Package: hJAM (via r-universe)

September 15, 2024

Encoding UTF-8

Type Package

Title Hierarchical Joint Analysis of Marginal Summary Statistics

Version 1.0.1

Author Lai Jiang <jian848@usc.edu>

Maintainer Lai Jiang <jian848@usc.edu>

Description Provides functions to implement a hierarchical approach which is designed to perform joint analysis of summary statistics using the framework of Mendelian Randomization or transcriptome analysis. Reference: Lai Jiang, Shujing Xu, Nicholas Mancuso, Paul J. Newcombe, David V. Conti (2020). ``A Hierarchical Approach Using Marginal Summary Statistics for Multiple Intermediates in a Mendelian Randomization or Transcriptome Analysis.'' <bioRxiv><doi:10.1101/2020.02.03.924241>.

License MIT + file LICENSE

LazyData true

RoxygenNote 7.2.3

Suggests knitr, rmarkdown

VignetteBuilder knitr

URL https://github.com/USCbiostats/hJAM

BugReports https://github.com/USCbiostats/hJAM/issues

Depends R (>= 3.5.0)

Imports ggplot2, reshape2, glmnet, susieR, dplyr, utils, tibble, matrixcalc, Rmpfr

Remotes github::lailylajiang/susieR

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

RemoteUrl https://github.com/uscbiostats/hjam

RemoteRef HEAD

RemoteSha a15efa0fcc216f2e0b79e240dd85636cc9f29fc8

Contents

`EN.hJAM` *Elastic net hJAM*

Description

Function to implement regularized hJAM, including elastic net hJAM and lasso hJAM.

Usage

```
EN.hJAM(
  betas.Gy,
  N.Gy,
  eaf.Gy = NULL,
  Geno,
  A,
  tune_glmnet = 0.5,
  ridgeTerm = FALSE
)
```

Arguments

`betas.Gy`	The betas in the paper: the marginal effects of SNPs on the phenotype (Gy)
`N.Gy`	The sample size of the GWAS where you obtain the betas.Gy and betas_se.Gy
`eaf.Gy`	The effect allele frequency of the SNPs in betas.Gy
`Geno`	The individual level data of the reference panel. Must have the same order of SNPs as in the betas.Gy.
`A`	The conditional A matrix.
`tune_glmnet`	The α used in the glmnet R package to tune the shrinkage parameter. Default is 0.5.
`ridgeTerm`	Add a small elelment to the diagnoal of X'X to make the matrix invertable.

Value

An object of the Regularized hJAM

numSNP The number of SNPs that the user use in the instrument set.

Selected_variable_length The number of selected intermediates, regardless of the credible sets.

Selected_variable_name The label/name for each selected intermediates.

Coefficients The coefficients of selected intermediates. Otherwise will be zero.

Author(s)

Lai Jiang

Examples

```
data(ENhJAM.SimulationSet)
EN.hJAM(betas.Gy = Simulation.betas.gwas, N.Gy = 5000, eaf.Gy = Simulation.maf.gwas,
Geno = Simulation.Geno, A = Simulation.Amatrix, ridgeTerm = FALSE)
```

ENhJAM.SimulationSet *Simulation data for EN-hJAM*

Description

Simulation data for EN-hJAM

Format

The `ENhJAM.SimulationSet` is a set of simulation data sets for the example of elastic net hJAM

Simulation.Amatrix The conditional $\hat{A}$ matrix with 118 metabolites and 144 SNPs, which was composed by SuSiE JAM and the marginal $\hat{A}$ matrix.

Simulation.Geno The reference genotype data for the 144 SNPs from the European-ancestry population in 1000 Genome Project (Consortium, 2015).

Simulation.betas.gwas The b vector. The association estimates between selected SNPs and the risk of prostate cancer from (Schumacher et al., 2018)

Simulation.betas.se.gwas The se(b) vector from (Schumacher et al., 2018)

Simulation.maf.gwas The vector of the effect allele frequency of the SNPs from (Schumacher et al., 2018)

get_XtX *Get transformed statistics: XtX*

Description

To calculate sufficient statistics based on summary statistics

To calculate sufficient statistics based on summary statistics

Usage

```
get_XtX(N_outcome, Gl, maf)

get_XtX(N_outcome, Gl, maf)
```

Arguments

`N_outcome`	Sample size in the GWAS where we obtained ‘betas‘
`Gl`	A matrix of reference dosage, columns are SNPs and rows are individuals.
`maf`	A vector of minor allele frequencies

Value

a variance covariance matrix of scaled Gl

a variance covariance matrix of scaled Gl

get_yty *Get transformed statistics: yty*

Description

To calculate sufficient statistics based on summary statistics. This yty estimate follows Yang et al. (2012) Nat Gen. Marginal estimates from one SNP will produce one yty estimates. Yang suggests taking the median across all SNPs to obtain a robust estimate. Here we record all yty estimates and output both the median and the entire vector.

To calculate sufficient statistics based on summary statistics. This yty estimate follows Yang et al. (2012) Nat Gen. Marginal estimates from one SNP will produce one yty estimates. Yang suggests taking the median across all SNPs to obtain a robust estimate. Here we record all yty estimates and output both the median and the entire vector.

Usage

```
get_yty(maf, N_outcome, betas, betas.se)

get_yty(maf, N_outcome, betas, betas.se)
```

Arguments

`maf`	A vector of minor allele frequencies
`N_outcome`	Sample size in the GWAS where we obtained 'betas'
`betas`	A vector of marginal estimates of effect sizes (betas for continuous outcome; logOR for binary outcome)
`betas.se`	A vector of the standard errors of marginal effect estimates ('betas').

