Package 'MOJOV'

February 19, 2015

Type Package

Title Mojo Variants: Rare Variants analysis

Version 1.0.1

Date 2013-02-25

Author Ke-Hao Wu

Maintainer Ke-Hao Wu <kehao.wu@gmail.com>

Description A package for analysis between rare variants and quantitative traits by CMC (the combined multivariate and collapsing method).

License GPL (>= 2)

Depends methods,aod,survey,saws,lattice

NeedsCompilation no

Repository CRAN

Date/Publication 2013-02-28 16:37:22

R topics documented:

`MOJOV-package` *Mojo Variants: Rare Variants analysis tool.*

Description

Genome-wide association analysis has effectively identified lots of common loci which is related to some diseases, but a large amount of heritability is still unexplained. To find the missing heritability, Morris and Zeggini proposed the combined multivariate and collapsing (CMC) method. In addition, in light of most recent findings, linear regression is still the gold standard to detect the association between rare variants and continuous traits.

To make the CMC method and linear regression analysis more convenient for all people who want to investigate the secret in rare variants and to construct a public platform for rare variants analysis, we developed MOJOV package.MOJOV is from Mojo Variants. It means rare variants is a fullfil of magic power.MOJOV package has some objects as follows:

(1) Providing a complete system to detect rare variants using CMC method and linear regression model.

(2) Providing more convenient method to scan all genes in your squencing data.

(3) Providing more than one options for analysis.

(4) Providing summary statistic information for your data.

Details

Package:	MOJOV
Type:	Package
Version:	1.0.1
Date:	2013-02-25
License:	GPL(>=2)
Depends:	methods,aod,saws,survey,lattice

This package contains two important class and four main functions:

"MOJOV" class: This class will be used from the beginning to the end of your analysis. It will record all your infomation for your analysis. The details you can see MOJOV-class.

"MOJOV.Summary" class: This class is created by MOJOV.summary function. The details you can see MOJOV.summary.

MOJOV.read: It reads genotype and phenotype data from your file. And it create "MOJOV" class for your later analysis.

MOJOV.phenotype: It adjusts your phenotype.

MOJOV.analysis: This is the most important function in this package.

MOJOV.summary: It will help you to know you data deeply by export some import summary imformation.

To finish a complete analysis using MOJOV package, you can do as following steps:

(1) Reading your file by MOJOV.read, and it will return a "MOJOV" class. You should store this class into a variable for the next step.

(2) Passing the previous step result to MOJOV.phenotype. It will process your phenotype. The same as the step 1, you need store the result for the next step.

(3) Throw the result from the step 2 to MOJOV.analysis, it will help you accomplish all analysis. And it also return a "MOJOV" class. Of course, you can specify a file name to store your result.

The MOJOV.analysis is designed for more than one function. It can be run for whole genome, one single gene and specify some regions of interest. If you run this function, it will call MOJOV.genoMatrix, MOJOV.genoVector and MOJOV.linerRegAnalysis automatically. The details is described as follows:

At first, the MOJOV.genoMatrix read a "MOJOV" class which is from MOJOV.read, and export a matrix for genotype. And then, it passes the matrix to MOJOV.genoVector. It will create a vector called genotype vector by using different coding methods, such as proportion and indicator. At last, the genotype vector and the adjusted phenotype data is recieved by MOJOV.liearRegAnalysis and export the analysis result.

Also, MOJOV provides a function to simulate, but it will call ms by system() function to creating analysis cohort.

Author(s)

Ke-Hao Wu

Maintainer: Ke-Hao Wu <kehao.wu@gmail.com>

References

Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol. 2009;34:188-193.

Dering C, Pugh E, Ziegler A. Statistical analysis of rare sequence variants: an overview of collapsing methods. Genet Epidemiol. 2011;35(Suppl 8):12-17.

See Also

MOJOV

Examples

```
file1<-system.file("data","geno.csv.gz",package="MOJOV")
file2<-system.file("data","phen.csv.gz",package="MOJOV")
x<-MOJOV.read(genofile=file1,phenofile=file2)
x
```

geno *A data file contains genotype.*

Description

This data is created at random.

