Package ‘EthSEQ’

January 10, 2023

Type Package

Title Ethnicity Annotation from Whole-Exome and Targeted Sequencing Data

Version 3.0.1

Description Reliable and rapid ethnicity annotation from whole exome and targeted sequencing data.

License GPL-3

Depends R (>= 2.15)

Imports graphics, utils, parallel, grDevices, MASS (>= 7.3-47), geometry (>= 0.3-6), data.table (>= 1.10.0), SNPRelate (>= 1.8.0), gdsfmt (>= 1.10.1), plot3D (>= 1.1), Rcpp (>= 0.11.0)

LinkingTo Rcpp

RoxygenNote 7.2.1

biocViews

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation yes

Author Alessandro Romanel [aut, cre],
Davide Dalfovo [aut]

Maintainer Alessandro Romanel <`alessandro.romanel@unitn.it`>

Repository CRAN

Date/Publication 2023-01-10 16:43:11 UTC

R topics documented:

ethseq.Analysis 2
ethseq.RM 3
getModelsList 4
getSamplesInfo 4

Index 5

`ethseq.Analysis` *Ancestry analysis from whole-exome and targeted sequencing data*

Description

This function performs ancestry analysis of a set of samples ad reports the results.

Usage

```
ethseq.Analysis(
  target.vcf = NA,
  target.gds = NA,
  bam.list = NA,
  out.dir = tempdir(),
  model.gds = NA,
  model.available = NA,
  model.assembly = "hg38",
  model.pop = "All",
  model.folder = tempdir(),
  run.genotype = FALSE,
  aseq.path = tempdir(),
  mbq = 20,
  mrq = 20,
  mdc = 10,
  cores = 1,
  verbose = TRUE,
  composite.model.call.rate = 1,
  refinement.analysis = NA,
  space = "2D",
  bam.chr.encoding = FALSE
)
```

Arguments

`target.vcf`	Path to the sample's genotypes in VCF format
`target.gds`	Path to the sample's genotypes in GDS format
`bam.list`	Path to a file containing a list of BAM files paths
`out.dir`	Path to the folder where the output of the analysis is saved
`model.gds`	Path to a GDS file specifying the reference model
`model.available`	String specifying the pre-computed reference model to use
`model.assembly`	String value indicating the assembly version to download for the pre-build models
`model.pop`	String value indicating the population to download for the pre-build models

`model.folder`	Path to the folder where reference models are already present or downloaded when needed
`run.genotype`	Logical values indicating whether the ASEQ genotype should be run
`aseq.path`	Path to the folder where ASEQ binary is available or is downloaded when needed
`mbq`	Minmum base quality used in the pileup by ASEQ
`mrq`	Minimum read quality used in the piluep by ASEQ
`mdc`	Minimum read count acceptable for genotype inference by ASEQ
`cores`	Number of parallel cores used for the analysis
`verbose`	Print detailed information
`composite.model.call.rate`	SNP call rate used to run Principal Component Analysis (PCA)
`refinement.analysis`	Matrix specifying a tree of ancestry sets
`space`	Dimensions of PCA space used to infer ancestry (2D or 3D)
`bam.chr.encoding`	Logical value indicating whether input BAM files have chromosomes encoded with "chr" prefix

Value

Logical value indicating the success of the analysis

`ethseq.RM` *Create Reference Model for Ancestry Analysis*

Description

This function creates a GDS reference model that can be used to performe EthSEQ ancestry analysis

Usage

```
ethseq.RM(
  vcf.fn,
  annotations,
  out.dir = "./",
  model.name = "Reference.Model",
  bed.fn = NA,
  verbose = TRUE,
  call.rate = 1,
  cores = 1
)
```

Arguments

`vcf.fn`	vector of paths to genotype files in VCF format
`annotations`	data.frame with mapping of all samples names, ancestries and gender
`out.dir`	Path to output folder
`model.name`	Name of the output model
`bed.fn`	path to BED file with regions of interest
`verbose`	Print detailed information
`call.rate`	SNPs call rate cutoff for inclusion in the final reference model
`cores`	How many parallel cores to use in the reference model generation

Value

Logical value indicating the success of the analysis

`getModelsList` *List the models available*

Description

This function prints the list of all available models.

Usage

```
getModelsList()
```

Value

data.frame of all available models to use with specified assembly and population

---

`getSamplesInfo` *List the samples annotation*

Description

This function prints the list of all 1,000 Genomes Project samples used to build the reference models.

Usage

```
getSamplesInfo()
```

Index

`ethseq.Analysis`, 2
`ethseq.RM`, 3

`getModelsList`, 4
`getSamplesInfo`, 4