Package: colocPropTest (via r-universe)

September 10, 2024

Title Proportional Testing for Colocalisation Analysis

Version 0.9.1

Description Colocalisation analysis tests whether two traits share a causal genetic variant in a specified genomic region. Proportional testing for colocalisation has been previously proposed [Wallace (2013) <doi:10.1002/gepi.21765>], but is reimplemented here to overcome barriers to its adoption. Its use is complementary to the fine- mapping based colocalisation method in the 'coloc' package, and may be used in particular to identify false ``H3" conclusions in 'coloc'.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.2.3

VignetteBuilder knitr

Depends magrittr, data.table

Imports car, coloc, graphics, stats, utils

Suggests knitr, plotrix, testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Chris Wallace [aut, cre] (<https://orcid.org/0000-0001-9755-1703>)

Maintainer Chris Wallace <chris.x.wallace@gsk.com>

Repository CRAN

Date/Publication 2024-06-11 11:30:02 UTC

Contents

`adjust_LD` *adjust LD for variable sample size*

Description

adjust LD for variable sample size

Usage

```
adjust_LD(S, LD)
```

Arguments

`S`	coloc style dataset, with additional entries n0 and n1 which are *vectors* giving the number of cases and controls genotyped at each SNP
`LD`	matrix of LD, with dimnames given by snps in S$snp

Value

adjusted LD matrix

Examples

```
library(coloc)

data(coloc_test_data)
attach(coloc_test_data)
LD=D1$LD
dimnames(LD)=list(D1$snp,D1$snp)
D1$type="cc"
D1$s=.5
D1$n1=D1$N * sample(c(0.25,.5),length(D1$snp), replace=TRUE)
D1$n0=rep(0.5*D1$N,length(D1$snp))
aLD=colocPropTest::adjust_LD(D1,LD)
LD[1:6,1:6]
aLD[1:6,1:6]
detach(coloc_test_data)
```

estprop — *Proportional colocalisation testing supplying only a pair of regression coefficients.*

Description

Proportional colocalisation testing supplying only a pair of regression coefficients.

Usage

```
estprop(b1, b2, V1, V2)
```

Arguments

b1	regression coefficients for trait 1, expect length(b1)=2
b2	regression coefficients for trait 2, expect length(b2)=2
V1	2x2 variance-covariance matrix for trait 1
V2	2x2 variance-covariance matrix for trait 2

Value

a list, containing * result: the test statistic * plot.data: dataset for plotting the input data * plot.eta: dataset for plotting chisq as a function of theta or eta

Author(s)

Chris Wallace

estprop_slow — *Proportional colocalisation testing*

Description

This should return the same as estprop for a pair of snps, but is slower. Left here for checking. Also accomodates more than two snps.

Usage

```
estprop_slow(b1, b2, V1, V2)
```

Arguments

b1	regression coefficients for trait 1
b2	regression coefficients for trait 2
V1	variance-covariance matrix for trait 1
V2	variance-covariance matrix for trait 2

Value

a list, containing the test statistic and two datasets for plotting the input data or eta

Author(s)

Chris Wallace

---

keep_from_S *keep snp subset of coloc dataset*

Description

keep snp subset of coloc dataset

Usage

```
keep_from_S(S, keep)
```

Arguments

S	coloc dataset
keep	snps to keep

Value

subset of coloc dataset

---

lp *logp*

Description

uses logs in calculation to avoid numerical issues with very small std errors / p values

Usage

```
lp(beta, se)
```

Arguments

beta	coefficient
se	std error of coefficient

Value

-log10 p

marg_with_V — *create variance-covariance matrix for pair of marginal beta + vbeta, given estimate of r between snps*

Description

create variance-covariance matrix for pair of marginal beta + vbeta, given estimate of r between snps

Usage

```
marg_with_V(beta, vbeta, rho)
```

Arguments

beta	vector of two coefficients at two snps
vbeta	vector of two variance of coefficients at the same two snps
rho	LD (r) between the two snps

Value

list of coefficient & variance-covariance matrix

nform — *Helper function to adjust LD parameter r for differential sample size between snps*

Description

Estimate the r between effect estimates at snps which were genotyped on different sets of cases and controls. The adjusted r will be nform(...) * r (where r is the population correlation between snps).

