Package 'cosmicsig'

October 19, 2021

Title Repository of Mutational Signatures Data from COSMIC (the Catalogue of Somatic Mutations in Cancer)

Version 1.0.1

Author Steven G. Rozen, Nanhai Jiang

Maintainer Steven G. Rozen <steverozen@gmail.com>

Description Contains mutational signatures data from the COSMIC (the Catalogue Of Somatic Mutations In Cancer) website <https://cancer.sanger.ac.uk/signatures/>.

License GPL-3 | file LICENSE

URL https://github.com/Rozen-Lab/cosmicsig

BugReports https://github.com/Rozen-Lab/cosmicsig/issues

Language en-US

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.2

Depends R (>= 3.5)

Suggests testthat (>= 3.0.0), ICAMS, usethis

Config/testthat/edition 3

R topics documented:

`cosmicsig` *cosmicsig: A package of mutational signatures data from COSMIC (the Catalogue Of Somatic Mutations In Cancer) website* https://cancer.sanger.ac.uk/signatures/

Description

This is a data package with 2 main package variables: `signature` and `etiology`.

Details

There are also two functions for handling COSMIC signatures:

- `get_etiology`
- `SBS96_ID_to_SBS192_ID`

Source

COSMIC mutational signatures data were downloaded from https://cancer.sanger.ac.uk/signatures/downloads/.

`COSMIC_v3.0` *Mutational signatures data from COSMIC, the Catalogue Of Somatic Mutations In Cancer, (v3.0 - May 2019)*

Description

Mutational signatures data from COSMIC, the Catalogue Of Somatic Mutations In Cancer, (v3.0 - May 2019)

Usage

```
COSMIC_v3.0
```

Format

A list with one element `signature`.

- `signature` is a list with the elements:
  - `SBS96`: Strand-agnostic single-base substitutions in trinucleotide context.
  - `DBS78`: Strand-agnostic doublet-base substitutions.
  - `ID`: Strand-agnostic indels.

Remark

`SBS96` and `DBS78` signatures are from Human **GRCh38** reference genome. `SBS192` and `ID` signatures are from Human **GRCh37** reference genome. See CatalogRowOrder for the classification of mutation types.

Source

https://cancer.sanger.ac.uk/signatures/downloads/.

Examples

```
# As the abundances of the source sequence of the mutations vary between genome
# and exome and between species, users can use ICAMS to do the transformations.
if (!requireNamespace("ICAMS", quietly = TRUE)) {
  install.packages("ICAMS")
}
library(ICAMS)
SBS96_sigs_GRCh38_genome <- COSMIC_v3.0$signature$SBS96

# Transform SBS96 GRCh38 genome signatures to GRCh38 exome signatures
SBS96_sigs_GRCh38_exome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                            target.ref.genome = "GRCh38",
                                            target.region = "exome")

# Transform SBS96 GRCh38 genome signatures to GRCh37 genome signatures
SBS96_sigs_GRCh37_genome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                             target.ref.genome = "GRCh37",
                                             target.region = "genome")

# Transform SBS96 GRCh38 genome signatures to mm10 genome signatures
SBS96_sigs_mm10_genome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                           target.ref.genome = "mm10",
                                           target.region = "genome")
```

COSMIC_v3.1 — *Mutational signatures data from COSMIC, the Catalogue Of Somatic Mutations In Cancer, (v3.1 - June 2020)*

Description

Mutational signatures data from COSMIC, the Catalogue Of Somatic Mutations In Cancer, (v3.1 - June 2020)

Usage

```
COSMIC_v3.1
```

Format

A list with one element `signature`.

- `signature` is a list with the elements:
  - `SBS96`: Strand-agnostic single-base substitutions in trinucleotide context.
  - `SBS192`: Transcriptionally stranded single-base substitutions in trinucleotide context.
  - `DBS78`: Strand-agnostic doublet-base substitutions.
  - `ID`: Strand-agnostic indels.

