Package 'RVFam'

March 10, 2015

Version 1.1

Date 2015-02-06

Title Rare Variants Association Analyses with Family Data

Author Ming-Huei Chen <mhchen@bu.edu> and Qiong Yang <qyang@bu.edu>

Maintainer Ming-Huei Chen <mhchen@bu.edu>

Depends R (>= 3.0.0)

Imports coxme, survival, lme4, kinship2, MASS, Matrix

LazyData true

LazyDataCompression gzip

Description The RVFam package provides functions to perform single SNP association analyses and gene-based tests for continuous, binary and survival traits against sequencing data (e.g. exome chip) using family data.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2015-03-10 19:41:17

R topics documented:

RVFam-package *Rare Variants Association Analyses with Family Data*

Description

RVFam package provides functions to perform single SNP association analyses and gene-based tests for continuos, binary and survival traits against sequencing variant genotypes (e.g. exome chip and whole genome sequencing data) using family data. The gene-based tests include two burden tests, most powerful when effects are in the same direction across all included variants (Li and Leal 2008 and Madsen and Browning 2009), and one directional insensitive test (Wei 2009). For single SNP association analyses of continuous traits, RVFam fits linear mixed effects (LME) model with relationship coefficcient matrix as within pedigree correlation matrix to account for familial correlation; for binary traits, RVFam fits generalized linear mixed effects (GLMM) model that treats each pedigree as a cluster; while for survival traits, RVFam fits Cox proportional hazards regression model (COXPH) with frailty that adds a random effect for family clusters.

Details

Package:	RVFam
Type:	Package
Version:	1.0
Date:	2015-02-06
License:	GPL (>= 2)

Author(s)

Qiong Yang <qyang@bu.edu> and Ming-Huei Chen <mhchen@bu.edu>

Maintainer: Ming-Huei Chen <mhchen@bu.edu>

coxph.EC *function for testing a single/pooled variant for survival traits with family data using Cox proportional hazards regression model*

Description

Fit Cox proportional hazards regression model to test a single/pooled variant for associations against a survival phenotype with family data. The coxph function from package survival is used.

Usage

```
coxph.EC(snp,phen,test.dat,covar,chr,time)
```

Arguments

snp	a numeric vector with genotype of a single/pooled variant
phen	a character string for the phenotype name of a continuous trait of interest in test.dat
test.dat	the product of merging phenotype, genotype and pedigree data
covar	a character vector for covariates in test.dat
chr	chromosome number
time	the character string of variable named for survival time

Details

The coxph.EC function fits a Cox proportional hazards regression model with shared frailty (random effect) in each pedigree to test association between a survival phenotype and a single/pooled genetic variant with additive model. The trait-SNP association test is carried out by the coxph function from package survival. P-value from likelihood ratio test (LRT) is reported. This function is called in coxph.ped function to test all single/pooled variants.

Value

ntotal	number of individuals with genotype, phenotype and covariates
nmiss	number of individuals with missing genotype among ntotal
maf_ntotal	minor allele frequency based on ntotal
beta	regression coefficient of single SNP test or burden test
se	standard error of beta
Z	Z statistic based on signed LRT
remark	additional information of the analysis
p	LRT p-value of a single variant test or burden test
MAC	minor allele count
n0	the number of individuals with 0 copy of coded alleles
n1	the number of individuals with 1 copy of coded alleles
n2	the number of individuals with 2 copies of coded alleles

Author(s)

Ming-Huei Chen <mhchen@bu.edu> and Qiong Yang <qyang@bu.edu>

References

Therneau T (2014). A Package for Survival Analysis in S. R package version 2.37-7, http://CRAN.R-project.org/package=survival.

Terry M. Therneau and Patricia M. Grambsch (2000). Modeling Survival Data: Extending the Cox Model. Springer, New York. ISBN 0-387-98784-3.

