Package 'mtPRS'

January 13, 2023

Type Package

Title Construct multi-trait PRS (mtPRS) via performing principal component analysis (PCA) on genetic correlaton matrix among traits

Version 0.1.0

Maintainer Song Zhai <zsviolet1993@gmail.com>

Description Construct multi-trait PRS (mtPRS) via performing principal component analysis (PCA) on genetic correlaton matrix among traits. Package is based on "Integrating multiple traits for improving polygenic risk prediction in disease and pharmacogenomics GWAS" by Zhai, S., Guo, B., Wu, B., Mehrotra, D.V., and Shen, J., 2023 (submitted).

License GPL (>= 2)

Depends R (>= 4.0.0)

Imports dplyr (>= 1.0.2), mvtnorm (>= 1.1.1), data.table (>= 1.13.2), Matrix (>= 1.2.18), glmnet (>= 4.0.2), stats (>= 4.0.3), ACAT (>= 0.91)

Suggests knitr, rmarkdown

VignetteBuilder knitr

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

NeedsCompilation yes

Author Song Zhai [aut, cre]

Repository CRAN

R topics documented:

`generate_dis_data` *Simulate disease GWAS data for mtPRS analysis*

Description

Simulate disease GWAS data for mtPRS analysis, including both disease GWAS summary statistics in base cohort and individual-level data in target cohort

Usage

```
generate_dis_data(
  structure = "CS",
  sparseness = "no",
  rho_DT,
  rho_T,
  rho_C,
  K = 4,
  m = 10000,
  pcausal = 0.01,
  blocksize = 100,
  samplesize = 700,
  h2_base = 0.5,
  h2_target = 0.5
)
```

Arguments

`structure`	genetic correlation structure, either uniformly correlated (i.e., "CS" or "AR1") or clustered correlated ("clustered")
`sparseness`	effect sparseness, either "no", or "half", or "more"
`rho_DT`	effect correlations between traits in base cohort and the phenotype in target cohort
`rho_T`	genetic correlation among traits in base cohort
`rho_C`	between-cluster correlation when the genetic correlation structure is clustered correlated
`K`	number of traits
`m`	number of SNPs
`pcausal`	proportion of causal SNPs
`blocksize`	size of each LD block (i.e., the number of SNPs in each LD block)
`samplesize`	sample size of target cohort
`h2_base`	heritability in base cohort
`h2_target`	heritability in target cohort

Details

Simulate disease GWAS data for mtPRS analysis into 12 (2x3x2) different genetic architectures

Value

a list of disease GWAS summary statistics in base cohort, individual-level data in target cohort, genetic correlation matrix, and true effect sizes in base and target cohorts

Author(s)

Song Zhai

References

Zhai, S., Guo, B., Wu, B., Mehrotra, D.V., and Shen, J., 2023. Integrating multiple traits for improving polygenic risk prediction in disease and pharmacogenomics GWAS (submitted).

Examples

```
dat <- generate_dis_data(structure = "clustered", sparseness = "more",
rho_DT = c(0.5,0.5,0.5,0.5), rho_T = 0.5, rho_C = 0.1,
K=4, m=1000, pcausal=0.1, blocksize=100,
samplesize=700, h2_base=0.3, h2_target=0.3)
plot(dat$truesize[,1])
print(dat$corr)
```

---

generate_pgx_data *Simulate PGx GWAS data for mtPRS analysis*

Description

Simulate PGx GWAS data for mtPRS analysis, including both disease GWAS summary statistics in base cohort and individual-level data in target cohort

Usage

```
generate_pgx_data(
  structure = "CS",
  sparseness = "no",
  rho_DT,
  rho_T,
  rho_E,
  rho_C,
  K = 4,
  m = 10000,
  pcausal = 0.01,
  blocksize = 100,
  gamma = 1,
  samplesize = 700,
  h2_base = 0.5,
  h2_target = 0.5
)
```

Arguments

structure	genetic correlation structure, either uniformly correlated (i.e., "CS" or "AR1") or clustered correlated ("clustered")
sparseness	effect sparseness, either "no", or "half", or "more"
rho_DT	effect correlations between traits in base cohort and the phenotype in target cohort
rho_T	genetic correlation among traits in base cohort
rho_E	correlation between prognostic and predictive effects in target cohort
rho_C	between-cluster correlation when the genetic correlation structure is clustered correlated
K	number of traits
m	number of SNPs
pcausal	proportion of causal SNPs
blocksize	size of each LD block (i.e., the number of SNPs in each LD block)
gamma	prognostic-to-predictive effect size ratio
samplesize	sample size of target cohort
h2_base	heritability in base cohort
h2_target	heritability in target cohort

Details

Simulate PGx GWAS data for mtPRS analysis into 12 (2x3x2) different genetic architectures

Value

a list of disease GWAS summary statistics in base cohort, individual-level data in target cohort, genetic correlation matrix, and true effect sizes in base and target cohorts

Author(s)

Song Zhai

References

Zhai, S., Guo, B., Wu, B., Mehrotra, D.V., and Shen, J., 2023. Integrating multiple traits for improving polygenic risk prediction in disease and pharmacogenomics GWAS (submitted).

