Package ‘gtWAS’

July 22, 2025

Type Package

Title Genome and Transcriptome Wide Association Study

Version 1.1.0

Date 2019-06-01

Author JunhuiLi WenxinLiu

Maintainer JunhuiLi<`junhuili@cau.edu.cn`>

Description Quantitative trait loci mapping and genome wide association analysis are used to find candidate molecular marker or region associated with phenotype based on linkage analysis and linkage disequilibrium. Gene expression quantitative trait loci mapping is used to find candidate molecular marker or region associated with gene expression. In this package, we applied the method in Liu W. (2011) <doi:10.1007/s00122-011-1631-7> and Gusev A. (2016) <doi:10.1038/ng.3506> to genome and transcriptome wide association study, which is aimed at revealing the association relationship between phenotype and molecular markers, expression levels, molecular markers nested within different related expression effect and expression effect nested within different related molecular marker effect. F test based on full and reduced model are performed to obtain p value or likelihood ratio statistic. The best linear model can be obtained by stepwise regression analysis.

License GPL (>= 2)

Depends R (>= 2.10)

NeedsCompilation no

Repository CRAN

Date/Publication 2019-06-01 09:50:03 UTC

Contents

`gtWAS-package` *Genome and Transcriptome Wide Association Study*

Description

Quantitative trait loci mapping and genome wide association analysis are used to find candidate molecular marker or region associated with phenotype based on linkage analysis and linkage disequilibrium. Expression quantitative trait loci mapping is used to find candidate molecular marker or region associated with gene expression. This package is aimed at revealing the association relationship between phenotype and molecular markers, expression levels, molecular markers with different related expression levels and expression levels with different related molecular marker. F test based on full and reduced model are performed to obtain p value or likelihood ratio statistic. The best linear model can be obtained by stepwise regression analysis.

Details

Package:	gtWAS
Type:	Package
Version:	1.1.0
Date:	2019-06-01
License:	GPL (>= 2)

Author(s)

JunhuiLi WeninLiu

Maintainer: JunhuiLi<junhuili@cau.edu.cn>

References

Junhui Li, Haixiao Hu, Yujie Meng, Kun Cheng, Guoliang Li, Wenxin Liu, and Shaojiang Chen.(2016)Pleiotropic QTL detection for stalk traits in maize and related R package programming. Journal of China Agricultural University. DOI 10.11841/j.issn.1007-4333.2016.06.00

Liu W., Maurer H.P., Reif J.C., Melchinger A.E., Utz F., Tucker M.R., Ranc N., Della Porta G., Wurschum T. (2013) Optimum Design of Family Structure and Allocation of Resources in Association Mapping with Lines from Multiple Crosses. Heredity 110: 71-79

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., & Penninx, B. W., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nature Genetics, 48(3), 245.

Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307.

Judge, & GeorgeG. (1985). The Theory and practice of econometrics /-2nd ed. The Theory and practice of econometrics /. Wiley.

Mcquarrie, A. D. R., & Tsai, C. L. (1998). Regression and Time Series Model Selection. Regression and time series model selection /. World Scientific.

Sawa, T. (1978). Information criteria for discriminating among alternative regression models. Econometrica, 46(6), 1273-1291.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6(2), pags. 15-18.

`alldata` *Data including base and expression data*

Description

Data including base and expression data

Usage

```
data("alldata")
```

Format

A data frame with 100 observations on the following 200 variables.

The first 100th variables are SNP and the second are expression data

Examples

```
data(alldata)
```

`Association` *Genome and Transcriptome Wide Association*

Description

Reveal the association relationship between phenotype and molecular marker, expression effect, expression effect nested within molecular marker and molecular marker effect nested within expression effect

Usage

```
Association(Tdata,alldata,independent="B(E)",Elevels=c(0.05,0.95),selection="stepwise",
      select="SL",Choose=NULL,SL=c(0.05,0.15,0.15),correct="Bonferroni")
```

Arguments

`Tdata`	Phenotye data
`alldata`	Independent variables including molecular marker or corresponding expression effect related to marker on transcriptome level
`independent`	Indicator of independent variable to be used in linear model. 'B' is molecular marker effect, 'E' is expression effect, 'B(E)' is molecular marker nesting expression effect and 'E(B)' is expression effect nesting molecular marker effect
`Elevels`	Percentage of threshold value for different expression levels
`selection`	Model selection method including "forward" and "stepwise",forward selection starts with no effects in the model and adds effects, while stepwise regression is similar to the forward method except that effects already in the model do not necessarily stay there
`select`	Specifies the criterion that uses to determine the order in which effects enter and/or leave at each step of the specified selection method including Akaike Information Criterion(AIC), the Corrected form of Akaike Information Criterion(AICc),Bayesian Information Criterion(BIC),Schwarz criterion(SBC), Significant Levels(SL) and so on
`Choose`	Chooses from the list of models at the steps of the selection process the model that yields the best value of the specified criterion. If the optimal value of the specified criterion occurs for models at more than one step, then the model with the smallest number of parameters is chosen. If you do not specify the Choose option, then the model selected is the model at the final step in the selection process
`SL`	Thresholds for significant levels of association and stepwise regression
`correct`	Bonferroni correct or the p value method for significant levels, default is bonferroni

