

Package ‘DNaseR’

February 28, 2015

Version 1.5.0

Date 2014-05-12

Type Package

Title DNase I footprinting analysis of DNase-seq data

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

Description Strand-specific digital genomic footprinting in 'double-hit' DNase-seq data. The cumulative Skellam distribution function (package 'skellam') is used to detect significant normalized count differences of opposed sign at each DNA strand. This is done in order to determine the protein-binding footprint flanks. Preprocessing of the mapped reads is recommended before running DNaseR (e.g., quality checking and removal of sequence-specific biases).

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

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

Imports GenomeInfoDb, GenomicRanges, Rsamtools

Suggests RUnit, BiocGenerics

License GPL-2 + file LICENSE

biocViews Transcription, Genetics

LazyLoad yes

LazyData yes

NeedsCompilation no

R topics documented:

DNaseR-package	2
footprints	3
Index	6

DNaseR-package

DNase I footprinting analysis of DNase-seq data

Description

Strand-specific digital genomic footprinting in 'double-hit' DNase-seq data. The cumulative Skellam distribution function (package 'skellam') is used to detect significant normalized count differences of opposed sign at each DNA strand. This is done in order to determine the protein-binding footprint flanks. Preprocessing of the mapped reads is recommended before running DNaseR (e.g., quality checking and removal of sequence specific bias).

Details

Package:	DNaseR
Type:	Package
Version:	1.3.2
Date:	2014-05-12
License:	GPL-2 + file LICENSE
LazyLoad:	yes

Author(s)

Pedro Madrigal,

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

References

Madrigal P, Krajewski P (2012) Current bioinformatic approaches to identify DNase I hypersensitive sites and genomic footprints from DNase-seq data. *Front Genet* 3: 230.

Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, Maurano MT, Humbert R, Rynes E, Wang H, Vong S, Lee K, Bates D, Diegel M, Roach V, Dunn D, Neri J, Schafer A, Hansen RS, Kutayavin T, Giste E, Weaver M, Canfield T, Sabo P, Zhang M, Balasundaram G, Byron R, MacCoss MJ, Akey JM, Bender MA, Groudine M, Kaul R, Stamatoyannopoulos JA (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489: 83-90.

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.

Song L, Crawford GE (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* 2: pdb.prot5384.

Examples

```
## hg18. chrY:1 - 3000Kb reads from DNase-seq dataset wgEncodeUwDgfTh1A1n.bam
## from the ENCODE Project.
##
## Downloaded from:
## http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDgf/
## release1/wgEncodeUwDgfTh1A1n.bam

owd <- setwd(tempdir())

bamfile <- "chrY_3Kb_wgEncodeUwDgfTh1A1n.bam"

f <- system.file("extdata", bamfile, package="DNaseR", mustWork = TRUE)

dgf <- footprints(bam=f, chrN="chrY", chrL=3e6, p=1e-9, width=c(6,40), N=2e6)

setwd(owd)
```

footprints

DNase I footprinting analysis of DNase-seq data

Description

DNase I footprinting analysis of DNase-seq data.

Usage

```
footprints(bam, chrN, chrL, p=1e-9, width=c(6,40), N=5e6, correction="BH")
```

Arguments

bam	BAM file of DNase-seq mapped reads.
chrN	Vector of chromosome names.
chrL	Vector of chromosome sizes (bp).
p	p-value cutoff for the footprint events (default 1e-9).
width	Max. and min. footprint width (bp).
N	Genome is divided in blocks of N bp. for processing. N must be not higher than the size of the smallest chromosome.
correction	Multiple comparison adjustment method for the p-values at each flank of a footprint. Allowed values are same as in p.adjust: "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none" (default "BH").

Details

Strand-specific digital genomic footprinting in DNase-seq data. The cumulative Skellam distribution function (package 'skellam') is used to detect significant normalized count differences of opposed sign at each DNA strand. This is done in order to determine the footprint flanks. Preprocessing of the mapped reads is recommended before running DNaseR (e.g., quality checking and removal of sequence-specific bias). Initially, one p-value is calculated at each flank of the footprint. To control for multiple testing, the p-values delimiting each flank of the footprint (pval.forward and pval.reverse) are corrected using Benjamini-Hochberg procedure ("BH", default). The column pval.footprint.event stores the sum 'pval.forward + pval.reverse'. 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

footprints.events

A data.frame with the location of each footprint ("chr","start","end"), the width of the footprint ("length"), the corrected p-values at each flank ("pval.forward","pval.reverse"), the final p-value of the footprint ("pval.footprint.event") and its $-\log_{10}$ ("log10.pval.footprint.event").

Author(s)

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

References

Madrigal P, Krajewski P (2012) Current bioinformatic approaches to identify DNase I hypersensitive sites and genomic footprints from DNase-seq data. *Front Genet* 3: 230.

Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, Maurano MT, Humbert R, Rynes E, Wang H, Vong S, Lee K, Bates D, Diegel M, Roach V, Dunn D, Neri J, Schafer A, Hansen RS, Kutayavin T, Giste E, Weaver M, Canfield T, Sabo P, Zhang M, Balasundaram G, Byron R, MacCoss MJ, Akey JM, Bender MA, Groudine M, Kaul R, Stamatoyannopoulos JA (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489: 83-90.

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.

Song L, Crawford GE (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* 2: pdb.prot5384.

See Also

[DNaseR-package](#)

Examples

```
## hg18. chrY:1 - 3000Kb reads from DNase-seq dataset wgEncodeUwDgfTh1A1n.bam
## from the ENCODE Project.
##
## Downloaded from:
## http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwDgf/
## release1/wgEncodeUwDgfTh1A1n.bam

owd <- setwd(tempdir())

bamfile <- "chrY_3Kb_wgEncodeUwDgfTh1A1n.bam"

f <- system.file("extdata", bamfile, package="DNaseR", mustWork = TRUE)

dgf <- footprints(bam=f, chrN="chrY", chrL=3e6, p=1e-9, width=c(6,40), N=2e6)

setwd(owd)
```

Index

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

DNaseR-package, [2](#)

footprints, [3](#)