Package 'GUIDEseq'

October 27, 2015

Type Package

Title GUIDE-seq analysis pipeline

Version 1.0.3

Date 2015-10-21

Author Lihua Julie Zhu, Michael Lawrence, Ankit Gupta, Alper Kucukural, Manuel Garber, Scot A. Wolfe

Maintainer Lihua Julie Zhu `<julie.zhu@umassmed.edu>`

Depends R (>= 3.2.0), IRanges, BiocGenerics, S4Vectors

Imports BiocParallel, Biostrings, CRISPRseek, ChIPpeakAnno, GenomicRanges, data.table, matrixStats, BSgenome, parallel

biocViews GeneRegulation, Sequencing, WorkflowStep

Suggests knitr, RUnit, BiocStyle, BSgenome.Hsapiens.UCSC.hg19

VignetteBuilder knitr

Description The package implements GUIDE-seq analysis workflow including functions for obtaining unique cleavage events, estimating the locations of the cleavage sites, aka, peaks, merging estimated cleavage sites from plus and minus strand, and performing off target search of the extended regions around cleavage sites.

License GPL (>= 2)

LazyLoad yes

NeedsCompilation no

R topics documented:

`GUIDEseq-package` *Analysis of GUIDE-seq*

Description

The package includes functions to retain one read per unique molecular identifier (UMI), filter reads lacking integration oligo sequence, identify peak locations (cleavage sites) and heights, perform target and off target search of the input gRNA. This package leverages CRISPRseek and ChIPpeakAnno packages.

Details

Package:	GUIDEseq
Type:	Package
Version:	1.0
Date:	2015-09-04
License:	GPL (>= 2)

Function GUIDEseqAnalysis integrates all steps of GUIDE-seq analysis into one function call

Author(s)

Lihua Julie Zhu Maintainer:julie.zhu@umassmed.edu

References

Shengdar Q Tsai and J Keith Joung et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature Biotechnology 33, 187 to 197 (2015)

See Also

GUIDEseqAnalysis

Examples

```
    if(interactive())
    {
        umiFile <- system.file("extdata", "UMI-HEK293_site4_R1.txt",
            package = "GUIDEseq")
        alignFile <- system.file("extdata","bowtie2.HEK293_site4.sort.bed" ,
            package = "GUIDEseq")
        gRNA.file <- system.file("extdata","gRNA.fa", package = "GUIDEseq")
        guideRes <- GUIDEseqAnalysis(
            alignment.inputfile = alignFile ,
            umi.inputfile = umiFile, gRNA.file = gRNA.file)
        names(cleavages)
   }
```

`getPeaks` *Obtain peaks from GUIDE-seq*

Description

Obtain strand-specific peaks from GUIDE-seq

Usage

```
getPeaks(gr, window.size = 20L, step = 20L, bg.window.size = 5000L,
    min.reads = 10L, min.SNratio = 2, maxP = 0.05, multi.core = TRUE,
    stats = c("poisson", "nbinom"), p.adjust.methods =
    c("none", "BH", "holm", "hochberg", "hommel", "bonferroni", "BY", "fdr"))
```

Arguments

`gr`	GRanges with cleavage sites, output from getUniqueCleavageEvents
`window.size`	window size to calculate coverage
`step`	step size to calculate coverage
`bg.window.size`	window size to calculate local background
`min.reads`	minimum number of reads to be considered as a peak
`min.SNratio`	minimum signal noise ratio, which is the coverage normalized by local background
`maxP`	Maximum p-value to be considered as significant
`multi.core`	Indicating whether run in multi core mode, i.e., parallel processing, default TRUE
`stats`	Statistical test, default poisson
`p.adjust.methods`	Adjustment method for multiple comparisons, default none

Value

`peaks`	GRanges with count (peak height), bg (local background), SNratio (signal noise ratio), p-value, and option adjusted p-value
`summarized.count`	A data frame contains the same information as peaks except that it has all the sites without filtering.

