Package 'CaMutQC'

May 13, 2024

Type Package

Title An R Package for Comprehensive Filtration and Selection of Cancer Somatic Mutations

Version 1.1.0

Description CaMutQC is able to filter false positive mutations generated due to technical issues, as well as to select candidate cancer mutations through a series of well-structured functions by labeling mutations with various flags. And a detailed and vivid filter report will be offered after completing a whole filtration or selection section. Also, CaMutQC integrates serveral methods and gene panels for Tumor Mutational Burden (TMB) estimation.

biocViews Software, QualityControl, GeneTarget

Encoding UTF-8

LazyData false

Suggests knitr, rmarkdown, BiocStyle

VignetteBuilder knitr

RoxygenNote 7.3.1

NeedsCompilation no

BugReports https://github.com/likelet/CaMutQC/issues

URL https://github.com/likelet/CaMutQC

Depends R (>= 4.0.0)

Imports ggplot2, dplyr, org.Hs.eg.db, vcfR, clusterProfiler, stringr, DT, MesKit, maftools, data.table, utils, stats, methods, tidyr

License GPL-3

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

git_branch devel

git_last_commit f2d131d

git_last_commit_date 2024-04-30

Repository Bioconductor 3.20

Date/Publication 2024-05-13

Author Xin Wang [aut, cre] (<https://orcid.org/0000-0002-6072-599X>)

Maintainer Xin Wang <sylviawang555@gmail.com>

Contents

`calTMB` *calTMB*

Description

Calculate Tumor Mutational Burden (TMB) in specific regions.

Usage

```
calTMB(
  maf,
  bedFile = NULL,
  bedHeader = FALSE,
  assay = "MSK-v3",
  genelist = NULL,
  mutType = "nonsynonymous",
  bedFilter = TRUE
)
```

Arguments

maf	An MAF data frame, generated by vcfToMAF function.
bedFile	A file in bed format that contains region information. Default: NULL.
bedHeader	Whether the input bed file has a header or not. Default: FALSE.
assay	Methodology and assay will be applied as a reference, including 'MSK-v3', 'MSK-v2', 'MSK-v1', 'FoundationOne', 'Pan-Cancer Panel' and 'Customized'. Default: 'MSK-v3'.
genelist	A vector of panel gene list, only useful when assay is set to 'Customized'.
mutType	A group of variant classifications that will be kept, only useful when assay is set to 'Pan-Cancer Panel' or 'Customized', including 'exonic', 'nonsynonymous'. and 'all' Default: 'nonsynonymous'.
bedFilter	Whether to filter the information in bed file or not, which only leaves segments in Chr1-Ch22, ChrX and ChrY. Default: TRUE.

Value

A TMB value.

Examples

```
maf <- vcfToMAF(system.file("extdata", "WES_EA_T_1_mutect2.vep.vcf",
package="CaMutQC"))
TMB_value <- calTMB(maf, bedFile=system.file("extdata/bed/panel_hg38",
"FlCDx-hg38.rds", package="CaMutQC"))
```

---

mutFilterAdj *mutFilterAdj*

Description

Filter SNVs with adjacent indels

Usage

```
mutFilterAdj(maf, maxIndelLen = 50, minInterval = 10)
```

Arguments

maf	An MAF data frame, generated by vcfToMAF function.
maxIndelLen	Maximum length of indel accepted to be included. Default: 50
minInterval	Minimum length of interval between an SNV and an indel accepted to be included. Default: 10

Value

An MAF data frame after filtration for adjacent variants.

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf", package="CaMutQC"))
mafF <- mutFilterAdj(maf)
```

---

mutFilterCan *mutFilterCan*

Description

Apply common filtering strategies on a MAF data frame for different cancer types.

Usage

```
mutFilterCan(
  maf,
  cancerType,
  PONfile,
  PONformat = "vcf",
  panel = "Customized",
  tumorDP = 0,
  normalDP = 0,
  tumorAD = 0,
  normalAD = Inf,
  VAF = 0,
  VAFratio = 0,
  SBmethod = "SOR",
  SBscore = Inf,
  maxIndelLen = Inf,
  minInterval = 0,
  tagFILTER = NULL,
  dbVAF = 0.01,
  ExAC = FALSE,
  Genomesprojects1000 = FALSE,
  ESP6500 = FALSE,
  gnomAD = FALSE,
  dbSNP = FALSE,
  keepCOSMIC = FALSE,
  keepType = "all",
  bedFile = NULL,
  bedFilter = TRUE,
  bedHeader = FALSE,
  mutFilter = FALSE,
  selectCols = FALSE,
  report = TRUE,
  reportFile = "FilterReport.html",
  reportDir = "./",
```