Examples

```
## Not run:
coxph.EC(snp=rsnps.dat[,"snp1"],snp1,phen="trait2",test.dat=rsnps.dat,
covar=c("age","sex"),chr=1,time="survival_time")
## End(Not run)
```

---

coxph.ped *function of single SNP analysis and gene-based tests for survival traits with family data using Cox proportional hazards regression model*

Description

Fit Cox proportional hazards regression model with shared frailty (random effect) in each pedigree for single SNP analysis that tests associations between a survival phenotype and each genotyped SNP on a chromosome in a genotype file and for gene-based tests in family data. The association test is carried out by `coxph.EC` function. Likelihood ratio test (LRT) result is reported. In each test, the `coxph` function from package `survival` is used.

Usage

```
coxph.ped(phenfile,phen,covars=NULL,mafRange=c(0,0.05),chr,genfile,
pedfile,snpinfoRdata,sep.ped=",",sep.phe=",",sep.gen=" ",time,
aggregateBy="SKATgene",maf.file,snp.cor,ssq.beta.wts=c(1,25),
singleSNP.outfile=F)
```

Arguments

`genfile`	a character string naming the genotype file for reading
`phenfile`	a character string naming the phenotype file for reading
`pedfile`	a character string naming the pedigree file for reading
`phen`	a character string for the phenotype name of a survival trait of interest in `test.dat`
`covars`	a character vector for covariates in `phenfile`
`sep.ped`	the field separator character for pedigree file
`sep.phe`	the field separator character for phenotype file
`sep.gen`	the field separator character for genotype file
`time`	the character string of variable named for survival time
`mafRange`	range of MAF to include SNPs for gene-based burden tests, default is c(0,0.05)
`chr`	chromosome number that can be 1,2,...,22, and 'X'
`snpinfoRdata`	a character string naming the RData containing SNP info to be loaded, this should at least include 'Name' (for SNP name), 'Chr', and aggregateBy (default='SKATgene') columns
`aggregateBy`	the column of SNP info on which single SNPs are to be aggregated for burden tests, default is 'SKATgene'

maf.file	a character string naming the comma delimited file containing 'Name' for SNP name and 'maf' for MAF
snp.cor	a character string naming the RData containing lists of SNP correlation matrix within each 'SKATgene'
ssq.beta.wts	a vector of parameters of beta weights used in proposed sum of squares test, default=c(1,25) as in SKAT
singleSNP.outfile	a logical value, TRUE indicating single SNP analysis has been done and result files are available for computing SSQ using a different mafRange

Details

The coxph.ped function reads in and merges phenotype, genotype, and pedigree files to perform single SNP analysis, two burden tests (weight=1 for Li & Leal 2008; weight=1/(MAF)/(1-MAF) for Madsen & Browning 2009), and one sum of squares (SSQ) test (Wei 2009) using Cox proportional hazards regression model with shared frailty (random effect) in each family as implemented in coxph function in survival R package and to output an RData that is computed based on single SNP results and that is compatible with seqMeta R package for conducting meta-analysis. For burden tests and SSQ test, SNPs genotypes/results are aggregated by aggregateBy (default = "SKATgene") using SNPs selected according to user specified mafRange within each gene (by default). genfile contains unique individual numerical id and genotype data on a chromosome, with the column names being "id" and SNP names. For each SNP, the genotype data should be coded as 0, 1, 2 indicating the numbers of the coded alleles. The SNP name in genotype file should not have any dash, '-' and other special characters(dots and underscores are OK). phenfile contains unique individual id, phenotype and covariates data, with the column names being "id" and phenotype and covaraite names. pedfile contains pedigree informaion, with the column names being "famid","id","fa","mo","sex". LRT is used in all genetic association tests.

Value

No value is returned. Instead, tab delimited result files and an RData are generated. A single SNP result file, named with phen and singleSNP, contains columns: gene, Name, maf, ntotal, nmiss, maf_ntotal, beta, se, Z, remark, p (p-value from LRT), MAC, n0, n1, and n2. A burden test result file, named with phen and T/MB for Li & Leal 2008/Madsen & Browning 2009 respectively, contains columns: gene, beta, se, Z, cmafTotal, cmafUsed, nsnpsTotal, nsnpsUsed, nmiss, remark, and p. A SSQ test result file, named with phen and SSQ, contains columns: gene, SSQ, cmafTotal, cmafUsed, nsnpsTotal, nsnpsUsed, nmiss, df, and p. A generated RData that is a list that contains scores, cov, n, maf and sey for each gene with gene names being the names of the list. Note maf in RData is MAF based on ntotal.

