

# Using the TENE<sup>T</sup>.AnnotationHub datasets

Rhie Lab at the University of Southern California

2024-08-28

## Abstract

This vignette describes the basic usage of the TENE<sup>T</sup>.AnnotationHub package, which contains datasets for use in the TENE<sup>T</sup> package in the form of GenomicRanges objects representing putative enhancer, promoter, and open chromatin regions from a variety of sources. See our GitHub repository (<https://github.com/rhielab/TENE<sup>T</sup>.AnnotationHub>) for more information and to view the manifest files for each of the datasets in this package. All included datasets are aligned to the human hg38 genome.

## Contents

<b>Introduction</b>	<b>1</b>
<b>Acquiring and installing TENE<sup>T</sup>.AnnotationHub</b>	<b>2</b>
<b>Loading TENE<sup>T</sup>.AnnotationHub</b>	<b>2</b>
<b>Using the included datasets</b>	<b>2</b>
<b>Included datasets</b>	<b>3</b>
ENCODE_dELS_regions . . . . .	3
ENCODE_pELS_regions . . . . .	4
ENCODE_PLS_regions . . . . .	4
TENE <sup>T</sup> _10_cancer_panel_enhancer_regions . . . . .	5
TENE <sup>T</sup> _10_cancer_panel_open_chromatin_regions . . . . .	6
TENE <sup>T</sup> _10_cancer_panel_promoter_regions . . . . .	7
TENE <sup>T</sup> _consensus_enhancer_regions . . . . .	8
TENE <sup>T</sup> _consensus_open_chromatin_regions . . . . .	9
TENE <sup>T</sup> _consensus_promoter_regions . . . . .	10
<b>Session info</b>	<b>11</b>

## Introduction

The TENE<sup>T</sup>.AnnotationHub package contains 9 GenomicRanges datasets for use in the TENE<sup>T</sup> package, which are all aligned to the hg38 human genome. These datasets include regions of putative enhancers, promoters, and open chromatin such as the ENCODE Registry of cCREs V3 datasets, consensus datasets derived from a wide variety of tissues, cells, and patient samples from sources including Roadmap Epigenomics ChromHMM annotations, FANTOM5 putative enhancers, the ENCODE DNaseI Hypersensitive Site Master List, and TCGA tumor samples, as well as datasets relevant to 10 unique cancer types (BLCA, BRCA, COAD, ESCA, HNSC, KIRP, LIHC, LUAD, LUSC, and THCA) we personally curated from hundreds of GEO datasets and relevant TCGA tumor samples (see Mullen et al).

Manifests for the last two categories of datasets, which are derived from a variety of different sources instead of a single source, are available in the data-raw subdirectory of the package GitHub repository

(<https://github.com/rhielab/TENET.AnnotationHub/tree/dev/data-raw>). These manifests detail, among other information, the ENCODE/GEO experiments where the files originate.

The raw `.bed.gz` and `.narrowPeak.gz` files we downloaded/processed, respectively, are available in a separate `TENET.AnnotationHub_files` repository at [https://github.com/rhielab/TENET.AnnotationHub\\_files](https://github.com/rhielab/TENET.AnnotationHub_files), which also contains copies of the manifests for the datasets containing these files.

## Acquiring and installing `TENET.AnnotationHub`

R 4.4 or a newer version is required.

On Ubuntu 22.04, installation was successful in a fresh R environment after adding the official R Ubuntu repository using the instructions at <https://cran.r-project.org/bin/linux/ubuntu/> and running the following command in a terminal:

```
sudo apt-get install r-base-core r-base-dev libcurl4-openssl-dev libfreetype6-dev  
libfribidi-dev libfontconfig1-dev libharfbuzz-dev libtiff5-dev libxml2-dev
```

No dependencies other than R are required on macOS or Windows.

Two versions of this package are available.

To install the stable version from Bioconductor, start R and run:

```
## Install BiocManager, which is required to install packages from Bioconductor  
if (!requireNamespace("BiocManager", quietly = TRUE)) {  
  install.packages("BiocManager")  
}  
  
BiocManager::install("TENET.AnnotationHub", version = "devel")
```

The development version containing the most recent updates is available from our GitHub repository (<https://github.com/rhielab/TENET.AnnotationHub>).

