GeneSummary

September 21, 2022

---

loadGeneSummary   *Load gene summary table*

Description

Load gene summary table

Usage

```
loadGeneSummary(organism = 9606,
    status = c("reviewed", "predicted", "inferred", "validated", "provisional"))
```

Arguments

organism   Name of the organism (e.g. "Homo sapiens") or the corresponding taxon ID
               (e.g. 9606).

status   Status of the review of the genes.

Details

The gene summaries are extracted from RefSeq database (https://ftp.ncbi.nih.gov/refseq/release/complete/*.rna.gbff.gz).
Gene summaries are available in the "COMMENT" section of the *rna.gbff.gz files.

The Perl script used to parse *rna.gbff.gz files is available in system.file("extdata", "parse_refseq.pl",
package = "GeneSummary").

The data is based on RefSeq release 207 (https://ftp.ncbi.nih.gov/refseq/release/RELEASE_NUMBER).

Value

A data frame with the following columns:

RefSeq_accession RefSeq accession. Note there is also the version number after the accession.

Organism Full organism name.

Taxon_ID Taxon ID of the organism.

Gene_ID Entrez gene ID.

Review_status Status of the review of the gene.

Gene_summary A long description of the gene.

Examples

```
df = loadGeneSummary(9606)
dim(df)
head(df)
```

Index