gene	gene name
Name	SNP name
maf	minor allele frequency based on genotyped sample
ntotal	number of individuals with genotype, phenotype and covariates
nmiss	number of individuals with missing genotype among ntotal
maf_ntotal	minor allele frequency based on ntotal
beta	regression coefficient of single SNP test or burden test

se	standard error of beta
Z	signed likelihood ratio statistic
remark	additional information of the analysis
p	p-value of single SNP test or burden test by LRT
camfTotal	sum of maf_ntotal of SNPs in a gene
cmafUsed	sum of maf_ntotal of SNPs selected with mafRange in a gene for burden tests or SSQ test
nsnpsTotal	total number of SNPs in a gene
nsnpsUsed	number of SNPs selected and used in burden tests and SSQ test
SSQ	sum of squares statistic
df	degrees of freedom of SSQ
MAC	minor allele count
n0	the number of individuals with 0 copy of coded alleles
n1	the number of individuals with 1 copy of coded alleles
n2	the number of individuals with 2 copies of coded alleles
scores	beta/se^2 in output RData, where beta and se are vectors
cov	diag(1/se)*LD matrix*diag(1/se) in output RData
n	maximum ntotal in a gene in output RData
sey	1 in output RData

Author(s)

Ming-Huei Chen <mhchen@bu.edu> and Qiong Yang <qyang@bu.edu>

References

Therneau T (2014). A Package for Survival Analysis in S. R package version 2.37-7, http://CRAN.R-project.org/package=survival.

Terry M. Therneau and Patricia M. Grambsch (2000). Modeling Survival Data: Extending the Cox Model. Springer, New York. ISBN 0-387-98784-3.

Li, B. and Leal, S. M (2008). Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. *Am J Hum Genet*, **83(3)**, 311-321.

Madsen, B. E. and Browning, S. R (2009). A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. *PLoS Genet*, **5(2)** e1000384.

Wei P (2009). Asymptotic Tests of Association with Multiple SNPs in Linkage Disequilibrium. *Genet Epidemiol*, **33(6)**, 497-507.

Examples

```
## Not run:
coxph.ped(genfile="EC_chr1.txt",phenfile="trait1.csv",pedfile="ped.csv",
phen="trait1",covars=NULL,sep.ped=",",sep.phe=",",sep.gen=" ",
mafRange=c(0,0.01),chr=1,snpinfoRdata="SNPinfo_EC.RData",
aggregateBy="SKATgene",time="survival_time",maf.file="EC_MAF.csv",
snp.cor="EC_SNPcor.RData")

## End(Not run)
```

`gc.fun` *function that does genomic control correction to single SNP analysis, sum of square test and RData for survival trait analysis*

Description

When high genomic control (GC) parameter (lambda) estimate is observed, `gc.fun` applies GC correction to SNPs with minor allele counts (`MAC`) less than a user specified threshold that may have inflated type I error rate for survival traits in particular, adjusts RData output accordingly, and recomputes sum of square statistic.

Usage

```
gc.fun(path,phen,snpinfoRdata,snp.cor,mac,aggregateBy="SKATgene",
maf.file,mafRange,ssq.beta.wts=c(1,25))
```

Arguments

`path`	path to directory that saves all 23 tab delimited single SNP analysis result files
`phen`	a character string for the phenotype name of a trait of interest
`snpinfoRdata`	a character string naming the RData containing SNP info to be loaded, this should at least include 'Name' (for SNP name), 'Chr', and aggregateBy (default='SKATgene') columns
`snp.cor`	a character string naming the RData containing lists of SNP correlation matrix within each 'SKATgene'
`mac`	user specified MAC threshold for applying GC correction to SNPs with MAC under it
`aggregateBy`	the column of SNP info on which single SNPs are to be aggregated for burden tests, default is 'SKATgene'
`maf.file`	a character string naming the comma delimited file containing 'snp.names' for SNP name and 'maf' for MAF
`mafRange`	range of MAF to include SNPs for gene-based burden tests, default is c(0,0.05)
`ssq.beta.wts`	a vector of parameters of beta weights used in proposed sum of squares test, default=c(1,25) as in SKAT

Details

When high lambda is observed from survival trait single SNP analysis, the gc.fun function applies GC correction to SNPs with user defined MAC, adjusts RData output based on GC corrected single SNP analysis results, recomputes sum of squares statistic and then outputs corrected single SNP analysis results, SSQ analysis results and RData. Initial single SNP analysis result files are required and the input arguments should be identical to the ones used in initial analysis (except for path).