MOJOV-class *Class "MOJOV"*

Description

A class for MOJOV package, and you can create it by call `new("MOJOV",...)`

Slots

genoFile `"character"` Genotype file name.

genoIID `"character"` Individual ID in genotype file.

genoChr `"character"` Chromosome name in genotype file.

genoPosi `"numeric"` Physical position in genotype file.

genoVariant `"character"` Variants labels in genotype file.

genoH `"numeric"` Genotype code in genotype file. 1 stands for heterozygote, 0 stands for no variant copy, 2 stands for two variant copies.

genoGene `"character"` Gene labels in genotype file.

phenoFile `"character"` Phenotype file name.

phenoIID `"character"` Individuals ID in phenotype.

phenoGender `"numeric"` Gender information.

phenoAge `"numeric"` Age for all individuals.

phenoHeight `"numeric"` Height for all individuals.

phenoWeight `"numeric"` Weight for all individuals.

phenoLabel `"character"` Phenotype name.

phenoData `"numeric"` Phenotype data.

phenoColumn `"numeric"` Indicating which column will be read.

phenoIndNum `"numeric"` Number of individuals.

adjustAuto `"logical"` An arguments for MOJOV.phenotype.

adjustData `"numeric"` Phenotype data after adjusted.

adjustGender `"logical"` An arguments for MOJOV.phenotype.

adjustPower `"numeric"` The exponent for adjust.

adjustPowerPvalue `"numeric"` The result of shapiro test.

adjustTerms "character" Which term has been adjusted.

adjustPvalue "numeric" The p value for adjust.

varMAF "numeric" Minor allele frequency.

varFreq "numeric" Frequency for all variants.

varTot "character" ~~

varRare "character" ~~

regionFile "character" File name for ROI file.

regionType "character" Indicating the ROI type.

regionChr "character" Chromosome infomation in ROI file.

regionPStart "numeric" Physical position (start) in ROI file.

regionPStop "numeric" Physical position (stop) in ROI file.

regionGene "character" Gene information in ROI file.

analyCode "character" Coding method.

analyWeighted "character" Weighted method.

resultMethod "character" Test method.

resultPvalue "numeric" P value from linear regression analysis.

Author(s)

Ke-Hao Wu

See Also

MOJOV.read

`MOJOV.analysis` *Analysis between rare variants and quantatitive traits using CMC.*

Description

Analysis between rare variants and quantatitive traits using CMC.

Usage

```
MOJOV.analysis(x = NULL, MAF = 0.01, ROI = "scan", savefile = NULL,
      codeMethod = c("Proportion", "Indicator", "ChuanhuaXin"),
      weightMethod = NULL,testMethod = c("FTest", "WaldTest",
      "LRT", "Sandwich", "all"), ...)
```

Arguments

x	A "MOJOV" class, it must be edited by MOJOV.phenotype function.
MAF	Specify the minor allele frequency, and the default value is 0.01.
ROI	Specify the region of interest.It can be a gene symbol, character "scan" or a ROI file name (see details). The default value is "scan", and it will scan all gene in your input "MOJOV" class.
savefile	You can specify a file name here, if your want to export the result to a file.
codeMethod	Specify the coding method for CMC, It can be "Indicator", "Proportion" and "ChuanhuaXin". And the default value is "Proportion".Details see MOJOV.genoVector.
weightMethod	Specify the weighted method for CMC, It only provide "ChuanhuaXin" method in this version.
testMethod	Specify the test method for linear regression, It can be "FTest", "WaldTest", "LRT", "Sandwich" and "all". And the default value is "FTest".
...	You can specify other arguments here, for example "method" for regTermTest.

Details

This function is the most important one in MOJOV. And it provide more than one method to code your data, to find the rare variants which are casual variants.

At first, you can change the MAF argument, it can be 0.05 or 0.01 and any value you want. And then, You can choose different region for your analysis.

To achieve the objective described, you can handle the ROI argument. It can be NULL, a gene symbol, a region of interest file and a word "scan". If it is NULL, it will analysis the whole data as one set. This way is best for the file which only contains one gene data. Otherwise, it will judge whether the ROI is a gene symbol from your file. If it is, it analysis the gene. And if it is not, it will find a file whose name is you described in your argument, the format of the file will be showed as below. And at last, if it is a word "scan", it will seperate data for different genes automatically.

The file for ROI should be four columns, They are gene symbols, chromosomes, start positions and stop positions. And the start position should be less than the stop position.