To install the development version from GitHub, start R and run:

```
## Install prerequisite packages to install the development version from GitHub  
if (!requireNamespace("BiocManager", quietly = TRUE)) {  
  install.packages("BiocManager")  
}  
if (!requireNamespace("remotes", quietly = TRUE)) {  
  install.packages("remotes")  
}  
  
BiocManager::install("rhielab/TENET.AnnotationHub")
```

## Loading `TENET.AnnotationHub`

To load the `TENET.AnnotationHub` package, start R and run:

```
library(TENET.AnnotationHub)
```

## Using the included datasets

Wrapper functions are provided to allow easy access to all included datasets. Usage of each wrapper function is demonstrated below.

## Included datasets

### ENCODE\_dELS\_regions

A GRanges object containing regions of candidate cis-regulatory elements with distal enhancer-like signatures as identified by the ENCODE SCREEN project. These consist of regions with high H3K27ac and DNase signal, but low H3K4me3 signal, and located more than 2kb from GENCODE transcription start sites. **Citation:** ENCODE Project Consortium; Moore JE, Purcaro MJ, Pratt HE, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020 Jul;583(7818):699-710. doi: 10.1038/s41586-020-2493-4. Epub 2020 Jul 29. Erratum in: Nature. 2022 May;605(7909):E3. PMID: 32728249; PMCID: PMC7410828.

This dataset consists of 786,756 ranges and has no metadata columns.

```
## Retrieve the AnnotationHub metadata for the object
ENCODE_dELS_regions(metadata = TRUE)
#> AnnotationHub with 1 record
#> # snapshotDate(): 2024-08-01
#> # names(): AH116727
#> # $dataprovder: ENCODE
#> # $species: Homo sapiens
#> # $rdataclass: GRanges
#> # $rdatadateadded: 2024-04-29
#> # $title: ENCODE_dELS_regions
#> # $description: A GRanges object containing regions of candidate cis-regulat...
#> # $taxonomyid: 9606
#> # $genome: hg38
#> # $sourcetype: BED
#> # $sourceurl: https://screen.encodeproject.org/
#> # $sourcesize: NA
#> # $tags: c("DnaseSeq", "ENCODE", "GenomicSequence", "H3K27ac",
#> # "Homo_sapiens", "peaks")
#> # retrieve record with 'object[["AH116727"]]'

## Retrieve the object itself
ENCODE_dELS_regions()
#> require("GenomicRanges")
#> GRanges object with 786756 ranges and 0 metadata columns:
#>           seqnames           ranges strand
#>           <Rle>             <IRanges> <Rle>
#> [1]      chr1      271227-271468      *
#> [2]      chr1      274330-274481      *
#> [3]      chr1      605331-605668      *
#> [4]      chr1      727122-727350      *
#> [5]      chr1      807737-807916      *
#> ...           ...           ...           ...
#> [786752] chrY 26319537-26319816      *
#> [786753] chrY 26319848-26320155      *
#> [786754] chrY 26363541-26363721      *
#> [786755] chrY 26670949-26671287      *
#> [786756] chrY 26671293-26671478      *
#> -----
#> seqinfo: 24 sequences from an unspecified genome; no seqlengths
```

## ENCODE\_pELS\_regions

A GenomicRanges dataset of proximal enhancer-like elements from the ENCODE A GRanges object containing regions of candidate cis-regulatory elements with proximal enhancer-like signatures as identified by the ENCODE SCREEN project. These consist of regions with high H3K27ac and DNase signal, but low H3K4me3 signal, and located 2kb or less from GENCODE transcription start sites. **Citation:** ENCODE Project Consortium; Moore JE, Purcaro MJ, Pratt HE, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020 Jul;583(7818):699-710. doi: 10.1038/s41586-020-2493-4. Epub 2020 Jul 29. Erratum in: Nature. 2022 May;605(7909):E3. PMID: 32728249; PMCID: PMC7410828.

