Package 'GMAC'

December 15, 2017

Type Package

Title Genomic Mediation Analysis with Adaptive Confounding Adjustment

Version 3.0

Date 2017-12-14

Author Fan Yang, Jiebiao Wang and Lin Chen

Maintainer Fan Yang <yfcarolyn@gmail.com>

Description Performs genomic mediation
analysis with adaptive confounding adjustment (GMAC) pro-
posed by Yang et al. (2017) <doi:10.1101/078683>. It implements large scale
mediation analysis and adaptively selects potential confounding variables to
adjust for each mediation test from a pool of candidate confounders. The package
is tailored for but not limited to genomic mediation analysis (e.g., cis-gene
mediating trans-gene regulation pattern where an eQTL, its cis-linking gene
transcript, and its trans-gene transcript play the roles as treatment, mediator
and the outcome, respectively), restricting to scenarios with the presence of
cis-association (i.e., treatment-mediator association) and random eQTL (i.e.,
treatment).

License GPL

Depends parallel

LazyData TRUE

NeedsCompilation no

Repository CRAN

RoxygenNote 6.0.1

Date/Publication 2017-12-15 17:45:23 UTC

R topics documented:

---

dat   *Example data*

Description

This simulated data list is for demonstration.

Value

A list containing

known.conf	The known confounders matrix which is adjusted in all mediation tests. Each row is a confounder, each column is a sample.
cov.pool	The pool of candidate confounding variables from which potential confounders are adaptively selected to adjust for each mediation test. Each row is a covariate, each column is a sample.
exp.dat	The gene expression matrix. Each row is for one gene, each column is a sample.
snp.dat.cis	The cis-eQTL genotype matrix. Each row is an eQTL, each column is a sample.
trios.idx	The matrix of selected trios indexes (row numbers) for mediation tests. Each row consists of the index (i.e., row number) of the eQTLs in snp.dat.cis, the index of cis-gene transcript in exp.dat, and the index of trans-gene transcript in exp.dat. The dimension is the number of trios by three.

Examples

```
data(example)
```

---

gmac   *Genomic Mediation analysis with Adaptive Confounding adjustment*

Description

The gmac function performs genomic mediation analysis with adaptive confounding adjustment. It tests for mediation effects for a set of user specified mediation trios (e.g., eQTL, cis- and trans-genes) in the genome with the assumption of the presence of cis-association. The gmac function considers either a user provided pool of potential confounding variables, real or constructed by other methods, or all the PCs based on the expression data as the potential confounder pool. It returns the mediation p-values and the proportions mediated (e.g., the percentage of reduction in trans-effects after accounting for cis-mediation), based on the mediation tests i) adjusting for known confounders only, and ii) adjusting for known confounders and adaptively selected potential confounders for each mediation trio. It also provides plots of mediation p-values (in the negative of log base of 10) versus the proportions mediated based on the above two adjustments.

Usage

```
gmac(cl = NULL, known.conf, cov.pool = NULL, exp.dat, snp.dat.cis, trios.idx,
     nperm = 10000, fdr = 0.05, fdr_filter = 0.1, nominal.p = FALSE)
```

Arguments

`cl`	Parallel backend if it is set up. It is used for parallel computing.
`known.conf`	A known confounders matrix which is adjusted in all mediation tests. Each row is a confounder, each column is a sample.
`cov.pool`	The pool of candidate confounding variables from which potential confounders are adaptively selected to adjust for each mediation test. Each row is a covariate, each column is a sample.
`exp.dat`	A gene expression matrix. Each row is for one gene, each column is a sample.
`snp.dat.cis`	The cis-eQTL genotype matrix. Each row is an eQTL, each column is a sample.
`trios.idx`	A matrix of selected trios indexes (row numbers) for mediation tests. Each row consists of the index (i.e., row number) of the eQTL in `snp.dat.cis`, the index of cis-gene transcript in `exp.dat`, and the index of trans-gene in `exp.dat`. The dimension is the number of trios by three.
`nperm`	The number of permutations for testing mediation.
`fdr`	The false discovery rate to select confounders. We set `fdr=0.05` as default.
`fdr_filter`	The false discovery rate to filter common child and intermediate variables. We set `fdr_filter=0.1` as default.
`nominal.p`	An option to obtain the nominal p-value or permutation-based p-value, which is the default.

Details

In genomic studies, a large number of mediation tests are often performed, and it is challenging to adjust for unmeasured confounding effects for the cis- and trans-genes (i.e., mediator-outcome) relationship. The current function adaptively selects the variables to adjust for each mediation trio given a large pool of constructed or real potential confounding variables. The function allows the input of variables known to be potential cis- and trans-genes (mediator-outcome) confounders in all mediation tests (`known.conf`), and the input of the pool of candidate confounders from which potential confounders for each mediation test will be adaptively selected (`cov.pool`). When no pool is provided (`cov.pool = NULL`), all the PCs based on expression data (`exp.dat`) will be constructed as the potential confounder pool.

