Package 'mrMLM'

August 26, 2018

Type Package

Title Multi-Locus Random-SNP-Effect Mixed Linear Model Tools for Genome-Wide Association Study

Version 3.1

Date 2018-8-25

Author Zhang Ya-Wen, Li Pei, Ren Wen-Long, Ni Yuan-Li, and Zhang Yuan-Ming

Maintainer Yuanming Zhang<soyzhang@mail.hzau.edu.cn>

Description Conduct multi-locus genome-wide association study under the framework of random-SNP-effect mixed linear model (mrMLM). First, each marker on the genome is scanned. Bonferroni correction is replaced by a less stringent selection criterion for significant test. Then, all the markers that are potentially associated with the trait are included in a multi-locus model, their effects are estimated by empirical Bayes and true QTN are identified by likelihood ratio test. Wen YJ, Zhang H, Ni YL, Huang B, Zhang J, Feng JY, Wang SB, Dunwell JM, Zhang YM, Wu R (2018) <doi:10.1093/bib/bbw145>.

Depends MASS,data.table,doParallel,foreach

Imports methods,openxlsx,stringr,qqman,ggplot2,lars,ncvreg,coin,sampling,bigmemory

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2018-08-25 22:10:03 UTC

R topics documented:

`DoData` *process raw data*

Description

process raw data for later use

Usage

```
DoData(genRaw,Genformat,pheRaw1q,kkRaw,psmatrixRaw,trait,type)
```

Arguments

`genRaw`	raw genotype matrix.
`Genformat`	genotype format.
`pheRaw1q`	raw phenotype matrix.
`kkRaw`	raw kinship matrix.
`psmatrixRaw`	raw population structure matrix.
`trait`	which trait to analysis.
`type`	which type to transform.

Author(s)

Zhang Ya-Wen, Li Pei, Ren Wen-Long, Ni Yuan-Li, Zhang Yuan-Ming
Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

Examples

```
G1=data(Gen)
P1=data(Phe)
readraw=ReadData(fileGen=Gen,filePhe=Phe,fileKin=NULL,filePS =NULL,
Genformat=1)
result=DoData(readraw$genRaw,Genformat=3,readraw$pheRaw1q,readraw$kkRaw,
readraw$psmatrixRaw,trait=1,type=2)
```

FASTmrEMMA *To perform GWAS with FASTmrEMMA method*

Description

FAST multi-locus random-SNP-effect EMMA

Usage

```
FASTmrEMMA(gen,phe,outATCG,genRaw,kk,psmatrix,svpal,svmlod,Genformat,Likelihood,CLO)
```

Arguments

gen	genotype matrix.
phe	phenotype matrix.
outATCG	genotype for code 1.
genRaw	raw genotype.
kk	kinship matrix.
psmatrix	population structure matrix.
svpal	Critical P-value for selecting variable.
svmlod	Critical LOD score for significant QTN.
Genformat	Format for genotypic codes.
Likelihood	restricted maximum likelihood (REML) and maximum likelihood (ML).
CLO	number of CPU.

Author(s)

Zhang Ya-Wen, Li Pei, Ren Wen-Long, Ni Yuan-Li, Zhang Yuan-Ming
Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

Examples

```
G1=data(Gen)
P1=data(Phe)
Readraw=ReadData(fileGen=Gen,filePhe=Phe,fileKin=NULL,filePS =NULL,
Genformat=1)
InputData=inputData(readraw=Readraw,Genformat=1,method="FASTmrEMMA",trait=1)
result=FASTmrEMMA(InputData$doFME$gen,InputData$doFME$phe,
InputData$doFME$outATCG,InputData$doFME$genRaw,
InputData$doFME$kk,InputData$doFME$psmatrix,0.005,
svmlod=3,Genformat=1,Likelihood="REML",CLO=1)
```

FASTmrMLM *To perform GWAS with FASTmrMLM method*

Description

FAST multi-locus random-SNP-effect Mixed Linear Model

Usage

```
FASTmrMLM(gen,phe,outATCG,genRaw,kk,psmatrix,svpal,svrad,svmlod,Genformat,CLO)
```

Arguments

gen	genotype matrix.
phe	phenotype matrix.
outATCG	genotype for code 1.
genRaw	raw genotype.
kk	kinship matrix.
psmatrix	population structure matrix.
svpal	Critical P-value for selecting variable.
svrad	Search Radius in search of potentially associated QTN.
svmlod	Critical LOD score for significant QTN.
Genformat	Format for genotypic codes.
CLO	number of CPU.