Author(s)

Lihua Julie Zhu

Examples

```
if (interactive())
{
    data(uniqueCleavageEvents)
    peaks <- getPeaks(uniqueCleavageEvents$cleavage.gr,
        min.reads = 80)
    peaks$peaks
}
```

`getUniqueCleavageEvents`

Using UMI sequence to obtain the starting sequence library

Description

PCR amplification often leads to biased representation of the starting sequence population. To track the sequence tags present in the initial sequence library, a unique molecular identifier (UMI) is added to the 5 prime of each sequence in the staring library. This function uses the UMI sequence plus the first few sequence from R1 reads to obtain the starting sequence library.

Usage

```
getUniqueCleavageEvents(alignment.inputfile, umi.inputfile,
alignment.format = "bed", umi.header = FALSE, read.ID.col = 1,
umi.col = 2, umi.sep = "\t", keep.R1only = TRUE, keep.R2only = TRUE,
paired.direction = "opposite.strand", max.paired.distance = 1000,
min.mapping.quality = 30, max.R1.len = 130, max.R2.len = 130,
apply.both.max.len = FALSE, same.chromosome = TRUE,
distance.inter.chrom = -1, min.R1.mapped = 30, min.R2.mapped = 30,
apply.both.min.mapped = FALSE, max.duplicate.distance = 0,
umi.plus.R1start.unique = TRUE, umi.plus.R2start.unique = TRUE)
```

Arguments

`alignment.inputfile`	The alignment file. Currently supports bed output file with CIGAR information. Suggest run the workflow binReads.sh, which sequentially runs barcode binning, adaptor removal, alignment to genome, alignment quality filtering, and bed file conversion. Please download the workflow function and its helper scripts at http://mccb.umassmed.edu/GUIDE-seq/binReads/
`umi.inputfile`	A text file containing at least two columns, one is the read identifier and the other is the UMI or UMI plus the first few bases of R1 reads. Suggest use getUMI.sh to generate this file. Please download the script and its helper scripts at http://mccb.umassmed.edu/GUIDE-seq/getUMI/
`alignment.format`	The format of the alignment input file. Default bed file format. Currently only bed file format is supported, which is generated from binReads.sh
`umi.header`	Indicates whether the umi input file contains a header line or not. Default to FALSE
`read.ID.col`	The index of the column containing the read identifier in the umi input file, default to 1
`umi.col`	The index of the column containing the umi or umi plus the first few bases of sequence from the R1 reads, default to 2
`umi.sep`	column separator in the umi input file, default to tab
`keep.R1only`	Specify whether to include alignment with only R1 without paired R2. Default TRUE
`keep.R2only`	Specify whether to include alignment with only R2 without paired R1. Default TRUE

`paired.direction`	Specify whether the R1 and R2 should be aligned to the same strand or opposite strand. Default opposite.strand
`max.paired.distance`	Specify the maximum distance allowed between paired R1 and R2 reads. Default 1000 bp
`min.mapping.quality`	Specify min.mapping.quality of acceptable alignments
`max.R1.len`	The maximum retained R1 length to be considered for downstream analysis, default 130 bp. Please note that default of 130 works well when the read length 150 bp. Please also note that retained R1 length is not necessarily equal to the mapped R1 length
`max.R2.len`	The maximum retained R2 length to be considered for downstream analysis, default 130 bp. Please note that default of 130 works well when the read length 150 bp. Please also note that retained R2 length is not necessarily equal to the mapped R2 length
`apply.both.max.len`	Specify whether to apply maximum length requirement to both R1 and R2 reads, default FALSE
`same.chromosome`	Specify whether the paired reads are required to align to the same chromosome, default TRUE
`distance.inter.chrom`	Specify the distance value to assign to the paired reads that are aligned to different chromosome, default -1
`min.R1.mapped`	The maximum mapped R1 length to be considered for downstream analysis, default 30 bp.
`min.R2.mapped`	The maximum mapped R2 length to be considered for downstream analysis, default 30 bp.
`apply.both.min.mapped`	Specify whether to apply minimum mapped length requirement to both R1 and R2 reads, default FALSE
`max.duplicate.distance`	Specify the maximum distance apart for two reads to be considered as duplicates, default 0. Currently only 0 is supported
`umi.plus.R1start.unique`	To specify whether two mapped reads are considered as unique if both containing the same UMI and same alignment start for R1 read, default TRUE.
`umi.plus.R2start.unique`	To specify whether two mapped reads are considered as unique if both containing the same UMI and same alignment start for R2 read, default TRUE.

