

# Package ‘CexoR’

April 9, 2015

**Version** 1.4.0

**Date** 2014-01-01

**Type** Package

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

**Author** Pedro Madrigal <pm12@sanger.ac.uk>

**Description** 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'.

**Depends** R (>= 2.10.0), S4Vectors, IRanges

**Maintainer** Pedro Madrigal <pm12@sanger.ac.uk>

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

**Suggests** RUnit, BiocGenerics, BiocStyle

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

**LazyLoad** yes

**LazyData** yes

**biocViews** Transcription, Genetics, Sequencing

## R topics documented:

CexoR-package . . . . .	2
cexor . . . . .	3

<b>Index</b>	<b>6</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 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'.

## Details

Package: CexoR  
Type: Package  
Version: 1.3.2  
Date: 2014-01-01  
License: Artistic-2.0 | GPL-2 + file LICENSE  
LazyLoad: yes

## Author(s)

Pedro Madrigal,  
Maintainer: Pedro Madrigal <pm12@sanger.ac.uk>

## References

Madrigal P (in preparation) CexoR: An R/Bioconductor package to uncover high-resolution protein-DNA interactions in ChIP-exo replicates.  
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)

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

setwd(owd)

```

cexor

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

ChIP-exo peak-pair calling with replicates.

**Usage**

```

cexor(bam, chrN, chrL, p=1e-12, dpeaks=c(0,150), dpairs=100, idr=0.01,
N=5e6, bedfile=TRUE)

```

**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 to allow the correct estimation of the irreproducible discovery rate (idr). See the vignette for more information.)
dpeaks	Main. 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.

**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.
- `pairedPeaksRepl`  
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, <pm12@sanger.ac.uk>

**References**

- Madrigal P (in preparation) CexoR: An R/Bioconductor package to uncover high-resolution protein-DNA interactions in ChIP-exo replicates.
- 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.

**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)
```

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

# Index

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

cexor, [3](#)

CexoR-package, [2](#)