Package 'PlasmaMutationDetector2'

October 12, 2022

Type Package

Title Tumor Mutation Detection in Plasma using Barcoding

Version 1.1.11

Date 2022-04-07

Author Yves Rozenholc [cre, aut]
Oddmund Nordgård [con, aut]
Nicolas Pécuchet [con]
Pierre-Laurent Puig [con]

Maintainer Rozenholc <yves.rozenholc@u-paris.fr>

Description Aims at detecting single nucleotide variation (SNV) and insertion/deletion (INDEL) in circulating tumor DNA (ctDNA), used as a surrogate marker for tumor, at each base position of an Next Generation Sequencing (NGS) analysis using barcoding. Mutations are assessed by comparing the minor-allele frequency at each position to the measured PER in control samples.
This package has been used for Kjersti Tjensvoll, Morten Lapin, Bjørnar Gilje, Herish Garresori, Satu Oltedal, Rakel Brendsdal Forthun, Anders Molven, Yves Rozenholc and Oddmund N\o{o}rdgaard (2022) <https://www.nature.com/articles/s41598-022-09698-5>.

License MIT + file LICENSE

Imports S4Vectors (>= 0.16.0), Rsamtools (>= 1.30.0), rtracklayer (>= 1.38.0), robustbase (>= 0.92-8), SummarizedExperiment (>= 1.8.0)

Depends R (>= 3.5.0), ggplot2 (>= 2.2.0), grid (>= 3.4.0), GenomicRanges (>= 1.30.0), VariantAnnotation (>= 1.24.0)

Encoding UTF-8

RoxygenNote 7.1.2

LazyData true

NeedsCompilation no

Repository CRAN

Date/Publication 2022-05-03 10:00:08 UTC

R topics documented:

`background_error_rate` *The package provide the SNV and INDEL PERs computed for the Ion AmpliSeq™ Colon and Lung Cancer Panel v2 from 29 controls in a table available in the data file* `background_error_rate.txt`.

Description

This table contains 9 variables for each genomic position

- `chrpos`, char, of the form chrN:XXXXXXXXX defining genomic position
- `N0`, integer, the coverture in the controls
- `E0`, integer, the number of errors in the controls
- `p.sain`, numeric, the ratio E0/N0
- `up.sain`, numeric, the 95th quantile of the Binomial with parameter N0 and E0/N0
- `E0indel`, integer, the amount of indel
- `indel.p.sain`, numeric, the ration E0indel/N0
- `indel.up.sain`, numeric, the 95th quantile of the Binomial with parameter N0 and E0indel/N0
- `hotspot`, char, either 'Non-hotspot' or 'Hotspot' depending if the genomic position is known as hotspot or not.

Usage

```
data(background_error_rate)
```

Author(s)

N. Pécuchet, P. Laurent-Puig, O. Nordgård and Y. Rozenholc

References

Analysis of base-position error rate of next-generation sequencing to detect tumor mutations in circulating DNA N. Pécuchet, Y. Rozenholc, E. Zonta, D. Pietraz, A. Didelot, P. Combe, L. Gibault, J-B. Bachet, V. Taly, E. Fabre, H. Blons, P. Laurent-Puig in *Clinical Chemistry*

Novel hybridization- and tag-based error-corrected method for sensitive ctDNA mutation detection using ion semiconductor sequencing Kjersti Tjensvoll, Morten Lapin, Bjørnar Gilje, Herish Garresori, Satu Oltedal, Rakel Brendsdal Forthun, Anders Molven, Yves Rozenholc and Oddmund Nordgård in *Scientific Reports*

See Also

`BuildCtrlErrorRate`

`BuildCtrlErrorRate` *function BuildCtrlErrorRate*

Description

Compute the SNV Position-Error Rates and INDEL Position-Error Rates from control samples (available in the control directory `ctrl.dir`). This function requires MAF files, that will be automatically generated if not present in the specified control folder. SNV PER is computed as the sum in control samples of SNV background counts / sum in control samples of depths where SNV background counts = depth - major allele count. INDEL PER is computed as sum in control samples of INDEL background counts / sum in control samples of depths where INDEL background counts = sum of insertion and deletion counts.

