Package 'CYCLeR'

July 1, 2022

Type Package

Title CircRNA transcriptome assembly tool

Version 1.4.2

Description CYCLeR is a software package for assembly of circRNA transcripts from RNA-seq data. Takes a set of BSJ prediction files and RNA-seq BAM files as an input and outputs circRNA trascnripts as FASTA, GTF and flat annotation files. The tools also outputs a padded FASTA to serve as an index for transcript EM abundance estimation.

License GNU General Public License (v3)

LazyData yes

Depends R (>= 4.0),
Rsamtools (>= 1.31.2),
SummarizedExperiment,
methods,
tidyverse (>= 1.3.1),
SGSeq (>= 1.23.2),
igraph (>= 1.2.6),
DEXSeq,
rtracklayer (>= 1.52.1),
polyester,
Rsubread (>= 2.8.2)

Imports AnnotationDbi,
BiocGenerics,
Biostrings,
GenomicAlignments,
GenomicFeatures,
GenomeInfoDb,
RUnit,
S4Vectors,
grDevices,
graphics,
igraph,
parallel,
rtracklayer,
stats,
tidyverse,
SGSeq,
igraph,
DEXSeq,
polyester

Suggests BiocStyle,
BSgenome.Hsapiens.UCSC.hg38,
TxDb.Hsapiens.UCSC.hg38.knownGene,
knitr,
rmarkdown

VignetteBuilder knitr

biocViews TranscriptomeAssembly, RNASeq, Transcription

RoxygenNote 7.2.0

Encoding UTF-8

R topics documented:

`combine_two_BSJ_tables` *combine BSJs*

Description

Combine 2 BSJ tables

Usage

```
combine_two_BSJ_tables(ce_bsjs, ciri_bsjs)
```

Arguments

`ce_bsjs`	BSJ table 1
`ciri_bsjs`	BSJ table 2

Details

Just a combination of BSJ tables to make sure we have a complete set of BSJs. The variable names do not actually matter since the all tables have the same formatting.

Value

`Tibble` object with combined filtered BSJ coordinate and number of junction spanning reads across sample.

Author(s)

Stefan Stefanov

`filter_bam` *BAM file filter*

Description

A wrapper function for samtools use to trim the files

Usage

```
filter_bam(BSJ_gr, sample_table, samtools_prefix)
```

Arguments

`BSJ_gr`	a GRange of BSJ cooredinates
`sample_table`	sample table formatted according to the manual, Must contain “sample_name” “treatment” “file_bam” “lib_size” “read_len”; NB the values in column “treatment” can only be “control” and “enriched”
`samtools_prefix`	a string that corresponds to user’s samtools run prefix

Details

This function removes the BAM file reads that do not overlap with the BSJ loci. This significantly speeds us the feature detection and lowers the virtual memory requirements

Value

`BAMFileList` object with info on the trimmed files

Author(s)

Stefan Stefanov

find_depleted_features *CircRNA feature selection*

Description

CircRNA feature selection

Usage

```
find_depleted_features(circ_fc_adj, sample_table, circ_sg, test = "DEX")
```

Arguments

circ_fc_adj	count matrix corresponding to the circRNA features
sample_table	sample table formatted according to the manual, Must contain “sample_name” “treatment” “file_bam” “lib_size” “read_len”; NB the values in column “treatment” can only be “control” and “enriched”
circ_sg	SGSeq object supplying feature info
test	either “DEX” for DEXSeq based feature selection or “comparison” simple average comaparison

Details

This function works in 2 ways: direct comparison of average quantities or as a wrapper of DEXSeq. In case of dataset with replicates, the suggested approach is the use of DEXSeq statistical test.

