Package 'KnockoffHybrid'

June 11, 2024

Title Hybrid Analysis of Population and Trio Data with Knockoff
Statistics for FDR Control

Version 1.0.1

Description

Identification of putative causal variants in genome-wide association studies using hybrid analysis of both the trio and population designs. The package implements the method in the paper: Yang, Y., Wang, Q., Wang, C., Buxbaum, J., & Ionita-Laza, I. (2024). KnockoffHybrid: A knockoff framework for hybrid analysis of trio and population designs in genome-wide association studies. The American Journal of Human Genetics, in press.

Imports glmnet, SPAtest

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

Depends R (>= 2.10)

NeedsCompilation no

Author Yi Yang [aut, cre] (<https://orcid.org/0000-0002-1471-4026>)

Maintainer Yi Yang <yi.yang@cityu.edu.hk>

Repository CRAN

Date/Publication 2024-06-11 15:10:02 UTC

Contents

| calculate_weight | *Calculate weight from population designs* |

Description

Calculate weight using population genotype or summary statistics.

Usage

```
calculate_weight(
  pval = NULL,
  beta = NULL,
  method = "score",
  geno = NULL,
  y = NULL,
  phetype = "D",
  PCs = NULL
)
```

Arguments

pval	A numeric vector of length p for p-values. P-values must be between 0 and 1. If not NULL, weight will be calculated as -log10(p-value).
beta	A numeric vector of length p for beta coefficients. If not NULL, weight will be calculated as the absolute value of beta coefficients.
method	A character string for the name of the weight estimation method. Must not be NULL if population genotype is used to calculate weight. Weight can be calculated using "score" (i.e., single variant score test) or "lasso" (i.e., least absolute shrinkage and selection operator). The default is "score".
geno	A n*p matrix for the population genotype data, in which n is the number of subjects and p is the number of variants. The genotypes must be coded as 0, 1, or 2.
y	A numeric vector of length n for the phenotype data for the n subjects.
phetype	A character for the variable type of the phenotype. The type can be "C" (i.e., continuous) or "D" (i.e., dichotomous). The default is "D".
PCs	A n*k matrix for the principal components of population structure, in which n is the number of subjects and k is the number of (top) principal components. If not NULL, principal components will be included as covariates when calculating weight from population genotype.

Value

A numeric vector of length p for the weight.

Examples

```
data(KnockoffHybrid.example)
weight<-calculate_weight(geno=KnockoffHybrid.example$dat.pop,y=KnockoffHybrid.example$y.pop)
```

---

causal_loci	*Identification of causal loci*

Description

Identification of causal loci using KnockoffHybrid's feature statistics

Usage

```
causal_loci(window, M = 10, fdr = 0.2)
```

Arguments

window	The result window from KnockoffHybrid. If there are multiple windows, please use rbind to combine the windows.
M	A positive integer for the number of knockoffs. The default is 10.
fdr	A real number in a range of (0,1) indicating the target FDR level. The default is 0.2.

Value

A list that contains:

window A data frame for an updated window that includes an extra column for KnockoffHybrid's Q-values. A locus with a Q-value <= the target FDR level, i.e., window$q<=fdr, is considered as causal.

thr.w A positive real number indicating the significance threshold for KnockoffTrio's feature statistics. A locus with a feature statistic >= thr.w, i.e., window$w>=thr.w is considered as causal. The loci selected by window$w>=thr.w are equivalent to those by window$q<=fdr. No loci are selected at the target FDR level if thr.w=Inf.

Examples

```
data(KnockoffHybrid.example)
dat.ko<-create_knockoff(KnockoffHybrid.example$dat.hap,KnockoffHybrid.example$pos,M=10)
weight<-calculate_weight(geno=KnockoffHybrid.example$dat.pop,y=KnockoffHybrid.example$y.pop)
window<-KnockoffHybrid(dat=KnockoffHybrid.example$dat,dat.ko=dat.ko,
        pos=KnockoffHybrid.example$pos,weight=weight)
result<-causal_loci(window,M=10,fdr=0.2)
```

---

create_knockoff   *Create knockoff genotype data in trio designs*

Description

Create knockoff genotype data using phased haplotype data in trio designs.

Usage

```
create_knockoff(
  dat.hap,
  pos,
  M = 10,
  maxcor = 0.7,
  maxbp = 90000,
  phasing.dad = NULL,
  phasing.mom = NULL,
  seed = 100
)
```

Arguments

dat.hap	A 6n*p matrix for the haplotype data, in which n is the number of trios and p is the number of variants. Each trio must consist of father, mother, and offspring (in this order). The haplotypes must be coded as 0 or 1. Missing haplotypes are not allowed.
pos	A numeric vector of length p for the position of p variants.
M	A positive integer for the number of knockoffs. The default is 10.
maxcor	A real number in a range of [0,1] for the correlation threshold in hierarchical clustering, such that variants from two different clusters do not have a correlation greater than maxcor when constructing knockoff parents. The default is 0.7.
maxbp	A positive integer for the size of neighboring base pairs used to generate knockoff parents. The default is 90000.
phasing.dad	A numeric vector of length n that contains 1 or 2 to indicate which paternal haplotype was transmitted to offspring in each trio. If NA, the function will calculate the phasing information based on the input haplotype matrix.
phasing.mom	A numeric vector of length n that contains 1 or 2 to indicate which maternal haplotype was transmitted to offspring in each trio. If NA, the function will calculate the phasing information based on the input haplotype matrix.
seed	An integer for the random seed in the knockoff generation. The default is 100.

