Package 'ICAMSxtra'

August 26, 2020

Title ICAMSxtra

Version 0.0.0.9002

Author Steve Rozen, Nanhai Jiang, Arnoud Boot, Mo Liu

Maintainer Steve Rozen <steverozen@gmail.com>

Description ICAMSxtra.

Imports data.table,
ICAMS

Depends R (>= 3.5)

License GPL-3

Encoding UTF-8

LazyData true

Language en-US

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Suggests BSgenome.Hsapiens.1000genomes.hs37d5,
BSgenome.Hsapiens.UCSC.hg38,
BSgenome.Mmusculus.UCSC.mm10,
testthat

R topics documented:

AnnotateIDVCFsWithTransRanges 2
as.catalog.for.ID115 3
Canonicalize1Del115 4
CatalogRowOrder 5
Collapse115CatalogTo83 5
CollapseID115CatalogsToID83s 6
GRCh37.proportions 6
ID115_PlotTransBias 7
ID115_PlotTransBiasToPdf 8
PlotExposure 9
PlotExposureToPdf 10
PlotID115AsID83ToPdf 11
PlotID115Catalog 12
PlotID115CatalogToPdf 12

PlotTranscriptionAssociatedDamageToPdf 13
ReadExposure 14
ReadID115Catalog 14
Reverse 15
Reverse_pooled 16
SortExposure 16
Target 17
Target_pooled 17
VCFsToID115Catalogs 17
VCFsToID115CatalogsAndPlotToPdf 18
WriteExposure 19
WriteID115Catalog 19

Index **20**

`AnnotateIDVCFsWithTransRanges` *Add sequence context and transcript information to an in-memory ID VCF*

Description

Add sequence context and transcript information to an in-memory ID VCF

Usage

```
AnnotateIDVCFsWithTransRanges(
  ID.vcfs,
  ref.genome,
  trans.ranges = NULL,
  vcf.names = NULL
)
```

Arguments

`ID.vcfs`	A list of in-memory ID VCF as a `data.frame`.
`ref.genome`	A `ref.genome` argument as described in `ICAMS`.
`trans.ranges`	Optional. If `ref.genome` specifies one of the `BSgenome` object 1. `BSgenome.Hsapiens.1000genomes.hs37d5` 2. `BSgenome.Hsapiens.UCSC.hg38` 3. `BSgenome.Mmusculus.UCSC.mm10` then the function will infer `trans.ranges` automatically. Otherwise, user will need to provide the necessary `trans.ranges`. Please refer to `TranscriptRanges` for more details. If `is.null(trans.ranges)` do not add transcript range information.
`vcf.names`	list of names of the vcfs

Value

A list of in-memory ID VCFs each a `data.table`. These have been annotated with the sequence context (column name `seq.context`) and with transcript information in the form of a gene symbol (e.g. `"TP53"`) and transcript strand. This information is in the columns `trans.start.pos`, `trans.end.pos` , `trans.strand`, `trans.Ensembl.gene.ID` and `trans.gene.symbol` in the output. These columns are not added if `is.null(trans.ranges)`.

Examples

```
file <- c(system.file("extdata/Strelka-ID-vcf",
                      "Strelka.ID.GRCh37.s1.vcf",
                      package = "ICAMSxtra"))
list.of.vcfs <- ICAMS::ReadStrelkaIDVCFs(file)
if (requireNamespace("BSgenome.Hsapiens.1000genomes.hs37d5", quietly = TRUE)) {
  annotated.ID.vcfs <- AnnotateIDVCFsWithTransRanges(list.of.vcfs, ref.genome = "hg19",
                                                     trans.ranges = ICAMS::trans.ranges.GRCh37)
  }
```

`as.catalog.for.ID115` *Create a catalog from a* `matrix`, `data.frame`, *or* `vector`

Description

Create a catalog from a `matrix`, `data.frame`, or `vector`

Usage

```
as.catalog.for.ID115(
  object,
  ref.genome = NULL,
  region = "unknown",
  catalog.type = "counts",
  abundance = NULL,
  infer.rownames = FALSE
)
```

