Package 'ChIC.data'

April 14, 2020

Title ChIC package data

Version 1.6.0

Maintainer Carmen Maria Livi <carmen.livi@ifom.eu>

Description This package contains annotation and metagene profile data for the ChIC package.

Depends R (>= 3.5)

Imports caret (>= 6.0-78)

biocViews ExperimentData, ENCODE

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

NeedsCompilation no

git_url https://git.bioconductor.org/packages/ChIC.data

git_branch RELEASE_3_10

git_last_commit 53ac712

git_last_commit_date 2019-10-29

Date/Publication 2020-04-14

Author Carmen Maria Livi [aut, cre, dtc],
Endre Sebestyen [aut]

R topics documented:

Index   **13**

---

chipSubset   *ChIP-seq bam file stored as spp tag-list for a subset of chromosomes*
                            for the chip

Description

Example data for manual and vignette. tag-list created with the read.bam.tags() function from spp
package. The original bam file has been dowloaded from ENCODE (ID: ENCFF000BFX).

Usage

```
data(chipSubset)
```

Format

list of 2 elements containing the reads and the read quality of the ChIP.

- tags : list containing the start coordinates of each read aligned (ChIP) (3'end)
- quality: list containing the read quality of each read

---

classesDefList   *List with definitions*

Description

List containing information about chromatin marks, transcription factors, definitions for broad and
sharp class.

Usage

```
data(classesDefList)
```

Format

A list of 5 containing a vector with chromatin mark names, TF names, the members of sharp binding
marks, broad binding marks and RNApol2

- Hlist: vector with chromatin marks
- allSharp: vector with sharp binding marks
- allBroad: vector with broad binding marks
- RNAPol2 : RNApol2
- TF: vector with TF names

---

compendium_db   *Histone mark compendium*

Description

Histone mark ChIP-seq compendium for the ChIC package. The compendium contains quality control metrics and metadata for 2329 histone mark samples analysed from ENCODE and Roadmap Epigenomics.

Usage

```
data(compendium_db)
```

Format

Data frame with 366 variables (quality control metrics, metadata) for each analysed sample.

Source

XX

---

compendium_db_tf   *Transcription factor compendium*

Description

Transcription factor ChIP-seq compendium for the ChIC package. The compendium contains quality control metrics and metadata for 1427 transcription factors analysed from ENCODE.

Usage

```
data(compendium_db_tf)
```

Format

Data frame with 366 variables (quality control metrics, metadata) for each analysed sample.

Source

XX

---

compendium_profiles *Metagene profile data for chromatin marks*

Description

Compendium of averaged metagene profiles for the ChIC package. Contains averaged metagene profiles for following ChIP-seq datasets from ENCODE and Roadmap Epigenomics.

Usage

```
data(compendium_profiles)
```

Format

A list of data frames with the coordinates of the metagene profile of the respective chromatin mark.

x : genomic coordinates mean : the mean of the signal intensity in the compendium sd : the standard deviation of the signal intensity in the compendium q1..q5 : being the respective quantile of the value distribution sderr : standard error

Details

- H2A.Z
- H2AFZ
- H2AK5ac
- H2AK9ac
- H2BK120ac
- H2BK12ac
- H2BK15ac
- H2BK20ac
- H2BK5ac
- H3K14ac
- H3K18ac
- H3K23ac
- H3K23me2
- H3K27ac
- H3K27me3
- H3K36me3
- H3K4ac
- H3K4me1
- H3K4me2
- H3K4me3
- H3K56ac
- H3K79me1
- H3K79me2

- H3K9ac
- H3K9me1
- H3K9me3
- H3T11ph
- H4K12ac
- H4K20me1
- H4K5ac
- H4K8ac
- H4K91ac
- POLR2A
- POLR2AphosphoS2
- POLR2AphosphoS5

Source

XX

---

compendium_profiles_TF

Metagene profile data for transcription factors

Description

Compendium of averaged metagene profiles for the ChIC package. Contains averaged metagene profiles for following ChIP-seq datasets from ENCODE.

Usage

```
data(compendium_profiles_TF)
```

Format

A list of data frames with the coordinates of the metagene profile of the respective transcription factor.

x : genomic coordinates mean : the mean of the signal intensity in the compendium sd : the standard deviation of the signal intensity in the compendium q1..q5 : being the respective quantile of the value distribution sderr : standard error

Details

- ARID3A
- ATF2
- ATF3
- BACH1
- BATF
- BCL11A

- BCL3
- BCLAF1
- BHLHE40
- BRCA1
- CBX2
- CBX3
- CEBPB
- CEBPD
- CEBPZ
- CHD1
- CHD2
- CHD4
- CREB1
- CREBBP
- CTCF
- CUX1
- E2F4
- E2F6
- EBF1
- EGR1
- ELF1
- ELK1
- EP300
- ESRRA
- ETS1
- EZH2
- FOSL1
- FOSL2
- FOS
- FOXA1
- FOXA2
- FOXM1
- GABPA
- GATA2
- GATA3
- GTF2F1
- HA
- HDAC2
- HDAC6
- HNF4A

