Package 'FRGEpistasis'

October 12, 2016

Type Package

Title Epistasis Analysis for Quantitative Traits by Functional Regression Model

biocViews Genetics, NetworkInference, GeneticVariability, Software

Version 1.8.0

Date 2014-03-24

Author Futao Zhang

Maintainer Futao Zhang <futoaz@gmail.com>

Depends R (>= 2.15), MASS, fda, methods, stats

Imports utils

Description A Tool for Epistasis Analysis Based on Functional Regression Model

License GPL-2

NeedsCompilation no

R topics documented:

`FRGEpistasis-package` *Package of Epistasis Detection by Functional Regression Model*

Description

FRGEpistasis is designed to detect the epistasis between genes or genomic regions for both common variants and rare variants. Currently FRGEpistasis was developed by Futao Zhang with R language and maintained in Xiong lab at UTSPH. This tool is friendly, convenient and memory efficient.

Details

Package:	FRGEpistasis
Type:	Package
Version:	0.99.5
Date:	2014-03-22
License:	GPL-2

It currently has the following functional modules: Functional Regression Model (FRG) for Testing Interaction; Regression on Principal Components Analysis (PCA) for Testing Interaction; Pointwise interaction Test.

Author(s)

Futao Zhang Maintainer: Futao Zhang <futoaz@gmail.com>

`fourierExpansion` *Fourier Expansion of Genotype*

Description

This function aims to expand the genotype of one gene (or genomic region) with Fourier Expansion.

Usage

```
fourierExpansion(gene_idx, geno, gene_list, snp_map, rng)
```

Arguments

`gene_idx`	The expansion gene index in the gene annotation list.
`geno`	Genotype of all the genes in the gene annotation list.
`gene_list`	Gene annotation list which includes gene name, chromosome, start position and end position.

snp_map	SNP genetic map includes chromosome, snp indentifier, genetic distance and base-pair position.
rng	A numeric value which represents gene region extensible scope.

Details

This function reduces the dimension of one gene(or genomic region) with Fourier Expansion. Fist extract out the genotype of this gene with the gene annotation and the SNP map information. Then expanse the gene with the genotypes and SNP positions if the number of SNPs in the gene is over 3. Otherwise the raw genotypes of the gene would be returned. The number of Fourier Basis is selected to explain 80 percent of genetic variation.

Value

If the SNPs number of the gene is over 3, returns the expansion of the genotype which is a matrix with the dimension Sample number * Fourier Basis number. If the SNPs number of the gene is no more than 3, returns the raw genotype of the gene which is a matrix with the dimension Sample number * SNP number.

Author(s)

Futao Zhang

Examples

```
gLst<-read.csv(system.file("extdata", "gene.list.csv", package="FRGEpistasis"))
fdata<-read.table(system.file("extdata", "simGeno-chr1.raw", package="FRGEpistasis"),header=TRUE)
geno<-fdata[,-1:-6]
snp_map<-read.table(system.file("extdata", "chr1.map", package="FRGEpistasis"))
fourierExpansion(1, geno, gLst, snp_map, 0)
```

---

fRGEpistasis *Genome Wide Epistasis Study by Functional Regress Model*

Description

This function is the entrance of the software package. It tests the Genome Wide Epistasis by Functional Regression Model.

Usage

```
fRGEpistasis(wDir, phenoInfo, gnoFiles, mapFiles, gLst, fdr, rng)
```

Arguments

wDir	The dataset directory. If the dataset is in the working directory, wDir is ".".
phenoInfo	It is a matrix with two columns. One column is the individual ID and the other is the phenotype. The phenotype can be quantitative trait or binary trait.
gnoFiles	The vector of genotype file names. It contains the genotype file names indicating where to read the genotype files.
mapFiles	The vector of SNP genetic map file names. It contains the map file names indicating where to read the genetic map files.
gLst	Gene annotation which includes gene name, chromosome, start position and end position.
fdr	FDR control threshold, When this value == 1, turn FDR control off.
rng	A numeric value which represents gene region extensible scope.

