PLACO+ Software Documentation

June 19, 2025

Version 0.2.0
Date June 19, 2025
Title Pleiotropic analysis under composite null hypothesis
Correspondence Debashree Ray, Ph.D. <dray@jhu.edu>
Description PLACO+ implements a variant-level formal statistical test of pleiotropy of any two traits using summary-level GWAS data, and can account for potential correlation across traits, such as that arising due to sample overlap.
Depends R (>= 3.0.1)

`cor.pearson` *To estimate correlation between Z-scores from two traits*

Description

The R function `cor.pearson` allows the user to estimate the correlation between Z-scores from two traits. It is estimated only once irrespective of the number p of genetic variants to be tested for pleiotropic association.

Usage

```
cor.pearson(Z.matrix, P.matrix, p.threshold=1e-4, returnMatrix=TRUE)
```

Arguments

`Z.matrix` It is the $p \times 2$ matrix of Z-scores where p is total number of variants in the GWAS dataset. The 2 columns correspond to the 2 traits. For a trait, Z-score is the ratio of estimated genetic effect to its standard error ($Z = \hat{\beta}/\hat{se}$).

`P.matrix` It is the $p \times 2$ matrix of p-values of the 2 traits where p is total number of variants in the GWAS dataset. Like the Z-scores (summary statistics) in `Z.matrix`, the p-values of `P.matrix` correspond to individual test of each trait against each genetic variant. The order in which traits and the genetic variants are arranged in `Z.matrix` and `P.matrix` must be same.

`p.threshold` The p-value threshold used to determine which genetic variants are likely not associated. A liberal threshold needs to be used to screen out any signal that may affect the estimate of the correlation matrix `R`. Genetic variants (here, rows) with p-values smaller than this threshold for any trait are removed before estimating `R`. Default value is 10^{-4}.

`returnMatrix` This logical argument determines if the estimated 2×2 correlation matrix will be returned (default) or just the correlation parameter estimate (scalar quantity).

Value

`R` [`returnMatrix=TRUE`] The estimated 2×2 correlation matrix of the GWAS summary statistics of the 2 traits under the complete null hypothesis of no association.

`R[1,2]` [`returnMatrix=FALSE`] The estimated correlation between the 2 traits.

`var.placo` *To estimate the variance parameters needed to implement PLACO+*

Description

The R function `var.placo` estimates the variances of the Z-scores of the two traits under the composite null hypothesis of no pleiotropy. Output from this function goes as input for

R function `placo` (original PLACO for independent traits) or `placo.plus` (PLACO+ for any two traits). This estimation procedure is done only once for a given study using the single-trait Z-scores and p-values (or GWAS summary statistics) that are usually publicly available.

Usage

```
var.placo(Z.matrix, P.matrix, p.threshold=1e-4)
```

Arguments

Arguments for `var.placo` are the same as those for `cor.pearson`.

Value

`VarZ`	A vector of estimated variances for the Z-scores of 2 traits.

`placo` *Pleiotropic association test of two independent or uncorrelated traits using GWAS summary statistics*

Description

PLACO uses genome-wide summary statistics (Z-scores and p-values) on two *independent* or *uncorrelated* traits to test for variant-level pleiotropic association between a genetic marker and the two traits.

Usage

```
placo(Z, VarZ, AbsTol=.Machine$double.eps^0.8)
```

Arguments

`Z`	The vector of Z-scores of 2 traits for a given genetic variant.
`VarZ`	The vector of estimated variances for the Z-scores of 2 traits as estimated by the function `var.placo()`.
`AbsTol`	The user can specify the absolute tolerance value used in the numerical integration for evaluating PLACO p-value. Default value is 3×10^{-13}. Function `integrate()` is used for numerical integration.

Details

Consider two genome-wide studies of traits Y_1 and Y_2 on n_1 and n_2 individuals respectively who were genotyped and/or imputed or sequenced at p genetic variants. Let $\boldsymbol{Y}_k$ and $\boldsymbol{X}_k$ be the vectors of k-th trait values and genotypes at a given genetic variant respectively on all n_k individuals ($k = 1, 2$). For the k-th trait, suppose β_k is the genetic effect and the corresponding summary statistic for testing no genetic association of the trait is $Z_k = \hat{\beta}_k / \operatorname{se}(\hat{\beta}_k)$, where $\hat{\beta}_k$ is the maximum likelihood estimate (MLE) of β_k and $\operatorname{se}(\hat{\beta}_k)$ is its standard error. Publicly available GWAS data usually have information on $\hat{\beta}_k$ and $\operatorname{se}(\hat{\beta}_k)$, and/or Z_k and the corresponding p-value p_k, $k = 1, 2$.

