Package 'katdetectr'

October 17, 2024

Title Detection, Characterization and Visualization of Kataegis in Sequencing Data

Version 1.6.0

Date 2023-02-14

Description Kataegis refers to the occurrence of regional hypermutation and is a phenomenon observed in a wide range of malignancies. Using changepoint detection katdetectr aims to identify putative kataegis foci from common data-formats housing genomic variants. Katdetectr has shown to be a robust package for the detection, characterization and visualization of kataegis.

License GPL-3 + file LICENSE

URL https://doi.org/doi:10.18129/B9.bioc.katdetectr

BugReports https://github.com/ErasmusMC-CCBC/katdetectr/issues

Depends R (>= 4.2)

Imports Biobase (>= 2.54.0), BiocParallel (>= 1.26.2), BSgenome (>= 1.62.0), BSgenome.Hsapiens.UCSC.hg19 (>= 1.4.3), BSgenome.Hsapiens.UCSC.hg38 (>= 1.4.4), changepoint (>= 2.2.3), changepoint.np (>= 1.0.3), checkmate (>= 2.0.0), dplyr (>= 1.0.8), GenomeInfoDb (>= 1.28.4), GenomicRanges (>= 1.44.0), ggplot2 (>= 3.3.5), ggtext (>= 0.1.1), IRanges (>= 2.26.0), maftools (>= 2.10.5), methods (>= 4.1.3), plyranges (>= 1.17.0), Rdpack (>= 2.3.1), rlang (>= 1.0.2), S4Vectors (>= 0.30.2), scales (>= 1.2.0), tibble (>= 3.1.6), tidyr (>= 1.2.0), tools, utils, VariantAnnotation (>= 1.38.0)

Suggests BiocStyle (>= 2.26.0), knitr (>= 1.37), rmarkdown (>= 2.13), stats, testthat (>= 3.0.0)

VignetteBuilder knitr

RdMacros Rdpack

biocViews WholeGenome, Software, SNP, Sequencing, Classification, VariantAnnotation

Config/testthat/edition 3

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

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

git_branch RELEASE_3_19

git_last_commit abc118b

git_last_commit_date 2024-04-30

Repository Bioconductor 3.19

Date/Publication 2024-10-16

Author Daan Hazelaar [aut, cre] (<https://orcid.org/0000-0002-7513-6813>),
Job van Riet [aut] (<https://orcid.org/0000-0001-7767-7923>),
Harmen van de Werken [ths] (<https://orcid.org/0000-0002-9794-1477>)

Maintainer Daan Hazelaar <daanhazelaar@gmail.com>

Contents

`constructKatdetect` *constructKatdetect*

Description

Constructor function for a KatDetect object.

Usage

```
constructKatdetect(
  genomicVariants = VariantAnnotation::VRanges(),
  segments = GenomicRanges::GRanges(),
  kataegisFoci = GenomicRanges::GRanges(),
  info = list()
)
```

Arguments

genomicVariants	(VRanges)
segments	(GRanges)
kataegisFoci	(GRanges)
info	(list)

Value

(KatDetect): Returns a KatDetect object.

Author(s)

Daan Hazelaar

Examples

```
constructKatdetect()
```

`detectKataegis` *detectKataegis*

Description

Detection of kataegis foci using changepoint detection.

Changepoint detection is performed on the intermutation distance (IMD) of the variants using the changepoint (killick 2014) package.

Note that we recommend using the default parameters for the detection of kataegis.

Usage

```
detectKataegis(
  genomicVariants,
  refSeq = "hg19",
  minSizeKataegis = 6,
  IMDcutoff = 1000,
  test.stat = "Exponential",
  penalty = "BIC",
  pen.value = 0,
  method = "PELT",
  minseglen = 2,
  BPPARAM = BiocParallel::SerialParam(),
  aggregateRecords = FALSE
)
```

Arguments

`genomicVariants`
: (VRanges, VCF or MAF): VRanges, or path to VCF or MAF file containing genomic variants.

`refSeq`
: (character or data.frame): The used reference genome for variant calling. Choose: "hg19" or "hg38". For analysis of non standard sequences: provide a data.frame containing the length of the sequences.

