Package 'riboSeq'

September 11, 2014

Type Package

Title Analysis of sequencing data from ribosome profiling experiments.

Description Plotting functions, frameshift detection and parsing of sequencing
data from ribosome profiling experiments.

Version 0.99.14

Date 2014-08-14

Author Thomas J. Hardcastle

Maintainer Thomas J. Hardcastle <tjh48@cam.ac.uk>

License GPL-3

LazyLoad yes

Depends R (>= 3.0.2), methods, GenomicRanges, abind

Suggests baySeq, BiocStyle, RUnit, BiocGenerics

biocViews Sequencing,Genetics,Visualization

R topics documented:

---

riboSeq-package   *Analysis of sequencing data from ribosome profiling experiments.*

Description

Plotting functions, frameshift detection and parsing of sequencing data from ribosome profiling experiments.

Details

Package:	riboSeq
Type:	Package
Version:	1.0.0
Date:	2014-05-02
License:	GPL-3
LazyLoad:	yes
Depends:	R (>= 3.0.2), methods, baySeq, GenomicRanges, abind
Imports:	GenomicRanges, methods, abind
Packaged:	2014-05-16 10:49:24 UTC; tjh48
Built:	R 3.0.2; ; 2014-05-16 10:49:44 UTC; unix

Further information is available in the following vignettes:

riboSeq   riboSeq (source, pdf)

Author(s)

Thomas J. Hardcastle

Maintainer: Thomas J. Hardcastle <tjh48@cam.ac.uk>

---

filterHits   *Filters framecalled data based on the mean number of hits observed across a replicate group.*

Description

If no ribosomal footprints of the correct lengths (and frames) are seen at a coding sequence in any replicate group, this sequence is unlikely to be translated (and is therefore likely to be uninteresting). This function filters out these coding sequences, leaving only those with a minimum number of hits in at least one replicate group, and a minimum number of unique sequences aligning in at least one replicate group (to exclude single stacks of sequenced reads passing the filtering).

Usage

```
filterHits(fCs, lengths = 27, frames, hitMean = 10, unqhitMean = 1)
```

Arguments

fCs	riboCoding object defining the number of ribosome footprint reads observed over a set of coordinates, their lengths, and their frame relative to coding start sites.
lengths	The lengths of ribosome footprint reads to be used in filtering.
frames	If given, the frames of the ribosome footprint reads to be used in filtering. Should be of equal length to the 'lengths' parameter - see examples.
hitMean	The mean number of hits within a replicate group that should be observed to pass filtering.
unqhitMean	The mean number of unique sequences within a replicate group that should be observed to pass filtering. This parameter is intended to avoid cases where a coding sequence is deemed to be expressed based on a few highly expressed sequences.

Details

Frames can be given as a single vector (which specifies the frames used for all lengths of footprints, or as a list, specifying the frame for each length given in 'lengths'.

Value

riboCoding object containing the filtered coding sequences and the associated numbers of ribosome footprint reads.

Author(s)

Thomas J. Hardcastle

See Also

frameCounting

Examples

```
#ribosomal footprint data
datadir <- system.file("extdata", package = "riboSeq")
ribofiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(17,3,5,7), sep = "")
rnafiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(10,12,14,16), sep = "")

riboDat <- readRibodata(ribofiles, rnafiles, replicates = c("WT", "WT",
"M", "M"))

# CDS coordinates
```

```
chlamyFasta <- paste(datadir, "/rsem_chlamy236_deNovo.transcripts.fa", sep = "")
fastaCDS <- findCDS(fastaFile = chlamyFasta,
                    startCodon = c("ATG"),
                    stopCodon = c("TAG", "TAA", "TGA"))

# frame calling
fCs <- frameCounting(riboDat, fastaCDS)

# filter coding sequences. 27-mers are principally in the 0-frame,
# 28-mers are principally in the 2-frame relative to coding start (see
# readingFrame function).

ffCs <- filterHits(fCs, lengths = c(27, 28), frames = list(0, 2),
                   hitMean = 50, unqhitMean = 10)
```

---

findCDS   *Parses a transcriptome file looking for start/stop codons in frame.*

Description

Looks in the fastaFile for defined start and stop codons in frame with one another. Reports the discovered coordinates as a GRanges object with a 'frame' value.

