Package 'Neoantimon'

October 5, 2017

Type Package

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

Version 1.1

Date 2017-06-09

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 TRUE

Imports utils

RoxygenNote 6.0.1

R topics documented:

CopyNum *Prices of 50,000 round cut diamonds*

Description

A dataset containing the prices and other attributes of almost 54,000 diamonds. The variables are as follows:

Usage

```
data(CopyNum)
```

Format

A data frame with 53940 rows and 10 variables

Details

- price. price in US dollars (\$326–\$18,823)

hla_table *Prices of 50,000 round cut diamonds*

Description

A dataset containing the prices and other attributes of almost 54,000 diamonds. The variables are as follows:

Usage

```
data(hla_table)
```

Format

A data frame with 53940 rows and 10 variables

Details

- price. price in US dollars (\$326–\$18,823)

hla_table2 *Prices of 50,000 round cut diamonds*

Description

A dataset containing the prices and other attributes of almost 54,000 diamonds. The variables are as follows:

Usage

```
data(hla_table2)
```

Format

A data frame with 53940 rows and 10 variables

Details

- price. price in US dollars (\$326–\$18,823)

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,
  hmdir = getwd(), job_ID = "NO_JOB_ID", RNAseq_file = NA, RNA_bam = NA,
  CNV = NA, ccfp_dir = "lib/ccfp.jar", Purity = NA,
  netMHCpan_dir = "lib/netMHCpan-3.0/netMHCpan", refDNA = "lib/GRCh37.fa",
  refFlat_file = paste(hmdir, "/data/refFlat.txt", sep = ""),
  refMrna_1 = paste(hmdir, "/data/refMrna.cut1.fa", sep = ""),
  refMrna_3 = paste(hmdir, "/data/refMrna.cut3.fa", sep = ""),
  Chr_Column = 1, Mutation_Start_Column = 2, Mutation_End_Column = 3,
  Mutation_Ref_Column = 4, Mutation_Alt_Column = 5, NM_ID_Column = 10,
  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	An input file annotated by ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) or other softwares to be analyzed (Required). See by data(sample_annovar.txt); sample_annovar; data(sample_genomon); sample_genomon;
HLA_file	A tab separated file indicating HLA types (Required). The 1st column is input_file name, and the following columns indicate HLA types. See by data(hla_table); hla_table;

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

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

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

`RNAseq_file`
: (Default=NA) A file including RNA expressions. The 1st, 2nd and 3rd columns are "GeneSymbol Chr:ExonStart-ExonEnd(locus) Expression Amount", respectively. The 1st row should be any header. See by data(RNAseq.txt); RNAseq;

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

`CNV`
: A file including copy number variation to calculate cancer cell fraction probability (CCFP) (Default=NA). The format is according to ASCAT (https://www.crick.ac.uk/peter-van-loo/software/ASCAT) output files. The columns are "Chromosome Position Log R segmented LogR BAF segmented BAF Copy number Minor allele Raw copy number" The 1st row should be the above header. See data(CopyNum); head(CopyNum);

`ccfp_dir`
: The file directory to CCFP.pl (Default="lib/ccfp.jar").

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

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

`refDNA`
: refDNA information to be used to calculate RNA VAF (Default="ib/GRCh37.fa").

`refFlat_file`
: refFlat file to be used in constructing peptide. (Default=paste(hmdir,"ib/refFlat.txt",sep="") This file is automaticalluy generated by Main_Preparation.sh

`refMrna_1`
: refMrna file to be used in constructing peptide (Default=paste(hmdir,"ib/refMrna.merge.cut1.fa",sep= This file is automaticalluy generated by Main_Preparation.sh

`refMrna_3`
: refMrna file to be used in constructing peptide (Default=paste(hmdir,"ib/refMrna.merge.cut3.fa",sep= This file is automaticalluy generated by Main_Preparation.sh

`Chr_Column`
: The column number describing Chromosome number in input_file (Default=1).

`Mutation_Start_Column`
: The column number describing Mutation Start Position in input_file (Default=2) .

`Mutation_End_Column`
: The column number describing Mutation End Position in input_file (Default=3).

`Mutation_Ref_Column`
: The column number describing Mutation Ref in input_file (Default=4).

`Mutation_Alt_Column`
: The column number describing Mutation Alt in input_file (Default=5).

`NM_ID_Column`
: The column number describing NM IDs in input_file (Default=10).

