

# Package ‘Cogito’

December 20, 2024

**Type** Package

**Title** Compare genomic intervals tool - Automated, complete, reproducible and clear report about genomic and epigenomic data sets

**Version** 1.13.0

**Date** 2022-06-20

**Description** Biological studies often consist of multiple conditions which are examined with different laboratory set ups like RNA-sequencing or ChIP-sequencing. To get an overview about the whole resulting data set, Cogito provides an automated, complete, reproducible and clear report about all samples and basic comparisons between all different samples. This report can be used as documentation about the data set or as starting point for further custom analysis.

**License** LGPL-3

**Encoding** UTF-8

**LazyData** TRUE

**Depends** R (>= 4.1), GenomicRanges, jsonlite, GenomicFeatures, entropy

**Imports** BiocManager, rmarkdown, GenomeInfoDb, S4Vectors, AnnotationDbi, graphics, stats, utils, methods, magrittr, ggplot2, TxDb.Mmusculus.UCSC.mm9.knownGene

**biocViews** FunctionalGenomics, GeneRegulation, Software, Sequencing

**Suggests** BiocStyle, knitr, markdown, testthat (>= 3.0.0)

**VignetteBuilder** knitr

**NeedsCompilation** no

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

**git\_branch** devel

**git\_last\_commit** 55b1c5a

**git\_last\_commit\_date** 2024-10-29

**Repository** Bioconductor 3.21

**Date/Publication** 2024-12-20

**Author** Annika Bürger [cre, aut]

**Maintainer** Annika Bürger <a.buerger@uni-muenster.de>

## Contents

|                             |   |
|-----------------------------|---|
| aggregateRanges . . . . .   | 2 |
| MurEpi.ChIP.small . . . . . | 3 |
| MurEpi.RNA.small . . . . .  | 4 |
| MurEpi.RRBS.small . . . . . | 5 |
| summarizeRanges . . . . .   | 6 |

|              |          |
|--------------|----------|
| <b>Index</b> | <b>8</b> |
|--------------|----------|

---

|                 |   |
|-----------------|---|
| aggregateRanges | <i>Aggregate GRanges with columns of attached values to genes</i> |
|-----------------|---|

---

### Description

Aggregates multiple GRanges objects with present columns of attached values (mcols) to genes or ranges of given reference of given organism.

### Usage

```
aggregateRanges(ranges, configfile = NULL, organism = NULL,
                referenceRanges = NULL, name = "", verbose = FALSE)
```

### Arguments

|                 |   |
|-----------------|---|
| ranges          | list of GRanges, GRangesList or CompressedGRangesList with names in "RRBS DNA CNV RNA ChIP"                     |
| configfile      | character, path to configuration file in json format  |
| organism        | TxDb or OrganismDb object, default value NULL   |
| referenceRanges | list of GRanges, GRangesList or CompressedGRangesList with length one and name of reference, default value NULL |
| name            | character, default value ""   |
| verbose         | logical, default value FALSE  |

### Value

List object with three members: One GRanges object with one gene or range of given reference per line and one column per sample, configuration information, and name.

