

Package ‘SIPI’

December 18, 2017

Type Package

Title SIPI

Version 0.51

Date 2017-12-18

Depends SNPassoc, car, lmtest, parallel, pscl

Author Hui-Yi Lin and Po-Yu Huang

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

Description SNP Interaction Pattern Identifier (SIPI) evaluates SNP-SNP interactions associated with a binary or continuous outcome. The primary strengths of SIPI are (1) taking non-hierarchical models, reverse coding and inheritance modes (dominant, recessive and additive mode) into consideration and (2) using BIC to search for a best interaction pattern. For each SNP pair, the SIPI evaluates 45 interaction models. The best interaction pattern is the one with the lowest value of the Bayesian information criterion (BIC).

Reference Lin HY, Chen DT, Huang PY, Liu YH, Ochoa A, Zabaleta J, Mercante DE, Fang Z, Sellers TA, Pow-Sang JM, Cheng CH, Eeles R, Easton D, Kote-Jarai Z, Amin Al Olama A, Benlloch S, Muir K, Giles GG, Wiklund F, Gronberg H, Haiman CA, Schleutker J, Nordestgaard BG, Travis RC, Hamdy F, Pashayan N, Khaw KT, Stanford JL, Blot WJ, Thibodeau SN, Maier C, Kibel AS, Cybulski C, Cannon-Albright L, Brenner H, Kaneva R, Batra J, Teixeira MR, Pandha H, Lu YJ, Consortium P, Park JY. SNP interaction pattern identifier (SIPI): an intensive search for SNP-SNP interaction patterns. *Bioinformatics*. 2016. PubMed PMID: 28039167.

LazyData true

License GPL-2

R topics documented:

AA9int	2
Grid3by3	4
MAFinfo	5
parAA9int	6
parSIPI	7
simData	9
SIPI	9
Index	12

AA9int	AA9int
--------	--------

Description

AA9int, a mini version of SIPI, evaluates SNP-SNP interactions associated with a binary or continuous outcome. For each SNP pair, the SIPI_mini evaluates 9 interaction models (AA set). AA9int treats a SNP as an additive mode and takes reverse coding and non-hierarchical models into consideration. The best interaction pattern is the one with the lowest value of the Bayesian information criterion (BIC). The details of the 9 models/patterns are listed in the SIPI published paper.

Usage

```
AA9int(Outcome, SNPdata, PairInfo,
       X=NULL, categXNames=NULL,
       TestType="WaldTest", ModelType="binomial")
```

Arguments

Outcome	Binary (1: event of interest; 0: reference) or continuous variable.
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. Invalid character or field blank is considered to be missing values.
PairInfo	3 types of PairInfo: (1) 2d-vector: names of the given SNP pair for one-pair analyses. ex: c('SNP1','SNP2') (2) 2d-matrix or 2d-dataframe: names for candidate SNP pair. (3) "all": for pairwise analyses.
X	Covariate(s) to be adjusted in the model (for missing values, keep the field blank), NULL=without covariate. Default is NULL.
categXNames	The variable names of categorical variables, NULL=without categorical covariate. Default is NULL.
TestType	Specify the statistical test type: "WaldTest" = the Wald test; "LRT" = the likelihood ratio test. Default is "WaldTest".
ModelType	Model type: "binomial"=logistic regression; "gaussian"=linear regression. Default is "binomial".

Value

Returns a list with the following attributes:

selectedModel	The results of the best model with the lowest BIC value among the 9 models.
res9Models	For one-pair analyses only: Detailed results with all 9 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

Lin HY, Chen DT, Huang PY, Liu YH, Ochoa A, Zabaleta J, Mercante DE, Fang Z, Sellers TA, Pow-Sang JM, Cheng CH, Eeles R, Easton D, Kote-Jarai Z, Amin Al Olama A, Benlloch S, Muir K, Giles GG, Wiklund F, Gronberg H, Haiman CA, Schleutker J, Nordestgaard BG, Travis RC, Hamdy F, Pashayan N, Khaw KT, Stanford JL, Blot WJ, Thibodeau SN, Maier C, Kibel AS, Cybulski C, Cannon-Albright L, Brenner H, Kaneva R, Batra J, Teixeira MR, Pandha H, Lu YJ, Consortium P, Park JY. SNP interaction pattern identifier (SIPI): an intensive search for SNP-SNP interaction patterns. *Bioinformatics*. 2016. PubMed PMID: 28039167.