Author(s)

Zhang Ya-Wen, Li Pei, Ren Wen-Long, Ni Yuan-Li, Zhang Yuan-Ming
Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

Examples

```
G1=data(Gen)
P1=data(Phe)
Readraw=ReadData(fileGen=Gen,filePhe=Phe,fileKin=NULL,filePS =NULL,
Genformat=1)
InputData=inputData(readraw=Readraw,Genformat=1,method="FASTmrMLM",trait=1)
result=FASTmrMLM(InputData$doMR$gen,InputData$doMR$phe,
InputData$doMR$outATCG,InputData$doMR$genRaw,
InputData$doMR$kk,InputData$doMR$psmatrix,0.01,svrad=20,
svmlod=3,Genformat=1,CLO=1)
```

Gen *Genotype data*

Description

Numeric format of genotype dataset.

Usage

```
data(Gen)
```

Details

Dataset input of Genotype for mrMLM function.

Author(s)

Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

inputData *Input data which have been transformed*

Description

Input all the dataset which have been transformed

Usage

```
inputData(readraw,Genformat,method,trait)
```

Arguments

readraw	genotype matrix.
Genformat	genotype format.
method	which method to analysis.
trait	which trait to analysis.

Author(s)

Zhang Ya-Wen, Li Pei, Ren Wen-Long, Ni Yuan-Li, Zhang Yuan-Ming
Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

Examples

```
G1=data(Gen)
P1=data(Phe)
Readraw=ReadData(fileGen=Gen,filePhe=Phe,fileKin=NULL,filePS =NULL,
Genformat=1)
result=inputData(readraw=Readraw,Genformat=3,method="mrMLM",trait=1)
```

ISIS *To perform GWAS with ISIS EM-BLASSO method*

Description

Iterative Sure Independence Screening EM-Bayesian LASSO

Usage

```
ISIS(gen,phe,outATCG,genRaw,kk,psmatrix,svpal,svmlod,Genformat,CLO)
```

Arguments

gen	genotype matrix.
phe	phenotype matrix.
outATCG	genotype for code 1.
genRaw	raw genotype.
kk	kinship matrix.
psmatrix	population structure matrix.
svpal	Critical P-value for selecting variable.
svmlod	Critical LOD score for significant QTN.
Genformat	Format for genotypic codes.
CLO	number of CPU.

Author(s)

Zhang Ya-Wen, Li Pei, Ren Wen-Long, Ni Yuan-Li, Zhang Yuan-Ming
Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

Examples

```
G1=data(Gen)
P1=data(Phe)
Readraw=ReadData(fileGen=Gen,filePhe=Phe,fileKin=NULL,filePS =NULL,
Genformat=1)
InputData=inputData(readraw=Readraw,Genformat=1,method="ISIS EM-BLASSO",
trait=1)
result=ISIS(InputData$doMR$gen,InputData$doMR$phe,InputData$doMR$outATCG,
InputData$doMR$genRaw,InputData$doMR$kk,InputData$doMR$psmatrix,
0.01,svmlod=3,Genformat=1,CLO=1)
```

mrMLM *Multi-Locus Random-SNP-Effect Mixed Linear Model Tools for Genome-Wide Association Study*

Description

Conduct multi-locus genome-wide association study under the framework of random-SNP-effect mixed linear model (mrMLM). First, each marker on the genome is scanned. Bonferroni correction is replaced by a less stringent selection criterion for significant test. Then, all the markers that are potentially associated with the trait are included in a multi-locus model, their effects are estimated by empirical Bayes and true QTNs are identified by likelihood ratio test.

