Package 'iECAT'

May 11, 2017

Type Package

Title Integrating External Controls into Association Test

Version 0.8

Date 2017-05-10

Author Seunggeun (Shawn) Lee

Maintainer Seunggeun (Shawn) Lee <leeshawn@umich.edu>

Description
Functions for single-variant and region-based tests with external control samples. These methods use external study samples as control samples with adjusting for possible batch effects.

License GPL (>= 2)

Depends R (>= 2.13.0)

Imports MetaSKAT, SKAT (>= 1.2.0)

NeedsCompilation no

Repository CRAN

Date/Publication 2017-05-11 05:59:26 UTC

R topics documented:

Close_SSD_wEC *Close SNP set data (SSD) and external control (EC) info files*

Description

Close the opened SNP Set data (SSD) and external control (EC) info files.

Usage

```
Close_SSD_wEC()
```

Author(s)

Seunggeun Lee

Example *Example dataset for iECAT*

Description

Example dataset for iECAT.

Format

Example contains the following objects:

Z.list a list of 10 genes

tbl.external.all.list a list of 10 allele count tables of external control samples for the 10 genes in Z.list

Y a numeric vector of binary phenotypes

`Generate_SSD_SetID_wEC` *Generate SNP set data file (SSD) with using external control information*

Description

Generate SNP set data (SSD) and external control information files

Usage

```
Generate_SSD_SetID_wEC(File.Bed, File.Bim, File.Fam, File.SetID
, File.EC, File.SSD, File.Info, File.EC.Info)
```

Arguments

`File.Bed`	name of the binary ped file (BED).
`File.Bim`	name of the binary map file (BIM).
`File.Fam`	name of the FAM file (FAM).
`File.SetID`	name of the SNP set ID file that defines SNP sets. The first column must be Set ID, and the second column must be SNP ID. There should be no header!!
`File.EC`	name of the file that has allele count information of the external controls.
`File.SSD`	name of the SSD file generated.
`File.Info`	name of the SSD info file generated.
`File.EC.Info`	name of the external control info file generated.

Details

The SetID file is a white-space (space or tab) delimitered file with 2 columns: SetID and SNP_ID.

Please keep in mind that there should be no header! The SNP_IDs and SetIDs should be less than 50 characters, otherwise, it will return an error message.

File.EC should be a text formated file

The SSD file is a binary formated file with genotypes. The SSD info file is a text file with general information on data and SNP sets (first 6 rows), and information on each set (after 8th row).

Author(s)

Seunggeun Lee, Larisa Miropolsky

iECAT *Integrating External Controls to Association Tests*

Description

Test for association of a set of variants with integrating external study samples to improve power

Usage

```
iECAT(Z, obj, tbl.external.all, weights.beta=c(1,25), weights = NULL,
r.corr=0, method="davies", missing_cutoff=0.15, MAC.lowlimit=3,
MAF.limit=1)

iECAT.SSD.OneSet_SetIndex(EC.INFO, SetIndex,  obj, ..., obj.SNPWeight=NULL)
```

Arguments

Z	a numeric genotype matrix with each row as a different individual and each column as a separate gene/snp. Each genotype should be coded as 0, 1, 2, and 9 (or NA) for AA, Aa, aa, and missing. A does not need to be a major allele.
obj	an output object of the SKAT_Null_Model function.
tbl.external.all	a p x 2 matrix of allele count of the matrix. Each row should be matched with the each column in Z.
weights.beta	a numeric vector of parameters for the beta weights for the weighted kernels. If you want to use your own weights, please use the "weights" parameter. It will be ignored if "weights" parameter is not null.
weights	a numeric vector of weights for the weighted kernels. It is $\sqrt{w}$ in the SKAT paper. So if you want to use the Madsen and Browning (2009) weight, you should set each element of weights as $1/\sqrt{p(1-p)}$, not $1/p(1-p)$. When it is NULL, the beta weight with the "weights.beta" parameter is used.
r.corr	the ρ parameter for the compound symmetric correlation structure kernels (default= 0). If you give a vector value, SKAT will conduct the optimal test. It will be ignored if method="optimal" or method="optimal.adj". See details.
method	a method to compute the p-value (default= "davies"). "davies" represents an exact method that computes the p-value by inverting the characteristic function of the mixture chisq, "liu" represents an approximation method that matches the first 3 moments, "liu.mod" represents modified "liu" method that matches kurtosis instead of skewness to improve tail probability approximation, "optimal" represents a SKAT-O based on an unified approach.

missing_cutoff	a cutoff of the missing rates of SNPs (default=0.15). Any SNPs with missing rates higher than the cutoff will be excluded from the analysis.
MAC.lowlimit	a cutoff for MAC low limit (default=3). Variants with internal study MAC <= MAC.lowlimit will be excluded from the analysis.
MAF.limit	a cutoff for MAF upper limit (default=1). Variants with internal study MAF > MAF.limit will be excluded from the analysis.
EC.INFO	an EC_INFO object returned from Open_SSD_wEC.
SetIndex	a numeric value of Set index. A set index of each set can be found from SetInfo object in EC.INFO$SSD.INFO.
...	further arguments to be passed to "iECAT"
obj.SNPWeight	an output object of Read_SNP_WeightFile (default=NULL). If NULL, the beta weight with the "weights.beta" parameter will be used.