Usage

```
BuildCtrlErrorRate(
  ctrl.dir = "Plasma ctrl/",
  bai.ext = ".bai",
  pos_ranges.file = NULL,
  hotspot.file = NULL,
  cov.min = 5000,
  force = FALSE,
  output.dir = ctrl.dir,
  n.trim = 0
)
```

Arguments

`ctrl.dir,`	char, foldername containing the control files (default 'Plasma ctrl/'). The typical folder hierarchy will consist of 'Plasma ctrl/rBAM'
`bai.ext,`	char, filename extension of the bai files (default '.bai')

`pos_ranges.file`,
: char, name of the Rdata file containing the three variables `pos_ind`, `pos_snp` and `pos_ranges` as build by the function `PrepareLibrary`. Default NULL, use the position_ranges.rda provided, used for our analysis.

`hotspot.file`,
: char, name of the text file containing a list of the genomic positions of the hotspots (default NULL, read the provide hotspot.txt, see `hotspot`)

`cov.min`,
: integer, minimal coverture to take into account a position (default 5000)

`force`,
: boolean, (default FALSE) if TRUE force all computations to all files including already processed ones

`output.dir`,
: char, name of the folder to save results (default `ctrl.dir`).

`n.trim`,
: integer, number of base positions trimmed at the ends of each amplicon (default 8)

Value

the number of processed files

Author(s)

N. Pécuchet, P. Laurent-Puig, O. Nordgård and Y. Rozenholc

References

Analysis of base-position error rate of next-generation sequencing to detect tumor mutations in circulating DNA N. Pécuchet, Y. Rozenholc, E. Zonta, D. Pietraz, A. Didelot, P. Combe, L. Gibault, J-B. Bachet, V. Taly, E. Fabre, H. Blons and P. Laurent-Puig in *Clinical Chemistry*

Novel hybridization- and tag-based error-corrected method for sensitive ctDNA mutation detection using ion semiconductor sequencing Kjersti Tjensvoll, Morten Lapin, Bjørnar Gilje, Herish Garresori, Satu Oltedal, Rakel Brendsdal Forthun, Anders Molven, Yves Rozenholc and Oddmund Nordgård in *Scientific Reports*

Examples

```
## Not run:
  ctrl.dir = system.file("extdata", "4test_only/ctrl/", package = "PlasmaMutationDetector2")
   if (substr(ctrl.dir,nchar(ctrl.dir),nchar(ctrl.dir))!='/')
     ctrl.dir = paste0(ctrl.dir,'/') # TO RUN UNDER WINDOWS
   BuildCtrlErrorRate(ctrl.dir,output.dir=paste0(tempdir(),'/'))

## End(Not run)
```

`DetectPlasmaMutation` *function DetectPlasmaMutation*

Description

This is the main function of the package that calls mutations by comparing at each genomic position the SNV or INDEL frequencies computed in one tested sample to the SNV or INDEL Position-Error Rates computed from several control samples by a binomial test. An outlier detection is performed among all intra-sample p-values to call a mutation. For users wishing to develop their own analysis for other sequencing panel, it requires recalibrated BAM files control samples to be processed to compute the Position-Error Rates stored in a file specified in `ber.ctrl.file`.

Usage

```
DetectPlasmaMutation(
  patient.dir = "./",
  patient.name = NULL,
  pos_ranges.file = NULL,
  ber.ctrl.file = NULL,
  bai.ext = ".bai",
  alpha = 0.05,
  n.trim = 0,
  force = FALSE,
  show.more = FALSE,
  qcutoff.snv = 1,
  qcutoff.indel = 1,
  cutoff.sb.hotspot = Inf,
  cutoff.sb.nonhotspot = cutoff.sb.hotspot,
  cutoff.sb.indel = cutoff.sb.hotspot,
  cutoff.sb.ref = 0.9,
  hotspot.indel = "chr7:55227950:55249171",
  output.dir = patient.dir
)
```

Arguments

`patient.dir,`	char, foldername containing the rBAM folder of the patients. The typical folder hierarchy will consist of 'Plasma/rBAM'
`patient.name,`	char, filename of the patient .bam file(s) (default NULL read all patients in folder `patient.dir`)
`pos_ranges.file,`	char, name of the Rdata file containing the three variables `pos_ind`, `pos_snp`, `pos_ranges` as build by the function `PrepareLibrary`. Default NULL, use the position_ranges.rda provides that we used for our analysis.
`ber.ctrl.file,`	char, pathname of the file providing the background error rates obtained from the controls (default NULL use the provided background error rates obtained from

our 29 controls). See `background_error_rate.txt` data and `BuildCtrlErrorRate` function.

