1 Introduction

The COSMIC 67 package provides the curated mutations published with the COSMIC release version 67 (2013-10-24). Both variants found in coding and non-coding regions are included and offered as a single object of class 'CollapsedVCF' and a bgzipped and tabix-index 'VCF' file.

Additionally, the package contains the Cancer Gene Census, a list of genes causally linked to cancer.

2 Accessing and Using the Data

```r
library(VariantAnnotation)

Loading required package: BiocGenerics

Attaching package: 'BiocGenerics'

The following objects are masked from 'package:stats':

  IQR, mad, sd, var, xtabs

The following objects are masked from 'package:base':
```
Filter, Find, Map, Position, Reduce, anyDuplicated, aperm, append, as.data.frame, basename, cbind, colnames, dir.name, do.call, duplicated, eval, evalq, get, grep, grepl, intersect, is.unsorted, lapply, mapply, match, mget, order, paste, pmax, pmax.int, pmin, pmin.int, rank, rbind, row.names, sapply, setdiff, table, tapply, union, unique, unsplit, which.max, which.min

Loading required package: MatrixGenerics

Loading required package: matrixStats

Attaching package: ‘MatrixGenerics’
The following objects are masked from ‘package:matrixStats’:


Loading required package: GenomeInfoDb

Loading required package: S4Vectors

Loading required package: stats4

Attaching package: ‘S4Vectors’
The following object is masked from ‘package:utils’:

findMatches

The following objects are masked from ‘package:base’:

I, expand.grid, unname

Loading required package: IRanges

Loading required package: GenomicRanges
Loading required package: SummarizedExperiment
Loading required package: Biobase

Welcome to Bioconductor

Vignettes contain introductory material; view with 'browseVignettes()'. To cite Bioconductor, see 'citation("Biobase")', and for packages 'citation("pkgname")'.

Attaching package: ‘Biobase’
The following object is masked from ‘package:MatrixGenerics’:
   rowMedians
The following objects are masked from ‘package:matrixStats’:
   anyMissing, rowMedians
Loading required package: Rsamtools
Loading required package: Biostrings
Loading required package: XVector

Attaching package: ‘Biostrings’
The following object is masked from ‘package:base’:
   strsplit

Attaching package: ‘VariantAnnotation’
The following object is masked from ‘package:base’:
   tabulate

library(GenomicRanges)

data(package = "COSMIC.67")
data(cosmic_67, package = "COSMIC.67")

tp53_range = GRanges("17", IRanges(7565097, 7590856))
vcf_path = system.file("vcf", "cosmic_67.vcf.gz", package = "COSMIC.67")
cosmicTp53 = readVcf(vcf_path, genome = "GRCh37", ScanVcfParam(which = tp53_range))
cosmicTp53
class: CollapsedVCF
dim: 5892 0
rowRanges(vcf):
   GRanges with 5 metadata columns: paramRangeID, REF, ALT, QUAL, FILTER
info(vcf):
   DataFrame with 5 columns: GENE, STRAND, CDS, AA, CNT
info(header(vcf)):
   Number Type Description
Data Provenance

3.1 COSMIC Mutations

The following steps are performed for importing and processing of the VCF data:

1. Downloading of the VCF files 'CosmicCodingMuts_v67_20131024.vcf.gz' and 'Cosmic-NonCodingVariants_v67_20131024.vcf.gz' from 'ftp://ngs.sanger.ac.uk/production/cosmic/' to 'inst/raw/'.
2. Importing of both files to R using 'readVcf'.
3. Sorting of the seqlevels and adding 'seqinfo' data for the toplevel chromosomes of 'GRCh37'.
4. Merging of both objects, sorting according to genomic position.
5. Converting the object to class VariantAnnotation::VRanges.
6. Converting the 'character' columns to 'factors'.
7. Saving the merged object to 'data/cosmic_v67_vcf.rda'.
8. Exporting the merged object as a bgzipped and tabix-indexed 'VCF' to 'inst/vcf/cosmic_v67.vcf.gz'.

