Package 'KnockoffTrio'

March 7, 2022

Title Trio Data Analysis with Knockoff Statistics for FDR Control

Version 1.0.0

Description
Identification of putative causal variants in genome-wide association studies with trio design.

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 <yy3136@cumc.columbia.edu>

Repository CRAN

Date/Publication 2022-03-07 20:30:14 UTC

R topics documented:

causal_loci *Identification of causal loci*

Description

Identification of causal loci using KnockoffTrio's feature statistics

Usage

```
causal_loci(window, M = 10, fdr = 0.15)
```

Arguments

window	The result window from KnockoffTrio. 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 0.15.

Value

A list that contains:

window A data frame for an updated window that includes an extra column for KnockoffTrio'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(KnockoffTrio.example)
dat.ko<-create_knockoff(KnockoffTrio.example$dat.hap,KnockoffTrio.example$pos,M=10)
window<-KnockoffTrio(KnockoffTrio.example$dat,dat.ko,KnockoffTrio.example$pos)
result<-causal_loci(window,M=10,fdr=0.15)
```

create_knockoff *Create knockoff genotype data*

Description

Create knockoff genotype data using phased haplotype data.

Usage

```
create_knockoff(
  dat.hap,
  pos,
  M = 10,
  maxcor = 0.7,
  maxbp = 80000,
  phasing.dad = NA,
  phasing.mom = NA
)
```

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 80000.
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.

Value

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

Examples

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

KnockoffTrio *Calculate KnockoffTrio's feature statistics*

Description

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

Usage

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

Arguments

dat	A 3n*p matrix for the original 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 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.
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 KnockoffTrio'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".
xchr	A logical value indicating whether the analysis is for the X chromosome. When xchr is TRUE, the analysis is for the X chromosome and sex is required. When xchr is FALSE, the analysis is for the autosomes.
sex	A numeric vector of length n for the sex of offspring. 0s indicate females and 1s indicate males. Sex is required when xchr is TRUE.

Value

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

Examples

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

KnockoffTrio.example *Example data for KnockoffTrio*

Description

A toy example of the haplotype and genotype data for original trios

Usage

```
KnockoffTrio.example
```

Format

KnockoffTrio.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.

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

`meta_analysis` *Meta-analysis for KnockoffTrio*

Description

Meta-analysis for KnockoffTrio

Usage

```
meta_analysis(window, n = NA, M = 10)
```

Arguments

`window`	A list of windows for the analysis results from different cohorts/studies.
`n`	A positive integer vector for the number of trios in each cohort/study. For weighted meta-analysis, a study's weight is based on the number of trios. The default is NA for unweighted meta-analysis.
`M`	A positive integer for the number of knockoffs. The default is 10.

Value

A data frame for the meta-analysis results.

Examples

```
data(KnockoffTrio.example)
dat.ko<-create_knockoff(KnockoffTrio.example$dat.hap,KnockoffTrio.example$pos,M=10)
window<-data.matrix(KnockoffTrio(KnockoffTrio.example$dat,dat.ko,KnockoffTrio.example$pos))
window.list<-list(window,window)
window.meta<-meta_analysis(window.list,M=10)
result<-causal_loci(window.meta,M=10,fdr=0.15)
```

Index