Value

p value of all effect and significant ones

Author(s)

JunhuiLi

References

Junhui Li, Haixiao Hu, Yujie Meng, Kun Cheng, Guoliang Li, Wenxin Liu, and Shaojiang Chen.(2016)Pleiotropic QTL detection for stalk traits in maize and related R package programming. Journal of China Agricultural University. DOI 10.11841/j.issn.1007-4333.2016.06.00

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., & Penninx, B. W., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nature Genetics, 48(3), 245.

Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307.

Judge, & GeorgeG. (1985). The Theory and practice of econometrics /-2nd ed. The Theory and practice of econometrics /. Wiley.

Mcquarrie, A. D. R., & Tsai, C. L. (1998). Regression and Time Series Model Selection. Regression and time series model selection /. World Scientific.

Sawa, T. (1978). Information criteria for discriminating among alternative regression models. Econometrica, 46(6), 1273-1291.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6(2), pags. 15-18.

Examples

```
data(Tdata)
data(alldata)
Edata <- alldata[,1:100+100]
Bdata <- alldata[,1:100]
BE <- "B(E)"
EB <- "E(B)"
B <- "B"
E <- "E"

#for "B(E)"
#Association(Tdata,alldata,BE,Elevels=c(0.05,0.95),selection='stepwise',
#select="SL",Choose=NULL,SL=c(0.05,0.15,0.15),correct = "Bonferroni")

#for "E(B)" with Elevels = null
#Association(Tdata,alldata,EB,Elevels=NULL,selection='stepwise',
#select="SL",Choose=NULL,SL=c(0.05,0.15,0.15),correct = "Bonferroni")

#for "E" with Elevels = null
#Association(Tdata,Edata,E,Elevels=NULL,selection='stepwise',
#select="SL",Choose=NULL,SL=c(0.05,0.15,0.15),correct = "Bonferroni")

#for "B"
#Association(Tdata,Bdata,B,Elevels=NULL,selection='stepwise',
#select="SL",Choose=NULL,SL=c(0.05,0.15,0.15),correct = "Bonferroni")
```

`ModelFit` *Compute model fit statistics*

Description

Compute model fit statistics based on a given criteria for linear model function

Usage

```
ModelFit(criteria, lmresult, nObs, sigma_sqr)
```

Arguments

`criteria`	The class of criteria including Akaike information criterion(AIC), the corrected form of Akaike information criterion(AICc), Bayesian information criterion(BIC), Schwarz criterion(SBC) and significant levels(SL)

`lmresult`	Result of linear model function
`nObs`	Number of observation
`sigma_sqr`	The estimation of pure error variance for the full model in regression

Value

A numeric of model fit statistics

Author(s)

JunhuiLi

References

Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307.

Judge, & GeorgeG. (1985). The Theory and practice of econometrics /-2nd ed. The Theory and practice of econometrics /. Wiley.

Mcquarrie, A. D. R., & Tsai, C. L. (1998). Regression and Time Series Model Selection. Regression and time series model selection /. World Scientific.

R.S. Sparks, W. Zucchini, & D. Coutsourides. (1985). On variable selection in multivariate regression. Communication in Statistics- Theory and Methods, 14(7), 1569-1587.

Sawa, T. (1978). Information criteria for discriminating among alternative regression models. Econometrica, 46(6), 1273-1291.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6(2), pags. 15-18.

Examples

```
set.seed(4)
YX <- matrix(rnorm(200,20,4),20,10)
YX <- as.data.frame(YX)
colnames(YX) <- c("Y1","Y2",paste("X",c(1:8),sep=""))
lm_formula <- as.formula("Y1~X1+X2+X3+X4+X5")
lmresult <- lm(lm_formula,data=YX)
ModelFit("SBC", lmresult, nrow(YX), 0)
```

`StepOne` *Compute minimum p value and information criteria statistics in one step*

Description

Compute minimum p value and information criteria statistics in one step by adding or removing a variable