```
  TMB = FALSE,
  progressbar = TRUE,
  codelog = FALSE,
  codelogFile = "mutFilterCan.log",
  verbose = TRUE
)
```

Arguments

maf	An MAF data frame.
cancerType	Type of cancer whose filtering parameters need to be referred to. Options are: "COADREAD", "BRCA", "LIHC", "LAML", "LCML", "UCEC", "UCS", "BLCA", "KIRC" and "KIRP"
PONfile	Panel-of-Normals files, which can be either obtained through GATK (https://gatk.broadinstitute.org/hc/en-us/articles/360035890631-Panel-of-Normals-PON-) or generated by users. Should have at least four columns: CHROM, POS, REF, ALT
PONformat	The format of PON file, either "vcf" or "txt". Default: "vcf"
panel	The sequencing panel applied on the dataset. Parameters for mutFilterQual function are set differently for different panels. Default: "Customized". Options: "MSKCC", "WES".
tumorDP	Threshold of tumor total depth. Default: 20
normalDP	Threshold of normal total depth. Default: 10
tumorAD	Threshold of tumor alternative allele depth. Default:5
normalAD	Threshold of normal alternative allele depth. Default: Inf
VAF	Threshold of VAF value. Default: 0.05
VAFratio	Threshold of VAF ratio (tVAF/nVAF). Default: 0
SBmethod	Method will be used to detect strand bias, including 'SOR' and 'Fisher'. Default: 'SOR'. SOR: StrandOddsRatio (https://gatk.broadinstitute.org/hc/en-us/articles/360041849111-StrandOddsRatio)
SBscore	Cutoff strand bias score used to filter variants. Default: 3
maxIndelLen	Maximum length of indel accepted to be included. Default: 50
minInterval	Maximum length of interval between an SNV and an indel accepted to be included. Default: 10
tagFILTER	Variants with spcific tag in the FILTER column will be kept, Default: 'PASS'
dbVAF	Threshold of VAF of certain population for variants in database. Default: 0.01.
ExAC	Whether to filter variants listed in ExAC with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
Genomesprojects1000	Whether to filter variants listed in Genomesprojects1000 with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
ESP6500	Whether to filter variants listed in ESP6500 with VAF higher than cutoff(set in VAF parameter). Default: TRUE.

gnomAD	Whether to filter variants listed in gnomAD with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
dbSNP	Whether to filter variants listed in dbSNP. Default: FALSE.
keepCOSMIC	Whether to keep variants in COSMIC even they have are present in germline database. Default: TRUE.
keepType	A group of variant classifications will be kept, including 'exonic', 'nonsynonymous' and 'all'. Default: 'exonic'.
bedFile	A file in bed format that contains region information. Default: NULL
bedFilter	Whether to filter the information in bed file or not, which only leaves segments in Chr1-Ch22, ChrX and ChrY. Default: TRUE
bedHeader	Whether the input bed file has a header or not. Default: FALSE.
mutFilter	Whether to directly return a filtered MAF data frame. If FALSE, a simulation filtration process will be run, and the original MAF data frame with tags in CaTag column, and a filter report will be returned. If TRUE, a filtered MAF data frame and a filter report will be generated. Default: FALSE
selectCols	Columns will be contained in the filtered data frame. By default (TRUE), the first 13 columns and 'Tumor_Sample_Barcode' column. Or a vector contains column names will be kept.
report	Whether to generate report automatically. Default: TRUE
reportFile	File name of the report. Default: 'FilterReport.html'
reportDir	Path to the output report file. Default: './'
TMB	Whether to calculate TMB. Default: TRUE
progressbar	Whether to show progress bar when running this function Default: TRUE
codelog	If TRUE, your code, along with the parameters you set, will be export in a log file. It will be convenient for users to repeat experiments. Default: FALSE
codelogFile	Where to store the codelog, only useful when codelog is set to TRUE. Default: "mutFilterCan.log"
verbose	Whether to generate message/notification during the filtration process. Default: TRUE.

Value

An MAF data frame after common strategy filtration for a cancer type.

A filter report in HTML format

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf", package="CaMutQC"))
mafF <- mutFilterCan(maf, cancerType='BRCA',
PONfile=system.file("extdata", "PON_test.txt", package="CaMutQC"),
PONformat="txt", TMB=FALSE)
```

mutFilterCom *mutFilterCom*

Description

Apply common filtering strategies on a MAF data frame.

Usage

```
mutFilterCom(
  maf,
  PONfile,
  PONformat = "vcf",
  panel = "Customized",
  tumorDP = 20,
  normalDP = 10,
  tumorAD = 5,
  normalAD = Inf,
  VAF = 0.05,
  VAFratio = 0,
  SBmethod = "SOR",
  SBscore = 3,
  maxIndelLen = 50,
  minInterval = 10,
  tagFILTER = "PASS",
  dbVAF = 0.01,
  ExAC = TRUE,
  Genomesprojects1000 = TRUE,
  gnomAD = TRUE,
  dbSNP = FALSE,
  keepCOSMIC = TRUE,
  keepType = "exonic",
  bedFile = NULL,
  bedHeader = FALSE,
  bedFilter = TRUE,
  mutFilter = FALSE,
  ESP6500 = TRUE,
  selectCols = TRUE,
  report = TRUE,
  assay = "MSK-v3",
  genelist = NULL,
  mutType = "nonsynonymous",
  reportFile = "FilterReport.html",
  reportDir = "./",
  TMB = TRUE,
  cancerType = NULL,
  reference = NULL,
```

```
  progressbar = TRUE,
  codelog = FALSE,
  codelogFile = "mutFilterCom.log",
  verbose = TRUE
)
```