Usage

```
findCDS(fastaFile, startCodon = c("ATG"), stopCodon = c("TAG", "TAA", "TGA"))
```

Arguments

fastaFile	Fasta file of transcriptome sequences.
startCodon	Vector of possible start codons. Defaults to "ATG".
stopCodon	Vector of possible stop codons. Defaults to c("TAG", "TAA", "TGA").

Value

A GRanges object.

Author(s)

Thomas J. Hardcastle

Examples

```
datadir <- system.file("extdata", package = "riboSeq")
chlamyFasta <- paste(datadir, "/rsem_chlamy236_deNovo.transcripts.fa", sep = "")
fastaCDS <- findCDS(fastaFile = chlamyFasta,
                    startCodon = c("ATG"),
                    stopCodon = c("TAG", "TAA", "TGA"))
```

---

frameCounting	*Counts aligned reads within coding sequence regions by frame and footprint size, splitting by frame and footprint size.*

Description

Ribosome footprint sequencing reads aligning within coding sequence regions may align in the same frame (relative to start codon) as the coding sequence, or frame shifted by 1 or 2 frames. This function calls the number of aligning reads within the coding sequence, split by frame and footprint size.

Usage

```
frameCounting(riboDat, fastaCDS, lengths = 25:30)
```

Arguments

riboDat	A `riboData` object containing the ribosome footprints to be counted.
fastaCDS	A `GenomicRanges` object containing the coordinates of the coding sequences.
lengths	Lengths of ribosome footprints to be included in the `riboData` object.

Value

A `riboCoding` object.

Author(s)

Thomas J. Hardcastle

See Also

`riboData`

Examples

```
#ribosomal footprint data
datadir <- system.file("extdata", package = "riboSeq")
ribofiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(17,3,5,7), sep = "")
rnafiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(10,12,14,16), sep = "")

riboDat <- readRibodata(ribofiles, rnafiles, replicates = c("WT", "WT",
"M", "M"))

# CDS coordinates
chlamyFasta <- paste(datadir, "/rsem_chlamy236_deNovo.transcripts.fa", sep = "")
```

```
fastaCDS <- findCDS(fastaFile = chlamyFasta,
                    startCodon = c("ATG"),
                    stopCodon = c("TAG", "TAA", "TGA"))

# frame calling
fCs <- frameCounting(riboDat, fastaCDS)
```

---

plotCDS	*Plots average ribosome footprint alignment to coding sequences at 5'* *and 3' ends.*

Description

For each sample, the average (normalised by translation abundance over transcript) of the ribosome footprints of a given length alignments at the 5' and 3' ends of all specified transcripts beginning at each base relative to coding start/end are plotted. The bases are colour coded relative to start codon.

Usage

```
plotCDS(coordinates, riboDat, lengths = 27, min5p = -20, max5p = 200,
min3p = -200, max3p = 20, cap, main = "", plot = TRUE)
```

Arguments

coordinates	Coordinates (as a `GRanges` object) of the coding sequences.
riboDat	[riboData](#) object containing ribosome footprint data.
lengths	Lengths of footprints to be plotted. May be given as a vector, in which case multiple plots will be produced.
min5p	The distance upstream of the translation start to be plotted.
max5p	The distance downstream of the translation start to be plotted.
min3p	The distance upstream of the translation end to be plotted.
max3p	The distance downstream of the translation end to be plotted.
cap	If given, caps the height of plotted values.
main	Title of the plot.
plot	Should the acquired matrix of mean expression be plotted? Defaults to TRUE.

Value

Invisibly returns lists of lists of matrices containing weighted averages plotted for each sample/length combination.

Author(s)

Thomas J. Hardcastle

Examples

```
#ribosomal footprint data
datadir <- system.file("extdata", package = "riboSeq")
ribofiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(17,3,5,7), sep = "")
rnafiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(10,12,14,16), sep = "")

riboDat <- readRibodata(ribofiles, rnafiles, replicates = c("WT", "WT",
"M", "M"))

# CDS coordinates
chlamyFasta <- paste(datadir, "/rsem_chlamy236_deNovo.transcripts.fa", sep = "")
fastaCDS <- findCDS(fastaFile = chlamyFasta,
                    startCodon = c("ATG"),
                    stopCodon = c("TAG", "TAA", "TGA"))

# frame calling
fCs <- frameCounting(riboDat, fastaCDS)

# filter coding sequences. 27-mers are principally in the 0-frame,
# 28-mers are principally in the 2-frame relative to coding start (see
# readingFrame function).

ffCs <- filterHits(fCs, lengths = c(27, 28), frames = list(0, 2),
                    hitMean = 50, unqhitMean = 10)

# plotCDS(coordinates = ffCs@CDS, riboDat = riboDat, lengths = 27)
```

---

plotTranscript	*Plots ribosome footprint abundance and mRNA coverage (if available) for a specific transcript.*

Description

Abundances of ribosomal footprints of a given size class are plotted on a transcript. The footprints are colour coded according to the first base of the transcript, and not any coding start site, to allow for multiple coding start sites on a given transcript. Coding regions may simultaneously be plotted and colour coded under the same scheme.