Usage

```
nform(n0a, n1a, n0b, n1b, n0ab = pmin(n0a, n0b), n1ab = pmin(n1a, n1b))
```

Arguments

n0a	number of controls with data at snp a
n1a	number of cases with data at snp a
n0b	number of controls with data at snp b
n1b	number of cases with data at snp b
n0ab	number of controls with data at both snps a and b
n1ab	number of cases with data at both snps a and b

Value

proportionality constant that depends on sample size.

`plot_ellipses` *draw two ellipses*

Description

draw two ellipses

Usage

```
plot_ellipses(
  b1,
  vb1,
  b2,
  vb2,
  legend = c("inferred", "observed"),
  include_origin = FALSE,
  ...
)
```

Arguments

`b1`	ellipse 1 centre (2d)
`vb1`	ellipse 1 vcov matrix
`b2`	ellipse 2 centre (2d)
`vb2`	ellipse 2 vcov matrix
`legend`	character vector length 2 naming ellipse 1 and 2
`include_origin`	if TRUE, ensure plot includes (0,0)
`...`	arguments passed to plot()

Value

draw ellipses on current graphics device

Author(s)

Chris Wallace

Examples

```
plot_ellipses(b1=c(5,5), vb1=diag(2),
              b2=c(2,2), vb2=matrix( c(1,0.5,0.5,1), 2, 2 ),
              legend=c("circle", "ellipse"),
              include.origin=TRUE)
```

`run_proptests` *run proportional tests on extreme subset of snp pairs from two coloc style datasets. Of all functions in this package, this is the main one that should be used.*

Description

run proportional tests on extreme subset of snp pairs from two coloc style datasets. Of all functions in this package, this is the main one that should be used.

Usage

```
run_proptests(
  S1,
  S2,
  LD,
  topsnps = "auto",
  r2.thr = 0.95,
  maxtests = 10000,
  nauto = 200,
  adjust_n = FALSE
)
```

Arguments

`S1`	coloc dataset 1
`S2`	coloc dataset 2
`LD`	LD matrix - rownames, colnames capture the snps and S1$snp[j] must be represented
`topsnps`	list of topsnps to be considered for testing or, if "auto", will be automatically selected
`r2.thr`	r2 threshold for initial tagging step - includes only one of any set of snps in mutually high LD with r2 > r2.thr
`maxtests`	maximum number of test pairs to consider. if more than maxtests pairs available, will select a random sample.
`nauto`	number of snps to use when automatically defining topsnps. only has an effect if topsnps=="auto"
`adjust_n`	TRUE if you want to adjust for variable sample size between snps. This is only set up for case control data at the moment (ask if you need quantitative) and requires that you supply separately the number of cases and controls at each snp in each dataset, as vector elements of the lists called n1 (cases) and n0 (controls)

Value

data.table containing the tests run

Author(s)

Chris Wallace

Examples

```
library(colocPropTest)
library(coloc)
data(coloc_test_data)
attach(coloc_test_data)
LD=D1$LD
dimnames(LD)=list(D1$snp,D1$snp)
results=run_proptests(D1,D2,LD=LD,topsnps=D1$snp,maxtests=100)
min(results$fdr)
```

`tag` *Derive tag SNPs using heirarchical clustering*

Description

Uses complete linkage and the `hclust` function to define clusters, then cuts the tree at 1-tag.threshold

Usage

```
tag(r2, r2_threshold = 0.95, quiet = FALSE, method = "complete")
```

Arguments

`r2`	matrix of rsquared values
`r2_threshold`	only 1 of a set of snps with r2 > r2_threshold will be kept
`quiet`	if FALSE (default), show progress messages
`method`	method used for heirarchical clustering. See hclust for options.

Value

character vector, names are `snps`, values are the tag for each SNP

Author(s)

Chris Wallace

`tester_marg` *run proportional test directly on marginal test stats from coloc datasets*

Description

run proportional test directly on marginal test stats from coloc datasets

Usage

```
tester_marg(j, S1, S2, LD1, LD2 = LD1)
```

Arguments

`j`	indices of thw two snps
`S1`	coloc dataset 1
`S2`	coloc dataset 2
`LD1`	LD matrix for dataset 1 - rownames, colnames capture the snps and S1$snp[j] must be represented
`LD2`	LD matrix for dataset 2 - rownames, colnames capture the snps and S2$snp[j] must be represented. if not supplied, defaults to LD1

Value

result from estprop

Index