Package ‘PRSPGx’

October 12, 2022

**Type** Package

**Title** Construct PGx PRS

**Version** 0.3.0

**Maintainer** Song Zhai <zsviolet1993@gmail.com>

**Description** Construct pharmacogenomics (PGx) polygenic risk score (PRS) with PRS-PGx-Unadj (unadjusted), PRS-PGx-CT (clumping and thresholding), PRS-PGx-L, -GL, -SGL (penalized regression), PRS-PGx-Bayes (Bayesian regression). Package is based on "Pharmacogenomics Polygenic Risk Score for Drug Response Prediction Using PRS-PGx Methods" by Zhai, S., Zhang, H., Mehrotra, D.V., and Shen, J., 2021 (submitted).

**License** GPL (>= 2)

**Depends** R (>= 4.0.0)

**Imports** gglasso (>= 1.5.0), SGL (>= 1.3.0), glmnet (>= 4.0.2), bigsnpr

(>= 1.5.2), Matrix (>= 1.2.18), GIGrvg (>= 0.5.0), MCMCpack (>= 1.4.6), bdsmatrix (>= 1.3.4), bigparser (>= 0.4.0), lmtest (>= 0.9.37), mvtnorm (>= 1.3.4), bigstatsr (>= 3.6.3), bigstatsr (>= 1.2.3), Rfast (>= 1.9.9), matrixcalc (>= 1.0-3)

**Suggests** knitr, rmarkdown

**VignetteBuilder** knitr

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.2

**NeedsCompilation** yes

**Author** Song Zhai [aut, cre]

**Repository** CRAN

**Date/Publication** 2022-07-20 15:00:02 UTC

**R topics documented:**

PRSPGx.example .......................... 2
PRSPGx.example

Simulated example data

Description

Simulated example data required by PRS-DIS and PRS-PGx functions.

Usage

data(PRSPGx.example)

Format

A list with 8 sublists:

**PGx_GWAS** PGx GWAS including SNP ID, MAF, position, $\beta$, $\alpha$, 2-df p-value, and N; SD(Y), and mean(T)

**DIS_GWAS** disease GWAS including SNP ID, MAF, position, $\beta$, SE($\beta$), p-value, and N

**G_reference** simulated individual-level genotype from the reference panel matched with the simulated sample PGx genotype

**Y** simulated phenotype (continuous)

**T** simulated treatment assignment, 1 = treatment, 0 = placebo

**G** simulated sample PGx genotype with 100 SNPs and 4000 subjects

**beta** simulated prognostic effect sizes (i.e., the underlying true prognostic effect sizes)

**alpha** simulated predictive effect sizes (i.e., the underlying true predictive effect sizes)
Construct disease PRS unadjusted or using clumping and thresholding

Description

Shrink prognostic effect sizes by p-value cutoff (PRS-Dis-CT turns out to be PRS-Dis-Unadj when setting p-value cutoff = 1)

Usage

PRS_Dis_CT(
  DIS_GWAS,
  G_reference,
  pcutoff = 1e-05,
  clumping = TRUE,
  p1 = 1e-04,
  d1 = 250000,
  r1 = 0.8
)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIS_GWAS</td>
<td>a numeric matrix containing disease GWAS summary statistics, including SNP ID, position, $\beta$, SE($\beta$), p-value, N, and MAF</td>
</tr>
<tr>
<td>G_reference</td>
<td>a numeric matrix containing the individual-level genotype information from the reference panel (e.g., 1KG)</td>
</tr>
<tr>
<td>pcutoff</td>
<td>a numeric value indicating the p-value cutoff</td>
</tr>
<tr>
<td>clumping</td>
<td>a logical flag indicating should clumping be performed</td>
</tr>
<tr>
<td>p1</td>
<td>a numeric value indicating p-value threshold to decide flag SNPs in clumping</td>
</tr>
<tr>
<td>d1</td>
<td>a numeric value indicating window size in clumping</td>
</tr>
<tr>
<td>r1</td>
<td>a numeric value indicating correlation in clumping</td>
</tr>
</tbody>
</table>

Details

PRS-Dis-CT automatically sets predictive effect sizes equivalent to the prognostic effect sizes; and only need disease GWAS summary statistics

Value

A numeric list, the first sublist contains estimated prognostic effect sizes, the second sublist contains estimated predictive effect sizes

Author(s)

Song Zhai
References


Examples

```r
data(PRSPGx.example); attach(PRSPGx.example)
coef_est <- PRS_Dis_CT(DIS_GWAS, G_reference, pcutoff = 0.01, clumping = TRUE)
summary(coef_est$coef.G)
summary(coef_est$coef.TG)
```

PRS_Dis_LDpred2

*Construct disease PRS using LDpred2*

**Description**

Using snp_ldpred2_grid function from bigsnpr function

**Usage**

```r
PRS_Dis_LDpred2(DIS_GWAS, G_reference, pcausal, h2)
```

**Arguments**

- **DIS_GWAS**: a numeric matrix containing disease GWAS summary statistics, including SNP ID, position, $\beta$, SE($\beta$), p-value, N, and MAF
- **G_reference**: a numeric matrix containing the individual-level genotype information from the reference panel (e.g., 1KG)
- **pcausal**: a numeric value indicating the hyper-parameter as the proportion of causal variants
- **h2**: a numeric value indicating the estimated heritability

**Details**

PRS-Dis-LDpred2 automatically sets predictive effect sizes equivalent to the prognostic effect sizes; and only need disease GWAS summary statistics and external reference genotype

**Value**

A numeric list, the first sublist contains estimated prognostic effect sizes, the second sublist contains estimated predictive effect sizes
Construct PGx PRS using Bayesian regression

**Description**

Flexibly shrink prognostic and predictive effect sizes simultaneously with glocal-local shrinkage parameters.

