Package 'Neoantimon'

January 5, 2018

Type Package

Title Neoantimon: An R package for automatic identification of tumor-specific neoantigens from sequencing data

Version 1.2

Date 2018-01-04

Author Takanori Hasegawa

Maintainer Takanori Hasegawa <t-hasegw@ims.u-tokyo.ac.jp>

Description This Package is develped to calculate candidates neoantigens from Mutation Data (.vcf) requiring netMHCpan3.0, netMHCIIpan3.1, human refMrna, and refFlat. If you do not have some of these files, see README.md.

License MIT + file LICENSE

VignetteBuilder knitr

Suggests knitr, rmarkdown

LazyData FALSE

Imports utils

RoxygenNote 6.0.1

R topics documented:

`InstallRefFlat` *Get refFlat file*

Description

Get refFlat file

Usage

```
InstallRefFlat(url = NA, export_dir = "lib")
```

Arguments

`url`	Url for getting the corresponding refFlat.txt.gz (Default = "http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/refFlat.txt.gz").
`export_dir`	Export directory (Default = "lib").

Value

void

`InstallRefMrnaFile` *Get refMrna file*

Description

Get refMrna file

Usage

```
InstallRefMrnaFile(url = NA, export_dir = "lib")
```

Arguments

`url`	Url for getting the corresponding refMrna.fa.gz (Default = "http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/refMrna.fa.gz").
`export_dir`	Export directory (Default = "lib").

Value

void

InstallSampleFiles *Get Sample Files for Neoantimon*

Description

Get Sample Files for Neoantimon

Usage

```
InstallSampleFiles(export_dir = "lib")
```

Arguments

export_dir	Export directory (Default = "lib").

Value

void

InstallSamtools *Install Samtools*

Description

Install Samtools

Usage

```
InstallSamtools(url = NA, export_dir = "lib")
```

Arguments

url	Url for getting samtools (Default = "https://github.com/hase62/Neoantimon/raw/master/lib/samtools-0.1.19.tar.bz2").
export_dir	Export directory (Default = "lib").

Value

void

MainINDELClass1 *Calculate Neoantigen Candidates on Indels for MHC Class1*

Description

Calculate Neoantigen Candidates on Indels for MHC Class1

Usage

```
MainINDELClass1(input_file, hla_file, file_name_in_hla_table = input_file,
  refflat_file = paste(hmdir, "lib/refFlat.txt", sep = "/"),
  refmrna_file = paste(hmdir, "lib/refMrna.merge.fa", sep = "/"),
  hmdir = getwd(), job_id = "NO_job_id", export_dir = paste("result",
  file_name_in_hla_table, job_id, sep = "."), rnaexp_file = NA,
  rnabam_file = NA, cnv_file = NA, purity = 1,
  netMHCpan_dir = paste(hmdir, "lib/netMHCpan-3.0/netMHCpan", sep = "/"),
  refdna_file = NA, samtools_dir = NA, bcftools_dir = NA,
  chr_column = NA, mutation_start_column = NA, mutation_end_column = NA,
  mutation_ref_column = NA, mutation_alt_column = NA, nm_id_column = NA,
  depth_normal_column = NA, depth_tumor_column = NA,
  ambiguous_between_exon = 0, ambiguous_codon = 0, peptide_length = c(8,
  9, 10, 11, 12, 13))
```

Arguments

input_file	(Required) An input vcf file annotated by, e.g., ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) or other softwares. See by data(sample_vcf); sample_vcf;
hla_file	(Required) A tab separated file indicating HLA types. The 1st column is input_file name, and the following columns indicate HLA types. See by data(sample_hla_table_c1); sample_hla_table_c1;
file_name_in_hla_table	If the name (1st column) in HLA table is not the same as input_file, indicate the corresponding name (Default=input_file).
refflat_file	refFlat file to be used in constructing peptide. (Default=paste(hmdir, "lib/refFlat.txt", sep="""). See "https://github.com/hase62/Neoantimon"
refmrna_file	refMrna file to be used in constructing peptide (Default=paste(hmdir, "lib/refMrna.fa", sep="""). See "https://github.com/hase62/Neoantimon"
hmdir	Home directory for the analysis (Default = getwd()).
job_id	Job-Id to be attached in output files (Default = "NO_job_id").
export_dir	The directory will be stored results (Default = "paste("result", file_name_in_hla_table, job_id, sep=".")")
rnaexp_file	A file including RNA expressions (Default=NA). The 1st, 2nd and 3rd columns are "GeneSymbol Chr:Exonstart-Exonend (locus) ExpressionAmount", respectively. The 1st row should be any header. See by data(sample_rna_exp); sample_rna_exp;

`rnabam_file` RNA bam file to calculate variant allele frequency of RNA at each mutation (Default=NA).

`cnv_file` A file including copy number variation to calculate cancer cell fraction probability (CCFP) (Default=NA). The format is according to ASCAT output files. The columns are "SNPName Chromosome Position LogR segmentedLogR BAF segmentedBAF CopyNumber MinorAllele RawCopyNumber" The 1st row should be the above header.

See data(sample_copynum); sample_copynum;

`purity` Tumor purity or tumor contents ratio required to calculate CCFP (Default=1).

`netMHCpan_dir` The file directory to netMHCpan (Default="lib/netMHCpan-3.0/netMHCpan").

`refdna_file` refdna_file information to be used to calculate RNA VAF (Default=NA).

See "https://github.com/hase62/Neoantimon"

`samtools_dir` The file directory to samtools_0_x_x (Default="samtools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF.

`bcftools_dir` The file directory to netMHCpan (Default="bcftools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF. samtools 0_x_x includes bcftools in the directory.

`chr_column` The column number describing chromosome number in input_file (Default=NA, but will automatically search "Chr" in header).

`mutation_start_column`
The column number describing mutation start Position in input_file (Default=NA, but will automatically search "Start" in header) .