Value

median of yty estimates across all SNPs; and a vector of all yty estimates

median of yty estimates across all SNPs; and a vector of all yty estimates

get_z *Get transformed statistics: z, or Xty*

Description

To calculate sufficient statistics based on summary statistics

To calculate sufficient statistics based on summary statistics

Usage

```
get_z(maf, betas, N_outcome)

get_z(maf, betas, N_outcome)
```

Arguments

`maf`	A vector of minor allele frequencies
`betas`	A vector of marginal estimates of effect sizes (betas for continuous outcome; logOR for binary outcome)
`N_outcome`	Sample size in the GWAS where we obtained 'betas'

Value

a numeric vector of calculated z statistic

a numeric vector of calculated z statistic

`GTEx.PrCa` *Real data for selecting the genes on chromosome 10 for the risk of prostate cancer*

Description

Real data for selecting the genes on chromosome 10 for the risk of prostate cancer

Format

The `GTEx.PrCa` is a set of data sets which was applied for selecting the genes on chromosome 10 for the risk of prostate cancer

GTEx.PrCa.IVWmarginal.A The marginal $\hat{A}$ matrix with 158 genes and 182 eQTLs. The raw data was downloaded from GTEx analysis v7 (https://gtexportal.org/home/datasets). Priority Pruner was used to select the independent eQTLs. We used this matrix for MR-BMA implementation.)

GTEx.PrCa.marginal.A The marginal $\hat{A}$ matrix with 167 genes and 447 eQTLs. The raw data was downloaded from GTEx analysis v7 (https://gtexportal.org/home/datasets). This is the raw $\hat{A}$ matrix for constructing the conditional weight matrix for SHA-JAM analysis.

GTEx.PrCa.marginal.A.se The standard errors of the marginal $\hat{A}$ effects for the SNP-gene pairs (167 genes, 447 eQTLs). The raw data was downloaded from GTEx analysis v7 (https://gtexportal.org/home/datasets).

GTEx.PrCa.inclusion.indicator The inclusion indicator for the significant SNP-gene pairs (167 genes, 447 eQTLs). Significant as 1; otherwise 0. This matrix is for composing the conditional weight matrix using the raw data.

GTEx.PrCa.Amatrix The conditional $\hat{A}$ matrix with 167 genes and 447 eQTLs, which was composed by SuSiE JAM and the raw data of $\hat{A}$ matrix.

GTEx.PrCa.Geno The reference genotype data for the 447 eQTLs from the European-ancestry population in 1000 Genome Project (Consortium, 2015)

GTEx.PrCa.betas.gwas The b vector. The association estimates between eQTLs and the risk of prostate cancer from (Schumacher et al., 2018)

GTEx.PrCa.betas.se.gwas The se(b) vector from (Schumacher et al., 2018)

GTEx.PrCa.pvalue.gwas The pvalues vector of the association estimates between selected SNPs and the risk of prostate cancer from (Schumacher et al., 2018)

GTEx.PrCa.maf.gwas The vector of the effect allele frequency of the SNPs from (Schumacher et al., 2018)

References

Consortium GP. A global reference for human genetic variation. Nature 2015; 526: 68.

Lonsdale, John, et al. The genotype-tissue expression (GTEx) project. Nature genetics 45.6 (2013): 580-585.

Schumacher, Fredrick R., et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nature genetics 50.7 (2018): 928-936.

`hJAM` *hJAM Fit hJAM with linear regression*

Description

The hJAM function is to get the results from the hJAM model using input data

Usage

```
hJAM(betas.Gy, N.Gy, Geno, A, ridgeTerm = FALSE)
```

Arguments

`betas.Gy`	The betas in the paper: the marginal effects of SNPs on the phenotype (Gy)
`N.Gy`	The sample size of Gy
`Geno`	The reference panel (Geno), such as 1000 Genome
`A`	The A matrix in the paper: the marginal/conditional effects of SNPs on the exposures (Gx)
`ridgeTerm`	ridgeTerm = TRUE when the matrix L is singular. Matrix L is obtained from the cholesky decomposition of G0'G0. Default as FALSE.

Value

An object of the hJAM with linear regression results.

Exposure The intermediates, such as the modifiable risk factors in Mendelian Randomization and gene expression in transcriptome analysis.

numSNP The number of SNPs that the user use in the instrument set.

Estimate The conditional estimates of the associations between intermediates and the outcome.

StdErr The standard error of the conditional estimates of the associations between intermediates and the outcome.

Lower.CI The lower bound of the 95% confidence interval of the estimates.

Upper.CI The upper bound of the 95% confidence interval of the estimates.

Pvalue The p value of the estimates with a type-I error equals 0.05.

Author(s)

Lai Jiang

References

Lai Jiang, Shujing Xu, Nicholas Mancuso, Paul J. Newcombe, David V. Conti (2020). A Hierarchical Approach Using Marginal Summary Statistics for Multiple Intermediates in a Mendelian Randomization or Transcriptome Analysis. *bioRxiv* https://doi.org/10.1101/2020.02.03.924241.

Examples

```
data(MI)
hJAM(betas.Gy = MI.betas.gwas, Geno = MI.Geno, N.Gy = 459324, A = MI.Amatrix, ridgeTerm = TRUE)
```

`hJAM_egger` *hJAM_egger Fit hJAM with Egger regression*

Description

The hJAM_egger function is to get the results from the hJAM model with Egger regression. It is for detecting potential pleiotropy

Usage

```
hJAM_egger(betas.Gy, N.Gy, Geno, A, ridgeTerm = TRUE)
```

Arguments

`betas.Gy`	The betas in the paper: the marginal effects of SNPs on the phenotype (Gy)
`N.Gy`	The sample size of Gy
`Geno`	The reference panel (Geno), such as 1000 Genome
`A`	The A matrix in the paper: the marginal/conditional effects of SNPs on the exposures (Gx)
`ridgeTerm`	ridgeTerm = TRUE when the matrix L is singular. Matrix L is obtained from the cholesky decomposition of G0'G0. Default as TRUE

Value

An object of the hJAM with egger regression results.

Exposure The intermediates, such as the modifiable risk factors in Mendelian Randomization and gene expression in transcriptome analysis.

numSNP The number of SNPs that the user use in the instrument set.

Estimate The conditional estimates of the associations between intermediates and the outcome.

StdErr The standard error of the conditional estimates of the associations between intermediates and the outcome.

Lower.CI The lower bound of the 95% confidence interval of the estimates.

Upper.CI The upper bound of the 95% confidence interval of the estimates.

Pvalue The p value of the estimates with a type-I error equals 0.05.

Est.Int The intercept of the regression of intermediates on the outcome.

StdErr.Int The standard error of the intercept of the regression of intermediates on the outcome.

Lower.CI.Int The lower bound of the 95% confidence interval of the intercept.

Upper.CI.Int The upper bound of the 95% confidence interval of the intercept.

Pvalue.Int The p value of the intercept with a type-I error equals 0.05.

