Package 'ECLIPSE'

November 20, 2026

Title Enhanced CLonotypic Inference via Prediction of Single-cell Expression

Version 0.99.3

Description
This package is used for the analysis of single-cell T cell receptor sequencing data (scTCR-seq). ECLIPSE maps chain pairings across a sample to predict the true clonotype of each cell, unlike other methods which only consider the chains that are sequenced in each cell. Seurat/SingleCellExperiment objects and Cell Ranger vdj/multi contig files are required as inputs.

License GPL (>= 3)

Encoding UTF-8

Roxygen list(markdown = TRUE)

Imports dplyr, ggplot2, scRepertoire, stringr, tidyr, VDJdive, methods, rlang, tidyselect, magrittr, utils

URL https://github.com/eburns10/ECLIPSE

biocViews Software, GeneExpression, BiomedicalInformatics, ImmunoOncology, Transcriptomics, Classification, SingleCell, Annotation, Sequencing

BugReports https://github.com/eburns10/ECLIPSE/issues

Depends R (>= 4.4.0)

Config/roxygen2/version 8.0.0

Suggests knitr, rmarkdown, ragg, tidyverse, testthat, BiocStyle, Seurat, SingleCellExperiment, S4Vectors, SummarizedExperiment, SeuratObject, tibble

VignetteBuilder knitr

RoxygenNote 8.0.0

git_url https://git.bioconductor.org/packages/ECLIPSE

git_branch devel

git_last_commit 9ca87aa

git_last_commit_date 2026-11-10

Repository Bioconductor 3.24

Date/Publication 2026-11-19

Author Ethan Burns [aut, cre] (ORCID: <https://orcid.org/0009-0009-8061-3686>),
Zhaochen Ye [ctb] (ORCID: <https://orcid.org/0009-0000-4422-3861>),
Kelly Street [ctb] (ORCID: <https://orcid.org/0000-0001-6379-5013>),
David Braun [aut, fnd] (ORCID: <https://orcid.org/0000-0003-4543-5553>)

Maintainer Ethan Burns <ethan.burns@yale.edu>

Contents

ECLIPSE-package . . . 2
.assignFinalClonotypes . . . 3
.findMissingMonotypicChains . . . 4
.identifyThreeChainClones . . . 4
.modifyThreeChainContigs . . . 5
.qcFilterContigs . . . 6
.readContigs . . . 7
.reformatSeuratObject . . . 7
.runVDJdive . . . 8
ECLIPSE . . . 9
findThreeChainClones . . . 11
post_qc_contigs . . . 11
sceECLIPSE . . . 12
tcrArtifactDetect . . . 14
test_seurat . . . 15
test_seurat_with_tcr . . . 15
unit_test_object . . . 16
vignette_object . . . 16

Index **17**

`ECLIPSE-package` *ECLIPSE: Enhanced CLonotypic Inference via Prediction of Single-cell Expression*

Description

This package is used for the analysis of single-cell T cell receptor sequencing data (scTCR-seq). ECLIPSE maps chain pairings across a sample to predict the true clonotype of each cell, unlike other methods which only consider the chains that are sequenced in each cell. Seurat/SingleCellExperiment objects and Cell Ranger vdj/multi contig files are required as inputs.

Author(s)

Maintainer: Ethan Burns <ethan.burns@yale.edu> (ORCID)

Authors:

- Ethan Burns <ethan.burns@yale.edu> (ORCID)
- David Braun <david.braun@yale.edu> (ORCID) [funder]

Other contributors:

- Zhaochen Ye <zhaochen.ye@yale.edu> (ORCID) [contributor]
- Kelly Street <kelly.street@gmail.com> (ORCID) [contributor]

See Also

Useful links:

- https://github.com/eburns10/ECLIPSE
- Report bugs at https://github.com/eburns10/ECLIPSE/issues

`.assignFinalClonotypes` *Assign final clonotypes*

Description

This internal function determines and adds the final EM-based clonotype to the Seurat object. It also adds the chains that were detected and how the cell would have been called with standard analysis. It is a helper function for ECLIPSE.