`mutation_end_column`
The column number describing mutation end Position in input_file (Default=NA, but will automatically search "End" in header).

`mutation_ref_column`
The column number describing mutation Ref in input_file (Default=NA, but will automatically search "Ref" in header).

`mutation_alt_column`
The column number describing mutation Alt in input_file (Default=NA, but will automatically search "Alt" in header).

`nm_id_column` (Required) The column number describing NM IDs in input_file such as "SLCO1C1:NM_001145944:exon7:c.692_693insG:p.L231fs" (Default=NA).

`depth_normal_column`
The column number describing the read count from normal cells (Default = NA).

`depth_tumor_column`
The column number describing the read count from tumor cells (Default = NA).

`ambiguous_between_exon`
The maximum number to permit the differences between Exon-Lengths from refFlat and refMrna (Default=0).

`ambiguous_codon`
The maximum number to permit the differences between inputfile- and refMrna-oriented translation start/end position (Default=0).

`peptide_length` Peptide Length to be generated (Default = 8,9,10,11,12,13).

Value

void (Calculated Neoantigen Files will be generated as .tsv files.)

MainINDELClass2 *Calculate Neoantigen Candidates on Indels for MHC Class2*

Description

Calculate Neoantigen Candidates on Indels for MHC Class2

Usage

```
MainINDELClass2(input_file, hla_file, file_name_in_hla_table = input_file,
  refflat_file = paste(hmdir, "lib/refFlat.txt", sep = "/"),
  refmrna_file = paste(hmdir, "lib/refMrna.merge.fa", sep = "/"),
  hmdir = getwd(), job_id = "NO_job_id", export_dir = paste("result",
  file_name_in_hla_table, job_id, sep = "."), rnaexp_file = NA,
  rnabam_file = NA, cnv_file = NA, purity = 1,
  netMHCIIpan_dir = paste(hmdir, "lib/netMHCIIpan-3.1/netMHCIIpan", sep =
  "/"), refdna_file = NA, samtools_dir = NA, bcftools_dir = NA,
  chr_column = NA, mutation_start_column = NA, mutation_end_column = NA,
  mutation_ref_column = NA, mutation_alt_column = NA, nm_id_column = NA,
  depth_normal_column = NA, depth_tumor_column = NA,
  ambiguous_between_exon = 0, ambiguous_codon = 0, peptide_length = c(15))
```

Arguments

input_file	(Required) An input vcf file annotated by, e.g., ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) or other softwares. See by data(sample_vcf); sample_vcf;
hla_file	(Required) A tab separated file indicating HLA types. The 1st column is input_file name, and the following columns indicate HLA types. See by data(sample_hla_table_c2); sample_hla_table_c2;
file_name_in_hla_table	If the name (1st column) in HLA table is not the same as input_file, indicate the corresponding name (Default=input_file).
refflat_file	refFlat file to be used in constructing peptide. (Default=paste(hmdir, "lib/refFlat.txt", sep=""). See "https://github.com/hase62/Neoantimon"
refmrna_file	refMrna file to be used in constructing peptide (Default=paste(hmdir, "lib/refMrna.fa", sep=""). See "https://github.com/hase62/Neoantimon"
hmdir	Home directory for the analysis (Default = getwd()).
job_id	Job-Id to be attached in output files (Default = "NO_job_id").
export_dir	The directory will be stored results (Default = "paste("result", file_name_in_hla_table, job_id, sep=".")")
rnaexp_file	A file including RNA expressions (Default=NA). The 1st, 2nd and 3rd columns are "GeneSymbol Chr:Exonstart-Exonend (locus) ExpressionAmount", respectively. The 1st row should be any header. See by data(sample_rna_exp); sample_rna_exp;

`rnabam_file` RNA bam file to calculate variant allele frequency of RNA at each mutation (Default=NA).

`cnv_file` A file including copy number variation to calculate cancer cell fraction probability (CCFP) (Default=NA). The format is according to ASCAT output files. The columns are "SNPName Chromosome Position LogR segmentedLogR BAF segmentedBAF CopyNumber MinorAllele RawCopyNumber" The 1st row should be the above header.

See data(sample_copynum); sample_copynum;

`purity` Tumor purity or tumor contents ratio required to calculate CCFP (Default=1).

`netMHCIIpan_dir`
The file directory to netMHCpan (Default="lib/netMHCIIpan-3.1/netMHCpan").

`refdna_file` refdna_file information to be used to calculate RNA VAF (Default=NA).

See "https://github.com/hase62/Neoantimon"

`samtools_dir` The file directory to samtools_0_x_x (Default="samtools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF.

`bcftools_dir` The file directory to netMHCpan (Default="bcftools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF. samtools 0_x_x includes bcftools in the directory.

`chr_column` The column number describing chromosome number in input_file (Default=NA, but will automatically search "Chr" in header).

`mutation_start_column`
The column number describing mutation start Position in input_file (Default=NA, but will automatically search "Start" in header) .

`mutation_end_column`
The column number describing mutation end Position in input_file (Default=NA, but will automatically search "End" in header).

`mutation_ref_column`
The column number describing mutation Ref in input_file (Default=NA, but will automatically search "Ref" in header).

`mutation_alt_column`
The column number describing mutation Alt in input_file (Default=NA, but will automatically search "Alt" in header).

`nm_id_column` (Required) The column number describing NM IDs in input_file such as "SLCO1C1:NM_001145944:exon7:c.692_693insG:p.L231fs" (Default=NA).

`depth_normal_column`
The column number describing the read count from normal cells (Default = NA).

`depth_tumor_column`
The column number describing the read count from tumor cells (Default = NA).

`ambiguous_between_exon`
The maximum number to permit the differences between Exon-Lengths from refFlat and refMrna (Default=0).

`ambiguous_codon`
The maximum number to permit the differences between inputfile- and refMrna-oriented translation start/end position (Default=0).

`peptide_length` Peptide Length to be generated (Default = 15 in HLA Class2).