Usage

```
mrMLM(fileGen,filePhe,fileKin,filePS,Genformat,method,Likelihood,trait,
SearchRadius,CriLOD,SelectVariable,Bootstrap,DrawPlot,Plotformat,Resolution,dir)
```

Arguments

fileGen	File path and name in your computer of Genotype.
filePhe	File path and name in your computer of Phenotype.
fileKin	File path and name in your computer of Kinship.
filePS	File path and name in your computer of Population Structure.
Genformat	Format for genotypic codes, Num (number), Cha (character) and Hmp (Hapmap).
method	Six multi-locus GWAS methods. Users may select one to six methods, including mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB and ISIS EM-BLASSO.
Likelihood	This parameter is only for FASTmrEMMA, including restricted maximum likelihood (REML) and maximum likelihood (ML).
trait	Traits analyzed from number 1 to number 2.
SearchRadius	This parameter is only for mrMLM and FASTmrMLM, indicating Search Radius in search of potentially associated QTN.
CriLOD	Critical LOD score for significant QTN.
SelectVariable	This parameter is only for pLARmEB. SelectVariable=50 indicates that 50 potentially associated variables are selected from each chromosome. Users may change this number in real data analysis in order to obtain the best results as final results.
Bootstrap	This parameter is only for pLARmEB, including FASLE and TRUE, Bootstrap=FALSE indicates the analysis of only real dataset, Bootstrap=TRUE indicates the analysis of both real dataset and four resampling datasets.
DrawPlot	This parameter is for all the six methods, including FALSE and TRUE, DrawPlot=FALSE indicates no figure output, DrawPlot=TRUE indicates the output of the Manhattan, QQ and LOD score against genome position figures.

Plotformat	This parameter is for all the figure files, including *.jpeg, *.png, *.tiff and *.pdf.
Resolution	This parameter is for all the figure files, including Low and High.
dir	This parameter is for the save path.

Details

Package:	mrMLM
Type:	Package
Version:	3.1
Date:	2018-8-4
Depends:	MASS,data.table,doParallel,foreach
Imports:	methods,openxlsx,stringr,qqman,ggplot2,lars,ncvreg,coin
License:	GPL version 2 or newer
LazyLoad:	yes

Author(s)

Zhang Ya-Wen, Li Pei, Ren Wen-Long, Ni Yuan-Li, Zhang Yuan-Ming
Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

References

Methodological implementation of mixed linear models in multi-locus genome-wide association studies. Yang-Jun Wen,Hanwen Zhang,Yuan-Li Ni,Bo Huang,Jin Zhang,Jian-Ying Feng,Shi-Bo Wang,Jim M.Dunwell,Yuan-Ming Zhang,Rongling Wu.

Examples

```
G1=data(Gen)
P1=data(Phe)
mrMLM(fileGen=Gen,filePhe=Phe,fileKin=NULL,filePS=NULL,Genformat="Num",
method=c("ISIS EM-BLASSO"),Likelihood="REML",trait=1:1,SearchRadius=20,
CriLOD=3,SelectVariable=50,Bootstrap=FALSE,
DrawPlot=FALSE,Plotformat="jpeg",Resolution="Low",dir=tempdir())
```

---

mrMLMFun *To perform GWAS with mrMLM method*

Description

multi-locus random-SNP-effect Mixed Linear Model

Usage

```
mrMLMFun(gen,phe,outATCG,genRaw,kk,psmatrix,svpal,svrad,svmlod,Genformat,CLO)
```

Arguments

gen	genotype matrix.
phe	phenotype matrix.
outATCG	genotype for code 1.
genRaw	raw genotype.
kk	kinship matrix.
psmatrix	population structure matrix.
svpal	Critical P-value for selecting variable
svrad	Search Radius in search of potentially associated QTN.
svmlod	Critical LOD score for significant QTN.
Genformat	Format for genotypic codes.
CLO	number of CPU.

Author(s)

Zhang Ya-Wen, Li Pei, Ren Wen-Long, Ni Yuan-Li, Zhang Yuan-Ming
Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

Examples

```
G1=data(Gen)
P1=data(Phe)
Readraw=ReadData(fileGen=Gen,filePhe=Phe,fileKin=NULL,filePS =NULL,
Genformat=1)
InputData=inputData(readraw=Readraw,Genformat=1,method="mrMLM",trait=1)
result=mrMLMFun(InputData$doMR$gen,InputData$doMR$phe,InputData$doMR$outATCG,
InputData$doMR$genRaw,InputData$doMR$kk,InputData$doMR$psmatrix,
0.01,svrad=20,svmlod=3,Genformat=1,CLO=1)
```

Phe *Phenotype dataset*

Description

Phenotype dataset of multiple traits.