In terms of codeMethod argument, it can be "Indicator", "Proportion" and "ChuanhuaXin". When it is "Indicator", it will result in a dichotomous variable that indicates presence or absence of any rare variant whin the ROI. When it is "Proportion", it counts the number of variants of subject over all sites; that is, it is proportion of all variants for all subjects. If it is "ChuanhuaXin", it calls a method from Xing et al.

In this version, the weightMethod is only provided for "ChuanhuaXin" method. In light of testMethod, the "FTest" will call a test named " F test". The "WaldTest" will call a function termed wald.test from aod package. The "LRT" will call regTermTest from survey package, and the "Sandwich" will call saws from saws package, this function will use sandwich estimator. At last, if it is "all", it will call all function above, and the result returned is a matrix.

Value

MOJOV-class MOJOV.genoMatrix MOJOV.genoVector

Author(s)

Ke-Hao Wu

References

Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol. 2009;34:188-193.

Dering C, Pugh E, Ziegler A. Statistical analysis of rare sequence variants: an overview of collapsing methods. Genet Epidemiol. 2011;35(Suppl 8):12-17.

See Also

MOJOV-class

Examples

```
file1<-system.file("data","geno.csv.gz",package="MOJOV")
file2<-system.file("data","phen.csv.gz",package="MOJOV")
x<-MOJOV.read(genofile=file1,phenofile=file2)
x<-MOJOV.phenotype(x=x)
x<-MOJOV.analysis(x=x)
```

MOJOV.export *Export test result from a MOJOV class to file.*

Description

Export test result from a MOJOV class to file.

Usage

```
MOJOV.export(x = NULL, file = "MOJOV.result.csv", p = 0.001)
```

Arguments

x	A "MOJOV" class, and it must be edited by MOJOV.analysis.
file	A file name, and the default value is "MOJOV.result.csv".
p	Specify the significance level, and the default value is 0.001.

Author(s)

Ke-Hao Wu

See Also

MOJOV-class MOJOV.analysis

Examples

```
file1<-system.file("data","geno.csv.gz",package="MOJOV")
file2<-system.file("data","phen.csv.gz",package="MOJOV")
x<-MOJOV.read(genofile=file1,phenofile=file2)
x<-MOJOV.phenotype(x=x)
x<-MOJOV.analysis(x=x)
MOJOV.export(x=x,file="example.result.csv",p=0.05)
```

`MOJOV.genoMatrix` *A internal function. It generates a genotype matrix.*

Description

A internal function. It generates a genotype matrix.

Usage

```
MOJOV.genoMatrix(genoIID = NULL, genoVariant = NULL, genoH = NULL,
      phenoIID = NULL, MAF = NULL)
```

Arguments

`genoIID`	A character vector for individuals ID.
`genoVariant`	A character vector for variant labels.
`genoH`	A character vector for genotype.
`phenoIID`	A character vector for all individuals from phenotype file.
`MAF`	Specify the minor allele frequency.

Details

See `MOJOV.analysis`

Value

This function will return a matrix, it is individuals multiply sites. If the specified variant site of one individual contains one copy of rare variant, it will code 1. If two copy, it will code 2, and others are coded 0.

Author(s)

Ke-Hao Wu

See Also

`MOJOV.genoVector` `MOJOV.analysis`

Examples

```
#Generating individuals ID for genotype file at random.
genoIID<-paste("ID",floor(runif(2000,1,101)),sep="")
#Generating variants labels for genotype file at random.
genoVariant<-paste("rs",floor(runif(2000,223,250)),sep="")
#Generating genotype code for genotype file at random.
genoH<-floor(runif(2000,1,3))
#Generating individuals ID for phenotype file.
phenoIID<-paste("ID",1:100,sep="")
#Generating genotype matrix.
x<-MOJOV.genoMatrix(genoIID=genoIID,genoVariant=genoVariant,genoH=genoH,
         phenoIID=phenoIID)
```

`MOJOV.genoVector` *A internal function. It makes genotype matrix become a genotype vector.*

Description

It makes genotype matrix become a genotype vector using different coding methods.

Usage

```
MOJOV.genoVector(x = NULL, y = NULL, codeMethod = c("Proportion", "Indicator",
      "ChuanhuaXin"), weightMethod = c("ChuanhuaXin"))
```

Arguments

`x`	A matrix for genotype, and it can be from MOJOV.matrix.
`y`	A vector for phenotype, it's length should be equal to the rows of x.
`codeMethod`	Specify the coding method for CMC, It can be "Indicator", "Proportion" and "ChuanhuaXin". And the default value is "Proportion".
`weightMethod`	Specify the weighted method for CMC, It only provide "ChuanhuaXin" method in this version.