Value

void (Calculated Neoantigen Files will be generated as .tsv files.)

MainMergeINDELSVClass1 *Merge Results from MainINDELClass1.R or MainSVFUSIONClass1.R*

Description

Merge Results from MainINDELClass1.R or MainSVFUSIONClass1.R

Usage

```
MainMergeINDELSVClass1(hmdir = getwd(), annotation_file, input_dir,
  file_prefix)
```

Arguments

hmdir	Home directory for the analysis (Default=getwd()).
annotation_file	The result annotation file ($vcf.$job_id.peptide.txt) generated by MainINDELClass1() or MainSVFUSIONClass1(). For example, sample_vcf.txt.NO_job_id.peptide.txt.
input_dir	(Required) Directory storing netMHCpan Results.
file_prefix	(Required) File prefix of netMHCpan Results. If you have "sample_annovar.txt.NO_JOB_ID.HLACLASS1.1.peptide.txt", please set "sample_annovar.txt.NO_JOB_ID".

Value

void (Calculated Neoantigen Files will be generated as .tsv files.):

HLA: HLA type used to calculate neoantigen.

Pos: The position of the fraction of peptide used to be evaluated from the full-length peptide.

Gene: Gene symbol used to be evaluated in NetMHCpan.

Evaluated_Mutant_Peptide_Core: The core peptide of the mutant peptide to be evaluated in NetMHCpan.

Evaluated_Mutant_Peptide: The mutant peptide to be evaluated.

Mut_IC50: IC50 value for evaluated mutant peptide.

Mut_Rank: Rank value for evaluated mutanat peptide.

Chr: Chromosome Number of the mutation.

NM_ID: NM_ID used to construct peptides from the mutation.

Change: The annotation to be described in .vcf file.

Ref: reference type nucleic acid base.

Alt: alternative type nucleic acid base.

Prob: A probability of reference nucleic acid base described in .vcf file.

Mutation_Prob: A probability of alternative nucleic acid base described in .vcf file.

Exon_Start: The exon start position of the corrsponding NM_ID.

Exon_End: The exon end position of the corrsponding NM_ID.

Mutation_Position: The mutation position of the corrsponding NM_ID.

Total_Depth: The depth of the reference nucleic acid base.

Tumor_Depth: The depth of the alternative nucleic acid base.

Wt_Peptide: The full-length of the wild-type peptide.

Mutant_Peptide: The full-length of the mutant peptide.

Total_RNA: The expression amount of the corresponding RNA.

Tumor_RNA_Ratio: The variant allele frequency of the corresponding RNA.

Tumor_RNA: The modified amount of the corresponding RNA level based on RNA Reads.

Tumor_RNA_based_on_DNA: The modified amount of the corresponding RNA level based on DNA Reads.

MutRatio: The mean value of the cancer cell fraction probability.

MutRatio_Min: The 1% percentile of the cancer cell fraction probability.

MutRatio_Max: The 99% percentile of the cancer cell fraction probability.

`MainMergeINDELSVClass2` *Merge Results from MainINDELClass2.R or MainSVFUSIONClass2.R*

Description

Merge Results from MainINDELClass2.R or MainSVFUSIONClass2.R

Usage

```
MainMergeINDELSVClass2(hmdir = getwd(), annotation_file, input_dir,
  file_prefix)
```

Arguments

`hmdir`	Home directory for the analysis (Default=getwd()).
`annotation_file`	The result annotation file ($vcf.$job_id.peptide.txt) generated by MainINDELClass2() or MainSVFUSIONClass2(). For example, sample_vcf.txt.NO_job_id.peptide.txt.
`input_dir`	(Required) Directory storing netMHCpan Results.
`file_prefix`	(Required) File prefix of netMHCpan Results. If you have "sample_annovar.txt.NO_JOB_ID.HLACLASS2.1.peptide.txt", please set "sample_annovar.txt.NO_JOB_ID".

Value

void (Calculated Neoantigen Files will be generated as .tsv files.):

HLA: HLA type used to calculate neoantigen.

Pos: The position of the fraction of peptide used to be evaluated from the full-length peptide.

Gene: Gene symbol used to be evaluated in NetMHCpan.

Evaluated_Mutant_Peptide_Core: The core peptide of the mutant peptide to be evaluated in NetMHCpan.

Evaluated_Mutant_Peptide: The mutant peptide to be evaluated.

Mut_IC50: IC50 value for evaluated mutant peptide.

Mut_Rank: Rank value for evaluated mutanat peptide.

Chr: Chromosome Number of the mutation.

NM_ID: NM_ID used to construct peptides from the mutation.

Change: The annotation to be described in .vcf file.

Ref: reference type nucleic acid base.

Alt: alternative type nucleic acid base.

Prob: A probability of reference nucleic acid base described in .vcf file.

Mutation_Prob: A probability of alternative nucleic acid base described in .vcf file.

Exon_Start: The exon start position of the corrsponding NM_ID.

Exon_End: The exon end position of the corrsponding NM_ID.

Mutation_Position: The mutation position of the corrsponding NM_ID.

Total_Depth: The depth of the reference nucleic acid base.

Tumor_Depth: The depth of the alternative nucleic acid base.

Wt_Peptide: The full-length of the wild-type peptide.

Mutant_Peptide: The full-length of the mutant peptide.

Total_RNA: The expression amount of the corresponding RNA.

Tumor_RNA_Ratio: The variant allele frequency of the corresponding RNA.

Tumor_RNA: The modified amount of the corresponding RNA level based on RNA Reads.

Tumor_RNA_based_on_DNA: The modified amount of the corresponding RNA level based on DNA Reads.

MutRatio: The mean value of the cancer cell fraction probability.

MutRatio_Min: The 1% percentile of the cancer cell fraction probability.

MutRatio_Max: The 99% percentile of the cancer cell fraction probability.