Usage

```
StepOne(findIn, independent, criteria, varIn, TMdata, sigma)
```

Arguments

`findIn`	Logical value for adding or removing independent variables in regression model, the parameter is ture for removing a variable otherwise adding a variable
`independent`	Indicator of independent variable to be used in linear model. 'B' is molecular marker effect, 'E' is expression effect, 'B(E)' is expression effect nested within molecular marker effect and 'E(B)' is molecular marker effect nested within expression effect
`criteria`	Specifies the criterion that uses to determine the order in which effects enter and/or leave at each step of the specified selection method including Akaike Information Criterion(AIC), the Corrected form of Akaike Information Criterion(AICc),Bayesian Information Criterion(BIC),Schwarz criterion(SBC),Hannan and Quinn Information Criterion(HQ), Significant Levels(SL) and so on
`varIn`	Sequence of vector for every independent variables, 1 indicates this independent variable stays in the regression model, and 0 is not in the model
`TMdata`	Phenotype data
`sigma`	The estimation of pure error variance from the full model in regression

Value

A list of minimum p value or information criteria statistics, sequence id of independent variable staying in the model, linear model regression and rank of last step linear model

Author(s)

JunhuiLi

References

Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307.

Judge, & GeorgeG. (1985). The Theory and practice of econometrics /-2nd ed. The Theory and practice of econometrics /. Wiley.

Mcquarrie, A. D. R., & Tsai, C. L. (1998). Regression and Time Series Model Selection. Regression and time series model selection /. World Scientific.

R.S. Sparks, W. Zucchini, & D. Coutsourides. (1985). On variable selection in multivariate regression. Communication in Statistics- Theory and Methods, 14(7), 1569-1587.

Sawa, T. (1978). Information criteria for discriminating among alternative regression models. Econometrica, 46(6), 1273-1291.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6(2), pags. 15-18.

Examples

```
data(Tdata)
data(alldata)
TMdata <- cbind(Tdata,alldata[,1:100])
findIn = FALSE
independent = "B"
varIn <- rep(0,100)
StepOne(findIn,independent,criteria="SBC",varIn,TMdata,sigma=0)
```

`stp` *stepwise regression*

Description

Stepwise regression for model selection using linear model

Usage

```
stp(AllData, independent, selection = "stepwise", select = "SL",
sle = 0.15, sls = 0.15, Choose = NULL)
```

Arguments

`AllData`	Data about dependent and independent variable data
`independent`	Indicator of independent variable to be used in linear model. 'B' is molecular marker effect, 'E' is expression data, 'B(E)' is expression effect nested within molecular marker effect and 'E(B)' is molecular marker effect nested within expression effect
`selection`	Model selection method including "forward" and "stepwise",forward selection starts with no effects in the model and adds effects, while stepwise regression is similar to the forward method except that effects already in the model do not necessarily stay there
`select`	Specifies the criterion that uses to determine the order in which effects enter and/or leave at each step of the specified selection method including Akaike Information Criterion(AIC), the Corrected form of Akaike Information Criterion(AICc),Bayesian Information Criterion(BIC),Schwarz criterion(SBC),Hannan and Quinn Information Criterion(HQ), Significant Levels(SL) and so on
`sle`	Specifies the significance level for entry, default is 0.15
`sls`	Specifies the significance level for staying in the model, default is 0.15
`Choose`	Chooses from the list of models at the steps of the selection process the model that yields the best value of the specified criterion. If the optimal value of the specified criterion occurs for models at more than one step, then the model with the smallest number of parameters is chosen. If you do not specify the Choose option, then the model selected is the model at the final step in the selection process

Author(s)

JunhuiLi

References

Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307.

Judge, & GeorgeG. (1985). The Theory and practice of econometrics /-2nd ed. The Theory and practice of econometrics /. Wiley.

Mcquarrie, A. D. R., & Tsai, C. L. (1998). Regression and Time Series Model Selection. Regression and time series model selection /. World Scientific.

R.S. Sparks, W. Zucchini, & D. Coutsourides. (1985). On variable selection in multivariate regression. Communication in Statistics- Theory and Methods, 14(7), 1569-1587.

Sawa, T. (1978). Information criteria for discriminating among alternative regression models. Econometrica, 46(6), 1273-1291.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6(2), pags. 15-18.

Examples

```
data(Tdata)
data(alldata)
independent <- "B"
nbase <- 100
AllData <- cbind(Tdata[colnames(Tdata)[1]],alldata[,1:nbase])
AllData <- sapply(AllData, as.numeric)
AllData <- as.data.frame(AllData)
stp(AllData,independent,selection="stepwise",select="SBC",sle=0.05,sls=0.05,Choose=NULL)
```

`Tdata` *Phenotype data*

Description

Phenotype data by rnorm function

Usage

```
data("Tdata")
```

Format

A data frame with 100 observations on the following variable.

`Trait1` a numeric vector

Examples

```
data(Tdata)
```

Index