Package 'DNAfusion'

April 16, 2024

Title Identification of gene fusions using paired-end sequencing

Version 1.5.0

biocViews TargetedResequencing, Genetics, GeneFusionDetection, Sequencing

Description DNAfusion can identify gene fusions such as EML4-ALK based on paired-end sequencing results. This package was developed using position deduplicated BAM files generated with the AVENIO Oncology Analysis Software. These files are made using the AVENIO ctDNA surveillance kit and Illumina Nextseq 500 sequencing. This is a targeted hybridization NGS approach and includes ALK-specific but not EML4-specific probes.

License GPL-3

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Suggests knitr, rmarkdown, testthat, sessioninfo, BiocStyle

VignetteBuilder knitr

Imports bamsignals, GenomicRanges, IRanges, Rsamtools, GenomicAlignments, BiocBaseUtils, S4Vectors, GenomicFeatures, TxDb.Hsapiens.UCSC.hg38.knownGene, BiocGenerics

Depends R (>= 4.2.0)

BugReports https://github.com/CTrierMaansson/DNAfusion/issues

URL https://github.com/CTrierMaansson/DNAfusion

git_url https://git.bioconductor.org/packages/DNAfusion

git_branch devel

git_last_commit 633918e

git_last_commit_date 2023-10-25

Repository Bioconductor 3.19

Date/Publication 2024-04-16

Author Christoffer Trier Maansson [aut, cre]
(<https://orcid.org/0000-0002-3071-3437>),
Emma Roger Andersen [ctb, rev],
Maiken Parm Ulhoi [dtc],
Peter Meldgaard [dtc],
Boe Sandahl Sorensen [rev, fnd]

Maintainer Christoffer Trier Maansson <ctm@clin.au.dk>

Contents

ALK_sequence 2
break_position 3
break_position_depth 4
EML4_ALK_analysis 5
EML4_ALK_detection 6
EML4_sequence 7
find_variants 8
introns_ALK_EML4 9

Index **10**

`ALK_sequence` *Identification of ALK breakpoint bases*

Description

This function identifies the basepairs following the ALK breakpoint.

Usage

```
ALK_sequence(reads, basepairs = 20, genome = "hg38")
```

Arguments

`reads`	`GAlignments` returned by `EML4_ALK_detection()`.
`basepairs`	`integer`, number of basepairs identified from the EML4-ALK fusion. Default=20.
`genome`	`Character string` representing the reference genome. Can be either "hg38" or "hg19". Default="hg38".

Value

If EML4-ALK is detected, returns a `table` of identified ALK basepairs with the number of corresponding reads for each sequence. If no spanning reads in ALK is detected an empty `GAlignments` object is returned.If no EML4-ALK is detected "No EML4-ALK was detected" is returned.

Examples

```
H3122_bam <- system.file("extdata",
"H3122_EML4.bam",
package="DNAfusion")
HCC827_bam <-  system.file("extdata",
"HCC827_EML4.bam",
package="DNAfusion")

ALK_sequence(EML4_ALK_detection(file=H3122_bam,
                                genome="hg38",
                                mates=2),
             basepairs=20,
             genome="hg38")
ALK_sequence(EML4_ALK_detection(file=HCC827_bam,
                                genome="hg38",
                                mates=2),
             basepairs=20,
             genome="hg38")
```

---

break_position *EML4-ALK breakpoint*

Description

This function identifies the genomic position in EML4 or ALK, where the breakpoint has happened.

Usage

```
break_position(reads, gene, genome = "hg38")
```

Arguments

reads	GAlignments object returned by EML4_ALK_detection().
gene	Character string representing the gene. Can be either "ALK" or "EML4".
genome	Character string representing the reference genome. Can be either "hg38" or "hg19". Default="hg38".

Value

If EML4-ALK is detected, it returns a table of genomic positions with the number of corresponding reads for each sequence. If no spanning reads in EML4 or ALK is detected an empty GAlignments object is returned. If no EML4-ALK is detected "No EML4-ALK was detected" is returned.