The conventional cross-phenotype association methods test the global null hypothesis that none of the traits is associated with the given genetic variant (i.e., $\beta_1 = \beta_2 = 0$). Rejection of this global null can be due to one associated trait ($\beta_1 \neq 0, \beta_2 = 0$ or $\beta_1 = 0, \beta_2 \neq 0$). Here, we are interested in identifying the genetic variants that are associated with both the traits or outcomes (i.e., $\beta_1 \neq 0, \beta_2 \neq 0$). The effects of such a genetic variant on the traits may or may not be equal. Formally, our null hypothesis of no pleiotropy is H_0 : at most 1 trait is associated with the genetic variant while the alternative hypothesis is H_a : both traits are associated. Mathematically, our null hypothesis of no pleiotropy is a composite null hypothesis, and can simply be written as $H_0 : \beta_1\beta_2 = 0$ vs. the alternative hypothesis $H_a : \beta_1\beta_2 \neq 0$.

The PLACO test statistic and approximate asymptotic p-value for testing the composite null hypothesis H_0, *assuming the two traits are independent*, are

$$
\begin{aligned}
T_{\text{PLACO}} &= Z_1 Z_2 \\
p_{\text{PLACO}} &= \mathbb{F}\left(z_1 z_2 / \sqrt{\text{Var}(Z_1)}\right) + \mathbb{F}\left(z_1 z_2 / \sqrt{\text{Var}(Z_2)}\right) - \mathbb{F}\left(z_1 z_2\right)
\end{aligned}
$$

where z_1 and z_2 are the observed Z-scores for the two traits at a given genetic variant; $\text{Var}(Z_1)$ and $\text{Var}(Z_2)$ are the estimated marginal variances of the Z-scores (as estimated by the function `var.placo()`); and $\mathbb{F}(u) = 2\int_{|u|}^{\infty} \mathbb{f}(x)dx$ is the two-sided tail probability of a (symmetric) normal product distribution at value u.

If the two traits come from studies with overlapping samples, either partially (e.g., case-control traits with shared controls) or completely, then the Z-scores will be correlated. We recommend that PLACO be applied to uncorrelated traits, and PLACO+ (implemented using

the `placo.plus()` function) for correlated traits. PLACO is only applicable genome-wide and not to a selected set of genetic variants.

For more details on how PLACO may be used, please refer Ray and Chatterjee (2020).

Value

`T.placo`	The PLACO statistic for the test of pleiotropic association of a single variant and two traits.
`p.placo`	The approximate asymptotic p-value of PLACO.

Reference

Ray, D. and Chatterjee, N. A powerful method for pleiotropic analysis under composite null hypothesis identifies novel shared loci between type 2 diabetes and prostate cancer. *PLoS Genetics*, **16(12)**: e1009218, 2020.

Example

```
#-------- Download or directly source PLACO
# require(devtools)
# source_url("https://github.com/RayDebashree/PLACO/blob/master/
PLACO_v0.2.0.R?raw=TRUE")
#--------

source("PLACO_v0.2.0.R")
set.seed(1)
## For an example, let's first simulate a toy set of GWAS summary
## statistics on 2 uncorrelated traits and 1000 variants
require(MASS)
k <- 2
p <- 1000
Z.matrix <- mvrnorm(n=p, mu=rep(0,k), Sigma=diag(1,k))
P.matrix <- matrix(NA, nrow=p, ncol=k)
for(j in 1:k){
    P.matrix[,j] <- sapply(1:nrow(Z.matrix),
        function(i) pchisq(Z.matrix[i,j]^2,df=1,ncp=0,lower.tail=F))
}
colnames(Z.matrix) <- paste("Z",1:k,sep="")
colnames(P.matrix) <- paste("P",1:k,sep="")
```

```
## Steps to implementing PLACO
# Step 1: Obtain the variance parameter estimates (only once)
VarZ <- var.placo(Z.matrix, P.matrix, p.threshold=1e-4)
# Step 2: Apply test of pleiotropy for each variant
out <- sapply(1:p, function(i) placo(Z=Z.matrix[i,], VarZ=VarZ))
# Check the output for say variant 100
dim(out)
out[,100]$T.placo
out[,100]$p.placo
```

`placo.plus` *A more general, robust pleiotropic association test of any two traits using GWAS summary statistics, where the traits may or may not be correlated, and may have unknown extent of sample overlap*

Description

PLACO+ uses genome-wide summary statistics (Z-scores and p-values) on *two possibly correlated traits* to test for variant-level pleiotropic association between a genetic marker and the two traits. PLACO, originally proposed by Ray and Chatterjee (2020), is a special case of PLACO+.

Usage

```
placo.plus(Z, VarZ, CorZ, AbsTol=.Machine$double.eps^0.8)
```

Arguments

`Z`	The vector of Z-scores of 2 traits for a given genetic variant.
`VarZ`	The vector of estimated variances for the Z-scores of 2 traits as estimated by the function `var.placo()`.
`CorZ`	The correlation parameter estimate, as estimated by the function `cor.pearson()` with `returnMatrix=FALSE` using the Z-scores of the 2 traits.
`AbsTol`	The user can specify the absolute tolerance value used in the numerical integration for evaluating PLACO+ p-value. Default value is 3×10^{-13}. Function `integrate()` is used for numerical integration.