### Author(s)

Annika Bürger

**See Also**[summarizeRanges](#)**Examples**

```

mm9 <- TxDb.Mmusculus.UCSC.mm9.knownGene::TxDb.Mmusculus.UCSC.mm9.knownGene

### small artificial example ###
ranges.RNA.control <-
  GRanges(seq = "chr10",
           IRanges(c(41023369, 41211825, 41528287, 41994926, 42301673,
                    43256520, 43618919, 49503584, 51349066, 52099001),
              c(41023544, 41212385, 41528663, 41995357, 42302290,
                43257075, 43619492, 49504033, 51349425, 52099521)),
           seqinfo = GenomeInfoDb::seqinfo(mm9),
           expr = runif(5, 0, 1))
ranges.RNA.condition <-
  GRanges(seq = "chr10",
           IRanges(c(41013942, 41208731, 41535166, 41999999, 42292275,
                    43256194, 43615562, 49497888, 51347046, 52092180),
              c(41014274, 41209664, 41536039, 42000182, 42292965,
                43256430, 43615866, 49498362, 51347969, 52092733)),
           seqinfo = GenomeInfoDb::seqinfo(mm9),
           expr = runif(5, 0, 1))
ranges.ChIP.control <-
  GRanges(seq = "chr10",
           IRanges(c(41022835, 41307587, 42197924, 42302387, 42893825,
                    43259749, 43620352, 43721891, 44248812, 45207572,
                    49508713, 51309978, 51348779, 52101900, 52265513),
              c(41022954, 41307745, 42198201, 42302555, 42893974,
                43259889, 43620604, 43722051, 44248920, 45207704,
                49508859, 51310187, 51348921, 52102030, 52265689)),
           seqinfo = GenomeInfoDb::seqinfo(mm9),
           score = round(runif(15, 5, 90)))

example.dataset <- list(RNA = GRangesList(control = ranges.RNA.control,
                                          condition = ranges.RNA.condition),
                      ChIP = ranges.ChIP.control)

aggregated.ranges <- aggregateRanges(ranges = example.dataset,
                                     organism = mm9,
                                     name = "art.example",
                                     verbose = TRUE)

names(aggregated.ranges)
head(aggregated.ranges$genes)

```

**Description**

This murine data from King et al. was downloaded from the NCBI GEO database under accession number GSE77004.

The available ChIP-seq data (GSE77002) was then processed as described in King et al.: After alignment with bowtie with parameters selecting for uniquely mapped, best-matching reads and a maximum of two mismatches per read, the peak calling was done with homer findPeaks algorithm and an input control. Subsequently, the raw peaks were filtered with the following parameters: -F 8 for H3K4me3, -size 1000 -minDist 3000 -F 4 -tagThreshold 32 for H3K27me3, -F 4 for H3K27ac and -size 1000 -minDist1000 -nfr for H3K4me1.

To reduce the storage size and complexity of the murine example data, the data only contains data of chr5 and four sample conditions (3x TFX and 1x mut) were removed.

**Usage**

```
MurEpi.ChIP.small
```

**Format**

A GRanges Object containing a lot of Ranges with scores.

**Source**

NCBI GEO database accession number GSE77002 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77002>

**References**

King AD, Huang K, Rubbi L, Liu S, Wang CY, Wang Y, Pellegrini M, Fan G. Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. Cell Rep. 2016 Sep 27;17(1):289-302. doi: 10.1016/j.celrep.2016.08.083

---

MurEpi.RNA.small

*Example data set: Murine RNA-seq RPKM values of GSE77004*

---

**Description**

This murine data from King et al. was downloaded from the NCBI GEO database under accession number GSE77004.

The available RNA-seq RPKM values per gene from the same study, provided under the accession number GSE77003.

To reduce the storage size and complexity of the murine example data, the data only contains data of chr5 and four sample conditions (3x TFX and 1x mut) were removed.

**Usage**

```
MurEpi.RNA.small
```

**Format**

A GRanges Object containing a lot of Ranges with expression values scores.

**Source**

NCBI GEO database accession number GSE77003 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77003>

**References**

King AD, Huang K, Rubbi L, Liu S, Wang CY, Wang Y, Pellegrini M, Fan G. Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. *Cell Rep.* 2016 Sep 27;17(1):289-302. doi: 10.1016/j.celrep.2016.08.083

---

MurEpi.RRBS.small

*Example data set: Murine methylation status data set of GSE77004*

---

**Description**

This murine data from King et al. was downloaded from the NCBI GEO database under accession number GSE77004.

The methylation status, measured by RRBS, was similarly taken from the published files (accession number GSE84103), which contain the fraction of methylated cytosine for every CpG context supported by a minimum of 5 reads.

To reduce the storage size and complexity of the murine example data, the data only contains data of chr5 and four sample conditions (3x TFX and 1x mut) were removed.