3.2 Cancer Gene Census

The following steps are performed for importing and processing of the Cancer Gene Census data:

2. Import of the files as a data frame.
3. Annotation of the 'HGNC' and 'ENSEMBLID' identifiers, using the 'ENTREZ gene ID'
as query with the 'org.Hs.eg.db' object.
4. Saving the object to 'data/cgc_67.rda'.

4 Data Source

The mutation data was obtained from the Sanger Institute Catalogue Of Somatic Mutations
In Cancer web site, http://www.sanger.ac.uk/cosmic

Bamford et al (2004):
The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website.
Br J Cancer, 91,355-358.

For details on the usage and redistribution of the data, please see ftp://ftp.sanger.ac.uk/
pub/CGP/cosmic/GUIDELINES_ON_THE_USE_OF_THIS_DATA.txt.

5 References

• http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/
• http://nar.oxfordjournals.org/content/39/suppl_1/D945.long
• ftp://ftp.sanger.ac.uk/pub/CGP/cosmic/GUIDELINES_ON_THE_USE_OF_THIS_DATA.txt

6 Session Info

R version 4.4.0 beta (2024-04-15 r86425)
Platform: x86_64-pc-linux-gnu
Running under: Ubuntu 22.04.4 LTS

Matrix products: default
BLAS: /home/biocbuild/bbs-3.19-bioc/R/lib/libRblas.so
LAPACK: /usr/lib/x86_64-linux-gnu/liblapack.so.3.10.0

locale:
[1] LC_CTYPE=en_US.UTF-8     LC_NUMERIC=C
[3] LC_TIME=en_GB         LC_COLLATE=C
[5] LC_MONETARY=en_US.UTF-8 LC_MESSAGES=en_US.UTF-8
[7] LC_PAPER=en_US.UTF-8   LC_NAME=C
[9] LC_ADDRESS=C          LC_TELEPHONE=C

time zone: America/New_York
tzcode source: system (glibc)

attached base packages:
[1] stats4   stats   graphics  grDevices  utils   datasets
COSMIC 67

[7] methods base

other attached packages:
[1] VariantAnnotation_1.50.0 Rsamtools_2.20.0
[3] Biostrings_2.72.0 XVector_0.44.0
[5] SummarizedExperiment_1.34.0 Biobase_2.64.0
[7] GenomicRanges_1.56.0 GenomeInfoDb_1.40.0
[9] IRanges_2.38.0 S4Vectors_0.42.0
[11] MatrixGenerics_1.16.0 matrixStats_1.3.0
[13] BiocGenerics_0.50.0 knitr_1.46

loaded via a namespace (and not attached):
[1] SparseArray_1.4.0 bitops_1.0-7
[3] RSQLite_2.3.6 lattice_0.22-6
[5] digest_0.6.35 evaluate_0.23
[7] grid_4.4.0 fastmap_1.1.1
[9] blob_1.2.4 jsonlite_1.8.8
[11] Matrix_1.7-0 AnnotationDbi_1.66.0
[13] restfulr_0.0.15 DBI_1.2.2
[15] BiocManager_1.30.22 httr_1.4.7
[17] BSgenome_1.72.0 UCSC.utils_1.0.0
[19] XML_3.99-0.16.1 codetools_0.2-20
[21] abind_1.4-5 cli_3.6.2
[23] rlang_1.1.3 crayon_1.5.2
[25] BiocStyle_2.32.0 bit64_4.0.5
[27] cachem_1.0.8 DelayedArray_0.30.0
[29] yaml_2.3.8 GenomicFeatures_1.56.0
[31] S4Arrays_1.4.0 tools_4.4.0
[33] parallel_4.4.0 BiocParallel_1.38.0
[35] memoise_2.0.1 GenomeInfoDbData_1.2.12
[37] curl_5.2.1 png_0.1-8
[39] vctrs_0.6.5 R6_2.5.1
[41] BiocIO_1.14.0 rtracklayer_1.64.0
[43] zlibbioc_1.50.0 KEGGREST_1.44.0
[45] bit_4.0.5 highr_0.10
[47] GenomicAlignments_1.40.0 xfun_0.43
[49] rjson_0.2.21 htmltools_0.5.8.1
[51] rmarkdown_2.26 compiler_4.4.0
[53] RCurl_1.98-1.14