The algorithm assumes the presence of cis-association (treatment-mediator association), random eQTL (treatment) and the standard identification assumption in causal mediation literature that no effect of eQTL (treatment) that confounds the cis- and trans-genes (mediator-outcome) relationship. The algorithm will first filter out common child (Figure 1.B) and intermediate variables (Figure 1.C) from `cov.pool` for each mediation trio at a pre-specified significance threshold of FDR (`fdr_filter`) by utilizing their associations with the eQTL (treatment). Then, confounder (Figure 1.A) set for each mediation trio will be selected from the retained pool of candidate variables using a stratified FDR approach. Specifically, for each trio, the p-values of association for each candidate variable to the cis-gene (mediator) and trans-gene (outcome) pairs are obtained based on

the F-test for testing the joint association to either the cis-gene (mediator) or the trans-gene (outcome). For each candidate variable, a pre-specified FDR (`fdr`) threshold is applied to the p-values corresponding to the joint associations of this variable to all the potential mediation trios. Lastly, mediation is tested for each mediation trio. Adjusting for the adaptively selected confounder set, we calculate the mediation statistic as the Wald statistic for testing the indirect mediation effect $H_0 : \beta_1 = 0$ based on the regression $T_j = \beta_0 + \beta_1 C_i + \beta_2 L_i + \tau X_{ij} + \epsilon$ where L_i, C_i, T_j and X_{ij} are the eQTL genotype (treatment), the cis-gene expression level (mediator), the trans-gene expression level (outcome) and the selected set of potential confounding variables. P-values are calculated based on within-genotype group permutation on the cis-gene expression level which maintains the cis- and trans-associations while breaks the potential mediation effect from the cis- to the trans-gene transcript.
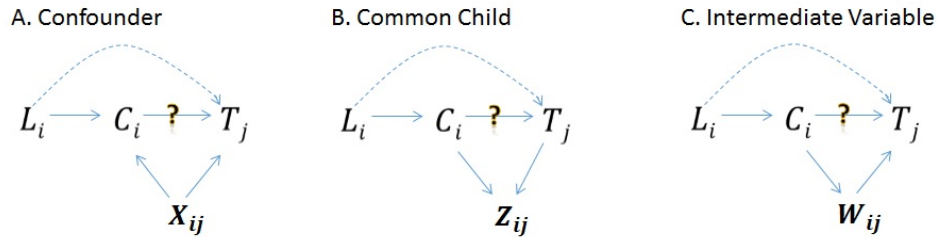

Figure 1. Graphical illustrations of (A) a potential mediation relationship among an eQTL L_i, its cis-gene transcript C_i, and a trans-gene transcript T_j, with confounders X_{ij}(i.e., variables affecting both C_i and T_j), allowing L_i to affect T_j via a pathway independent of C_i. For the mediation effect tests to have a causal interpretation, adjustment must be made for the confounders. (B) A potential mediation trio with common child variables, Z_{ij} (i.e., variables affected by both C_i and T_j). Adjusting for common child variables in mediation analysis would "marry" C_i and T_j and make C_i appearing to be regulating T_j even if there is no such effect. (C) A potential mediation trio with intermediate variables W_{ij} (i.e., variables affected by C_i and affecting T_j). Adjusting for intermediate variables in mediation analysis would prevent the detection of the true mediation effect from C_i to T_j.

The algorithm returns the mediation p-values (`pvals`) and the proportions mediated (`beta.change`, i.e., the percentage of reduction in trans-effects after accounting for cis-mediation), based on the mediation tests i) adjusting for known confounders only, and ii) adjusting for known confounders and adaptively selected potential confounders for each mediation trio. It also returns indicator matrix for the selected potential confounders (`sel.conf.ind`). Plots of mediation p-values (in the negative of log base of 10) versus the proportions mediated based on the adjustments i) and ii) are provided. The plot could further be used as a diagnostic check for sufficiency in confounding adjustment in scenarios such as cis-gene mediating trans-gene regulation pattern, where we expect the trios with very significant mediation p-values to have positive proportions mediated. Therefore, a J shape pattern is expected when most if not all confounding effects have been well adjusted, whereas a U shape pattern may indicate the presence of unadjusted confounders.

Value

The algorithm will return a list of p-values, beta changes, and indicator matrix for confounders selected.

pvals   The mediation p-values. A matrix with dimension of the number of trios by two ("Adjust Known Covariates Only", "Adjust Known + Selected Covariates").

beta.change	The proportions mediated. A matrix with dimension of the number of trios by two ("Adjust Known Covariates Only", "Adjust Known + Selected Covariates").
sel.conf.ind	An indicator matrix with dimension of the number of trios by the number of covariates in `cov.pool` or `pc.matrix` if the principal components (PCs) based on expression data are used as the pool of potential confounders.
pc.matrix	PCs will be returned if the PCs based on expression data are used as the pool of potential confounders. Each column is a PC.

References

Fan Yang, Jiebiao Wang, the GTEx consortium, Brandon L. Pierce, and Lin S. Chen. (2017) Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis. Genome Research. Volume 27, pp. 1859-1871. doi: 10.1101/078683

John D. Storey with contributions from Andrew J. Bass, Alan Dabney and David Robinson (2015). qvalue: Q-value estimation for false discovery rate control. R package version 2.8.0. doi: 10.18129/B9.bioc.qvalue

Examples

```
data(example)

# a fast example with only 50 permutations
output <- gmac(known.conf = dat$known.conf, cov.pool = dat$cov.pool,
    exp.dat = dat$exp.dat, snp.dat.cis = dat$snp.dat.cis,
    trios.idx = dat$trios.idx[1:40, ], nperm = 50, nominal.p = TRUE)

plot(output)


## Not run:
## the construction of PCs as cov.pool
pc <- prcomp(t(dat$exp.dat), scale = T)
cov.pool <- t(pc$x)


## generate a cluster with 2 nodes for parallel computing
cl <- makeCluster(2)
output <- gmac(cl = cl, known.conf = dat$known.conf, cov.pool = cov.pool,
    exp.dat = dat$exp.dat, snp.dat.cis = dat$snp.dat.cis,
    trios.idx = dat$trios.idx, nominal.p = TRUE)
stopCluster(cl)

## End(Not run)
```

Index