

# Package ‘ChIC.data’

October 16, 2021

**Title** ChIC package data

**Version** 1.12.0

**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), genomeIntervals, randomForest

**Imports** caret (>= 6.0-78)

**biocViews** ExperimentData, ENCODE

**License** GPL-2

**Encoding** UTF-8

**LazyData** true

**LazyDataCompression** bzip2

**RoxygenNote** 7.1.1

**NeedsCompilation** no

**git\_url** <https://git.bioconductor.org/packages/ChIC.data>

**git\_branch** RELEASE\_3\_13

**git\_last\_commit** 557c32b

**git\_last\_commit\_date** 2021-05-19

**Date/Publication** 2021-10-16

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

## R topics documented:

chipSubset . . . . .	2
classesDefList . . . . .	2
compendium_db . . . . .	3
compendium_db_tf . . . . .	3
compendium_profiles . . . . .	4
compendium_profiles_TF . . . . .	5
dm3_chrom_info . . . . .	9
dm3_refseq_genes_filtered_granges . . . . .	9
EM_Results . . . . .	9
hg19_chrom_info . . . . .	10

hg19_refseq_genes_filtered_granges . . . . .	10
hg38_chrom_info . . . . .	11
hg38_refseq_genes_filtered_granges . . . . .	11
inputSubset . . . . .	12
mm10_chrom_info . . . . .	12
mm10_refseq_genes_filtered_granges . . . . .	13
mm9_chrom_info . . . . .	13
mm9_refseq_genes_filtered_granges . . . . .	14
rf_models . . . . .	14

<b>Index</b>	<b>15</b>
--------------	-----------

---

chipSubset	<i>ChIP-seq bam file stored as spp tag-list for a subset of chromosomes for the chip</i>
------------	--

---

### 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 downloaded 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	<i>List with definitions</i>
----------------	------------------------------

---

### 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	<i>Histone mark compendium</i>
---------------	--------------------------------

---

**Description**

Histone mark ChIP-seq compendium for the ChIC package. The compendium contains quality control metrics and metadata for 2204 histone mark samples analysed from ENCODE and Roadmap Epigenomics.

**Usage**

```
data(compendium_db)
```

**Format**

Data frame with 251 variables (quality control metrics, metadata) for each analysed sample.

**Source**

XX

---

compendium_db_tf	<i>Transcription factor compendium</i>
------------------	--

---

**Description**

Transcription factor ChIP-seq compendium for the ChIC package. The compendium contains quality control metrics and metadata for 1350 transcription factors analysed from ENCODE.

**Usage**

```
data(compendium_db_tf)
```

**Format**

Data frame with 251 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    setSize : setsize analysed

### 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 setSize : setsize analysed

**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

---

dm3_chrom_info	<i>dm3 chromosome information</i>
----------------	-----------------------------------

---

**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	<i>Filtered RefSeq gene annotation as GRanges for dm3</i>
-----------------------------------	---

---

**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.

---

EM_Results	<i>Example results from the qualityScores_EM() function based on the example ChIP-seq data for the Vignette for ChIC package</i>
------------	--

---

**Description**

Example data to be used in the vignette for ChIC package

**Usage**

```
data(EM_Results)
```

**Format**

list containing EM scores and filtered reads

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.

---

hg38_chrom_info	<i>hg38 chromosome information</i>
-----------------	------------------------------------

---

**Description**

hg38 chromosome information for the ChIC package.

**Usage**

```
data(hg38_chrom_info)
```

**Format**

A named list of int vectors with 2 elements, the start and end position of all hg38 chromosomes.

**Source**

<http://hgdownload.cse.ucsc.edu/goldenPath/hg38/bigZips/hg38.chrom.sizes>

---

hg38_refseq_genes_filtered_granges	<i>Filtered RefSeq gene annotation as GRanges</i>
------------------------------------	---

---

**Description**

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

**Usage**

```
data(hg38_refseq_genes_filtered_granges)
```

**Format**

A GRanges object.