`bai.ext`, char, filename extension of the bai files (default '.bai')

`alpha`, num, global false positive rate = global test level (default 0.05)

`n.trim`, integer, number of base positions trimmed at the ends of each amplicon (default 0)

`force`, boolean, (default FALSE) if TRUE force all computations to all files including already processed ones

`show.more`, boolean, (default FALSE show only detected positions) if TRUE additional annotations on result plots are given for non-significant mutations

`qcutoff.snv`, numeric, proportion of kept base positions ranged by increasing percentile SNV PER in control samples (default 1)

`qcutoff.indel`, numeric, proportion of kept base positions ranged by increasing percentile INDEL PER in control samples (default 1)

`cutoff.sb.hotspot`, numeric, exclude hotspot positions without Symmetric Odds Ratio test < cutoff (default 1)

`cutoff.sb.nonhotspot`, numeric, exclude non-hotspot positions without Symmetric Odds Ratio test < cutoff (default cutoff.sb.hotspot)

`cutoff.sb.indel`, numeric, exclude indel positions without Symmetric Odds Ratio test < cutoff (default cutoff.sb.hotspot)

`cutoff.sb.ref`, numeric, exclude ref positions without Symmetric Odds Ratio test < cutoff (default cutoff = 0.9)

`hotspot.indel`, char, a vector containing the known positions of hotspot deletion/insertion defined as chrX:start:end (default 'chr7:55227950:55249171')

`output.dir`, char, name of the folder to save results (default `patient.dir`).

Value

the number of processed patients

Author(s)

N. Pécuchet, P. Laurent-Puig, O. Nordgård and Y. Rozenholc

References

Analysis of base-position error rate of next-generation sequencing to detect tumor mutations in circulating DNA N. Pécuchet, Y. Rozenholc, E. Zonta, D. Pietraz, A. Didelot, P. Combe, L. Gibault, J-B. Bachet, V. Taly, E. Fabre, H. Blons and P. Laurent-Puig in *Clinical Chemistry*

Novel hybridization- and tag-based error-corrected method for sensitive ctDNA mutation detection using ion semiconductor sequencing Kjersti Tjensvoll, Morten Lapin, Bjørnar Gilje, Herish Garresori, Satu Oltedal, Rakel Brendsdal Forthun, Anders Molven, Yves Rozenholc and Oddmund Nordgård in *Scientific Reports*

Examples

```
patient.dir=system.file("extdata","4test_only/case/",package="PlasmaMutationDetector2")
  if (substr(patient.dir,nchar(patient.dir),nchar(patient.dir))!='/')
    patient.dir = paste0(patient.dir,'/') # TO RUN UNDER WINDOWS
  DetectPlasmaMutation(patient.dir,output.dir=paste0(tempdir(),'/'))
```

`hotspot` *The package provide a list of known hotspot positions located on the amplicons of the Ion AmpliSeq™ Colon and Lung Cancer Panel v2 as a txt file* `hotspot.txt` *which contains a vector/variable —named chrpos (first row)— of chars, of the form chrN:XXXXXXXXX defining genomic positions.*

Description

The package provide a list of known hotspot positions located on the amplicons of the Ion AmpliSeq™ Colon and Lung Cancer Panel v2 as a txt file `hotspot.txt` which contains a vector/variable —named chrpos (first row)— of chars, of the form chrN:XXXXXXXXX defining genomic positions.

Usage

```
data(hotspot)
```

Author(s)

N. Pécuchet, P. Laurent-Puig, O. Nordgård and Y. Rozenholc

References

Analysis of base-position error rate of next-generation sequencing to detect tumor mutations in circulating DNA N. Pécuchet, Y. Rozenholc, E. Zonta, D. Pietraz, A. Didelot, P. Combe, L. Gibault, J-B. Bachet, V. Taly, E. Fabre, H. Blons, P. Laurent-Puig in *Clinical Chemistry*

Novel hybridization- and tag-based error-corrected method for sensitive ctDNA mutation detection using ion semiconductor sequencing Kjersti Tjensvoll, Morten Lapin, Bjørnar Gilje, Herish Garresori, Satu Oltedal, Rakel Brendsdal Forthun, Anders Molven, Yves Rozenholc and Oddmund Nordgård in *Scientific Reports*

`LoadBackgroundErrorRate` *function LoadBackgroundErrorRate*

Description

This function will load the background error rates created from the controls using the function BuildCtrlErrorRate

Usage

```
LoadBackgroundErrorRate(pos_ranges.file, ber.ctrl.file)
```

Arguments

`pos_ranges.file`,
: char, name of the Rdata file containing the three variables `pos_ind`, `pos_snp`, `pos_ranges` as build by the function `PrepareLibrary`. Default NULL, use the position_ranges.rda provides that we used for our analysis.