This dataset consists of 171,292 ranges and has no metadata columns.

```
## Retrieve the AnnotationHub metadata for the object
ENCODE_pELS_regions(metadata = TRUE)
#> AnnotationHub with 1 record
#> # snapshotDate(): 2024-08-01
#> # names(): AH116728
#> # $dataprovder: ENCODE
#> # $species: Homo sapiens
#> # $rdataclass: GRanges
#> # $rdatadateadded: 2024-04-29
#> # $title: ENCODE_pELS_regions
#> # $description: A GRanges object containing regions of candidate cis-regulat...
#> # $taxonomyid: 9606
#> # $genome: hg38
#> # $sourcetype: BED
#> # $sourceurl: https://screen.encodeproject.org/
#> # $sourcesize: NA
#> # $tags: c("DnaseSeq", "ENCODE", "GenomicSequence", "H3K27ac",
#> # "Homo_sapiens", "peaks")
#> # retrieve record with 'object[["AH116728"]]'

## Retrieve the object itself
ENCODE_pELS_regions()
#> GRanges object with 171292 ranges and 0 metadata columns:
#>           seqnames           ranges strand
#>           <Rle>             <IRanges> <Rle>
#> [1]      chr1      138867-139134      *
#> [2]      chr1      779223-779432      *
#> [3]      chr1      779736-780028      *
#> [4]      chr1      804782-805061      *
#> [5]      chr1      817974-818323      *
#> ...           ...           ...           ...
#> [171288] chrY 26352858-26353207      *
#> [171289] chrY 26353316-26353520      *
#> [171290] chrY 26359880-26360225      *
#> [171291] chrY 26408967-26409315      *
#> [171292] chrY 26409364-26409710      *
#> -----
#> seqinfo: 24 sequences from an unspecified genome; no seqlengths
```

## ENCODE\_PLS\_regions

A GRanges object containing regions of candidate cis-regulatory elements with promoter-like signatures as identified by the ENCODE SCREEN project. These consist of regions with high H3K4me3 and DNase

signal, and located within 200 bp of a GENCODE transcription start site. **Citation:** ENCODE Project Consortium; Moore JE, Purcaro MJ, Pratt HE, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020 Jul;583(7818):699-710. doi: 10.1038/s41586-020-2493-4. Epub 2020 Jul 29. Erratum in: *Nature*. 2022 May;605(7909):E3. PMID: 32728249; PMCID: PMC7410828.

This dataset consists of 40,734 ranges and has no metadata columns.

```
## Retrieve the AnnotationHub metadata for the object
ENCODE_PLS_regions(metadata = TRUE)
#> AnnotationHub with 1 record
#> # snapshotDate(): 2024-08-01
#> # names(): AH116729
#> # $dataprotider: ENCODE
#> # $species: Homo sapiens
#> # $rdataclass: GRanges
#> # $rdatadateadded: 2024-04-29
#> # $title: ENCODE_PLS_regions
#> # $description: A GRanges object containing regions of candidate cis-regulat...
#> # $taxonomyid: 9606
#> # $genome: hg38
#> # $sourcetype: BED
#> # $sourceurl: https://screen.encodeproject.org/
#> # $sourcesize: NA
#> # $tags: c("DnaseSeq", "ENCODE", "GenomicSequence", "H3K4me3",
#> # "Homo_sapiens", "peaks")
#> # retrieve record with 'object[["AH116729"]]'

## Retrieve the object itself
ENCODE_PLS_regions()
#> GRanges object with 40734 ranges and 0 metadata columns:
#>      seqnames      ranges strand
#>      <Rle>         <IRanges> <Rle>
#> [1] chr1      778571-778919      *
#> [2] chr1      779027-779180      *
#> [3] chr1      817081-817403      *
#> [4] chr1      827418-827767      *
#> [5] chr1      870121-870448      *
#> ...      ...      ...      ...
#> [40730] chrY 19503067-19503228      *
#> [40731] chrY 19567039-19567384      *
#> [40732] chrY 19744632-19744981      *
#> [40733] chrY 20574962-20575286      *
#> [40734] chrY 20575518-20575867      *
#> -----
#> seqinfo: 24 sequences from an unspecified genome; no seqlengths
```

### TENET\_10\_cancer\_panel\_enhancer\_regions

A composite GRanges object containing regions of putative enhancer elements from 10 different cancer types (BRCA, BLCA, COAD, ESCA, HNSC, KIRP, LIHC, LUAD, LUSC, and THCA) primarily for use in the TENET Bioconductor package. This dataset is composed of H3K27ac and H3K4me1 peaks from ChIP-seq datasets collected from Cistrome.org and processed using the ENCODE pipelines. For additional information on component datasets, see the manifest file hosted at [https://github.com/rhielab/TENET.AnnotationHub/blob/devel/data-raw/TENET\\_10\\_cancer\\_panel\\_enhancer\\_regions\\_manifest.tsv](https://github.com/rhielab/TENET.AnnotationHub/blob/devel/data-raw/TENET_10_cancer_panel_enhancer_regions_manifest.tsv).