Value

`cleavage.gr`	Cleavage sites with one site per UMI as GRanges with metadata column total set to 1 for each range
`unique.umi.plus.R2`	a data frame containing unique cleavage site from R2 reads mapped to plus strand with the following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2

	read) strand.x (strand of readSide.x/R1 read) start.y (start of readSide.y/R2 read) end.x (start of readSide.x/R1 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)
`unique.umi.minus.R2`	a data frame containing unique cleavage site from R2 reads mapped to minus strand with the following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2 read) strand.x (strand of readSide.x/R1 read) end.y (end of readSide.y/R2 read) start.x (start of readSide.x/R1 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)
`unique.umi.plus.R1`	a data frame containing unique cleavage site from R1 reads mapped to minus strand without corresponding R2 reads mapped to the plus strand, with the following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2 read) strand.x (strand of readSide.x/R1 read) start.x (start of readSide.x/R1 read) start.y (start of readSide.y/R2 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)
`unique.umi.minus.R1`	a data frame containing unique cleavage site from R1 reads mapped to plus strand without corresponding R2 reads mapped to the minus strand, with the following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2 read) strand.x (strand of readSide.x/R1 read) end.x (end of readSide.x/R1 read) end.y (end of readSide.y/R2 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)
`all.umi`	a data frame containing all the mapped reads with the following columns. readName (read ID), chr.x (chromosome of readSide.x/R1 read), start.x (start of eadSide.x/R1 read), end.x (end of eadSide.x/R1 read), mapping.qual.x (mapping quality of readSide.x/R1 read), strand.x (strand of readSide.x/R1 read), cigar.x (CIGAR of readSide.x/R1 read) , readSide.x (1/R1), chr.y (chromosome of readSide.y/R2 read) start.y (start of readSide.y/R2 read), end.y (end of readSide.y/R2 read), mapping.qual.y (mapping quality of readSide.y/R2 read), strand.y (strand of readSide.y/R2 read), cigar.y (CIGAR of readSide.y/R2 read), readSide.y (2/R2) R1.base.kept (retained R1 length), R2.base.kept (retained R2 length), distance (distance between mapped R1 and R2), UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)

Author(s)

Lihua Julie Zhu

References

Shengdar Q Tsai and J Keith Joung et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature Biotechnology 33, 187 to 197 (2015)

See Also

getPeaks

Examples

```
if(interactive())
{
    umiFile <- system.file("extdata", "UMI-HEK293_site4_R1.txt",
        package = "GUIDEseq")
    alignFile <- system.file("extdata","bowtie2.HEK293_site4.sort.bed",
        package = "GUIDEseq")
    cleavages <- getUniqueCleavageEvents(
        alignment.inputfile = alignFile , umi.inputfile = umiFile)
    names(cleavages)
}
```

`GUIDEseqAnalysis` *Analysis pipeline for GUIDE-seq dataset*

Description

A wrapper function that uses the UMI sequence plus the first few bases of each sequence from R1 reads to estimate the starting sequence library, piles up reads with a user defined window and step size, identify the cleavage sites, merge cleavage sites from plus strand and minus strand, followed by off target analysis of extended regions around the identified cleavage sites.