Arguments

maf	An MAF data frame.
PONfile	Panel-of-Normals files, which can be either obtained through GATK (https://gatk.broadinstitute.org/hc/en-us/articles/360035890631-Panel-of-Normals-PON-) or generated by users. Should have at least four columns: CHROM, POS, REF, ALT
PONformat	The format of PON file, either "vcf" or "txt". Default: "vcf"
panel	The sequencing panel applied on the dataset. Parameters for mutFilterQual function are set differently for different panels. Default: "Customized". Options: "MSKCC", "WES".
tumorDP	Threshold of tumor total depth. Default: 20
normalDP	Threshold of normal total depth. Default: 10
tumorAD	Threshold of tumor alternative allele depth. Default: 5
normalAD	Threshold of normal alternative allele depth. Default: Inf
VAF	Threshold of VAF value. Default: 0.05
VAFratio	Threshold of VAF ratio (tVAF/nVAF). Default: 0.
SBmethod	Method will be used to detect strand bias, including 'SOR' and 'Fisher'. Default: 'SOR'. SOR: StrandOddsRatio (https://gatk.broadinstitute.org/hc/en-us/articles/360041849111-StrandOddsRatio)
SBscore	Cutoff strand bias score used to filter variants. Default: 3.
maxIndelLen	Maximum length of indel accepted to be included. Default: 50.
minInterval	Maximum length of interval between an SNV and an indel accepted to be included. Default: 10.
tagFILTER	Variants with spcific tag in the FILTER column will be kept, Default: 'PASS'.
dbVAF	Threshold of VAF value for databases. Default: 0.01.
ExAC	Whether to filter variants listed in ExAC with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
Genomesprojects1000	Whether to filter variants listed in Genomesprojects1000 with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
gnomAD	Whether to filter variants listed in gnomAD with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
dbSNP	Whether to filter variants listed in dbSNP. Default: FALSE.
keepCOSMIC	Whether to keep variants in COSMIC even they have are present in germline database. Default: TRUE.
keepType	A group of variant classifications will be kept, including 'exonic', 'nonsynonymous' and 'all'. Default: 'exonic'.

bedFile	A file in bed format that contains region information. Default: NULL.
bedHeader	Whether the input bed file has a header or not. Default: FALSE.
bedFilter	Whether to filter the information in bed file or not, which only leaves segments in Chr1-Ch22, ChrX and ChrY. Default: TRUE.
mutFilter	Whether to directly return a filtered MAF data frame. If FALSE, a simulation filtration process will be run, and the original MAF data frame with tags in CaTag column, and a filter report will be returned. If TRUE, a filtered MAF data frame and a filter report will be generated. Default: FALSE.
ESP6500	Whether to filter variants listed in ESP6500 with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
selectCols	Columns will be contained in the filtered data frame. By default (TRUE), the first 13 columns and 'Tumor_Sample_Barcode' column. Or a vector contains column names will be kept.
report	Whether to generate report automatically. Default: TRUE
assay	Methodology and assay will be applied as a reference, including 'MSK-v3', 'MSK-v2', 'MSK-v1', 'FoundationOne', 'Pan-Cancer Panel' and 'Customized'. Default: 'MSK-v3'.
genelist	A vector of panel gene list, only useful when assay is set to 'Customized'.
mutType	A group of variant classifications that will be kept, only useful when assay is set to 'Pan-Cancer Panel' or 'Customized', including 'exonic' and 'nonsynonymous'. Default: 'nonsynonymous'.
reportFile	File name of the report. Default: 'FilterReport.html'
reportDir	Path to the output report file. Default: './'.
TMB	Whether to calculate TMB. Default: TRUE.
cancerType	Type of cancer whose filtering parameters need to be referred to. Options are: "COADREAD", "BRCA", "LIHC", "LAML", "LCML", "UCEC", "UCS", "BLCA", "KIRC" and "KIRP"
reference	A specific study whose filtering strategies need to be referred to. Format: "Last_name_of_the_first_author Journal-Year-Cancer_type" Options are: "Haraldsdottir_et_al-Gastroenterology-2014-UCEC", "Cherniack_et_al-Cancer_Cell-2017-UCS", "Mason_et_al-Leukemia-2015-LCML", "Gerlinger_et_al-Engl_J_Med-2012-KIRC" "Zhu_et_al-Nat_Commun-2020-KIRP"
progressbar	Whether to show progress bar when running this function Default: TRUE
codelog	If TRUE, your code, along with the parameters you set, will be export in a log file. It will be convenient for users to repeat experiments. Default: FALSE
codelogFile	Where to store the codelog, only useful when codelog is set to TRUE. Default: "mutFilterCom.log"
verbose	Whether to generate message/notification during the filtration process. Default: TRUE.