Value

No value is returned. Instead, tab delimited result files and an RData are generated. A single SNP result file, named with phen and singleSNP, contains columns: gene, Name, maf, ntotal, nmiss, maf_ntotal, beta, se, Z, remark, p (p-value from LRT), MAC, n0, n1, and n2. A SSQ test result file, named with phen and SSQ, contains columns: gene, SSQ, cmafTotal, cmafUsed, nsnpsTotal, nsnpsUsed, nmiss, df, and p. A generated RData that is a list that contains scores, cov, n, maf and sey for each gene with gene names being the names of the list. Note maf in RData is MAF based on ntotal.

Author(s)

Ming-Huei Chen <mhchen@bu.edu> and Qiong Yang <qyang@bu.edu>

Examples

```
## Not run:
gc.fun(path="/home/mhchen/",phen="trait1",mafRange=c(0,0.01),
snpinfoRdata="SNPinfo_EC.RData",aggregateBy="SKATgene",
maf.file="EC_MAF.csv",snp.cor="EC_SNPcor.RData",ssq.beta.wts=c(1,25))

## End(Not run)
```

---

glmm.binped *function of single SNP analysis and gene-based tests for binary traits with family data using generalized linear mixed effects model*

Description

Fit generalized linear mixed effects model (GLMM) with logistic link that treats each pedigree as a cluster for single SNP analysis that tests associations between a binary phenotype and each genotyped SNP on a chromosome in a genotype file and for gene-based tests in family data. The association test is carried out by glmm.EC function. In each test, the glmer function from package lme4 is used.

Usage

```
glmm.binped(phenfile,genfile,pedfile,phen,covars=NULL,
mafRange=c(0,0.05),chr,snpinfoRdata,sep.ped=",",sep.phe=",",
sep.gen=" ",aggregateBy="SKATgene",maf.file,
snp.cor,ssq.beta.wts=c(1,25),singleSNP.outfile=F)
```

Arguments

phenfile	a character string naming the phenotype file for reading
genfile	a character string naming the genotype file for reading
pedfile	a character string naming the pedigree file for reading
phen	a character string for the phenotype name of a binary trait of interest in test.dat
covars	a character vector for covariates in phenfile
mafRange	range of MAF to include SNPs for gene-based burden tests, default is c(0,0.05)
chr	chromosome number that can be 1,2,...,22, and 'X'
snpinfoRdata	a character string naming the RData containing SNP info to be loaded, this should at least include 'Name' (for SNP name), 'Chr', and aggregateBy (default='SKATgene') columns
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file
aggregateBy	the column of SNP info on which single SNPs are to be aggregated for burden tests, default is 'SKATgene'
maf.file	a character string naming the comma delimited file containing 'Name' for SNP name and 'maf' for MAF
snp.cor	a character string naming the RData containing lists of SNP correlation matrix within each 'SKATgene'
ssq.beta.wts	a vector of parameters of beta weights used in proposed sum of squares test, default=c(1,25) as in SKAT
singleSNP.outfile	a logical value, TRUE indicating single SNP analysis has been done and result files are available for computing SSQ using a different mafRange