Details

Firstly this package reduces the dimension of genotype of all the genomic regions. Secondly this package tests the epistasis of genomic regions both of which are on the same chromosome(file). Thirdly this package tests the epistasis of genomic regions which are on different chromosomes(files).

This function is memory efficient with high performance. Memory efficiency: Only store reduced expansion data of genotypes instead of raw data of genotypes. This package reduces the dimension of genotype of all the genomic regions(see details of function "reduceGeno"). In real dataset the genotypes on different chromosome are always organized into different files. And each genotype file is very large. Reading all the files into memory is unacceptable. This package reads the files one by one and reduces the genotype dimension with Fourier expansion. In order to inform the package how many files and where to read, we need two data structures "gnoFiles" and "mapFiles" to store the file names.

high performance: Each data file only needs to read once and reduce dimension once. So I/O times are reduced and repeated computing of data reduction was avoided. This method is a kind of group test. We take a gene(or genomic region) as the test unit. The number of Test is sharply reduced comparing with point-wise interaction (SNP-SNP) test. The dimension of genotype is reduced by functional expansion, So the time of each test is reduced.

Value

Return a data frame which contains all the names of gene pairs and the p values of chi-square test for their epistasis.

Author(s)

Futao Zhang

Examples

```
wDir <-paste(system.file("extdata", package="FRGEpistasis"),"/",sep="")
gnoFiles<-read.table(system.file("extdata", "list_geno.txt", package="FRGEpistasis"))
mapFiles<-read.table(system.file("extdata", "list_map.txt", package="FRGEpistasis"))
phenoInfo <- read.csv(system.file("extdata", "phenotype.csv", package="FRGEpistasis"),header=TRUE)
```

```
gLst<-read.csv(system.file("extdata", "gene.list.csv", package="FRGEpistasis"))
rng=0
fdr=0.05
out_epi <- data.frame( )
phenoInfo [,2]=log(phenoInfo [,2])
out_epi = fRGEpistasis(wDir,phenoInfo,gnoFiles,mapFiles,gLst,fdr,rng)
```

`frgEpistasisTest` *Epistasis Test by Functional Regression Model*

Description

This function is used to analyse the epistasis of genomic region A and genomic region B.

Usage

```
frgEpistasisTest(pheno, geno_A, pos_A, geno_B, pos_B)
```

Arguments

`pheno`	A vector of phenotype which can be quantitative trait or binary trait.
`geno_A`	Genotype matrix of gene (or genomic region) A.
`pos_A`	Vector of physical positions of SNPs in gene (or genomic region) A.
`geno_B`	Genotype matrix of gene (or genomic region) B.
`pos_B`	Vector of physical positions of SNPs in gene (or genomic region) B.

Details

This function is independent with other functions in this package. It is designed for small dataset test. It takes phenotype, genotype and Physical positions as the input. If the position information is NULL, this function considers the SNPs in this gene to be uniformly filled in the gene scope. First this function expanses the genotypes of gene A and gene B. Then it analyses their epistasis.

Value

It returns the p value of chi-square test for epistasis detection between gene A and gene B.

Author(s)

Futao Zhang

Examples

```
smp_num=1000
number_snp_A=25
number_snp_B=20
pheno<-sample(c(0:500),smp_num,replace=TRUE)
smpl=rep(0,number_snp_A*smp_num)
idx_1=sample(c(1:(number_snp_A*smp_num)),ceiling(number_snp_A*smp_num/100))
idx_2=sample(c(1:(number_snp_A*smp_num)),ceiling(number_snp_A*smp_num/200))
smpl[idx_1]=1
smpl[idx_2]=2
geno_A=matrix(smpl,smp_num,number_snp_A)

smpl=sample(c(0,1,2),number_snp_B*smp_num,replace=TRUE)
geno_B=matrix(smpl,smp_num,number_snp_B)

frgEpistasisTest(pheno,geno_A,pos_A=NULL,geno_B,pos_B=NULL)
```

---

fRGInteraction *Interaction Test by Functional Regression Model*

Description

Test interaction between two gene (or genomic regions) with chi-squared test.

Usage

```
fRGInteraction(phenoData, x_A, x_B)
```

Arguments

phenoData	Vector of phenotype data which can be quantitative trait or binary trait.
x_A	Expansion data matrix of Genotype of gene A.
x_B	Expansion data matrix of Genotype of gene B.