Value

An MAF data frame after common strategy filtration

A filter report in HTML format

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf", package="CaMutQC"))
mafF <- mutFilterCom(maf,
PONfile=system.file("extdata", "PON_test.txt", package="CaMutQC"),
TMB=FALSE, report=FALSE, PONformat="txt", verbose=FALSE)
```

`mutFilterDB` *mutFilterDB*

Description

Filter variants in germline database.

Usage

```
mutFilterDB(
  maf,
  dbVAF = 0.01,
  ExAC = TRUE,
  Genomesprojects1000 = TRUE,
  ESP6500 = TRUE,
  gnomAD = TRUE,
  dbSNP = FALSE,
  keepCOSMIC = TRUE,
  verbose = TRUE
)
```

Arguments

`maf`	An MAF data frame, generated by `vcfToMAF` function.
`dbVAF`	Threshold of VAF value for database annotations. Default: 0.01.
`ExAC`	Whether to filter variants listed in ExAC with VAF higher than cutoff (set in dbVAF parameter). Default: TRUE.
`Genomesprojects1000`	Whether to filter variants listed in Genomesprojects1000 with VAF higher than cutoff (set in dbVAF parameter). Default: TRUE.
`ESP6500`	Whether to filter variants listed in ESP6500 with VAF higher than cutoff (set in dbVAF parameter). Default: TRUE.
`gnomAD`	Whether to filter variants listed in gnomAD with VAF higher than cutoff (set in dbVAF parameter). Default: TRUE.
`dbSNP`	Whether to filter variants listed in dbSNP. Default: FALSE.
`keepCOSMIC`	Whether to keep variants in COSMIC even they are present in germline database. Default: TRUE.
`verbose`	Whether to generate message/notification during the filtration process. Default: TRUE.

Value

An MAF data frame after filtration for database and clinical significance

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf", package="CaMutQC"))
mafF <- mutFilterDB(maf)
```

`mutFilterNormalDP` *mutFilterNormalDP*

Description

Filter dbsnp/non-dbsnp variants based on their normal depth. Variants in dbSNP database should have normal depth >= 19, while non-dbSNP variants should have normal depth >= 8 to avoid being filtered.

Usage

```
mutFilterNormalDP(maf, dbsnpCutoff = 19, nonCutoff = 8, verbose = TRUE)
```

Arguments

`maf`	An MAF data frame, generated by `vcfToMAF` function.
`dbsnpCutoff`	Cutoff of normal depth for dnSNP variants. Default: 19.
`nonCutoff`	Cutoff of normal depth for non-dnSNP variants. Default: 8.
`verbose`	Whether to generate message/notification during the filtration process. Default: TRUE.

Value

An MAF data frame where some variants has N tag in CaTag column for Normal depth filtration.

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf", package="CaMutQC"))
mafF <- mutFilterNormalDP(maf)
```

mutFilterPON *mutFilterPON*

Description

Filter variants based on Panel of Normals

Usage

```
mutFilterPON(maf, PONfile, PONformat = "vcf", verbose = TRUE)
```

Arguments

maf	An MAF data frame, generated by vcfToMAF function.
PONfile	Panel-of-Normals files, which can be either obtained through GATK (https://gatk.broadinstitute.org/hc/en-us/articles/360035890631-Panel-of-Normals-PON-) or generated by users. Should have at least four columns: CHROM, POS, REF, ALT
PONformat	The format of PON file, either "vcf" or "txt". Default: "vcf"
verbose	Whether to generate message/notification during the filtration process. Default: TRUE.

Value

An MAF data frame where some variants have P tag in CaTag column for PON filtration.

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf", package="CaMutQC"))
mafF <- mutFilterPON(maf, PONfile=system.file("extdata",
"PON_test.txt", package="CaMutQC"), PONformat="txt")
```

mutFilterQual *mutFilterQual*

Description

Filter variants in low sequencing quality or low confidence.

Usage

```
mutFilterQual(
  maf,
  panel = "Customized",
  tumorDP = 20,
  normalDP = 10,
  tumorAD = 5,
  normalAD = Inf,
  VAF = 0.05,
  VAFratio = 0
)
```

Arguments

maf	An MAF data frame, generated by vcfToMAF function.
panel	The sequencing panel applied on the dataset. Parameters for mutFilterQual function are set differently for different panels. Default: "Customized". Options: "MSKCC", "WES".
tumorDP	Threshold of tumor total depth. Default: 20
normalDP	Threshold of normal total depth. Default: 10
tumorAD	Threshold of tumor alternative allele depth. Default: 5
normalAD	Threshold of normal alternative allele depth. Default: Inf
VAF	Threshold of VAF value. Default: 0.05
VAFratio	Threshold of VAF ratio (tVAF/nVAF). Default: 0

Value

An MAF data frame where some variants have Q tag in CaTag column for sequencing quality filtration

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf", package="CaMutQC"))
mafF <- mutFilterQual(maf)
```

---

mutFilterRef *mutFilterRef*

Description

Use the same filtering strategies that a specific study used, or top-rated strategies shared by users.

Usage