Usage

```
GUIDEseqAnalysis(alignment.inputfile, umi.inputfile,
    alignment.format = "bed",
    umi.header = FALSE, read.ID.col = 1,
    umi.col = 2, umi.sep = "\t",
    BSgenomeName,
    gRNA.file,
    outputDir,
    keep.R1only = TRUE, keep.R2only = TRUE,
    paired.direction = "opposite.strand",
    max.paired.distance = 1000, min.mapping.quality = 30,
    max.R1.len = 130, max.R2.len = 130,
    apply.both.max.len = FALSE, same.chromosome = TRUE,
    distance.inter.chrom = -1, min.R1.mapped = 30,
    min.R2.mapped = 30, apply.both.min.mapped = FALSE,
    max.duplicate.distance = 0,
    umi.plus.R1start.unique = TRUE, umi.plus.R2start.unique = TRUE,
    window.size = 20L, step = 20L, bg.window.size = 5000L,
    min.reads = 5L, min.SNratio = 2, maxP = 0.05,
    stats = c("poisson", "nbinom"),
    p.adjust.methods =
    c( "none", "BH", "holm", "hochberg", "hommel", "bonferroni", "BY", "fdr"),
    distance.threshold = 40,
    plus.strand.start.gt.minus.strand.end = TRUE,
    gRNA.format = "fasta",
    overlap.gRNA.positions = c(17,18),
    upstream = 50, downstream = 50, PAM.size = 3, gRNA.size = 20,
    PAM = "NGG", PAM.pattern = "(NAG|NGG|NGA)$", max.mismatch = 6,
```

```
allowed.mismatch.PAM = 2, overwrite = TRUE,
weights = c(0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079,
0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615,0.804, 0.685, 0.583))
```

Arguments

`alignment.inputfile`	The alignment file. Currently supports bed output file with CIGAR information. Suggest run the workflow binReads.sh, which sequentially runs barcode binning, adaptor removal, alignment to genome, alignment quality filtering, and bed file conversion. Please download the workflow function and its helper scripts at http://mccb.umassmed.edu/GUIDE-seq/binReads/
`umi.inputfile`	A text file containing at least two columns, one is the read identifier and the other is the UMI or UMI plus the first few bases of R1 reads. Suggest use getUMI.sh to generate this file. Please download the script and its helper scripts at http://mccb.umassmed.edu/GUIDE-seq/getUMI/
`alignment.format`	The format of the alignment input file. Default bed file format. Currently only bed file format is supported, which is generated from binReads.sh
`umi.header`	Indicates whether the umi input file contains a header line or not. Default to FALSE
`read.ID.col`	The index of the column containing the read identifier in the umi input file, default to 1
`umi.col`	The index of the column containing the umi or umi plus the first few bases of sequence from the R1 reads, default to 2
`umi.sep`	column separator in the umi input file, default to tab
`BSgenomeName`	BSgenome object. Please refer to available.genomes in BSgenome package. For example, BSgenome.Hsapiens.UCSC.hg19 for hg19, BSgenome.Mmusculus.UCSC.mm10 for mm10, BSgenome.Celegans.UCSC.ce6 for ce6, BSgenome.Rnorvegicus.UCSC.rn5 for rn5, BSgenome.Drerio.UCSC.danRer7 for Zv9, and BSgenome.Dmelanogaster.UCSC.dm3 for dm3
`gRNA.file`	gRNA input file path or a DNAStringSet object that contains gRNA plus PAM sequences used for genome editing
`outputDir`	the directory where the off target analysis and reports will be written to
`keep.R1only`	Specify whether to include alignment with only R1 without paired R2. Default TRUE
`keep.R2only`	Specify whether to include alignment with only R2 without paired R1. Default TRUE
`paired.direction`	Specify whether the R1 and R2 should be aligned to the same strand or opposite strand. Default opposite.strand
`max.paired.distance`	Specify the maximum distance allowed between paired R1 and R2 reads. Default 1000 bp
`min.mapping.quality`	Specify min.mapping.quality of acceptable alignments
`max.R1.len`	The maximum retained R1 length to be considered for downstream analysis, default 130 bp. Please note that default of 130 works well when the read length 150 bp. Please also note that retained R1 length is not necessarily equal to the mapped R1 length