- HNF4G
- IRF3
- IRF4
- JUND
- JUN
- KAT2A
- KAT2B
- KDM1A
- KDM4A
- KDM5A
- KDM5B
- MAFF
- MAFK
- MAX
- MAZ
- MBD4
- MEF2A
- MEF2C
- MTA3
- MXI1
- MYBL2
- MYC
- NANOG
- NCOR1
- NFATC1
- NFIC
- NR2F2
- NR3C1
- NRF1
- PAX5
- PBX3
- PHF8
- PML
- POU2F2
- PRDM1
- RAD21
- RBBP5
- RCOR1
- REST
- RFX5

- RNF2
- RUNX3
- RXRA
- SAP30
- SETDB1
- SIN3A
- SIRT6
- SIX5
- SMC3
- SP1
- SP2
- SP4
- SPI1
- SREBF1
- SREBF2
- SRF
- STAT3
- STAT5A
- SUPT20H
- SUZ12
- TAF1
- TAF7
- TAL1
- TBP
- TCF12
- TCF3
- TEAD4
- USF1
- USF2
- WRNIP1
- YY1
- ZBTB33
- ZBTB7A
- ZC3H11A
- ZEB1
- ZMIZ1
- ZNF143
- ZNF274
- ZNF384

Source

XX

---

crossvalues_Chip *CrossCorrelation values for example ChIP-seq data for Vignette*

Description

Example data to be used in the vignette for chrom2

Usage

```
data(crossvalues_Chip)
```

Format

list of 20 elements containing EM scores

---

dm3_chrom_info *dm3 chromosome information*

Description

dm3 chromosome information for the ChIC package.

Usage

```
data(dm3_chrom_info)
```

Format

A named list of int vectors with 2 elements, the start and end position of all dm3 chromosomes.

Source

http://hgdownload.cse.ucsc.edu/goldenPath/dm3/database/ chromInfo.txt.gz

---

dm3_refseq_genes_filtered_granges
Filtered RefSeq gene annotation as GRanges for dm3

Description

RefSeq gene annotation data for the ChIC package in GRanges format, filtered by the gene length and by overlaps.

Usage

```
data(dm3_refseq_genes_filtered_granges)
```

Format

A GRanges object.

hg19_chrom_info   *hg19 chromosome information*

Description

hg19 chromosome information for the ChIC package.

Usage

```
data(hg19_chrom_info)
```

Format

A named list of int vectors with 2 elements, the start and end position of all hg19 chromosomes.

Source

http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/ hg19.chrom.sizes

hg19_refseq_genes_filtered_granges

Filtered RefSeq gene annotation as GRanges

Description

RefSeq gene annotation data for the ChIC package in GRanges format, filtered by the gene length and by overlaps.

Usage

```
data(hg19_refseq_genes_filtered_granges)
```

Format

A GRanges object.

inputSubset	*ChIP-seq bam file stored as spp tag-list for a subset of chromosomes for the input*

Description

Example data for manual and vignette. Tag-list created with the read.bam.tags() function from spp package for the input data. The original bam file has been dowloaded from ENCODE (ID: ENCFF000BDQ).

Usage

```
data(inputSubset)
```

Format

list of 2 elements containing the reads and the read quality of the input.

- tags : list containing the start coordinates of each read aligned (input) (3'end)
- quality: list containing the read quality of each read

mm9_chrom_info	*mm9 chromosome information*

Description

mm9 chromosome information for the ChIC package.

Usage

```
data(mm9_chrom_info)
```

Format

A named list of int vectors with 2 elements, the start and end position of all mm9 chromosomes.

Source

http://hgdownload.cse.ucsc.edu/goldenPath/mm9/bigZips/ mm9.chrom.sizes

mm9_refseq_genes_filtered_granges

Filtered RefSeq gene annotation as GRanges for mm9

Description

RefSeq gene annotation data for the ChIC package in GRanges format, filtered by the gene length and by overlaps.

Usage

```
data(mm9_refseq_genes_filtered_granges)
```

Format

A GRanges object.

rf_models

Random forest models for chromatin marks ChIP-seq experiment classification

Description

Random forest models based on ENCODE and Roadmap data for ChIP-seq experiment classification using the ChIC package.

Usage

```
data(rf_models)
```

Format

A list of 7 random forest models for the different chromatin marks and transcription factors:

- broadEncode : model for broad binding marks
- H3K9Encode : model for H3K9me3
- H3K27Encode : model for H3K27me3
- H3K36Encode : model for H3K36me3
- RNAPol2Encode : model for RNAPol2
- sharpEncode : model for sharp binding marks
- TFmodel : model for transcription factors

Index