

# Package ‘CexoR’

September 18, 2024

**Version** 1.43.0

**Date** 2022-05-28

**Type** Package

**Title** An R package to uncover high-resolution protein-DNA interactions in ChIP-exo replicates

**Description** Strand specific peak-pair calling in ChIP-exo replicates. The cumulative Skellam distribution function is used to detect significant normalised count differences of opposed sign at each DNA strand (peak-pairs). Then, irreproducible discovery rate for overlapping peak-pairs across biological replicates is computed.

**Depends** R (>= 4.2.0), S4Vectors, IRanges

**Imports** Rsamtools, GenomeInfoDb, GenomicRanges, rtracklayer, idr, RColorBrewer, genomation

**Suggests** RUnit, BiocGenerics, BiocStyle, knitr, rmarkdown

**License** Artistic-2.0 | GPL-2 + file LICENSE

**biocViews** FunctionalGenomics, Sequencing, Coverage, ChIPSeq, PeakDetection

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

**git\_branch** devel

**git\_last\_commit** 16301bb

**git\_last\_commit\_date** 2024-04-30

**Repository** Bioconductor 3.20

**Date/Publication** 2024-09-18

**Author** Pedro Madrigal [aut, cre] (<<https://orcid.org/0000-0003-1959-8199>>)

**Maintainer** Pedro Madrigal <[pmadrigal@ebi.ac.uk](mailto:pmadrigal@ebi.ac.uk)>

## Contents

|                         |          |
|-------------------------|----------|
| CexoR-package . . . . . | 2        |
| CexoR . . . . .         | 3        |
| cexor . . . . .         | 3        |
| plotcexor . . . . .     | 5        |
| <b>Index</b>            | <b>7</b> |

---

CexoR-package

*An R package to uncover high-resolution protein-DNA interactions in ChIP-exo replicates*

---

## Description

Strand specific peak-pair calling in ChIP-exo replicates. The cumulative Skellam distribution function (package 'skellam') is used to detect significant normalised count differences of opposed sign at each DNA strand (peak-pairs). Irreproducible discovery rate for overlapping peak-pairs across biological replicates is estimated using the package 'idr'.

## Details

|           |                                     |
|-----------|-------------------------------------|
| Package:  | CexoR                               |
| Type:     | Package                             |
| Version:  | 1.35.1                              |
| Date:     | 2022-05-28                          |
| License:  | Artistic-2.0   GPL-2 + file LICENSE |
| LazyLoad: | yes                                 |

## Author(s)

Pedro Madrigal,

Maintainer: Pedro Madrigal <pmadrigal@ebi.ac.uk>

## References

Madrigal P (2015) CexoR: an R/Bioconductor package to uncover high-resolution protein-DNA interactions in ChIP-exo replicates. *EMBNet.journal* 21: e837.

Skellam JG (1946) The frequency distribution of the difference between two Poisson variates belonging to different populations. *J R Stat Soc Ser A* 109: 296.

Li Q, Brown J, Huang H, Bickel P (2011) Measuring reproducibility of high-throughput experiments. *Ann Appl Stat* 5: 1752-1779.

Rhee HS, Pugh BF (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* 147: 1408-1419.

## Examples

```
## hg19. chr2:1-1,000,000. CTCF data from Rhee and Pugh (2011)
owd <- setwd( tempdir() )

rep1 <- "CTCF_rep1_chr2_1-1e6.bam"
rep2 <- "CTCF_rep2_chr2_1-1e6.bam"
rep3 <- "CTCF_rep3_chr2_1-1e6.bam"
r1 <- system.file( "extdata", rep1, package="CexoR", mustWork = TRUE )
r2 <- system.file( "extdata", rep2, package="CexoR", mustWork = TRUE )
r3 <- system.file( "extdata", rep3, package="CexoR", mustWork = TRUE )
```

```
chipexo <- cexor( bam=c(r1,r2,r3), chrN="chr2", chrL=1e6, idr=0.01, p=1e-12, N=3e4 )
plotcexor( bam=c(r1,r2,r3), peaks=chipexo, EXT=500 )
setwd( owd )
```

CexoR

*CexoR internal functions***Description**

Internal undocumentation functions

cexor

*ChIP-exo peak-pair calling with replicates***Description**

ChIP-exo peak-pair calling with replicates.

**Usage**

```
cexor(bam, chrN, chrL, p=1e-9, dpeaks=c(0,150), dpairs=100, idr=0.01,
N=5e6, bedfile=TRUE, mu=2.6, sigma=1.3, rho=0.8, prop=0.7)
```

**Arguments**

|         |  |
|---------|--|
| bam     | BAM alignment files of biological replicates.  |
| chrN    | Vector of chromosome names.  |
| chrL    | Vector of chromosome sizes (bp).   |
| p       | P-value cutoff (should be relaxed, e.g. 1e-3, to allow the correct estimation of the irreproducible discovery rate (idr). However, this depends on the sequencing depth. For datasets with high number of tag counts, 1e-9 can be appropriate. See the vignette for more information.) |
| dpeaks  | Min. and max. allowed distance between peak pairs located at opposed strands in a replicate (bp).  |
| dpairs  | Max. allowable distance between peak-pair centres across replicates (bp).  |
| idr     | Irreproducible discovery rate cutoff [0-1].  |
| N       | Genome is divided in blocks of N bp. for processing. N must be not higher than the size of the smallest chromosome.  |
| bedfile | Generate BED files of ChIP-exo reproducible peak pairs.  |
| mu      | A starting value for the mean of the reproducible component (see 'idr' package).   |
| sigma   | A starting value for the standard deviation of the reproducible component (see 'idr' package).   |
| rho     | A starting value for the correlation coefficient of the reproducible component (see 'idr' package).  |
| prop    | A starting value for the proportion of reproducible component (see 'idr' package).   |