`max.R2.len`	The maximum retained R2 length to be considered for downstream analysis, default 130 bp. Please note that default of 130 works well when the read length 150 bp. Please also note that retained R2 length is not necessarily equal to the mapped R2 length
`apply.both.max.len`	Specify whether to apply maximum length requirement to both R1 and R2 reads, default FALSE
`same.chromosome`	Specify whether the paired reads are required to align to the same chromosome, default TRUE
`distance.inter.chrom`	Specify the distance value to assign to the paired reads that are aligned to different chromosome, default -1
`min.R1.mapped`	The maximum mapped R1 length to be considered for downstream analysis, default 30 bp.
`min.R2.mapped`	The maximum mapped R2 length to be considered for downstream analysis, default 30 bp.
`apply.both.min.mapped`	Specify whether to apply minimum mapped length requirement to both R1 and R2 reads, default FALSE
`max.duplicate.distance`	Specify the maximum distance apart for two reads to be considered as duplicates, default 0. Currently only 0 is supported
`umi.plus.R1start.unique`	To specify whether two mapped reads are considered as unique if both containing the same UMI and same alignment start for R1 read, default TRUE.
`umi.plus.R2start.unique`	To specify whether two mapped reads are considered as unique if both containing the same UMI and same alignment start for R2 read, default TRUE.
`window.size`	window size to calculate coverage
`step`	step size to calculate coverage
`bg.window.size`	window size to calculate local background
`min.reads`	minimum number of reads to be considered as a peak
`min.SNratio`	minimum signal noise ratio, which is the coverage normalized by local background
`maxP`	Maximum p-value to be considered as significant
`stats`	Statistical test, default poisson
`p.adjust.methods`	Adjustment method for multiple comparisons, default none
`distance.threshold`	Specify the maximum gap allowed between the plus stranded and the negative stranded peak, default 40. Suggest set it to twice of window.size used for peak calling.
`plus.strand.start.gt.minus.strand.end`	Specify whether plus strand peak start greater than the paired negative strand peak end. Default to TRUE
`gRNA.format`	Format of the gRNA input file. Currently, fasta is supported
`PAM.size`	PAM length, default 3

`gRNA.size`	The size of the gRNA, default 20
`PAM`	PAM sequence after the gRNA, default NGG
`overlap.gRNA.positions`	The required overlap positions of gRNA and restriction enzyme cut site, default 17 and 18 for SpCas9.
`max.mismatch`	Maximum mismatch allowed in off target search, default 6
`PAM.pattern`	Regular expression of protospacer-adjacent motif (PAM), default (NAG|NGG|NGA)$ for off target search
`allowed.mismatch.PAM`	Number of degenerative bases in the PAM sequence, default to 2 for N[A|G]G PAM
`upstream`	upstream offset from the peak start to search for off targets, default 50
`downstream`	downstream offset from the peak end to search for off targets, default 50
`weights`	a numeric vector size of gRNA length, default c(0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583) for SPcas9 system, which is used in Hsu et al., 2013 cited in the reference section. Please make sure that the number of elements in this vector is the same as the gRNA.size, e.g., pad 0s at the beginning of the vector.
`overwrite`	overwrite the existing files in the output directory or not, default FALSE

Value

`cleavage.gr`	Cleavage sites with one site per UMI as GRanges with metadata column total set to 1 for each range
`unique.umi.plus.R2`	a data frame containing unique cleavage site from R2 reads mapped to plus strand with the following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2 read) strand.x (strand of readSide.x/R1 read) start.y (start of readSide.y/R2 read) end.x (start of readSide.x/R1 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)
`unique.umi.minus.R2`	a data frame containing unique cleavage site from R2 reads mapped to minus strand with the following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2 read) strand.x (strand of readSide.x/R1 read) end.y (end of readSide.y/R2 read) start.x (start of readSide.x/R1 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)
`unique.umi.plus.R1`	a data frame containing unique cleavage site from R1 reads mapped to minus strand without corresponding R2 reads mapped to the plus strand, with the following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2 read) strand.x (strand of readSide.x/R1 read) start.x (start of readSide.x/R1 read) start.y (start of readSide.y/R2 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)
`unique.umi.minus.R1`	a data frame containing unique cleavage site from R1 reads mapped to plus strand without corresponding R2 reads mapped to the minus strand, with the

