Integrating multiple traits for improving polygenic risk prediction in disease and pharmacogenomics GWAS

```
library(mtPRS)
```

Contents

Overview

mtPRS package (Zhai et al., 2023) implements two novel multi-trait polygenic risk score (mtPRS) methods, mtPRS-PCA and mtPRS-O. Specifically, mtPRS-PCA combines individual single-trait PRSs (stPRSs) with weights calculated by performing principal component analysis (PCA) on the genetic correlation matrix among traits. mtPRS-O aggregates p-values from mtPRS-PCA, mtPRS-ML (Machine Learning; Krapohl et al., 2018), and all stPRSs using Cauchy Combination Test (CCT) (Liu et al., 2019) to provide a robust way for the multi-trait PRS association test.

System Requirements

The package development version is tested on the following systems:

Mac OSX: Mojave version 10.14.6 (R version 3.6.3)

Windows 10 (R version 3.6.3)

The CRAN package should be compatible with Windows and Mac operating systems.

Installing Guide

`mtPRS` package requires R with version 3.6.3 or higher, which can be downloaded and installed from here.

Package dependencies

Users should install the following packages prior to installing `mtPRS`, from an `R` terminal:

```
install.packages(c('glmnet','Matrix','mvtnorm','stats','ACAT','dplyr','data.table'))
```

Package Installation

To install `mtPRS`, type the following code from an `R` session:

```
library(devtools)
devtools::install_github("zhaiso1/mtPRS")

library(mtPRS)
```

Demo

Step 1: Prepare disease GWAS summary statistics in base cohort and individual-level data in target cohort (we take PGx GWAS data (i.e., phenotype = drug response) for example in this Readme file)

In this section, we will simulate an example data with our simulation algorithm, in which the list includes the following elements:

- **base**: disease GWAS summary statistics of K traits in base cohort;
- **target**: individual-level data in target cohort, including the drug response, the treatment assignment, and the genotype for PGx GWAS;
- **corr**: underlying genetic correlation matrix among traits;
- **truesize**: the simulated true effects of K traits in base cohort (μ), and the simulated true prognostic (β) and predictive (α) effects in target cohort.

```
## Simulate data for PGx GWAS mtPRS analysis

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)
#> [1] "Step 1. Prepare correlation matrix"
#> [1] "Step 2. Simulate true effect size"
#> [1] "Step 3. Simulate base cohort"
#> [1] "Step 4. Simulate target cohort"
#> [1] "Step 5. Output results"
```

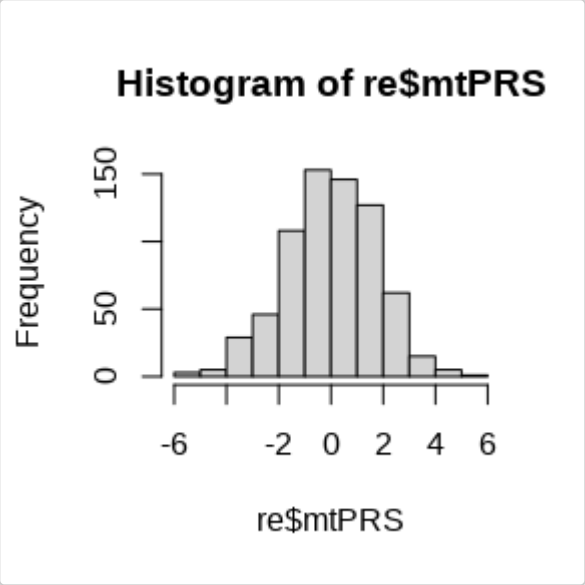
Step 2. Run mtPRS-PCA method

mtPRS-PCA will output a list of

- weights
- stPRSs (standardized)
- mtPRS-PCA

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

hist(re$mtPRS)
```



Step 3. Run mtPRS-O method

If the phenotype is from disease GWAS, then **mtPRS-O** will output

- main prognostic p-value.

If the phenotype is from PGx GWAS (with both T and C arms), then **mtPRS-O** will output a vector of

- main prognostic p-value from 2df test;
- interaction predictive p-value from 2df test;
- main prognostic p-value in T arm only;
- main prognostic p-value in C arm only.

```
re <- mtPRS_O(dat, pcut = 0.05, varcut = 0.8, K = 4, phenotype = "pgx")

print(re)
#>   p_2df_prog   p_2df_pred     p_T_prog     p_C_prog
#> 2.471470e-05 2.375257e-04 1.583040e-21 5.452460e-06
```

References

Krapohl E, Patel H, Newhouse S, et al. Multi-polygenic score approach to trait prediction. Mol. psychiatry 2018; 23: 1368-1374

Liu Y, Chen S, Li Z, et al. ACAT: a fast and powerful p-value combination method for rare-variant analysis in sequencing studies. Am. J. Hum. Genet. 2019; 104: 410-421.

Zhai S, Guo B, Wu B, Mehrotra DV, and Shen J. Integrating multiple traits for improving polygenic risk prediction in disease and pharmacogenomics GWAS. Submitted to Briefings in Bioinformatics, 2023.