An object of hJAM with egger regression results.

Author(s)

Lai Jiang

References

Lai Jiang, Shujing Xu, Nicholas Mancuso, Paul J. Newcombe, David V. Conti (2020). A Hierarchical Approach Using Marginal Summary Statistics for Multiple Intermediates in a Mendelian Randomization or Transcriptome Analysis. *bioRxiv* https://doi.org/10.1101/2020.02.03.924241.

Examples

```
data(MI)
hJAM_egger(betas.Gy = MI.betas.gwas, Geno = MI.Geno, N.Gy = 459324, A = MI.Amatrix)
```

JAM_A *Compute conditional A matrix*

Description

The JAM_A function is to get the conditional A matrix by using marginal A matrix

Usage

```
JAM_A(marginalA, Geno, N.Gx, eaf_Gx = NULL, ridgeTerm = TRUE)
```

Arguments

marginalA	the marginal effects of SNPs on the exposures (Gx).
Geno	the reference panel (Geno), such as 1000 Genome
N.Gx	the sample size of each Gx. It can be a scalar or a vector. If there are multiple X's from different Gx, it should be a vector including the sample size of each Gx. If all alphas are from the same Gx, it could be a scalar.
eaf_Gx	the effect allele frequency of the SNPs in the Gx data.
ridgeTerm	ridgeTerm = TRUE when the matrix L is singular. Matrix L is obtained from the cholesky decomposition of G0'G0. Default as TRUE.

Value

A matrix with conditional estimates which are converted from marginal estimates using the JAM model.

Author(s)

Lai Jiang

Examples

```
data(MI)
JAM_A(marginalA = MI.marginal.Amatrix, Geno = MI.Geno, N.Gx = c(339224, 659316), ridgeTerm = TRUE)
JAM_A(marginalA = MI.marginal.Amatrix, Geno = MI.Geno, N.Gx = c(339224, 659316),
eaf_Gx = MI.SNPs_info$ref_frq)
```

`JAM_alphas` *Compute conditional alphas*

Description

The `JAM_alphas` function is to compute the conditional alpha vector for each X If only one X in the model, please use JAM_alphas instead of JAM_A A sub-step in the JAM_A function

Usage

```
JAM_alphas(marginalA, Geno, N.Gx, eaf_Gx = NULL, ridgeTerm = TRUE)
```

Arguments

`marginalA`	the marginal effects of SNPs on one exposure (Gx).
`Geno`	the reference panel (Geno), such as 1000 Genome
`N.Gx`	the sample size of the Gx. It can be a scalar.
`eaf_Gx`	the effect allele frequency of the SNPs in the Gx data.
`ridgeTerm`	ridgeTerm = TRUE when the matrix L is singular. Matrix L is obtained from the cholesky decomposition of G0'G0. Default as TRUE.

Value

A vector with conditional estimates which are converted from marginal estimates using the JAM model.

Author(s)

Lai Jiang

References

Lai Jiang, Shujing Xu, Nicholas Mancuso, Paul J. Newcombe, David V. Conti (2020). A Hierarchical Approach Using Marginal Summary Statistics for Multiple Intermediates in a Mendelian Randomization or Transcriptome Analysis. *bioRxiv* https://doi.org/10.1101/2020.02.03.924241.

Examples

```
data(MI)
JAM_alphas(marginalA = MI.marginal.Amatrix[, 1], Geno = MI.Geno, N.Gx = 339224)
JAM_alphas(marginalA = MI.marginal.Amatrix[, 1], Geno = MI.Geno, N.Gx = 339224,
eaf_Gx = MI.SNPs_info$ref_frq)
```

`LogisticToLinearEffects` *Transform log odds ratios to linear effects*

Description

Adopted from R2BGLiMS::JAM_LogisticToLinearEffects. Reference: Benner 2015, FINEMAP

Adopted from R2BGLiMS::JAM_LogisticToLinearEffects. Reference: Benner 2015, FINEMAP

Usage

```
LogisticToLinearEffects(
  log.ors = NULL,
  log.or.ses = NULL,
  snp.genotype.sds = NULL,
  mafs = NULL,
  n = NULL,
  p.cases = NULL
)

LogisticToLinearEffects(
  log.ors = NULL,
  log.or.ses = NULL,
  snp.genotype.sds = NULL,
  mafs = NULL,
  n = NULL,
  p.cases = NULL
)
```

Arguments

`log.ors`	A vector of log odds ratios
`log.or.ses`	A vector of the standard errors of the log ORs
`snp.genotype.sds`	A vector of standard deviations of genotypes (optional if 'mafs' is provided)
`mafs`	A vector of effective allele frequencies (optional if 'snp.genotype.sds' is provided)
`n`	Sample size in the GWAS where we obtained 'log.ors'
`p.cases`	A numeric value of the proportion of cases in the GWAS.

Value

Transformed linear effect estimates, and transformed standards errors of linear effects.

Transformed linear effect estimates, and transformed standards errors of linear effects.

`MI` *Example data of hJAM*

Description

Real data for BMI/T2D on the risk of Myocardial infarction

Format

The `MI` object is a set of data sets which was used to estimate the causal effect of body mass index and type 2 diabetes on the risk of myocardial infarction.

MI.marginal.Amatrix The marginal $\hat{A}$ matrix. Column one and two are the marginal estimates of the SNPs on body mass index from GIANT consortium (n = 339,224) (Locke et al., 2015) and type 2 diabetes from DIAGRAM+GERA+UKB (n = 659,316) (Xue et al., 2018), respectively

MI.Amatrix The conditional $\hat{A}$ matrix composed by JAM and the marginal $\hat{A}$ matrix. Column one and two are the conditional effect estimates of the SNPs on body mass index and type 2 diabetes, respectively.

MI.Geno The reference genotype data from the European-ancestry population in 1000 Genome Project (Consortium, 2015).

MI.betas.gwas The b vector. The association estimates between selected SNPs and the risk of myocardial infarction from UK Biobank (Sudlow et al., 2015).

MI.SNPs_info The SNP information. Five columns included: the RSID, reference allele, reference allele frequency, if BMI significant and if T2D significant. The last two columns are indicator variables for the SNPs which are genome-wide significant associated with BMI/T2D.

References

Consortium GP. A global reference for human genetic variation. Nature 2015; 526: 68.