	following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2 read) strand.x (strand of readSide.x/R1 read) end.x (end of readSide.x/R1 read) end.y (end of readSide.y/R2 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)
`all.umi`	a data frame containing all the mapped reads with the following columns. readName (read ID), chr.x (chromosome of readSide.x/R1 read), start.x (start of eadSide.x/R1 read), end.x (end of eadSide.x/R1 read), mapping.qual.x (mapping quality of readSide.x/R1 read), strand.x (strand of readSide.x/R1 read), cigar.x (CIGAR of readSide.x/R1 read) , readSide.x (1/R1), chr.y (chromosome of readSide.y/R2 read) start.y (start of readSide.y/R2 read), end.y (end of readSide.y/R2 read), mapping.qual.y (mapping quality of readSide.y/R2 read), strand.y (strand of readSide.y/R2 read), cigar.y (CIGAR of readSide.y/R2 read), readSide.y (2/R2), R1.base.kept (retained R1 length), R2.base.kept (retained R2 length), distance (distance between mapped R1 and R2), UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)

Author(s)

Lihua Julie Zhu

References

Shengdar Q Tsai and J Keith Joung et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature Biotechnology 33, 187 to 197 (2015)

See Also

getPeaks

Examples

```
if(interactive())
    {
        library("BSgenome.Hsapiens.UCSC.hg19")
        umiFile <- system.file("extdata", "UMI-HEK293_site4_R1.txt",
           package = "GUIDEseq")
        alignFile <- system.file("extdata","bowtie2.HEK293_site4.sort.bed" ,
            package = "GUIDEseq")
        gRNA.file <- system.file("extdata","gRNA.fa", package = "GUIDEseq")
        guideSeqRes <- GUIDEseqAnalysis(
            alignment.inputfile = alignFile,
            umi.inputfile = umiFile, gRNA.file = gRNA.file,
            BSgenomeName = Hsapiens, min.reads = 80)
        names(guideSeqRes)
   }
```

`mergePlusMinusPeaks` *Merge peaks from plus strand and minus strand*

Description

Merge peaks from plus strand and minus strand with required orientation and within certain distance apart

Usage

```
mergePlusMinusPeaks(peaks.gr, peak.height.mcol = "count",
    bg.height.mcol = "bg", distance.threshold = 40, step = 20,
    plus.strand.start.gt.minus.strand.end = TRUE, output.bedfile)
```

Arguments

`peaks.gr`	Specify the peaks as GRanges object, which should contain peaks from both plus and minus strand. In addition, it should contain peak height metadata column to store peak height and optionally background height.
`peak.height.mcol`	Specify the metadata column containing the peak height, default to count
`bg.height.mcol`	Specify the metadata column containing the background height, default to bg
`distance.threshold`	Specify the maximum gap allowed between the plus stranded and the negative stranded peak, default 40. Suggest set it to twice of window.size used for peak calling.
`step`	Specify the cushion distance if plus.strand.start.gt.minus.strand.end is set to TRUE. Default 20. Suggest set it to be the step size used for the peak calling.
`plus.strand.start.gt.minus.strand.end`	Specify whether plus strand peak start greater than the paired negative strand peak end. Default to TRUE
`output.bedfile`	Specify the bed output file name, which is used for off target analysis subsequently.

Value

output a list and a bed file containing the merged peaks a data frame of the bed format

`mergedPeaks.gr`	merged peaks as GRanges
`mergedPeaks.bed`	merged peaks in bed format

Author(s)

Lihua Julie Zhu

References

Zhu L.J. et al. (2010) ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics 2010, 11:237doi:10.1186/1471-2105-11-237. Zhu L.J. (2013) Integrative analysis of ChIP-chip and ChIP-seq dataset. Methods Mol Biol. 2013;1067:105-24. doi: 10.1007/978-1-62703-607-8_8.

