Package: rvHPDT (via r-universe)

November 4, 2024

Type Package

Title Calling Haplotype-Based and Variant-Based Pedigree Disequilibrium Test for Rare Variants in Pedigrees

Version 4.0

Date 2022-5-9

Author Wei Guo <wei.guo3@nih.gov>

Maintainer Wei Guo <wei.guo3@nih.gov>

Depends gtools, R(>= 2.15.0)

Imports stats, utils

Description To detecting rare variants for binary traits using general pedigrees, the pedigree disequilibrium tests are proposed by collapsing rare haplotypes/variants with/without weights. To run the test, MERLIN is needed in Linux for haplotyping.

License GPL (>= 2)

Repository https://weiguonimh.r-universe.dev

RemoteUrl https://github.com/weiguonimh/rvhpdt

RemoteRef HEAD

RemoteSha 573d948af8844e9f6a88f10756416d1ca34aa1ea

Contents

`convert.factors.to.strings.in.dataframe` *Convert data.frame columns from factors to characters*

Description

Convert data.frame columns from factors to characters

Usage

```
convert.factors.to.strings.in.dataframe(dataframe)
```

Arguments

`dataframe`	dataframe with columns of factors

Value

`dataframe`	dataframe with columns of characters

`fullPedigree` *Internal function.*

Description

Complete family members as requested by Merlin software.

`HapMendlianCheck` *Internal function.*

Description

Check Mendlian error in families.

`mydaoshu` *Internal function.*

Description

function of 1/x

`mypaste` *Internal function.*

Description

Modified paste function.

`randomPDT` *Internal function.*

Description

Calculate PDT statistics by permuting the transmission and non-transmission status for each child based on parents' genotype.

`rchild` *Internal function.*

Description

Generate child's genotype by permuting the transmission and non-transmission status based on parents' genotype.

`rhapPDT` *Calling haplotype-based and variant-based pedigree disequilibrium test for rare variants in pedigrees.*

Description

To detecting rare variants for binary traits using general pedigrees, the pedigree disequilibrium tests are proposed by collapsing rare haplotypes/variants with/without weights.

Usage

```
 rhapPDT(ped, map, aff=2, unaff=1, mu=1.04,
merlinFN.prefix="merlin", nperm=1000, trace=TRUE)
```

Arguments

ped	input data, has same format with PLINK but having column names. The PED file is a white-space (space or tab) delimited file, and the first six columns are mandatory: FID: Family ID; IID: Individual ID; FA: Paternal ID; MO: Maternal ID; SEX: Sex (1=male; 2=female; other=unknown); PHENO: Phenotype; Genotypes (column 7 onwards) should also be white-space delimited; they are coded as 0, 1 and 2, indicating the number of coding allele, and NA is for missing genotype.
map	input data, has same format with MAP file required by MERLIN. The MAP file is a white-space (space or tab) delimited file with 3 columns as follows, CHROMOSOME: chromosome (1-22, X, Y or 0 if unplaced) MARKER: marker name in PED file that is usually rs# or snp identifier POSITION: Genetic distance (morgans) The data file and map file can include different sets of markers, but markers that are absent from the map file will be ignored by MERLIN.
aff	indicates the values that represents affected status in "PHENO" column of PED data; default is 2.
unaff	indicates the values that represents unaffected status in "PHENO" column of PED data; default is 1.
mu	indicates mu value that defines causal in the training data; default is 1.04.
merlinFN.prefix	Requests that output file of MERLIN names should be derived from outFN.prefix. For example, when it is set to be "merlin" as default, estimated haplotypes should be stored in a file called merlin.chr.
nperm	indicates the times of permutation; default is 1000.
trace	Indicates whether or not the intermediate outcomes should be printed; default is FALSE.

Value

hPDT_v0	P value of unweighted haplotype PDT test statistic.
hPDT_v1	P value of weighted haplotype PDT test statistic.
rvPDT_v0	P value of unweighted rvPDT test statistic.
rvPDT_v1	P value of weighted rvPDT test statistic.

References

Guo W , Shugart YY, Does Haplotype-based Collapsing Tests Gain More Power than Variant-based Collapsing Tests for Detecting Rare Variants in Pedigrees (manuscript).

Examples

```
#ped<-read.table("MLIP.ped",head=1,stringsAsFactors=FALSE)
#map<-read.table("MLIP.map",head=1,stringsAsFactors=FALSE)
#test<-rhapPDT(ped, map, trace=TRUE)
#test
#$hPDT_v0
```

```
#[1] 0.4231359

#$hPDT_v1
#[1] 0.1481145

#$rvPDT_v0
#[1] 0.03237073

#$rvPDT_v1
#[1] 0.162997
```

`rvPDT.test` *Variants-based pedigree disequilibrium test for rare variants in pedigrees.*

Description

To detecting rare variants for binary traits using general pedigrees, the pedigree disequilibrium tests are extended by collapsing rare variants with/without weights.

Usage

```
rvPDT.test(seed=NULL,ped, aff=2,unaff=1, snpCol, hfreq=NULL,
training=0.3, mu=1.28,useFamWeight=TRUE,trace=FALSE)
```

Arguments

`seed`	indicates the seed for randomly selectiong training data.
`ped`	input data, has same format with PLINK but having column names. The PED file is a white-space (space or tab) delimited file: the first six columns are mandatory: FID: Family ID; IID: Individual ID; FA: Paternal ID; MO: Maternal ID; SEX: Sex (1=male; 2=female; other=unknown); PHENO: Phenotype; Genotypes (column 7 onwards) should also be white-space delimited; they are coded as 0, 1 and 2, indicating the number of coding allele, and NA is for missing genotype.
`aff`	indicates the values that represents affected status in ped data; default is 2.
`unaff`	indicates the values that represents unaffected status in ped data; default is 1.
`snpCol`	indicates the columns of variants in ped data.
`hfreq`	indicates the frequencies of variants that used in calculating weights; when it is NULL, the frequencies are estimated by ped data.
`training`	indicates the proportion of training data; default is 0.3.
`mu`	indicates mu value that defines causal in the training data; default is 1.04.
`useFamWeight`	indicates whether the family weights need to be used in the test.
`trace`	indicates whether or not the intermediate outcomes should be printed; default is FALSE.

