCE_ID: Sequence identifier
- SEQUENCE_IMGT: IMGT-gapped observed sequence.
- GERMLINE_IMGT_D_MASK: IMGT-gapped germline sequence with N, P and D regions masked.
- V_CALL: V region allele assignments.
- V_CALL_GENOTYPED: TIgGER corrected V region allele assignment.
- D_CALL: D region allele assignments.
- J_CALL: J region allele assignments.
- JUNCTION: Junction region sequence.
- JUNCTION_LENGTH: Length of the junction region in nucleotides.
- NP1_LENGTH: Combined length of the N and P regions proximal to the V region.
- NP2_LENGTH: Combined length of the N and P regions proximal to the J region.
- SAMPLE: Sample identifier. Time in relation to vaccination.
- ISOTYPE: Isotype assignment.
- DUPCOUNT: Copy count (number of duplicates) of the sequence.

References

1. Laserson U and Vigneault F, et al. High-resolution antibody dynamics of vaccine-induced immune responses. Proc Natl Acad Sci USA. 2014 111:4928-33.

`scoper` *The SCOPer package*

Description

Provides a computational framework for unsupervised identification B cell clones from adaptive immune receptor repertoire sequencing (AIRR-Seq) datasets. This method is based on spectral clustering of the junction sequences of B cell receptors (BCRs, Immunoglobulins) that share the same V gene, J gene and junction length.

Spectral Clustering for clOne Partitioning (SCOPer)

- defineClonesScoper: Clustering sequences into clonal groups.
- analyzeClones: Summary statistics and visualization of the clonal clustering results.

References

1. Nouri N and Kleinstein SH (2018). A spectral clustering-based method for identifying clones from high-throughput B cell repertoire sequencing data. Bioinformatics, 34(13):i341-i349.

Index