Details

The glmm.binped function reads in and merges phenotype, genotype, and pedigree files to perform single SNP analysis, two burden tests (weight=1 for Li & Leal 2008; weight=1/(MAF)/(1-MAF) for Madsen & Browning 2009), and one sum of squares (SSQ) test (Wei 2009) using GLMM with logistic link that treats each pedigree as a cluster as implemented in glmer function in lme4 R package and to output an RData that is computed based on single SNP results and that is compatible with seqMeta for conducting meta-analysis. For burden tests and SSQ test, SNPs genotypes/results are aggregated by aggregateBy (default = "SKATgene") using SNPs selected according to user specified mafRange within each gene (by default). genfile contains unique individual numerical id and genotype data on a chromosome, with the column names being "id" and SNP names. For each SNP, the genotype data should be coded as 0, 1, 2 indicating the numbers of the coded alleles. The SNP name in genotype file should not have any dash, '-' and other special characters(dots and underscores are OK). phenfile contains unique individual id, phenotype and covariates data, with the column names being "id" and phenotype and covaraite names. pedfile contains pedigree informaion, with the column names being "famid","id","fa","mo","sex". Wald chi-square test is used in all genetic association tests.

Value

No value is returned. Instead, tab delimited result files and an RData are generated. A single SNP result file, named with phen and singleSNP, contains columns: gene, Name, maf, ntotal, nmiss, maf_ntotal, beta, se, Z, remark, p, MAC, n0, n1, and n2. A burden test result file, named with phen and T/MB for Li & Leal 2008/Madsen & Browning 2009 respectively, contains columns: gene, beta, se, Z, cmafTotal, cmafUsed, nsnpsTotal, nsnpsUsed, nmiss, remark, and p. A SSQ test result file, named with phen and SSQ, contains columns: gene, SSQ, cmafTotal, cmafUsed, nsnpsTotal, nsnpsUsed, nmiss, df, and p. A generated RData that is a list that contains scores, cov, n, maf and sey for each gene with gene names being the names of the list. Note maf in RData is MAF based on ntotal.

gene	gene name
Name	SNP name
maf	minor allele frequency based on genotyped sample
ntotal	number of individuals with genotype, phenotype and covariates
nmiss	number of individuals with missing genotype among ntotal
maf_ntotal	minor allele frequency based on ntotal
beta	regression coefficient of single SNP test or burden test
se	standard error of beta
Z	signed likelihood ratio statistic
remark	additional information of the analysis
p	p-value of single SNP test or burden test
camfTotal	sum of maf_ntotal of SNPs in a gene
cmafUsed	sum of maf_ntotal of SNPs selected with mafRange in a gene for burden tests or SSQ test
nsnpsTotal	total number of SNPs in a gene
nsnpsUsed	number of SNPs selected and used in burden tests and SSQ test
SSQ	sum of squares statistics
df	degree of freedom of SSQ
MAC	minor allele count
n0	the number of individuals with 0 copy of coded alleles
n1	the number of individuals with 1 copy of coded alleles
n2	the number of individuals with 2 copies of coded alleles
scores	beta/se^2 in output RData, where beta and se are vectors
cov	diag(1/se)*LD matrix*diag(1/se) in output RData
n	maximum ntotal in a gene in output RData
sey	1 in output RData

Author(s)

Ming-Huei Chen <mhchen@bu.edu> and Qiong Yang <qyang@bu.edu>

References

Bates D, Maechler M, Bolker B and Walker S (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, http://CRAN.R-project.org/package=lme4.

Li, B. and Leal, S. M (2008). Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. *Am J Hum Genet*, **83(3)**, 311-321.

Madsen, B. E. and Browning, S. R (2009). A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. *PLoS Genet*, **5(2)** e1000384.

Wei P (2009). Asymptotic Tests of Association with Multiple SNPs in Linkage Disequilibrium. *Genet Epidemiol*, **33(6)**, 497-507.

Examples

```
## Not run:
glmm.binped(genfile="EC_chr1.txt",phenfile="trait1.csv",pedfile="ped.csv",
phen="trait1",covars=c("age"),sep.ped=",",sep.phe=",",sep.gen=" ",
mafRange=c(0,0.01),chr=1,snpinfoRdata="SNPinfo_EC.RData",aggregateBy="SKATgene",
maf.file="EC_MAF.csv",snp.cor="EC_SNPcor.RData",ssq.beta.wts=c(1,25))

## End(Not run)
```

`glmm.EC` *function for testing a single/pooled variant for continuous traits with family data using generalized linear mixed effects model*

Description

Fit generalized linear mixed effects model (GLMM) with logistic link that treats each pedigree as a cluster to test a single/pooled variant for associations against a continuous phenotype with family data. The `glmer` function from package `lme4` is used.