`MainMergeSNVClass1` *Merge Results from MainSNVClass1.R*

Description

Merge Results from MainSNVClass1.R

Usage

```
MainMergeSNVClass1(hmdir = getwd(), annotation_file, input_dir, file_prefix)
```

Arguments

hmdir	Home directory for the analysis (Default=getwd()).
annotation_file	The result annotation file ($vcf.$job_id.peptide.txt) generated by MainSNVClass1(). For example, sample_vcf.txt.NO_job_id.peptide.txt.
input_dir	(Required) Directory storing netMHCpan Results.
file_prefix	(Required) File prefix of netMHCpan Results. If you have "sample_annovar.txt.NO_JOB_ID.HLACLASS1.1.peptide.txt", please set "sample_annovar.txt.NO_JOB_ID".

Value

void (Calculated Neoantigen Files will be generated as .tsv files.):

HLA: HLA type used to calculate neoantigen.

Pos: The position of the fraction of peptide used to be evaluated from the full-length peptide.

Gene: Gene symbol used to be evaluated in NetMHCpan.

Evaluated_Mutant_Peptide: The mutant peptide to be evaluated.

Mut_IC50: IC50 value for evaluated mutant peptide.

Mut_Rank: Rank value for evaluated mutanat peptide.

Evaluated_Wt_Peptide: The wild-type peptide to be evaluated.

Wt_IC50: IC50 value for evaluated wild-type peptide.

Wt_Rank: Rank value for evaluated wild-type peptide.

Chr: Chromosome Number of the mutation.

NM_ID: NM_ID used to construct peptides from the mutation.

Change: The annotation to be described in .vcf file.

Ref: reference type nucleic acid base.

Alt: alternative type nucleic acid base.

Prob: A probability of reference nucleic acid base described in .vcf file.

Mutation_Prob: A probability of alternative nucleic acid base described in .vcf file.

Exon_Start: The exon start position of the corrsponding NM_ID.

Exon_End: The exon end position of the corrsponding NM_ID.

Mutation_Position: The mutation position of the corrsponding NM_ID.

Total_Depth: The depth of the reference nucleic acid base.

Tumor_Depth: The depth of the alternative nucleic acid base.

Wt_Peptide: The full-length of the wild-type peptide.

Mutant_Peptide: The full-length of the mutant peptide.

Total_RNA: The expression amount of the corresponding RNA.

Tumor_RNA_Ratio: The variant allele frequency of the corresponding RNA.

Tumor_RNA: The modified amount of the corresponding RNA level based on RNA Reads.

Tumor_RNA_based_on_DNA: The modified amount of the corresponding RNA level based on DNA Reads.

MutRatio: The mean value of the cancer cell fraction probability.

MutRatio_Min: The 1% percentile of the cancer cell fraction probability.

MutRatio_Max: The 99% percentile of the cancer cell fraction probability.

`MainMergeSNVClass2` *Merge Results from MainSNVClass2.R*

Description

Merge Results from MainSNVClass2.R

Usage

```
MainMergeSNVClass2(hmdir = getwd(), annotation_file, input_dir, file_prefix)
```

Arguments

`hmdir`	Home directory for the analysis (Default=getwd()).
`annotation_file`	The result annotation file ($vcf.$job_id.peptide.txt) generated by MainSNVClass2(). For example, sample_vcf.txt.NO_job_id.peptide.txt.
`input_dir`	(Required) Directory storing netMHCpan Results.
`file_prefix`	(Required) File prefix of netMHCpan Results. If you have "sample_annovar.txt.NO_JOB_ID.HLACLASS2.1.peptide.txt", please set "sample_annovar.txt.NO_JOB_ID".

Value

void (Calculated Neoantigen Files will be generated as .tsv files.):

HLA: HLA type used to calculate neoantigen.

Pos: The position of the fraction of peptide used to be evaluated from the full-length peptide.

Gene: Gene symbol used to be evaluated in NetMHCpan.

Evaluated_Mutant_Peptide: The mutant peptide to be evaluated.

Mut_IC50: IC50 value for evaluated mutant peptide.

Mut_Rank: Rank value for evaluated mutanat peptide.

Evaluated_Wt_Peptide: The wild-type peptide to be evaluated.

Wt_IC50: IC50 value for evaluated wild-type peptide.

Wt_Rank: Rank value for evaluated wild-type peptide.

Chr: Chromosome Number of the mutation.

NM_ID: NM_ID used to construct peptides from the mutation.

Change: The annotation to be described in .vcf file.

Ref: reference type nucleic acid base.

Alt: alternative type nucleic acid base.

Prob: A probability of reference nucleic acid base described in .vcf file.

Mutation_Prob: A probability of alternative nucleic acid base described in .vcf file.

Exon_Start: The exon start position of the corrsponding NM_ID.

Exon_End: The exon end position of the corrsponding NM_ID.

Mutation_Position: The mutation position of the corrsponding NM_ID.

Total_Depth: The depth of the reference nucleic acid base.

Tumor_Depth: The depth of the alternative nucleic acid base.

Wt_Peptide: The full-length of the wild-type peptide.

Mutant_Peptide: The full-length of the mutant peptide.

Total_RNA: The expression amount of the corresponding RNA.

Tumor_RNA_Ratio: The variant allele frequency of the corresponding RNA.

Tumor_RNA: The modified amount of the corresponding RNA level based on RNA Reads.

Tumor_RNA_based_on_DNA: The modified amount of the corresponding RNA level based on DNA Reads.

MutRatio: The mean value of the cancer cell fraction probability.

MutRatio_Min: The 1% percentile of the cancer cell fraction probability.

MutRatio_Max: The 99% percentile of the cancer cell fraction probability.