`ber.ctrl.file`,
: char, pathname of the file providing the background error rates obtained from the controls (default NULL use the provided background error rates obtained from our 29 controls). See `background_error_rate.txt` data and `BuildCtrlErrorRate` function.

Value

the adapted background error rate

Author(s)

N. Pécuchet, P. Laurent-Puig, O. Nordgård and Y. Rozenholc

References

Analysis of base-position error rate of next-generation sequencing to detect tumor mutations in circulating DNA N. Pécuchet, Y. Rozenholc, E. Zonta, D. Pietraz, A. Didelot, P. Combe, L. Gibault, J-B. Bachet, V. Taly, E. Fabre, H. Blons and P. Laurent-Puig in *Clinical Chemistry*

Novel hybridization- and tag-based error-corrected method for sensitive ctDNA mutation detection using ion semiconductor sequencing Kjersti Tjensvoll, Morten Lapin, Bjørnar Gilje, Herish Garresori, Satu Oltedal, Rakel Brendsdal Forthun, Anders Molven, Yves Rozenholc and Oddmund Nordgård in *Scientific Reports*

MAF_from_BAM *function MAF_from_BAM*

Description

Read BAM files and create MAF file. BAMfiles are stored in a sub-folder '/rBAM'. MAF files are intermediate files stored in a sub-folder '/BER'. MAF files contain the raw counts of A,T,C,G, insertion, deletion, insertion>2bp, deletion >2bp for strand plus and stand minus. Note : we strongly recommand to externally recalibrate BAM files using tools like GATK.

Usage

```
MAF_from_BAM(
  study.dir = "Plasma/",
  input.filenames = NULL,
  bai.ext = ".bai",
  pos_ranges.file = NULL,
  force = FALSE,
  output.dir = study.dir,
  n.trim = 8
)
```

Arguments

study.dir,	char, name of the folder containing the rBAM directory (default 'Plasma/'). The typical folder hierarchy will consist of 'Plasma/rBAM'
input.filenames,	a vector of char (default NULL), the names of the BAM files to process. If NULL all BAM files in the rBAM folder will be processed
bai.ext,	char, filename extension of the bai files (default '.bai')
pos_ranges.file,	char, name of the Rdata file containing the three variables pos_ind, pos_snp and pos_ranges as build by the function PrepareLibrary. Default NULL, use the position_ranges.rda provided, used for our analysis.
force,	boolean, (default FALSE) if TRUE force all computations to all files including already processed ones
output.dir,	char, name of the folder to save results (default study.dir)
n.trim,	integer, number of base positions trimmed at the ends of each amplicon (default 8)

Value

the path/names of the MAF files

Author(s)

N. Pécuchet, P. Laurent-Puig, O. Nordgård and Y. Rozenholc

References

Analysis of base-position error rate of next-generation sequencing to detect tumor mutations in circulating DNA N. Pécuchet, Y. Rozenholc, E. Zonta, D. Pietraz, A. Didelot, P. Combe, L. Gibault, J-B. Bachet, V. Taly, E. Fabre, H. Blons, P. Laurent-Puig in *Clinical Chemistry*

Novel hybridization- and tag-based error-corrected method for sensitive ctDNA mutation detection using ion semiconductor sequencing Kjersti Tjensvoll, Morten Lapin, Bjørnar Gilje, Herish Garresori, Satu Oltedal, Rakel Brendsdal Forthun, Anders Molven, Yves Rozenholc and Oddmund Nordgård in *Scientific Reports*

Examples

```
  ## Not run:
     ctrl.dir = system.file("extdata", "4test_only/ctrl/",
       package = "PlasmaMutationDetector2")
     if (substr(ctrl.dir,nchar(ctrl.dir),nchar(ctrl.dir))!='/')
       ctrl.dir = paste0(ctrl.dir,'/') # TO RUN UNDER WINDOWS
     MAF_from_BAM(ctrl.dir,force=TRUE,output.dir=paste0(tempdir(),'/'))

## End(Not run)
```

`positions_ranges` *The package provide the positions and ranges computed for the Ion AmpliSeq™ Colon and Lung Cancer Panel v2 as a Rdata file* `positions_ranges.rda`.

