Package ‘toprdata’

August 19, 2022

Title Gene and Exon Data from Ensembl
Version 1.0.1
Description Gene and exon information from Ensembl genome builds GRCh38.p13 (104) and GRCh37 (v40) to use with the ‘topr’ package.
License LGPL (>= 3)
Encoding UTF-8
LazyData true
LazyDataCompression xz
RoxygenNote 7.2.0
Suggests testthat (>= 3.0.0), knitr, rmarkdown, markdown
Config/testthat/edition 3

Imports
Depends R (>= 3.5.0)
NeedsCompilation no

Author Genuity Science [cph],
Thorhildur Juliusdottir [aut, cre],
Andri Stefansson [aut]

Maintainer Thorhildur Juliusdottir <thorhildur@genuitysci.com>
Repository CRAN
Date/Publication 2022-08-19 13:40:02 UTC

R topics documented:

ENSEXONS ................................................................. 2
ENSEXONS_37 .............................................................. 2
ENSGENES ................................................................. 3
ENSGENES_37 .............................................................. 3
toprdata ................................................................. 4

Index 5

1
ENSEXONS

Ensembl exons build HG38-104-5-2

Description

https://www.ensembl.info/2021/05/05/ensembl-104-has-been-released/

Usage

ENSEXONS

Format

A data frame with 40,122 rows and 7 variables:

- **chrom** Chromosome on build version 38 (GRCh38/hg38)
- **gene_start** genetic position of gene start on build version 38
- **gene_end** genetic position of gene end on build version 38
- **gene_symbol** The name of the gene
- **exon_chromstart** genetic positions of exon start
- **exon_chromend** genetic position of exon end

Details

exons on chrY and chrM were excluded from the exon dataset

ENSEXONS_37

Ensembl exons build HG37 (GRCh37) v40 from gencode

Description

exons on chrY and chrM were excluded from the exon dataset

Usage

ENSEXONS_37

Format

A data frame with 62,739 rows and 7 variables:

- **chrom** Chromosome on build version 38 (GRCh38/hg38)
- **gene_start** genetic position of gene start on build version 38
- **gene_end** genetic position of gene end on build version 38
- **gene_symbol** The name of the gene
- **exon_chromstart** genetic positions of exon start
- **exon_chromend** genetic position of exon end
**ENSGENES**

*Ensembl genes build HG38.104-5-2*

**Description**

https://www.ensembl.info/2021/05/05/ensembl-104-has-been-released/

**Usage**

ENSGENES

**Format**

A data frame with 40,122 rows and 5 variables:

- **chrom** Chromosome on build version 38 (GRCh38/hg38)
- **gene_start** genetic position of gene start on build version 38
- **gene_end** genetic position of gene end on build version 38
- **gene_symbol** The name of the gene
- **biotype** the biotype of the gene

**Details**

genes on chrY and chrM were excluded

---

**ENSGENES_37**

*Ensembl genes build HG37 (GRCh37) v40 from gencode*

**Description**

genes on chrY and chrM were excluded

**Usage**

ENSGENES_37

**Format**

A data frame with 62,743 rows and 5 variables:

- **chrom** Chromosome on build version 38 (GRCh38/hg38)
- **gene_start** genetic position of gene start on build version 38
- **gene_end** genetic position of gene end on build version 38
- **gene_symbol** The name of the gene
- **biotype** the biotype of the gene
**Description**

Gene and exon datasets to use with the 'topr' package

**datasets**

The two datasets are:

- **ENSGENES** Gene information from genome build GRCh38
- **ENSEXONS** Exon information from genome build GRCh38
- **ENSGENES_37** Gene information from genome build GRCh37
- **ENSEXONS_37** Exon information from genome build GRCh37

**Examples**

```r
library(toprdata)
head(ENSGENES)
head(ENSEXONS)
```
Index

* datasets
  ENSEXONS, 2
  ENSEXONS_37, 2
  ENSGENES, 3
  ENSGENES_37, 3
  
  ENSEXONS, 2, 4
  ENSEXONS_37, 2, 4
  ENSGENES, 3, 4
  ENSGENES_37, 3, 4
  
toprdata, 4