`MainSNVClass1` *Calculate Neoantigen Candidates on SNVs for MHC Class1*

Description

Calculate Neoantigen Candidates on SNVs for MHC Class1

Usage

```
MainSNVClass1(input_file, hla_file, file_name_in_hla_table = input_file,
  refflat_file = paste(hmdir, "lib/refFlat.txt", sep = "/"),
  refmrna_file = paste(hmdir, "lib/refMrna.merge.fa", sep = "/"),
  hmdir = getwd(), job_id = "NO_job_id", export_dir = paste("result",
  file_name_in_hla_table, job_id, sep = "."), rnaexp_file = NA,
  rnabam_file = NA, cnv_file = NA, purity = 1,
  netMHCpan_dir = paste(hmdir, "lib/netMHCpan-3.0/netMHCpan", sep = "/"),
  refdna_file = NA, samtools_dir = NA, bcftools_dir = NA,
  chr_column = NA, mutation_start_column = NA, mutation_end_column = NA,
  mutation_ref_column = NA, mutation_alt_column = NA, nm_id_column = NA,
  depth_normal_column = NA, depth_tumor_column = NA,
  ambiguous_between_exon = 0, ambiguous_codon = 0, peptide_length = c(8,
  9, 10, 11, 12, 13))
```

Arguments

`input_file`	(Required) An input vcf file annotated by, e.g., ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) or other softwares. See by data(sample_vcf); sample_vcf;
`hla_file`	(Required) A tab separated file indicating HLA types. The 1st column is input_file name, and the following columns indicate HLA types. See by data(sample_hla_table_c1); sample_hla_table_c1;

`file_name_in_hla_table`
: If the name (1st column) in HLA table is not the same as input_file, indicate the corresponding name (Default=input_file).

`refflat_file`
: refFlat file to be used in constructing peptide. (Default=paste(hmdir, "lib/refFlat.txt", sep="").
See "https://github.com/hase62/Neoantimon"

`refmrna_file`
: refMrna file to be used in constructing peptide (Default=paste(hmdir, "lib/refMrna.fa", sep="").
See "https://github.com/hase62/Neoantimon"

`hmdir`
: Home directory for the analysis (Default = getwd()).

`job_id`
: Job-Id to be attached in output files (Default = "NO_job_id").

`export_dir`
: The directory will be stored results (Default = "paste("result", file_name_in_hla_table, job_id, sep=".")")

`rnaexp_file`
: A file including RNA expressions (Default=NA). The 1st, 2nd and 3rd columns are "GeneSymbol Chr:Exonstart-Exonend (locus) ExpressionAmount", respectively. The 1st row should be any header.
See by data(sample_rna_exp); sample_rna_exp;

`rnabam_file`
: RNA bam file to calculate variant allele frequency of RNA at each mutation (Default=NA).

`cnv_file`
: A file including copy number variation to calculate cancer cell fraction probability (CCFP) (Default=NA). The format is according to ASCAT output files. The columns are "SNPName Chromosome Position LogR segmentedLogR BAF segmentedBAF CopyNumber MinorAllele RawCopyNumber" The 1st row should be the above header.
See data(sample_copynum); sample_copynum;

`purity`
: Tumor purity or tumor contents ratio required to calculate CCFP (Default=1).

`netMHCpan_dir`
: The file directory to netMHCpan (Default="lib/netMHCpan-3.0/netMHCpan").

`refdna_file`
: refdna_file information to be used to calculate RNA VAF (Default=NA).
See "https://github.com/hase62/Neoantimon"

`samtools_dir`
: The file directory to samtools_0_x_x (Default="samtools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF .

`bcftools_dir`
: The file directory to netMHCpan (Default="bcftools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF . samtools 0_x_x includes bcftools in the directory.

`chr_column`
: The column number describing chromosome number in input_file (Default=NA, but will automatically search "Chr" in header).

`mutation_start_column`
: The column number describing mutation start Position in input_file (Default=NA, but will automatically search "Start" in header) .

`mutation_end_column`
: The column number describing mutation end Position in input_file (Default=NA, but will automatically search "End" in header).

`mutation_ref_column`
: The column number describing mutation Ref in input_file (Default=NA, but will automatically search "Ref" in header).

`mutation_alt_column`
: The column number describing mutation Alt in input_file (Default=NA, but will automatically search "Alt" in header).

nm_id_column (Required) The column number describing NM IDs in input_file such as "SLCO1C1:NM_001145944:exon7:c.692_693insG:p.L231fs" (Default=NA).

depth_normal_column
The column number describing the read count from normal cells (Default = NA).

depth_tumor_column
The column number describing the read count from tumor cells (Default = NA).

ambiguous_between_exon
The maximum number to permit the differences between Exon-Lengths from refFlat and refMrna (Default=0).

ambiguous_codon
The maximum number to permit the differences between inputfile- and refMrna-oriented translation start/end position (Default=0).

peptide_length Peptide Length to be generated (Default = 8,9,10,11,12,13).

Value

void (Calculated Neoantigen Files will be generated as .tsv files.)

MainSNVClass2 *Calculate Neoantigen Candidates on SNVs for MHC Class2*

Description

Calculate Neoantigen Candidates on SNVs for MHC Class2

Usage

```
MainSNVClass2(input_file, hla_file, file_name_in_hla_table = input_file,
  refflat_file = paste(hmdir, "lib/refFlat.txt", sep = "/"),
  refmrna_file = paste(hmdir, "lib/refMrna.merge.fa", sep = "/"),
  hmdir = getwd(), job_id = "NO_job_id", export_dir = paste("result",
  file_name_in_hla_table, job_id, sep = "."), rnaexp_file = NA,
  rnabam_file = NA, cnv_file = NA, purity = 1,
  netMHCIIpan_dir = paste(hmdir, "lib/netMHCIIpan-3.1/netMHCIIpan", sep =
  "/"), refdna_file = NA, samtools_dir = NA, bcftools_dir = NA,
  chr_column = NA, mutation_start_column = NA, mutation_end_column = NA,
  mutation_ref_column = NA, mutation_alt_column = NA, nm_id_column = NA,
  depth_normal_column = NA, depth_tumor_column = NA,
  ambiguous_between_exon = 0, ambiguous_codon = 0, peptide_length = c(15))
```

Arguments

input_file (Required) An input vcf file annotated by,
e.g., ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) or other softwares.
See by data(sample_vcf); sample_vcf;

hla_file (Required) A tab separated file indicating HLA types. The 1st column is input_file name, and the following columns indicate HLA types.
See by data(sample_hla_table_c2); sample_hla_table_c2;

`file_name_in_hla_table`
: If the name (1st column) in HLA table is not the same as input_file, indicate the corresponding name (Default=input_file).