`minSizeKataegis`
: (integer): Minimal number of variants required within a segment for classification as a kataegis foci.

`IMDcutoff`
: (numberic or function): When a numeric is supplied this represents the max mean IMD within a segment for classification as a kataegis foci. When a custom function is supplied by the user a IMD cutoff value is determined for each segment.

`test.stat`
: (character): Distribution that is fitted to the data (Exponential or Empirical). See cpt.meanvar.

`penalty`
: (character): Penalty used to guard against overfitting (BIC or Manual). See cpt.meanvar.

`pen.value`
: (integer): Only needed for manual penalty. See cpt.meanvar.

`method`
: (character): The search method used in changepoint analysis. Choice of: "PELT", "AMOC", "SegNeigh" or "BinSeg".

`minseglen`
: (integer): Min. size of segments (no. of variants).

`BPPARAM`
: (BiocParallelParam): Can be used for parallelization purposes.

`aggregateRecords`
: (logical): Aggregate multiple samples and treat as-if all records originate from a single sample.

Value

(KatDetect): Returns a KatDetect object including putative kataegis foci.

Author(s)

Daan Hazelaar

Job van Riet

References

Killick R, Eckley I (2014). "changepoint: An R package for changepoint analysis." Journal of statistical software, 58(3), 1–19.

Examples

```
syntheticData <- generateSyntheticData()
kd <- detectKataegis(syntheticData)
```

`generateSyntheticData` *Generate genomic variants with pre-defined kataegis foci.*

Description

This function generates a synthetic dataset (VRanges) containing background genomic variants and a no. of desired interjected kataegis foci.

Usage

```
generateSyntheticData(
  genome = BSgenome.Hsapiens.UCSC.hg19::BSgenome.Hsapiens.UCSC.hg19,
  nBackgroundVariants = 100,
  seqnames = NULL,
  probMutationType = c(0.8, 0.1, 0.1),
  nKataegisFoci = 5,
  nKataegisVariants = 20,
  expectedIMD = 100,
  sampleName = "syntheticData",
  removeValidationColumns = TRUE
)
```

Arguments

`genome`	(BSgenome): The genome (DNA) which will be sampled for genomic variants.
`nBackgroundVariants`	(integer): The no. of generated background genomic variants.
`seqnames`	(character): The sequences on which genomic variants will be sampled. If NULL, then all human autosomes and sex chromosomes will be used.
`probMutationType`	(numeric): The probability of a generated variant being an SNV, MNV, Deletion or Insertion, respectively.
`nKataegisFoci`	(integer): The no. of generated and interjected kataegis foci.
`nKataegisVariants`	(integer): The no. of genomic variants within each kataegis foci.
`expectedIMD`	(integer): The expected mean intermutational distance (IMD) of the generated kataegis foci.
`sampleName`	(character): The name of the sample
`removeValidationColumns`	(logical): Include columns with extra information regarding mutation sampling?

Value

(VRanges): VRanges containing background genomic variants and pre-defined kataegis foci.

Author(s)

Daan Hazelaar

Job van Riet

Examples

```
syntheticData1 <- generateSyntheticData()

syntheticData2 <- generateSyntheticData(
    genome = BSgenome.Hsapiens.UCSC.hg38::BSgenome.Hsapiens.UCSC.hg38,
    nBackgroundVariants = 75,
    seqnames = c("chr1", "chrX"),
    nKataegisFoci = 1,
    nKataegisVariants = 25,
    sampleName = "testSample",
    removeValidationColumns = FALSE
)
```

---

getGenomicVariants *Retrieve genomic variants from KatDetect object.*

Description

Retrieve genomic variants from KatDetect object.

Usage

```
getGenomicVariants(x)

## S4 method for signature 'KatDetect'
getGenomicVariants(x)
```

Arguments

x	(KatDetect): KatDetect object.

Value

(VRanges): Returns a VRanges with annotated genomic variants.

Examples

```
kd <- constructKatdetect()
getGenomicVariants(kd)
```

getInfo *Retrieve model parameters from a KatDetect object.*

Description

Retrieve model parameters from a KatDetect object.

