Package ‘INTACT’

October 11, 2024

Type Package

Title Integrate TWAS and Colocalization Analysis for Gene Set Enrichment Analysis

Version 1.5.2

Description This package integrates colocalization probabilities from colocalization analysis with transcriptome-wide association study (TWAS) scan summary statistics to implicate genes that may be biologically relevant to a complex trait. The probabilistic framework implemented in this package constrains the TWAS scan z-score-based likelihood using a gene-level colocalization probability. Given gene set annotations, this package can estimate gene set enrichment using posterior probabilities from the TWAS-colocalization integration step.

Depends R (>= 4.3.0)

Imports SQUAREM, bdsmatrix, numDeriv, stats, tidyr, ggplot2

License GPL-3 + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

VignetteBuilder knitr

Suggests BiocStyle, knitr, rmarkdown, testthat

BugReports https://github.com/jokamoto97/INTACT/issues

URL https://github.com/jokamoto97/INTACT

biocViews Bayesian, GeneSetEnrichment

LazyData false

Config/testthat/edition 3

git_url https://git.bioconductor.org/packages/INTACT

git_branch devel

git_last_commit d0c030b

git_last_commit_date 2024-10-05

Repository Bioconductor 3.20

Date/Publication 2024-10-10

Author Jeffrey Okamoto [aut, cre] (<https://orcid.org/0000-0001-9988-1618>), Xiaoquan Wen [aut] (<https://orcid.org/0000-0001-8990-2737>)

Maintainer Jeffrey Okamoto <jokamoto@umich.edu>

Contents

.bf_em_coloc_pi0 . . . 2
.bf_loglik_coloc . . . 3
.bf_weighted_sum_coloc . . . 3
.em_est . . . 4
.enrich_bootstrap_se . . . 4
.enrich_res . . . 5
.logistic_em . . . 5
.logistic_em_nopseudo . . . 6
.logistic_loglik . . . 6
.multi_em_posteriors . . . 7
.multi_prior_estimation . . . 7
.pi1_fun . . . 8
.wakefield_bf_z_ln . . . 9
chisq_sumstat . . . 9
expit . . . 10
exprwt_sumstats . . . 11
fdr_rst . . . 11
gene_set_list . . . 12
hybrid . . . 12
intact . . . 13
intactGSE . . . 14
ld_sumstats . . . 15
linear . . . 16
multi_intact . . . 16
multi_simdat . . . 18
protwt_sumstats . . . 19
simdat . . . 19
step . . . 20
z_sumstats . . . 20

Index **22**

`.bf_em_coloc_pi0` *Multi-INTACT EM algorithm fixed-point function.*

Description

Multi-INTACT EM algorithm fixed-point function.

Usage

```
.bf_em_coloc_pi0(w, bf, fp_coloc)
```

Arguments

`w`	A vector of prior parameter estimates.
`bf`	A vector of Bayes factors.
`fp_coloc`	A vector of aggregated colocalization evidence. The default is the max of all GLCPs, truncated at 0.05.

Value

A vector of updated parameter estimates.

.bf_loglik_coloc *Multi-INTACT EM algorithm log-likelihood function.*

Description

Multi-INTACT EM algorithm log-likelihood function.

Usage

```
.bf_loglik_coloc(w, bf, fp_coloc)
```

Arguments

w	A vector of prior parameter estimates.
bf	A vector of Bayes factors.
fp_coloc	A vector of aggregated colocalization evidence. The default is the max of all GLCPs, truncated at 0.05.

Value

A scalar log likelihood.

.bf_weighted_sum_coloc *Helper function for EM algorithm to compute weighted sum of Bayes factors.*

Description

Helper function for EM algorithm to compute weighted sum of Bayes factors.

Usage

```
.bf_weighted_sum_coloc(w, bf, fp_coloc, i)
```

Arguments

w	A vector of prior parameter estimates.
bf	A vector of Bayes factors.
fp_coloc	A vector of aggregated colocalization evidence. The default is the max of all GLCPs, truncated at 0.05.
i	The index for the ith gene.

Value

A scalar weighted sum of Bayes factors

.em_est *Compute gene set enrichment estimates.*

Description

Compute gene set enrichment estimates.

