# Package 'Genefusiondiscover'

August 9, 2022

**Title** Identification of gene fusions using paired-end sequencing

**Version** 0.99.1

**biocViews** TargetedResequencing, Genetics, GeneFusionDetection, Sequencing

**Description** Paired-end sequencing of cfDNA generated BAM files can be used as input to discover EML4-ALK variants. 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** 7.2.1

**Suggests** knitr,
rmarkdown,
usethis,
devtools,
testthat,
BiocStyle

**VignetteBuilder** knitr

**Imports** bamsignals,
GenomicRanges,
IRanges,
Rsamtools

**Depends** dplyr,
R (>= 4.2.0)

**BugReports** <https://github.com/CTrierMaansson/Genefusiondiscover/issues>

## R topics documented:

---

ALK_sequence                    *Identification of ALK breakpoint bases*

---

**Description**

This function identifies the basepairs following the ALK breakpoint.

**Usage**

```
ALK_sequence(reads, basepairs = 20)
```

**Arguments**

| | |
|---|---|
| reads | data.frame returned by EML4_ALK_detection(). |
| basepairs | integer, number of basepairs identified from the EML4-ALK fusion. Default=20. |

**Value**

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

**Examples**

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

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

---

breakPosition                    *EML4-ALK breakpoint*

---

**Description**

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

**Usage**

```
breakPosition(reads)
```

## Arguments

reads            Data.frame returned by EML4_ALK_detection().

## Value

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

## Examples

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

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

---

breakPositionDepth        *Read depth at breakpoint*

---

## Description

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

## Usage

```
breakPositionDepth(file, reads)
```

## Arguments

file               The name of the file which the data are to be read from.

reads           data.frame returned by EML4_ALK_detection().

## Value

If EML4-ALK is detected a single integer corresponding to the read depth at the breakpoint is returned. Otherwise "No EML4-ALK was detected" is returned

## Examples

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

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

---

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 corresponding to breakPosition(), and read_depth corresponding to breakPositionDepth().

## Examples

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

```
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 EML4-ALK variants*

---

## Description

This function looks for 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 EML4-ALK mate pairs needed to be detected in order to call a variant. Default=2. |

## Value

If EML4-ALK is detected a `data.frame` with soft-clipped reads representing EML4-ALK is returned. Otherwise "No EML4-ALK was detected" is returned.

## Examples

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

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)
```

**Arguments**

| | |
|---|---|
| reads | Data.frame returned by EML4_ALK_detection(). |
| basepairs | Integer, number of basepairs identified from the EML4-ALK fusion. Default=20. |

**Value**

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

**Examples**

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

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

# Index