Usage

```
getInfo(x)

## S4 method for signature 'KatDetect'
getInfo(x)
```

Arguments

x	(KatDetect): KatDetect object.

Value

(list): Returns a list with all model parameters used for kataegis detection.

Examples

```
kd <- constructKatdetect()
getInfo(kd)
```

getKataegisFoci *Retrieve kateagis foci from a KatDetect object.*

Description

Retrieve kateagis foci from a KatDetect object.

Usage

```
getKataegisFoci(x)

## S4 method for signature 'KatDetect'
getKataegisFoci(x)
```

Arguments

x	(KatDetect): KatDetect object.

Value

(GRanges): Returns a GRanges with annotated kataegis foci.

Examples

```
kd <- constructKatdetect()
getKataegisFoci(kd)
```

`getSegments` *Retrieve segments from KatDetect a object.*

Description

Retrieve segments from KatDetect a object.

Usage

```
getSegments(x)

## S4 method for signature 'KatDetect'
getSegments(x)
```

Arguments

x	(KatDetect): KatDetect object.

Value

(GRanges): Returns a GRanges with annotated segments.

Examples

```
kd <- constructKatdetect()
getSegments(kd)
```

`KatDetect-class` *KatDetect-class: KatDetect objects*

Description

The katdetectr package introduces a new S4 object which stores all relevant information regarding kataegis detection.

Value

(KatDetect) KatDetect object.

Slots

`kataegisFoci` (GRanges): Contains all annotated putative kataegis foci.

`genomicVariants` (VRanges): Contains all processed and annotated genomic variants.

`segments` (GRanges): Contains all segments detected after changepoint analysis.

`info` (list): Contains some general information and model parameters used for kataegis detection.

Author(s)

Daan Hazelaar

Job van Riet

Examples

```
syntheticData <- generateSyntheticData()
kd <- detectKataegis(syntheticData)

getKataegisFoci(kd)
getGenomicVariants(kd)
getSegments(kd)
getInfo(kd)
```

`katdetectr` *Katdetectr: A package for kataegis detection, characterization and visualization*

Description

The katdetectr package provides three main functions: detectKataegis(), rainfallPlot() and generateSyntheticData()

Details

See the vignette for more details, a step by step explanation of these function, and the general katdetectr workflow.

`rainfallPlot` *Rainfall plot.*

Description

Visualize the IMD, segments and putative kataegis foci using a rainfall plot.

Y-axis represents the 5' intermutation distance (IMD) of each genomic variant in a log10-scale. X-axis represent variant ID.

Variants within kataegis foci are bold with the kataegis foci shown in a blue rectangle (if showKataegis = TRUE). Color represent the mutation type whilst horizontal lines represent the mean IMD of a segments and Vertical lines depict the detected changepoints (if showSegmentation = TRUE).

Usage

```
rainfallPlot(
  kd,
  showSequence = "All",
  showKataegis = TRUE,
  showSegmentation = FALSE
)
```

Arguments

`kd`	(KatDetect): KatDetect object.
`showSequence`	(character): Which sequence(s) should be visualized? Choice of: 'All', 'Kataegis', c('seqname1', 'seqname2').
`showKataegis`	(logical): Highlight putative kataegis foci and underlying genomic variants?
`showSegmentation`	(logical): Show changepoints and mean IMD of each segment?

Value

(ggplot): Returns rainfall plot.

Author(s)

Daan Hazelaar

Job van Riet

Examples

```
syntheticData <- generateSyntheticData(nBackgroundVariants = 200)
kd <- detectKataegis(syntheticData)

# Visualize the IMD of the genomic variants by constructing a rainfall plot
katdetectr::rainfallPlot(kd)
```

```
# Show the chromosomes which contain one or more kataegis foci
katdetectr::rainfallPlot(kd, showSequence = "Kataegis")

# Show only chromosome 1 and 2
katdetectr::rainfallPlot(kd = kd, showSequence = c("chr1", "chr2"))

# Display changepoints and mean IMD per segment
katdetectr::rainfallPlot(kd = kd, showSequence = c("chr1", "chr2"), showSegmentation = TRUE)
```

Index