Examples

```
if (interactive())
{
    data(peaks.gr)
    mergedPeaks <- mergePlusMinusPeaks(peaks.gr = peaks.gr,
        output.bedfile = "mergedPeaks.bed")
    mergedPeaks$mergedPeaks.gr
    head(mergedPeaks$mergedPeaks.bed)
}
```

`offTargetAnalysisOfPeakRegions` *Offtarget Analysis of GUIDE-seq peaks*

Description

Finding offtargets around peaks from GUIDE-seq or around any given genomic regions

Usage

```
offTargetAnalysisOfPeakRegions(gRNA, peaks,
    format=c("fasta", "bed"),
    peaks.withHeader = FALSE, BSgenomeName, overlap.gRNA.positions = c(17,18),
    upstream = 50, downstream =50, PAM.size = 3, gRNA.size = 20,
    PAM = "NGG", PAM.pattern = "(NAG|NGG|NGA)$", max.mismatch = 6,
    outputDir, allowed.mismatch.PAM = 2, overwrite = TRUE,
    weights = c(0, 0, 0.014, 0, 0, 0.395,
    0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615,
    0.804, 0.685, 0.583)
    )
```

Arguments

`gRNA`	gRNA input file path or a DNAStringSet object that contains gRNA plus PAM sequences used for genome editing
`peaks`	peak input file path or a GenomicRanges object that contains genomic regions to be searched for potential offtargets
`format`	Format of the gRNA and peak input file. Currently, fasta and bed are supported for gRNA and peak input file respectively
`peaks.withHeader`	Indicate whether the peak input file contains header, default FALSE
`PAM.size`	PAM length, default 3
`gRNA.size`	The size of the gRNA, default 20
`PAM`	PAM sequence after the gRNA, default NGG
`BSgenomeName`	BSgenome object. Please refer to available.genomes in BSgenome package. For example, BSgenome.Hsapiens.UCSC.hg19 for hg19, BSgenome.Mmusculus.UCSC.mm10 for mm10, BSgenome.Celegans.UCSC.ce6 for ce6, BSgenome.Rnorvegicus.UCSC.rn5 for rn5, BSgenome.Drerio.UCSC.danRer7 for Zv9, and BSgenome.Dmelanogaster.UCSC.dm3 for dm3

`overlap.gRNA.positions`	The required overlap positions of gRNA and restriction enzyme cut site, default 17 and 18 for SpCas9.
`max.mismatch`	Maximum mismatch allowed in off target search, default 6
`PAM.pattern`	Regular expression of protospacer-adjacent motif (PAM), default (NAG|NGG|NGA)$ for off target search
`allowed.mismatch.PAM`	Number of degenerative bases in the PAM sequence, default to 2 for N[A|G]G PAM
`outputDir`	the directory where the off target analysis and reports will be written to
`upstream`	upstream offset from the peak start to search for off targets, default 50
`downstream`	downstream offset from the peak end to search for off targets, default 50
`weights`	a numeric vector size of gRNA length, default c(0, 0, 0.014, 0, 0, 0.395, 0.317, 0, 0.389, 0.079, 0.445, 0.508, 0.613, 0.851, 0.732, 0.828, 0.615, 0.804, 0.685, 0.583) for SPcas9 system, which is used in Hsu et al., 2013 cited in the reference section. Please make sure that the number of elements in this vector is the same as the gRNA.size, e.g., pad 0s at the beginning of the vector.
`overwrite`	overwrite the existing files in the output directory or not, default FALSE

Value

a tab-delimited file offTargetsInPeakRegions.tsv, containing all input peaks with potential gRNA binding sites, mismatch number and positions, alignment to the input gRNA and predicted cleavage score.

Author(s)

Lihua Julie Zhu

References

Patrick D Hsu, David A Scott, Joshua A Weinstein, F Ann Ran, Silvana Konermann, Vineeta Agarwala, Yinqing Li, Eli J Fine, Xuebing Wu, Ophir Shalem,Thomas J Cradick, Luciano A Marraffini, Gang Bao & Feng Zhang (2013) DNA targeting specificity of rNA-guided Cas9 nucleases. Nature Biotechnology 31:827-834 Lihua Julie Zhu, Benjamin R. Holmes, Neil Aronin and Michael Brodsky. CRISPRseek: a Bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. Plos One Sept 23rd 2014 Lihua Julie Zhu (2015). Overview of guide RNA design tools for CRISPR-Cas9 genome editing technology. Frontiers in Biology August 2015, Volume 10, Issue 4, pp 289-296