Details

PLACO+ is a robust general statistical method for testing pleiotropic effects of a genetic variant that is applicable to any two arbitrary traits (e.g., quantitative, case-control, count traits) with or without overlapping samples that may or may not be known. The strengths and directions of effects of a pleiotropic genetic variant on the traits can vary, and need not be equal. Being based on GWAS summary statistics, PLACO+ is applicable to traits that may come from study designs not involving random sampling of unrelated individuals. The original PLACO proposed by Ray and Chatterjee (2020) is a special case of PLACO+.

PLACO+ tests the null hypothesis of no pleiotropy H_0 : at most 1 trait is associated with the genetic variant while the alternative hypothesis is H_a : both traits are associated. Mathematically, our null hypothesis of no pleiotropy is a composite null hypothesis, and can simply be written as $H_0 : \beta_1\beta_2 = 0$ vs. the alternative hypothesis $H_a : \beta_1\beta_2 \neq 0$.

The PLACO+ test statistic and approximate asymptotic p-value for testing the composite null hypothesis H_0 are

$$T_{\text{PLACO+}} = Z_1 Z_2$$
$$p_{\text{PLACO+}} = \mathbb{F}_\rho\left(z_1 z_2 / \sqrt{\text{Var}(Z_1)}\right) + \mathbb{F}_\rho\left(z_1 z_2 / \sqrt{\text{Var}(Z_2)}\right) - \mathbb{F}_\rho\left(z_1 z_2\right)$$

where z_1 and z_2 are the observed Z-scores for the two traits at a given genetic variant; $\text{Var}(Z_1)$ and $\text{Var}(Z_2)$ are the estimated marginal variances of the Z-scores (as estimated by the function `var.placo()`); $\mathbb{F}_\rho(u) = \int_{-\infty}^{-|u|} \mathbb{f}_\rho(x) + \int_{|u|}^{\infty} \mathbb{f}_\rho(x)dx$ is the two-sided tail

probability of a (asymmetric) bivariate normal product distribution $\mathbb{f}_\rho(x)$ at value u; and ρ is the correlation between Z-scores (as estimated by the function `cor.pearson()`).

Although PLACO and PLACO+ test statistics are identical, these methods differ in their p-values based on how the trait correlation is accommodated through the null distribution. When the traits are uncorrelated ($\rho = 0$), PLACO+ and PLACO are identical. We suggest that `placo()` be used for uncorrelated traits since the more general `placo.plus()` can take longer time. PLACO+ is only applicable genome-wide and not to a selected set of genetic variants.

For more details on how PLACO+ may be used, please refer Park and Ray (2025+). We request that the references for Ray and Chatterjee (2020), and Park and Ray (2025+) be cited if this software is used in any publication.

Value

`T.placo.plus`	The PLACO+ statistic for the test of pleiotropic association of a single variant and two traits.
`p.placo.plus`	The approximate asymptotic p-value of PLACO+.

Reference

Park, J. and Ray, D. A robust pleiotropy method with applications to lipid traits and to inflammatory bowel disease subtypes with sample overlap. *Submitted*, 2025+.

Example

```
#-------- Download or directly source PLACO+
# require(devtools)
# source_url("https://github.com/RayDebashree/PLACO/blob/master/
PLACO_v0.2.0.R?raw=TRUE")
#--------

source("PLACO_v0.2.0.R")
set.seed(1)
## For an example, let's first simulate a toy set of GWAS summary
## statistics on 2 correlated traits and 1000 variants
require(MASS)
k <- 2
```

```
p <- 1000
r <- 0.5    # Desired correlation between traits
# Covariance matrix with specified correlation
Sigma <- matrix(r, nrow=k, ncol=k)
diag(Sigma) <- 1

Z.matrix <- mvrnorm(n=p, mu=rep(0,k), Sigma=Sigma)
P.matrix <- matrix(NA, nrow=p, ncol=k)
for(j in 1:k){
    P.matrix[,j] <- sapply(1:nrow(Z.matrix),
        function(i) pchisq(Z.matrix[i,j]^2,df=1,ncp=0,lower.tail=F))
}
colnames(Z.matrix) <- paste("Z",1:k,sep="")
colnames(P.matrix) <- paste("P",1:k,sep="")

## Steps to implementing PLACO+
# Step 1: Obtain the variances and correlation parameter estimates
# (done only once using all variants)
VarZ <- var.placo(Z.matrix, P.matrix, p.threshold=1e-4)
CorZ <- cor.pearson(Z.matrix, P.matrix, p.threshold = 1e-4, returnMatrix=F)
# Step 2: Apply test of pleiotropy for each variant
out <- sapply(1:p, function(i) placo.plus(Z=Z.matrix[i,], VarZ=VarZ, CorZ=CorZ))
# Check the output for say variant 100
dim(out)
out[,100]$T.placo.plus
out[,100]$p.placo.plus
```