```
mutFilterRef(
  maf,
  reference,
  PONfile,
  PONformat = "vcf",
  tumorDP = 0,
  normalDP = 0,
  tumorAD = 0,
  normalAD = Inf,
  VAF = 0,
  VAFratio = 0,
  SBmethod = "SOR",
  SBscore = Inf,
  maxIndelLen = Inf,
  minInterval = 0,
  tagFILTER = NULL,
  dbVAF = 0.01,
  ExAC = FALSE,
  Genomesprojects1000 = FALSE,
  ESP6500 = FALSE,
  gnomAD = FALSE,
  dbSNP = FALSE,
  keepCOSMIC = FALSE,
  keepType = "all",
  bedFile = NULL,
  bedFilter = TRUE,
  mutFilter = FALSE,
  selectCols = FALSE,
  report = TRUE,
  reportFile = "FilterReport.html",
  reportDir = "./",
  TMB = FALSE,
  progressbar = TRUE,
  codelog = FALSE,
  codelogFile = "mutFilterCom.log",
  verbose = TRUE
)
```

Arguments

maf	An MAF data frame.
reference	A specific study whose filtering strategies need to be referred to. Format: "Last_name_of_the_first_author Journal-Year-Cancer_type" Options are: "Haraldsdottir_et_al-Gastroenterology-2014-UCEC", "Cherniack_et_al-Cancer_Cell-2017-UCS", "Mason_et_al-Leukemia-2015-LCML", "Gerlinger_et_al-Engl_J_Med-2012-KIRC", "Zhu_et_al-Nat_Commun-2020-KIRP"

PONfile	Panel-of-Normals files, which can be either obtained through GATK (https://gatk.broadinstitute.org/hc/en-us/articles/360035890631-Panel-of-Normals-PON-) or generated by users. Should have at least four columns: CHROM, POS, REF, ALT
PONformat	The format of PON file, either "vcf" or "txt". Default: "vcf"
tumorDP	Threshold of tumor total depth. Default: 0
normalDP	Threshold of normal total depth. Default: 0
tumorAD	Threshold of tumor alternative allele depth. Default: 0
normalAD	Threshold of normal alternative allele depth. Default: Inf
VAF	Threshold of VAF value. Default: 0
VAFratio	Threshold of VAF ratio (tVAF/nVAF). Default: 0
SBmethod	Method will be used to detect strand bias, including 'SOR' and 'Fisher'. Default: 'SOR'. SOR: StrandOddsRatio (https://gatk.broadinstitute.org/hc/en-us/articles/360041849111-StrandOddsRatio)
SBscore	Cutoff strand bias score used to filter variants. Default: 3
maxIndelLen	Maximum length of indel accepted to be included. Default: Inf
minInterval	Maximum length of interval between an SNV and an indel accepted to be included. Default: 0
tagFILTER	Variants with spcific tag in the FILTER column will be kept, Default: NULL
dbVAF	Threshold of VAF of certain population for variants in database. Default: 0.01
ExAC	Whether to filter variants listed in ExAC with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
Genomesprojects1000	Whether to filter variants listed in Genomesprojects1000 with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
ESP6500	Whether to filter variants listed in ESP6500 with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
gnomAD	Whether to filter variants listed in gnomAD with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
dbSNP	Whether to filter variants listed in dbSNP. Default: FALSE.
keepCOSMIC	Whether to keep variants in COSMIC even they have are present in germline database. Default: FALSE.
keepType	A group of variant classifications will be kept, including 'exonic', 'nonsynonymous' and 'all'. Default: 'all'.
bedFile	A file in bed format that contains region information. Default: NULL.
bedFilter	Whether to filter the information in bed file or not, which only leaves segments in Chr1-Ch22, ChrX and ChrY. Default: TRUE
mutFilter	Whether to directly return a filtered MAF data frame. If FALSE, a simulation filtration process will be run, and the original MAF data frame with tags in CaTag column, and a filter report will be returned. If TRUE, a filtered MAF data frame and a filter report will be generated. Default: FALSE

selectCols	Columns will be contained in the filtered data frame. By default (TRUE), the first 13 columns and 'Tumor_Sample_Barcode' column. Or a vector contains column names will be kept.
report	Whether to generate report automatically. Default: TRUE
reportFile	File name of the report. Default: 'FilterReport.html'
reportDir	Path to the output report file. Default: './'
TMB	Whether to calculate TMB. Default: TRUE
progressbar	Whether to show progress bar when running this function Default: TRUE
codelog	If TRUE, your code, along with the parameters you set, will be export in a log file. It will be convenient for users to repeat experiments. Default: FALSE
codelogFile	Where to store the codelog, only useful when codelog is set to TRUE. Default: "mutFilterCom.log"
verbose	Whether to generate message/notification during the filtration process. Default: TRUE.

Value

An MAF data frame after applied filtering strategies in another study

A filter report in HTML format

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf",package="CaMutQC"))
mafR <- mutFilterRef(maf, reference="Zhu_et_al-Nat_Commun-2020-KIRP",
PONfile=system.file("extdata","PON_test.txt", package="CaMutQC"),
PONformat="txt", TMB=FALSE, verbose=FALSE, report=FALSE)
```

`mutFilterReg` *mutFilterReg*

Description

Filter variants not in specific regions.