Arguments

`object`	A numeric `matrix`, numeric `data.frame`, or `vector`. If a `vector`, converted to a 1-column `matrix` with rownames taken from the element names of the `vector` and with column name `"Unknown"`. If argument `infer.rownames` is `FALSE` than this argument must have rownames to denote the mutation types. See `CatalogRowOrder` for more details.
`ref.genome`	A `ref.genome` argument as described in `ICAMS`.
`region`	A character string designating a region, one of `genome`, `transcript`, `exome`, `unknown`; see `ICAMS`.
`catalog.type`	One of "counts", "density", "counts.signature", "density.signature".
`abundance`	If `NULL`, then inferred if `ref.genome` is one of the reference genomes known to ICAMS and `region` is not `unknown`. See `ICAMS`. The argument `abundance` should contain the counts of different source sequences for mutations in the same format as the numeric vectors in `all.abundance`.

infer.rownames If TRUE, and object has no rownames, then assume the rows of object are in the correct order and add the rownames implied by the number of rows in object (e.g. rownames for SBS 192 if there are 192 rows). If TRUE, **be sure the order of rows is correct.**

Value

A catalog as described in ICAMS.

Canonicalize1Del115 *Given a deletion and its sequence context, categorize it*

Description

This function is primarily for internal use, but we export it to document the underlying logic.

Usage

```
Canonicalize1Del115(context, del.seq, pos, trace = 0, strand)
```

Arguments

context	The deleted sequence plus ample surrounding sequence on each side (at least as long as del.seq).
del.seq	The deleted sequence in context.
pos	The position of del.sequence in context.
trace	If > 0, then generate messages tracing how the computation is carried out.
strand	NULL by default. But when called by PlotTransBiasInternal, strand is either + or -. The return value will include :trans or :nontrans indicating whether the deletion occurred on the transcribed or non-transcribed strand.

Details

See https://github.com/steverozen/ICAMS/raw/master/data-raw/PCAWG7_indel_classification_2017_12_08.xlsx for additional information on deletion mutation classification.

This function first handles deletions in homopolymers, then handles deletions in simple repeats with longer repeat units, (e.g. CACACACA, see FindMaxRepeatDel), and if the deletion is not in a simple repeat, looks for microhomology (see FindDelMH).

See the code for unexported function CanonicalizeID and the functions it calls for handling of insertions.

Value

A string that is the canonical representation of the given deletion type. Return NA and raise a warning if there is an un-normalized representation of the deletion of a repeat unit. See FindDelMH for details. (This seems to be very rare.)

@keywords internal

CatalogRowOrder *Standard order of row names in a catalog*

Description

This data is designed for those who need to create their own catalogs from formats not supported by this package. The rownames denote the mutation types. For example, for SBS96 catalogs, the rowname AGAT represents a mutation from AGA > ATA.

Usage

```
catalog.row.order
```

Format

A list of character vectors indicating the standard orders of row names in catalogs.

Note

In ID (small insertion and deletion) catalogs, deletion repeat sizes range from 0 to 5+, but for plotting and end-user documentation deletion repeat sizes range from 1 to 6+. In ID83 catalogs, deletion repeat sizes range from 0 to 5.

Examples

```
catalog.row.order$ID115
# TODO Jia Geng ...
# There are altogether 115 row names to denote the mutation types
# in ID115 catalog.
```

Collapse115CatalogTo83 *"Collapse" a catalog*

Description

`Collapse115CatalogTo83` Collapse a ID 115 catalog to a ID 83 catalog.

Usage

```
Collapse115CatalogTo83(catalog)
```

Arguments

`catalog`	A catalog as defined in ICAMS.

Value

A catalog as defined in ICAMS.

`CollapseID115CatalogsToID83s` *"Collapse" a matrix of ID 115 catalogs to ID 83 catalog*

Description

"Collapse" a matrix of ID 115 catalogs to ID 83 catalog

Usage

```
CollapseID115CatalogsToID83s(catalogs)
```

Arguments

`catalogs`	A catalog as defined in `ICAMS`.

Value

A catalog as defined in `ICAMS`.