**Usage**

```r
PRS_PGx_Bayes(
  PGx_GWAS,
  G_reference,
  n.itr = 1000,
  n.burnin = 500,
  n.gap = 10,
  paras,
  standardize = TRUE
)
```

**Arguments**

- `PGx_GWAS` a numeric list containing PGx GWAS summary statistics (with SNP ID, position, $\beta$, $\alpha$, 2-df p-value, MAF and N), SD(Y), and mean(T).
- `G_reference` a numeric matrix containing the individual-level genotype information from the reference panel (e.g., 1KG).
- `n.itr` a numeric value indicating the total number of MCMC iteration.
n.burnin  
... a numeric value indicating the number of burn in
n.gap  
... a numeric value indicating the MCMC gap
paras  
... a numeric vector containing hyper-parameters $(v, \phi)$
standardize  
... a logical flag indicating should phenotype and genotype be standardized

**Details**

PRS-PGx-Bayes only needs PGx summary statistics and external reference genotype

**Value**

A numeric list, the first sublist contains estimated prognostic effect sizes, the second sublist contains estimated predictive effect sizes

**Author(s)**

Song Zhai

**References**


**Examples**

```r
data(PRSPGx.example); attach(PRSPGx.example)
paras = c(3, 5)
coef_est <- PRS_PGx_Bayes(PGx_GWAS, G_reference, paras = paras, n.itr = 10, n.burnin = 5, n.gap = 1)
summary(coef_est$coef.G)
summary(coef_est$coef.TG)
```

**PRS_PGx_CT**

*Construct PGx PRS unadjusted or using clumping and thresholding*

**Description**

Shrink prognostic and predictive effect sizes simultaneously by 2-df (main and interaction) p-value cutoff (PRS-PGx-CT turns out to be PRS-PGx-Unadj when setting p-value cutoff = 1)
Usage

PRS_PGx_CT(
    PGx_GWAS,
    G_reference,
    pcutoff = 1e-04,
    clumping = TRUE,
    p1 = 1e-04,
    d1 = 250000,
    r1 = 0.8
)

Arguments

  PGx_GWAS    a numeric matrix containing PGx GWAS summary statistics, including SNP ID, MAF, position, \( \beta \), \( \alpha \), 2-df p-value, and N
  G_reference a numeric matrix containing the individual-level genotype information from the reference panel (e.g., 1KG)
  pcutoff     a numeric value indicating the p-value cutoff
  clumping    a logical flag indicating should clumping be performed
  p1          a numeric value indicating p-value threshold to decide flag SNPs in clumping
  d1          a numeric value indicating window size in clumping
  r1          a numeric value indicating correlation in clumping

Details

PRS-PGx-CT only needs PGx summary statistics

Value

A numeric list, the first sublist contains estimated prognostic effect sizes, the second sublist contains estimated predictive effect sizes, the third sublist contains 2-df p-values

Author(s)

Song Zhai

References


Examples

data(PRSPGx.example); attach(PRSPGx.example)
coef_est <- PRS_PGx_CT(PGx_GWAS, G_reference, pcutoff = 0.01, clumping = TRUE)
summary(coef_est$coef.G)
summary(coef_est$coef.TG)
**Description**

Shrink prognostic and predictive effect sizes simultaneously via the penalized term. With different assumptions on the relationship between the two effects, can be PRS-PGx-L (Lasso), PRS-PGx-GL (Group Lasso), and PRS-PGx-SGL (Sparse Group Lasso).

**Usage**

```r
PRS_PGx_Lasso(Y, Tr, G, intercept = TRUE, lambda, method, alpha = 0.5)
```

**Arguments**

- `Y` a numeric vector containing the quantitative trait
- `Tr` a numeric vector containing the treatment assignment
- `G` a numeric matrix containing genotype information
- `intercept` a logical flag indicating should intercept be fitted (default=TRUE) or set to be FALSE
- `lambda` a numeric value indicating the penalty
- `method` a logical flag for different penalized regression methods: 1 = PRS-PGx-L, 2 = PRS-PGx-GL, 3 = PRS-PGx-SGL
- `alpha` a numeric value indicating the mixing parameter (only used when method = 3).
  - `alpha = 1` is the lasso penalty.
  - `alpha = 0` is the group lasso penalty

**Details**

PRS-PGx-Lasso requires individual-level data

**Value**

A numeric list, the first sublist contains estimated prognostic effect sizes, the second sublist contains estimated predictive effect sizes

**Author(s)**

Song Zhai

**References**


Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Fit a GLM (or cox model) with a combination of lasso and group lasso regularization. R package version, 1 (2015).

Examples

data(PRSPGx.example); attach(PRSPGx.example)
coef_est <- PRS_PGx_Lasso(Y, Tr, G, lambda = 1, method = 1)
summary(coef_est$coef.G)
summary(coef_est$coef.TG)
Index

* datasets
  - PRSPGx.example, 2
  - PRS_DiS_CT, 3
  - PRS_DiS_LDpred2, 4
  - PRS_PGx_Bayes, 5
  - PRS_PGx_CT, 6
  - PRS_PGx_Lasso, 8
  - PRSPGx.example, 2