Locke, Adam E., et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 518.7538 (2015): 197-206.

Xue, Angli, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nature communications 9.1 (2018): 1-14.

Sudlow, Cathie, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. Plos med 12.3 (2015): e1001779.

`mJAM_build_CS` *Construct mJAM credible set based for selected index SNP*

Description

Construct mJAM credible set based for selected index SNP

Usage

```
mJAM_build_CS(
  X_id,
  prev_X_list = NULL,
  All_id,
  PrCS_weights = "Pr(M_C)",
  coverage = 0.95,
  GItGI_curr,
  GIty_curr,
  yty_curr,
  yty_med,
  N_GWAS,
  rare_SNPs = NULL,
  Pr_Med_cut = 0.1,
  use_robust_var_est = FALSE
)
```

Arguments

`X_id`	A character specifying the ID of the index SNP; should be found in 'All_id'.
`prev_X_list`	A list of character vector of the ID(s) of previously selected index SNP(s).
`All_id`	A list of character vector of the ID(s) of all SNP(s) remaining in the analysis, including all previously selected SNP(s) and the current index SNP.
`PrCS_weights`	An option to specify what weights to apply on Pr(Med). Default is "Pr(M_C)".
`coverage`	A number between 0 and 1 specifying the "coverage" of the estimated confidence sets.
`GItGI_curr`	A list of GItGI statistics at the current stage (after pruning out SNPs correlated with previously selected index SNPs).
`GIty_curr`	A list of GIty estimates of all remaining SNPs at the current stage (after pruning out SNPs correlated with previously selected index SNPs).
`yty_curr`	A list of yty estimates of all remaining SNPs at the current stage (after pruning out SNPs correlated with previously selected index SNPs).
`yty_med`	A list of median yty across all SNPs.

`N_GWAS` A vector of sample sizes in all original GWAS studies.

`rare_SNPs` A numeric vector of ID(s) for rare SNP(s) which we do not apply weighting. Instead, we use the individual estimate of yty for these SNPs for robustness.

`Pr_Med_cut` The cutoff for Pr(Mediation); SNPs with Pr(Mediation) smaller than this cutoff will be assigned a Pr(CS) = 0 and thus not included in the credible set for the current index

`use_robust_var_est`
whether to use linear combination of median yty and individual yty.

Value

A table with the following columns:

CS_SNP SNP name.

Post_Model_Prob The posterior Pr(Model) of this SNP on its absolute scale.

Post_Model_Prob_Ratio The posterior Pr(Model) of this SNP divided by the posterior Pr(Model) of index SNP. It should be <= 1.

Post_Model_Prob_Ratio2 If 'Post_Model_Prob_Ratio' is greater than 1, set 'Post_Model_Prob_Ratio2' to 1. Otherwise, it's the same as 'Post_Model_Prob_Ratio'.

Med_Effect_Size The posterior mediation effect size.

Post_Med_Prob The posterior Pr(Mediation) of this SNP.

Post_Med_Prob2 If 'Post_Med_Prob' is less than 'Pr_Med_cut', set 'Post_Med_Prob2' to 0. Otherwise, it's the same as 'Post_Med_Prob'.

SD_Post_CS_Prob Standardized Pr(CS) where Pr(CS) = Pr(Model)*Pr(Mediation)

CumSum_Porb The cumulative 'SD_Post_CS_Prob'. Note that the table is ordered by descending 'SD_Post_CS_Prob'.

EmpiricalCut The empirical coverage of this CS (should be >= requested 'coverage').

CS_in A logical variable indicating whether this CS_SNP is included in this CS or not.

index_SNP The name of the index SNP.

Author(s)

Jiayi Shen

`mJAM_Forward` *Run mJAM with Forward Selection*

Description

fitting mJAM-Forward

Usage

```
mJAM_Forward(
  N_GWAS,
  X_ref,
  Marg_Result,
  EAF_Result,
  condp_cut = NULL,
  index_snps = NULL,
  within_pop_threshold = 0.5,
  across_pop_threshold = 0.2,
  coverage = 0.95,
  Pr_Med_cut = 0,
  filter_rare = FALSE,
  rare_freq = NULL,
  filter_unstable_est = FALSE,
  use_robust_var_est = FALSE
)
```

Arguments

N_GWAS	A vector of sample sizes in all original GWAS studies.
X_ref	A list of matrices with individual-level SNP dosage data in each study/population. Each column corresponds to a SNP. Note that the columns name should match exactly to the SNP column in 'Marg_Result' and 'EAF_Result'. If certain SNP(s) is missing in dosage, then insert NAs in corresponding column(s).
Marg_Result	A data frame with marginal summary statistics from all studies. Col1: SNP name; Col2: Effect sizes from study #1; Col3: Std Errors of effect sizes from study #1; ...
EAF_Result	A data frame with effect allele frequency (EAF) from all studies. Col1: SNP name; Col2: EAF from study #1; Col3: EAF from study #2; ...
condp_cut	Threshold of conditional p-value to be considered as significant. No default specified. Usually recommend 5e-8.
index_snps	User-defined index SNP(s), if any. Default is 'NULL' which means mJAM-Forward will automatically select index variants.
within_pop_threshold	Threshold of r2 with selected index SNP(s) within a single population. If a SNP's correlation with any selected index SNP is greater than this threshold in at least one population, it will be excluded from subsequent rounds of index SNP selection.
across_pop_threshold	Threshold of r2 with selected index SNP(s) across all populations. If a SNP's correlation with any selected index SNP is greater than this threshold in all populations, it will be excluded from subsequent rounds of index SNP selection.
coverage	The required coverage of credible sets. Default is 0.95.

`Pr_Med_cut` Cut off of mJAM posterior mediation probability (P(Med)) during credible set construction. Low P(Med) may indicate low correlation between the candidate SNP and the index SNP. Any candidate credible set SNPs with P(Med) < Pr_Med_cut will be not be considered for credible set. Default is 0.

`filter_rare` A logical variable indicating whether to filter rare SNPs before the analysis. Default is 'FALSE.' If 'TRUE', then please specify 'rare_freq'.