`GRCh37.proportions` *TODO Jia Geng*

Description

TODO Jia Geng

Usage

```
GRCh37.proportions
```

Format

TODO Jia Geng

`ID115_PlotTransBias` *Plot transcription strand bias*

Description

Plot transcription strand bias

Usage

```
ID115_PlotTransBias(annotated.ID.vcf, pool, damaged.base = NULL, ymax = NULL)
```

Arguments

`annotated.ID.vcf`	An ID VCF annotated by `AnnotateIDVCFsWithTransRanges`. It **must** have transcript range information added.
`pool`	TODO Jia Geng
`damaged.base`	One of `NULL`, `"purine"` or `"pyrimidine"`. This function allocates approximately equal numbers of mutations from `damaged.base` into each of `num.of.bins` bin by expression level. E.g. if `damaged.base` is `"purine"`, then mutations from A and G will be allocated in approximately equal numbers to each expression-level bin. The rationale for the name `damaged.base` is that the direction of strand bias is a result of whether the damage occurs on a purine or pyrimidine. If `NULL`, the function attempts to infer the `damaged.base` based on mutation counts.
`ymax`	Limit for the y axis. If not specified, it defaults to NULL and the y axis limit equals 1.5 times of the maximum mutation counts in a specific mutation type.

Value

A list whose first element is a logic value indicating whether the plot is successful. The second element is a named numeric vector containing the p-values printed on the plot.

Note

The strand bias statistics are Benjamini-Hochberg q-values based on two-sided binomial tests of the mutation counts on the transcribed and untranscribed strands relative to the actual abundances of C and T on the transcribed strand. On the plot, asterisks indicate q-values as follows *, $Q < 0.05$; **, $Q < 0.01$; ***, $Q < 0.001$.

Examples

```
file <- c(system.file("extdata/Strelka-ID-vcf/",
                      "Strelka.ID.GRCh37.s1.vcf",
                      package = "ICAMSxtra"))
ID.vcf <- ICAMS::ReadStrelkaIDVCFs(file)
if (requireNamespace("BSgenome.Hsapiens.1000genomes.hs37d5", quietly = TRUE)) {
  annotated.ID.vcf <- AnnotateIDVCFsWithTransRanges(ID.vcf, ref.genome = "hg19",
                                                    trans.ranges = ICAMS::trans.ranges.GRCh37,
                                                       vcf.names = "Strelka.ID.GRCh37.s1.vcf")
  #' alternatively run below code to skip call to AnnotateIDVCFsWithTransRanges
```

```
  #' load(c(system.file("extdata/annotated.ID.vcf.rda",package = "ICAMSxtra")))
  ID115_PlotTransBias(annotated.ID.vcf = annotated.ID.vcf[[1]],
                      pool = TRUE)
}
```

`ID115_PlotTransBiasToPdf` *Plot transcription strand bias to a PDF file*

Description

Plot transcription strand bias to a PDF file

Usage

```
ID115_PlotTransBiasToPdf(annotated.ID.vcfs, file, pool, damaged.base = NULL)
```

Arguments

`annotated.ID.vcfs`	TODO Jia Geng
`file`	The name of output file.
`pool`	TODO Jia Geng
`damaged.base`	One of `NULL`, `"purine"` or `"pyrimidine"`. This function allocates approximately equal numbers of mutations from `damaged.base` into each of `num.of.bins` bin by expression level. E.g. if `damaged.base` is `"purine"`, then mutations from A and G will be allocated in approximately equal numbers to each expression-level bin. The rationale for the name `damaged.base` is that the direction of strand bias is a result of whether the damage occurs on a purine or pyrimidine. If `NULL`, the function attempts to infer the `damaged.base` based on mutation counts.

Value

A list whose first element is a logic value indicating whether the plot is successful. The second element is a named numeric vector containing the p-values printed on the plot.

Note

The strand bias statistics are Benjamini-Hochberg q-values based on two-sided binomial tests of the mutation counts on the transcribed and untranscribed strands relative to the actual abundances of C and T on the transcribed strand. On the plot, asterisks indicate q-values as follows *, $Q < 0.05$; **, $Q < 0.01$; ***, $Q < 0.001$.

Examples

```
library(ICAMS)
file <- c(system.file("extdata/Strelka-ID-vcf/",
                      "Strelka.ID.GRCh37.s1.vcf",
                      package = "ICAMSxtra"))
ID.vcf <- ICAMS::ReadStrelkaIDVCFs(file)
if (requireNamespace("BSgenome.Hsapiens.1000genomes.hs37d5", quietly = TRUE)) {
```

```
  annotated.ID.vcf <- AnnotateIDVCFsWithTransRanges(ID.vcf, ref.genome = "hg19",
                                      trans.ranges = ICAMS::trans.ranges.GRCh37,
                                      vcf.names = "Strelka.ID.GRCh37.s1")
  ID115_PlotTransBiasToPdf(annotated.ID.vcfs = annotated.ID.vcf,
                           file = file.path(tempdir(), "test.pdf"),
                           pool = TRUE)
}
```