`refflat_file`
: refFlat file to be used in constructing peptide. (Default=paste(hmdir, "lib/refFlat.txt", sep="").
See "https://github.com/hase62/Neoantimon"

`refmrna_file`
: refMrna file to be used in constructing peptide (Default=paste(hmdir, "lib/refMrna.fa", sep="").
See "https://github.com/hase62/Neoantimon"

`hmdir`
: Home directory for the analysis (Default = getwd()).

`job_id`
: Job-Id to be attached in output files (Default = "NO_job_id").

`export_dir`
: The directory will be stored results (Default = "paste("result", file_name_in_hla_table, job_id, sep=".")")

`rnaexp_file`
: A file including RNA expressions (Default=NA). The 1st, 2nd and 3rd columns are "GeneSymbol Chr:Exonstart-Exonend (locus) ExpressionAmount", respectively. The 1st row should be any header.
See by data(sample_rna_exp); sample_rna_exp;

`rnabam_file`
: RNA bam file to calculate variant allele frequency of RNA at each mutation (Default=NA).

`cnv_file`
: A file including copy number variation to calculate cancer cell fraction probability (CCFP) (Default=NA). The format is according to ASCAT output files. The columns are "SNPName Chromosome Position LogR segmentedLogR BAF segmentedBAF CopyNumber MinorAllele RawCopyNumber" The 1st row should be the above header.
See data(sample_copynum); sample_copynum;

`purity`
: Tumor purity or tumor contents ratio required to calculate CCFP (Default=1).

`netMHCIIpan_dir`
: The file directory to netMHCpan (Default="lib/netMHCIIpan-3.1/netMHCpan").

`refdna_file`
: refdna_file information to be used to calculate RNA VAF (Default=NA).
See "https://github.com/hase62/Neoantimon"

`samtools_dir`
: The file directory to samtools_0_x_x (Default="samtools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF .

`bcftools_dir`
: The file directory to netMHCpan (Default="bcftools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF . samtools 0_x_x includes bcftools in the directory.

`chr_column`
: The column number describing chromosome number in input_file (Default=NA, but will automatically search "Chr" in header).

`mutation_start_column`
: The column number describing mutation start Position in input_file (Default=NA, but will automatically search "Start" in header) .

`mutation_end_column`
: The column number describing mutation end Position in input_file (Default=NA, but will automatically search "End" in header).

`mutation_ref_column`
: The column number describing mutation Ref in input_file (Default=NA, but will automatically search "Ref" in header).

mutation_alt_column
: The column number describing mutation Alt in input_file (Default=NA, but will automatically search "Alt" in header).

nm_id_column
: (Required) The column number describing NM IDs in input_file such as "SLCO1C1:NM_001145944:exon7:c.692_693insG:p.L231fs" (Default=NA).

depth_normal_column
: The column number describing the read count from normal cells (Default = NA).

depth_tumor_column
: The column number describing the read count from tumor cells (Default = NA).

ambiguous_between_exon
: The maximum number to permit the differences between Exon-Lengths from refFlat and refMrna (Default=0).

ambiguous_codon
: The maximum number to permit the differences between inputfile- and refMrna-oriented translation start/end position (Default=0).

peptide_length
: Peptide Length to be generated (Default=15 in HLA Class2).

Value

void (Calculated Neoantigen Files will be generated as .tsv files.)

`MainSVFUSIONClass1` *Calculate Neoantigen Candidates on SV fusions for MHC Class1*

Description

Calculate Neoantigen Candidates on SV fusions for MHC Class1

Usage

```
MainSVFUSIONClass1(input_file, hla_file, file_name_in_hla_table = input_file,
  refflat_file = paste(hmdir, "lib/refFlat.txt", sep = "/"),
  refmrna_file = paste(hmdir, "lib/refMrna.merge.fa", sep = "/"),
  hmdir = getwd(), job_id = "NO_job_id", export_dir = paste("result",
  file_name_in_hla_table, job_id, sep = "."), rnaexp_file = NA,
  rnabam_file = NA, cnv_file = NA, purity = 1,
  netMHCpan_dir = paste(hmdir, "lib/netMHCpan-3.0/netMHCIIpan", sep = "/"),
  refdna_file = NA, samtools_dir = NA, bcftools_dir = NA,
  chr_column = NA, mutation_start_column = NA, mutation_end_column = NA,
  mutation_ref_column = NA, mutation_alt_bnd_column = NA,
  depth_normal_column = NA, depth_tumor_column = NA, nm_id_column = NA,
  ambiguous_between_exon = 0, ambiguous_codon = 0, peptide_length = c(8,
  9, 10, 11, 12, 13), gene_symbol_column = NA, mate_id_column = NA)
```

Arguments

input_file
: (Required) An input vcf file (BND format) annotated by,
e.g., ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) or other softwares.
See by data(sample_sv_bnd); sample_sv_bnd;

`hla_file` (Required) A tab separated file indicating HLA types. The 1st column is input_file name, and the following columns indicate HLA types.
See by data(sample_hla_table_c1); sample_hla_table_c1;

`file_name_in_hla_table`
If the name (1st column) in HLA table is not the same as input_file, indicate the corresponding name (Default=input_file).