Usage

```
glmm.EC(snp,phen,test.dat,covar,chr)
```

Arguments

`snp`	a numeric vector with genotype of a single/pooled variant
`phen`	a character string for the phenotype name of a binary trait of interest in `test.dat`
`test.dat`	the product of merging phenotype, genotype and pedigree data, should be ordered by "famid"
`covar`	a character vector for covariates in `test.dat`
`chr`	chromosome number

Details

The `glmm.EC` function fits a generalized linear mixed effects model (GLMM) with logistic link that treats each pedigree as a cluster to test association between a binary trait and a single/pooled genetic variant with additive model. The trait-variant association test is carried out by the `glmer` function from package `lme4`. P-value from likelihood ratio test (LRT) is reported. This function is called in `glmm.ped` function to test all single/pooled variants.

Value

`ntotal`	number of individuals with genotype, phenotype and covariates
`nmiss`	number of individuals with missing genotype among `ntotal`
`maf_ntotal`	minor allele frequency based on `ntotal`
`beta`	regression coefficient of single SNP test or burden test
`se`	standard error of `beta`
`Z`	Z statistic based on signed LRT
`remark`	additional information of the analysis
`p`	LRT p-value of a single variant test or burden test
`MAC`	minor allele count
`n0`	the number of individuals with 0 copy of coded alleles
`n1`	the number of individuals with 1 copy of coded alleles
`n2`	the number of individuals with 2 copies of coded alleles

Author(s)

Ming-Huei Chen <mhchen@bu.edu> and Qiong Yang <qyang@bu.edu>

References

Bates D, Maechler M, Bolker B and Walker S (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, http://CRAN.R-project.org/package=lme4.

Examples

```
## Not run:
glmm.EC(snp=rsnps.dat[,"snp1"],phen="disease",test.dat=rsnps.dat,
covar=c("sex"),chr=1)

## End(Not run)
```

`lme.EC` *function for testing a single/pooled variant for continuous traits with family data using Linear Mixed Effects model*

Description

Fit linear mixed effects (LME) model to test a single/pooled variant for associations against a continuous phenotype with family data. The `lmekin` function from package `coxme` is used.

Usage

```
lme.EC(snp,phen,test.dat,covar,kmat,chr)
```

Arguments

`snp`	a numeric vector with genotype of a single/pooled variant
`phen`	a character string for the phenotype name of a continuous trait of interest in `test.dat`
`test.dat`	the product of merging phenotype, genotype and pedigree data
`covar`	a character vector for covariates in `test.dat`
`kmat`	relationship coefficient (twice of kinship coefficient) matrix based on pedigree file
`chr`	chromosome number

Details

The `lme.EC` function fits a Linear Mixed Effects model (LME) that uses relationship coefficient matrix as within pedigree correlation matrix to test association between a continuous phenotype and a single/pooled genetic variant with additive model. The trait-SNP association test is carried out by the `lmekin` function from package `coxme`. This function is called in `lme.ped` function to test all single/pooled variants.

Value

`ntotal`	number of individuals with genotype, phenotype and covariates
`nmiss`	number of individuals with missing genotype among `ntotal`
`maf_ntotal`	minor allele frequency based on `ntotal`
`beta`	regression coefficient of single SNP test or burden test
`se`	standard error of `beta`
`Z`	Wald Z statistic
`remark`	additional information of the analysis
`p`	p-value of single SNP test or burden test
`MAC`	minor allele count

n0	the number of individuals with 0 copy of coded alleles
n1	the number of individuals with 1 copy of coded alleles
n2	the number of individuals with 2 copies of coded alleles

Author(s)

Ming-Huei Chen <mhchen@bu.edu> and Qiong Yang <qyang@bu.edu>

References

coxme package: mixed-effects Cox models, sparse matrices, and modeling data from large pedigrees. Beth Atkinson (atkinson@mayo.edu) for pedigree functions.Terry Therneau (therneau@mayo.edu) for all other functions. 2007. Ref Type: Computer Program. http://cran.r-project.org/web/packages/coxme/.

