Package: prop.coloc (via r-universe)

January 16, 2025

Title A frequentist test of proportional colocalization after selecting relevant genetic variants

Version 1.1.0

Description A proportional colocalization test that accounts for uncertainty in variant selection using summary data.

License use_mit_license()

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Imports ggplot2 (>= 3.3.0)

Depends R (>= 2.10)

LazyData true

Suggests knitr, rmarkdown

VignetteBuilder knitr

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

RemoteUrl https://github.com/ash-res/prop-coloc

RemoteRef HEAD

RemoteSha cd00a469fc1149709db8960e0d722d9df6eb9e03

Contents

GLP1R *GLP1R summary data included in the package*

Description

GLP1R summary data included in the package

References

https://gtexportal.org/home/tissueSummaryPage

prop.coloc *A frequentist test of proportional colocalization after selecting relevant genetic variants*

Description

A proportional colocalization test that accounts for uncertainty in variant selection using summary data.

Usage

```
prop.coloc(
  b1,
  se1,
  b2,
  se2,
  n,
  ld,
  tau = NULL,
  alpha = NULL,
  prune = NULL,
  J = NULL,
  figs = NULL,
  traits = NULL,
  seed = 100
)
```

Arguments

b1	beta coefficients for trait 1
se1	standard errors for trait 1
b2	beta coefficients for trait 2
se2	standard errors for trait 2

n	sample size; a one-sample analysis is performed if only one sample size is entered. If two sample sizes are entered, then a two-sample analysis is performed under the assumption that genetic associations with traits 1 and 2 are measured from non-overlapping samples
ld	genetic variant correlation matrix
tau	(optional) correlation between trait 1 and trait 2; if traits are measured in separate samples, then this should be set to 0 (default value is 0)
alpha	(optional) nominal size of the conditional proportional colocalization test; if unspecified, p-value is reported to 2 decimal places
prune	(optional) R^2 pruning threshold for variant correlations; if unspecified (default value is R^2 = 0.6)
J	(optional) the top J variants for at least one trait are used to fit multivariable trait-variant linear models (default value is J=10)
figs	(optional) logical: if TRUE, return plots of the univariable and fitted multivariable variant–trait associations of top J variants for each trait (default value is FALSE)
traits	(optional) a character vector of the two trait names
seed	(optional) the seed for random sampling involved in computing conditional critical values for the prop-coloc-cond p-value (default value is 100)

Details

The method computes two different proportional colocalization tests. Each test is a frequentist hypothesis test where the null hypothesis is colocalization (i.e. the genetic associations with traits 1 and 2 are proportional), and the alternative hypothesis is failure to colocalize (the genetic associations are not proportional).

First, prop-coloc-cond is a conditional test based on lead variants for each trait that accounts for uncertainty in variant selection. By contrast, prop-coloc-full tests the proportional colocalization hypothesis using a larger set of variant associations (by default, the top 10 strongest variant associations for each trait).

Compared with prop-coloc-full, the prop-coloc-cond test is shown to achieve good type I error control, and be less sensitive to measurement errors in estimated genetic correlations. However, prop-coloc-cond and prop-coloc-full test slightly different null hypotheses; whereas prop-coloc-cond focuses on the evidence from lead variants, prop-coloc-full considers the evidence across more variants.

We note that an optional input of "tau", which is the correlation between the two traits, is typically not provided in usual genetic association summary data. Therefore, sensitivity of analyses to the choice of tau is recommended.

Both proportional colocalization tests require specification of the variant correlation matrix. This is a signed matrix, and the correlations must be provided with respect to the same effect alleles as the beta coefficients for each trait.

Value

Output is a list containing:

- `p_cond`
  p-value of prop-coloc-cond test to two decimal places. Null hypothesis is that the beta-coefficients are proportional (this represents proportional colocalization), rejection indicates failure to proportionately colocalize.
  If `alpha` is specified, then `p_cond` is logical `TRUE` if the proportional colocalization null hypothesis is rejected or `FALSE` if the proportional colocalization null hypothesis is not rejected at the specified `alpha` significance threshold.
- `eta_cond`
  proportionality constant based on lead variants for each trait
- `LM_cond`
  p-value of Lagrange Multiplier test based on lead variants for each trait. Null hypothesis is that the proportionality constant is zero, rejection indicates proportionality constant is non-zero. Colocalization is only meaningful if the proportionality constant is non-zero.
- `p_full`
  p-value of prop-coloc-full test based on top J variants for each trait. Null hypothesis is that the beta-coefficients are proportional (this represents proportional colocalization), rejection indicates failure to colocalize.
- `eta_full`
  proportionality constant based on top J variants for each trait
- `LM_full`
  p-value of Lagrange Multiplier test based on top J variants for each trait. Null hypothesis is that the proportionality constant is zero, rejection indicates proportionality constant is non-zero. Colocalization is only meaningful if the proportionality constant is non-zero.
- `Q`
  naive test statistic for proportional colocalization hypothesis from prop-coloc-cond
- `alpha`
  nominal size of the conditional proportional colocalization test (if specified)
- `convergence`
  whether uniroot function converged to find a conditional critical value for the prop-coloc-cond test
- `top`
  two most relevant variants selected for the prop-coloc-cond test
- `variants`
  set of variants used for the prop-coloc-full test
- `fig_uni`
  plot of the univariable variant–trait associations of top J variants for each trait (if `figs` is specified and `TRUE`)
- `fig_multi`
  plot of the fitted multivariable variant–trait associations of top J variants for each trait (if `figs` is specified and `TRUE`)

Author(s)

Stephen Burgess and Ashish Patel

References

A frequentist test of proportional colocalization after selecting relevant genetic variants. Preprint. https://arxiv.org/abs/2402.12171

Examples

```
res1 <- prop.coloc(b1=GLP1R$atrial_appendage$beta, se1=GLP1R$atrial_appendage$se, b2=GLP1R$pancreas$beta, se2=GL
# here the LM test is rejected for both methods (p<0.01), so there is a non-zero slope in the scatterplot of genetic a
# the proportionality test is rejected in the prop-coloc-full method (p<0.01), but not in the prop-coloc-cond method
# hence there is evidence for failure to colocalize
# note that when alpha is not specified, the code takes much longer to run
res2 <- prop.coloc(b1=GLP1R$thyroid$beta, se1=GLP1R$thyroid$se, b2=GLP1R$lung$beta, se2=GLP1R$lung$se, ld=GLP1R$
# here the LM test is not rejected, so there is no clear evidence for a non-zero slope in the scatterplot of genetic a
# additionally, the proportionality test is rejected in both methods
# hence there is evidence for failure to colocalize
res3 <- prop.coloc(b1=GLP1R$atrial_appendage$beta, se1=GLP1R$atrial_appendage$se, b2=GLP1R$left_ventricle$beta,
# here the LM test is rejected for both methods (p<0.01), so there is a non-zero slope in the scatterplot of genetic a
# the proportionality test is not rejected by either method: p>0.05 for prop-coloc-cond and p=0.72 for prop-coloc-fu
# hence there is no evidence against colocalization
res3$fig_uni   # scatterplot of univariable associations (beta coefficients)
res3$fig_multi # scatterplot of multivariable associations
```

Index