`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,
  hmdir = getwd(), job_ID = "NO_JOB_ID", RNAseq_file = NA, RNA_bam = NA,
  CNV = NA, ccfp_dir = "lib/ccfp.jar", Purity = NA,
  netMHCpan_dir = "lib/netMHCIIpan-3.1/netMHCIIpan",
  refDNA = "lib/GRCh37.fa", refFlat_file = paste(hmdir, "/data/refFlat.txt",
  sep = ""), refMrna_1 = paste(hmdir, "/data/refMrna.cut1.fa", sep = ""),
  refMrna_3 = paste(hmdir, "/data/refMrna.cut3.fa", sep = ""),
  Chr_Column = 1, Mutation_Start_Column = 2, Mutation_End_Column = 3,
  Mutation_Ref_Column = 4, Mutation_Alt_Column = 5, NM_ID_Column = 10,
  Depth_Normal_Column = NA, Depth_Tumor_Column = NA,
  ambiguous_between_exon = 0, ambiguous_codon = 0, peptide_length = c(15))
```

Arguments

input_file	An input file annotated by ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) or other softwares to be analyzed (Required). See by data(sample_annovar.txt); sample_annovar; data(sample_genomon); sample_genomon;
HLA_file	A tab separated file indicating HLA types (Required). The 1st column is input_file name, and the following columns indicate HLA types. See by data(hla_table2); hla_table2;
file_name_in_HLA_table	If the name (1st column) in HLA table is not input_file, indicate the corresponding name (Default=input_file).
hmdir	Home directory for the analysis (Default=getwd()).
job_ID	Job-Id to be attached in output files (Default="NO_JOB_ID").
RNAseq_file	(Default=NA) A file including RNA expressions. The 1st, 2nd and 3rd columns are "GeneSymbol Chr:ExonStart-ExonEnd(locus) Expression Amount", respectively. The 1st row should be any header. See by data(RNAseq.txt); RNAseq;
RNA_bam	RNA bam file to calculate variant allele frequency of RNA at each mutation (Default=NA).
CNV	A file including copy number variation to calculate cancer cell fraction probability (CCFP) (Default=NA). The format is according to ASCAT (https://www.crick.ac.uk/peter-van-loo/software/ASCAT) output files. The columns are "Chromosome Position Log R segmented LogR BAF segmented BAF Copy number Minor allele Raw copy number" The 1st row should be the above header. See data(CopyNum); head(CopyNum);

ccfp_dir	The file directory to CCFP.pl (Default="lib/ccfp.jar").
Purity	Tumor purity or tumor contents ratio required to calculate CCFP (Default=NA).
netMHCpan_dir	The file directory to netMHCpan (Default="lib/netMHCpan-3.0/netMHCpan").
refDNA	refDNA information to be used to calculate RNA VAF (Default="lib/GRCh37.fa").
refFlat_file	refFlat file to be used in constructing peptide. (Default=paste(hmdir,"lib/refFlat.txt",sep="") This file is automaticalluy generated by Main_Preparation.sh
refMrna_1	refMrna file to be used in constructing peptide (Default=paste(hmdir,"lib/refMrna.merge.cut1.fa",sep= This file is automaticalluy generated by Main_Preparation.sh
refMrna_3	refMrna file to be used in constructing peptide (Default=paste(hmdir,"lib/refMrna.merge.cut3.fa",sep= This file is automaticalluy generated by Main_Preparation.sh
Chr_Column	The column number describing Chromosome number in input_file (Default=1).
Mutation_Start_Column	The column number describing Mutation Start Position in input_file (Default=2) .
Mutation_End_Column	The column number describing Mutation End Position in input_file (Default=3).
Mutation_Ref_Column	The column number describing Mutation Ref in input_file (Default=4).
Mutation_Alt_Column	The column number describing Mutation Alt in input_file (Default=5).
NM_ID_Column	The column number describing NM IDs in input_file (Default=10).
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.)

`MainMergeClass1` *Merge Results from MainSnvClass1.R*

Description

Merge Results from MainSnvClass1.R

Usage

```
MainMergeClass1(hmdir = getwd(), input_dir, input_file_prefix,
  Tumor_RNA_BASED_ON_DNA = TRUE, INDEL = FALSE)
```

Arguments

hmdir	Home directory for the analysis (Default=getwd()).
input_dir	Directory storing netMHCpan Results (Required).
input_file_prefix	File prefix of netMHCpan Results (Required). If you have "sample_annovar.txt.NO_JOB_ID.HLACL please set "sample_annovar", "sample_annovar.txt" or "sample_annovar.txt.NO_JOB_ID".
Tumor_RNA_BASED_ON_DNA	In calculating tumor specific RNA expression, TRUE uses variant allele frequency on DNA. Otherwise, use VAF on RNA (Default=TRUE).
INDEL	If the targeting results are generated from Indels, Please check TRUE.

Value

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

MainMergeClass2 *Merge Results from MainSnvClass2.R*

Description

Merge Results from MainSnvClass2.R

Usage