Usage

```
.assignFinalClonotypes(
  em_results,
  Contigs,
  Seurat_object,
  Orphan_contigs,
  Donor,
  Group,
  CombineTcrArgs,
  Verbose
)
```

Arguments

`em_results`	the results from running .runVDJdive.
`Contigs`	post-QC filtered contigs.
`Seurat_object`	Seurat object that contains scRNA-seq data.
`Orphan_contigs`	orphan TCR contigs determined by .qcFilterContigs.
`Donor`	name of column in Seurat meta.data. The column should contain the donor, regardless of treatment/timepoint/condition.
`Group`	name of column in Seurat meta.data. This column should specify how the final statistics for each clone should be calculated. For example, should they be calculated per sample, per donor, or among all cells? This can and often will be the same column as donor.
`CombineTcrArgs`	arguments that are provided to scRepertoire combineTCR. This should be a list of 4 logical elements, with the elements corresponding to the combineTCR arguments. 1 is removeNA, 2 is removeMulti, 3 is filterMulti, 4 is filterNonproductive. If no clonotype can be predicted with high confidence, then combineTCR using these arguments will determine the final clonotype.
`Verbose`	specifies whether to print an important QC graph, a small table, and also notes through the pipeline.

Value

A Seurat object with the clonotype information added.

`.findMissingMonotypicChains`

Find missing monotypic chains

Description

This internal function finds cells that may be in clones with 3 TCR chains and have no monotypic chain detected. If likely, it assigns them to these clones. It is a helper function for findThreeChainClones.

Usage

```
.findMissingMonotypicChains(Contigs, specialTCRs, Chain, Mode, print_stats)
```

Arguments

`Contigs`	modified TCR contigs from .modifyThreeChainContigs.
`specialTCRs`	a vector of bitypic chains.
`Chain`	"alpha" or "beta". This tells it whether clones with two of which chain should be looked for.
`Mode`	modified contigs are returned with standard. If "show" is provided, two tables are returned. The first shows the final set of clones with a double chain of one type, and the second shows how/why some of those clones were called as such. The second is pre-filtering, so some clones may not have been called as having 3 total chains in the end.
`print_stats`	specifies whether notes/stats should be printed.

Value

a tibble of modified TCR contigs.

`.identifyThreeChainClones`

Identify three TCR chain clones

Description

This internal function finds which T cell clones likely have 3 total TCR chains. It is a helper function for findThreeChainClones.

Usage

```
.identifyThreeChainClones(Contigs, Chain, Mode)
```

Arguments

Contigs	tibble of contigs from cellranger all or filtered contig files.
Chain	"alpha" or "beta". This tells it whether clones with two of which chain should be looked for.
Mode	standard leads to modified contigs being returned. If "show" is provided, two tables are returned. The first shows the final set of clones with a double chain of one type, and the second shows how/why some of those clones were called as such. The second is pre-filtering, so some clones may not have been called as having 3 total chains in the end.

Value

A list of first the clones that have 3 TCR chains, then the clonotypes of select cells, then all bitypic chains. If show is provided to Mode, then a list of two tables are returned showing the clones with 3 TCR chains.

.modifyThreeChainContigs

Modify three TCR chain contigs

Description

This internal function takes the clones that have 3 TCR chains and modifies their contigs so that these clones can be retained as having 3 TCR chains downstream. It is a helper function for findThreeChainClones.

Usage

```
.modifyThreeChainContigs(Contigs, three_chain_clones, Clones, Chain, Mode)
```

Arguments

Contigs	tibble of TCR contigs.
three_chain_clones	a tibble of the three chain clones.
Clones	the clonotypes of select cells made by .identifyThreeChainClones.
Chain	"alpha" or "beta". This tells it whether clones with two of which chain should be looked for.
Mode	modified contigs are returned with standard. If "show" is provided, two tables are returned. The first shows the final set of clones with a double chain of one type, and the second shows how/why some of those clones were called as such. The second is pre-filtering, so some clones may not have been called as having 3 total chains in the end.

Value

a tibble of modified TCR contigs.

`.qcFilterContigs` *QC filter contigs*

Description

This internal function filters contigs that do not pass QC. It is a helper function for ECLIPSE.