`PlotExposure` *Plot exposures in multiple plots each with a manageable number of samples*

Description

Plot exposures in multiple plots each with a manageable number of samples

Usage

```
PlotExposure(
  exposure,
  samples.per.line = 30,
  plot.proportion = FALSE,
  xlim = NULL,
  ylim = NULL,
  legend.x = NULL,
  legend.y = NULL,
  cex.legend = 0.9,
  ...
)
```

Arguments

`exposure`	Exposures as a numerical `matrix` (or `data.frame`) with signatures in rows and samples in columns. Rownames are taken as the signature names and column names are taken as the sample IDs. If you want `exposure` sorted from largest to smallest, use `SortExposure`. Do not use column names that start with multiple underscores. The exposures will often be mutation counts, but could also be e.g. mutations per megabase.
`samples.per.line`	Number of samples to show in each plot.
`plot.proportion`	Plot exposure proportions rather than counts.
`xlim, ylim`	Limits for the x and y axis. If `NULL`(default), the function tries to do something reasonable.
`legend.x, legend.y`	The x and y co-ordinates to be used to position the legend.
`cex.legend`	A numerical value giving the amount by which legend plotting text and symbols should be magnified relative to the default.
`...`	Other arguments passed to `barplot`. If `ylab` is not included, it defaults to a value depending on `plot.proportion`. If `col` is not supplied the function tries to do something reasonable.

Value

An **invisible** list whose first element is a logic value indicating whether the plot is successful. The second element is a numeric vector giving the coordinates of all the bar midpoints drawn, useful for adding to the graph.

Examples

```
file <- system.file("extdata",
                    "synthetic.exposure.csv",
                    package = "ICAMSxtra")
exposure <- ReadExposure(file)
PlotExposure(exposure[, 1:30])
```

`PlotExposureToPdf` *Plot exposures in multiple plots each with a manageable number of samples to PDF*

Description

Plot exposures in multiple plots each with a manageable number of samples to PDF

Usage

```
PlotExposureToPdf(
  exposure,
  file,
  mfrow = c(2, 1),
  mar = c(6, 4, 3, 2),
  oma = c(3, 2, 0, 2),
  samples.per.line = 30,
  plot.proportion = FALSE,
  xlim = NULL,
  ylim = NULL,
  legend.x = NULL,
  legend.y = NULL,
  cex.legend = 0.9,
  ...
)
```

Arguments

`exposure`	Exposures as a numerical `matrix` (or `data.frame`) with signatures in rows and samples in columns. Rownames are taken as the signature names and column names are taken as the sample IDs. If you want `exposure` sorted from largest to smallest, use `SortExposure`. Do not use column names that start with multiple underscores. The exposures will often be mutation counts, but could also be e.g. mutations per megabase.
`file`	The name of the PDF file to be produced.
`mfrow`	A vector of the form `c(nr,nc)`. Subsequent figures will be drawn in an `nr`-by-`nc` array on the device by rows.

mar	A numerical vector of the form c(bottom,left,top,right) which gives the number of lines of margin to be specified on the four sides of the plot.
oma	A vector of the form c(bottom,left,top,right) giving the size of the outer margins in lines of text.
samples.per.line	Number of samples to show in each plot.
plot.proportion	Plot exposure proportions rather than counts.
xlim, ylim	Limits for the x and y axis. If NULL(default), the function tries to do something reasonable.
legend.x, legend.y	The x and y co-ordinates to be used to position the legend.
cex.legend	A numerical value giving the amount by which legend plotting text and symbols should be magnified relative to the default.
...	Other arguments passed to barplot. If ylab is not included, it defaults to a value depending on plot.proportion. If col is not supplied the function tries to do something reasonable.

Value

An **invisible** list whose first element is a logic value indicating whether the plot is successful. The second element is a numeric vector giving the coordinates of all the bar midpoints drawn, useful for adding to the graph.