The peaks within each cancer type are reduced, but the final dataset with peaks across all 10 cancer types is not reduced, and consists of 4,798,784 ranges, with a single metadata column TYPE, which lists which of the ten cancer types each range represents.

```
## Retrieve the AnnotationHub metadata for the object
TENET_10_cancer_panel_enhancer_regions(metadata = TRUE)
#> AnnotationHub with 1 record
#> # snapshotDate(): 2024-08-01
#> # names(): AH116721
#> # $dataprovder: NA
#> # $species: Homo sapiens
#> # $rdataclass: GRanges
#> # $rdatadateadded: 2024-04-29
#> # $title: TENET_10_cancer_panel_enhancer_regions
#> # $description: A composite GRanges object containing regions of putative en...
#> # $taxonomyid: 9606
#> # $genome: hg38
#> # $sourcetype: Multiple
#> # $sourceurl: https://github.com/rhielab/TENET.AnnotationHub/raw/devel/data-...
#> # $sourcesize: NA
#> # $tags: c("ChIPseq", "ENCODE", "GenomicSequence", "GEO",
#> # "Homo_sapiens", "peaks", "TENET")
#> # retrieve record with 'object[["AH116721"]]'

## Retrieve the object itself
TENET_10_cancer_panel_enhancer_regions()
#> GRanges object with 4798784 ranges and 1 metadata column:
#>           seqnames           ranges strand |           TYPE
#>           <Rle>             <IRanges> <Rle> | <character>
#> [1]      chr1      181470-181624      * |      BLCA
#> [2]      chr1      605275-605445      * |      BLCA
#> [3]      chr1      777756-778239      * |      BLCA
#> [4]      chr1      778309-779527      * |      BLCA
#> [5]      chr1      779619-780128      * |      BLCA
#> ...           ...           ...           ...           ...
#> [4798780] chrY 19744006-19745833      * |      THCA
#> [4798781] chrY 20559489-20559923      * |      THCA
#> [4798782] chrY 20574894-20575499      * |      THCA
#> [4798783] chrY 20575745-20576459      * |      THCA
#> [4798784] chrY 22731803-22732201      * |      THCA
#> -----
#> seqinfo: 25 sequences from an unspecified genome; no seqlengths
```

### TENET\_10\_cancer\_panel\_open\_chromatin\_regions

A composite GRanges object containing regions of open chromatin from 10 different cancer types (BRCA, BLCA, COAD, ESCA, HNSC, KIRP, LIHC, LUAD, LUSC, and THCA) primarily for use in the TENET Bioconductor package. This dataset is composed of peaks from DNase I and ATAC-seq datasets collected from Cistrome.org and processed using the ENCODE guidelines, along with additional TCGA ATAC-seq peaks from cancer samples of these ten types. For additional information on component datasets, see the manifest file hosted at [https://github.com/rhielab/TENET.AnnotationHub/blob/devel/data-raw/TENET\\_10\\_cancer\\_panel\\_open\\_chromatin\\_regions\\_manifest.tsv](https://github.com/rhielab/TENET.AnnotationHub/blob/devel/data-raw/TENET_10_cancer_panel_open_chromatin_regions_manifest.tsv).

