Package 'MTAR'

January 20, 2025

Type Package

Title Multi-Trait Analysis of Rare-Variant Association Study

Version 0.1.1

Maintainer Lan Luo <lanluoruc@gmail.com>

Description Perform multi-trait rare-variant association tests using the summary statistics and adjust for possible sample overlap. Package is based on ``Multi-Trait Analysis of Rare-Variant Association Summary Statistics using MTAR'' by Luo, L., Shen, J., Zhang, H., Chhibber, A. Mehrotra, D.V., Tang, Z., 2019 (submitted).

License GPL (>= 2)

Depends R (>= 3.4.0)

Imports CompQuadForm (>= 1.4.3), MASS (>= 7.3-51.4), Matrix (>= 1.2.2), stats, utils

Suggests knitr, rmarkdown

VignetteBuilder knitr

Encoding UTF-8

LazyData true

RoxygenNote 7.0.2

NeedsCompilation no

Author Lan Luo [aut, cre]

Repository CRAN

Date/Publication 2020-04-22 20:44:10 UTC

Contents

---

beta.example   *Example3: the genetic effect estimates and their standard errors*

Description

An example of calculating U and V for MTAR given the genetic effect estimates and their standard errors.

Usage

```
data(beta.example)
```

Format

A list with 3 sublists:

Beta a list of genetic effect estimates, each sublist is $m \times K$ matrix for each study

Beta.se a list of standard errors of genetic effect estimates, each sublist is $m \times K$ matrix for each study

R a SNP correlation matrix for the union of SNPs appearing in all the studies

---

Get_UV_from_beta   *Compute the summary statistics given the genetic effect estimates and their standard errors*

Description

This function allows you to calculate the score summary statistics U and their covariance matrix V for MTAR, given the genetic effect estimates and their standard errors

Usage

```
Get_UV_from_beta(Beta, Beta.se, R)
```

Arguments

Beta	a numeric list, each sublist containing estimation information of genetic effect estimates β for each study. In each study, a numeric $m \times K$ matrix with each row as a SNP and each column as a separate trait. The number of traits and the number of SNPs in each study can be different but their names are required.
Beta.se	a numeric list, each sublist containing the standard error of estimators information for each study. In each study, a numeric $m \times K$ matrix with each row as a SNP and each column as a separate trait.
R	a SNP correlation matrix, which should contain the correlation of all the SNPs in these studies.

Value

A list containing summary statistics for each traits, including the score summary statistics U and their covariance matrix V.

Author(s)

Lan Luo

Examples

```
data("beta.example")
attach(beta.example)
obs.stat <- Get_UV_from_beta(Beta = Beta, Beta.se = Beta.se, R = R)
detach(beta.example)
```

Get_UV_from_data	*Compute the summary statistics given the individual-level data*

Description

This function allows you to calculate the score summary statistics U and their covariance matrix V for MTAR, given the traits, covariates and genotype data sets. If one trait only takes no more than two values, this function will treat this trait as a binary trait.

Usage

```
Get_UV_from_data(traits, covariates, genotype, covariance = TRUE)
```

Arguments

traits	a numeric list, each sublist containing trait information for each study. In each study, a numeric $n \times K$ matrix with each row as an independent individual and each column as a separate trait. If subject i is without trait k, the corresponding value is set as NA. The number of traits in each study can be different but the names of traits are required.

covariates	a numeric list, each sublist containing covariates information for each study. In each study, a numeric $n \times D$ matrix with each row as an independent individual and each column as a covariate.
genotype	a numeric list, each sublist containing genotype information for each study. In each study, a numeric $n \times m$ matrix with each row as an independent individual and each column as a SNP Each genotype should be coded as 0, 1, 2, and NA for AA, Aa, aa, and missing, where A and a represents a major and minor allele, respectively. The number of SNPs in each study can be different but the names of SNPs are required. Also, the number of studies must be the same in genotype, covariates and traits lists. The order of subject ID must be the same among traits, covariates, and genotype within each study.
covariance	a logical value indicating whether to calculate the covariance matrix of score summary statistics U. The default value is TRUE. If covariance is set as FALSE, then only the diagonal values of the covairance matrix are calculated, which is faster. In estimating the zeta matrix correcting for overlap samples, we recommend set covariance as FALSE in calculating the summary statistics for common variants. Since the number of common variants may be large and time-consuming.

Value

A list containing summary statistics for each trait. If covariance is TRUE, the score summary statistics U and its covariance matrix V are returned. Otherwise, only U and the diagonal elements of covariance matrix are returned.