See Also

GUIDEseq

Examples

```
if (interactive()) {
    library("BSgenome.Hsapiens.UCSC.hg19")
    peaks <- system.file("extdata", "T2plus100OffTargets.bed",
        package = "CRISPRseek")
    gRNAs <- system.file("extdata", "T2.fa",
        package = "CRISPRseek")
```

```
    outputDir = getwd()
    offTargets <- offTargetAnalysisOfPeakRegions(gRNA = gRNAs, peaks = peaks,
        format=c("fasta", "bed"),
        peaks.withHeader = TRUE, BSgenomeName = Hsapiens,
        upstream = 50, downstream =50, PAM.size = 3, gRNA.size = 20,
        PAM = "NGG", PAM.pattern = "NNN$", max.mismatch = 2,
        outputDir = outputDir,
        allowed.mismatch.PAM = 3, overwrite = TRUE
   )
}
```

`peaks.gr` *example cleavage sites*

Description

An example data set containing cleavage sites (peaks) from getPeaks

Usage

```
data("peaks.gr")
```

Format

GRanges with count (peak height), bg (local background), SNratio (signal noise ratio), p-value, and option adjusted p-value

Value

`peaks.gr`	GRanges with count (peak height), bg (local background), SNratio (signal noise ratio), p-value, and option adjusted p-value

Source

http://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR1695644

Examples

```
data(peaks.gr)
names(peaks.gr)
peaks.gr
```

`uniqueCleavageEvents` *example unique cleavage sites*

Description

An example data set containing cleavage sites with unique UMI, generated from getUniqueCleavageEvents

Usage

```
data("uniqueCleavageEvents")
```

Value

cleavage.gr Cleavage sites with one site per UMI as GRanges with metadata column total set to 1 for each range

unique.umi.plus.R2 a data frame containing unique cleavage site from R2 reads mapped to plus strand with the following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2 read) strand.x (strand of readSide.x/R1 read) start.y (start of readSide.y/R2 read) end.x (start of readSide.x/R1 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)

unique.umi.minus.R2 a data frame containing unique cleavage site from R2 reads mapped to minus strand with the following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2 read) strand.x (strand of readSide.x/R1 read) end.y (end of readSide.y/R2 read) start.x (start of readSide.x/R1 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)

unique.umi.plus.R1 a data frame containing unique cleavage site from R1 reads mapped to minus strand without corresponding R2 reads mapped to the plus strand, with the following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2 read) strand.x (strand of readSide.x/R1 read) start.x (start of readSide.x/R1 read) start.y (start of readSide.y/R2 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)

unique.umi.minus.R1 a data frame containing unique cleavage site from R1 reads mapped to plus strand without corresponding R2 reads mapped to the minus strand, with the following columns chr.y (chromosome of readSide.y/R2 read) chr.x (chromosome of readSide.x/R1 read) strand.y (strand of readSide.y/R2 read) strand.x (strand of readSide.x/R1 read) end.x (end of readSide.x/R1 read) end.y (end of readSide.y/R2 read) UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)

all.umi a data frame containing all the mapped reads with the following columns. readName (read ID), chr.x (chromosome of readSide.x/R1 read), start.x (start of eadSide.x/R1 read), end.x (end of eadSide.x/R1 read), mapping.qual.x (mapping quality of readSide.x/R1 read), strand.x (strand of readSide.x/R1 read), cigar.x (CIGAR of readSide.x/R1 read) , readSide.x (1/R1), chr.y (chromosome of readSide.y/R2 read) start.y (start of readSide.y/R2 read), end.y (end of readSide.y/R2 read), mapping.qual.y (mapping quality of readSide.y/R2 read), strand.y (strand of readSide.y/R2 read), cigar.y (CIGAR of readSide.y/R2 read), readSide.y (2/R2) R1.base.kept (retained R1 length), R2.base.kept (retained R2 length), distance (distance between mapped R1 and R2), UMI (unique molecular identifier (umi) or umi with the first few bases of R1 read)

Source

http://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR1695644

Examples

```
data(uniqueCleavageEvents)
names(uniqueCleavageEvents)
str(uniqueCleavageEvents)
```

Index