Examples

```
file <- system.file("extdata",
                    "synthetic.exposure.csv",
                    package = "ICAMSxtra")
exposure <- ReadExposure(file)
PlotExposureToPdf(exposure, file = file.path(tempdir(), "exposure.pdf"))
```

PlotID115AsID83ToPdf *Plot an ID 115 signatures (default) or catalog as standard ID83 and save as pdf file*

Description

Plot an ID 115 signatures (default) or catalog as standard ID83 and save as pdf file

Usage

```
PlotID115AsID83ToPdf(catalog, file, ylim = NULL)
```

Arguments

catalog	A catalog as defined in ICAMS.
file	The name of the PDF file to be produced.
ylim	Has the usual meaning. Only implemented for SBS96Catalog and IndelCatalog.

Value

A list whose first element is a logic value indicating whether the plot is successful. For **SBS192Catalog** with "counts" catalog.type and non-null abundance and `plot.SBS12 = TRUE`, the list will have a second element which is a list containing the strand bias statistics.

`PlotID115Catalog` *Plot* **one** *spectrum or signature*

Description

Plot the spectrum of **one** sample or plot **one** signature. The type of graph is based on one attribute("catalog.type") of the input catalog. You can first use `TransformCatalog` to get different types of catalog and then do the plotting.

Usage

```
PlotID115Catalog(catalog, ylim = NULL)
```

Arguments

`catalog`	A catalog as defined in `ICAMS` with attributes added. See `as.catalog` for more details.
`ylim`	Has the usual meaning. Only implemented for SBS96Catalog and IndelCatalog.

`PlotID115CatalogToPdf` *Plot catalog to a PDF file*

Description

Plot catalog to a PDF file. The type of graph is based on one attribute("catalog.type") of the input catalog. You can first use `TransformCatalog` to get different types of catalog and then do the plotting.

Usage

```
PlotID115CatalogToPdf(catalog, file, ylim = NULL)
```

Arguments

`catalog`	A catalog as defined in `ICAMS` with attributes added. See `as.catalog` for more details.
`file`	The name of the PDF file to be produced.
`ylim`	Has the usual meaning. Only implemented for SBS96Catalog and IndelCatalog.

Value

A list whose first element is a logic value indicating whether the plot is successful. For **SBS192Catalog** with "counts" catalog.type and non-null abundance and `plot.SBS12 = TRUE`, the list will have a second element which is a list containing the strand bias statistics.

Note

The sizes of repeats involved in deletions range from 0 to 5+ in the mutational-spectra and signature catalog rownames, but for plotting and end-user documentation deletion repeat sizes range from 1 to 6+.

`PlotTranscriptionAssociatedDamageToPdf`
Plot indel counts on transcribed and nontranscribed strands to pdf

---

Description

Plot indel counts on transcribed and nontranscribed strands to pdf

Usage

```
PlotTranscriptionAssociatedDamageToPdf(
  list.of.vcfs,
  ref.genome,
  names.of.vcfs,
  file
)
```

Arguments

`list.of.vcfs`	List of in-memory ID VCFs. The list names will be the sample ids in the output catalog.
`ref.genome`	A `ref.genome` argument as described in ICAMS.
`names.of.vcfs`	list of names of vcfs
`file`	The name of the PDF file to be produced.

Value

a list of tables of p-values for each vcf

Note

The strand bias statistics are Benjamini-Hochberg q-values based on two-sided binomial tests of the mutation counts on the transcribed and untranscribed strands relative to the actual abundances of C and T on the transcribed strand. On the plot, asterisks indicate q-values as follows *, $Q < 0.05$; **, $Q < 0.01$; ***, $Q < 0.001$.

Examples

```
dirs <- c(system.file("extdata/Strelka-ID-vcf", "Strelka.ID.GRCh37.s1.vcf", package="ICAMSxtra"),
       system.file("extdata/Strelka-ID-vcf", "Strelka.ID.GRCh37.s2.vcf", package="ICAMSxtra"))

list.of.vcfs <- ICAMS::ReadStrelkaIDVCFs(dirs, names.of.VCFs = c("s1","s2"))
PlotTranscriptionAssociatedDamageToPdf(list.of.vcfs = list.of.vcfs,
                                       ref.genome = "hg19",
                                       names.of.vcfs = c("s1","s2"),
                                       file = file.path(tempdir(), "test.pdf"))
```

ReadExposure *Read an exposure matrix from a file*

Description

Read an exposure matrix from a file

Usage

```
ReadExposure(file, check.names = FALSE)
```

Arguments

file	CSV file containing an exposure matrix.
check.names	Passed to read.csv. **IMPORTANT**: If TRUE this will replace the double colon in identifiers of the form <tumor_type>::<sample_id> with two periods (i.e. <tumor_type>..<sample_id>. If check.names is true, generate a warning if double colons were present.