Usage

```
.em_est(pprobs, d_vec)
```

Arguments

pprobs	A vector of posterior probabilities for each gene estimated from the intact function. Gene order should match d_vec.
d_vec	A vector of gene set annotations for the genes of interest. Entries should be integer(1) if the gene is annotated and integer(0) otherwise.

Value

Maximum likelihood estimates for alpha0 and alpha1; convergence indicator.

.enrich_bootstrap_se *Compute bootstrap standard errors for alpha MLEs.*

Description

Compute bootstrap standard errors for alpha MLEs.

Usage

```
.enrich_bootstrap_se(pprobs, d_vec, reps = 100)
```

Arguments

pprobs	A vector of posterior probabilities for each gene estimated from the intact function. Gene order should match d_vec.
d_vec	A vector of gene set annotations for the genes of interest. Entries should be integer(1) if the gene is annotated and integer(0) otherwise.
reps	Number of bootstrap samples.

Value

MLEs for alpha0 and alpha1 from bootstrap samples.

.enrich_res — *Compute gene set enrichment estimates with standard errors.*

Description

Compute gene set enrichment estimates with standard errors.

Usage

```
.enrich_res(sig_lev, pprobs, d_vec, SE_type = "NDS", boot_rep = NULL)
```

Arguments

sig_lev	A significance threshold for gene set enrichment hypothesis testing.
pprobs	A vector of posterior probabilities for each gene estimated from the intact function. Gene order should match d_vec.
d_vec	A vector of gene set annotations for the genes of interest. Entries should be integer(1) if the gene is annotated and integer(0) otherwise.
SE_type	A method to compute standard errors of the gene set enrichment estimates. Possible methods are "profile_likelihood," "bootstrap," and "NDS". NDS performs numerical differentiation of the Fisher score vector.
boot_rep	Number of bootstrap samples, if boostrap standard errors are specified for SE_type.

Value

A data frame with the alpha1 estimate, standard error, z-score, p-value, (1-sig_lev)% CI limits, and convergence indicator.

.logistic_em — *A fixed-point mapping for the expectation-maximization algorithm. Used as an argument for fixptfn in the squarem function.*

Description

A fixed-point mapping for the expectation-maximization algorithm. Used as an argument for fixptfn in the squarem function.

Usage

```
.logistic_em(d_vec, pprobs, alpha)
```

Arguments

d_vec	A vector of gene set annotations for the genes of interest. Entries should be integer(1) if the gene is annotated and integer(0) otherwise.
pprobs	A vector of posterior probabilities for each gene estimated from the intact function. Gene order should match d_vec.
alpha	A vector containing the current estimates of the enrichment parameters alpha0 and alpha1$.

Value

Updated estimates of alpha0 and alpha1.

.logistic_em_nopseudo *Similar to logistic_em(), but does not use pseudocounts to stablize the algorithm.*

Description

Similar to logistic_em(), but does not use pseudocounts to stablize the algorithm.

Usage

```
.logistic_em_nopseudo(d_vec, pprobs, alpha)
```

Arguments

d_vec	A vector of gene set annotations for the genes of interest. Entries should be integer(1) if the gene is annotated and integer(0) otherwise.
pprobs	A vector of posterior probabilities for each gene estimated from the intact function. Gene order should match d_vec.
alpha	A vector containing the current estimates of the enrichment parameters alpha0 and alpha1.

Value

Updated estimates of alpha0 and alpha1.

.logistic_loglik *A log likelihood function for the expectation-maximization algorithm. Used as an argument for objfn in the squarem function.*

Description

A log likelihood function for the expectation-maximization algorithm. Used as an argument for objfn in the squarem function.

Usage

```
.logistic_loglik(alpha, d_vec, pprobs)
```

Arguments

alpha	A vector containing the current estimates of the enrichment parameters alpha0 and alpha1.
d_vec	A vector of gene set annotations for the genes of interest. Entries should be integer(1) if the gene is annotated and integer(0) otherwise.
pprobs	A vector of posterior probabilities for each gene estimated from the intact function. Gene order should match d_vec.

Value

Log likelihood evaluated at the current estimates of alpha0 and alpha1.

.multi_em_posteriors *Compute gene product relevance probabilities using prior parameter estimates and Bayes factors.*

Description

Compute gene product relevance probabilities using prior parameter estimates and Bayes factors.