Usage

```
.qcFilterContigs(
  Contigs,
  Seurat_object,
  Batch,
  Donor,
  Group,
  Original_barcode,
  Format,
  ReturnFilteredContigs,
  Verbose
)
```

Arguments

`Contigs`	the TCR contigs data.frame made by .readContigs.
`Seurat_object`	Seurat object that contains scRNA-seq data.
`Batch`	name of column in Seurat meta.data. The column should contain integers, not characters, that tell which position in the contig_folder/file_path vectors contains the TCR contigs for the cell.
`Donor`	name of column in Seurat meta.data. The column should contain the donor, regardless of treatment/timepoint/condition.
`Group`	name of column in Seurat meta.data. This column should specify how the final statistics for each clone should be calculated. For example, should they be calculated per sample, per donor, or among all cells? This can and often will be the same column as donor.
`Original_barcode`	name of column in Seurat meta.data. This must contain the original 10X barcode for the cell. Typically this will end with -1, but users should check their TCR contig files to be sure.
`Format`	This tells how missing data is stored in the TCR contig files. Current files likely contain "", but if older contig files are used where any missing data is stored as "None", then "None" should be provided here.
`ReturnFilteredContigs`	specifies if filtered contigs should be returned instead of running the whole ECLIPSE pipeline.
`Verbose`	specifies whether to print an important QC graph, a small table, and also notes through the pipeline.

Value

A list of 3, with the filtered contigs first, the seurat object, then contigs that are only seen in one cell aka orphan contigs. Alternatively, if ReturnFilteredContigs == TRUE, then just the contigs are returned.

.readContigs *Read TCR contigs*

Description

This internal function reads in TCR contigs and joins them. It is a helper function for ECLIPSE.

Usage

```
.readContigs(Contig_type, Contig_folder_paths, Contig_file_paths, Verbose)
```

Arguments

Contig_type	tells whether contig_folder_paths or contig_file_paths should be used to read in contig files and how. To provide file paths, this should be "manual". To provide folders, this should be "all" or "filtered" to search for all or filtered contig annotations files.
Contig_folder_paths	vector of paths to directories that contain TCR files. If you use this argument, provide either "all" or "filtered" to contig_type. The order should match the numbers provided in batch.
Contig_file_paths	vector of TCR file paths. The order should match the numbers provided in batch. The files can have any name, but they must be all_contig_annotations.csv or filtered_contig_annotations.csv format.
Verbose	specifies whether to print an important QC graph, a small table, and also notes through the pipeline.

Value

A data.frame of all contigs

.reformatSeuratObject *Reformat Seurat object*

Description

This internal function renames columns in the Seurat meta.data, removes unneeded columns, and formats the columns for use with scRepertoire downstream. It is a helper function for ECLIPSE.

Usage

```
.reformatSeuratObject(Seurat_object, Group, DetailedEclipseStats, Verbose)
```

Arguments

Seurat_object	Seurat object that contains scRNA-seq data.
Group	name of column in Seurat meta.data. This column should specify how the final statistics for each clone should be calculated. For example, should they be calculated per sample, per donor, or among all cells? This can and often will be the same column as donor.
DetailedEclipseStats	Provides detailed stats relating to EM assignment in Seurat meta.data for each cell if TRUE.
Verbose	specifies whether to print an important QC graph, a small table, and also notes through the pipeline.

Value

A Seurat object that is properly formatted and has all TCR information added.

.runVDJdive *Run VDJdive*

Description

This internal function runs the EM algorithm of VDJdive and extracts the results from the assignment matrix. It is a helper function for ECLIPSE.

Usage

```
.runVDJdive(Contigs, Verbose, Donor)
```

Arguments

Contigs	TCR contigs made from .qcFilterContigs and findThreeChainClones.
Verbose	specifies whether to print an important QC graph, a small table, and also notes through the pipeline.
Donor	name of column in Seurat meta.data. The column should contain the donor, regardless of treatment/timepoint/condition.

Value

A data.frame describing the results from the EM algorithm.

ECLIPSE *Determines the true clonotype of each cell*

Description

Analyzes scTCR-seq data and returns results in Seurat object. Specifically, chain pairings across the sample are analyzed to predict chains lost to sequencing dropout and to decipher whether cells with >2 total TCR chains are caused by technical artifacts or true biology.

