Package 'MetaSKAT'

October 8, 2024

Type Package

Title Meta Analysis for SNP-Set (Sequence) Kernel Association Test

Version 0.90

Date 2024-10-08

Author Seunggeun (Shawn) Lee [aut, cre]

Maintainer Seunggeun (Shawn) Lee <lee7801@snu.ac.kr>

Description Functions for Meta-
analysis Burden Test, Sequence Kernel Association Test (SKAT) and Optimal SKAT (SKAT-O) by Lee et al. (2013) <doi:10.1016/j.ajhg.2013.05.010>. These methods use summary-level score statistics to carry out gene-based meta-analysis for rare variants.

License GPL (>= 2)

SystemRequirements Little Endian

Depends R (>= 2.13.0)

Imports SKAT (>= 2.0.1)

NeedsCompilation yes

Repository CRAN

Date/Publication 2024-10-08 13:30:07 UTC

Contents

Example   *Example dataset*

Description

Example dataset

Format

This example dataset has the following objects:

y.list a list object of binary phenotypes. It has 3 elements for 3 study cohorts. Each element is a vector of binary phenotypes.

x.list a list object of covariates. It has 3 elements for 3 study cohorts. Each element is a matrix of covariates. The first and last elements have two covariates (two columns), and the second element has one covariate (one column).

n.g a numeric value of the number of cohorts (3).

Z.list a list object of genotypes of all samples. It has 10 elements for 10 genes. Each element is an nxp matrix with n being the total sample size (3000) and p being the number of SNPs.

Generate_Meta_Files   *Generate summary statistics files*

Description

Generate Meta SSD (MSSD) and Meta Info (MInfo) files. Both files are needed to run MetaSKAT with summary statistics.

Usage

```
Generate_Meta_Files(obj, File.Bed, File.Bim
, File.SetID, File.MSSD, File.MInfo, N.Sample
, File.Permu = NULL, data=NULL, impute.method="fixed")

Generate_Meta_Files_FromDosage(obj, File.Dosage
, File.SetID, File.MSSD, File.MInfo, N.Sample
, File.Permu=NULL, data=NULL, impute.method="fixed")
```

Arguments

`obj`	returned object from SKAT_Null_Model.
`File.Bed`	name of the binary ped file (BED).
`File.Bim`	name of the binary bim file (BIM).
`File.SetID`	name of the SNP set ID file. The first column must be Set ID, and the second column must be SNP ID. There should be no header!!
`File.MSSD`	name of the MSSD file that will be generated.
`File.MInfo`	name of the MInfo file that will be generated.
`N.Sample`	number of samples.
`File.Permu`	name of a file that will have score statistics from permuted phenotypes (currently internal use only).
`data`	an optional data frame containing the variables in the model (default=NULL). If it is NULL, the variables are taken from environment(formula).
`impute.method`	a method to impute missing genotypes (default= "fixed"). "bestguess" imputes missing genotypes as the most likely values(0,1,2), "random" imputes missing genotypes by generating binomial(2,p) random variables (p = MAF), and "fixed" imputes missing genotypes by assigning the mean genotype values (2p).
`File.Dosage`	name of the dosage file. The dosage file must not have a header.

Details

These functions generate summary statistic files (MSSD and MInfo files) from plink binary files. To run meta analysis, each study should provide both MSSD and MInfo files. The MSSD is a binary file with between-SNP information matrices, and MInfo is a text file with information on study cohorts and SNPsets.

If users want to use dosages instead of hard call genotypes, Generate_Meta_Files_FromDosage should be used instead of Generate_Meta_Files. The dosage file should follow the plink dosage file format with a single dosage value per each SNP (Format=1 in plink). The first three columns should be SNP ID, allele type1 (a1) and allele type2 (a2). After the first three columns, there should be N.Sample columns of dosage data. Each column represents each sample, and the order of samples should be matched with the order in phenotypes and covariates used in SKAT_Null_Model.

ex)

rs0001 A T 0.1 0.2

rs0002 C G 1.2 0

Dosage value is the expected number of a2 copies, and 0 .. 2 scale. So the value 0.1 indicates that the expected number of copy of a2 is 0.1.

Author(s)

Seunggeun Lee

MetaSKAT_MSSD_OneSet   *Meta analysis SKAT with summary data from each study cohort.*

Description

Meta analysis SKAT with Meta SSD (MSSD) and Info (MInfo) files. MetaSKAT_MSSD_OneSet computes a p-value for a given set, and MetaSKAT_MSSD_ALL computes p-values for all sets.