Usage

```
.multi_em_posteriors(w, bf, fp_coloc)
```

Arguments

w	A vector of prior parameter estimates.
bf	A vector of Bayes factors.
fp_coloc	A vector of aggregated colocalization evidence. The default is the max of all GLCPs, truncated at 0.05.

Value

A vector of posterior probabilities.

.multi_prior_estimation *Compute Multi-INTACT prior parameter estimates and gene product relevance probabilities.*

Description

Compute Multi-INTACT prior parameter estimates and gene product relevance probabilities.

Usage

```
.multi_prior_estimation(
  df,
  pi_init,
  chisq_vec,
  chisq_dof = 2,
  z_1,
  z_2,
  fp_coloc
)
```

Arguments

df	A data frame with marginal z-scores for the TWAS analyses of gene products 1 and 2, as well as the multivariate Wald chi-square statistic from the joint analysis.
pi_init	Initialization of prior parameters to be estimated. The first entry should be the qvalue estimate for pi0. The second parameter should be the gene-product-1-only parameter; the third should be the gene-product-2-only parameter, and the fourth should be the gene-product-1+2-parameter.
chisq_vec	A numeric vector of multivariate Wald chi-square test statistics. The order of this vector should match z_1, z_2, and fp_coloc.
chisq_dof	A numeric vector containing the chi-square test statistic degrees of freedom under the null. Default is 2.
z_1	A numeric vector of TWAS test statistics for gene product 1. The order of this vector should match chisq_vec, z_2, and fp_coloc.
z_2	A numeric vector of TWAS test statistics for gene product 2. The order of this vector should match chisq_vec, z_1, and fp_coloc.
fp_coloc	A vector of aggregated colocalization evidence. The default is the max of all GLCPs, truncated at 0.05. The order should match chisq_vec, z_1, and z_2.

Value

A list containing a data frame with the model posteriors (posterior_1, posterior_2, and posterior_12),gene product relevance probabilities (GPRP_1 and GPRP_2), a vector of prior parameter estimates and a Boolean indicating convergence of the EM algorithm.

.pi1_fun *Estimate pi1 from TWAS scan z-scores.*

Description

Estimate pi1 from TWAS scan z-scores.

Usage

```
.pi1_fun(z_vec, lambda = 0.5)
```

Arguments

z_vec	A vector of TWAS scan z-scores.
lambda	A value between 0 and 1. The density of TWAS scan z-scores should be flat at lambda. Set to 0.5 as default.

Value

A scalar estimate for pi1.

.wakefield_bf_z_ln — *A function to compute log Bayes factors from z-statistics using the Wakefield formula*

Description

A function to compute log Bayes factors from z-statistics using the Wakefield formula

Usage

```
.wakefield_bf_z_ln(z_vec, K = c(1, 2, 4, 8, 16))
```

Arguments

z_vec	A vector of TWAS scan z-scores for the genes of interest.
K	A vector of values over which Bayesian model averaging is performed.

Value

log Bayes factors

chisq_sumstat — *Compute a gene-level multivariate Wald chi-square statistic using summary-level genetic association and LD data.*

Description

Compute a gene-level multivariate Wald chi-square statistic using summary-level genetic association and LD data.

Usage

```
chisq_sumstat(z_vec, w, R)
```

Arguments

z_vec	A 2-vector for which the first entry is the TWAS z-score for the first gene product, and the second entry is the TWAS z-score for the second gene product (e.g. TWAS and PWAS z-scores).
w	A 2 x p matrix of TWAS prediction model weights that are used to generate the statistics required for the z_vec argument. The order of the columns must match the order of the z_vec statistics. Weights should be generated from standardized genotypes.
R	A p x p LD correlation matrix containing pairwise correlations for each SNP included in w.

Value

The value of approximated multivariate chi-square statistic.

Examples

```
data(z_sumstats)
data(protwt_sumstats)
data(exprwt_sumstats)
data(ld_sumstats)
chisq_sumstat(z_vec = z_sumstats,w = cbind(protwt_sumstats,exprwt_sumstats),
R = ld_sumstats)
```

`expit` *Transform a gene colocalization probability (GLCP) to a prior to be used in the evidence integration procedure. There are four prior function options, including expit, linear, step, and expit-linear hybrid.*

Description

Transform a gene colocalization probability (GLCP) to a prior to be used in the evidence integration procedure. There are four prior function options, including expit, linear, step, and expit-linear hybrid.