Details

This function takes phenotype vector and expansed genotype matrices as input. It is the most important part of this software package. It is called by functions "innerEpi" and "innerEpi" of this package. The interaction between gene A and gene B is tested with chi-squared test.

Value

It returns the p value of chi-squared test for epistasis detection between gene A and gene B.

Author(s)

Futao Zhang

Examples

```
x_A<-as.matrix(rnorm(1000,mean=0,sd=1))
x_B<-as.matrix(rnorm(1000,mean=0,sd=1))
phenoData<-runif(1000,15,60)
fRGInteraction(phenoData,x_A,x_B)
```

---

innerEpi *Epistasis Detection Inner one Chromosome*

Description

Detect epistasis between 2 genes (or genomic regions) both of which are on the same chromosome.

Usage

```
innerEpi(pheno, gstnd, geno_expn, gname, gchr)
```

Arguments

pheno	Vector of phenotype data which can be quantitative trait or binary trait.
gstnd	Vector of indexes which indicates the start indexes and end indexes of expansed genotype of each gene on current chromosome in the matrix "geno_expn".
geno_expn	Matrix of expansed genotype data of all the genes.
gname	Vector of gene names on current chromosome.
gchr	Vector of Chromosome number of current chromosome.

Details

This function tests the epistasis between 2 genes both of which are on the same chromosome. It takes expansed genotype data as the input. First the data of the gene are extracted from "geno_expn" with "gstnd" and "gname". Then the function "fRGInteraction" will be called.

Value

Return a matrix which contains the gene names of the gene pairs and the p values of chi-squared test for the epistasis of the gene pairs.

Author(s)

Futao Zhang

Examples

```
smp_num<-1000
number_basis<-9
pheno<-sample(c(0:500),smp_num,replace=TRUE)
gname<-c("g1","g2")
gstnd<-c(0,5,9)
smpl<-runif(number_basis*smp_num, 0.0, 1.0)
geno_expn<-matrix(smpl,smp_num,number_basis)
gchr<-c(1,1)
innerEpi(pheno,gstnd,geno_expn,gname,gchr)
```

---

innerSnpListInteraction

Pairwise Interaction Test Inner The Same SNP List

Description

Test the SNP-SNP interaction. And the SNPs are organized into one data structure.

Usage

```
innerSnpListInteraction(pheno, snpList)
```

Arguments

pheno	Vector of phenotype data.
snpList	Matrix of the genotypes of all the SNPs for testing the pairwise interactions.

Details

This function aims to test the pairwise interactions between the SNPs organized into the same data structure. It takes phenotype and genotypes of the SNPs as the input. And output all the p values for the interactions of SNP pairs.

Value

Return a frame contains names of all the SNPs pairs and p values for interactions of these pairs.

Author(s)

Futao Zhang

Examples

```
pheno<- round(runif(1000,40,60))
geno<- as.data.frame(matrix(round(runif(5000,0,2)),1000,5))
innerSnpListInteraction(pheno,geno)
```

`logTransPheno` *logarithmic transformation*

Description

Logarithmic Transformation of Phenotype

Usage

```
logTransPheno(pheno)
```

Arguments

pheno	Vector of phenotype which is the quantitative trait.

Details

Some variables are not normally distributed. And using statistical tests on this data can give misleading results because they do not meet the statistical assumptions. Many variables have log-normal distributions.

Value

Return vector of transformed phenotype.

Examples

```
smp_num=100
pheno<-sample(c(0:500),smp_num,replace=TRUE)
logTransPheno(pheno)
```

`outerEpi` *Epistasis Detection Outer Chromosomes*

Description

Detect epistasis between 2 genes (or genomic regions) which are on different chromosomes.

