Package 'cosmicsig'

December 17, 2021

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

Version 1.0.5

Description A data package with 2 main package variables: 'signature' and 'etiology'.
The 'signature' variable contains the latest mutational signature profiles
released on COSMIC <https://cancer.sanger.ac.uk/signatures/> for 4 mutation
types:
* single base substitutions in the context of preceding and following bases,
* Strand bias single base substitutions: single base substitutions
from transcribed regions, that take into consideration the
the transcribed versus non-transcribed strand,
* Doublet base substitutions, and
* Small insertions and deletions.
The 'etiology' variable provides the known or hypothesized causes of signatures.
'cosmicsig' stands for COSMIC signatures. Please run ?'cosmicsig' for more
information.

License GPL-3

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 from COSMIC (Catalogue Of Somatic Mutations In Cancer)* https://cancer.sanger.ac.uk/signatures/.

Description

For a general introduction to mutational signatures and the techniques used to discover them, see Alexandrov et al., 2020 doi: 10.1038/s4158602019433.

Details

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

The `signature` variable contains the latest mutational signature profiles released on https://cancer.sanger.ac.uk/signatures/ for 4 mutation types:

- SBS (single base substitutions in the context of preceding and following bases, called SBS96 in this package)
- Strand bias SBS: SBSs from transcribed regions (exons and introns), that take into into consideration the transcriptional strand.
- DBS (doublet base substitutions, called DBS78 in this package)
- ID (small insertions and deletions)

The package variable `etiology` contains information on known or hypothesized causes of mutational signatures. In general, it is better to use `get_etiology`.

Earlier releases are available in the variables COSMIC_*version*, e.g. `COSMIC_v3.1`.

The profiles of SBSs signatures depend on the frequencies of trinucleotides in a genome and profiles of DBS signatures depend on the frequencies of dinucleotides in a genome. Therefore COSMIC and this package provide slightly different signatures for different reference genomes. COSMIC and this package offer versions of SBS and DBS signatures for human GRCh37 (also known as hg19) and GRCh38, and for mouse and rat. However strand bias SBS signatures are only available for GRCh37. ID signatures do not take into consideration differing nucleotide composition between reference genomes because relating this to the ID mutational categories would be extremely complicated. There are also some minor differences in identifiers for stranded SBSs, and a helper function to deal with these: `SBS96_ID_to_SBS192_ID`.

Some signatures are due to experimental or laboratory artifacts. Function `possible_artifacts` returns these.

Source

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

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

Description

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

Usage

```
COSMIC_v3.0
```

Format

A list with one element `signature`, with the same structure as `signature`, except that sub-element GRCh37 does contain `SBS192`.

Remark

The signatures are all genome signatures. See examples below for transforming genome signatures to exome signatures.
See CatalogRowOrder in package ICAMS for the classification of mutation types.

Source

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

Examples

```
# As the abundances of the source sequence of the mutations vary between genome
# and exome, users can use package ICAMS to do the transformations.
if (!requireNamespace("ICAMS")) {
  install.packages("ICAMS")
  library(ICAMS)

  SBS96_sig_GRCh37_genome <- COSMIC_v3.0$signature$GRCh37$SBS96

  # Transform SBS96 GRCh37 genome signatures to GRCh37 exome signatures
  SBS96_sig_GRCh37_exome <- TransformCatalog(
    catalog = SBS96_sig_GRCh37_genome,
    target.ref.genome = "GRCh37",
    target.region = "exome"
  )
}
```

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

Description

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

Usage

```
COSMIC_v3.1
```

Format

A list with one element `signature`, with the same structure as `signature`.

Remark

The signatures are all genome signatures. See examples below for transforming genome signatures to exome signatures.
See CatalogRowOrder in package ICAMS for the classification of mutation types.

Source

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

Examples

```
# As the abundances of the source sequence of the mutations vary between genome
# and exome, users can use package ICAMS to do the transformations.
if (!requireNamespace("ICAMS")) {
  install.packages("ICAMS")
  library(ICAMS)

  SBS96_sig_GRCh37_genome <- COSMIC_v3.1$signature$GRCh37$SBS96

  # Transform SBS96 GRCh37 genome signatures to GRCh37 exome signatures
  SBS96_sig_GRCh37_exome <- TransformCatalog(
    catalog = SBS96_sig_GRCh37_genome,
    target.ref.genome = "GRCh37",
    target.region = "exome"
  )
}
```

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

Description

Mutational signatures data from COSMIC, 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 same structure as `signature`.
- `etiology` is a list with the same structure as `etiology`.