```
MainMergeClass2(hmdir = getwd(), input_dir, input_file_prefix,
  Tumor_RNA_BASED_ON_DNA = TRUE, INDEL = FALSE)
```

Arguments

hmdir	Home directory for the analysis (Default=getwd()).
input_dir	Directory storing netMHCpan Results (Required).
input_file_prefix	File prefix of netMHCpan Results (Required). If you have "sample_annovar.txt.NO_JOB_ID.HLACL please set "sample_annovar", "sample_annovar.txt" or "sample_annovar.txt.NO_JOB_ID".
Tumor_RNA_BASED_ON_DNA	In calculating tumor specific RNA expression, TRUE uses variant allele frequency on DNA. Otherwise, use VAF on RNA (Default=TRUE).
INDEL	If the targeting results are generated from Indels, Please check TRUE.

Value

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

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,
  hmdir = getwd(), job_ID = "NO_JOB_ID", RNAseq_file = NA, RNA_bam = NA,
  CNV = NA, ccfp_dir = "lib/ccfp.jar", Purity = NA,
  netMHCpan_dir = "lib/netMHCpan-3.0/netMHCpan", refDNA = "lib/GRCh37.fa",
  refFlat_file = paste(hmdir, "/data/refFlat.txt", sep = ""),
  refMrna_1 = paste(hmdir, "/data/refMrna.cut1.fa", sep = ""),
  refMrna_3 = paste(hmdir, "/data/refMrna.cut3.fa", sep = ""),
  Chr_Column = 1, Mutation_Start_Column = 2, Mutation_End_Column = 3,
  Mutation_Ref_Column = 4, Mutation_Alt_Column = 5, NM_ID_Column = 10,
  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	An input file annotated by ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) or other softwares to be analyzed (Required). See by data(sample_annovar.txt); sample_annovar; data(sample_genomon); sample_genomon;
HLA_file	A tab separated file indicating HLA types (Required). The 1st column is input_file name, and the following columns indicate HLA types. See by data(hla_table); hla_table;
file_name_in_HLA_table	If the name (1st column) in HLA table is not input_file, indicate the corresponding name (Default=input_file).
hmdir	Home directory for the analysis (Default=getwd()).
job_ID	Job-Id to be attached in output files (Default="NO_JOB_ID").
RNAseq_file	(Default=NA) A file including RNA expressions. The 1st, 2nd and 3rd columns are "GeneSymbol Chr:ExonStart-ExonEnd(locus) Expression Amount", respectively. The 1st row should be any header. See by data(RNAseq.txt); RNAseq;
RNA_bam	RNA bam file to calculate variant allele frequency of RNA at each mutation (Default=NA).
CNV	A file including copy number variation to calculate cancer cell fraction probability (CCFP) (Default=NA). The format is according to ASCAT (https://www.crick.ac.uk/peter-van-loo/software/ASCAT) output files. The columns are "Chromosome Position Log R segmented LogR BAF segmented BAF Copy number Minor allele Raw copy number" The 1st row should be the above header. See data(CopyNum); head(CopyNum);
ccfp_dir	The file directory to CCFP.pl (Default="lib/ccfp.jar").
Purity	Tumor purity or tumor contents ratio required to calculate CCFP (Default=NA).

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

`refDNA` refDNA information to be used to calculate RNA VAF (Default="lib/GRCh37.fa").

`refFlat_file` refFlat file to be used in constructing peptide. (Default=paste(hmdir,"lib/refFlat.txt",sep="")
This file is automaticalluy generated by Main_Preparation.sh

`refMrna_1` refMrna file to be used in constructing peptide (Default=paste(hmdir,"lib/refMrna.merge.cut1.fa",sep=
This file is automaticalluy generated by Main_Preparation.sh

`refMrna_3` refMrna file to be used in constructing peptide (Default=paste(hmdir,"lib/refMrna.merge.cut3.fa",sep=
This file is automaticalluy generated by Main_Preparation.sh

`Chr_Column` The column number describing Chromosome number in input_file (Default=1).

`Mutation_Start_Column`
The column number describing Mutation Start Position in input_file (Default=2)
.

`Mutation_End_Column`
The column number describing Mutation End Position in input_file (Default=3).

`Mutation_Ref_Column`
The column number describing Mutation Ref in input_file (Default=4).

`Mutation_Alt_Column`
The column number describing Mutation Alt in input_file (Default=5).

`NM_ID_Column` The column number describing NM IDs in input_file (Default=10).

`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,
  hmdir = getwd(), job_ID = "NO_JOB_ID", RNAseq_file = NA, RNA_bam = NA,
  CNV = NA, ccfp_dir = "lib/ccfp.jar", Purity = NA,
  netMHCpan_dir = "lib/netMHCIIpan-3.1/netMHCIIpan",
  refDNA = "lib/GRCh37.fa", refFlat_file = paste(hmdir, "/data/refFlat.txt",
  sep = ""), refMrna_1 = paste(hmdir, "/data/refMrna.cut1.fa", sep = ""),
  refMrna_3 = paste(hmdir, "/data/refMrna.cut3.fa", sep = ""),
  Chr_Column = 1, Mutation_Start_Column = 2, Mutation_End_Column = 3,
  Mutation_Ref_Column = 4, Mutation_Alt_Column = 5, NM_ID_Column = 10,
  Depth_Normal_Column = NA, Depth_Tumor_Column = NA,
  ambiguous_between_exon = 0, ambiguous_codon = 0, peptide_length = c(15))
```