Author(s)

Lan Luo

References

Hu, Y.J., Berndt, S.I., Gustafsson, S., Ganna, A., Mägi, R., Wheeler, E., Feitosa, M.F., Justice, A.E., Monda, K.L., Croteau-Chonka, D.C. and Day, F.R., 2013. Meta-analysis of gene-level associations for rare variants based on single-variant statistics. The American Journal of Human Genetics, 93(2), pp.236-248.

Tang, Z.Z. and Lin, D.Y., 2015. Meta-analysis for discovering rare-variant associations: statistical methods and software programs. The American Journal of Human Genetics, 97(1), pp.35-53.

Examples

```
data(rawdata)
attach(rawdata)
obs.stat <- Get_UV_from_data(traits = traits.dat,
                             covariates = cov.dat,
                             genotype = geno.dat,
                             covariance = TRUE)
obs.stat
detach(rawdata)
```

Get_UV_from_varU   *Compute the summary statistics given the score statistics and their variance.*

Description

This function allows you to calculate the score summary statistics U and their covariance matrix V for MTAR, given the score summary statistics and their variance.

Usage

```
Get_UV_from_varU(U, varU, R)
```

Arguments

U
a numeric list, each sublist containing score summary statistics U for each study. In each study, a numeric $m \times K$ matrix with each row as a SNP and each column as a separate trait. The number of traits and the number of SNPs in each study can be different but their names are required.

varU
a numeric list, each sublist containing the variance of score summary statistics information for each study. In each study, a numeric $m \times K$ matrix with each row as a SNP and each column as a separate trait.

R
a SNP correlation matrix, which should contain the correlation of all the SNPs in these studies.

Value

A list containing summary statistics for each traits, the score summary statistics U and their covariance matrix V.

Author(s)

Lan Luo

Examples

```
data("varU.example")
attach(varU.example)
obs.stat <- Get_UV_from_varU(U = U, varU = varU, R= R)
obs.stat
detach(varU.example)
```

Get_zeta	*Calculate Covariances of Z-scores between Traits from Overlapping Samples*

Description

This function allows you to estimate the matrix ζ to adjust for the potential sample overlap in the data set. Here we applied LD pruning ($r^2 < 0.1$ in 500kb region) on 1000 Genome genotype dataset (hg19) as a list of reference independent SNPs. The SNP ID is chr:pos.

Usage

```
Get_zeta(Zscore, pval_cutoff = 0.05, Indp_common_snp)
```

Arguments

Zscore	a numeric list, each sublist containing a vector of Z scores of SNPs with minor allele frequency (MAF) larger than 0.05. The chr:pos for each SNP is required.
pval_cutoff	a numeric value indicating the cutoff threshold of p-values. The default value is 0.05. Variants with p-value less than or equal to this threshold will be automatically removed.
Indp_common_snp	
	a numeric list of independent common SNPs

Value

A $K \times K$ matrix ζ, where K is the number of traits.

Author(s)

Lan Luo

Examples

```
data(zeta.example)
attach(zeta.example)
# Downloading independent common SNPs from 1000Genome data set.
githubURL <- "https://github.com/lan/MTAR/blob/master/indp_snps.1KG.rda?raw=true"
utils::download.file(githubURL,"1kgfile")
load("1kgfile")
zeta1 <- Get_zeta(Zscore = Zscore, Indp_common_snp = indp_snps.1KG)
zeta1
detach(zeta.example)
```

MTAR   *Multiple-Traits Analysis of Rare-Variant Association Test*

Description

Test for association between a set of rare SNPs and multiple traits with input of summary statistics, possibly from overlap samples. The input number of SNPs in each trait can be different, MTAR function will analyze the union of SNPs that show up in at least one trait, and automatically handle the non-polymorphic SNPs.

Usage

```
MTAR(
  U,
  V,
  MAF,
  MAF_UB = 0.05,
  zeta = NULL,
  genetic_cor.trait = NULL,
  rho.SNP = c(0, 0.5, 1),
  rho.trait = c(0, 0.5, 1),
  weight.SNP = c(1, 25)
)
```