Usage

```
MetaSKAT_MSSD_OneSet(Cohort.Info, SetID, combined.weight=TRUE, weights.beta=c(1,25),
method="davies", r.corr=0, is.separate = FALSE, Group_Idx=NULL, MAF.cutoff=1,
missing_cutoff=0.15)

MetaSKAT_MSSD_ALL(Cohort.Info, ...)
```

Arguments

Cohort.Info	output object from Open_MSSD_File_2Read.
SetID	a character value of set id to test.
combined.weight	
	a logical value (default=TRUE) for a type of weighting. See MetaSKAT_wZ page for details.
weights.beta	a numeric vector of parameters for the beta weights (default=c(1,25))
method	a method to compute a p-value (default= "davies"). See MetaSKAT_wZ page for details.
r.corr	the ρ parameter for the compound symmetric correlation structure kernels (default= 0). See MetaSKAT_wZ page for details.
is.separate	a logical value (default=FALSE) for homogeneous(=FALSE) or heterogeneous(=TRUE) genetic effects of a SNP set across studies. See MetaSKAT_wZ page for details.
Group_Idx	a vector of group indicator (default=NULL). See MetaSKAT_wZ page for details.
MAF.cutoff	a cutoff of the MAFs of SNPs (default=1). Any SNPs with MAFs > MAF.cutoff will be excluded from the analysis.
missing_cutoff	a cutoff of the missing rates of SNPs (default=0.15). See MetaSKAT_wZ page for details.
...	the same parameters of MetaSKAT_MSSD_OneSet after SetID.

Details

Please see MetaSKAT_wZ for details.

Value

MetaSKAT_MSSD_OneSet and MetaSKAT_wZ return the same object. See MetaSKAT_wZ for details. MetaSKAT_MSSD_ALL returns a dataframe with SetIDs (first column) and p-values (second column).

Author(s)

Seunggeun Lee

References

Lee, S., Teslovich, T., Boehnke, M., Lin, X. (2013) General framework for meta-analysis of rare variants in sequencing association studies. *American Journal of Human Genetics*, 93, 42-53.

MetaSKAT_wZ   *Meta analysis SKAT with individual level genotype data*

Description

Meta analysis SKAT with individual level genotype data.

Usage

```
MetaSKAT_wZ(Z, obj, combined.weight=TRUE, weights.beta=c(1,25),
method="davies", r.corr=0, is.separate = FALSE, Group_Idx=NULL,
impute.method="fixed",impute.estimate.maf=1, missing_cutoff=0.15)
```

Arguments

Z	a numeric genotype matrix with each row as a different individual and each column as a separate snp. Each genotype should be coded as 0, 1, 2, and 9 (or NA) for AA, Aa, aa, and missing, where A is a major allele and a is a minor allele. Missing genotypes will be imputed using observed MAFs.
obj	an output object from the Meta_Null_Model function.
combined.weight	
	a logical value (default=TRUE) for the type of weighting. If it is TRUE, a weight for each SNP is computed using MAFs that are common across studies. If it is FALSE, group specific weights will be used based on group specific MAFs.
weights.beta	a numeric vector of parameters of beta weights (default=c(1,25))
method	a method to compute a p-value (default= "davies"). "davies" represents an exact method that computes a p-value by inverting the characteristic function of the mixture chisq dist., and "optimal" represents the optimal test (SKAT-O) that is based on an optimal linear combination of burden and SKAT statistics. "optimal" is equivalent to "optimal.adj" in the SKAT function.

r.corr   the ρ parameter for the compound symmetric correlation structure kernel (de-
                fault= 0). If r.corr=0, it does the SKAT test. If r.corr=1, it does the burden test.
                If r.corr=(a vector of grid values between 0 and 1), it does SKAT-O. r.corr will
                be ignored if method="optimal". See the manual of SKAT.

is.separate   a logical value (default=FALSE) for homogeneous(=FALSE) or heterogeneous(=TRUE)
                genetic effects of a SNP set across studies. When FALSE, it is assumed that all
                studies share the same causal variants with the same effect size. When TRUE, it
                is assumed that studies/groups may have different causal variants.

Group_Idx   a vector of group indicator (default=NULL). If a vector of integers are speci-
                fied, it assumes causal variants are the same for studies with the same group in-
                dex, and different for studies with different group indexes. When NULL, stud-
                ies are assumed to be in different groups with different group indexes. When
                is.separate=FALSE, it will be ignored.

impute.method   a method to impute missing genotypes (default= "fixed"). "bestguess" imputes
                missing genotypes as the most likely values(0,1,2), "random" imputes missing
                genotypes by generating binomial(2,p) random variables (p = MAF), and "fixed"
                imputes missing genotypes by assigning the mean genotype value (2p).

impute.estimate.maf
                a numeric value indicating how to estimate MAFs for the imputation. If im-
                pute.estimate.maf=1 (default), MetaSKAT uses study-specific MAFs, in which
                each study MAFs will be used for the imputation. If impute.estimate.maf=2, all
                samples in the Z matrix will be used to calculate MAFs for the imputation. If im-
                pute.estimate.maf=3, MetaSKAT uses group-specific MAFs. Previous versions
                (< ver 0.6) used impute.estimate.maf=2 as a default.

missing_cutoff a cutoff of the missing rates of SNPs (default=0.15). If the first study has SNPs
                with missing rates higher than the cutoff, these SNPs in the study will be ex-
                cluded from the analysis. However, the same SNPs in other studies will not be
                excluded, if their missing rates are lower than the cutoff. The missing rates are
                calculated study by study.