The peaks within each cancer type are reduced, but the final dataset with peaks across all 10 cancer types is not reduced, and consists of 7,514,441 ranges, with a single metadata column TYPE, which lists which of

the ten cancer types each range represents.

```
## Retrieve the AnnotationHub metadata for the object
TENET_10_cancer_panel_open_chromatin_regions(metadata = TRUE)
#> AnnotationHub with 1 record
#> # snapshotDate(): 2024-08-01
#> # names(): AH116722
#> # $dataProvider: NA
#> # $species: Homo sapiens
#> # $rdataclass: GRanges
#> # $rdatadateadded: 2024-04-29
#> # $title: TENET_10_cancer_panel_open_chromatin_regions
#> # $description: A composite GRanges object containing regions of open chroma...
#> # $taxonomyid: 9606
#> # $genome: hg38
#> # $sourcetype: Multiple
#> # $sourceurl: https://github.com/rhielab/TENET.AnnotationHub/raw/devel/data-...
#> # $sourcesize: NA
#> # $tags: c("ENCODE", "GenomicSequence", "GEO", "Homo_sapiens", "peaks",
#> # "TCGA", "TENET")
#> # retrieve record with 'object[["AH116722"]]'

## Retrieve the object itself
TENET_10_cancer_panel_open_chromatin_regions()
#> GRanges object with 7514441 ranges and 1 metadata column:
#>           seqnames      ranges strand |      TYPE
#>           <Rle>        <IRanges> <Rle> | <character>
#> [1]      chr1      10368-10523      * |      BLCA
#> [2]      chr1      28641-28790      * |      BLCA
#> [3]      chr1      63263-63412      * |      BLCA
#> [4]      chr1      63833-63982      * |      BLCA
#> [5]      chr1     121325-121474      * |      BLCA
#> ...           ...           ...           ...           ...
#> [7514437] chrM      11188-12155      * |      THCA
#> [7514438] chrM      12400-12982      * |      THCA
#> [7514439] chrM      13089-13434      * |      THCA
#> [7514440] chrM      13519-14588      * |      THCA
#> [7514441] chrM      14731-16554      * |      THCA
#> -----
#> seqinfo: 25 sequences from an unspecified genome; no seqlengths
```

## TENET\_10\_cancer\_panel\_promoter\_regions

A composite GRanges object containing regions of putative promoter elements from 10 different cancer types (BRCA, BLCA, COAD, ESCA, HNSC, KIRP, LIHC, LUAD, LUSC, and THCA) primarily for use in the TENET Bioconductor package. This dataset is composed of H3K27me3 peaks from ChIP-seq datasets collected from Cistrome.org and processed using the ENCODE guidelines. For additional information on component datasets, see the manifest file hosted at [https://github.com/rhielab/TENET.AnnotationHub/blob/devel/data-raw/TENET\\_10\\_cancer\\_panel\\_promoter\\_regions\\_manifest.tsv](https://github.com/rhielab/TENET.AnnotationHub/blob/devel/data-raw/TENET_10_cancer_panel_promoter_regions_manifest.tsv).

The peaks within each cancer type are reduced, but the final dataset with peaks across all 10 cancer types is not reduced, and consists of 2,627,647 ranges, with a single metadata column TYPE, which lists which of the ten cancer types each range represents.



```

## Retrieve the AnnotationHub metadata for the object
TENET_10_cancer_panel_promoter_regions(metadata = TRUE)
#> AnnotationHub with 1 record
#> # snapshotDate(): 2024-08-01
#> # names(): AH116723
#> # $dataProvider: NA
#> # $species: Homo sapiens
#> # $rdataclass: GRanges
#> # $rdatadateadded: 2024-04-29
#> # $title: TENET_10_cancer_panel_promoter_regions
#> # $description: A composite GRanges object containing regions of putative pr...
#> # $taxonomyid: 9606
#> # $genome: hg38
#> # $sourcetype: Multiple
#> # $sourceurl: https://github.com/rhielab/TENET.AnnotationHub/raw/develop/data-...
#> # $sourcesize: NA
#> # $tags: c("ChIPseq", "ENCODE", "GenomicSequence", "GEO", "H3K4me3",
#> # "Homo_sapiens", "peaks", "TENET")
#> # retrieve record with 'object[["AH116723"]]'

## Retrieve the object itself
TENET_10_cancer_panel_promoter_regions()
#> GRanges object with 2627647 ranges and 1 metadata column:
#>           seqnames           ranges strand |           TYPE
#>           <Rle>           <IRanges> <Rle> | <character>
#> [1]      chr1      778400-778764      * |      BLCA
#> [2]      chr1      778866-779304      * |      BLCA
#> [3]      chr1      826789-826976      * |      BLCA
#> [4]      chr1      827289-827528      * |      BLCA
#> [5]      chr1      869975-870217      * |      BLCA
#> ...           ...           ...           ...           ...
#> [2627643] chrY 19076176-19077636      * |      THCA
#> [2627644] chrY 19566617-19568512      * |      THCA
#> [2627645] chrY 19743897-19745898      * |      THCA
#> [2627646] chrY 20574698-20575443      * |      THCA
#> [2627647] chrY 20575763-20576197      * |      THCA
#> -----
#> seqinfo: 25 sequences from an unspecified genome; no seqlengths