Description

This file contains 4 variables

- `pos_ind`, vector of chars, of the form chrN:XXXXXXXXX defining genomic positions of the Ion AmpliSeq™ Colon and Lung Cancer Panel v2
- `pos_snp`, vector of chars, of the form chrN:XXXXXXXXX defining the known snp genomic positions
- `pos_ranges`, GRanges object, describing the 92 amplicons of the Ion AmpliSeq™ Colon and Lung Cancer Panel v2

Usage

```
data(positions_ranges)
```

Author(s)

N. Pécuchet, P. Laurent-Puig, O. Nordgård and Y. Rozenholc

References

Analysis of base-position error rate of next-generation sequencing to detect tumor mutations in circulating DNA N. Pécuchet, Y. Rozenholc, E. Zonta, D. Pietraz, A. Didelot, P. Combe, L. Gibault, J-B. Bachet, V. Taly, E. Fabre, H. Blons, P. Laurent-Puig in *Clinical Chemistry*

Novel hybridization- and tag-based error-corrected method for sensitive ctDNA mutation detection using ion semiconductor sequencing Kjersti Tjensvoll, Morten Lapin, Bjørnar Gilje, Herish Garresori, Satu Oltedal, Rakel Brendsdal Forthun, Anders Molven, Yves Rozenholc and Oddmund Nordgård in *Scientific Reports*

See Also

`Prepare_Library`

`PrepareLibrary` *function PrepareLibrary*

Description

Define the Genomic Ranges and Genomic Positions covered by the AmpliSeq™ Panel to include in the study and define SNP positions to exclude from the study. Trimming amplicon ends is performed if specified. This function is mostly useful if you want to add some SNP positions which are not existing in the positions_ranges.rda file provided within the package. It is provided to be able to reconstruct `positions_ranges.rda` data.

Usage

```
PrepareLibrary(
  info.dir = "Info/",
  bed.filename = "PACT-ACT_iDES_1_Regions.bed",
  snp.filename = "ExAC.r1.sites.vep.vcf.gz",
  snp.extra = NULL,
  output.name = "positions_ranges.rda",
  output.dir = info.dir
)
```

Arguments

`info.dir,`	char, name of the folder containing the library information files (default 'Info/')
`bed.filename,`	char, name of a BED table (tab-delimited) describing the Panel (with first 3 columns: "chr" (ex:chr1), "start position" (ex:115252190), "end position" (ex:115252305), i.e. the Ion AmpliSeq™ Colon and Lung Cancer Research Panel v2 (default 'lungcolonV2.bed.txt' as provided in the inst/extdata/Info folder of the package).
`snp.filename,`	char, name of the vcf file describing known SNP positions, obtained from ftp://ftp.broadinstitute.org/pub/E (default 'ExAC.r0.3.sites.vep.vcf.gz'). It requires a corresponding TBI file to be in the same folder (obtained from ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.3/ExAC.r0.3.site

snp.extra,	a vector of char, a vector of extra known snp positions manually curated (ex:"chrN:XXXXXXXXX")
output.name,	char, filename to save pos_ind and pos_snp (default 'positions_ranges.rda')
output.dir,	char, directory where to save pos_ind and pos_snp (default info.dir)

Value

Save the following variables in a .rda file defined by output.name in the folder defined by output.dir:

- pos_ranges, a GRanges descriptor of amplicon positions
- pos_ind, a vector of char "chrN:XXXXXXXXX", defining ALL index positions
- pos_snp, a vector of char "chrN:XXXXXXXXX", defining SNP positions

Author(s)

N. Pécuchet, P. Laurent-Puig, O. Nordgård and Y. Rozenholc

References

Analysis of base-position error rate of next-generation sequencing to detect tumor mutations in circulating DNA N. Pécuchet, Y. Rozenholc, E. Zonta, D. Pietraz, A. Didelot, P. Combe, L. Gibault, J-B. Bachet, V. Taly, E. Fabre, H. Blons, P. Laurent-Puig in *Clinical Chemistry*

Novel hybridization- and tag-based error-corrected method for sensitive ctDNA mutation detection using ion semiconductor sequencing Kjersti Tjensvoll, Morten Lapin, Bjørnar Gilje, Herish Garresori, Satu Oltedal, Rakel Brendsdal Forthun, Anders Molven, Yves Rozenholc and Oddmund Nordgård in *Scientific Reports*

See Also

positions_ranges,

Examples

```
bad.pos = "chr7:15478"
PrepareLibrary(info.dir='./',snp.extra=bad.pos,output.dir=paste0(tempdir(),'/'))
```

Index