Usage

```
ECLIPSE(
  seurat_object,
  contig_type = "all",
  contig_folder_paths = NULL,
  contig_file_paths = NULL,
  batch,
  donor,
  group,
  original_barcode,
  format = "Blank",
  detailedEclipseStats = FALSE,
  combineTcrArgs = list(FALSE, FALSE, FALSE, TRUE),
  allowDoubleChains = "both",
  returnFilteredContigs = FALSE,
  verbose = TRUE
)
```

Arguments

seurat_object	Seurat object that contains scRNA-seq data.
contig_type	tells whether contig_folder_paths or contig_file_paths should be used to read in contig files and how. To provide file paths, this should be "manual". To provide folders, this should be "all" or "filtered" to search for all or filtered contig annotations files.
contig_folder_paths	vector of paths to directories that contain TCR files. If you use this argument, provide either "all" or "filtered" to contig_type. The order should match the numbers provided in batch.
contig_file_paths	vector of TCR file paths. The order should match the numbers provided in batch. The files can have any name, but they must be all_contig_annotations.csv or filtered_contig_annotations.csv format.
batch	name of column in Seurat meta.data. The column should contain integers, not characters, that tell which position in the contig_folder/file_path vectors contains the TCR contigs for the cell.
donor	name of column in Seurat meta.data. The column should contain the donor, regardless of treatment/timepoint/condition.

group name of column in Seurat meta.data. This column should specify how the final statistics for each clone should be calculated. For example, should they be calculated per sample, per donor, or among all cells? This can and often will be the same column as donor.

original_barcode
name of column in Seurat meta.data. This must contain the original 10X barcode for the cell. Typically this will end with -1, but users should check their TCR contig files to be sure.

format default is blank. This tells how missing data is stored in the TCR contig files. Current files likely contain "" so "blank" is the default, but if older contig files are used where any missing data is stored as "None", then "None" should be provided here.

detailedEclipseStats
default is FALSE. Provides detailed stats relating to EM assignment in Seurat meta.data for each cell if TRUE.

combineTcrArgs arguments that are provided to scRepertoire combineTCR. This should be a list of 4 logical elements, with the elements corresponding to the combineTCR arguments. 1 is removeNA, 2 is removeMulti, 3 is filterMulti, 4 is filterNonproductive. If no clonotype can be predicted with high confidence, then combineTCR using these arguments will determine the final clonotype. The default is the combineTCR default where only filterNonproductive is TRUE.

allowDoubleChains
specifies whether 2 alpha chains, 2 beta chains, or either 2 alpha or 2 beta chains are allowed. Default is both.

returnFilteredContigs
specifies if filtered contigs should be returned instead of running the whole ECLIPSE pipeline. Default is FALSE.

verbose specifies whether to print an important QC graph, a small table, and also notes through the pipeline. Default is TRUE.

Value

A Seurat object with TCR data provided in meta.data.

Examples

```
data(test_seurat)

files <- system.file("extdata/examples", package = "ECLIPSE")
tcr_paths <- list.files(
                files,
                pattern = "\\.csv.gz$",
                full.names = TRUE)

cd8_tcr <- ECLIPSE(test_seurat,
                   contig_file_paths = tcr_paths,
                   contig_type = "manual",
                   batch = "index",
                   donor = "donor",
                   group = "donor",
                   original_barcode = "Cell_Barcode")
```

findThreeChainClones *Finds clones with 3 total TCR chains and modifies their contigs*

Description

Analyzes chain pairings across a sample to find clones with 2 TCR alpha or 2 TCR beta chains. Then, the contigs are modified so that these cells only have 1 alpha and 1 beta contig, with ::: separating each CDR3 sequence. These contigs are then returned.

Usage

```
findThreeChainClones(contigs, chain, mode = "standard", printStats = TRUE)
```

Arguments

contigs	tibble of contigs from cellranger all or filtered contig files.
chain	"alpha" or "beta". This tells it whether clones with two of which chain should be looked for.
mode	"standard" is the default, where modified contigs are returned. If "show" is provided, two tables are returned. The first shows the final set of clones with a double chain of one type, and the second shows how/why some of those clones were called as such. The second is pre-filtering, so some clones may not have been called as having 3 total chains in the end.
printStats	specifies whether notes/stats should be printed. Default is TRUE.

Value

a tibble of modified contigs, except if mode show is selected.