Value

TDT	Transmission/disequilibrium matrix for each pedigrees.
Sib	Discordant sib pairs matrix for each pedigrees.
PDT	Pedigree disequilibrium matrix for each pedigrees, which is the sum of TDT and Sib.
W	Weights used in Weighted rvPDT test.
test.v1	Weighted rvPDT test statistic with weights W.
test.v0	Unweighted rvPDT test statistic with weights=1.
pvalue.v1	P value of weighted rvPDT test statistic (test.v1).
pvalue.v0	P value of unweighted rvPDT test statistic (test.v0).

References

Guo W , Shugart YY, Does Haplotype-based Collapsing Tests Gain More Power than Variant-based Collapsing Tests for Detecting Rare Variants in Pedigrees (manuscript).

---

rvPDT.test.permu *Variants-based pedigree disequilibrium test for rare variants in pedigrees.*

Description

To detecting rare variants for binary traits using general pedigrees, the pedigree disequilibrium tests are extended by collapsing rare variants with/without weights.

Usage

```
rvPDT.test.permu(ped, aff=2,unaff=1, snpCol, hfreq=NULL,
useFamWeight=TRUE, nperm=1000,trace=FALSE)
```

Arguments

ped	input data, has same format with PLINK but having column names. The PED file is a white-space (space or tab) delimited file: the first six columns are mandatory: FID: Family ID; IID: Individual ID; FA: Paternal ID; MO: Maternal ID; SEX: Sex (1=male; 2=female; other=unknown); PHENO: Phenotype; Genotypes (column 7 onwards) should also be white-space delimited; they are coded as 0, 1 and 2, indicating the number of coding allele, and NA is for missing genotype.
aff	indicates the values that represents affected status in ped data; default is 2.
unaff	indicates the values that represents unaffected status in ped data; default is 1.
snpCol	indicates the columns of variants in ped data.
hfreq	indicates the frequencies of variants that used in calculating weights; when it is NULL, the frequencies are estimated by ped data.

useFamWeight	indicates whether the family weights need to be used in the test.
nperm	indicates the times of permutation; default is 1000.
trace	indicates wether or not the intermediate outcomes should be printed; default is FALSE.

Value

TDT	Transmission/disequilibrium matrix for each pedigrees.
Sib	Discordant sib pairs matrix for each pedigrees.
PDT	Pedigree disequilibrium matrix for each pedigrees, which is the sum of TDT and Sib.
W	Weights used in Weighted rvPDT test.
test.v1	Weighted rvPDT test statistic with weights W.
test.v0	Unweighted rvPDT test statistic with weights=1.
pvalue.v1	P value of weighted rvPDT test statistic (test.v1).
pvalue.v0	P value of unweighted rvPDT test statistic (test.v0).

References

Guo W , Shugart YY, Does Haplotype-based Collapsing Tests Gain More Power than Variant-based Collapsing Tests for Detecting Rare Variants in Pedigrees (manuscript).

---

rvPDT.test.sub *Internal function.*

Description

Internal function of testing rare variants for binary traits using general pedigrees.

---

whap.prehap *Prepare haplotype pairs for hPDT tests in pedigree data.*

Description

Before running hPDT test, haplotype pairs are inferred by calling MERLIN in linux for all pedigree members, and then perpare some internal statistics. Require the R package of "gregmisc" and MERLIN software.

Usage

```
whap.prehap(ped,map, merlinDir="", outFN.prefix="merlin",aff=2,trace=FALSE)
```

Arguments

ped	input data, has same format with PLINK but having column names. The PED file is a white-space (space or tab) delimited file, and the first six columns are mandatory: FID: Family ID; IID: Individual ID; FA: Paternal ID; MO: Maternal ID; SEX: Sex (1=male; 2=female; other=unknown); PHENO: Phenotype; Genotypes (column 7 onwards) should also be white-space delimited; they are coded as 0, 1 and 2, indicating the number of coding allele, and NA is for missing genotype.
map	input data, has same format with MAP file required by MERLIN. The MAP file is a white-space (space or tab) delimited file with 3 columns as follows, CHROMOSOME: chromosome (1-22, X, Y or 0 if unplaced) MARKER: marker name in PED file that is usually rs# or snp identifier POSITION: Genetic distance (morgans) The data file and map file can include different sets of markers, but markers that are absent from the map file will be ignored by MERLIN.
merlinDir	indicates the directory of Merlin, for example, merlinDir="./Merlin/"; use the default="" when Merlin is in current directory or your bin directory.
outFN.prefix	Requests that output file of MERLIN names should be derived from outFN.prefix. For example, when it is set to be "merlin" as default, estimated haplotypes should be stored in a file called merlin.chr.
aff	indicates the values that represents affected status in ped data; default is 2.
trace	indicates whether or not the intermediate outcomes should be printed; default is FALSE.

Value

SNPname	SNP names of testing.
hapData	Haplotype data for each individuals.
freq	Estimated frequencies of haplotypes.
trans	Transmission matrix of haplotypes.
hapScore	Score matrix of haplotypes.

References

Guo W , Shugart YY, Does Haplotype-based Collapsing Tests Gain More Power than Variant-based Collapsing Tests for Detecting Rare Variants in Pedigrees (manuscript).

Index