Details

The rows of Z should be matched with phenotypes and covariates. If there are 3 studies, and study
1,2, and 3 have n1, n2, and n3 samples, the first n1, n2, and n3 rows of Z should be genotypes of
the first, second, and third studies, respectively.

Group_Idx is a vector of group index. Suppose the first two studies are European-based and the
last study is African American-based. If you want to run MetaSKAT with assuming ancestry group
specific heterogeneity, you can set Group_Idx=c(1,1,2), which indicates the first two studies belong
to the same group.

The four methods in the MetaSKAT paper can be run with the following parameters:

1. Hom-Meta-SKAT: combined.weight=TRUE, is.separate=FALSE

2. Hom-Meta-SKAT-O: combined.weight=TRUE, is.separate=FALSE, method="optimal"

3. Het-Meta-SKAT: combined.weight=FALSE, is.separate=TRUE

4. Het-Meta-SKAT-O: combined.weight=FALSE, is.separate=TRUE, method="optimal"

Value

p.value   p-value.

param   estimated parameters of each method.

param$Is_Converged

(only with method="davies") an indicator for the convergence. 1 indicates convergence and 0 otherwise. When 0 (not converged), "liu" method will used to compute a p-value.

Author(s)

Seunggeun Lee

Examples

```
data(Example)
attach(Example)

###############################################################
# Compute a p-value of the first gene

obj<-Meta_Null_Model(y.list, x.list, n.cohort=3, out_type="D")

# rho=0
MetaSKAT_wZ(Z.list[[1]], obj)$p.value

# rho=1 (burden test)
MetaSKAT_wZ(Z.list[[1]], obj, r.corr=1)$p.value


# optimal test
MetaSKAT_wZ(Z.list[[1]], obj, method="optimal")$p.value

# cohort specific weights
MetaSKAT_wZ(Z.list[[1]], obj, combined.weight=FALSE)$p.value


# Seperate = TRUE
# Assume heterogeneous genetic effect
MetaSKAT_wZ(Z.list[[1]], obj, combined.weight=FALSE, is.separate = TRUE)$p.value


# Group

# the first two cohorts are in the same group.
Group_Idx=c(1,1,2)
MetaSKAT_wZ(Z.list[[1]], obj, combined.weight=FALSE, is.separate = TRUE,Group_Idx=Group_Idx)$p.value

# all three cohorts are in different group.
Group_Idx=c(1,2,3)
MetaSKAT_wZ(Z.list[[1]], obj, combined.weight=FALSE, is.separate = TRUE,Group_Idx=Group_Idx)$p.value
```

Meta_Null_Model   *Get parameters and residuals from H0*

Description

Compute model parameters and residuals under the null model (H0) of no associations. It can be used only when individual level data are available.

Usage

```
Meta_Null_Model(y.list, x.list, n.cohort, out_type="C",  n.Resampling=0)
Meta_Null_Model_EmmaX(y.list, x.list, n.cohort, K.list=NULL, Kin.File.list=NULL)
```

Arguments

y.list   a list object for phenotypes. Each element should be a vector of phenotypes. If you have 3 cohorts, it should have 3 elements.

x.list   a list object for covariates. Each element should be a vector or a matrix of covariates. If there are 3 cohorts, it should have 3 elements. If there are no covariates to adjust for, the element should be "intercept". See the examples.

n.cohort   a numeric value of the number of cohort.

out_type   an indicator for the outcome type. "C" for continuous outcomes and "D" for dichotomous outcomes.

n.Resampling   internal use only.

K.list   a list object of the kinship matrices. If K.list=NULL, the function reads files in Kin.File.list to obtain kinship matrices.

Kin.File.list   a list object of emmax-kin output file names.

Value

It returns an object that has model parameters and residuals. The returned object will be used to run MetaSKAT_wZ.

Author(s)

Seunggeun Lee

Examples

```
data(Example)
attach(Example)

#############################################################
# Compute a p-value of the first gene

obj<-Meta_Null_Model(y.list, x.list, n.cohort=3, out_type="D")
MetaSKAT_wZ(Z.list[[1]], obj)$p.value

#############################################################
# If you want to use the intercept-only model for the 2nd cohort

x.list[[2]]<-"intercept"
obj<-Meta_Null_Model(y.list, x.list, n.cohort=3, out_type="D")
MetaSKAT_wZ(Z.list[[1]], obj)$p.value
```

Open_MSSD_File_2Read   *Read Meta SSD and Info files*

Description

Read Meta SSD (MSSD) and Meta Info (MInfo) files.

Usage

```
Open_MSSD_File_2Read(File.MSSD.vec, File.MInfo.vec)
```

Arguments

File.MSSD.vec   a vector of MSSD files. Each element represents a MSSD file of each study.

File.MInfo.vec   a vector of MInfo files. Each element represents a MInfo file of each study.

Details

Users should open MSSD and MInfo files to run MetaSKAT_MSSD_OneSet or MetaSKAT_MSSD_ALL. If all individual level data are available, use MetaSKAT_wZ instead.

Value

This function returns an data object. The returned object will be used to run MetaSKAT_MSSD_OneSet or MetaSKAT_MSSD_ALL.

Author(s)

Seunggeun Lee

Index