

Package ‘SIPI’

September 23, 2016

Type Package

Title SIPI

Version 0.2

Date 2016-9-22

Depends logregperm, SNPassoc, sm, car, lmtest, parallel

Author Hui-Yi Lin and Po-Yu Huang

Maintainer Po-Yu Huang<jettly32@gmail.com>

Description SNP Interaction Pattern Identifier (SIPI)

License GPL-2

LazyData true

R topics documented:

Grid3by3	1
MAFinfo	2
parSIPI	3
simData	4
SIPI	5
Index	7

Grid3by3

Grid3by3

Description

Outcome prevalence by the 3-by-3 genotypes of a given SNP pair

Usage

Grid3by3(Outcome, SNPdata, PairInfo)

Arguments

Outcome	Binary outcome variable name: a binary variable with "1" as the event of interest and "0" as the reference.
SNPdata	SNP data
PairInfo	c('SNP1','SNP2'): names of the given SNP pairs for one-pair analyses

Value

Returns a cross-table of the outcome prevalence by the genotype combinations of the given SNP pair

Author(s)

Hui-Yi Lin and Po-Yu Huang

Examples

```
##load data
data(simData)

#### define SNP data
SNPdata = simData[,3:12]

#####
### run Grid3by3
#####
Grid3by3(simData$D, SNPdata, c('SNP1', 'SNP2'))
```

MAFinfo

MAFinfo

Description

Obtain major and minor allele and minor allele frequency (MAF) for testing SNPs

Usage

```
MAFinfo(SNPdata)
```

Arguments

SNPdata	SNP data
---------	----------

Value

maj/min	major and minor allele
MAF	minor allele frequency
Missing (%)	missing vales

Author(s)

Hui-Yi Lin and Po-Yu Huang

References

JR Gonzalez, L Armengol, X Sole, E Guino, JM Mercader, X Estivill, V Moreno. SNPassoc: an R package to perform whole genome association studies. *Bioinformatics*, 2007;23(5):654-5.

Examples

```
data(simData)
SNPdata = simData[,3:12]
MAFinfo(SNPdata)
```

parSIPI

Parallel computing for SIPI

Description

parSIPI is a parallel computing version of SIPI. This function can decrease computing time, which is useful for large-scale data.

Usage

```
parSIPI(Outcome, SNPdata, X, TestType)
```

Arguments

Outcome	Binary outcome variable name: a binary variable with "1" as the event of interest and "0" as the reference.
SNPdata	SNP data: All SNP variables should have a character variable attribute and contain two of four letters (C, T, A, and G). No other letters or numbers should be used. For missing SNP values, keep the field blank.
X	Covariate(s) to be adjusted in the model, NULL=without covariate.
TestType	Specify the statistical test type: "WaldTest" = the Wald test; "LRT" = the likelihood ratio test.

Value

Returns a data frame of the results of the best model with the lowest BIC value among the 45 models.

Author(s)

Hui-Yi Lin and Po-Yu Huang

References

HY Lin, DT Chen, PY Huang et al. SNP Interaction Pattern Identifier (SIPI): An Intensive Search for SNP-SNP Interaction Patterns. (under review).

Examples

```
##load data
data(simData)

#### define SNP data
SNPdata = simData[,3:12]

#####
### run parSIPI pairwise analyses <Wald test>
#####

res_all = parSIPI(simData$D,SNPdata,NULL,"WaldTest")

#####
###run parSIPI pairwise analyses <Wald test> adjusted for covariates using SIPI with parallel computing
###[age(numeric), gender (binary), and group (categorical)]
#####

X1 = simData[,c('age','gender','group')]

## define categorical variables
X1$gender = factor(X1$gender)
X1$group = factor(X1$group)

res_all_X = parSIPI(simData$D,SNPdata,X,"WaldTest")
```

simData

An example data set

Description

An example dataset with one binary outcome variable (D), the 10 SNPs, and three covariates [age (numeric), gender (binary), and group (categorical)].