Arguments

input_file	An input file annotated by ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) or other softwares to be analyzed (Required). See by data(sample_annovar.txt); sample_annovar; data(sample_genomon); sample_genomon;
HLA_file	A tab separated file indicating HLA types (Required). The 1st column is input_file name, and the following columns indicate HLA types. See by data(hla_table2); hla_table2;
file_name_in_HLA_table	If the name (1st column) in HLA table is not input_file, indicate the corresponding name (Default=input_file).
hmdir	Home directory for the analysis (Default=getwd()).
job_ID	Job-Id to be attached in output files (Default="NO_JOB_ID").
RNAseq_file	(Default=NA) A file including RNA expressions. The 1st, 2nd and 3rd columns are "GeneSymbol Chr:ExonStart-ExonEnd(locus) Expression Amount", respectively. The 1st row should be any header. See by data(RNAseq.txt); RNAseq;
RNA_bam	RNA bam file to calculate variant allele frequency of RNA at each mutation (Default=NA).
CNV	A file including copy number variation to calculate cancer cell fraction probability (CCFP) (Default=NA). The format is according to ASCAT (https://www.crick.ac.uk/peter-van-loo/software/ASCAT) output files. The columns are "Chromosome Position Log R segmented LogR BAF segmented BAF Copy number Minor allele Raw copy number" The 1st row should be the above header. See data(CopyNum); head(CopyNum);
ccfp_dir	The file directory to CCFP.pl (Default="lib/ccfp.jar").
Purity	Tumor purity or tumor contents ratio required to calculate CCFP (Default=NA).
netMHCpan_dir	The file directory to netMHCpan (Default="lib/netMHCpan-3.0/netMHCpan").
refDNA	refDNA information to be used to calculate RNA VAF (Default="lib/GRCh37.fa").
refFlat_file	refFlat file to be used in constructing peptide. (Default=paste(hmdir,"lib/refFlat.txt",sep="") This file is automaticalluy generated by Main_Preparation.sh
refMrna_1	refMrna file to be used in constructing peptide (Default=paste(hmdir,"lib/refMrna.merge.cut1.fa",sep= This file is automaticalluy generated by Main_Preparation.sh
refMrna_3	refMrna file to be used in constructing peptide (Default=paste(hmdir,"lib/refMrna.merge.cut3.fa",sep= This file is automaticalluy generated by Main_Preparation.sh
Chr_Column	The column number describing Chromosome number in input_file (Default=1).

`Mutation_Start_Column`
: The column number describing Mutation Start Position in input_file (Default=2).

`Mutation_End_Column`
: The column number describing Mutation End Position in input_file (Default=3).

`Mutation_Ref_Column`
: The column number describing Mutation Ref in input_file (Default=4).

`Mutation_Alt_Column`
: The column number describing Mutation Alt in input_file (Default=5).

`NM_ID_Column`
: The column number describing NM IDs in input_file (Default=10).

`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.)

`Neoantimon` *Neoantimon*

Description

Calculate Lists of Candidate Neoantingens on SNVs and Indels to MHC Class1 and Class2

`RNAseq` *Prices of 50,000 round cut diamonds*

Description

A dataset containing the prices and other attributes of almost 54,000 diamonds. The variables are as follows:

Usage

```
data(RNAseq)
```

Format

A data frame with 53940 rows and 10 variables

Details

- price. price in US dollars (\$326–\$18,823)

`sample` *Prices of 50,000 round cut diamonds*

Description

A dataset containing the prices and other attributes of almost 54,000 diamonds. The variables are as follows:

Usage

```
data(sample)
```

Format

A data frame with 53940 rows and 10 variables

Details

- price. price in US dollars (\$326–\$18,823)

Index