Examples

```
##load data
data(simData)

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

#####
### run AA9int one-pair analyses <Wald test>
#####

## For a SNP pair of SNP2 and SNP8
res_snp_2v8 = AA9int(simData$D,SNPdata,c('SNP2','SNP8'))

## print out the best model
res_snp_2v8$selectedModel

## list of 9 models for one SNP pair
res_snp_2v8$res9Models

#####
### run AA9int for a list of multiple SNP pairs <Wald test>
#####

## For 5 SNP pairs(1 v.s. 6, 2 v.s. 7, 3 v.s. 8, 4 v.s. 9, and 5 v.s. 10)
pairMatrix = c("SNP1","SNP2", "SNP3", "SNP4","SNP5","SNP6","SNP7","SNP8","SNP9","SNP10")

pairMatrix = matrix(pairMatrix,5)

pairMatrix = as.data.frame(pairMatrix)

res_snp = AA9int(simData$D,SNPdata,pairMatrix)

#####
### run AA9int pairwise analyses <Wald test>
#####

res_all = AA9int(simData$D,SNPdata,"all")
```

```
#####
### run AA9int pairwise analyses <Wald test> adjusted for covariates
### [age(numeric), gender (binary), and group (categorical)]
#####

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

res_all_cov = AA9int(simData$D,SNPdata,"all",X1,c("gender","group"))
```

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. 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. Invalid character or field blank is considered to be missing values.
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, and the information of major/minor allele (maj: major allele; min: minor allele).

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	<i>MAFinfo</i>
---------	----------------

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)
```

parAA9int	<i>Parallel computing for AA9int</i>
-----------	--------------------------------------

Description

parAA9int is a parallel computing version of AA9int. This function can decrease computing time, which is useful for large-scale data.

Usage

```
parAA9int(Outcome, SNPdata, PairInfo,
          X=NULL, categXNames=NULL,
          TestType="WaldTest", ModelType="binomial", core_ratio=0.9)
```

Arguments

Outcome	Binary (1: event of interest; 0: reference) or continuous variable.
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. No other letters or numbers should be used. Invalid character or field blank is considered to be missing values.
PairInfo	3 types of PairInfo: (1) 2d-vector: names of the given SNP pair for one-pair analyses. ex: c('SNP1','SNP2') (2) 2d-matrix or 2d-dataframe: names for candidate SNP pair. (3) "all": for pairwise analyses.
X	Covariate(s) to be adjusted in the model (for missing values, keep the field blank), NULL=without covariate. Default is NULL.
categXNames	The variable names of categorical variables, NULL=without categorical covariate. Default is NULL.
TestType	Specify the statistical test type: "WaldTest" = the Wald test; "LRT" = the likelihood ratio test. Default is "WaldTest".
ModelType	Model type: "binomial"=logistic regression; "gaussian"=linear regression. Default is "binomial".
core_ratio	The ratio of total cores for parallel computing. Default is 0.9.

Value

Returns a data frame of the results of the best model with the lowest BIC value among the 9 models.

Author(s)

Hui-Yi Lin and Po-Yu Huang

References

Lin HY, Chen DT, Huang PY, Liu YH, Ochoa A, Zabaleta J, Mercante DE, Fang Z, Sellers TA, Pow-Sang JM, Cheng CH, Eeles R, Easton D, Kote-Jarai Z, Amin Al Olama A, Benlloch S, Muir K, Giles GG, Wiklund F, Gronberg H, Haiman CA, Schleutker J, Nordestgaard BG, Travis RC, Hamdy F, Pashayan N, Khaw KT, Stanford JL, Blot WJ, Thibodeau SN, Maier C, Kibel AS, Cybulski C, Cannon-Albright L, Brenner H, Kaneva R, Batra J, Teixeira MR, Pandha H, Lu YJ, Consortium P, Park JY. SNP interaction pattern identifier (SIPI): an intensive search for SNP-SNP interaction patterns. *Bioinformatics*. 2016. PubMed PMID: 28039167.

Examples

```
##load data
data(simData)

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

#####
### run parAA9int pairwise analyses <Wald test>
#####

res_all = parAA9int(simData$D,SNPdata,"all")

#####
###run parAA9int 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')]
res_all_X = parAA9int(simData$D,SNPdata,"all",X1,c("gender","group"))
```

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, PairInfo,
        X=NULL, categXNames=NULL,
        TestType="WaldTest", ModelType="binomial", core_ratio=0.9)
```

Arguments

Outcome	Binary (1: event of interest; 0: reference) or continuous variable.
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. Invalid character or field blank is considered to be missing values.

PairInfo	3 types of PairInfo: (1) 2d-vector: names of the given SNP pair for one-pair analyses. ex: c('SNP1','SNP2') (2) 2d-matrix or 2d-dataframe: names for candidate SNP pair. (3) "all": for pairwise analyses.
X	Covariate(s) to be adjusted in the model (for missing values, keep the field blank), NULL=without covariate. Default is NULL.
categXNames	The variable names of categorical variables, NULL=without categorical covariate. Default is NULL.
TestType	Specify the statistical test type: "WaldTest" = the Wald test; "LRT" = the likelihood ratio test. Default is "WaldTest".
ModelType	Model type: "binomial"=logistic regression; "gaussian"=linear regression. Default is "binomial".
core_ratio	The ratio of total cores for parallel computing. Default is 0.9.