Details

See `MOJOV.analysis`

Value

It will return a vector of genotype processed by different coding methods.

Author(s)

Ke-Hao WU

References

Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol. 2009;34:188-193.

Dering C, Pugh E, Ziegler A. Statistical analysis of rare sequence variants: an overview of collapsing methods. Genet Epidemiol. 2011;35(Suppl 8):12-17.

See Also

MOJOV.analysis MOJOV.genoMatrix MOJOV.weight

Examples

```
#Generating individuals ID for genotype file at random.
genoIID<-paste("ID",floor(runif(2000,1,101)),sep="")
#Generating variants labels for genotype file at random.
genoVariant<-paste("rs",floor(runif(2000,223,250)),sep="")
#Generating genotype code for genotype file at random.
genoH<-floor(runif(2000,1,3))
#Generating individuals ID for phenotype file.
phenoIID<-paste("ID",1:100,sep="")
#Generating genotype matrix.
x<-MOJOV.genoMatrix(genoIID=genoIID,genoVariant=genoVariant,genoH=genoH,
         phenoIID=phenoIID)
#Generating genotype vector.
xVector<-MOJOV.genoVector(x=x,y=phenoIID)
```

MOJOV.linearRegAnalysis

A internal function. Linear regression analysis between rare variants and phenotype.

Description

A internal function. Linear regression analysis between rare variants and phenotype.

Usage

```
MOJOV.linearRegAnalysis(x = NULL, y = NULL, testMethod = c("FTest",
      "WaldTest", "LRT", "Sandwich", "all"), ...)
```

Arguments

x	A vector for genotype, and it can be from MOJOV.genoVector.
y	A vector for phenotype.
testMethod	Specify the test method for linear regression, It can be "FTest", "WaldTest", "LRT", "Sandwich" and "all". And the default value is "FTest".
...	See MOJOV.analysis

Details

See MOJOV.analysis

Author(s)

Ke-Hao Wu

See Also

MOJOV.analysis

Examples

```
x<-runif(100,0,3)
y<-x+rnorm(100,1,1)
MOJOV.linearRegAnalysis(x=x,y=y)
```

MOJOV.phenotype *Ajusting phenotype data for covariates.*

Description

Ajusting phenotype data for covariates.

Usage

```
MOJOV.phenotype(x = NULL, auto = FALSE, gender = TRUE, power = 1,
      Terms = 1:3)
```

Arguments

x	A "MOJOV" class.
auto	If it is TRUE, it will automatically find the best arguments. And the default value is FALSE.
gender	If it is TRUE, it will adjust for different gender.
power	Specify the exponent. It can be a number or a numeric vector.
Terms	It specify which argument will be used for adjusted. The 1,2,3 stand for age, height and weight respectively.

Author(s)

Ke-Hao Wu

MOJOV.read *Reading genotype and phenotype and generating a "MOJOV" class.*

Description

Reading genotype and phenotype and generating a "MOJOV" class.

And it can read another phenotype for a existing "MOJOV" class.

Usage

```
MOJOV.read(x = NULL, genofile = NULL, phenofile = NULL,
      indelfile = NULL, header = T, column = 1, ...)
```

Arguments

x	If x is null, it will need both genotype and phenotype file, or x should be a "MOJOV" class and it only need phenotype file, see detials.
genofile	Specify genotype file name.
phenofile	Specify phenotype file name.
indelfile	Specify insert and delete variants file name, it is optional.
header	Specify whether the file contains a head line.
column	Which column phenotype will be read. This column is not include the previous 5 columns. The default value is 1.
...	You can specify other arguments here.

Author(s)

Ke-Hao Wu

See Also

MOJOV-class

Examples

```
#Reading both genotype and phenotype files.
file1<-system.file("data","geno.csv.gz",package="MOJOV")
file2<-system.file("data","phen.csv.gz",package="MOJOV")
x<-MOJOV.read(genofile=file1,phenofile=file2)

#Reading phenotype and switch different phenotype column to analysis.
x  #x is a existing "MOJOV" class. The phenotype is R_13_AREA.
x<-MOJOV.read(x=x,phenofile=file2,column=2) #The phenotype is _13_BMD.
```

MOJOV.regTermTest *A internal function. A test from survey package.*

Description

A test from survey package for linear regression.