Examples

```
data(post_qc_contigs)

modified_contigs <- findThreeChainClones(
                        contigs = post_qc_contigs,
                        chain = "a")
```

post_qc_contigs *Example Post-QC TCR contigs that would be generated within ECLIPSE and passed to findThreeChainClones.*

Description

Example Post-QC TCR contigs that would be generated within ECLIPSE and passed to findThree-ChainClones.

Usage

```
data(post_qc_contigs)
```

Format

A data frame in the all_contig_annotations.csv format that is produced by cellranger vdj.

Source

This is TCR contigs from cells from Zhao et al., Science Immunology, 2021, with doi 10.1126/sciimmunol.abf6692. Please view the README file in inst/ for more information. Code to reproduce this is available in inst/scripts/. The data is available on GEO with GSE167118.

---

sceECLIPSE *Determines the true clonotype of each cell*

Description

Analyzes scTCR-seq data and returns results as a SingleCellExperiment object. Unlike regular ECLIPSE, this function requires a SingleCellExperiment object as an input, not a Seurat object. Specifically, chain pairings across the sample are analyzed to predict chains lost to sequencing dropout and to decipher whether cells with >2 total TCR chains are caused by technical artifacts or true biology.

Usage

```
sceECLIPSE(
  sce_object,
  contig_type = "all",
  contig_folder_paths = NULL,
  contig_file_paths = NULL,
  batch,
  donor,
  group,
  original_barcode,
  format = "Blank",
  detailedEclipseStats = FALSE,
  combineTcrArgs = list(FALSE, FALSE, FALSE, TRUE),
  allowDoubleChains = "both",
  returnFilteredContigs = FALSE,
  verbose = TRUE
)
```

Arguments

sce_object	SingleCellExperiment object that contains scRNA-seq data.
contig_type	tells whether contig_folder_paths or contig_file_paths should be used to read in contig files and how. To provide file paths, this should be "manual". To provide folders, this should be "all" or "filtered" to search for all or filtered contig annotations files.
contig_folder_paths	vector of paths to directories that contain TCR files. If you use this argument, provide either "all" or "filtered" to contig_type. The order should match the numbers provided in batch.

`contig_file_paths`
: vector of TCR file paths. The order should match the numbers provided in batch. The files can have any name, but they must be all_contig_annotations.csv or filtered_contig_annotations.csv format.

`batch`
: name of column in colData. The column should contain integers, not characters, that tell which position in the contig_folder/file_path vectors contains the TCR contigs for the cell.

`donor`
: name of column in colData. The column should contain the donor, regardless of treatment/timepoint/condition.

`group`
: name of column in colData. This column should specify how the final statistics for each clone should be calculated. For example, should they be calculated per sample, per donor, or among all cells? This can and often will be the same column as donor.

`original_barcode`
: name of column in colData. This must contain the original 10X barcode for the cell. Typically this will end with -1, but users should check their TCR contig files to be sure.

`format`
: default is blank. This tells how missing data is stored in the TCR contig files. Current files likely contain "" so "blank" is the default, but if older contig files are used where any missing data is stored as "None", then "None" should be provided here.

`detailedEclipseStats`
: default is FALSE. Provides detailed stats relating to EM assignment in colData for each cell if TRUE.

`combineTcrArgs`
: arguments that are provided to scRepertoire combineTCR. This should be a list of 4 logical elements, with the elements corresponding to the combineTCR arguments. 1 is removeNA, 2 is removeMulti, 3 is filterMulti, 4 is filterNonproductive. If no clonotype can be predicted with high confidence, then combineTCR using these arguments will determine the final clonotype. The default is the combineTCR default where only filterNonproductive is TRUE.

`allowDoubleChains`
: specifies whether 2 alpha chains, 2 beta chains, or either 2 alpha or 2 beta chains are allowed. Default is both.

`returnFilteredContigs`
: specifies if filtered contigs should be returned instead of running the whole ECLIPSE pipeline. Default is FALSE.

`verbose`
: specifies whether to print an important QC graph, a small table, and also notes through the pipeline. Default is TRUE.

Value

A SingleCellExperiment object with TCR data provided in colData.