**Usage**

MurEpi.RRBS.small

**Format**

A GRanges Object containing a lot of Ranges with methylation status.

**Source**

NCBI GEO database accession number GSE84103 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84103>

**References**

King AD, Huang K, Rubbi L, Liu S, Wang CY, Wang Y, Pellegrini M, Fan G. Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. *Cell Rep.* 2016 Sep 27;17(1):289-302. doi: 10.1016/j.celrep.2016.08.083

---

|                 |                                     |
|-----------------|-------------------------------------|
| summarizeRanges | <i>Summarize Aggregated GRanges</i> |
|-----------------|-------------------------------------|

---

**Description**

Summarize GRanges with present columns of attached values (mcols).

**Usage**

```
summarizeRanges(aggregated.ranges, outputFormat = "pdf", verbose = FALSE)
```

**Arguments**

|                   |  |
|-------------------|--|
| aggregated.ranges | list of GRanges, cofiguration information, and name for example result from function Cogito::summarizeRanges |
| outputFormat      | character, can be pdf or html default value pdf  |
| verbose           | logical, default value FALSE   |

**Value**

No return value, only side effects: creation of a rmd, a pdf or html and a data file (RData).

**Author(s)**

Annika Bürger

**See Also**

[aggregateRanges](#)

**Examples**

```
mm9 <- TxDb.Mmusculus.UCSC.mm9.knownGene::TxDb.Mmusculus.UCSC.mm9.knownGene

### small artificial example ###
ranges.RNA.control <-
  GRanges(seq = "chr10",
          IRanges(c(41023369, 41211825, 41528287, 41994926, 42301673,
                    43256520, 43618919, 49503584, 51349066, 52099001),
          c(41023544, 41212385, 41528663, 41995357, 42302290,
            43257075, 43619492, 49504033, 51349425, 52099521)),
          seqinfo = GenomeInfoDb::seqinfo(mm9),
          expr = runif(5, 0, 1))
ranges.RNA.condition <-
  GRanges(seq = "chr10",
          IRanges(c(41013942, 41208731, 41535166, 41999999, 42292275,
                    43256194, 43615562, 49497888, 51347046, 52092180),
          c(41014274, 41209664, 41536039, 42000182, 42292965,
```

```
      43256430, 43615866, 49498362, 51347969, 52092733)),
  seqinfo = GenomeInfoDb::seqinfo(mm9),
  expr = runif(5, 0, 1))
ranges.ChIP.control <-
  GRanges(seq = "chr10",
    IRanges(c(41022835, 41307587, 42197924, 42302387, 42893825,
      43259749, 43620352, 43721891, 44248812, 45207572,
      49508713, 51309978, 51348779, 52101900, 52265513),
    c(41022954, 41307745, 42198201, 42302555, 42893974,
      43259889, 43620604, 43722051, 44248920, 45207704,
      49508859, 51310187, 51348921, 52102030, 52265689)),
  seqinfo = GenomeInfoDb::seqinfo(mm9),
  score = round(runif(15, 5, 90)))

example.dataset <- list(RNA = GRangesList(control = ranges.RNA.control,
  condition = ranges.RNA.condition),
  CHIP = ranges.ChIP.control)

aggregated.ranges <- aggregateRanges(ranges = example.dataset,
  organism = mm9,
  name = "art.example",
  verbose = TRUE)

# adding information about conditions
aggregated.ranges$config$conditions <- list(condition = c("RNA.condition.expr"),
  control = c("RNA.control.expr",
    "CHIP.score"))

summarizeRanges(aggregated.ranges = aggregated.ranges,
  outputFormat = "pdf",
  verbose = TRUE)
```

# Index

## \* datasets

MurEpi.ChIP.small, 3

MurEpi.RNA.small, 4

MurEpi.RRBS.small, 5

aggregateRanges, 2, 6

MurEpi.ChIP.small, 3

MurEpi.RNA.small, 4

MurEpi.RRBS.small, 5

summarizeRanges, 3, 6