Usage

```
data(simData)
```

Format

A data frame with 1000 observations on the following 12 variables.

id A numeric vector for identification

D A numeric vector for a binary outcome with "1" as the event of interest and "0" as the reference

SNP1 to SNP10 A factor with 3 levels for a SNP contain two of four letters (C, T, A, and G), such as CC, TC and TT.

age numeric age.

gender 0: female, 1: male.

group 1: group 1; 2: group 2, 3: group 3.

Examples

```
data(simData)
```

SIPI

SNP Interaction Pattern Identifier (SIPI)

Description

SIPI is used to evaluate pairwise SNP-SNP interactions associated with a binary outcome. For each SNP pair, the SIPI evaluates 45 logistic interaction models which take inheritance modes [additive, dominant, and recessive mode with the original (based on the minor allele) and reverse coding], and risk category grouping (model structure: hierarchical and non-hierarchical models) into consideration. The best interaction pattern is the one with the lowest value of the Bayesian information criterion (BIC). The significant test of the interaction is based on the Wald test (or the likelihood ratio test) of the interaction term in a model. There are two sub-functions of this SIPI macro: (1) one-pair only and (2) pairwise interaction analyses. The details of the 45 models/patterns are listed in the SIPI paper.

Usage

```
SIPI(Outcome, SNPdata, PairInfo, X, TestType)
```

Arguments

Outcome	Binary outcome variable name: a binary variable with "1" as the event of interest and "0" as the reference.
SNPdata	SNP data: All SNP variables should have a character variable attribute and contain two of four letters (C, T, A, and G). No other letters or numbers should be used. For missing SNP values, keep the field blank.
PairInfo	c('SNP1','SNP2'): names of the given SNP pair for one-pair analyses "all": for pairwise analyses.
X	Covariate(s) to be adjusted in the model, NULL=without covariate.
TestType	Specify the statistical test type: "WaldTest" = the Wald test; "LRT" = the likelihood ratio test.

Value

Returns a list with the following attributes:

selectedModel	The results of the best model with the lowest BIC value among the 45 models.
res45Models	For one-pair analyses only: Detailed results with all 45 models sorted by BIC (lowest first). Output variables: Var1: SNP 1; Var2: SNP 2 (the pattern/model labels are based on this order); Model: interaction model/pattern; Wald_Chisq: the Wald chi-square value of the interaction term; Wald_p: the Wald p-value of the interaction term; LRT_Chisq: the chi-square value of likelihood ratio test (LRT) for the interaction term LRT_p: the LRT p-value of the interaction term; BIC: the Bayesian information criterion. The model with the lowest BIC is preferred.

Author(s)

Hui-Yi Lin and Po-Yu Huang

References

HY Lin, DT Chen, PY Huang et al. SNP Interaction Pattern Identifier (SIPI): An Intensive Search for SNP-SNP Interaction Patterns. (under review).

Examples

```
##load data
data(simData)

#### define SNP data
SNPdata = simData[,3:12]

#####
### run SIPI one-pair analyses <Wald test>
#####

## For a SNP pair of SNP2 and SNP8
res_snp_2v8 = SIPI(simData$D,SNPdata,c('SNP2','SNP8'),NULL,"WaldTest")

## print out the best model
res_snp_2v8$selectedModel

## list of 45 models for one SNP pair
res_snp_2v8$res45Models

#####
### run SIPI pairwise analyses <Wald test>
#####

res_all = SIPI(simData$D,SNPdata,"all",NULL,"WaldTest")

#####
### run SIPI pairwise analyses <Wald test> adjusted for covariates
### [age(numeric), gender (binary), and group (categorical)]
#####

X1 = simData[,c('age','gender','group')]

## define categorical variables
X1$gender = factor(X1$gender)
X1$group = factor(X1$group)

res_all_cov = SIPI(simData$D,SNPdata,"all",X1,"WaldTest")
```

Index

*Topic **datasets**
simData, [4](#)

Grid3by3, [1](#)

MAFinfo, [2](#)

parSIPI, [3](#)

simData, [4](#)
SIPI, [5](#)