Value

vector of featureID

Author(s)

Stefan Stefanov

get_seqs *Extract sequence per exon bin*

Description

A wrapper function for BSgenome subsequencing

Usage

```
get_seqs(full_sg, bs_genome)
```

Arguments

full_sg	a SGSeq object of exon bins
bs_genome	“BSgenome” name

Details

Extracts sequence based on `SGSeq` object and "BSgenome" name

Value

sequence list

Author(s)

Stefan Stefanov

`make_BSJ_gr` *Convert BSJ string to GRanges obejct*

Description

Convert BSJ string to GRanges obejct

Usage

```
make_BSJ_gr(BSJ_set)
```

Arguments

`BSJ_set`	a list of BSJ ID records procudes by `process_BSJs` or `combine_two_BSJ_tables`

Details

Convert BSJ string to `GRanges` obejct

Value

`GRanges` object indicating BSJ loci

Author(s)

Stefan Stefanov

make_BSJ_sg *Preparation of the BSJ-specific splice graphs*

Description

Selection of the exons based on BSJ set

Usage

```
make_BSJ_sg(circ_sg, BSJ_gr)
```

Arguments

circ_sg	SGSeq prediction object
BSJ_gr	a GRange of BSJ coordinates

Details

Selection of the exons based on BSJ set

Value

SGSeq containing exons belonging to BSJ loci

Author(s)

Stefan Stefanov

merge_fasta *Merging 2* DNAStringSets

Description

Merging 2 DNAStringSets Merging 2 DNAStringSets

Usage

```
merge_fasta(qics_fa, known_fa)
```

Arguments

qics_fa	DNAStringSet 1
known_fa	DNAStringSet 2

Value

merged DNAStringSet file to be used for quantification

Author(s)

Stefan Stefanov

merge_qics *Merging 2 assemblies*

Description

Pair-wise merging 2 assemblies Pair-wise merging 2 assemblies

Usage

```
merge_qics(qics1, qics2, sgfc_pred)
```

Arguments

qics1	assembly 1
qics2	assembly 2
sgfc_pred	SGFC from the SGSeq feature detection

Value

data.frame of transcript information in flat format

Author(s)

Stefan Stefanov

overlap_SG_BSJ *Overlap of BSJ and a splice graph*

Description

Creates a disjointed set of exons based on a SGSeq obejct and a BSJ GRanges object

Usage

```
overlap_SG_BSJ(sgfc_pred, BSJ_gr, sg_annot)
```

Arguments

sgfc_pred	SGSeq prediction object
BSJ_gr	a GRange of BSJ cooredinates
sg_annot	SGSeq prediction object with annotation info

Details

Creates a disjointed set of exons based on a SGSeq obejct and a BSJ GRanges object. The function keeps the SGSeq metadata

Value

SGSeq with disjoint exon bins

Author(s)

Stefan Stefanov

parse_files *Parse BSJ input*

Description

Parse BSJ files from CIRI, CIRCexplorer2 or a TSV file

Usage

```
parse_files(file_list, file_path, input_type)
```

Arguments

`file_list`	list with file names
`file_path`	string object with file path, clould be an empty string
`input_type`	`CIRI` for CIRI2 input, `CE` for CIRCexplorer2 input and `tsv` for TSV formatted input

Details

This processes BSJ prediction files and prepares them for the next step of the pipeline. `input_type` is essential for the correct parsing of the files.

Value

`Tibble` object with combined BSJ coordinate and number of junction spanning reads across sample

Author(s)

Stefan Stefanov

plotRanges2 *Plot ranges*

Description

Plots GRanges objects

Usage

```
plotRanges2(...)
```

Details

ggplot of multiple GRanges object. Every object is auto assigned a colour from colorblind friendly scheme

Value

ggplot of multiple GRanges objects

Author(s)

Stefan Stefanov

`prep_circular_sg` *Preparing circular splice graph features*

Description

Preparing circular splice graph features Preparing circular splice graph features

Usage

```
prep_circular_sg(full_sg, full_fc, sgfc_pred, bs_genome, BSJ_gr, th = 15)
```

Arguments

full_sg	SGranges object supplying feature info
full_fc	count matrix corresponding to the features
sgfc_pred	SGFC from the SGSeq feature detection
bs_genome	BSGenome object used for extracting the sequences
BSJ_gr	GRange object of BSJ coordinates
th	minimum number of reads limiting exon selection

