

Package ‘simMP’

April 21, 2017

Type Package

Title Simulate Somatic Mutations in Cancer Genomes from Mutational Processes

Version 0.17.3

Author Nan Zhou

Maintainer Nan Zhou <zhnanx@gmail.com>

Description Simulates somatic single base substitutions carried in cancer genomes. By only providing a human reference genome, substitutions that result from mutational processes operative in every cancer genome can be generated.

License GPL-2

Encoding UTF-8

LazyData true

Imports Biostrings, BSgenome, GenomeInfoDb, GenomicRanges, IRanges, XVector

Depends R (>= 3.0.0), doParallel, utils, foreach, parallel, stats

Suggests BSgenome.Hsapiens.UCSC.hg38

NeedsCompilation no

Repository CRAN

Date/Publication 2017-04-21 06:03:09 UTC

R topics documented:

mutDistriWGS 2
simSBS 2

Index 5

mutDistriWGS *Distribution of single base substitutions*

Description

Distribution of single base substitutions over all currently available WGS genomes in ICGC data realse 23.

Usage

```
data("mutDistriWGS")
```

Format

A data frame with 3543 observations on the following variable.

X0 a numeric vector

Source

Zhou, Nan, et al. "Pan-cancer scale landscape of simple somatic mutations." bioRxiv (2017): 112367.

Examples

```
data(mutDistriWGS)
head(mutDistriWGS)

## Not run:
plot(1:nrow(mutDistriWGS), sort(c(t(mutDistriWGS))), decreasing = TRUE))

## End(Not run)
```

simSBS *Simulate single base substitutions*

Description

Given the number of genomes to be created, generate single base substitutions in those genomes from simulated mutational processes, by referring to a human reference genome.

Usage

```
simSBS(nSigs = NULL, nGenomes = NULL, refGenome = NULL,
       similarity = 0.6, noise = 0,
       presetSigs = NULL, chrs = NULL, nMutPerGenome = NULL,
       sigPrevalence = NULL, chrDistribution = NULL,
       parallel = TRUE, saveDir = './')
```

Arguments

nSigs	Required. The number of mutational processes to be created.
nGenomes	Required. The number of genomes in which to simulate single base substitutions.
refGenome	Required. A BSgenome object of human reference genome.
similarity	Optional. Limit the similarity between any two mutational processes. 0 indicates no similarity while 1 indicates the opposite. Lower similarity may require more time to simulate.
noise	Optional. The value should be between 0 and 1, indicating the amount of random mutations (noise) added to each simulated genome. 0 indicates no noise while 1 indicates the amount of noise is equal to the amount of mutation.
presetSigs	Optional. Use user defined mutational processes to simulate mutations in the genome. It should be a 96-by-n matrix, where 96 denotes the number of mutation motifs while n denotes the number of mutational processes. If presetSigs is given, nSigs = n.
chrs	Optional. On which chromosome(s) mutations simulated from. Default is c(1:22, 'X', 'Y'). This argument accepts a vector that indicates chromosomes, which should be a vector created by manual input or, for example, using R code like c(1:22, 'X', 'Y', 'M'), where 'X', 'Y', 'M' are case sensitive (upper case) and indicate chromosome X, Y and mitochondrial chromosome. Incompatible input could cause fatal errors cause of unidentifiable chromosome name.
nMutPerGenome	Optional. <i>NULL</i> or a numerical vector whose length equals nGenomes. Number of mutations on each genome to simulate. If not defined, Default will use the distribution of number of single base substitutions in all WGS projects of ICGC release 23.
sigPrevalence	Optional. Acceptable values are either <i>NULL</i> or a <i>numerical vector</i> . The prevalence of mutational processes in wild. The default uses known prevalences of 21 processes from Alexandrov et al. 's work.
chrDistribution	Optional. <i>NULL</i> or a <i>numerical vector</i> are acceptable. The percentage of mutations assigned to each chromosome in a genome. The default uses the distribution of length of chromosomes (chr1 to chr22 and chrX and chrY). If a numerical vector was given, its length should equal the length of <i>chrs</i> and values should sum up to 1.
parallel	Optional. TRUE or FALSE. Whether enable or disable parallel computing ability.
saveDir	Optional. The directory where to save simulation output. Default is the current working directory. Other paths should also be relative to the current working directory.

Value

If succeed, the return value is 1. Simulation results are saved in saveDir.

Examples

```
if(require(BSgenome.Hsapiens.UCSC.hg38)){
  simSBS(nSigs = 2, nGenomes = 2,
        refGenome = BSgenome.Hsapiens.UCSC.hg38::BSgenome.Hsapiens.UCSC.hg38,
        nMutPerGenome = sample(10:50, 2),
        parallel = FALSE)
}else{
  message('Cannot proceed without a valid reference genome.')
}
```

Index

*Topic **datasets**

mutDistriWGS, [2](#)

*Topic **simulate**

simSBS, [2](#)

BSgenome, [3](#)

mutDistriWGS, [2](#)

simSBS, [2](#)