Package 'EthSEQ'

February 18, 2021

Type Package

Title Ethnicity Annotation from Whole Exome Sequencing Data

Version 2.1.4

Description Reliable and rapid ethnicity annotation from whole exome 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)

RoxygenNote 6.0.1

biocViews

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

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

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

Repository CRAN

Date/Publication 2021-02-18 22:30:05 UTC

R topics documented:

---

ethseq.Analysis

Ethnicity analysis from whole-exome sequencing data

Description

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

Usage

```
ethseq.Analysis(target.vcf = NA, target.gds = NA, bam.list = NA,
  out.dir = "/tmp", model.gds = NA, model.available = NA,
  model.folder = "/tmp", run.genotype = FALSE, aseq.path = "/tmp",
  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 (SS2,SS4,NimblegenV3,HALO,Exonic)
model.folder	Path to the folder where reference models are already present of downloaded when needed
run.genotype	Logical values indicating wheter the ASEQ genotype should be run
aseq.path	Path to the folder where ASEQ binary is available of 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 accettable for genotype inference by ASEQ
cores	Number of parallel cores used for the analysis
verbose	Print detaild information
composite.model.call.rate	
	SNP call rate used to run Principal Component Analysis (PCA)
refinement.analysis	
	Matrix specyfing a tree of ethnicities
space	Dimensions of PCA space used to infer ethnicity (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 Ethnicity Analysis*

Description

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

Usage

```
ethseq.RM(vcf.fn, annotations, out.dir = "./",
  model.name = "Reference.Model", bed.fn = NA, 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, ethnicities 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
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

Index