Usage

```
plotTranscript(transcript, coordinates, annotation, riboData, length =
27, frameShift = 0, cap, riboScale, rnaScale, baseLim, main, note = "", ...)
```

Arguments

transcript	The name of the transcript to be plotted.
coordinates	A GRanges object containing any coding regions on the transcript.
annotation	A GRanges object containing annotated coding coordinates to be plotted as bars above the figure.
riboData	A [riboData](#) object containing the ribosome footprint (and optionally, RNA-seq) data.
length	Size class of ribosome footprint data to be plotted.
frameShift	Frameshift for the ribosome footprint data. See Details.
cap	Cap on the largest value that will be plotted as an abundance of the ribosome footprint data.
riboScale	Scale to be used on the ribosome footprint axis.
rnaScale	Scale to be used on the RNA-seq coverage axis.
baseLim	Limits of the bases of the transcript to be plotted (i.e., the x-axis). If missing, the full transcript will be plotted.
main	Optional title for the plot.
note	Additional note to be added to plot titles (in addition to transcript and sample names).
...	Additional arguments to be passed to plotting function.

Details

The readingFrame value allows the colour-coding of the ribosome footprints to be shifted so that the colours of the coding sequences match the colours of the ribosome footprint data. E.g., if 28-mers are predominantly in frame 2 relative to coding start, a value of 'readingFrame=2' will ensure that 28-mers in a coding region will take the same colour as that coding region if they are in the correct relative frame.

Value

NULL; plotting function.

Author(s)

Thomas J. Hardcastle

Examples

```
#ribosomal footprint data
datadir <- system.file("extdata", package = "riboSeq")
ribofiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(17,3,5,7), sep = "")
rnafiles <- paste(datadir,
                  "/chlamy236_plus_deNovo_plusOnly_Index", c(10,12,14,16), sep = "")

riboDat <- readRibodata(ribofiles, rnafiles, replicates = c("WT", "WT",
```

```
"M", "M"))

# CDS coordinates
chlamyFasta <- paste(datadir, "/rsem_chlamy236_deNovo.transcripts.fa", sep = "")
fastaCDS <- findCDS(fastaFile = chlamyFasta,
                    startCodon = c("ATG"),
                    stopCodon = c("TAG", "TAA", "TGA"))

# frame calling
fCs <- frameCounting(riboDat, fastaCDS)

# filter coding sequences. 27-mers are principally in the 0-frame,
# 28-mers are principally in the 2-frame relative to coding start (see
# readingFrame function).

ffCs <- filterHits(fCs, lengths = c(27, 28), frames = list(0, 2),
                   hitMean = 50, unqhitMean = 10)

# plot ribosome footprint profile for CUFF.37930.1

plotTranscript("CUFF.37930.1", coordinates = ffCs@CDS,
               riboData = riboDat, length = 27, cap = 200)
```

---

readingFrame	*Analyses frame called ribosome footprint data within coding sequences and identifies likely frame-shift of different length ribosome footprint reads.*

Description

Ribosome footprint data data can be used to identify the frame-shift relative to start codon of the different n-mers. For each readlength specified, the sum of alignments in the different frames is shown, together with the maximum likelihood frame.

Usage

```
readingFrame(coordinates, riboDat, rC, lengths = 26:30)
plotFS(fS, lengths, legend.text = c("Frame 0", "Frame 1", "Frame 2"), ...)
```

Arguments

coordinates	Coordinates of transcripts. Need not be given if a riboCoding object is specified in 'rC'.
riboDat	A riboData object containing the ribosome footprint data. Need not be given if a riboCoding object is specified in 'rC'.
rC	A riboCoding object.

lengths	Lengths of reads to be analysed for frame-shift, or to be plotted. May be omitted in plotting, in which case all lengths will be plotted.
fS	The output of the readingFrame function, to be plotted.
legend.text	Text for legend.
...	Additional arguments to be passed to barplot function.