`rare_freq` A vector of frequencies between 0 and 0.5 to specify the minor allele frequency cut-off if you want to filter rare SNPs before the analysis. Please also set 'filter_rare' to be TRUE. For example, if there are 3 populations, then rare_freq = c(0.01, 0, 0.01) means SNPs with MAF < 0.01 in pop 1 and MAF < 0.01 in pop 3 will be removed from analysis.

`filter_unstable_est`
whether to filter variants with inconsistent estimate between mJAM and meta-analysis.

`use_robust_var_est`
whether to use the robust estimate of residual variance (weighting between median and individual estimates).

Value

index A table listing all the selected index SNP(s) ('SNP'), along with their log10(p-value) conditional on all SNP(s) above ('cond_log10p'), the log10(p-value) conditional on all other index SNP(s) ('final_log10p'), and the p-value threshold used in this analysis ('pcut').

cs A table recording various posterior probabilities of all SNPs being considered for credible set SNPs.

mJAM_marg_est A table with the marginal effect estimates and standard errors of all SNPs under the mJAM model.

QC_marg_est The complete table of marginal effect estimates using fixed-effect model and mJAM model. For QC purpose only.

Author(s)

Jiayi Shen

`mJAM_get_condp` *Get conditional p-value under mJAM model*

Description

Get conditional p-value under mJAM model

Usage

```
mJAM_get_condp(
  GItGI,
  GIty,
  yty,
  yty_med,
  N_GWAS,
  g = NULL,
  selected_id,
  use_robust_var_est = FALSE,
  use_median_yty_ethnic = NULL,
  rare_id = NULL
)
```

Arguments

GItGI	A list of transformed statistics from 'get_XtX()' for each study.
GIty	A list of transformed statistics from 'get_z()' for each study.
yty	A list of transformed statistics from 'get_yty()' for each study.
yty_med	A numeric vector of median yty across all SNPs within each study.
N_GWAS	A numeric vector of GWAS sample size for each study.
g	Hyperparameter in g-prior. If 'NULL', it will be set to 'sum(N_GWAS)'.
selected_id	A numeric vector of IDs of previously selected index SNP(s).
use_robust_var_est	whether to use linear combination of median yty and individual yty.
use_median_yty_ethnic	A numeric vector of study index in which median_yty is used for all SNPs in 'selected_id'.
rare_id	A numeric vector of IDs for rare SNP(s) which we do not apply weighting. Instead, we use the individual estimate of yty for these SNPs for robustness.

Value

which_condp_min The index of which SNP has the smallest conditional p-value.

condp_min The smallest conditional p-value.

condp A vector of all conditional p-values.

effect_est A vector of all conditional effect estimates.

se_est A vector of standard errors of all the conditional effect estimates.

condp_mx A complete matrix recording all conditional effect est & se for testing SNPs and 'selected_id'.

Author(s)

Jiayi Shen

`mJAM_get_condp_selected` *Get conditional p-value for selected (index SNPs) under mJAM model*

Description

Get conditional p-value for selected (index SNPs) under mJAM model

Usage

```
mJAM_get_condp_selected(
  GItGI,
  GIty,
  yty,
  yty_med,
  N_GWAS,
  g = NULL,
  selected_id,
  use_robust_var_est = FALSE,
  use_median_yty_ethnic = NULL,
  rare_SNPs = NULL
)
```

Arguments

`GItGI`	A list of transformed statistics from 'get_XtX()' for each study.
`GIty`	A list of transformed statistics from 'get_z()' for each study.
`yty`	A list of transformed statistics from 'get_yty()' for each study.
`yty_med`	A numeric vector of median yty across all SNPs within each study.
`N_GWAS`	A numeric vector of GWAS sample size for each study.
`g`	Hyperparameter in g-prior. If 'NULL', it will be set to 'sum(N_GWAS)'.
`selected_id`	A numeric vector of IDs of previously selected index SNP(s).
`use_robust_var_est`	whether to use linear combination of median yty and individual yty. (only for mJAM-Forward)
`use_median_yty_ethnic`	A numeric vector of study index in which median_yty is used for all SNPs in 'selected_id'.
`rare_SNPs`	A character vector for rare SNP(s) which we do not apply weighting. Instead, we use the individual estimate of yty for these SNPs for robustness.

Value

b_joint The estimated conditional effect size when all SNPs in 'selected_id' are in one mJAM model.

b_joint_var The variance of 'b_joint'.

condp A vector of all conditional p-values for 'b_joint'.

Author(s)

Jiayi Shen

`mJAM_get_PrM` *Get Pr(Model) based on BF-type model probability*

Description

Also apply weighting to get robust estimates of yty

Usage

```
mJAM_get_PrM(
  GItGI,
  GIty,
  yty,
  yty_med,
  N_GWAS,
  C_id,
  prev_X_list = NULL,
  g = NULL,
  rare_SNPs = NULL,
  use_robust_var_est = FALSE
)
```

Arguments

`GItGI`	A list of transformed statistics from 'get_XtX()' for each study.
`GIty`	A list of transformed statistics from 'get_z()' for each study.
`yty`	A list of transformed statistics from 'get_yty()' for each study.
`yty_med`	A numeric vector of median yty across all SNPs within each study.
`N_GWAS`	A numeric vector of GWAS sample size for each study.
`C_id`	An ingeter vector of IDs for the SNPs to be tested.
`prev_X_list`	A numeric vector of the ID(s) of previously selected index SNP(s).
`g`	The pre-specified 'g' in 'g'-prior formulation.
`rare_SNPs`	A numeric vector of ID(s) for rare SNP(s) which we do not apply weighting. Instead, we use the individual estimate of yty for these SNPs for robustness.
`use_robust_var_est`	whether to use linear combination of median yty and individual yty.

Value

post_prob Posterior Pr(Model) for each SNPs in 'C_id'.

R2_est R2 estimates of every one-SNP model (one for each SNPs in 'C_id').

n_miss An integer vector of how many studies have missing values for each SNP.