Value

p.value	p-value of iECAT.
p.value.noadj	p-value without the adjustment of possible batch effects
p.value.internal	SKAT/SKAT-O p-values with only internal study samples
param	estimated parameters of each method.
param$n.marker	a number of variants in the genotype matrix (Z).
param$n.marker.test	a number of variants used for the test.

Author(s)

Seunggeun (Shawn) Lee

Examples

```
library(SKAT)

data(Example, package="iECAT")
attach(Example)

# iECAT-O
# test the first gene

obj<-SKAT_Null_Model(Y ~ 1, out_type="D")
Z = Z.list[[1]]
tbl.external.all = tbl.external.all.list[[1]]

iECAT(Z, obj, tbl.external.all, method="optimal")


# test for the first 3 genes in the Example dataset
p.value.all<-rep(0,3)
```

```
p.value.internal.all<-rep(0,3)
for(i in 1:3){

re<-iECAT(Z.list[[i]], obj, tbl.external.all.list[[i]], method="optimal")
p.value.all[i]<-re$p.value
p.value.internal.all[i]<-re$p.value.internal

}

# iECAT-O p-values
p.value.all

# SKAT-O p-values
p.value.internal.all
```

`iECAT.SSD.All` *Integrating External Controls to Association Tests*

Description

Iteratively carry out association tests with phenotypes and SNP sets in SSD file.

Usage

```
iECAT.SSD.All(EC.INFO, obj, ..., obj.SNPWeight=NULL)
```

Arguments

`EC.INFO`	EC_INFO object returned from Open_SSD_wEC.
`obj`	output object from SKAT_Null_Model.
`...`	further arguments to be passed to "iECAT".
`obj.SNPWeight`	output object from Read_SNP_WeightFile (default=NULL). If NULL, the beta weight with the "weights.beta" parameter will be used.

Details

Please see iECAT for details.

Value

`results`	dataframe that contains SetID, p-values (P.value, P.value.Noadj, and P.value.Internal), the number of markers in the SNP sets (N.Marker.All), and the number of markers to test for associations (N.Marker.Test).
`P.value.Resampling`	currently resampling p-values are not provided.

Author(s)

Seunggeun Lee

`iECAT_SingleVar` *Integrating External Controls to Association Tests*

Description

Test for association of a single variant with integrating external study samples to improve power

Usage

```
iECAT_SingleVar(Z, obj, tbl.external, Fisher.test=FALSE, weight.factor=NULL)
iECAT_SingleVar_Tbl(Tbl, Fisher.test=FALSE, weight.factor=NULL)
```

Arguments

`Z`	a numeric genotype vector for the variant. Each genotype should be coded as 0, 1, 2, and 9 (or NA) for AA, Aa, aa, and missing. A does not need to be a major allele.
`obj`	an output object of the SKAT_Null_Model function.
`tbl.external`	a vector of allele count from the external control.
`Fisher.test`	an indicator variable (default= "FALSE") to indicate whether to obtain p.value.internal, p.value.noadjust, and p.value.IvsE using Fisher's exact test. If FALSE, asymptotic tests will be used to calculate p-values.
`weight.factor`	internal use only.
`Tbl`	a 3 by 2 allele count table. The first column should have allele counts for the minor allele, and the second column should have allele counts for the major allele. The first, second and third row should correspond to case, internal control and external control, respectively.

Value

`OR`	OR estimate from iECAT.
`SE`	SE estimate of OR.
`p.value`	p-value of iECAT.
`p.value.noadj`	p-value without the adjustment of possible batch effects
`p.value.internal`	p-value with only internal study samples
`p.value.IvsE`	p-value of comparing internal vs external control samples

Author(s)

Seunggeun (Shawn) Lee

Examples

```
library(SKAT)

data(Example, package="iECAT")
attach(Example)

# iECAT-O
# test the second SNP in the first gene
idx.snp<-2

obj<-SKAT_Null_Model(Y ~ 1, out_type="D")
Z = Z.list[[1]][,idx.snp]
tbl.external= tbl.external.all.list[[1]][idx.snp,]


iECAT_SingleVar(Z, obj, tbl.external)
```

Open_SSD_wEC *Open SNP set data (SSD) and external control (EC) info files*

Description

Open a SNP Set data (SSD) and external control (EC) info files. After finishing using the files, please close them by callinsg Close_SSD_wEC function.

Usage

```
Open_SSD_wEC(File.SSD, File.Info, File.EC.Info)
```

Arguments

File.SSD	name of the SSD file .
File.Info	name of the SSD info file.
File.EC.Info	name of the external control (EC) info file.

Value

a list object of EC.INFO.

Author(s)

Seunggeun Lee

Index