Remark

SBS96 and DBS78 signatures are from Human **GRCh38** reference genome. SBS192 and ID signatures are from Human **GRCh37** reference genome. See CatalogRowOrder for the classification of mutation types.

Source

https://cancer.sanger.ac.uk/signatures/downloads/.

Examples

```
# As the abundances of the source sequence of the mutations vary between genome
# and exome and between species, users can use ICAMS to do the transformations.
if (!requireNamespace("ICAMS", quietly = TRUE)) {
  install.packages("ICAMS")
}
library(ICAMS)
SBS96_sigs_GRCh38_genome <- COSMIC_v3.1$signature$SBS96

# Transform SBS96 GRCh38 genome signatures to GRCh38 exome signatures
SBS96_sigs_GRCh38_exome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                            target.ref.genome = "GRCh38",
                                            target.region = "exome")

# Transform SBS96 GRCh38 genome signatures to GRCh37 genome signatures
SBS96_sigs_GRCh37_genome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                             target.ref.genome = "GRCh37",
                                             target.region = "genome")

# Transform SBS96 GRCh38 genome signatures to mm10 genome signatures
SBS96_sigs_mm10_genome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                           target.ref.genome = "mm10",
                                           target.region = "genome")
```

COSMIC_v3.2 — *Mutational signatures data from COSMIC, the Catalogue Of Somatic Mutations In Cancer, (v3.2 - March 2021)*

Description

Mutational signatures data from COSMIC, the Catalogue Of Somatic Mutations In Cancer, (v3.2 - March 2021)

Usage

```
COSMIC_v3.2
```

Format

A list with two elements, signature and etiology.

- signature is a list with the elements:
  - SBS96: Strand-agnostic single-base substitutions in trinucleotide context.

  - SBS192: Transcriptionally stranded single-base substitutions in trinucleotide context.
  - DBS78: Strand-agnostic doublet-base substitutions.
  - ID: Strand-agnostic indels.
- etiology is a list with elements:
  - SBS96
  - SBS192
  - DBS78
  - ID

  Each element in etiology is a single column matrix with rownames being the signature IDs and values being a short character string description of the proposed etiology.

  In general use get_etiology, which handles new signatures without elements in etiology.

Remark

SBS96 and DBS78 signatures are from Human **GRCh38** reference genome. SBS192 and ID signatures are from Human **GRCh37** reference genome. See CatalogRowOrder for the classification of mutation types.

Note

SBS10c, SBS10d, SBS91, SBS92, SBS93, SBS94 (total 6) new SBS signatures were added in COSMIC v3.2. See the news from COSMIC release for more details https://cosmic-blog.sanger.ac.uk/cosmic-mutational-signatures-release-v3-2/

Source

https://cancer.sanger.ac.uk/signatures/downloads/.

Examples

```
# As the abundances of the source sequence of the mutations vary between genome
# and exome and between species, users can use ICAMS to do the transformations.
if (!requireNamespace("ICAMS", quietly = TRUE)) {
  install.packages("ICAMS")
}
library(ICAMS)
SBS96_sigs_GRCh38_genome <- COSMIC_v3.2$signature$SBS96

# Transform SBS96 GRCh38 genome signatures to GRCh38 exome signatures
SBS96_sigs_GRCh38_exome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                            target.ref.genome = "GRCh38",
                                            target.region = "exome")

# Transform SBS96 GRCh38 genome signatures to GRCh37 genome signatures
SBS96_sigs_GRCh37_genome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                             target.ref.genome = "GRCh37",
                                             target.region = "genome")

# Transform SBS96 GRCh38 genome signatures to mm10 genome signatures
SBS96_sigs_mm10_genome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                           target.ref.genome = "mm10",
                                           target.region = "genome")
```

etiology — *List of mutational signature's proposed etiology summarized from COSMIC, the Catalogue Of Somatic Mutations In Cancer (v3.2 - March 2021)*

Description

List of mutational signature's proposed etiology summarized from COSMIC, the Catalogue Of Somatic Mutations In Cancer (v3.2 - March 2021)

Usage

```
etiology
```

Format

A list with the elements:

- SBS96
- SBS192
- DBS78
- ID

Each list element is a single column matrix with rownames being the signature IDs and values being a short character string description of the proposed etiology.

