

# BSgenome.Hsapiens.UCSC.hg38

February 28, 2017

---

BSgenome.Hsapiens.UCSC.hg38

*Full genome sequences for Homo sapiens (UCSC version hg38)*

---

## Description

Full genome sequences for Homo sapiens (Human) as provided by UCSC (hg38, Dec. 2013) and stored in Biostrings objects.

## Note

This BSgenome data package was made from the following source data files:

hg38.2bit from <http://hgdownload.cse.ucsc.edu/goldenPath/hg38/bigZips/>

See [?BSgenomeForge](#) and the BSgenomeForge vignette (`vignette("BSgenomeForge")`) in the **BSgenome** software package for how to make a BSgenome data package.

## Author(s)

The Bioconductor Dev Team

## See Also

- [BSgenome](#) objects and the `available.genomes` function in the **BSgenome** software package.
- [DNAString](#) objects in the **Biostrings** package.
- The BSgenomeForge vignette (`vignette("BSgenomeForge")`) in the **BSgenome** software package for how to make a BSgenome data package.

## Examples

```
BSgenome.Hsapiens.UCSC.hg38
genome <- BSgenome.Hsapiens.UCSC.hg38
seqlengths(genome)
genome$chr1 # same as genome[["chr1"]]
```

```
## -----
## Extract the upstream sequences
```

```
## -----
## The upstream sequences located in
## http://hgdownload.cse.ucsc.edu/goldenPath/hg38/bigZips/
## are based on RefSeq genes (RefSeq Genes track in the Genome Browser).
## Upstream sequences based on UCSC genes (UCSC Genes track in the
## Genome Browser) can easily be extracted from the full genome
## sequences with:

library(TxDb.Hsapiens.UCSC.hg38.knownGene)
knownGene_txdb <- TxDb.Hsapiens.UCSC.hg38.knownGene
knownGene_up1000seqs <- extractUpstreamSeqs(genome, knownGene_txdb)

## Or, to get upstream sequences based on RefSeq genes:

refGene_txdb <- makeTxDbFromUCSC("hg38", "refGene")
refGene_up1000seqs <- extractUpstreamSeqs(genome, refGene_txdb)

## Note that you can make a TxDb object from various annotation
## resources. See the makeTxDbFromUCSC(), makeTxDbFromBiomart(), and
## makeTxDbFromGFF() functions in the GenomicFeatures package for more
## information.
## IMPORTANT: Make sure you use a TxDb package (or TxDb object) that
## contains a gene model based on hg38 or on a compatible genome (i.e.
## a genome with sequences identical to the sequences in hg38). See
## ?extractUpstreamSeqs in the GenomicFeatures package for more
## information.

## -----
## Genome-wide motif searching
## -----
## See the GenomeSearching vignette in the BSgenome software
## package for some examples of genome-wide motif searching using
## Biostrings and the BSgenome data packages:
if (interactive())
  vignette("GenomeSearching", package="BSgenome")
```

# Index

\*Topic **data**

BSgenome.Hsapiens.UCSC.hg38, [1](#)

\*Topic **package**

BSgenome.Hsapiens.UCSC.hg38, [1](#)

available.genomes, [1](#)

BSgenome, [1](#)

BSgenome.Hsapiens.UCSC.hg38, [1](#)

BSgenome.Hsapiens.UCSC.hg38-package  
(BSgenome.Hsapiens.UCSC.hg38),  
[1](#)

BSgenomeForge, [1](#)

DNASTring, [1](#)

Hsapiens (BSgenome.Hsapiens.UCSC.hg38),  
[1](#)