Value

A 3n*p*M array for the knockoff genotype data.

Examples

```
data(KnockoffHybrid.example)
dat.ko<-create_knockoff(KnockoffHybrid.example$dat.hap,KnockoffHybrid.example$pos,M=10)
```

KnockoffHybrid	*Calculate KnockoffHybrid's feature statistics*

Description

Calculate KnockoffHybrid's feature statistics using original and knockoff genotype data.

Usage

```
KnockoffHybrid(
  dat,
  dat.ko = NA,
  pos,
  allele = NA,
  start = NA,
  end = NA,
  size = c(1, 1000, 5000, 10000, 20000, 50000),
  p_value_only = FALSE,
  adjust_for_cov = FALSE,
  y = NA,
  chr = "1",
  sex = NA,
  weight = NULL
)
```

Arguments

dat	A 3n*p matrix for the original trio genotype data, in which n is the number of trios and p is the number of variants. Each trio must consist of father, mother, and offspring (in this order). The genotypes must be coded as 0, 1, or 2. Missing genotypes are not allowed.
dat.ko	A 3n*p*M array for the knockoff trio genotype data created by function create_knockoff. M is the number of knockoffs.
pos	A numeric vector of length p for the position of p variants.
allele	A vector of length p for the minor allele at each position. Minor alleles for windows with multiple variants will be shown as "W" in the output.
start	An integer for the first position of sliding windows. If NA, start=min(pos). Only used if you would like to use the same starting position for different cohorts/analyses.
end	An integer for the last position of sliding windows. If NA, end=max(pos). Only used if you would like to use the same ending position for different cohorts/analyses.

size	A numeric vector for the size(s) of sliding windows when scanning the genome
p_value_only	A logical value indicating whether to perform the knockoff analysis. When p_value_only is TRUE, only the ACAT-combined p-values are to be calculated for each window. When p_value_only is FALSE, dat.ko is required and KnockoffHybrid's feature statistics are to be calculated for each window in addition to the p-values.
adjust_for_cov	A logical value indicating whether to adjust for covariates. When adjust_for_cov is TRUE, y is required.
y	A numeric vector of length n for the residual Y-Y_hat. Y_hat is the predicted value from the regression model in which the quantitative trait Y is regressed on the covariates. If Y is dichotomous, you may treat Y as quantitative when applying the regression model.
chr	A character for the name of the chromosome, e.g., "1", "2", ..., "22", and "X".
sex	A numeric vector of length n for the sex of offspring. 0s indicate females and 1s indicate males.
weight	A numeric vector of length p for the weight of p variants. The weight can be obtained via the function "calculate_weight" using population genotype or summary statistics. If NULL, the weight will be calculated based on minor allele frequencies.

Value

A data frame for the hybrid analysis results. Each row contains the p-values and, if p_value_only is FALSE, KnockoffHybrid's feature statistics for a window.

Examples

```
data(KnockoffHybrid.example)
dat.ko<-create_knockoff(KnockoffHybrid.example$dat.hap,KnockoffHybrid.example$pos,M=10)
weight<-calculate_weight(geno=KnockoffHybrid.example$dat.pop,y=KnockoffHybrid.example$y.pop)
window<-KnockoffHybrid(dat=KnockoffHybrid.example$dat,dat.ko=dat.ko,
        pos=KnockoffHybrid.example$pos,weight=weight)
```

KnockoffHybrid.example

Example data for KnockoffHybrid

Description

A toy example of the haplotype and genotype data for trio designs and the genotype and phenotype data for population designs

Usage

KnockoffHybrid.example

Format

KnockoffHybrid.example contains the following items:

dat A numeric genotype matrix of 3 trios and 5 variants. Each trio contains 3 rows in the order of father, mother and offspring. Each column represents a variant.

dat.hap A numeric haplotype matrix of 3 trios and 5 variants. Each trio contains 6 rows in the order of father, mother and offspring. Each column represents a variant.

dat.pop A numeric genotype matrix of 8 subjects and 5 variants. Each row represents a subject and each column represents a variant.

y.pop A numeric vector of length 8 for the dichotomous phenotypes of 8 subjects.

pos A numeric vector of length 5 for the position of 5 variants.

Index