Author(s)

Jiayi Shen

`mJAM_get_PrMed` *Get Pr(Mediation) based on causal mediation models*

Description

Also apply weighting to get robust estimates of yty

Usage

```
mJAM_get_PrMed(
  GItGI,
  GIty,
  yty,
  yty_med,
  N_GWAS,
  g = NULL,
  C_id,
  X_id,
  prev_X_list
)
```

Arguments

`GItGI`	A list of transformed statistics from 'get_XtX()' for each study.
`GIty`	A list of transformed statistics from 'get_z()' for each study.
`yty`	A list of transformed statistics from 'get_yty()' for each study.
`yty_med`	A numeric vector of median yty across all SNPs within each study.
`N_GWAS`	A numeric vector of GWAS sample size for each study.
`g`	The pre-specified 'g' in 'g'-prior formulation.
`C_id`	An ingeter vector of IDs for the SNPs to be tested.
`X_id`	An integer specifying the ID of the index SNP.
`prev_X_list`	A numeric vector of the ID(s) of previously selected index SNP(s).

Value

Post_Med_Prob Posterior Pr(Mediation) for each SNPs in C_id.
Med_Effect_Size Posterior mediation effect size for each SNPs in C_id.
Med_var_CX Posterior variance of mediation effect in models with both C and X.
Med_var_C Posterior variance of mediation effect in models with C only.

Author(s)

Jiayi Shen

---

mJAM_get_PrM_Wald *Get Pr(Model) based on Wald-type model probability*

Description

Also apply weighting to get robust estimates of yty

Usage

```
mJAM_get_PrM_Wald(
  GItGI,
  GIty,
  yty,
  yty_med,
  N_GWAS,
  C_id,
  prev_X_list = NULL,
  g = NULL,
  rare_SNPs = NULL,
  use_robust_var_est = FALSE
)
```

Arguments

GItGI	A list of transformed statistics from 'get_XtX()' for each study.
GIty	A list of transformed statistics from 'get_z()' for each study.
yty	A list of transformed statistics from 'get_yty()' for each study.
yty_med	A numeric vector of median yty across all SNPs within each study.
N_GWAS	A numeric vector of GWAS sample size for each study.
C_id	An ingeter vector of IDs for the SNPs to be tested.
prev_X_list	A numeric vector of the ID(s) of previously selected index SNP(s).
g	The pre-specified 'g' in 'g'-prior formulation.
rare_SNPs	A numeric vector of ID(s) for rare SNP(s) which we do not apply weighting. Instead, we use the individual estimate of yty for these SNPs for robustness.
use_robust_var_est	whether to use linear combination of median yty and individual yty.

Value

A numeric vector of posterior Pr(Model) for each SNPs in 'C_id'.

Author(s)

Jiayi Shen

`mJAM_LDpruning` *Pruning SNPs based on LD*

Description

Pruning SNPs based on LD

Usage

```
mJAM_LDpruning(target, testing, R, within_thre = 0.95, across_thre = 0.8)
```

Arguments

`target`	Target SNP ID.
`testing`	IDs of SNPs to be tested.
`R`	a list of correlation matrix of all SNPs.
`within_thre`	threshold of r2 with selected index SNP(s) within a single population. If a SNP's correlation with any selected index SNP is greater than this threshold in at least one population, it will be excluded from subsequent rounds of index SNP selection.
`across_thre`	threshold of r2 with selected index SNP(s) across all populations. If a SNP's correlation with any selected index SNP is greater than this threshold in all populations, it will be excluded from subsequent rounds of index SNP selection.

Value

remove_within SNP IDs to be pruned due to high within-population correlation

remove_across SNP IDs to be pruned due to high across-population correlation

Author(s)

Jiayi Shen

`mJAM_SuSiE` *Run mJAM with SuSiE*

Description

fitting mJAM-SuSiE

Usage

```
mJAM_SuSiE(
  Marg_Result = NULL,
  EAF_Result = NULL,
  N_GWAS,
  X_ref,
  filter_rare = FALSE,
  rare_freq = NULL,
  SuSiE_num_comp = 10,
  SuSiE_coverage = 0.95,
  SuSiE_min_abs_corr = 0.5,
  max_iter = 500,
  estimate_residual_variance = F
)
```

Arguments

`Marg_Result`	A data frame with marginal summary statistics from all studies. Col1: SNP name; Col2: Effect sizes from study #1; Col3: Std Errors of effect sizes from study #1; ...
`EAF_Result`	A data frame with effect allele frequency (EAF) from all studies. Col1: SNP name; Col2: EAF from study #1; Col3: EAF from study #2; ...
`N_GWAS`	A vector of sample sizes in all original GWAS studies.
`X_ref`	A list of matrices with individual-level SNP dosage data in each study/population.
`filter_rare`	A logical variable indicating whether to filter rare SNPs before the analysis. Default is 'FALSE.' If 'TRUE', then please specify 'rare_freq'.
`rare_freq`	A vector of frequencies between 0 and 0.5 to specify the minor allele frequency cut-off if you want to filter rare SNPs before the analysis. Please also set 'filter_rare' to be TRUE. For example, if there are 3 populations, then rare_freq = c(0.01, 0, 0.01) means SNPs with MAF < 0.01 in pop 1 and MAF < 0.01 in pop 3 will be removed from analysis.
`SuSiE_num_comp`	SuSiE argument. The maximum number of causal SNPs that you want to select. Default is 10.
`SuSiE_coverage`	SuSiE argument. The required coverage of credible sets. Default is 0.95.
`SuSiE_min_abs_corr`	SuSiE argument. Minimum absolute correlation allowed in a credible set.

`max_iter` SuSiE argument. Maximum iterations to perform.

`estimate_residual_variance`
SuSiE argument. If 'TRUE', then the susie algorithm is updating residual variance estimate during iterations. If 'FALSE', then use the residual variance is a fixed value, which is usually var(Y).

Value

summary A table of the SuSiE posterior inclusion probabilities (PIPs), posterior mean, and posterior sd of all SNPs.

fit SuSiE fit object.

Author(s)

Jiayi Shen

`mJAM_SuSiE_get_cs` *Get and tidy SuSiE credible sets*

Description

Get and tidy SuSiE credible sets

Usage

```
mJAM_SuSiE_get_cs(mjam_susie_res, coverage = 0.95)
```

Arguments

`mjam_susie_res`
The mJAM-SuSiE result returned from 'mJAM_SuSiE()'

`coverage` A number between 0 and 1 specifying the "coverage" of the estimated confidence sets.

Value

A table summary of SuSiE credible sets with the following columns:

#'

index The label for a distinct credible set.

coverage The empirical coverage of this credible set.

CS_size The number of SNPs in total in corresponding credible set.

index_SNP_id The name of the index SNP (SNP with highest posterior probability) in corresponding credible set.