Examples

```
H3122_bam <- system.file("extdata",
"H3122_EML4.bam",
package="DNAfusion")
HCC827_bam <-  system.file("extdata",
"HCC827_EML4.bam",
package="DNAfusion")

break_position(EML4_ALK_detection(file=H3122_bam,
                                  genome="hg38",
                                  mates=2),gene="EML4",genome="hg38")
break_position(EML4_ALK_detection(file=H3122_bam,
                                  genome="hg38",
                                  mates=2),gene="ALK",genome="hg38")
break_position(EML4_ALK_detection(file=HCC827_bam,
                                  genome="hg38",
                                  mates=2),gene="EML4",genome="hg38")
break_position(EML4_ALK_detection(file=HCC827_bam,
                                  genome="hg38",
                                  mates=2),gene="ALK",genome="hg38")
```

---

break_position_depth *Read depth at breakpoint*

Description

This function identifies the read depth at the basepair before the breakpoint in EML4 or ALK

Usage

```
break_position_depth(file, reads, gene, genome = "hg38")
```

Arguments

file	The name of the file which the data are to be read from.
reads	GAlignments object returned by EML4_ALK_detection().
gene	Character string representing the gene. Can be either "ALK" or "EML4".
genome	Character string representing the reference genome. Can be either "hg38" or "hg19". Default="hg38".

Value

If EML4-ALK is detected a single integer corresponding to the read depth at the breakpoint is returned. If no spanning reads in EML4 or ALK is detected an empty GAlignments object is returned. If no EML4-ALK is detected "No EML4-ALK was detected" is returned.

Examples

```
H3122_bam <- system.file("extdata",
"H3122_EML4.bam",
package="DNAfusion")
HCC827_bam <-  system.file("extdata",
"HCC827_EML4.bam",
package="DNAfusion")

break_position_depth(file=H3122_bam,
                        EML4_ALK_detection(file=H3122_bam,
                                           genome="hg38",
                                           mates=2),
                                           gene="ALK",genome="hg38")
break_position_depth(file=H3122_bam,
                        EML4_ALK_detection(file=H3122_bam,
                        genome="hg38",
                        mates=2),
                        gene="EML4",genome="hg38")
break_position_depth(file=HCC827_bam,
                        EML4_ALK_detection(file=HCC827_bam,
                                           genome="hg38",
                                           mates=2),
                                           gene="ALK",genome="hg38")
break_position_depth(file=H3122_bam,
                        EML4_ALK_detection(file=H3122_bam,
                                           genome="hg38",
                                           mates=2),
                                           gene="EML4",genome="hg38")
```

EML4_ALK_analysis *Complete EML4-ALK analysis*

Description

This functions collects the results from the other functions of the package.

Usage

```
EML4_ALK_analysis(file, genome = "hg38", mates = 2, basepairs = 20)
```

Arguments

file	The name of the file which the data are to be read from.
genome	character representing the reference genome. Can be either "hg38" or "hg19". Default="hg38".
mates	interger, the minimum number EML4-ALK mate pairs needed to be detected in order to call a variant. Default=2.
basepairs	integer, number of basepairs identified from the EML4-ALK fusion. Default=20.

Value

A list object with clipped_reads corresponding to EML4_ALK_detection(), last_EML4 corresponding to EML4_sequence(), first_ALK corresponding to ALK_sequence(), breakpoint_ALK corresponding to break_position(), gene = "ALK", breakpoint_EML4 corresponding to break_position(),gene = "EML4", read_depth_ALK corresponding to break_position_depth().gene = "ALK", and read_depth_EML4 corresponding to break_position_depth() ,gene = "EML4". If no EML4-ALK is detected an empty GAlignments is returned.

Examples

```
H3122_bam <- system.file("extdata",
"H3122_EML4.bam",
package="DNAfusion")
HCC827_bam <-  system.file("extdata",
"HCC827_EML4.bam",
package="DNAfusion")

EML4_ALK_analysis(file=H3122_bam,
                    genome="hg38",
                    mates=2,
                    basepairs=20)
EML4_ALK_analysis(file=HCC827_bam,
                    genome="hg38",
                    mates=2,
                    basepairs=20)
```

EML4_ALK_detection *Detection of ALK and EML4 breakpoint*

Description

This function identifies the genomic position in ALK and EML4 where the breakpoint has happened. This function looks for ALK-EML4 and EML4-ALK mate pair reads in the BAM file.

Usage

```
EML4_ALK_detection(file, genome = "hg38", mates = 2)
```

Arguments

file	The name of the file which the data are to be read from.
genome	Character string representing the reference genome. Can be either "hg38" or "hg19". Default="hg38".
mates	Interger, the minimum number ALK-EML4 mate pairs needed to be detected in order to call a variant. Default=2.

