Package ‘HumanTranscriptomeCompendium’

May 4, 2024

**Title**  Tools to work with a Compendium of 181000 human transcriptome sequencing studies

**Description**  Provide tools for working with a compendium of human transcriptome sequences (originally htxcomp).

**Version**  1.20.0

**Author**  Sean Davis, Vince Carey

**Suggests**  knitr, BiocStyle, beeswarm, tximportData, DT, tximport, dplyr, magrittr, BiocFileCache, testthat, rhdf5client, markdown

**Imports**  shiny, ssrch, S4Vectors, SummarizedExperiment, utils, BiocManager

**Depends**  R (>= 3.6)

**Maintainer**  VJ Carey <stvjc@channing.harvard.edu>

**License**  Artistic-2.0

**LazyLoad**  yes

**LazyData**  yes

**biocViews**  Transcriptomics, Infrastructure

**VignetteBuilder**  knitr

**RoxygenNote**  7.2.3

**Encoding**  UTF-8

**PackageStatus**  Deprecated

**git_url**  https://git.bioconductor.org/packages/HumanTranscriptomeCompendium

**git_branch**  RELEASE_3_19

**git_last_commit**  9e20604

**git_last_commit_date**  2024-04-30

**Repository**  Bioconductor 3.19

**Date/Publication**  2024-05-03
addRD

**Description**

add gene-level rowData derived from transcript level rowRanges

**Usage**

addRD(x)

**Arguments**

x result of htx_load()

**Value**

RangedSummarizedExperiment with enhanced rowData

**Examples**

# this function operates on a SummarizedExperiment that has
# transcript-level rowRanges but gene-level quantifications
addRD
ca43k

app to survey 43000 cancer transcriptomes

Description

app to survey 43000 cancer transcriptomes

Usage

cia43k()

Value

a SummarizedExperiment

Note

Copies source code and metadata to a temporary folder and executes shiny::runApp there; sets working directory to folder where ca43k was called when app is exited. Also will return either NULL or a SummarizedExperiment at conclusion.

Examples

cia43k

get_ds4841

return instance of ssrch::DocSet with metadata on 4841 human transcriptome studies in NCBI SRA

Description

return instance of ssrch::DocSet with metadata on 4841 human transcriptome studies in NCBI SRA

Usage

get_ds4841(

cache = BiocFileCache::BiocFileCache(),
csv_zip_path = path_doc4842()
)

Arguments

cache instance of ‘BiocFileCache’, defaults to ‘BiocFileCache::BiocFileCache()’
csv_zip_path a path leading to the zip file of CSV for metadata in the DocSet instance
Value

instance of DocSet as defined in ssrch package

Note

will bind the correct value of ‘zipf’ in ‘environment(ds4841@doc_retriever)’, which depends on details of installation

Examples

get_ds4841()

htx_app  

explore SRA metadata

Description

explore SRA metadata

Usage

htx_app()

Value

a SummarizedExperiment can be requested through an event

Note

This function deals with extraction of compendium elements. The overall scope is determined by HumanTranscriptomeCompendium::studTable which is the list of all studies with taxon 9606, strategy RNA-seq, source transcriptomic. Some studies will not have experiments in the compendium, and if such are selected, a warning will be generated in the session.

Examples

if (interactive()) htx_app()
htx_load

load a SummarizedExperiment shell for the Human Transcriptome Compendium

Description

load a SummarizedExperiment shell for the Human Transcriptome Compendium

Usage

```r
htx_load(
  remotePath = NULL,
  cache = BiocFileCache::BiocFileCache(),
  genesOnly = TRUE
)
```

Arguments

- `remotePath` defaults to NULL, will be removed in future version
- `cache` a BiocFileCache instance, defaulting to value of BiocFileCache()
- `genesOnly` logical(1) if TRUE return reference to SummarizedExperiment with gene-level quantifications; in this case the remotePath value is set to ‘https://s3.amazonaws.com/bcfound-bigrna/rangedHtxGeneSE.rds’. As of Bioc 3.17, changes to S4Vectors necessitate different remote images or redesign of handling HSDS instances. Thus remotePath parameter is dropped and version-specific code is added to the loader to select an appropriate path.