CS_SNP_id The names of individual SNPs selected in this credible set.

Author(s)

Jiayi Shen

`output.format` *Keep the output as three digits*

Description

Keep the output as three digits

Usage

```
output.format(x, ...)
```

Arguments

`x`	input
`...`	other options you want to put in

Author(s)

Lai Jiang

`PrCa.lipids` *Real data for selecting the metabolites for the risk of prostate cancer*

Description

Real data for selecting the metabolites for the risk of prostate cancer

Format

The `PrCa.lipids` is a set of data sets which was for selecting the metabolites for the risk of prostate cancer

PrCa.lipids.marginal.Amatrix The marginal $\hat{A}$ matrix with 118 metabolites and 144 SNPs. This data is directly adapted from https://github.com/verena-zuber/demo_AMD (Zuber et al., 2020)

PrCa.lipids.Amatrix The conditional $\hat{A}$ matrix with 118 metabolites and 144 SNPs, which was composed by SuSiE JAM and the marginal $\hat{A}$ matrix.

PrCa.lipids.Geno The reference genotype data for the 144 SNPs from the European-ancestry population in 1000 Genome Project (Consortium, 2015).

PrCa.lipids.betas.gwas The b vector. The association estimates between selected SNPs and the risk of prostate cancer from (Schumacher et al., 2018)

PrCa.lipids.betas.se.gwas The se(b) vector from (Schumacher et al., 2018)

PrCa.lipids.pvalue.gwas The pvalues vector of the association estimates between selected SNPs and the risk of prostate cancer from (Schumacher et al., 2018)

PrCa.lipids.maf.gwas The vector of the effect allele frequency of the SNPs from (Schumacher et al., 2018)

PrCa.lipids.rsid The RSID of the SNPs.

References

Consortium GP. A global reference for human genetic variation. Nature 2015; 526: 68.

Zuber, Verena, et al. Selecting likely causal risk factors from high-throughput experiments using multivariable Mendelian randomization. Nature communications 11.1 (2020): 1-11.

Schumacher, Fredrick R., et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nature genetics 50.7 (2018): 928-936.

`print.ENhJAM` *Print out for EN-hJAM*

Description

Print out for EN-hJAM

Usage

```
## S3 method for class 'ENhJAM'
print(x, ...)
```

Arguments

`x`	obejct output from ENhJAM
`...`	other options you want to put in

Author(s)

Lai Jiang

`print.hJAM` *Print out for hJAM*

Description

Print out for hJAM_lnreg

Usage

```
## S3 method for class 'hJAM'
print(x, ...)
```

Arguments

x	object output by hJAM
...	other options you want to put in

Author(s)

Lai Jiang

---

print.hJAM_egger *Print out for hJAM_egger*

Description

Print out for hJAM_egger

Usage

```
## S3 method for class 'hJAM_egger'
print(x, ...)
```

Arguments

x	obejct output from hJAM_egger
...	other options you want to put in

Author(s)

Lai Jiang

---

print.SHAJAM *Print out for SHA-JAM*

Description

Print out for SHA-JAM

Usage

```
## S3 method for class 'SHAJAM'
print(x, ...)
```

Arguments

x	obejct output from SHAJAM
...	other options you want to put in

Author(s)

Lai Jiang

`SHAJAM` *SHA-JAM Fit SHA-JAM*

Description

Function to implement SHA-JAM

Usage

```
SHAJAM(
  betas.Gy,
  betas_se.Gy = NULL,
  N.Gy,
  eaf.Gy = NULL,
  Geno,
  A,
  L.cs = NULL,
  min_abs_corr = NULL,
  coverage = 0.95,
  estimate_residual_variance = TRUE,
  max_iter = 500
)
```

Arguments

`betas.Gy`	The betas in the paper: the marginal effects of SNPs on the phenotype (Gy)
`betas_se.Gy`	The standard errors of the betas
`N.Gy`	The sample size of the GWAS where you obtain the betas.Gy and betas_se.Gy
`eaf.Gy`	The reference allele frequency of the SNPs in betas.Gy
`Geno`	The individual level data of the reference panel. Must have the same order of SNPs as in the betas.Gy.
`A`	The conditional A matrix.
`L.cs`	The largest number of credible set allowed in SHA-JAM. Required by SHA-JAM.
`min_abs_corr`	The requested minimum absolute correlation coefficient between intermediates within one credible set. Required by SHA-JAM.
`coverage`	The coverage of credible set. Default is 0.95. Required by SHA-JAM.
`estimate_residual_variance`	If estimate the residual variance in the fitting procedure of SHA-JAM. Default as TRUE. Required by SHA-JAM.
`max_iter`	The number of maximum iterations in fitting SHA-JAM. Required by SHA-JAM.

Value

An object of the SHAJAM

numSNP The number of SNPs used in the analysis.

numX The number of intermediates in the analysis.

Selected_variable_length The number of selected intermediates, regardless of the credible sets.

Selected_variable_name The label/name for each selected intermediates.

Coefficients The coefficients of selected intermediates.

Selected_variable_pip The posterior inclusion probability of each selected intermediate.

num_Credible_sets Number of credible sets.

all_variables The label/name for all candidate intermediates.

all_variable_pip The posterior inclusion probability of all candidate intermediates.

all_variable_coefficient The coefficients of all candidate intermediates.

cs_purity The purity of the credibel set selected.

Author(s)

Lai Jiang

SNPs_heatmap *Heatmap for all the SNPs used in the analysis*

Description

To generate the heatmap of all the SNPs that the user use in the analysis

Usage

```
SNPs_heatmap(Geno, show.variables = FALSE, x.axis.angel = 90)
```

Arguments

`Geno`	The reference panel (Geno) of the SNPs that the user use in the analysis, such as 1000 Genome
`show.variables`	Select to show the variables name or not. Default set to be FALSE.
`x.axis.angel`	The angel for displaying the X axis. Default set to be 90.