Abecasis, G. R., Cardon, L. R., Cookson, W. O., Sham, P. C., & Cherny, S. S (2001). Association analysis in a variance components framework. *Genet Epidemiol*, **21** Suppl 1, S341-S346.

Examples

```
## Not run:
lme.EC(snp=rsnps.dat$counts,phen="trait",test.dat=rsnps.dat,
covar=c("age","sex"),kmat=kmat,chr=1)
## End(Not run)
```

---

lme.ped *function of single SNP analysis and gene-based tests for continuous traits with family data using Linear Mixed Effects model*

Description

Fit linear mixed effects (LME) model for single SNP analysis that tests associations between a continuous phenotype and each genotyped SNP on a chromosome in a genotype file and for gene-based tests in family data. The association test is carried out by `lme.EC` function. In each test, the `lmekin` function from package `coxme` is used.

Usage

```
lme.ped(phenfile,genfile,pedfile,phen,covars=NULL,mafRange=c(0,0.05),chr,
snpinfoRdata,sep.ped=",",sep.phe=",",sep.gen=" ",aggregateBy="SKATgene",
maf.file,snp.cor,ssq.beta.wts=c(1,25),singleSNP.outfile=F)
```

Arguments

phenfile	a character string naming the phenotype file for reading
genfile	a character string naming the genotype file for reading
pedfile	a character string naming the pedigree file for reading
phen	a character string for the phenotype name of a continuous trait of interest in `test.dat`

covars	a character vector for covariates in phenfile
mafRange	range of MAF to include SNPs for gene-based burden tests, default is c(0,0.05)
chr	chromosome number that can be 1,2,...,22, and 'X'
snpinfoRdata	a character string naming the RData containing SNP info to be loaded, this should at least include 'Name' (for SNP name), 'Chr', and aggregateBy (default='SKATgene') columns
sep.ped	the field separator character for pedigree file
sep.phe	the field separator character for phenotype file
sep.gen	the field separator character for genotype file
aggregateBy	the column of SNP info on which single SNPs are to be aggregated for burden tests, default is 'SKATgene'
maf.file	a character string naming the comma delimited file containing 'Name' for SNP name and 'maf' for MAF
snp.cor	a character string naming the RData containing lists of SNP correlation matrix within each 'SKATgene'
ssq.beta.wts	a vector of parameters of beta weights used in proposed sum of squares test, default=c(1,25) as in SKAT
singleSNP.outfile	a logical value, TRUE indicating single SNP analysis has been done and result files are available for computing SSQ using a different mafRange

Details

The lme.ped function reads in and merges phenotype, genotype, and pedigree files, and creates a relationship coefficient matrix using pedfile and kinship2 package to perform single SNP analysis, two burden tests (weight=1 for Li & Leal 2008; weight=1/(MAF)/(1-MAF) for Madsen & Browning 2009), one sum of squares (SSQ) test (Wei 2009) using a LME model as implemented in lmekin function in coxme R package and to output an RData that is computed based on single SNP results and that is compatible with seqMeta for conducting meta-analysis. For burden tests and SSQ test, SNPs genotypes/results are aggregated by aggregateBy (default = "SKATgene") using SNPs selected according to user specified mafRange within each gene (by default). genfile contains unique individual numerical id and genotype data on a chromosome, with the column names being "id" and SNP names. For each SNP, the genotype data should be coded as 0, 1, 2 indicating the numbers of the coded alleles. The SNP name in genotype file should not have any dash, '-' and other special characters(dots and underscores are OK). phenfile contains unique individual id, phenotype and covariates data, with the column names being "id" and phenotype and covaraite names. pedfile contains pedigree informaion, with the column names being "famid","id","fa","mo","sex". Wald chi-square test is used in all genetic association tests.