Usage

```
data(Phe)
```

Details

Dataset input of phenotype in mrMLM function.

Author(s)

Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

---

pKWmEB *To perform GWAS with pKWmEB method*

Description

Kruskal-Wallis test with empirical Bayes under polygenic background control

Usage

```
pKWmEB(gen,phe,outATCG,genRaw,kk,psmatrix,svpal,svmlod,Genformat,CLO)
```

Arguments

gen	genotype matrix.
phe	phenotype matrix.
outATCG	genotype for code 1.
genRaw	raw genotype.
kk	kinship matrix.
psmatrix	population structure matrix.
svpal	Critical P-value for selecting variable.
svmlod	Critical LOD score for significant QTN.
Genformat	Format for genotypic codes.
CLO	number of CPU.

Author(s)

Zhang Ya-Wen, Li Pei, Ren Wen-Long, Ni Yuan-Li, Zhang Yuan-Ming
Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

Examples

```
G1=data(Gen)
P1=data(Phe)
Readraw=ReadData(fileGen=Gen,filePhe=Phe,fileKin=NULL,filePS =NULL,
Genformat=1)
InputData=inputData(readraw=Readraw,Genformat=1,method="pKWmEB",trait=1)
result=pKWmEB(InputData$doMR$gen,InputData$doMR$phe,InputData$doMR$outATCG,
InputData$doMR$genRaw,InputData$doMR$kk,InputData$doMR$psmatrix,
0.05,svmlod=3,Genformat=1,CLO=1)
```

pLARmEB *To perform GWAS with pLARmEB method*

Description

polygene-background-control-based least angle regression plus Empirical Bayes

Usage

```
pLARmEB(gen,phe,outATCG,genRaw,kk,psmatrix,CriLOD,lars1,Genformat,Bootstrap,CLO)
```

Arguments

gen	genotype matrix.
phe	phenotype matrix.
outATCG	genotype for code 1.
genRaw	raw genotype.
kk	kinship matrix.
psmatrix	population structure matrix.
CriLOD	Critical LOD score for significant QTN.
lars1	No. of potentially associated variables selected by LARS.
Genformat	Format for genotypic codes.
Bootstrap	Bootstrap=FALSE indicates the analysis of only real dataset, Bootstrap=TRUE indicates the analysis of both real dataset and four resampling datasets.
CLO	number of CPU.

Author(s)

Zhang Ya-Wen, Li Pei, Ren Wen-Long, Ni Yuan-Li, Zhang Yuan-Ming
Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

Examples

```
G1=data(Gen)
P1=data(Phe)
Readraw=ReadData(fileGen=Gen,filePhe=Phe,fileKin=NULL,filePS =NULL,
Genformat=1)
InputData=inputData(readraw=Readraw,Genformat=1,method="pLARmEB",trait=1)
result=pLARmEB(InputData$doMR$gen,InputData$doMR$phe,InputData$doMR$outATCG,
InputData$doMR$genRaw,InputData$doMR$kk,InputData$doMR$psmatrix,
CriLOD=3,lars1=50,Genformat=1,Bootstrap=FALSE,CLO=1)
```

ReadData *read raw data*

Description

read raw data which have not been transformed

Usage

```
ReadData(fileGen,filePhe,fileKin,filePS,Genformat)
```

Arguments

fileGen	genotype matrix.
filePhe	phenotype matrix.
fileKin	kinship matrix.
filePS	population structure matrix.
Genformat	genotype format.

Author(s)

Zhang Ya-Wen, Li Pei, Ren Wen-Long, Ni Yuan-Li, Zhang Yuan-Ming
Maintainer: Yuan-Ming Zhang<soyzhang@mail.hzau.edu.cn>

Examples

```
G1=data(Gen)
P1=data(Phe)
result=ReadData(fileGen=Gen,filePhe=Phe,fileKin=NULL,filePS =NULL,
Genformat=1)
```

Index