Value

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

Author(s)

Hui-Yi Lin and Po-Yu Huang

References

Lin HY, Chen DT, Huang PY, Liu YH, Ochoa A, Zabaleta J, Mercante DE, Fang Z, Sellers TA, Pow-Sang JM, Cheng CH, Eeles R, Easton D, Kote-Jarai Z, Amin Al Olama A, Benlloch S, Muir K, Giles GG, Wiklund F, Gronberg H, Haiman CA, Schleutker J, Nordestgaard BG, Travis RC, Hamdy F, Pashayan N, Khaw KT, Stanford JL, Blot WJ, Thibodeau SN, Maier C, Kibel AS, Cybulski C, Cannon-Albright L, Brenner H, Kaneva R, Batra J, Teixeira MR, Pandha H, Lu YJ, Consortium P, Park JY. SNP interaction pattern identifier (SIPI): an intensive search for SNP-SNP interaction patterns. *Bioinformatics*. 2016. PubMed PMID: 28039167.

Examples

```
##load data
data(simData)

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

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

res_all = parSIPI(simData$D,SNPdata,"all")

#####
###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')]
res_all_X = parSIPI(simData$D, SNPdata, "all", X1, c("gender", "group"))
```

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

SNP Interaction Pattern Identifier (SIPI) evaluates SNP-SNP interactions associated with a binary or continuous outcome. The primary strengths of SIPI are (1) taking non-hierarchical models, reverse coding and inheritance modes (dominant, recessive and additive mode) into consideration and (2) using BIC to search for a best interaction pattern. For each SNP pair, the SIPI evaluates 45 interaction models. The best interaction pattern is the one with the lowest value of the Bayesian information criterion (BIC). The details of the 45 models/patterns are listed in the SIPI published paper.

Usage

```
SIPI(Outcome, SNPdata, PairInfo,
     X=NULL, categXNames=NULL,
     TestType="WaldTest", ModelType="binomial")
```

Arguments

Outcome	Binary (1: event of interest; 0: reference) or continuous variable.
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. Invalid character or field blank is considered to be missing values.
PairInfo	3 types of PairInfo: (1) 2d-vector: names of the given SNP pair for one-pair analyses. ex: c('SNP1','SNP2') (2) 2d-matrix or 2d-dataframe: names for candidate SNP pair. (3) "all": for pairwise analyses.
X	Covariate(s) to be adjusted in the model (for missing values, keep the field blank), NULL=without covariate. Default is NULL.
categXNames	The variable names of categorical variables, NULL=without categorical covariate. Default is NULL.
TestType	Specify the statistical test type: "WaldTest" = the Wald test; "LRT" = the likelihood ratio test. Default is "WaldTest".
ModelType	Model type: "binomial"=logistic regression; "gaussian"=linear regression. Default is "binomial".

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

Lin HY, Chen DT, Huang PY, Liu YH, Ochoa A, Zabaleta J, Mercante DE, Fang Z, Sellers TA, Pow-Sang JM, Cheng CH, Eeles R, Easton D, Kote-Jarai Z, Amin Al Olama A, Benlloch S, Muir K, Giles GG, Wiklund F, Gronberg H, Haiman CA, Schleutker J, Nordestgaard BG, Travis RC, Hamdy F, Pashayan N, Khaw KT, Stanford JL, Blot WJ, Thibodeau SN, Maier C, Kibel AS, Cybulski C, Cannon-Albright L, Brenner H, Kaneva R, Batra J, Teixeira MR, Pandha H, Lu YJ, Consortium P, Park JY. SNP interaction pattern identifier (SIPI): an intensive search for SNP-SNP interaction patterns. *Bioinformatics*. 2016. PubMed PMID: 28039167.

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'))

## print out the best model
res_snp_2v8$selectedModel

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

#####
### run SIPI for a list of multiple SNP pairs <Wald test>
#####

## For 5 SNP pairs(1 v.s. 6, 2 v.s. 7, 3 v.s. 8, 4 v.s. 9, and 5 v.s. 10)
pairMatrix = c("SNP1","SNP2", "SNP3", "SNP4","SNP5","SNP6","SNP7","SNP8","SNP9","SNP10")

pairMatrix = matrix(pairMatrix,5)

pairMatrix = as.data.frame(pairMatrix)

res_snp = SIPI(simData$D,SNPdata,pairMatrix)

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

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

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

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

res_all_cov = SIPI(simData$D,SNPdata,"all",X1,c("gender","group"))

```

Index

*Topic **datasets**

simData, [9](#)

AA9int, [2](#)

Grid3by3, [4](#)

MAFinfo, [5](#)

parAA9int, [6](#)

parSIPI, [7](#)

simData, [9](#)

SIPI, [9](#)