`refflat_file` refFlat file to be used in constructing peptide. (Default=paste(hmdir, "lib/refFlat.txt",sep="").
See "https://github.com/hase62/Neoantimon"

`refmrna_file` refMrna file to be used in constructing peptide (Default=paste(hmdir, "lib/refMrna.fa", sep="").
See "https://github.com/hase62/Neoantimon"

`hmdir` Home directory for the analysis (Default = getwd()).

`job_id` Job-Id to be attached in output files (Default = "NO_job_id").

`export_dir` The directory will be stored results (Default = "paste("result", file_name_in_hla_table, job_id, sep=".")")

`rnaexp_file` A file including RNA expressions (Default=NA). The 1st, 2nd and 3rd columns are "GeneSymbol Chr:Exonstart-Exonend (locus) ExpressionAmount", respectively. The 1st row should be any header.
See by data(sample_rna_exp); sample_rna_exp;

`rnabam_file` RNA bam file to calculate variant allele frequency of RNA at each mutation (Default=NA).

`cnv_file` A file including copy number variation to calculate cancer cell fraction probability (CCFP) (Default=NA). The format is according to ASCAT output files. The columns are "SNPName Chromosome Position LogR segmentedLogR BAF segmentedBAF CopyNumber MinorAllele RawCopyNumber" The 1st row should be the above header.
See data(sample_copynum); sample_copynum;

`purity` Tumor purity or tumor contents ratio required to calculate CCFP (Default=1).

`netMHCpan_dir` The file directory to netMHCpan (Default="lib/netMHCpan-3.0/netMHCpan").

`refdna_file` (Required) refdna_file information to be used to create SVs Region (Default=NA).
See "https://github.com/hase62/Neoantimon"

`samtools_dir` The file directory to samtools_0_x_x (Default="samtools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF .

`bcftools_dir` The file directory to netMHCpan (Default="bcftools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF . samtools 0_x_x includes bcftools in the directory.

`chr_column` The column number describing chromosome number in input_file (Default=NA, but will automatically search "Chr" in header).

`mutation_start_column`
The column number describing mutation start Position in input_file (Default=NA, but will automatically search "Start" in header) .

`mutation_end_column`
The column number describing mutation end Position in input_file (Default=NA, but will automatically search "End" in header).

`mutation_ref_column`
The column number describing mutation Ref in input_file (Default=NA, but will automatically search "Ref" in header).

mutation_alt_bnd_column
: The column number describing mutation Alt (BND format) in input_file (Default=NA, but will automatically search "Alt" in header).

depth_normal_column
: The column number describing the read count from normal cells (Default = NA).

depth_tumor_column
: The column number describing the read count from tumor cells (Default = NA).

nm_id_column
: (Required if gene_symbol_column = NA) The column number describing NM IDs in input_file such as "SLCO1C1:NM_001145944:exon7:c.692_693insG:p.L231fs" (Default=NA).

ambiguous_between_exon
: The maximum number to permit the differences between Exon-Lengths from refFlat and refMrna (Default=0).

ambiguous_codon
: The maximum number to permit the differences between inputfile- and refMrna-oriented translation start/end position (Default=0).

peptide_length
: Peptide Length to be generated (Default = 8,9,10,11,12,13).

gene_symbol_column
: (Required if nm_id_column = NA) The column number describing gene symbol in input_file (Default=NA).

mate_id_column
: (Required) The column indicating mateIDs or svIDs such as "SVMERGE1_1" (Default=NA).

Value

void (Calculated Neoantigen Files will be generated as .tsv files.)

`MainSVFUSIONClass2` *Calculate Neoantigen Candidates on SV fusions for MHC Class2*

Description

Calculate Neoantigen Candidates on SV fusions for MHC Class2

Usage

```
MainSVFUSIONClass2(input_file, hla_file, file_name_in_hla_table = input_file,
  refflat_file = paste(hmdir, "lib/refFlat.txt", sep = "/"),
  refmrna_file = paste(hmdir, "lib/refMrna.merge.fa", sep = "/"),
  hmdir = getwd(), job_id = "NO_job_id", export_dir = paste("result",
  file_name_in_hla_table, job_id, sep = "."), rnaexp_file = NA,
  rnabam_file = NA, cnv_file = NA, purity = 1,
  netMHCIIpan_dir = paste(hmdir, "lib/netMHCIIpan-3.1/netMHCIIpan", sep =
  "/"), refdna_file = NA, samtools_dir = NA, bcftools_dir = NA,
  chr_column = NA, mutation_start_column = NA, mutation_end_column = NA,
  mutation_ref_column = NA, mutation_alt_bnd_column = NA,
  depth_normal_column = NA, depth_tumor_column = NA, nm_id_column = NA,
  ambiguous_between_exon = 0, ambiguous_codon = 0, peptide_length = c(15),
  gene_symbol_column = NA, mate_id_column = NA)
```