Usage

```
expit(GLCP, t = 0.05, D = 0.1, u = 1, thresholding = "hard")
```

Arguments

`GLCP`	A gene colocalization probability
`t`	A hard threshold for the GLCP. Values below this number will be shrunk to zero. Default is 0.05.
`D`	A curvature shrinkage parameter. Lower values of D will result in a steeper curve. Default is 0.1
`u`	A factor between 0 and 1 by which the prior function is scaled.
`thresholding`	An option to use hard thresholding or soft thresholding for the prior function. Default is "hard". For soft thresholding, set to "soft".

Value

The value of the prior.

Examples

```
expit(0.2, 0.05, 1)
```

exprwt_sumstats — *TWAS weights for a simulated gene.*

Description

TWAS weights 1500 SNPs in a simulated data set. This data is included to demonstrate how to compute a multivariate Wald statistic from summary-level data.

Usage

```
exprwt_sumstats
```

Format

A vector of length 1500.

Value

A vector of length 1500.

Examples

```
data(exprwt_sumstats)
```

fdr_rst — *Bayesian FDR control for INTACT output*

Description

Bayesian FDR control for INTACT output

Usage

```
fdr_rst(posterior, alpha = 0.05)
```

Arguments

posterior	A vector of posterior probabilities for each gene estimated from the intact function.
alpha	A numeric target FDR control level.

Value

An n x 2 data frame where the first column is the inputted posterior probabilities, and the second is a Boolean vector denoting significance at the specified target control level.

Examples

```
data(simdat)
fdr_rst(simdat$GLCP)
```

gene_set_list — *Simulated gene set list.*

Description

A list object containing two elements. Each is a character list of gene names.

Usage

```
gene_set_list
```

Format

A list with two items:

gene set 1 gene set with 503 gene members. Significantly enriched in simdat.

gene set 2 gene set with 200 members. ...

Value

A list with two items:

Examples

```
data(gene_set_list)
```

hybrid — *Transform a gene colocalization probability (GLCP) to a prior to be used in the evidence integration procedure. There are four prior function options, including expit, linear, step, and expit-linear hybrid.*

Description

Transform a gene colocalization probability (GLCP) to a prior to be used in the evidence integration procedure. There are four prior function options, including expit, linear, step, and expit-linear hybrid.

Usage

```
hybrid(GLCP, t = 0.05, D = 0.1, u = 1, thresholding = "hard")
```

Arguments

GLCP	A gene colocalization probability
t	A hard threshold for the GLCP. Values below this number will be shrunk to zero. Default is 0.05.
D	A curvature shrinkage parameter. Lower values of D will result in a steeper curve. Default is 0.1
u	A factor between 0 and 1 by which the prior function is scaled.
thresholding	An option to use hard thresholding or soft thresholding for the prior function. Default is "hard". For soft thresholding, set to "soft".

Value

The value of the prior.

Examples

```
hybrid(0.2, 0.05, 1)
```

---

intact — *Compute the posterior probability that a gene may be causal, given a gene's TWAS scan z-score (or Bayes factor) and colocalization probability.*

Description

Compute the posterior probability that a gene may be causal, given a gene's TWAS scan z-score (or Bayes factor) and colocalization probability.

Usage

```
intact(
  GLCP_vec,
  prior_fun = linear,
  z_vec = NULL,
  t = NULL,
  D = NULL,
  K = c(1, 2, 4, 8, 16),
  twas_priors = .pi1_fun(z_vec = z_vec, lambda = 0.5),
  twas_BFs = NULL
)
```

Arguments

GLCP_vec	A vector of colocalization probabilities for the genes of interest
prior_fun	A function to transform a colocalization probability into a prior. Options are linear, step, expit, and hybrid.
z_vec	A vector of TWAS scan z-scores for the genes of interest. The order of genes must match GLCP_vec.
t	A hard threshold for the GLCP. Values below this number will be shrunk to zero. This argument is used in the user-specified prior function. Default value for the step prior is 0.5. Default value is 0.05 for all other prior functions.
D	A curvature shrinkage parameter. Lower values of D will result in a steeper curve. Default is 0.1. This parameter should only be specified if the user selects the expit or hybrid prior function and does not wish to use the default value.
K	A vector of values over which Bayesian model averaging is performed.
twas_priors	An optional vector of user-specified gene-specific TWAS priors. If no input is supplied, INTACT computes a scalar prior using the TWAS data (see the corresponding manuscript for more details).
twas_BFs	A vector of TWAS Bayes factors for the genes of interest. This is an alternative option if the user wishes to directly specify Bayes factors instead of computing them from TWAS scan z-scores.