Examples

```
data(test_seurat)

files <- system.file("extdata/examples", package = "ECLIPSE")
tcr_paths <- list.files(
               files,
               pattern = "\\.csv.gz$",
```

```
                full.names = TRUE)

sce_obj <- Seurat::as.SingleCellExperiment(test_seurat)

cd8_tcr <- sceECLIPSE(sce_obj,
                    contig_file_paths = tcr_paths,
                    contig_type = "manual",
                    batch = "index",
                    donor = "donor",
                    group = "donor",
                    original_barcode = "Cell_Barcode")
```

`tcrArtifactDetect` *Finds and removes doublets/TCR contamination from Seurat objects*

Description

Filters out cells with excessive numbers of TCR chains and helps set thresholds for RNA counts based on TCR chain count.

Usage

```
tcrArtifactDetect(
  object,
  filter = TRUE,
  singleChainLimit = 2,
  totalChainLimit = -1,
  showPlots = TRUE,
  verbose = TRUE
)
```

Arguments

`object`	Seurat object.
`filter`	logical statement whether cells should be filtered or not. Default is TRUE.
`singleChainLimit`	number of alpha or beta chains that is allowed before filtering occurs. Must be a whole number, with a default of 2.
`totalChainLimit`	number of total TCR chains that is allowed before filtering occurs. Must be a whole number. Default is no threshold.
`showPlots`	logical statement that determines whether plots are provided that help set additional filtering thresholds based on RNA counts and TCR chain count. Default is TRUE.
`verbose`	specifies whether notes should be printed. Default is TRUE.

Value

a Seurat object that is filtered if desired.

Examples

```
data(test_seurat_with_tcr)

cd8_tcr_filtered <- tcrArtifactDetect(
                            test_seurat_with_tcr,
                            singleChainLimit = 2)
```

`test_seurat` *Example minimal Seurat object without TCR information.*

Description

Example minimal Seurat object without TCR information.

Usage

```
data(test_seurat)
```

Format

A Seurat object.

Source

This is from 10 cells taken from Zhao et al., Science Immunology, 2021, with doi 10.1126/sciimmunol.abf6692. Please view the README file in inst/ for more information. Code to reproduce this is available in inst/scripts/. The data is available on GEO with GSE167118.

`test_seurat_with_tcr` *Example Seurat object with TCR information generated by running the ECLIPSE function.*

Description

Example Seurat object with TCR information generated by running the ECLIPSE function.

Usage

```
data(test_seurat_with_tcr)
```

Format

A Seurat object.

Source

This is from 10 cells taken from Zhao et al., Science Immunology, 2021, with doi 10.1126/sciimmunol.abf6692. Please view the README file in inst/ for more information. Code to reproduce this is available in inst/scripts/. The data is available on GEO with GSE167118.

unit_test_object *Example small Seurat object without TCR information used for unit testing.*

Description

Example small Seurat object without TCR information used for unit testing.

Usage

```
data(unit_test_object)
```

Format

A Seurat object.

Source

This is from 1759 cells taken from Zhao et al., Science Immunology, 2021, with doi 10.1126/sciimmunol.abf6692. Please view the README file in inst/ for more information. Code to reproduce this is available in inst/scripts/. The data is available on GEO with GSE167118.

vignette_object *Example small Seurat object without TCR information.*

Description

Example small Seurat object without TCR information.

Usage

```
data(vignette_object)
```

Format

A Seurat object.

Source

This is from 1000 cells taken from Zhao et al., Science Immunology, 2021, with doi 10.1126/sciimmunol.abf6692. Please view the README file in inst/ for more information. Code to reproduce this is available in inst/scripts/. The data is available on GEO with GSE167118.

Index

∗ **datasets**
post_qc_contigs, 11
test_seurat, 15
test_seurat_with_tcr, 15
unit_test_object, 16
vignette_object, 16
∗ **internal**
ECLIPSE-package, 2
.assignFinalClonotypes, 3
.findMissingMonotypicChains, 4
.identifyThreeChainClones, 4
.modifyThreeChainContigs, 5
.qcFilterContigs, 6
.readContigs, 7
.reformatSeuratObject, 7
.runVDJdive, 8

ECLIPSE, 9
ECLIPSE-package, 2

findThreeChainClones, 11

post_qc_contigs, 11

sceECLIPSE, 12

tcrArtifactDetect, 14
test_seurat, 15
test_seurat_with_tcr, 15

unit_test_object, 16

vignette_object, 16