Value

Matrix of exposures.

Examples

```
file <- system.file("extdata",
                    "synthetic.exposure.csv",
                    package = "ICAMSxtra")
exposure <- ReadExposure(file)
```

ReadID115Catalog *Read catalog*

Description

Read a catalog in standardized format from path.

Usage

```
ReadID115Catalog(
  file,
  ref.genome = NULL,
  region = "unknown",
  catalog.type = "counts"
)
```

Arguments

file	Path to a catalog on disk in the standardized format.
ref.genome	A ref.genome argument as described in ICAMS.
region	region A character string designating a genomic region; see as.catalog and ICAMS.
catalog.type	One of "counts", "density", "counts.signature", "density.signature".

Details

See also WriteCatalog

Value

A catalog as an S3 object; see as.catalog.

Comments

To add or change attributes of the catalog, you can use function attr.
For example, attr(catalog,"abundance") <-custom.abundance.

Note

In ID (small insertion and deletion) catalogs, deletion repeat sizes range from 0 to 5+, but for plotting and end-user documentation deletion repeat sizes range from 1 to 6+.

Reverse *Transcription bias of indels classified into 115 categories (purine)*

Description

This data is designed to be used as an example in function
ID115_PlotTransBias and ID115_PlotTransBiasToPdf.

Usage

```
reverse
```

Format

A vector which contains the 115 categories of indel events, but in purine format

An object of class character of length 36.

Reverse_pooled — *Transcription bias of indels classified into 13 categories (purine)*

Description

This data is designed to be used as an example in function ID115_PlotTransBias and ID115_PlotTransBias when pool = TRUE.

Usage

```
reverse_pooled
```

Format

A vector which contains the 13 categories of indel events, standardised to purine format.

An object of class character of length 4.

SortExposure — *Sort columns of an exposure matrix from largest to smallest (or vice versa)*

Description

Sort columns of an exposure matrix from largest to smallest (or vice versa)

Usage

```
SortExposure(exposure, decreasing = TRUE)
```

Arguments

exposure	Exposures as a numerical matrix (or data.frame) with signatures in rows and samples in columns. Rownames are taken as the signature names and column names are taken as the sample IDs.
decreasing	If TRUE, sort from largest to smallest.

Value

The original exposure with columns sorted.

Examples

```
file <- system.file("extdata",
                    "synthetic.exposure.csv",
                    package = "ICAMSxtra")
exposure <- ReadExposure(file)
exposure.sorted <- SortExposure(exposure)
```

Target — *Transcription bias of indels classified into 115 categories (pyrimidine)*

Description

This data is designed to be used as an example in function ID115_PlotTransBias and ID115_PlotTransBiasToPdf.

Usage

```
target
```

Format

A vector which contains the 115 categories of indel events, standardised to pyrimidine format.

An object of class character of length 36.

Target_pooled — *Transcription bias of indels classified into 13 categories (pyrimidine)*

Description

This data is designed to be used as an example in function ID115_PlotTransBias and ID115_PlotTransBias

Usage

```
target_pooled
```

Format

A vector which contains the 13 categories of indel events, standardised to pyrimidine format.

An object of class character of length 4.

VCFsToID115Catalogs — *Create ID (small insertion and deletion) catalog from ID VCFs*

Description

Create ID (small insertion and deletion) catalog from ID VCFs

Usage

```
VCFsToID115Catalogs(
  list.of.vcfs,
  ref.genome,
  region = "unknown",
  flag.mismatches = 0
)
```

Arguments

list.of.vcfs	List of in-memory ID VCFs. The list names will be the sample ids in the output catalog.
ref.genome	A ref.genome argument as described in ICAMS.
region	A character string acting as a region identifier, one of "genome", "exome".
flag.mismatches	Optional. If > 0, then if there are mismatches to references in the ID (insertion/deletion) VCF, generate messages showing the mismatched rows and continue. Otherwise stop if there are mismatched rows. See AnnotateIDVCF for more details.

Value

A list of two elements. 1st element catalog is the ID (small insertion and deletion) catalog with attributes added. See as.catalog for more details. 2nd element annotated.vcfs is a list of data frames which contain the original VCF with three additional columns seq.context.width, seq.context and ID.class added. The category assignment of each ID mutation in VCF can be obtained from ID.class column.