Value

The vector of posteriors.

Examples

```
data(simdat)
intact(GLCP_vec=simdat$GLCP, z_vec = simdat$TWAS_z)
intact(GLCP_vec=simdat$GLCP, prior_fun=expit, z_vec = simdat$TWAS_z,
t = 0.02,D = 0.09)
intact(GLCP_vec=simdat$GLCP, prior_fun=step, z_vec = simdat$TWAS_z,
t = 0.49)
intact(GLCP_vec=simdat$GLCP, prior_fun=hybrid, z_vec = simdat$TWAS_z,
t = 0.49,D = 0.05)
```

---

intactGSE — *Perform gene set enrichment estimation and inference, given TWAS scan z-scores and colocalization probabilities.*

Description

Perform gene set enrichment estimation and inference, given TWAS scan z-scores and colocalization probabilities.

Usage

```
intactGSE(
  gene_data,
  prior_fun = linear,
  t = NULL,
  D = NULL,
  gene_sets,
  sig_lev = 0.05,
  SE_type = "NDS",
  boot_rep = NULL
)
```

Arguments

gene_data	A data frame containing gene names and corresponding colocalization probabilities and TWAS z-scores for each gene. Column names should be "gene", "GLCP", and "TWAS_z'. If the user wishes to specify TWAS Bayes factors instead of z-scores, use the column name "TWAS_BFs". If the user wishes to specify gene-specific TWAS priors, use the column name "TWAS_priors".
prior_fun	A function to transform a colocalization probability into a prior. Options are linear, step, expit, and hybrid.
t	A hard threshold for the GLCP. Values below this number will be shrunk to zero. This argument is used in the user-specified prior function. Default value for the step prior is 0.5. Default value is 0.05 for all other prior functions.
D	A curvature shrinkage parameter. Lower values of D will result in a steeper curve. Default is 0.1. This parameter should only be specified if the user selects the expit or hybrid prior function and does not wish to use the default value.

gene_sets	A named list of gene sets for which enrichment is to be estimated. List items should be character vectors of gene IDs. Gene ID format should match the gene column in gene_data.
sig_lev	A significance threshold for gene set enrichment hypothesis testing.
SE_type	A method to compute standard errors of the gene set enrichment estimates. Possible methods are "profile_likelihood" and "bootstrap."
boot_rep	Number of bootstrap samples.

Value

A data frame with the alpha1 estimate, standard error, z-score, p-value, (1-sig_lev)% CI limits, and convergence indicator for each gene set in gene_sets.

Examples

```
data(simdat)
data(gene_set_list)
intactGSE(gene_data = simdat,gene_sets = gene_set_list)
intactGSE(gene_data = simdat,prior_fun = step,t = 0.45,
gene_sets = gene_set_list)
intactGSE(gene_data = simdat,prior_fun = expit,t = 0.08,D = 0.08,
gene_sets = gene_set_list)
intactGSE(gene_data = simdat,prior_fun = hybrid,t = 0.08,D = 0.08,
gene_sets = gene_set_list)
```

---

ld_sumstats *LD correlation matrix from a simulated data set.*

Description

A data set containing correlation data for 1500 cis-SNPs for a simulated target gene. This data is included to demonstrate how to compute a multivariate Wald statistic from summary-level data.

Usage

```
ld_sumstats
```

Format

A matrix with 1500 rows and 1500 columns.

Value

A matrix with 1500 rows and 1500 columns.

Examples

```
data(ld_sumstats)
```

`linear` *Transform a gene colocalization probability (GLCP) to a prior to be used in the evidence integration procedure. There are four prior function options, including expit, linear, step, and expit-linear hybrid.*

Description

Transform a gene colocalization probability (GLCP) to a prior to be used in the evidence integration procedure. There are four prior function options, including expit, linear, step, and expit-linear hybrid.