```

## TENET\_consensus\_enhancer\_regions

A composite GRanges object containing regions of putative enhancer elements from a variety of sources, primarily for use in the TENET Bioconductor package. This dataset is composed of regions of strong enhancers as annotated by the Roadmap Epigenomics ChromHMM expanded 18-state model based on 98 reference epigenomes, lifted over to the hg38 genome (the following 4 states represent strong enhancers: 7: Genic enhancer1, 8: Genic enhancer2, 9: Active Enhancer 1, and 10: Active Enhancer 2), as well as regions of human permissive enhancers identified by the FANTOM5 project in phase 1 and phase 2. For additional information on component datasets, see the manifest file hosted at [https://github.com/rhielab/TENET.AnnotationHub/blob/develop/data-raw/TENET\\_consensus\\_datasets\\_manifest.tsv](https://github.com/rhielab/TENET.AnnotationHub/blob/develop/data-raw/TENET_consensus_datasets_manifest.tsv). **Citations:** Roadmap Epigenomics Consortium; Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015 Feb 19;518(7539):317-30. doi: 10.1038/nature14248. PMID: 25693563; PMCID: PMC4530010. Lizio M, Harshbarger J, Shimoji H, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol* 16(1), 22 (2015). Abugessaisa I, Ramilowski



JA, Lizio M, et al. FANTOM enters 20th year: expansion of transcriptomic atlases and functional annotation of non-coding RNAs. *Nucleic Acids Res.* 2021 Jan 8;49(D1):D892-D898. doi: 10.1093/nar/gkaa1054. PMID: 33211864; PMCID: PMC7779024.

This dataset consists of 403,602 reduced ranges and has no metadata columns.

```
## Retrieve the AnnotationHub metadata for the object
TENET_consensus_enhancer_regions(metadata = TRUE)
#> AnnotationHub with 1 record
#> # snapshotDate(): 2024-08-01
#> # names(): AH116724
#> # $dataprovder: NA
#> # $species: Homo sapiens
#> # $rdataclass: GRanges
#> # $rdatadateadded: 2024-04-29
#> # $title: TENET_consensus_enhancer_regions
#> # $description: A composite GRanges object containing regions of putative en...
#> # $taxonomyid: 9606
#> # $genome: hg38
#> # $sourcetype: Multiple
#> # $sourceurl: https://github.com/rhielab/TENET.AnnotationHub/raw/devel/data-...
#> # $sourcesize: NA
#> # $tags: c("EpigenomeRoadMap", "FANTOM5", "GenomicSequence",
#> # "Homo_sapiens", "TENET")
#> # retrieve record with 'object[["AH116724"]]'

## Retrieve the object itself
TENET_consensus_enhancer_regions()
#> GRanges object with 403602 ranges and 0 metadata columns:
#>           seqnames           ranges strand
#>           <Rle>             <IRanges> <Rle>
#> [1]      chr1             10001-10400      *
#> [2]      chr1             14801-15200      *
#> [3]      chr1             16001-16600      *
#> [4]      chr1             20001-20400      *
#> [5]      chr1             79001-79800      *
#> ...           ...           ...           ...
#> [403598] chrY 56856054-56856253      *
#> [403599] chrY 56873201-56873229      *
#> [403600] chrY 56873666-56873879      *
#> [403601] chrY 56879705-56879876      *
#> [403602] chrM             1-16398        *
#> -----
#> seqinfo: 25 sequences from an unspecified genome; no seqlengths
```