Value

No value is returned. Instead, tab delimited result files and an RData are generated. A single SNP result file, named with phen and singleSNP, contains columns: gene, Name, maf, ntotal, nmiss, maf_ntotal, beta, se, Z, remark, p (p-value from LRT), MAC, n0, n1, and n2. A burden test result file, named with phen and T/MB for Li & Leal 2008/Madsen & Browning 2009 respectively, contains columns: gene, beta, se, Z, cmafTotal, cmafUsed, nsnpsTotal, nsnpsUsed, nmiss,

remark, and p. A SSQ test result file, named with phen and SSQ, contains columns: gene, SSQ, cmafTotal, cmafUsed, nsnpsTotal, nsnpsUsed, nmiss, df, and p. A generated RData that is a list that contains scores, cov, n, maf and sey for each gene with gene names being the names of the list. Note maf in RData is MAF based on ntotal.

gene	gene name
Name	SNP name
maf	minor allele frequency based on genotyped sample
ntotal	number of individuals with genotype, phenotype and covariates
nmiss	number of individuals with missing genotype among ntotal
maf_ntotal	minor allele frequency based on ntotal
beta	regression coefficient of single SNP test or burden test
se	standard error of beta
Z	Wald Z statistic
remark	additional information of the analysis
p	p-value of single SNP test or burden test
camfTotal	sum of maf_ntotal of SNPs in a gene
cmafUsed	sum of maf_ntotal of SNPs selected with mafRange in a gene for burden tests or SSQ test
nsnpsTotal	total number of SNPs in a gene
nsnpsUsed	number of SNPs selected and used in burden tests and SSQ test
SSQ	sum of squares statistics
df	degree of freedom of SSQ
MAC	minor allele count
n0	the number of individuals with 0 copy of coded alleles
n1	the number of individuals with 1 copy of coded alleles
n2	the number of individuals with 2 copies of coded alleles
scores	beta/se^2 in output RData, where beta and se are vectors
cov	diag(1/se)*LD matrix*diag(1/se) in output RData
n	maximum ntotal in a gene in output RData
sey	residual standard error in output RData

Author(s)

Ming-Huei Chen <mhchen@bu.edu> and Qiong Yang <qyang@bu.edu>

References

coxme package: mixed-effects Cox models, sparse matrices, and modeling data from large pedigrees. Beth Atkinson (atkinson@mayo.edu) for pedigree functions.Terry Therneau (therneau@mayo.edu) for all other functions. 2007. Ref Type: Computer Program. http://cran.r-project.org/web/packages/coxme/.

Abecasis, G. R., Cardon, L. R., Cookson, W. O., Sham, P. C., & Cherny, S. S (2001). Association analysis in a variance components framework. *Genet Epidemiol*, **21** Suppl 1, S341-S346.

Li, B. and Leal, S. M (2008). Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data. *Am J Hum Genet*, **83(3)**, 311-321.

Madsen, B. E. and Browning, S. R (2009). A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic. *PLoS Genet*, **5(2)** e1000384.

Wei P (2009). Asymptotic Tests of Association with Multiple SNPs in Linkage Disequilibrium. *Genet Epidemiol*, **33(6)**, 497-507.

Examples

```
## Not run:
lme.ped(genfile="EC_chr1.txt",phenfile="trait1.csv",pedfile="ped.csv",
phen="trait1",covars=NULL,sep.ped=",",sep.phe=",",sep.gen=" ",mafRange=c(0,0.01),
chr=1,snpinfoRdata="SNPinfo_EC.RData",aggregateBy="SKATgene",maf.file="EC_MAF.csv",
snp.cor="EC_SNPcor.RData",ssq.beta.wts=c(1,25))

## End(Not run)
```

---

rsnpsingene.cor *SNP correlation matrix RData*

Description

This rsnpsingene.cor RData contains SNP correlation matrix for each SKATgene computed based on simulated data of 2671 exome chip SNPs on chromosome 21.

Usage

```
rsnpsingene.cor
```

Format

lists of SNP correlation matrix of 2671 simulated exome chip SNPs on chromosome 21

`snpinfo` *SNP information RData*

Description

This snpinfo RData contains "Name" (SNP name), "Chr" (chromosome number), and "SKATgene" of 2671 exome chip SNPs on chromosome 21.

Usage

```
snpinfo
```

Format

A matrix containing Name, Chr and SKATgene of 2671 simulated SNPs.

Index