Value

A matrix giving the number of aligned reads in each frame for each length, and the maximum likelihood frame.

Author(s)

Thomas J. Hardcastle

See Also

[frameCounting](#)

Examples

```
#ribosomal footprint data
datadir <- system.file("extdata", package = "riboSeq")
ribofiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(17,3,5,7), sep = "")
rnafiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(10,12,14,16), sep = "")

riboDat <- readRibodata(ribofiles, rnafiles, replicates = c("WT", "WT",
"M", "M"))

# CDS coordinates
chlamyFasta <- paste(datadir, "/rsem_chlamy236_deNovo.transcripts.fa", sep = "")
fastaCDS <- findCDS(fastaFile = chlamyFasta,
                    startCodon = c("ATG"),
                    stopCodon = c("TAG", "TAA", "TGA"))

# frame calling
rCs <- frameCounting(riboDat, fastaCDS)

# analysis of frame shift for 27 and 28-mers.
fS <- readingFrame(rC = rCs, lengths = 27:28)
plotFS(fS)
```

Description

Reads flat text files (or compressed versions of these) containing strand, transcript name, start and sequence information for each alignment.

Usage

```
readRibodata(riboFiles, rnaFiles, columns = c(strand = 1, seqname = 2,
start = 3, sequence = 4), zeroIndexed = TRUE, header = FALSE, replicates, seqnames)
```

Arguments

riboFiles	Filenames of ribosomal alignments.
rnaFiles	Filenames of RNA alignments.
columns	Columns of alignment files containing strand, transcript (seqname) name, start of alignment, and sequence.
zeroIndexed	Are the alignments zero-indexed (i.e., the first base in a sequence is 0). Defaults to TRUE, which will result in an adjustment to 1-indexed data.
header	Does the alignment file have a header line? Defaults to FALSE.
replicates	Replicate information for the files.
seqnames	Transcript (seqname) names to be read into the object.

Value

`riboData` object.

Author(s)

Thomas J. Hardcastle

Examples

```
datadir <- system.file("extdata", package = "riboSeq")
ribofiles <- paste(datadir,
                    "/chlamy236_plus_deNovo_plusOnly_Index", c(17,3,5,7), sep = "")
rnafiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(10,12,14,16), sep = "")

riboDat <- readRibodata(ribofiles, rnafiles, replicates = c("WT", "WT", "M", "M"))
```

---

riboCoding-class   *Class "riboCoding"*

Description

The `riboCoding` class contains a set of coordinates defining coding sequences, a set of replicate data
for the experimental samples, an array of ribosome footprint abundances for each coding sequence
split by size class and frame, and a similar array describing the abundance of unique sequences
aligning within each coding sequence.

Slots

CDS: Object of class `"GRanges"` defining the coordinates of coding sequences.

hits: An array describing the abundances of ribosome footprints, split by size class and frame
(relative to coding start) for each of the coding sequences defined in the 'CDS' slot.

unqHits: An array describing the abundances of unique sequences of ribosome footprints, split by
size class and frame (relative to coding start) for each of the coding sequences defined in the
'CDS' slot.

replicates: A factor defining the replicate structure of the samples for which data are held. Repli-
cate samples will have the same level in this factor.

Methods

Methods '[' and 'show' are defined for this class.

Author(s)

Thomas J. Hardcastle

---

riboData-class   *Class "riboData"*

Description

The `riboData` class contains a list of GRanges objects containing ribosome footprint alignment
data, a factor defining the replicate structure of the samples involved, and (optionally) a list of
GRanges objects containing RNA-seq alignment data (paired with the ribosome footprint data). It
will generally be created by the 'readRibodata' function and not directly by the user.

Slots

riboGR: List of `"GRanges"` objects (one for each sequenced sample) describing the alignments of
ribosomal footprint data to a transcriptome.

rnaGR: List of `"GRanges"` objects (one for each sequenced sample, paired with 'riboGR' slot)
describing the alignments of RNA-seq data to a transcriptome.

replicates: A factor defining the replicate structure of the samples for which data are held. Repli-
cate samples will have the same level in this factor.

Methods

No methods are currently defined for this class.

Author(s)

Thomas J. Hardcastle

---

rnaCounts	*Extracts mRNA counts from a riboDat object for a set of coding sequence coordinates.*

Description

Takes mRNA count data from riboDat object, maps them to coding sequences specified in GRanges object, and counts the total number of hits. This is a crude approach intended to quickly produce comparable data to ribosome footprint counts. More sophisticated alternatives, addressing coverage variation, isoforms, multireads &c. have been widely described in the literature on mRNA-seq analyses.