### TENET\_consensus\_open\_chromatin\_regions

A composite GRanges object containing regions of open chromatin from a variety of sources, primarily for use in the TENET Bioconductor package. This dataset is composed of DNase I hypersensitive regions from the master list compiled from 125 cell types by ENCODE, lifted over to the hg38 genome, along with TCGA ATAC-seq peaks from 410 cancer samples of 23 cancer types. For additional information on component datasets, see the manifest file hosted at [https://github.com/rhielab/TENET.AnnotationHub/blob/devel/data-raw/TENET\\_consensus\\_datasets\\_manifest.tsv](https://github.com/rhielab/TENET.AnnotationHub/blob/devel/data-raw/TENET_consensus_datasets_manifest.tsv). **Citations:** ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012 Sep 6;489(7414):57-74. doi: 10.1038/nature11247. PMID: 22955616; PMCID: PMC3439153. Thurman

RE, Rynes E, Humbert R, et al. The accessible chromatin landscape of the human genome. *Nature*. 2012 Sep 6;489(7414):75-82. doi: 10.1038/nature11232. PMID: 22955617; PMCID: PMC3721348. Corces MR, Granja JM, Shams S, et al. The chromatin accessibility landscape of primary human cancers. *Science*. 2018 Oct 26;362(6413):eaav1898. doi: 10.1126/science.aav1898. PMID: 30361341; PMCID: PMC6408149.

This dataset consists of 2,525,827 reduced ranges and has no metadata columns.

```
## Retrieve the AnnotationHub metadata for the object
TENET_consensus_open_chromatin_regions(metadata = TRUE)
#> AnnotationHub with 1 record
#> # snapshotDate(): 2024-08-01
#> # names(): AH116725
#> # $dataprotider: NA
#> # $species: Homo sapiens
#> # $rdataclass: GRanges
#> # $rdatadateadded: 2024-04-29
#> # $title: TENET_consensus_open_chromatin_regions
#> # $description: A composite GRanges object containing regions of open chroma...
#> # $taxonomyid: 9606
#> # $genome: hg38
#> # $sourcetype: Multiple
#> # $sourceurl: https://github.com/rhielab/TENET.AnnotationHub/raw/devel/data-...
#> # $sourcesize: NA
#> # $tags: c("ENCODE", "GenomicSequence", "Homo_sapiens", "TCGA",
#> # "TENET")
#> # retrieve record with 'object[["AH116725"]]'

## Retrieve the object itself
TENET_consensus_open_chromatin_regions()
#> GRanges object with 2525827 ranges and 0 metadata columns:
#>           seqnames           ranges strand
#>           <Rle>             <IRanges> <Rle>
#> [1]      chr1             10121-10270      *
#> [2]      chr1             10441-10590      *
#> [3]      chr1             16141-16290      *
#> [4]      chr1             17239-17739      *
#> [5]      chr1             20061-20210      *
#> ...           ...           ...           ...
#> [2525823] chrY 56885219-56885368      *
#> [2525824] chrY 56885479-56885628      *
#> [2525825] chrY 56886279-56886428      *
#> [2525826] chrY 56886539-56886688      *
#> [2525827] chrY 56886879-56887028      *
#> -----
#> seqinfo: 24 sequences from an unspecified genome; no seqlengths
```

## TENET\_consensus\_promoter\_regions

A composite GRanges object containing regions of putative promoter elements from a variety of sources, primarily for use in the TENET Bioconductor package. This dataset is composed of regions flanking transcription start sites as annotated by the Roadmap Epigenomics ChromHMM expanded 18-state model based on 98 reference epigenomes, lifted over to the hg38 genome (the following 4 states represent regions flanking transcription start sites: 1: Active TSS, 2: Flanking TSS, 3: Flanking TSS Upstream, and 4: Flanking TSS Downstream). For additional information on component datasets, see the manifest file hosted at [10](https://github.com/rhielab/TENET.AnnotationHub/blob/devel/data-</a></p>
</div>
<div data-bbox=)

raw/TENET\_consensus\_datasets\_manifest.tsv. **Citation:** Roadmap Epigenomics Consortium; Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015 Feb 19;518(7539):317-30. doi: 10.1038/nature14248. PMID: 25693563; PMCID: PMC4530010.