Note

In ID (small insertion and deletion) catalogs, deletion repeat sizes range from 0 to 5+, but for plotting and end-user documentation deletion repeat sizes range from 1 to 6+.

VCFsToID115CatalogsAndPlotToPdf

Read a list of vcfs and plot ID115 catalogs as pdf

Description

Read a list of vcfs and plot ID115 catalogs as pdf

Usage

```
VCFsToID115CatalogsAndPlotToPdf(
  list.of.vcfs,
  ref.genome,
  region = "unknown",
  flag.mismatches = 0,
  file,
  ylim = NULL
)
```

Arguments

list.of.vcfs	List of in-memory ID VCFs. The list names will be the sample ids in the output catalog.
ref.genome	A ref.genome argument as described in ICAMS.
region	A character string acting as a region identifier, one of "genome", "exome".

flag.mismatches
: Optional. If > 0, then if there are mismatches to references in the ID (insertion/deletion) VCF, generate messages showing the mismatched rows and continue. Otherwise stop if there are mismatched rows. See AnnotateIDVCF for more details.

file
: The name of the PDF file to be produced.

ylim
: Has the usual meaning. Only implemented for SBS96Catalog and IndelCatalog.

WriteExposure *Write an exposure matrix to a file*

Description

Write an exposure matrix to a file

Usage

```
WriteExposure(exposure, file)
```

Arguments

exposure
: Exposures as a numerical matrix (or data.frame) with signatures in rows and samples in columns. Rownames are taken as the signature names and column names are taken as the sample IDs.

file
: File to which to write the exposure matrix (as a CSV file).

Examples

```
file <- system.file("extdata",
                    "synthetic.exposure.csv",
                    package = "ICAMSxtra")
exposure <- ReadExposure(file)
WriteExposure(exposure, file = file.path(tempdir(), "synthetic.exposure.csv"))
```

WriteID115Catalog *Write a catalog to a file.*

Description

Write a catalog to a file.

Usage

```
WriteID115Catalog(catalog, file, strict = TRUE)
```

Arguments

catalog
: A catalog as defined in ICAMS with attributes added. See as.catalog for more details.

file
: The path of the file to be written.

strict
: If TRUE, then stop if additional checks on the input fail.

Index

∗ **datasets**
CatalogRowOrder, 5
GRCh37.proportions, 6
Reverse, 15
Reverse_pooled, 16
Target, 17
Target_pooled, 17

all.abundance, *3*
AnnotateIDVCF, *18, 19*
AnnotateIDVCFsWithTransRanges, 2
as.catalog, *12*, *15*, *18, 19*
as.catalog.for.ID115, 3
attr, *15*

barplot, *9*, *11*
BSgenome, *2*
BSgenome.Hsapiens.1000genomes.hs37d5, *2*
BSgenome.Hsapiens.UCSC.hg38, *2*
BSgenome.Mmusculus.UCSC.mm10, *2*

Canonicalize1Del115, 4
CanonicalizeID, *4*
catalog.row.order *(*CatalogRowOrder*)*, 5
CatalogRowOrder, *3*, 5
Collapse115CatalogTo83, 5
CollapseID115CatalogsToID83s, 6

FindDelMH, *4*
FindMaxRepeatDel, *4*

GRCh37.proportions, 6

ICAMS, *2–6*, *11–13*, *15*, *18, 19*
ID115_PlotTransBias, 7, *15–17*
ID115_PlotTransBiasToPdf, 8, *15*, *17*

PlotExposure, 9
PlotExposureToPdf, 10
PlotID115AsID83ToPdf, 11
PlotID115Catalog, 12
PlotID115CatalogToPdf, 12
PlotTranscriptionAssociatedDamageToPdf, 13

read.csv, *14*
ReadExposure, 14
ReadID115Catalog, 14
Reverse, 15
reverse *(*Reverse*)*, 15
Reverse_pooled, 16
reverse_pooled *(*Reverse_pooled*)*, 16

SortExposure, *9, 10*, 16

Target, 17
target *(*Target*)*, 17
Target_pooled, 17
target_pooled *(*Target_pooled*)*, 17
TranscriptRanges, *2*
TransformCatalog, *12*

VCFsToID115Catalogs, 17
VCFsToID115CatalogsAndPlotToPdf, 18

WriteCatalog, *15*
WriteExposure, 19
WriteID115Catalog, 19