Arguments

U	a numeric list, each sublist containing summary statistics U for each traits. The SNP IDs must be provided.
V	a numeric list, each sublist containing the corresponding covariance matrix of summary statistics. If your original summary statistics are other format, please use Get_UV_from_data, Get_UV_from_varU or Get_UV_from_beta to generate the summary statistics U and V for MTAR.
MAF	a numeric vector containing minor allele frequency for the SNPs show up in at least one trait. The SNP IDs must be provided.
MAF_UB	a numeric value indicating the cutoff threshold of minor allele frequency for SNPs The default value is 0.05.
zeta	a numeric matrix containing the sample correlation of Z-scores over a large number of independent null common SNPs across genome. The default value is NULL, where MTAR assumes there are no overlap samples. However, if there is overlapping in subjects, zeta must be provided. zeta can be estimated using MTAR::Get_zeta.
genetic_cor.trait	a numeric matrix containing the genetic correlation among traits. The default value of genetic_cor.trait is NULL, where an exchangeable correlation structure with the correlation coefficient denoted by rho.trait (ρ_2) is assumed. In this case, there is no difference between cMTAR and iMTAR.

rho.SNP	a numeric vector containing all the possible values of ρ_1. The default value is c(0, 0.5, 1).
rho.trait	a numeric vector containing all the possible values of ρ_2. The default value is c(0, 0.5, 1).
weight.SNP	a numeric vector containing the parameters in Beta density function to calculate the weight among SNPs. The default value is c(1, 25).

Details

MTAR assumes that the genetic effect estimates β has covariance matrix B, which is a Kronecker product of two pre-specified matrices: among-variant effect covariance B_1 and among-trait effect covariance B_2. An exchaneable correlation structure with the correlation coefficient denoted by rho.SNPs (ρ_1) for B_1 is assumed. The default MTAR requires the input of genetic correlation matrix genetoc_cor.trait, if missing, then an exchaneable correlation structure for rho.trait (ρ_2) is assumed. The default weight of B_1 is $dBeta(MAF, 1, 25)$, which can be changed freely by users.

Value

a list of p-values of MTAR-O, cMTAR, iMTAR and cctP as well as ancillary information. Here cctP is the Cauchy-combined p-value of SKAT and burden tests with default weight $dBeta(MAF, 1, 25)$.

Author(s)

Lan Luo

Lan Luo

References

Liu, Y., Chen, S., Li, Z., Morrison, A.C., Boerwinkle, E. and Lin, X., 2019. ACAT: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. The American Journal of Human Genetics, 104(3), pp.410-421.

Liu, Y. and Xie, J., 2019. Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. Journal of the American Statistical Association, pp.1-18.

Luo, L., Shen, J., Zhang, H., Chhibber, A., Mehrotra, D. V., Tang, Z., 2019. Multi-trait analysis of rare-variant association summary statistics using MTAR.

Examples

```
data(MTAR.example)
attach(MTAR.example)
pval <- MTAR(U = U, V = V, MAF = MAF, genetic_cor.trait = genetic_cor.trait,
            zeta = zeta)
pval
detach(MTAR.example)
```

MTAR.example   *PNPLA2 Gene*

Description

PNPLA2 Gene Inofmation required by MTAR function.

Usage

```
data(MTAR.example)
```

Format

A list with 6 sublists:

annotation some annotation information of gene PNPLA2

U a list containing the summary statistics for each trait

V a list containing the covariance matrix of summary statistics for each trait

MAF the minor allele frequency of all the rare variants in this gene

genetic_cor.trait the genetic correlation among traits

zeta the overlapping estimation matrix, which is approximated by the sample correlation matrix of the scaled summary statistics over a large number of independent null common SNPs

---

rawdata   *Example1: individual-level data*

Description

An example of calculating U and V for MTAR given the individual-level data set.

Usage

```
data(rawdata)
```

Format

A list with 3 sublists:

geno.dat a list of genotype data, each sublist is $n \times m$ matrix for each study

traits.dat a list of trait data, each sublist is $n \times K$ matrix for each study

cov.dat a list of covariates data, each sublist is $n \times D$ matrix for each study

There are 3 studies, 3 continuous traits and 10 rare variants. Specifically, there are 1500 subjects in study1, but each subject only has one trait measurement. In study2 and study3, the sample size is 500 and each subject has two or three traits measurements.

---

varU.example   *Example2: the summary statistics and their variance*

Description

An example of calculating U and V for MTAR given the summary statistics and its variance.

Usage

```
data(varU.example)
```

Format

A list with 3 sublists:

U a list of summary statistics, each sublist is $m \times K$ matrix for each study

varU a list of variance of summary statistics, each sublist is $m \times K$ matrix for each study

R a SNP correlation matrix for the union of SNPs appearing in all the studies

---

zeta.example   *Example4: the summary statistics of 737 common and null SNPs*

Description

An example of estimating matrix ζ given the summary statistics information of null and common SNPs.

Usage

```
data(zeta.example)
```

Format

A list with 1 sublist:

Zscore a list, each sublist contains the Z-scores of 737 null and common SNPs for each trait

Index