Value

A GAlignments object with soft-clipped reads representing ALK-EML4 and EML4-ALK is returned. If no ALK-EML4 or EML4-ALK is detected the GAlignmentsis empty.

Examples

```
H3122_bam <- system.file("extdata",
"H3122_EML4.bam",
package="DNAfusion")
HCC827_bam <-  system.file("extdata",
"HCC827_EML4.bam",
package="DNAfusion")

EML4_ALK_detection(file=H3122_bam,
                    genome="hg38",
                    mates=2)
EML4_ALK_detection(file=HCC827_bam,
                    genome="hg38",
                    mates=2)
```

EML4_sequence *Identification of EML4 breakpoint bases*

Description

This function identifies the basepairs leading up to the EML4 breakpoint.

Usage

```
EML4_sequence(reads, basepairs = 20, genome = "hg38")
```

Arguments

reads	GAlignments object returned by EML4_ALK_detection().
basepairs	Integer, number of basepairs identified from the EML4-ALK fusion. Default=20.
genome	Character string representing the reference genome. Can be either "hg38" or "hg19". Default="hg38".

Value

If EML4-ALK is detected, returns a table of identified EML4 basepairs with the number of corresponding reads for each sequence. If no EML4-ALK is detected "No EML4-ALK was detected" is returned.

Examples

```
H3122_bam <- system.file("extdata",
"H3122_EML4.bam",
package="DNAfusion")
HCC827_bam <-  system.file("extdata",
"HCC827_EML4.bam",
package="DNAfusion")

EML4_sequence(EML4_ALK_detection(file=H3122_bam,
                                 genome="hg38",
                                 mates=2),
              basepairs=20,
              genome="hg38")
EML4_sequence(EML4_ALK_detection(file=HCC827_bam,
                                 genome="hg38",
                                 mates=2),
              basepairs=20,
              genome="hg38")
```

`find_variants` *Detect the variants of ALK-EML4*

Description

This function identifies ALK-EML4 variants using the intron of the breakpoint of EML4

Usage

```
find_variants(file, genome = "hg38")
```

Arguments

`file`	The name of the file which the data are to be read from.
`genome`	`character` representing the reference genome. Can be either "hg38" or "hg19". Default="hg38".

Value

A `dataframe`of the ALK-EML4 variant is returned. If no variant is detected, "No ALK-EML4 was detected" is returned. If the variant is not classified a `list` with identified introns with breakpoints is returned. If the breakpoint could not be identified in either of the genes a `list` with identified introns with breakpoints is returned.

Examples

```
H3122_bam <- system.file("extdata",
"H3122_EML4.bam",
package="DNAfusion")
HCC827_bam <-  system.file("extdata",
"HCC827_EML4.bam",
package="DNAfusion")
find_variants(file=H3122_bam,genome="hg38")
find_variants(file=HCC827_bam,genome="hg38")
```

---

introns_ALK_EML4 *Detect ALK and EML4 introns of the breakpoint*

Description

This function identifies the introns of ALK and EML4 where the breakpoint has happened.

Usage

```
introns_ALK_EML4(file, genome = "hg38")
```

Arguments

file	The name of the file which the data are to be read from.
genome	character representing the reference genome. Can be either "hg38" or "hg19". Default="hg38".

Value

A dataframe of the ALK- and EML4-intron of the breakpoint is returned corresponding to the transcript ENST00000389048.8 for ALK and ENST00000318522.10 for EML4. If the breakpoint is not located in introns of ALK or EML4, "Breakpoint not located in intron of ALK" or "Breakpoint not located in intron of EML4" is returned. If no EML4-ALK is detected “No EML4-ALK was detected” is returned.

Examples

```
H3122_bam <- system.file("extdata",
"H3122_EML4.bam",
package="DNAfusion")
HCC827_bam <-  system.file("extdata",
"HCC827_EML4.bam",
package="DNAfusion")
introns_ALK_EML4(file=H3122_bam,genome="hg38")
introns_ALK_EML4(file=HCC827_bam,genome="hg38")
```

Index

ALK_sequence, 2

break_position, 3
break_position_depth, 4

EML4_ALK_analysis, 5
EML4_ALK_detection, 6
EML4_sequence, 7

find_variants, 8

introns_ALK_EML4, 9