Usage

```
mutFilterReg(
  maf,
  bedFile = NULL,
  bedHeader = FALSE,
  bedFilter = TRUE,
  verbose = TRUE
)
```

Arguments

maf	An MAF data frame, generated by vcfToMAF function.
bedFile	A bed file that contains region information. Default: NULL
bedHeader	Whether the input bed file has a header or not. Default: FALSE.
bedFilter	Whether to filter the information in bed file or not, which only leaves segments in Chr1-Ch22, ChrX and ChrY. Default: TRUE
verbose	Whether to generate message/notification during the filtration process. Default: TRUE.

Value

An MAF data frame where some variants have R tag in CaTag column for region filtration.

Examples

```
maf <- vcfToMAF(system.file("extdata", "WES_EA_T_1_mutect2.vep.vcf",
package="CaMutQC"))
mafF <- mutFilterReg(maf, bedFile=system.file("extdata/bed/panel_hg38",
"Pan-cancer-hg38.rds", package="CaMutQC"))
```

---

mutFilterSB *mutFilterSB*

Description

Filter variants based on strand bias.

Usage

```
mutFilterSB(maf, method = "SOR", SBscore = 3)
```

Arguments

maf	An MAF object, generated by vcfToMAF function.
method	Method will be used to detect strand bias, including 'SOR' and 'Fisher'. Default: 'SOR'. SOR: StrandOddsRatio (https://gatk.broadinstitute.org/hc/en-us/articles/360041849111-StrandOddsRatio) Fisher's Exat Test: Switch to Phred socre (https://gatk.broadinstitute.org/hc/en-us/articles/360035532152-Fisher- s-Exact-Test)
SBscore	Cutoff strand bias score used to filter variants. Default: 3

Value

An MAF data frame where some variants have S tag in CaTag column for strand bias filtration

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf", package="CaMutQC"))
mafF <- mutFilterSB(maf)
```

---

mutFilterTech *mutFilterTech*

Description

Filter potential artifacts produced through technical issue, including filtration for sequencing quality, strand bias, adjacent indel tag, normal depth, panel of normal (PON) and FILTER field.

Usage

```
mutFilterTech(
  maf,
  PONfile,
  PONformat = "vcf",
  panel = "Customized",
  tumorDP = 20,
  normalDP = 10,
  tumorAD = 5,
  normalAD = Inf,
  VAF = 0.05,
  VAFratio = 0,
  SBmethod = "SOR",
  SBscore = 3,
  maxIndelLen = 50,
  minInterval = 10,
  tagFILTER = "PASS",
  progressbar = TRUE,
  verbose = TRUE
)
```

Arguments

maf	An MAF data frame, generated by vcfToMAF function.
PONfile	Panel-of-Normals files, which can be either obtained through GATK (https://gatk.broadinstitute.org/hc/en-us/articles/360035890631-Panel-of-Normals-PON-) or generated by users. Should have at least four columns: CHROM, POS, REF, ALT
PONformat	The format of PON file, either "vcf" or "txt". Default: "vcf"
panel	The sequencing panel applied on the dataset. Parameters for mutFilterQual function are set differently for different panels. Default: "Customized". Options: "MSKCC", "WES".
tumorDP	Threshold of tumor total depth. Default: 20

normalDP	Threshold of normal total depth. Default: 10
tumorAD	Threshold of tumor alternative allele depth. Default: 5
normalAD	Threshold of normal alternative allele depth. Default: Inf
VAF	Threshold of VAF value. Default: 0.05
VAFratio	Threshold of VAF ratio (tVAF/nVAF). Default: 0
SBmethod	Method will be used to detect strand bias, including 'SOR' and 'Fisher'. Default: 'SOR'. SOR: StrandOddsRatio (https://gatk.broadinstitute.org/hc/en-us/articles/360041849111-StrandOddsRatio)
SBscore	Cutoff strand bias score used to filter variants. Default: 3
maxIndelLen	Maximum length of indel accepted to be included. Default: 50
minInterval	Minimum length of interval between an SNV and an indel accepted to be included. Default: 10
tagFILTER	Variants with specific tag in FILTER column will be kept, set to NULL if you want to skip this filter. Default: 'PASS'
progressbar	Whether to show progress bar when running this function Default: TRUE
verbose	Whether to generate message/notification during the filtration process. Default: TRUE.

Value

An MAF data frame after filtration for technical issue

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf", package="CaMutQC"))
mafF <- mutFilterTech(maf, PONfile=system.file("extdata",
"PON_test.txt", package="CaMutQC"), PONformat="txt")
```

---

mutFilterType *mutFilterType*

Description

Filter variants based on variant types

Usage

```
mutFilterType(maf, keepType = "exonic")
```

Arguments

maf	An MAF data frame, generated by vcfToMAF function.
keepType	A group of variant classifications will be kept, including 'exonic', 'nonsynonymous' and 'all'. Default: 'exonic'.

