

# Package ‘cosmicsig’

December 17, 2021

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

**Version** 1.0.7

**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

**NeedsCompilation** no

**Author** Steven Rozen [aut, cre] (<<https://orcid.org/0000-0002-4288-0056>>),  
Nanhai Jiang [aut] (<<https://orcid.org/0000-0003-4974-2753>>)

**Maintainer** Steven Rozen <[steverozen@pm.me](mailto:steverozen@pm.me)>

## R topics documented:

cosmicsig . . . . .	2
COSMIC_v3.0 . . . . .	3
COSMIC_v3.1 . . . . .	3
COSMIC_v3.2 . . . . .	4
etiology . . . . .	5
get_etiology . . . . .	5
possible_artifacts . . . . .	6
rare_signatures . . . . .	7
SBS96_ID_to_SBS192_ID . . . . .	7
signature . . . . .	8

<b>Index</b>	<b>9</b>
--------------	----------

---

cosmicsig	<i>cosmicsig: A package of mutational signatures from COSMIC (Catalogue Of Somatic Mutations In Cancer)</i> <a href="https://cancer.sanger.ac.uk/signatures/">https://cancer.sanger.ac.uk/signatures/</a> .
-----------	---

---

## Description

For a general introduction to mutational signatures and the techniques used to discover them, see Alexandrov et al., 2020 doi: [10.1038/s4158602019433](https://doi.org/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 [CatalogRowOrder](#) in package ICAMS for the classification of mutation types.

**Source**

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

**Examples**

```
sbs96_sig_v3.0 <- COSMIC_v3.0$signature$GRCh37$SBS96
```

---

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 [CatalogRowOrder](#) in package ICAMS for the classification of mutation types.

**Source**

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

**Examples**

```
sbs96_sig_v3.1 <- COSMIC_v3.1$signature$GRCh37$SBS96
```

---

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

---

**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 [CatalogRowOrder](#) in package ICAMS for the classification of mutation types.

**Source**

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

**Examples**

```
sbs96_sig_v3.2 <- COSMIC_v3.2$signature$GRCh37$SBS96
```

---

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

---

**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	<i>Get the proposed etiology of mutational signatures.</i>
--------------	--

---

**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"))
```

---

<code>possible_artifacts</code>	<i>Return a character vector of the names of possible SBS96 signature artifacts</i>
---------------------------------	---

---

**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	<i>Return a character vector of the names of rare SBS96 signatures</i>
-----------------	--

---

**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	<i>Translate SBS96 signature IDs to SBS192 signature IDs by adding "-E" if necessary</i>
-----------------------	--

---

**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 [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

```
sbs96_sig <- signature$GRCh37$SBS96
```



# Index

## \* datasets

COSMIC\_v3.0, [3](#)

COSMIC\_v3.1, [3](#)

COSMIC\_v3.2, [4](#)

etiology, [5](#)

signature, [8](#)

CatalogRowOrder, [3](#), [4](#), [8](#)

COSMIC\_v3.0, [3](#)

COSMIC\_v3.1, [2](#), [3](#)

COSMIC\_v3.2, [4](#)

cosmic\_sig, [2](#)

etiology, [2](#), [4](#), [5](#)

get\_etiology, [2](#), [5](#), [5](#), [6](#)

possible\_artifacts, [2](#), [6](#)

rare\_signatures, [7](#)

SBS96\_ID\_to\_SBS192\_ID, [2](#), [7](#)

signature, [2-4](#), [8](#)