Usage

```
outerEpi(pheno, gstnd, gStndp, geno_expn, gname, gNamep, gchr, gChrp)
```

Arguments

`pheno`	Vector of phenotype data which can be quantitative trait or binary trait.
`gstnd`	Vector of indexes which indicates the start indexes and end indexes of expansed genotype of each gene on one chromosome in the matrix "geno_expn".
`gStndp`	Vector of indexes which indicates the start indexes and end indexes of expansed genotype of each gene on the other chromosome in the matrix "geno_expn".
`geno_expn`	Matrix of expansed genotype data of all the genes.
`gname`	Vector of gene names on one chromosome.
`gNamep`	Vector of gene names on the other chromosome.
`gchr`	Vector of Chromosome number of one chromosome.
`gChrp`	Vector of Chromosome number of the other chromosome.

Details

This function tests the epistasis between 2 genes which are on different chromosomes. It takes expansed genotype data as the input. First the data of the gene are extracted from "geno_expn" with "gstnd" and "gname". Then the function "fRGInteraction" will be called.

Value

Return a matrix which contains the gene names of the gene pairs and the p values of chi-squared test for the epistasis of the gene pairs.

Author(s)

Futao Zhang

Examples

```
smp_num=1000
number_basis<-40
pheno<-sample(c(0:500),smp_num,replace=TRUE)
gname<-c("g1","g2")
gNamep<-c("r1","r2","r3")
gstnd<-c(0,5,9)
gStndp<-c(16,23,29,36)
smpl<-runif(number_basis*smp_num, 0.0, 1.0)
geno_expn<-matrix(smpl,smp_num,number_basis)
gchr<-c(1,1)
gchrp<-c(3,3,3)
outerEpi(pheno,gstnd,gStndp,geno_expn,gname,gNamep,gchr,gchrp)
```

---

outerSnpListInteraction

Pairwise Interaction Test Outer The SNP Lists

Description

Test the SNP-SNP interaction. And the SNPs are organized into two different SNP Lists.

Usage

```
outerSnpListInteraction(pheno, snpList1, snpList2)
```

Arguments

pheno	Vector of phenotype data.
snpList1	Matrix of the genotypes of all the SNPs on the first SNP list for testing the pairwise interactions.
snpList2	Matrix of the genotypes of all the SNPs on the second SNP list for testing the pairwise interactions.

Details

This function aims to test the pairwise interactions between the SNPs organized into different data structures. It takes phenotype and genotypes of the SNPs as the input. And output all the p values for the interactions of SNP pairs.

Value

Return a frame contains names of all the SNPs pairs and p values for interactions of these pairs.

Author(s)

Futao Zhang

Examples

```
snp_list_1 <- as.data.frame(matrix(round(runif(3000,0,2)),1000,3))
snp_list_2 <- as.data.frame(matrix(round(runif(5000,0,2)),1000,5))
colnames(snp_list_1 )<-c("rs10","rs11","rs12")
colnames(snp_list_2 )<-c("rs20","rs21","rs22","rs23","rs24")
pheno<- round(runif(1000,40,60))
outerSnpListInteraction(pheno,snp_list_1,snp_list_2)
```

---

pCAInteraction *Epistasis Test by Principal Component Analysis*

Description

Test the epistasis between two genes (or genomic regions) with the principal components analysis method.

Usage

```
pCAInteraction(phenoData, x_A, x_B)
```

Arguments

phenoData	Vector of phenotype data which can be quantitative trait or binary trait.
x_A	Matrix of genotype of gene A.
x_B	Matrix of genotype of gene B.

Details

This function takes phenotype vector and genotype matrices as input and tests the epistasis using PCA method. The number of principal components is determined by PCA to explain 80 percent of the genetic variation. The interaction between gene A and gene B is tested with chi-squared test.

Value

It returns the p value of chi-squared test for epistasis detection between gene A and gene B.

Author(s)

Futao Zhang

Examples

```
smp_num=1000
number_snp_A=25
number_snp_B=20
pheno<-sample(c(0:500),smp_num,replace=TRUE)
smpl=rep(0,number_snp_A*smp_num)
idx_1=sample(c(1:(number_snp_A*smp_num)),ceiling(number_snp_A*smp_num/100))
idx_2=sample(c(1:(number_snp_A*smp_num)),ceiling(number_snp_A*smp_num/200))
smpl[idx_1]=1
smpl[idx_2]=2
geno_A=matrix(smpl,smp_num,number_snp_A)

smpl=sample(c(0,1,2),number_snp_B*smp_num,replace=TRUE)
geno_B=matrix(smpl,smp_num,number_snp_B)
pCAInteraction(pheno,geno_A,geno_B)
```

---

pCAPiontwiseEpistasis *Epistasis Test by PCA Method and Piontwise Method*

Description

This function is the another entrance of the software package. It tests the Genome Wide Epistasis by PCA Method and Piontwise Method.