In general use `get_etiology`, which handles new signatures do not have an element in `etiology`.

Examples

```
SBS96_etiology <- etiology$SBS96
```

get_etiology — *Get the proposed etiology of COSMIC signature*

Description

The level of evidence supporting the proposed etiologies varies. In addition, some proposed etiologies are more akin to associations than specific, mechanistic causes.

Usage

```
get_etiology(mutation_type, sig_id)
```

Arguments

`mutation_type`	Character string, one of "SBS96", "SBS192", "DBS78", "ID".
`sig_id`	Character vector with signature ids, e.g. `c("SBS3","SBS5")`.

Value

A character vector of the same length as `sig_id`, each element of which is the etiology of the corresponding signature, if known, or else the empty string.

Note

The etiologies information is not versionized in COSMIC website compared to signatures.

Examples

```
get_etiology(mutation_type = "ID", sig_id = c("ID1", "foo", "ID3"))
```

SBS96_ID_to_SBS192_ID *Translate SBS96 signature IDs to SBS192 signature IDs by adding "-E" if necessary*

Description

"-E" added to the name of a transcriptional strand bias signature indicates that it was extracted only from exome sequencing data, and thus reflects transcriptional strand bias in the exome rather than in the entire transcript, including introns.

Usage

```
SBS96_ID_to_SBS192_ID(sig_ids)
```

Arguments

`sig_ids`	Character vector of SBS96 signature IDs.

Value

Character vector of corresponding SBS192 signature IDs; some have "-E" (for exome) post-pended.

Examples

```
SBS96_ids <- c("SBS1", "SBS23", "SBS25")
SBS192_ids <- SBS96_ID_to_SBS192_ID(SBS96_ids)
```

signature *Mutational signatures data from COSMIC, the Catalogue Of Somatic Mutations In Cancer (v3.2 - March 2021)*

Description

Mutational signatures data from COSMIC, the Catalogue Of Somatic Mutations In Cancer (v3.2 - March 2021)

Usage

```
signature
```

Format

A list with the following elements:

- SBS96: Strand-agnostic single-base substitutions in trinucleotide context.
- SBS192: Transcriptionally stranded single-base substitutions in trinucleotide context.
- DBS78: Strand-agnostic doublet-base substitutions.
- ID: Strand-agnostic indels.

Remark

SBS96 and DBS78 signatures are from Human **GRCh38** reference genome. SBS192 and ID signatures are from Human **GRCh37** reference genome. See CatalogRowOrder for the classification of mutation types.

Note

SBS10c, SBS10d, SBS91, SBS92, SBS93, SBS94 (total 6) new SBS signatures were added in COSMIC v3.2. See the news from COSMIC release for more details https://cosmic-blog.sanger.ac.uk/cosmic-mutational-signatures-release-v3-2/

Source

https://cancer.sanger.ac.uk/signatures/downloads/.

Examples

```
# As the abundances of the source sequence of the mutations vary between genome
# and exome and between species, users can use ICAMS to do the transformations.
if (!requireNamespace("ICAMS", quietly = TRUE)) {
  install.packages("ICAMS")
}
library(ICAMS)
SBS96_sigs_GRCh38_genome <- signature$SBS96

# Transform SBS96 GRCh38 genome signatures to GRCh38 exome signatures
SBS96_sigs_GRCh38_exome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                            target.ref.genome = "GRCh38",
                                            target.region = "exome")
```

```
# Transform SBS96 GRCh38 genome signatures to GRCh37 genome signatures
SBS96_sigs_GRCh37_genome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                             target.ref.genome = "GRCh37",
                                             target.region = "genome")

# Transform SBS96 GRCh38 genome signatures to mm10 genome signatures
SBS96_sigs_mm10_genome <- TransformCatalog(catalog = SBS96_sigs_GRCh38_genome,
                                           target.ref.genome = "mm10",
                                           target.region = "genome")
```

Index