Author(s)

Lai Jiang

Examples

```
data(MI.Rdata)
SNPs_heatmap(Geno = MI.Geno[, 1: 10], show.variable = TRUE, x.axis.angel = 90)
```

SNPs_scatter_plot *Scatter plot for SNPs vs. one intermediate in the analysis*

Description

To generate the scatter plot of the SNPs vs. one intermediate that the user use in the analysis

Usage

```
SNPs_scatter_plot(alphas, betas.Gy, X.label = NULL)
```

Arguments

alphas	The effects of SNPs on the intermediate (i.e. exposure/risk factor) (Gx).
betas.Gy	The betas in the paper: the marginal effects of SNPs on the phenotype (Gy)
X.label	The label of the intermediate (i.e. exposure/risk factor). Default is NULL.

Value

A set of scatter plots with x-axis being the conditional α estimates for each intermediate and y-axis being the β estimates.

Author(s)

Lai Jiang

Examples

```
data(MI)
SNPs_scatter_plot(alphas = MI.Amatrix[, 1], betas.Gy = MI.betas.gwas, X.label = "BMI")
```

susieJAM_A *Compute conditional A using SuSiE JAM*

Description

The susieJAM_A function is to get the conditional A matrix by using marginal A matrix

Usage

```
susieJAM_A(
  marginalA,
  marginalA_se,
  N.Gx,
  eaf.Gy = NULL,
  Geno,
  inclusion.indicator,
  L.cs,
  min_abs_corr,
  max_iter,
  coverage,
  estimate_residual_variance = TRUE
)
```

Arguments

`marginalA`	the marginal effects of SNPs on the exposures (Gx).
`marginalA_se`	the standard error of the marginal effects of SNPs on the exposures (Gx).
`N.Gx`	the sample size of each Gx. It can be a scalar or a vector. If there are multiple X's from different Gx, it should be a vector including the sample size of each Gx. If all alphas are from the same Gx, it could be a scalar.
`eaf.Gy`	the effect allele frequency of the SNPs in the Gx data.
`Geno`	the reference panel (Geno), such as 1000 Genome
`inclusion.indicator`	The matrix of inclusion indicator of SNPs for each intermediate. Included as 1; otherwise 0.
`L.cs`	A susie input parameter. Number of components (nonzero elements) in the SuSiE regression model. If L.cs is larger than the number of covariate (p), L.cs is set to p.
`min_abs_corr`	A susie input parameter. Minimum of absolute value of correlation allowed in a credible set. The default, 0.5, corresponds to squared correlation of 0.25, which is a commonly used threshold for genotype data in genetics studies.
`max_iter`	Maximum number of iterations in SuSiE fitting.
`coverage`	Default as 0.95.The coveralge level of the credible set.
`estimate_residual_variance`	Default as TRUE. Estimate the residual variance in each iteration of SuSiE fitting.

Value

A matrix with conditional estimates which are converted from marginal estimates using the susie JAM model.

Author(s)

Lai Jiang

Examples

```
data(GTEx.PrCa)
susieJAM_A(marginalA = GTEx.PrCa.marginal.A[, 1:9],
marginalA_se = GTEx.PrCa.marginal.A.se[, 1:9], eaf.Gy = GTEx.PrCa.maf.gwas,
Geno = GTEx.PrCa.Geno, inclusion.indicator = GTEx.PrCa.inclusion.indicator,
N.Gx = 620, L.cs = 10, min_abs_corr = 0.5)
```

---

susieJAM_alphas *Compute conditional alphas using SuSiE JAM*

Description

The `susieJAM_alphas` function is to perform the variable selection and compute the selected conditional alpha vector for one intermediate. If only one intermediate in the model, please use susieJAM_alphas instead of susieJAM_A

Usage

```
susieJAM_alphas(
  marginalA,
  marginalA_se,
  N.Gx,
  eaf.Gy = NULL,
  Geno,
  L.cs = 10,
  min_abs_corr = 0.6,
  max_iter = 100,
  coverage = 0.95,
  estimate_residual_variance = FALSE
)
```

Arguments

`marginalA`	the marginal effects of SNPs on one exposure (Gx).
`marginalA_se`	the standard error of the marginal effects of SNPs on one outcome (Gx).
`N.Gx`	the sample size of the Gx. It can be a scalar.
`eaf.Gy`	The vector of the minor allele frequency or effect allele frequency in the GWAS.
`Geno`	the reference panel (Geno), such as 1000 Genome. The reference data has to be centered.
`L.cs`	A susie input parameter. Number of components (nonzero elements) in the SuSiE regression model. If L.cs is larger than the number of covariate (p), L.cs is set to p.

`min_abs_corr` A susie input parameter. Minimum of absolute value of correlation allowed in a credible set. The default, 0.5, corresponds to squared correlation of 0.25, which is a commonly used threshold for genotype data in genetics studies.

`max_iter` Maximum number of iterations in SuSiE fitting.

`coverage` Default as 0.95.The coveralge level of the credible set.

`estimate_residual_variance` Default as TRUE. Estimate the residual variance in each iteration of SuSiE fitting.

Author(s)

Lai Jiang

Examples

```
data(GTEx.PrCa)
include.SNPs = which(GTEx.PrCa.inclusion.indicator[,1]==1)
susieJAM_alphas(marginalA = GTEx.PrCa.marginal.A[include.SNPs, 1],
marginalA_se = GTEx.PrCa.marginal.A.se[include.SNPs, 1], eaf.Gy = GTEx.PrCa.maf.gwas[include.SNPs],
Geno = GTEx.PrCa.Geno[, include.SNPs], N.Gx = 620, L.cs = 10, min_abs_corr = 0.5)
```

---

susie_get_posterior_mean_v2

Get SuSiE posterior mean

Description

Get SuSiE posterior mean

Usage

```
susie_get_posterior_mean_v2(res, prior_tol = 1e-09)
```

Arguments

`res` A SuSiE fit object

`prior_tol` When the prior variance is estimated, compare the estimated value to prior_tol at the end of the computation, and exclude a single effect from PIP computation if the estimated prior variance is smaller than this tolerance value.

Value

A vector of posterior mean effects

`susie_get_posterior_sd_v2` *Get SuSiE posterior sd*

Description

Get SuSiE posterior sd

Usage

```
susie_get_posterior_sd_v2(res, prior_tol = 1e-09)
```

Arguments

`res`	A SuSiE fit object
`prior_tol`	When the prior variance is estimated, compare the estimated value to prior_tol at the end of the computation, and exclude a single effect from PIP computation if the estimated prior variance is smaller than this tolerance value.

Value

A vector of posterior standard deviations

Index