Usage

```
pCAPiontwiseEpistasis(wDir, oEpi, phenoInfo, gnoFiles, mapFiles, gLst, rng)
```

Arguments

wDir	The dataset directory. If the dataset is in the working directory, wDir is ".".
oEpi	Output data frame which contains all the names of gene pairs and the p values for their epistasis.
phenoInfo	It is a matrix with two columns. One column is the individual ID and the other is the phenotype. The phenotype can be quantitative trait or binary trait.
gnoFiles	The vector of genotype file names. It contains the genotype file names indicating where to read the genotype files.
mapFiles	The vector of SNP genetic map file names. It contains the map file names indicating where to read the genetic map files.
gLst	Gene annotation which includes gene name, chromosome, start position and end position.
rng	A numeric value which represents gene region extensible scope.

Details

The genotypes on different chromosome are stored in different files. The full names of these files are listed in the index file that is taken as the input parameter. After the index file is loaded, the function knows where to read the genotypes files. This function analyses the genotypes files one by another. That means this function tests the epistasis of genomic regions both of which are on the same chromosome(file), then the epistasis of genomic regions which are on different chromosomes(files). This function can test epistasis both with PCA method and pointwise method. For a pair of genes, we assume that the total number of all possible SNP pairs is K, The minum p value for SNP-SNP interaction among the K pairs is output as the pointwise method result of the gene pair.

Value

Return a data frame which contains all the names of gene pairs and the p values of chi-square test for their epistasis.

Author(s)

Futao Zhang

Examples

```
work_dir <-paste(system.file("extdata", package="FRGEpistasis"),"/",sep="")
##read the list of genotype files
geno_files<-read.table(system.file("extdata", "list_geno.txt", package="FRGEpistasis"))
##read the list of map files
mapFiles<-read.table(system.file("extdata", "list_map.txt", package="FRGEpistasis"))
##read the phenotype file
phenoInfo <- read.csv(system.file("extdata", "phenotype.csv", package="FRGEpistasis"),header=TRUE)
##read the gene annotation file
gLst<-read.csv(system.file("extdata", "gene.list.csv", package="FRGEpistasis"))
##define the extension scope of gene region
rng=0

##log transformation
phenoInfo [,2]=log(phenoInfo[,2])
out_epi<-data.frame()

pCAPiontwiseEpistasis(work_dir,out_epi,phenoInfo,geno_files,mapFiles,gLst,rng)
```

---

pointwiseInteraction *Pointwise Interaction Test*

Description

Test the epistasis of the gene pair by pointwise method

Usage

```
pointwiseInteraction(phenoData, x_A, x_B)
```

Arguments

phenoData	Vector of phenotype data which can be quantitative trait or binary trait.
x_A	Matrix of genotype of gene A.
x_B	Matrix of genotype of gene B.

Details

This function takes phenotype vector and genotype matrices as input and tests the epistasis using pointwise method. For a pair of genes, we assume that the total number of all possible SNP pairs is K (one SNP from one gene and the other SNP from the other gene). The interaction of each SNP pair between the two genes is tested. The minum p value for SNP-SNP interaction among the K pairs is output as the pointwise method result of the gene pair.