This dataset consists of 361,315 reduced ranges and has no metadata columns.

```
## Retrieve the AnnotationHub metadata for the object
TENET_consensus_promoter_regions(metadata = TRUE)
#> AnnotationHub with 1 record
#> # snapshotDate(): 2024-08-01
#> # names(): AH116726
#> # $dataprotider: NA
#> # $species: Homo sapiens
#> # $rdataclass: GRanges
#> # $rdatadateadded: 2024-04-29
#> # $title: TENET_consensus_promoter_regions
#> # $description: A composite GRanges object containing regions of putative pr...
#> # $taxonomyid: 9606
#> # $genome: hg38
#> # $sourcetype: Multiple
#> # $sourceurl: https://github.com/rhielab/TENET.AnnotationHub/raw/devel/data-...
#> # $sourcesize: NA
#> # $tags: c("EpigenomeRoadMap", "GenomicSequence", "Homo_sapiens",
#> # "TENET")
#> # retrieve record with 'object[["AH116726"]]'

## Retrieve the object itself
TENET_consensus_promoter_regions()
#> GRanges object with 361315 ranges and 0 metadata columns:
#>           seqnames           ranges strand
#>           <Rle>             <IRanges> <Rle>
#> [1]      chr1             10001-10600      *
#> [2]      chr1             16001-16200      *
#> [3]      chr1             17401-17600      *
#> [4]      chr1             20201-20400      *
#> [5]      chr1             28401-28800      *
#> ...           ...           ...           ...
#> [361311] chrY 56761271-56761670      *
#> [361312] chrY 56763271-56763670      *
#> [361313] chrY 56770671-56771070      *
#> [361314] chrY 56855254-56856053      *
#> [361315] chrM              1-16398      *
#> -----
#> seqinfo: 25 sequences from an unspecified genome; no seqlengths
```

## Session info

```
sessionInfo()
#> R version 4.4.1 (2024-06-14)
#> Platform: x86_64-pc-linux-gnu
#> Running under: Ubuntu 22.04.4 LTS
#>
#> Matrix products: default
```

```

#> BLAS: /home/biocbuild/bbs-3.20-bioc/R/lib/libRblas.so
#> LAPACK: /usr/lib/x86_64-linux-gnu/lapack/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
#> [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
#>
#> time zone: America/New_York
#> tzcode source: system (glibc)
#>
#> attached base packages:
#> [1] stats4 stats graphics grDevices utils datasets methods
#> [8] base
#>
#> other attached packages:
#> [1] GenomicRanges_1.57.1 GenomeInfoDb_1.41.1
#> [3] IRanges_2.39.2 S4Vectors_0.43.2
#> [5] BiocGenerics_0.51.0 TENET.AnnotationHub_0.99.0
#>
#> loaded via a namespace (and not attached):
#> [1] rappdirs_0.3.3 utf8_1.2.4 generics_0.1.3
#> [4] BiocVersion_3.20.0 RSQLite_2.3.7 digest_0.6.37
#> [7] magrittr_2.0.3 evaluate_0.24.0 fastmap_1.2.0
#> [10] blob_1.2.4 AnnotationHub_3.13.3 jsonlite_1.8.8
#> [13] AnnotationDbi_1.67.0 DBI_1.2.3 BiocManager_1.30.24
#> [16] httr_1.4.7 purrr_1.0.2 fansi_1.0.6
#> [19] UCSC.utils_1.1.0 Biostrings_2.73.1 cli_3.6.3
#> [22] crayon_1.5.3 rlang_1.1.4 dbplyr_2.5.0
#> [25] XVector_0.45.0 Biobase_2.65.0 bit64_4.0.5
#> [28] withr_3.0.1 cachem_1.1.0 yaml_2.3.10
#> [31] tools_4.4.1 memoise_2.0.1 dplyr_1.1.4
#> [34] GenomeInfoDbData_1.2.12 filelock_1.0.3 curl_5.2.2
#> [37] mime_0.12 vctrs_0.6.5 R6_2.5.1
#> [40] png_0.1-8 BiocFileCache_2.13.0 lifecycle_1.0.4
#> [43] zlibbioc_1.51.1 KEGGREST_1.45.1 bit_4.0.5
#> [46] pkgconfig_2.0.3 pillar_1.9.0 glue_1.7.0
#> [49] xfun_0.47 tibble_3.2.1 tidyselect_1.2.1
#> [52] knitr_1.48 htmltools_0.5.8.1 rmarkdown_2.28
#> [55] compiler_4.4.1

```