---

inputSubset	<i>ChIP-seq bam file stored as spp tag-list for a subset of chromosomes for the input</i>
-------------	---

---

**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 downloaded 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

---

mm10_chrom_info	<i>mm10 chromosome information</i>
-----------------	------------------------------------

---

**Description**

mm10 chromosome information for the ChIC package.

**Usage**

```
data(mm10_chrom_info)
```

**Format**

A named list of int vectors with 2 elements, the start and end position of all mm10 chromosomes.

**Source**

<http://hgdownload.cse.ucsc.edu/goldenPath/mm10/bigZips/mm10.chrom.sizes>

---

`mm10_refseq_genes_filtered_granges`*Filtered RefSeq gene annotation as GRanges for mm10*

---

**Description**

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

**Usage**

```
data(mm10_refseq_genes_filtered_granges)
```

**Format**

A GRanges object.

---

`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:

- Broad : model for broad binding marks
- H3K9me3 : model for H3K9me3
- H3K27me3 : model for H3K27me3
- H3K36me3 : model for H3K36me3
- RNAPol2 : model for RNAPol2
- Sharp : model for sharp binding marks
- TF : model for transcription factors

# Index

- \* **EM\_Results**
  - EM\_Results, 9
- \* **Transcription**
  - compendium\_profiles\_TF, 5
- \* **annotation**
  - dm3\_refseq\_genes\_filtered\_granges, 9
  - hg19\_refseq\_genes\_filtered\_granges, 10
  - hg38\_refseq\_genes\_filtered\_granges, 11
  - mm10\_refseq\_genes\_filtered\_granges, 13
  - mm9\_refseq\_genes\_filtered\_granges, 14
- \* **bamfile**
  - chipSubset, 2
  - inputSubset, 12
- \* **chromatin**
  - compendium\_profiles, 4
- \* **classification**
  - rf\_models, 14
- \* **compendium**
  - compendium\_db, 3
  - compendium\_db\_tf, 3
  - compendium\_profiles, 4
  - compendium\_profiles\_TF, 5
- \* **control**
  - compendium\_db, 3
  - compendium\_db\_tf, 3
- \* **definitions**
  - classesDefList, 2
- \* **dm3**
  - dm3\_chrom\_info, 9
- \* **factors**
  - compendium\_profiles\_TF, 5
- \* **hg19**
  - hg19\_chrom\_info, 10
- \* **hg38**
  - hg38\_chrom\_info, 11
- \* **histone**
  - compendium\_db, 3
- \* **marks**
  - compendium\_db, 3
  - compendium\_profiles, 4
- \* **metagene**
  - compendium\_profiles, 4
  - compendium\_profiles\_TF, 5
- \* **metrics**
  - compendium\_db, 3
  - compendium\_db\_tf, 3
- \* **mm10**
  - mm10\_chrom\_info, 12
- \* **mm9**
  - mm9\_chrom\_info, 13
- \* **profiles**
  - compendium\_profiles, 4
  - compendium\_profiles\_TF, 5
- \* **quality**
  - compendium\_db, 3
  - compendium\_db\_tf, 3
- \* **tf**
  - compendium\_db\_tf, 3
- chipSubset, 2
- classesDefList, 2
- compendium\_db, 3
- compendium\_db\_tf, 3
- compendium\_profiles, 4
- compendium\_profiles\_TF, 5
- dm3\_chrom\_info, 9
- dm3\_refseq\_genes\_filtered\_granges, 9
- EM\_Results, 9
- hg19\_chrom\_info, 10
- hg19\_refseq\_genes\_filtered\_granges, 10
- hg38\_chrom\_info, 11
- hg38\_refseq\_genes\_filtered\_granges, 11
- inputSubset, 12
- mm10\_chrom\_info, 12
- mm10\_refseq\_genes\_filtered\_granges, 13
- mm9\_chrom\_info, 13
- mm9\_refseq\_genes\_filtered\_granges, 14
- rf\_models, 14