Value

a RangedSummarizedExperiment instance

Examples

```r
if (interactive()) {
  h = htx_load()
  h
  SummarizedExperiment::assay(h)
}
```
htx_query_by_study_accession

retrieve 'restfulSE' SummarizedExperiment instance for selected studies in htx compendium

Description

retrieve 'restfulSE' SummarizedExperiment instance for selected studies in htx compendium

Usage

htx_query_by_study_accession(studies, htxSE, ...)

Arguments

- studies: character vector of study accessions
- htxSE: SummarizedExperiment instance, typically the result of htx_load(), which we don’t want to repeat needlessly
- ...: passed to ‘htx_load’, ignored if ‘se’ is nonmissing

Value

SummarizedExperiment instance

Note

This function was designed to perform a single query on a 'fresh' compendium image from 'htx_load()'. However, one could consider iterating the process to build up metadata on multiple series of studies. This is not likely to succeed without careful manipulation of the colData of the input SummarizedExperiment. A message will be written if the input SummarizedExperiment appears to be other than a 'fresh' 'htx_load' result.

Examples

htx_query_by_study_accession("ERP011411")

htx_query_by_text

subset compendium through keyword lookup

Description

subset compendium through keyword lookup

Usage

htx_query_by_text(query, ..., tryGrep = TRUE, ignore.case = TRUE)
**Arguments**

- `query` character(1) to be found in `ls(srch::kw2docs(get_ds4841()))` passed to `htx_query_by_study_accession`

- `tryGrep` logical(1) if TRUE, ‘query’ does not match any keyword directly, it will be treated as a regular expression and the vector of keywords will be grepped for pattern ‘query’; defaults to TRUE

- `ignore.case` logical(1) used when tryGrep is TRUE, defaults to TRUE

**Value**

SummarizedExperiment instance

**Note**

The DocSet instance returned by `get_ds4841()` is used. Lookups are case-sensitive. Look carefully at note for `htx_query_by_study_accession` to understand logic of incrementing metadata on a given input SummarizedExperiment.

**Examples**

```
htx_query_by_text("HNRNPC")
```

---

**HumanTranscriptomeCompendium.colnames**

*character vector of available samples in HDF cloud assay representation*

---

**Description**

character vector of available samples in HDF cloud assay representation

**Usage**

```
HumanTranscriptomeCompendium.colnames
```

**Format**

character vector

**Source**

compendium processing

**Examples**

```
head(HumanTranscriptomeCompendium::HumanTranscriptomeCompendium.colnames)
```
load_bigrnaFiles

**obtain listing of contents of BigRNA compendium (salmon runs)**

**Description**

obtain listing of contents of BigRNA compendium (salmon runs)

**Usage**

load_bigrnaFiles(cache = BiocFileCache::BiocFileCache())

**Arguments**

- `cache` instance of `BiocFileCache`, defaults to `BiocFileCache::BiocFileCache()`

**Value**

a named vector

**Examples**

if (interactive()) head(load_bigrnaFiles())

load_experTable

**obtain listing of experiments and submission date/time in compendium**

**Description**

obtain listing of experiments and submission date/time in compendium

**Usage**

load_experTable(cache = BiocFileCache::BiocFileCache())

**Arguments**

- `cache` instance of `BiocFileCache`, defaults to `BiocFileCache::BiocFileCache()`

**Value**

a data.frame

**Examples**

if (interactive()) head(load_experTable())
**load_studTable**

---

| load_studTable | obtain listing of all studies in compendium |

---

**Description**

obtain listing of all studies in compendium

**Usage**

```r
load_studTable(cache = BiocFileCache::BiocFileCache())
```

**Arguments**

- `cache`: instance of `BiocFileCache`, defaults to `BiocFileCache::BiocFileCache()`