Value

An MAF data frame where some variants has T tag in CaTag column for variant type filtration

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf", package="CaMutQC"))
mafF <- mutFilterType(maf)
```

`mutSelection` *mutSelection*

Description

Select candidate variants for cancer research.

Usage

```
mutSelection(
  maf,
  dbVAF = 0.01,
  ExAC = TRUE,
  Genomesprojects1000 = TRUE,
  ESP6500 = TRUE,
  gnomAD = TRUE,
  dbSNP = FALSE,
  keepCOSMIC = TRUE,
  keepType = "exonic",
  bedFile = NULL,
  bedHeader = FALSE,
  bedFilter = TRUE,
  progressbar = TRUE,
  verbose = TRUE
)
```

Arguments

`maf`	An MAF data frame, generated by `vcfToMAF` function.
`dbVAF`	Threshold of VAF of certain population for variants in database. Default: 0.01
`ExAC`	Whether to filter variants listed in ExAC with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
`Genomesprojects1000`	Whether to filter variants listed in Genomesprojects1000 with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
`ESP6500`	Whether to filter variants listed in ESP6500 with VAF higher than cutoff(set in VAF parameter). Default: TRUE.

gnomAD	Whether to filter variants listed in gnomAD with VAF higher than cutoff(set in VAF parameter). Default: TRUE.
dbSNP	Whether to filter variants listed in dbSNP. Default: FALSE.
keepCOSMIC	Whether to keep variants in COSMIC even they have are present in germline database. Default: TRUE.
keepType	A group of variant classifications will be kept, including 'exonic', 'nonsynonymous' and 'all'. Default: 'exonic'.
bedFile	A file in bed format that contains region information. Default: NULL
bedHeader	Whether the input bed file has a header or not. Default: FALSE.
bedFilter	Whether to filter the information in bed file or not, which only leaves segments in Chr1-Ch22, ChrX and ChrY. Default: TRUE
progressbar	Whether to show progress bar when running this function Default: TRUE
verbose	Whether to generate message/notification during the filtration process. Default: TRUE.

Value

An MAF data frame with variants after selection.

Examples

```
maf <- vcfToMAF(system.file("extdata",
"WES_EA_T_1_mutect2.vep.vcf", package="CaMutQC"))
mafF <- mutSelection(maf)
```

`processMut` *processMut*

Description

Takes union or intersection on multiple MAF data frame, and return 7 important columns.

Usage

```
processMut(mafList, processMethod = "union")
```

Arguments

mafList	A list of MAF data frames after going through at least one CaMutQC filtration function, and the length of the list <= 3.
processMethod	Methods for processing mutations, including "union" and "intersection". Default: "union".

Value

A data frame includes mutations after taking union or intersection.

Examples

```
maf_MuSE <- vcfToMAF(system.file("extdata/Multi-caller",
"WES_EA_T_1.MuSE.vep.vcf", package="CaMutQC"))
maf_MuSE_f <- mutFilterCom(maf_MuSE, report=FALSE, TMB=FALSE,
PONfile=system.file("extdata", "PON_test.txt", package="CaMutQC"),
PONformat="txt")
maf_VarScan2 <- vcfToMAF(system.file("extdata/Multi-caller",
"WES_EA_T_1_varscan_filter_snp.vep.vcf", package="CaMutQC"))
maf_VarScan2_f <- mutFilterCom(maf_VarScan2, report=FALSE, TMB=FALSE,
PONfile=system.file("extdata", "PON_test.txt", package="CaMutQC"),
PONformat="txt")
mafs <- list(maf_MuSE_f, maf_VarScan2_f)
maf_union <- processMut(mafs, processMethod="union")
```

`tomaftools` *tomaftools*

Description

Transform a CaMutQC maf object to a maftools maf object.

Usage

```
tomaftools(
  maf,
  clinicalData = NULL,
  rmFlags = FALSE,
  removeDuplicatedVariants = TRUE,
  useAll = TRUE,
  gisticAllLesionsFile = NULL,
  gisticAmpGenesFile = NULL,
  gisticDelGenesFile = NULL,
  gisticScoresFile = NULL,
  cnLevel = "all",
  cnTable = NULL,
  isTCGA = FALSE,
  vc_nonSyn = NULL,
  verbose = TRUE
)
```

Arguments

`maf`	An MAF data frame, generated by `vcfToMAF` function.
`clinicalData`	Clinical data associated with each # sample/Tumor_Sample_Barcode in MAF. Could be a text file or a data.frame. Default NULL. Inherited from maftools.
`rmFlags`	Default FALSE. Can be TRUE or an integer. If TRUE, removes all the top 20 FLAG genes. If integer, remove top n FLAG genes. Inherited from maftools.

`removeDuplicatedVariants`
: removes repeated variants in a particuar sample, mapped to multiple transcripts of same Gene. See Description. Default TRUE. Inherited from maftools.

`useAll`
: logical. Whether to use all variants irrespective of values in Mutation_Status. Defaults to TRUE. If FALSE, only uses with values Somatic. Inherited from maftools.

`gisticAllLesionsFile`
: All Lesions file generated by gistic. e.g; all_lesions.conf_XX.txt, where XX is the confidence level. Default NULL. Inherited from maftools.

`gisticAmpGenesFile`
: Amplification Genes file generated by gistic. e.g; amp_genes.conf_XX.txt, where XX is the confidence level. Default NULL. Inherited from maftools.

`gisticDelGenesFile`
: Deletion Genes file generated by gistic. e.g; del_genes.conf_XX.txt, where XX is the confidence level. Default NULL. Inherited from maftools.

`gisticScoresFile`
: scores.gistic file generated by gistic. Default NULL Inherited from maftools.

`cnLevel`
: level of CN changes to use. Can be 'all', 'deep' or 'shallow'. Default uses all i.e, genes with both 'shallow' or 'deep' CN changes. Inherited from maftools.

`cnTable`
: Custom copynumber data if gistic results are not available. Input file or a data.frame should contain three columns in aforementioned order with gene name, Sample name and copy number status (either 'Amp' or 'Del'). Default NULL. Inherited from maftools.

`isTCGA`
: Is input MAF file from TCGA source. If TRUE uses only first 12 characters from Tumor_Sample_Barcode. Inherited from maftools.

`vc_nonSyn`
: NULL. Provide manual list of variant classifications to be considered as non-synonymous. Rest will be considered as silent variants. Default uses Variant Classifications with High/Moderate variant consequences. Inherited from maftools.

`verbose`
: TRUE logical. Default to be talkative and prints summary. Inherited from maftools.

Value

An maf object that can be recognized by maftools.

Examples