Arguments

input_file	(Required) An input vcf file (BND format) annotated by, e.g., ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) or other softwares. See by data(sample_sv_bnd); sample_sv_bnd;
hla_file	(Required) A tab separated file indicating HLA types. The 1st column is input_file name, and the following columns indicate HLA types. See by data(sample_hla_table_c1); sample_hla_table_c1;
file_name_in_hla_table	If the name (1st column) in HLA table is not the same as input_file, indicate the corresponding name (Default=input_file).
refflat_file	refFlat file to be used in constructing peptide. (Default=paste(hmdir, "lib/refFlat.txt",sep="") See "https://github.com/hase62/Neoantimon"
refmrna_file	refMrna file to be used in constructing peptide (Default=paste(hmdir, "lib/refMrna.fa", sep="") See "https://github.com/hase62/Neoantimon"
hmdir	Home directory for the analysis (Default = getwd()).
job_id	Job-Id to be attached in output files (Default = "NO_job_id").
export_dir	The directory will be stored results (Default = "paste("result", file_name_in_hla_table, job_id, sep=".")")
rnaexp_file	A file including RNA expressions (Default=NA). The 1st, 2nd and 3rd columns are "GeneSymbol Chr:Exonstart-Exonend (locus) ExpressionAmount", respectively. The 1st row should be any header. See by data(sample_rna_exp); sample_rna_exp;
rnabam_file	RNA bam file to calculate variant allele frequency of RNA at each mutation (Default=NA).
cnv_file	A file including copy number variation to calculate cancer cell fraction probability (CCFP) (Default=NA). The format is according to ASCAT output files. The columns are "SNPName Chromosome Position LogR segmentedLogR BAF segmentedBAF CopyNumber MinorAllele RawCopyNumber" The 1st row should be the above header. See data(sample_copynum); sample_copynum;
purity	Tumor purity or tumor contents ratio required to calculate CCFP (Default=1).
netMHCIIpan_dir	The file directory to netMHCpan (Default="lib/netMHCIIpan-3.1/netMHCpan").
refdna_file	(Required) refdna_file information to be used to create SVs Region (Default=NA). See "https://github.com/hase62/Neoantimon"
samtools_dir	The file directory to samtools_0_x_x (Default="samtools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF .
bcftools_dir	The file directory to netMHCpan (Default="bcftools"). It shouled be indicated when you indicate RNA-bam and try to calculate RNA VAF . samtools 0_x_x includes bcftools in the directory.
chr_column	The column number describing chromosome number in input_file (Default=NA, but will automatically search "Chr" in header).
mutation_start_column	The column number describing mutation start Position in input_file (Default=NA, but will automatically search "Start" in header) .

`mutation_end_column`
: The column number describing mutation end Position in input_file (Default=NA, but will automatically search "End" in header).

`mutation_ref_column`
: The column number describing mutation Ref in input_file (Default=NA, but will automatically search "Ref" in header).

`mutation_alt_bnd_column`
: The column number describing mutation Alt (BND format) in input_file (Default=NA, but will automatically search "Alt" in header).

`depth_normal_column`
: The column number describing the read count from normal cells (Default = NA).

`depth_tumor_column`
: The column number describing the read count from tumor cells (Default = NA).

`nm_id_column`
: (Required if gene_symbol_column = NA) The column number describing NM IDs in input_file such as "SLCO1C1:NM_001145944:exon7:c.692_693insG:p.L231fs" (Default=NA).

`ambiguous_between_exon`
: The maximum number to permit the differences between Exon-Lengths from refFlat and refMrna (Default=0).

`ambiguous_codon`
: The maximum number to permit the differences between inputfile- and refMrna-oriented translation start/end position (Default=0).

`peptide_length`
: Peptide Length to be generated (Default = 15 in HLA Class2).

`gene_symbol_column`
: (Required if nm_id_column = NA) The column number describing gene symbol in input_file (Default=NA).

`mate_id_column`
: (Required) The column indicating mateIDs or svIDs such as "SVMERGE1_1" (Default=NA).

Value

void (Calculated Neoantigen Files will be generated as .tsv files.)

`sample_copynum` *A Format / Sample file for Copy Number Information*

Description

A dataset containing the copy number information obtained by, e.g., ASCAT.

Usage

```
data(sample_copynum)
```

Format

A data frame with 7 rows and 9 variables

sample_hla_table_c1 *A Format / Sample file for HLA CLASS1 Table*

Description

A dataset containing the HLA types of patients in each row.

Usage

```
data(sample_hla_table_c1)
```

Format

A data frame with 3 rows and at most 7 variables

sample_hla_table_c2 *A Format / Sample file for HLA CLASS2 Table*

Description

A dataset containing the HLA types of patients in each row.

Usage

```
data(sample_hla_table_c2)
```

Format

A data frame with at least 3 row and at most 10 variables

sample_result_INDEL_CLASS1_ALL *Analyzed Result for INDEL CLASS1*

Description

Analyzed Result for INDEL CLASS1

Usage

```
data(sample_result_INDEL_CLASS1_ALL)
```

sample_result_INDEL_CLASS2_ALL *Analyzed Result for INDEL CLASS2*

Description

Analyzed Result for INDEL CLASS2

Usage

```
data(sample_result_INDEL_CLASS2_ALL)
```

sample_result_SNV_CLASS1_ALL *Analyzed Result for SNV CLASS1*

Description

Analyzed Result for SNV CLASS1

Usage

```
data(sample_result_SNV_CLASS1_ALL)
```

sample_result_SNV_CLASS2_ALL *Analyzed Result for SNV CLASS2*

Description

Analyzed Result for SNV CLASS2

Usage

```
data(sample_result_SNV_CLASS2_ALL)
```

sample_rna_exp *A Format / Sample file for RNA Expression Information*

Description

A dataset containing the RNA expression amount of patient for each gene.

Usage

```
data(sample_rna_exp)
```

Format

A data frame with 22 rows and 3 variables

sample_sv_bnd *A Format / Sample file for Annotated vcf file.*

Description

A dataset containing the variant information of a patient.

Usage

```
data(sample_sv_bnd)
```

Format

A data frame with 9 rows and variables including "Chr" "Start" "End" "Ref" "Alt (BND format)" "Func.refGene (exonic, intron, intergenic, ...)" "ExonicFunc.refGene (exonic nonsynonymous, synonymous, insertion, ...)" "mateID (e.g., SVMERGE1_1)"

sample_vcf *A Format / Sample file for Annotated vcf file.*

Description

A dataset containing the variant information of a patient.

Usage

```
data(sample_vcf)
```

Format

A data frame with 9 rows and variables including "Chr" "Start" "End" "Ref" "Alt" "Func.refGene (exonic, intron, intergenic, ...)" "ExonicFunc.refGene (exonic nonsynonymous, synonymous, insertion, ...)" "AAChange.refGene (e.g., SLCO1C1:NM_001145944:exon7:c.692_693insG:p.L231fs ...)"

TestAnalysis *Execute Sample Analysis*

Description

Execute Sample Analysis

Usage

```
TestAnalysis()
```

Value

void

Index