**Value**

a data.frame

**Examples**

```r
if (interactive()) head(load_studTable())
```

---

**path_doc4842**

---

| path_doc4842 | return path to metadata csvs in zip file |

---

**Description**

return path to metadata csvs in zip file

**Usage**

```r
path_doc4842(cache = BiocFileCache::BiocFileCache())
```

**Arguments**

- `cache`: instance of `BiocFileCache`, defaults to `BiocFileCache::BiocFileCache()`

**Value**

path to zipfile

**Note**

CSVs were retrieved using methods provided at [https://api-omicidx.cancerdatasci.org/sra/1.0/ui/](https://api-omicidx.cancerdatasci.org/sra/1.0/ui/) and zipped together. Function will lodge zipfile in `cache` if not present.
procExpToGene

Examples

path_doc4842()

procExpToGene  acquire a single sample from bigRNA compendium specified by accession and develop gene-level quantifications using tximport

Description

acquire a single sample from bigRNA compendium specified by accession and develop gene-level quantifications using tximport

Usage

procExpToGene(
  acc,
  tx2gene = tx2gene_gencode27(),
  urlprefix = "http://bigrna-test.cancerdatasci.org/data/?accession=",
  manifestdata = HumanTranscriptomeCompendium::load_bigrnaFiles(),
  regexp = "quant.sf.bz2|json"
)

Arguments

acc character(1) sample-level accession as defined in SRA

tax2gene a data.frame instance mapping transcript identifiers used in the compendium to gene identifiers. See note.

urlprefix character(1) where the salmon run outputs are lodged, with acc a subfolder defined through the manifestData parameter.

manifestdata a character vector defining folders (under results/human/27/ with salmon outputs.

regexp a character(1) regular expression for filtering filename elements in manifestdata to define which salmon output components in the bigrna compendium are retrieved.

Value

the result of a tximport run

Note

The tx2gene_gencode function supplied with this package uses the tximportData package contents to create the data.frame for use as tx2gene. The system2 function is used to generate folders to be used by tximport.
**Examples**

```r
# this example involves nontrivial internet communications
args(procExpToGene)

td = tempdir()
od = getwd()
setwd(td)
nn = procExpToGene(“ERX1097381”)
str(nn)
setwd(od)
```

---

tx2gene_gencode27 generate a data.frame mapping gencode 27 ensembl transcript identifiers to ensembl gene identifiers

---

**Description**

generate a data.frame mapping gencode 27 ensembl transcript identifiers to ensembl gene identifiers

**Usage**

```r
tx2gene_gencode27()
```

**Value**

a data.frame with 200401 rows mapping transcript identifiers in column 1 to 58288 gene symbols in column 2.

**Note**

Uses CSV in tximportData to acquire the information.

**Examples**

```r
head(tx2gene_gencode27())
```
| uniqueAcc_120518 | experiment accessions available in compendium as of may 12 2018 |

**Description**

experiment accessions available in compendium as of may 12 2018

**Usage**

uniqueAcc_120518

**Format**

data.frame

**Source**

SRAdbV2 may 12 2018

**Examples**

head(HumanTranscriptomeCompendium::uniqueAcc_120518)
Index

* datasets
  HumanTranscriptomeCompendium.colnames, 7
  uniqueAcc_120518, 12

addRD, 2
ca43k, 3
get_ds4841, 3
htx_app, 4
htx_load, 5
htx_query_by_study_accession, 6
htx_query_by_text, 6
HumanTranscriptomeCompendium.colnames, 7

load_bigrnaFiles, 8
load_experTable, 8
load_studTable, 9
path_doc4842, 9
procExpToGene, 10
tx2gene_gencode27, 11
uniqueAcc_120518, 12