```
maf_CaMutQC <- vcfToMAF(system.file("extdata/Multi-caller/",
package="CaMutQC"), multiVCF=TRUE)
maf_maftools <- tomaftools(maf_CaMutQC)
```

toMesKit *toMeskit*

Description

Transform a CaMutQC maf object to a MesKit maf object.

Usage

```
toMesKit(
  maf,
  clinicalFile,
  ccfFile = NULL,
  nonSyn.vc = NULL,
  use.indel.ccf = FALSE,
  ccf.conf.level = 0.95
)
```

Arguments

maf	An MAF data frame, generated by vcfToMAF function.
clinicalFile	A clinical data file includes Tumor_Sample_Barcode, Tumor_ID, Patient_ID. Tumor_Sample_Label is optional.
ccfFile	A CCF file of somatic mutations. Default NULL.
nonSyn.vc	List of Variant classifications which are considered as non-silent. Default NULL.
use.indel.ccf	Whether include indels in ccfFile. Default FALSE.
ccf.conf.level	The confidence level of CCF to identify clonal or subclonal. Only works when "CCF_std" or "CCF_CI_high" is provided in ccfFile. Default 0.95.

Value

An maf object that can be recognized by MesKit.

Examples

```
maf_CaMutQC <- vcfToMAF(system.file("extdata/Multi-caller/",
package="CaMutQC"), multiVCF=TRUE)
clin_file <- system.file("extdata", "clin.txt", package="CaMutQC")
maf_MesKit <- toMesKit(maf_CaMutQC, clinicalFile=clin_file)
```

vcfToMAF *vcfToMAF*

Description

Format transformation from VCF to MAF.

Usage

```
vcfToMAF(
  vcfFile,
  multiVCF = FALSE,
  inputStrelka = FALSE,
  writeFile = FALSE,
  MAFfile = "MAF.maf",
  MAFdir = "./",
  tumorSampleName = "Extracted",
  normalSampleName = "Extracted",
  ncbiBuild = "Extracted",
  MAFcenter = ".",
  MAFstrand = "+",
  filterGene = FALSE,
  simplified = FALSE
)
```

Arguments

vcfFile	Directory of a VCF file, or the path to several VCF files that is going to be transformed. Files should be in .vcf or .vcf.gz format.
multiVCF	Logical, whether the input is a path that leads to several VCFs that come from multi-region/sample/caller sequencing. Default: FALSE
inputStrelka	The type of variants ('INDEL' or 'SNV') in VCF file if it is from Strelka. Default: FALSE
writeFile	Whether to directly write MAF file to the disk. If FALSE, a MAF data frame will be returned. If TRUE, a MAF file will be saved. Default: FALSE.
MAFfile	File name of the exported MAF file, if writeFile is set as TRUE.
MAFdir	Directory of the exported MAF file, if writeFile is set as TRUE.
tumorSampleName	Name of the tumor sample(s) in the VCF file(s). If it is set as 'Extracted', tumorSampleName would be extracted automatically from the VCF file. Default: 'Extracted'.
normalSampleName	Name the normal sample in the VCF file. If it is set as 'Extracted', normalSampleName would be extracted automatically from the VCF file. Default: 'Extracted'.

ncbiBuild	The reference genome used for the alignment, which will be presented as value in 'NCBIbuild' column in MAF file. Default: 'GRCh38'.
MAFcenter	One or more genome sequencing center reporting the variant, which will be presented as value in 'Center' column in MAF. Default: '.'.
MAFstrand	Genomic strand of the reported allele, which will be presented as value in 'Strand' column in MAF file. Default: '+'.
filterGene	Logical. Whether to filter variants without Hugo Symbol. Default: FALSE
simplified	Logical. Whether to extract the first thirteen columns after converting to MAF file. Default: FALSE

Value

A detailed MAF data frame

Examples

```
maf <- vcfToMAF(system.file("extdata", "WES_EA_T_1_mutect2.vep.vcf",
package="CaMutQC"))
```

Index