Value

Return the minum p value for SNP-SNP interaction among the K pairs

Author(s)

Futao Zhang

Examples

```
smp_num=1000
number_snp_A=25
number_snp_B=20
pheno<-sample(c(0:500),smp_num,replace=TRUE)
smpl=rep(0,number_snp_A*smp_num)
idx_1=sample(c(1:(number_snp_A*smp_num)),ceiling(number_snp_A*smp_num/100))
idx_2=sample(c(1:(number_snp_A*smp_num)),ceiling(number_snp_A*smp_num/200))
smpl[idx_1]=1
smpl[idx_2]=2
geno_A=matrix(smpl,smp_num,number_snp_A)

smpl=sample(c(0,1,2),number_snp_B*smp_num,replace=TRUE)
geno_B=matrix(smpl,smp_num,number_snp_B)
pointwiseInteraction(pheno,geno_A,geno_B)
```

rankTransPheno *Rank-Based Inverse Normal Transformation*

Description

Rank-Based Inverse Normal Transformation of Phenotype

Usage

```
rankTransPheno(pheno, para_c)
```

Arguments

pheno	Vector of phenotype which is the quantitative trait.
para_c	Adjust parameter, commonly as 0,1/3,3/8 or 1/2.

Details

Some variables are not normally distributed. And using statistical tests on this data can give misleading results because they do not meet the statistical assumptions. This function implements Rank-Based Inverse Normal Transformation to make phenotype normally distributed.

Value

Return vector of rank-based inverse normal transformed phenotype.

Author(s)

Futao Zhang

References

T. Mark Beasley, Stephen Erickson and David B. Allison. Rank-Based Inverse Normal Transformations are Increasingly Used, But are They Merited? Behav Genet. 2009 Sep.;39(5):580-595.

Examples

```
c=0.5
smp_num=100
pheno<-sample(c(0:500),smp_num,replace=TRUE)
rankTransPheno(pheno,c)
```

`reduceGeno` *Reduction Dimension of Genotype*

Description

Reduce Dimension of Genotype using Functional Regression Model

Usage

```
reduceGeno(wDir, pheno, gnoFiles, mapFiles, gLst, rng)
```

Arguments

`wDir`	The dataset directory. If the dataset is in the working directory, wDir is ".".
`pheno`	It is a matrix with two columns. One column is the individual ID and the other is the phenotype. The phenotype can be quantitative trait or binary trait.
`gnoFiles`	The vector of genotype file names. It contains the genotype file names indicating where to read the genotype files.
`mapFiles`	The vector of SNP genetic map file names. It contains the map file names indicating where to read the genetic map files.
`gLst`	Gene annotation which includes gene name, chromosome, start position and end position.
`rng`	A numeric value which represents gene region extensible scope.

Details

This function reduces the dimension of genotypes of all the genes with Fourier expansion. In real dataset the genotypes on different chromosome are always organized into different files. And each genotype file is very large. This function processes the genotype files in turns. The reduced genotype data (the expansion data) of all the chromosomes are combined together. The expansion data and other information are organized into a list. During Fourier expansion, the physical position information of the SNPs are used. This is one of merits of our method.

Value

Return a list that includes reduced genotype data, gene names,chromosome information, start index and end index of each gene.

Author(s)

Futao Zhang

Examples

```
wDir <-paste(system.file("extdata", package="FRGEpistasis"),"/",sep="")
gnoFiles<-read.table(system.file("extdata", "list_geno.txt", package="FRGEpistasis"))
mapFiles<-read.table(system.file("extdata", "list_map.txt", package="FRGEpistasis"))
phenoInfo <- read.csv(system.file("extdata", "phenotype.csv", package="FRGEpistasis"),header=TRUE)
gLst<-read.csv(system.file("extdata", "gene.list.csv", package="FRGEpistasis"))
rng=0
gInfo=reduceGeno(wDir,phenoInfo,gnoFiles,mapFiles,gLst,rng)
```

---

snpPairInteraction *SNP-SNP interaction*

Description

Test the interaction of one SNP with another

Usage

```
snpPairInteraction(pheno, snp1, snp2)
```

Arguments

pheno	Vector of phenotype data which can be quantitative trait or binary trait.
snp1	Vector of genotype data of SNP1.
snp2	Vector of genotype data of SNP2.

Details

This function tests the interaction of one SNP with another.

Value

Return the p value for snp-snp interaction

Author(s)

Futao Zhang

Examples

```
pheno<- round(runif(1000,40,60))
snp1<-round(runif(1000,0,2))
snp2<-round(runif(1000,0,2))
pval=snpPairInteraction(pheno,snp1,snp2)
```

Index