Usage

```
linear(GLCP, t = 0.05, u = 1, thresholding = "hard")
```

Arguments

`GLCP`	A gene colocalization probability
`t`	A hard threshold for the GLCP. Values below this number will be shrunk to zero. Default is 0.05.
`u`	A factor between 0 and 1 by which the prior function is scaled.
`thresholding`	An option to use hard thresholding or soft thresholding for the prior function. Default is "hard". For soft thresholding, set to "soft".

Value

The value of the prior.

Examples

```
linear(0.2, 0.05, 1)
linear(c(0.01,0.2,0.9))
```

`multi_intact` *Compute Multi-INTACT prior parameter estimates and gene product relevance probabilities.*

Description

Compute Multi-INTACT prior parameter estimates and gene product relevance probabilities.

Usage

```
multi_intact(
  df,
  chisq_dof = 2,
  prior_fun = linear,
  t = 0.05,
  D = NULL,
 xwas_priors = .pi1_fun(z_vec = qnorm(pchisq(df$chisq, df = chisq_dof, lower.tail =
    FALSE)/2)),
  xwas_BFs = NULL,
  bf_type = "wakefield",
  K = c(1, 2, 4, 8, 16),
  glcp_aggreg = "max",
  em_algorithm = TRUE,
  pi0 = 1 - .pi1_fun(z_vec = qnorm(pchisq(df$chisq, df = chisq_dof, lower.tail =
    FALSE)/2)),
  pi_init = c(pi0, rep(1 - pi0, 3)/3),
  return_model_posteriors = FALSE
)
```

Arguments

df	A data frame containing at least the following six columns: (1)'gene' contains the gene ID, (2)'GLCP_1' contains the numeric pairwise colocalization probability for the complex trait and gene product 1, (3)'GLCP_2' contains the numeric pairwise colocalization probability for the complex trait and gene product 2, (4) 'z_1' A numeric vector of TWAS test statistics for gene product 1, (5) 'z_2' A numeric vector of TWAS test statistics for gene product 2, and (6)'chisq' A numeric vector of multivariate Wald chi-square test statistics.
chisq_dof	A numeric scalar or vector with the degrees of freedom of the chi-square test statistics under the null. Default is 2 for all genes.
prior_fun	A function to transform an aggregated colocalization probability into a prior. Options are linear, step, expit, and hybrid.
t	A hard threshold for the aggregated colocalization probability. Values below this number will be shrunk to zero. This argument is used in the user-specified prior function. Default value for the step prior is 0.5. Default value is 0.05 for all other prior functions.
D	A curvature shrinkage parameter. Lower values of D will result in a steeper curve. Default is 0.1. This parameter should only be specified if the user selects the expit or hybrid prior function and does not wish to use the default value.
xwas_priors	An optional vector of user-specified gene-specific priors for the joint regression of the complex trait on the two gene product levels. If no input is supplied, Multi-INTACT computes a scalar prior using the multivariate Wald test data (see the corresponding manuscript for more details).
xwas_BFs	A vector of Bayes factors for the joint regression of the complex trait on the two gene product levels for the gene of interest. This is an alternative option if the user wishes to directly specify Bayes factors instead of computing them from the chi-square statistics.
bf_type	Method used to compute Bayes factors. Default is 'wakefield', but a formula from Johnson (2005) 'johnson' is also available.

K	A vector of values over which Bayesian model averaging is performed.
glcp_aggreg	Function used to aggregate pairwise coloclization probabilities. Default is taking the max ('max'), but 1-prod(1-GLCP) 'one_minus_prod_one_minus' is also available.
em_algorithm	Should the function run the EM algorithm to compute gene product relevance probabilities (TRUE/FALSE)?
pi0	Estimate for the parameter pi0.
pi_init	Initialization of prior parameters to be estimated. The first entry should be the qvalue estimate for pi0. The second parameter should be the gene-product-1-only parameter; the third should be the gene-product-2-only parameter, and the fourth should be the gene-product-1+2-parameter. Entries must sum to 1.
return_model_posteriors	Should the function return model posteriors in addition to the gene product relevance probabilities (TRUE/FALSE)?

