Package 'eSCAN'

June 21, 2020

Type Package

Title Scan the Enhancers: Scan Regulatory Regions for Rare Variant Aggregate Association Testing using Whole Genome Sequencing Data

Version 0.1.1

Date 2020-06-21

Author Yingxi Yang

Maintainer Yingxi Yang <yangyx117@gmail.com>

Description eSCAN (or "Scan the Enhancers", with "enhancers" as a shorthand for any potential regulatory regions in the genome) is an R package for performing genome-wide assessment of rare variant enhancer regions in sequencing studies, combining the advantages of dynamic window selection with the advantages of increasing genomic annotation information, including chromatin accessibility, histone markers, and 3D chromatin conformation.

License GPL-3

Encoding UTF-8

Imports Rcpp (>= 1.0.4.6), Matrix, GENESIS

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 7.1.0

R topics documented:

eGenerator 1
eSCAN 2
fitNull 3
preprocess 4

Index **5**

`eGenerator` *Generate enhancer locations*

Description

The `eGenerator` function generates regulatory rgeion locations if not specified by users.

Usage

```
eGenerator(geno, maxgap = 10^4)
```

Arguments

geno	an n*p genotype matrix, where n is the sample size and p is the number of rare varints included.
maxgap	threshold to split independent loci (default=10^4)

Value

The function returns a data frame containing generated enhancer location. The first column is serial number of the enhancer. The next two columns are physical location of the enhancer. The nest two columns are adjusted location of the enhancer (the position in the geno matrix). The last column is enhancer length.

---

eSCAN *Scan the enhancers*

Description

The eSCAN function is the main function in the package. It takes in genotype matrix, null model object and annotation information which could be specified by users, and then detect rare variant association between a quantitative/dichotomous phenotype and a regulatory region in a sequence by using eSCAN procedure.

Usage

```
eSCAN(
  genotype,
  nullmod,
  new_enloc = NULL,
  gap = 10^4,
  times = 1000,
  alpha = 0.05,
  analy = FALSE
)
```

Arguments

genotype	an n*p genotype matrix, where n is the sample size and p is the number of rare varints included.
nullmod	a null model object returned from fitNull. Note that fitNull is a wrapper of fitNullModel function from the GENESIS package.
new_enloc	a data frame of annotation information with dimension q*6, where q is the number of the target regulatory regions. The six columns indicates no., physical start position, physical end position, start position in the genotype matrix, end position in the genotype matrix and region length, respectively. If annotation information is not specified by users (default=NULL), a data frame of enhancer location will be automatically created by eGenerator.

gap	if new_enloc is not specified by users, this parameter will be used to generate enhancer locations in functoin eGenerator, where gap is the threshold to split independent loci (default=10^4)
times	simulation times for MC (default=1000)
alpha	significance level (default=0.05)
analy	TRUE indicates analytical threshold, FALSE indicates empirical threshold (default=FALSE)

Value

The function returns a list containing the following elements:

res: A matrix of significant regions detected by eSCAN. The first column is the p-value of the detected region. The next two columns are the location of the detected rgeion (physical position on chromosome).

res0: A matrix to summarise all the regions included in the analysis. The first column is the p-value of the detected region. The next two columns are the location of the detected rgeion (physical position on chromosome).

thres: threshold of eSCAN to control the family-wise/genome-wide error at alpha level.

---

fitNull *Fit generalized linear model under the null hypothesis for unrelated samples.*

Description

The fitNull function is a wrapper of fitNullModel from the GENESIS package. It fits a regression model under the null hypothesis for unrelated samples which is the preliminary step for subsequent analysis.

Usage

```
fitNull(x, outcome = NULL, covars = NULL, fam = "gaussian")
```

Arguments

x	a data frame containing outcome variable and covariates.
outcome	a character string specifying the name of outcome variable in x.
covars	a vector of character strings specifying the name of covariates in x.
fam	Can be either "gaussian" for continuous phenotype or "binomial" for binary phenotype.

Value

The function returns an object of model fit from fitNullModel. See fitNullModel in the GENESIS package for more details.

References

Gogarten, S.M., Sofer, T., Chen, H., Yu, C., Brody, J.A., Thornton, T.A., Rice, K.M., and Conomos, M.P. (2019). Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics.

preprocess *Data preprocessing*

Description

This function is a preliminary step for subsequent analysis using eSCAN.

Usage

```
preprocess(geno, enhancer, gap)
```

Arguments

geno	an n*p genotype matrix, where n is the sample size and p is the number of rare varints included.
enhancer	a data frame of annotation information with dimension q*6, where q is the number of the target regulatory regions. The six columns indicates no., physical start position, physical end position, start position in the geno matrix, end position in the geno matrix and region length, respectively. If annotation information is not specified by users (default=NULL), a data frame of enhancer location will be automatically created by eGenerator function.
gap	threshold to split independent loci (default=10^4)

Value

The functoin returns a list with the following elements: genotype: An n*p genotype matrix by increasing order of variant position. MAF: A vector (length=p) of the minor allele frequency. new_enloc: A data frame of enhancer location of dimension q*6, where q is the number of the target regulatory regions. The six columns indicates No., physical start position, physical end position, start position in the genotype matrix, end position in the genotype matrix and region length, respectively.

Index

eGenerator, 1, *2–4*
eSCAN, 2

fitNull, *2*, 3
fitNullModel, *2, 3*

GENESIS, *2, 3*

preprocess, 4