Usage

```
MOJOV.regTermTest(x, y, ...)
```

Arguments

x	A genotype vector.]
y	A phenotype vector.
...	Other arguments for regTermTest from survey. See ?regTermTest

Author(s)

Ke-Hao Wu

MOJOV.saws *A test from saws package.*

Description

A test from saws package for linear regression.

Usage

```
MOJOV.saws(x, y, ...)
```

Arguments

x	A genotype vector.
y	A phenotype vector.
...	Other arguments for regTermTest from survey. See ?saws

MOJOV.simulation *Simulation for CMC.*

Description

Provide a simulation function using ms software. If you have no ms and no Linux, you can not run this function.

Usage

```
MOJOV.simulation(cohortSize = 500, nReps = 2, theta = 10, sites = NULL,
affectNum = NULL, MAF = 0.01, totalMAF = 0.05, lambda = 1, sd = NULL,
type = c("alpha","belta"), sampleNum = 100, outFile = NULL, plot = FALSE,
codeMethod = c("Proportion", "Indicator", "ChuanhuaXin"),
weightMethod = c("ChuanhuaXin"), testMethod = c("FTest", "WaldTest",
"LRT", "Sandwich", "all"), save = NULL)
```

Arguments

cohortSize	Specify the size for analysis cohort.
nReps	Specify the number of independent samples to generate.
theta	Specify the mutation parameter,see details.
sites	Specify the number of sites, it is optional.
affectNum	Specify the number or ratio for affect sites.If it is less than 1, it will be the ratio of affect rare variants sites to all rare variants.
MAF	Specify the minor allele frequency, and the default value is 0.01.
totalMAF	Specify the total minor allele frequency for affect variants. And the default value is 0.05.
lambda	Specify the lambda for phenotype standard error.See details.
sd	Specify the standard error.See details.
type	It can be "alpha" and "belta". The default value is "alpha".
sampleNum	Specify the number of sample. The default value is 100.
outFile	You can specify a file name to save your data from ms.
plot	You can set it as TRUE, it will plot the result using lattice package.
codeMethod	See MOJOV.analysis
weightMethod	See MOJOV.analysis
testMethod	See MOJOV.analysis
save	You can specify a file name to save your simulation result.

Author(s)

Ke-Hao Wu

MOJOV.summary *Exporting summary information for genotype.*

Description

Exporting summary information for genetype data. Including total information, such as how many genes, how many variants, how many rare variants. And some information for each gene.

Usage

```
MOJOV.summary(x = NULL, ROI = "scan", bin = c(0, 0.01, 0.05,
      0.1, 0.2, 0.3, 0.4, 0.5), MAF = 0.05, ...)
```

Arguments

x	A "MOJOV" class.
ROI	It can be "scan" or a gene symbol. If it is "scan", it will scan the whole genome for every gene. Or it only run for the specified gene.
bin	A bin for frequency statistic.The defalut value is c(0, 0.01, 0.05,0.1, 0.2, 0.3, 0.4, 0.5).
MAF	Specify the minor allele frequency, and the default value is 0.05.
...	...

Author(s)

Ke-Hao Wu

MOJOV.Summary-class *Class* "MOJOV.Summary"

Description

A class for MOJOV package, and you can create it by call `new("MOJOV.Summary",...)`

Author(s)

Ke-Hao Wu

Examples

```
showClass("MOJOV.Summary")
```

MOJOV.wald.test *A test from aod package.*

Description

A test from aod package for linear regression.

Usage

```
MOJOV.wald.test(x, y, ...)
```

Arguments

x	A genotype vector.
y	A phenotype vector.
...	Other arguments for regTermTest from survey. See ?regTermTest

Author(s)

Ke-Hao Wu

MOJOV.weight *Providing a weight vector.*

Description

A function to provide weight data for analsysis.

Usage

```
MOJOV.weight(x = NULL, y = NULL, weightMethod = "ChuanhuaXin")
```

Arguments

x	A genotype vector.
y	A phenotype vector.
weightMethod	Specify the weighted method for CMC, It only provide "ChuanhuaXin" method in this version.

Author(s)

Ke-Hao Wu

phen *A data file contains phenotype.*

Description

This data is created at random.

Index