Value

SummarizedExperiment object with info on the circular splice graph features

Author(s)

Stefan Stefanov

`prep_output_gtf` *Creation of GTF based on CYCLeR results*

Description

Creation of GTF based on CYCLeR results

Usage

```
prep_output_gtf(qics_out, sgfc, annot_list = NULL)
```

Arguments

`sgfc`	`SummarizedExperiment` object with info on the circular splice graph features
`annot_list`	ORG package; soon to be expanded
`qics`	CYCLeR `data.frame` of intermediate results

Details

This function takes the

Value

GTF-like table

Author(s)

Stefan Stefanov

`process_BSJs` *Process BSJs*

Description

process the BSJ table and select high confidence BSJs

Usage

```
process_BSJs(cdf, sample_table)
```

Arguments

`cdf`	tibble produced by `parse_files`
`sample_table`	sample table formatted according to the manual, Must contain “sample_name” “treatment” “file_bam” “lib_size” “read_len”; NB the values in column “treatment” can only be “control” and “enriched”
`file_path`	string object with file path, clould be an empty string

Details

Filters BSJ based on comparison of the average CPM values of BSJs

Value

`Tibble` object with combined filtered BSJ coordinate and number of junction spanning reads across sample.

Author(s)

Stefan Stefanov

`recount_features` *Re-count of the reads per exon bin*

Description

A wrapper function for Rsubread

Usage

```
recount_features(full_sg, sample_table, paired_end = T)
```

Arguments

`full_sg`	a `SGSeq` object of exon bins
`sample_table`	sample table formatted according to the manual, Must contain “sample_name” “treatment” “file_bam” “lib_size” “read_len”; NB the values in column “treatment” can only be “control” and “enriched”
`paired_end`	a binary for pair-end info

Details

This function performs requantification of the exon bins with specifically selected parameters

Value

`BAMFileList` object with info on the trimmed files

Author(s)

Stefan Stefanov

`RPKM_calc` *RPKM calculation for the genomic features*

Description

RPKM calculation for the genomic features

Usage

```
RPKM_calc(
  count_matrix,
  sg,
  bsj_granges,
  bs_genome,
  sample_table,
  feature_type,
  fsj_overhang = 3,
  bsj_overhang = 15,
  eff_length_correction = T,
  gc_correction = F
)
```

Arguments

`count_matrix`	count matrix corresponding to the features
`sg`	`SGSeq` object supplying feature info
`bsj_granges`	GRange of BSJ cooredinates
`bs_genome`	a `BSGenome` object used for extracting the sequences
`sample_table`	sample table formatted according to the manual, Must contain "sample_name" "treatment" "file_bam" "lib_size" "read_len"; NB the values in column "treatment" can only be "control" and "enriched"
`feature_type`	either "e" for exons ot "j" for junctions
`fsj_overhang`	the FJS overhand used in the mapping a.k.a. anchor
`bsj_overhang`	the BSJ overhand used in the chimeric detection
`eff_length_correction`	whether or not to apply effective length correction
`gc_correction`	whether or not to apply GC-content correction; requires further testing

Details

This function performs RPKM calculations for the exonic features. The RPKM calculation is performed based on the exact sequences for the exons. For junctions, the sequences are selected based on the exons, flanking the junction. The function takes into account the needed effective length corrections.

Value

`BAMFileList` object with info on the trimmed files

Author(s)

Stefan Stefanov

`transcripts_per_sample` *Transcript assembly*

Description

Transcript assembly per sample Transcript assembly per sample based on sample name in the "sample_table"

Usage

```
transcripts_per_sample(sgfc, BSJ_gr, sample_name)
```

Arguments

`sgfc`	`SummarizedExperiment` object with info on the circular splice graph features
`BSJ_gr`	`GRange` object of BSJ coordinates
`sample_name`	name of the sample as indicated in the sample table

Value

`data.frame` of transcript information in flat format

Author(s)

Stefan Stefanov

Index