Package 'subgxe'

October 14, 2022

Title Combine Multiple GWAS by Using Gene-Environment Interactions

Version 0.9.0

Description Classical methods for combining summary data from genome-wide association studies (GWAS) only use marginal genetic effects and power can be compromised in the presence of heterogeneity. 'subgxe' is a R package that implements p-value assisted subset testing for association (pASTA), a method developed by Yu et al. (2019) <doi:10.1159/000496867>. pASTA generalizes association analysis based on subsets by incorporating gene-environment interactions into the testing procedure.

License GPL-3

URL https://github.com/umich-cphds/subgxe

BugReports https://github.com/umich-cphds/subgxe/issues

Suggests lmtest, knitr, rmarkdown

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

VignetteBuilder knitr

NeedsCompilation no

Author Youfei Yu [aut],
Alexander Rix [cre]

Maintainer Alexander Rix <alexrix@umich.edu>

Repository CRAN

Date/Publication 2019-06-14 13:30:03 UTC

R topics documented:

pasta *pasta for multi-phenotype analysis*

Description

Search for the subset that yields the strongest evidence of association and calculate the meta-analytic p-value, possibly in the presence of gene-environmental interaction.

Usage

```
pasta(p.values, study.sizes, cor)
```

Arguments

p.values	The p.value of each study.
study.sizes	The sample size of each study.
cor	The correlation matrix of the studies. For example, if each study is independent, cor would be the identity matrix.

Value

A list containing the joint p value and the test statistic, which contains the optimal subset.

References

Yu Y, Xia L, Lee S, Zhou X, Stringham H, M, Boehnke M, Mukherjee B: Subset-Based Analysis Using Gene-Environment Interactions for Discovery of Genetic Associations across Multiple Studies or Phenotypes. Hum Hered 2019. doi: 10.1159/000496867

Examples

```
# grab synthetic study for example
data("studies")
n.studies <- 5
study.sizes <- c(nrow(studies[[1]]), nrow(studies[[2]]), nrow(studies[[3]]),
                 nrow(studies[[4]]), nrow(studies[[5]]))
study.pvals <- rep(0, n.studies)
# Correlations of p-values among the studies.
# In this case the studies were generated independently so its just I
cor.matrix <- diag(1, n.studies)
# load the lrtest() function to conduct the likelihood ratio test
# Used just to generate the input p-values, not required in pasta itself.

library(lmtest)

for(i in 1:n.studies) {
 # model with gene(G) by environment(E) interaction
 model <- glm(D ~ G + E + GbyE, data = studies[[i]], family = binomial)
```

```
 # model without G and GE interaction
 null.model <- glm(D ~ E, data = studies[[i]], family = binomial)
 # likelihood ratio test from the package lmtest
 study.pvals[i] = lmtest::lrtest(null.model, model)[2, 5]
}

pasta <- pasta(study.pvals, study.sizes, cor.matrix)

pasta$p.pasta
pasta$test.statistic$selected.subset
```

`studies` *Synthetic data for subgxe*

Description

Synthetic data for subgxe

Usage

```
studies
```

Format

A list of 5 `data.frames` with 12000 observations (6000 cases, 6000 controls) on 4 variables:

D Disease status. Numeric 0-1

G Genetic variant. Numeric 0-1

E Exposure. Numeric 0-1

GbyE `G * E`. Either 1 or 0.

Index