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/

Remark

The signatures are all genome signatures. See examples below for transforming genome signatures to exome signatures.
See CatalogRowOrder in package ICAMS for the classification of mutation types.

Source

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

Examples

```
# As the abundances of the source sequence of the mutations vary between genome
# and exome, users can use package ICAMS to do the transformations.
if (!requireNamespace("ICAMS")) {
  install.packages("ICAMS")
  library(ICAMS)

  SBS96_sig_GRCh37_genome <- COSMIC_v3.2$signature$GRCh37$SBS96

  # Transform SBS96 GRCh37 genome signatures to GRCh37 exome signatures
  SBS96_sig_GRCh37_exome <- TransformCatalog(
    catalog = SBS96_sig_GRCh37_genome,
    target.ref.genome = "GRCh37",
    target.region = "exome"
  )
}
```

etiology — *List of mutational signatures's proposed etiology summarized from COSMIC, Catalogue Of Somatic Mutations In Cancer (as of March 2021)*

Description

List of mutational signatures's proposed etiology summarized from COSMIC, Catalogue Of Somatic Mutations In Cancer (as of March 2021)

Usage

```
etiology
```

Format

A list with the elements:

- SBS96
- SBS192
- DBS78
- ID

Each element is a single-column matrix with rownames being the signature IDs and values being a short string describing the proposed etiology.

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

Source

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

Examples

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

get_etiology — *Get the proposed etiology of mutational signatures.*

Description

Return the known or hypothesized causes of mutational signatures. 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 available, or else the empty string.

Note

The etiology information is not versioned at the COSMIC website.

See Also

get_etiology

Examples

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

---

possible_artifacts *Return a character vector of the names of possible SBS96 signature artifacts*

Description

Return a character vector of the names of possible SBS96 signature artifacts

Usage

```
possible_artifacts()
```

Value

A character vector of the names of possible SBS96 signature artifacts.

Examples

```
artifact_sigs <- possible_artifacts()
```

rare_signatures *Return a character vector of the names of rare SBS96 signatures*

Description

Return a character vector of the names of rare SBS96 signatures

Usage

```
rare_signatures()
```

Value

A character vector of the names of rare SBS96 signatures.

Examples

```
rare_sigs <- rare_signatures()
```

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, Catalogue Of Somatic Mutations In Cancer (v3.2 - March 2021)*

Description

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

Usage

```
signature
```

Format

A list with the following elements:

```
* `GRCh37`: Homo sapiens (human) genome assembly GRCh37.
* `GRCh38`: Homo sapiens (human) genome assembly GRCh38.
* `mm9`: Mus musculus (house mouse) genome assembly mm9.
* `mm10`: Mus musculus (house mouse) genome assembly mm10.
* `rn6`: Rattus norvegicus (Norway rat) genome assembly rn6.

Each element contains the sub elements:

* `SBS96`: Strand-agnostic single-base substitutions in trinucleotide context.
* `DBS78`: Strand-agnostic doublet-base substitutions.

Element GRCh37 contains the additional sub elements:

* `SBS192`: Transcriptionally stranded single-base substitutions in trinucleotide context.
* `ID`: Strand-agnostic indels (short insertions and deletions).
```

Remark

The signatures are all genome signatures. See examples below for transforming genome signatures to exome signatures.
See CatalogRowOrder in package ICAMS 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/.

Examples

```
# As the abundances of the source sequence of the mutations vary between genome
# and exome, users can use package ICAMS to do the transformations.
if (!requireNamespace("ICAMS")) {
  install.packages("ICAMS")
  library(ICAMS)

  SBS96_sig_GRCh37_genome <- signature$GRCh37$SBS96

  # Transform SBS96 GRCh37 genome signatures to GRCh37 exome signatures
  SBS96_sig_GRCh37_exome <- TransformCatalog(
    catalog = SBS96_sig_GRCh37_genome,
    target.ref.genome = "GRCh37",
    target.region = "exome"
  )
}
```

Index