## Details

Strand specific peak-pair calling in ChIP-exo replicates. The cumulative Skellam distribution function (package 'skellam') is used to detect significant normalized count differences of opposed sign at each DNA strand (peak-pairs). Irreproducible discovery rate for overlapping peak-pairs across biological replicates is estimated using the package 'idr'. The internal functions `pskellam` and `pskellam.sp` from the Jerry W. Lewis' 'skellam' R package (version 0.0-8-7) are used to calculate the cumulative Skellam distribution (see LICENSE file).

## Value

A list containing the following elements:

- `bindingEvents` A GRanges object with reproducible peak pair locations. The metadata 'value' indicates the Irreproducible discovery rate (IDR) estimated at this region, while 'repI.neg.log10pvalue' indicates  $-\log_{10}(\text{p-value})$  for the replicate I. 'Stouffer.pvalue' and 'Fisher.pvalue' report the combined p-value considering they come from independent significance tests.
- `bindingCentres` A GRanges object with centre position of reproducible peak pair locations. The metadata 'value' indicates the Irreproducible discovery rate (IDR) estimated at this region, while 'repI.neg.log10pvalue' indicates  $-\log_{10}(\text{p-value})$  for the replicate I. 'Stouffer.pvalue' and 'Fisher.pvalue' report the combined p-value considering they come from independent significance tests.
- `pairedPeaksRep1` A GRangesList object with the location of peak pairs retrieved at each replicate. The metadata 'score' indicates  $-\log_{10}(\text{p-value})$ .

## Author(s)

Pedro Madrigal, <pmadrigal@ebi.ac.uk>

## References

Madrigal P (2015) CexoR: an R/Bioconductor package to uncover high-resolution protein-DNA interactions in ChIP-exo replicates. *EMBnet.journal* 21: e837.

## See Also

[CexoR-package](#)

## Examples

```
## hg19. chr2:1-1,000,000. CTCF data from Rhee and Pugh (2011)
owd <- setwd(tempdir())

rep1 <- "CTCF_rep1_chr2_1-1e6.bam"
rep2 <- "CTCF_rep2_chr2_1-1e6.bam"
rep3 <- "CTCF_rep3_chr2_1-1e6.bam"
r1 <- system.file("extdata", rep1, package="CexoR", mustWork = TRUE)
r2 <- system.file("extdata", rep2, package="CexoR", mustWork = TRUE)
r3 <- system.file("extdata", rep3, package="CexoR", mustWork = TRUE)

chipexo <- cexor(bam=c(r1,r2,r3), chrN="chr2", chrL=1e6, idr=0.01, p=1e-12, N=3e4)

plotcexor(bam=c(r1,r2,r3), peaks=chipexo, EXT=500)
```

```
setwd(owd)
```

---

plotcexor

*Visualization of ChIP-exo peak-pair calling with replicates*

---

## Description

Visualization of ChIP-exo peak-pair calling with replicates.

## Usage

```
plotcexor(bam, peaks, EXT=500)
```

## Arguments

|       |   |
|-------|---|
| bam   | BAM alignment files of biological replicates.   |
| peaks | Object (list) output of the function 'cexor'.   |
| EXT   | Extension (bp) upstream and downstream the central position of reproducible peak pair locations for visualization purposes. |

## Details

Visualization of ChIP-exo peak-pair calling with replicates.

## Value

R plot.

## Author(s)

Pedro Madrigal, <pmadrigal@ebi.ac.uk>

## References

Madrigal P (2015) CexoR: an R/Bioconductor package to uncover high-resolution protein-DNA interactions in ChIP-exo replicates. *EMBnet.journal* 21: e837.

## See Also

[CexoR-package](#)

**Examples**

```
## hg19. chr2:1-1,000,000. CTCF data from Rhee and Pugh (2011)
owd <- setwd(tempdir())

rep1 <- "CTCF_rep1_chr2_1-1e6.bam"
rep2 <- "CTCF_rep2_chr2_1-1e6.bam"
rep3 <- "CTCF_rep3_chr2_1-1e6.bam"
r1 <- system.file("extdata", rep1, package="CexoR",mustWork = TRUE)
r2 <- system.file("extdata", rep2, package="CexoR",mustWork = TRUE)
r3 <- system.file("extdata", rep3, package="CexoR",mustWork = TRUE)

chipexo <- cexor(bam=c(r1,r2,r3), chrN="chr2", chrL=1e6, idr=0.01, p=1e-12, N=3e4)

plotcexor(bam=c(r1,r2,r3), peaks=chipexo, EXT=500)

setwd(owd)
```

# Index

\* **internal**

CexoR, [3](#)

CexoR, [3](#)

CexoR (CexoR-package), [2](#)

cexor, [3](#)

CexoR-package, [2](#)

plotcexor, [5](#)

pskellam (CexoR), [3](#)