

Package: enshuman (via r-universe)

November 3, 2024

Title Human Gene Annotation Data from 'Ensembl'

Version 1.0.0

Description Gene information from 'Ensembl' genome builds 'GRCh38.p14' and 'GRCh37.p13' to use with the 'topr' package. The datasets were originally downloaded from https://ftp.ensembl.org/pub/current/gtf/homo_sapiens/Homo_sapiens.GRCh38.111.gtf.gz and https://ftp.ensembl.org/pub/grch37/current/gtf/homo_sapiens/Homo_sapiens.GRCh37.87.gtf.gz and converted into the format required by the 'topr' package. See <https://github.com/totajuliusd/topr?tab=readme-ov-file#how-to-use-topr-with-other-species-than-human> to see the required format.

License LGPL (>= 3)

Encoding UTF-8

LazyData true

LazyDataCompression xz

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Suggests testthat (>= 3.0.0), knitr, rmarkdown, markdown

Config/testthat/edition 3

Depends R (>= 3.5.0)

Repository <https://totajuliusd.r-universe.dev>

RemoteUrl <https://github.com/totajuliusd/enshuman>

RemoteRef HEAD

RemoteSha 0fafd3e685132e84db47addaaf320e5938219478

Contents

enshuman	2
hg37	2
hg38	3
Index	4

enshuman	<i>enshuman</i>
----------	-----------------

Description

Human gene annotation datasets to use with the 'topr' package

datasets

The two datasets are:

- [hg38](#) Gene information from genome build GRCh38.p14
- [hg37](#) Gene information from genome build GRCh37.p13

Examples

```
library(enshuman)
head(hg38)
head(hg37)
```

hg37	<i>Ensembl genes build GRCh37.87</i>
------	--------------------------------------

Description

`wget https://ftp.ensembl.org/pub/grch37/current/gtf/homo_sapiens/Homo_sapiens.GRCh37.87.gtf.gz`

Usage

hg37

Format

A data frame with 55,882 rows and 7 variables:

chrom chromosome
gene_start genetic position of gene start
gene_end genetic position of gene end
gene_symbol The name of the gene
biotype the biotype of the gene
exon_chromstart genetic positions of exon start
exon_chromend genetic position of exon end

hg38

Ensembl genes build GRCh38.111

Description

wget https://ftp.ensembl.org/pub/current/gtf/homo_sapiens/Homo_sapiens.GRCh38.111.gtf.gz

Usage

hg38

Format

A data frame with 41,016 rows and 7 variables:

chrom chromosome
gene_start genetic position of gene start
gene_end genetic position of gene end
gene_symbol the name of the gene
biotype the biotype of the gene
exon_chromstart genetic positions of exon start
exon_chromend genetic position of exon end

Index

* **datasets**

hg37, [2](#)

hg38, [3](#)

enshuman, [2](#)

hg37, [2](#), [2](#)

hg38, [2](#), [3](#)