Usage

```
rnaCounts(riboDat, CDS)
```

Arguments

riboDat	A riboData object containing the RNA-seq alignments.
CDS	A GRanges object defining the coordinates of the coding sequences for which to acquire counts.

Details

The count data thus acquired can be compared to counts of ribosomal footprint data through a beta-binomial analysis (see vignette) to discover differential translation.

Value

A matrix containing count data for the RNA-seq libraries.

Author(s)

Thomas J. Hardcastle

See Also

sliceCounts

Examples

```
#ribosomal footprint data
datadir <- system.file("extdata", package = "riboSeq")
ribofiles <- paste(datadir,
                    "/chlamy236_plus_deNovo_plusOnly_Index", c(17,3,5,7), sep = "")
rnafiles <- paste(datadir,
                    "/chlamy236_plus_deNovo_plusOnly_Index", c(10,12,14,16), sep = "")

riboDat <- readRibodata(ribofiles, rnafiles, replicates = c("WT", "WT",
"M", "M"))

# CDS coordinates
chlamyFasta <- paste(datadir, "/rsem_chlamy236_deNovo.transcripts.fa", sep = "")
fastaCDS <- findCDS(fastaFile = chlamyFasta,
                    startCodon = c("ATG"),
                    stopCodon = c("TAG", "TAA", "TGA"))

# frame calling
fCs <- frameCounting(riboDat, fastaCDS)

# filter coding sequences. 27-mers are principally in the 0-frame,
# 28-mers are principally in the 2-frame relative to coding start (see
# readingFrame function).

ffCs <- filterHits(fCs, lengths = c(27, 28), frames = c(0, 2),
                    hitMean = 50, unqhitMean = 10)

# Extract counts of RNA hits from riboCount data.
rnaCounts <- rnaCounts(riboDat, ffCs@CDS)
```

---

sliceCounts	*Slices out count data from riboCoding object for use in differential translation analyses.*

Description

For any given coding sequence, multiple lengths of reads in various frames (relative to coding start) may align. This function extracts specific size-classes and frames of ribosome footprint reads and sums them to give a single value for each coding sequence and each sequencing library, for use in downstream analysis.

Usage

```
sliceCounts(rC, lengths = 27, frames)
```

Arguments

rC   A [riboCoding](#) object containing the coordinates of the coding sequences and
            the number of ribosomal footprint reads of various classes to be found in each.

lengths	Lengths of ribosome footprints to inform count data.
frames	Frames of ribosome footprints (relative to coding start site). If omitted, all frames are used.

Details

Frames can be given as a single vector (which specifies the frames used for all lengths of footprints, or as a list, specifying the frame for each length given in 'lengths'.

The count data thus acquired can be compared to counts of RNA-seq data through a beta-binomial analysis (see vignette) to discover differential translation.

Value

A matrix containing counts of ribosomal footprint matches to coding sequences specified in riboCoding object 'rC'.

Author(s)

Thomas J. Hardcastle

See Also

rnaCounts

Examples

```
#ribosomal footprint data
datadir <- system.file("extdata", package = "riboSeq")
ribofiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(17,3,5,7), sep = "")
rnafiles <- paste(datadir,
                   "/chlamy236_plus_deNovo_plusOnly_Index", c(10,12,14,16), sep = "")

riboDat <- readRibodata(ribofiles, rnafiles, replicates = c("WT", "WT",
"M", "M"))

# CDS coordinates
chlamyFasta <- paste(datadir, "/rsem_chlamy236_deNovo.transcripts.fa", sep = "")
fastaCDS <- findCDS(fastaFile = chlamyFasta,
                    startCodon = c("ATG"),
                    stopCodon = c("TAG", "TAA", "TGA"))

# frame calling
fCs <- frameCounting(riboDat, fastaCDS)

# filter coding sequences. 27-mers are principally in the 0-frame,
# 28-mers are principally in the 2-frame relative to coding start (see
# readingFrame function).

ffCs <- filterHits(fCs, lengths = c(27, 28), frames = list(0, 2),
                   hitMean = 50, unqhitMean = 10)
```

```
# Extract counts of ribosomal footprints from riboCount data.

riboCounts <- sliceCounts(ffCs, lengths = c(27, 28), frames = list(0,
2))
```

Index