Examples

```
dat <- generate_pgx_data(structure = "clustered", sparseness = "more",
rho_DT = c(0.5,0.5,0.5,0.5), rho_T = 0.5, rho_E = 0.3, rho_C = 0.1,
K=4, m=1000, pcausal=0.1, blocksize=100,
gamma=1, samplesize=700, h2_base=0.3, h2_target=0.3)
plot(dat$truesize[,1])
print(dat$corr)
```

mtPRS_O *Multi-trait PRS association test*

Description

Aggregate p-values from mtPRS-PCA, mtPRS-ML (Machine Learning), and stPRSs together with Cauchy Combination Test (CCT) for the robust multi-trait PRS association test

Usage

```
mtPRS_O(dat, pcut, varcut, K, phenotype)
```

Arguments

dat	a list of inputs including disease GWAS summary statistics of K traits from base cohort, individual-level data from base cohort, genetic correlation matrix, and true effect sizes
pcut	p-value cutoff for C+T method to construct individual stPRS
varcut	choose top principal components (PCs) until explaining varcut-percent variance
K	number of traits
phenotype	the type of phenotype in target cohort, either "dis" or "pgx"

Details

mtPRS_O needs disease GWAS summary statistics from K traits, and genetic correlation matrix among K traits

Value

if the phenotype is from disease GWAS, then return main p-value; if the phenotype is from PGx GWAS (with both T and C arms), then return main p-value from 2df test, interaction p-value from 2df test, main p-value in T arm only, and main p-value in C arm only

Author(s)

Song Zhai

References

Zhai, S., Guo, B., Wu, B., Mehrotra, D.V., and Shen, J., 2023. Integrating multiple traits for improving polygenic risk prediction in disease and pharmacogenomics GWAS (submitted).

Examples

```
dat <- generate_pgx_data(structure = "clustered", sparseness = "more",
rho_DT = c(0.5,0.5,0.5,0.5), rho_T = 0.5, rho_E = 0.3, rho_C = 0.1,
K=4, m=1000, pcausal=0.1, blocksize=100,
gamma=1, samplesize=700, h2_base=0.3, h2_target=0.3)
re <- mtPRS_O(dat, pcut = 0.05, varcut = 0.8, K = 4, phenotype = "pgx")
print(re)
```

mtPRS_PCA *Construct mtPRS by performing PCA on genetic correlation matrix among traits*

Description

Combine multiple individual single-trait PRSs (stPRSs) with weights calculated by performing principal component analysis (PCA) on genetic correlation matrix among traits

Usage

```
mtPRS_PCA(dat, pcut = 0.05, varcut = 0.8, K = 4, phenotype = "pgx")
```

Arguments

dat	a list of inputs including disease GWAS summary statistics of K traits from base cohort, individual-level data from base cohort, genetic correlation matrix, and true effect sizes
pcut	p-value cutoff for C+T method to construct individual stPRS
varcut	choose top principal components (PCs) until explaining varcut-percent variance
K	number of traits
phenotype	the type of phenotype in target cohort, either "dis" or "pgx"

Details

mtPRS_PCA needs disease GWAS summary statistics from K traits, and genetic correlation matrix among K traits

Value

a list of weights, individual stPRSs, and mtPRS-PCA

Author(s)

Song Zhai

References

Zhai, S., Guo, B., Wu, B., Mehrotra, D.V., and Shen, J., 2023. Integrating multiple traits for improving polygenic risk prediction in disease and pharmacogenomics GWAS (submitted).

Examples

```
dat <- generate_pgx_data(structure = "clustered", sparseness = "more",
rho_DT = c(0.5,0.5,0.5,0.5), rho_T = 0.5, rho_E = 0.3, rho_C = 0.1,
K=4, m=1000, pcausal=0.1, blocksize=100,
gamma=1, samplesize=700, h2_base=0.3, h2_target=0.3)
re <- mtPRS_PCA(dat, pcut = 0.05, varcut = 0.8, K = 4, phenotype = "pgx")
hist(re$mtPRS)
```

Index