Value

A list containing: (1) a data frame with gene probabilities of putative causality (GPPC), model posteriors (posterior_1, posterior_2, and posterior_12) and gene product relevance probabilities (GPRP_1 and GPRP_2); (2) EM algorithm prior parameter estimates; (3) a Boolean indicating convergence of the EM algorithm.

Examples

```
data(multi_simdat)
multi_intact(df = multi_simdat)
```

---

multi_simdat *Simulated TWAS, PWAS, and pairwise colocalization summary data.*

Description

A data set containing pairwise fastENLOC GLCPs, TWAS and PWAS z-scores, and multivariate Wald chi-square statistics for 1197 simulated genes.

Usage

```
multi_simdat
```

Format

A data frame with 1197 rows and 6 variables:

gene gene Ensembl ID

GLCP_1 Pairwise colocalization probability for the complex trait and gene expression levels.

GLCP_2 Pairwise colocalization probability for the complex trait and encoded protein levels.

z_1 TWAS z-score.

z_2 PWAS z-score.

chisq Multivariate Wald chi-square statistic from the regression of the complex trait on the predicted expression and protein levels of the target gene. ...

Value

A data frame with 1197 rows and 6 variables:

Examples

```
data(multi_simdat)
```

---

protwt_sumstats *PWAS weights for a simulated gene.*

Description

PWAS weights 1500 SNPs in a simulated data set. This data is included to demonstrate how to compute a multivariate Wald statistic from summary-level data.

Usage

```
protwt_sumstats
```

Format

A vector of length 1500.

Value

A vector of length 1500.

Examples

```
data(protwt_sumstats)
```

---

simdat *Simulated TWAS and colocalization summary data.*

Description

A data set containing GLCP and TWAS z-score for 1197 simulated genes.

Usage

```
simdat
```

Format

A data frame with 1197 rows and 3 variables:

gene gene Ensembl ID

GLCP colocalization probability

TWAS_z TWAS z-score ...

Value

A data frame with 1197 rows and 3 variables

Examples

```
data(simdat)
```

---

step — *Transform a gene colocalization probability (GLCP) to a prior to be used in the evidence integration procedure. There are four prior function options, including expit, linear, step, and expit-linear hybrid.*

Description

Transform a gene colocalization probability (GLCP) to a prior to be used in the evidence integration procedure. There are four prior function options, including expit, linear, step, and expit-linear hybrid.

Usage

```
step(GLCP, t = 0.5, u = 1)
```

Arguments

GLCP	A gene colocalization probability
t	A hard threshold for the GLCP. Values below this number will be shrunk to zero. Default is 0.5.
u	A factor between 0 and 1 by which the prior function is scaled.

Value

The value of the prior.

Examples

```
step(0.2, 0.05, 1)
```

---

z_sumstats — *TWAS and PWAS z-score for a simulated gene.*

Description

A 2-vector for which the first entry is the TWAS z-score, and the second entry is the PWAS z-score for a simulated gene. This data is included to demonstrate how to compute a multivariate Wald statistic from summary-level data.

Usage

```
z_sumstats
```

Format

A 2-vector.

Value

A 2-vector.

Examples

```
data(z_sumstats)
```

Index

∗ **datasets**
    exprwt_sumstats, 11
    gene_set_list, 12
    ld_sumstats, 15
    multi_simdat, 18
    protwt_sumstats, 19
    simdat, 19
    z_sumstats, 20
.bf_em_coloc_pi0, 2
.bf_loglik_coloc, 3
.bf_weighted_sum_coloc, 3
.em_est, 4
.enrich_bootstrap_se, 4
.enrich_res, 5
.logistic_em, 5
.logistic_em_nopseudo, 6
.logistic_loglik, 6
.multi_em_posteriors, 7
.multi_prior_estimation, 7
.pi1_fun, 8
.wakefield_bf_z_ln, 9

chisq_sumstat, 9

expit, 10
exprwt_sumstats, 11

fdr_rst, 11

gene_set_list, 12

hybrid, 12

intact, 13
intactGSE, 14

ld_sumstats, 15
linear, 16

multi_intact, 16
multi_simdat, 18

protwt_sumstats, 19

simdat, 19
step, 20

z_sumstats, 20