Package 'ldscR'

September 12, 2025

Type Package

Title Heritability and Genetic Correlation Matrix Estimation Using Linkage Disequilibrium Score Regression

Version 0.1.0

Author Yihe Yang

Maintainer Yihe Yang <yxy1234@case.edu>

Description This package provides tools for estimating heritability and genetic correlation matrices using Linkage Disequilibrium Score Regression (LDSC). It automates allele harmonization, merges GWAS results from multiple traits, and estimates the genetic covariance matrix for these traits. The diagonal elements of the genetic covariance matrix represent heritability estimates, while the corresponding correlation matrix represents the genetic correlation coefficients.

License MIT

Encoding UTF-8

LazyData true

Depends R (>= 2.10)

Imports data.table (>= 1.12.0),
arrow,
CppMatrix

RoxygenNote 7.3.2

Config/testthat/edition 3

Contents

`download_7MLDSC` *Download 7M SNP LD Score Reference Panels*

Description

Downloads LD Score reference panels computed from UK Biobank imputed data containing approximately 7 million SNPs. These panels are derived from the SBayesRC project which provides high-density LD reference data for multiple ancestries and cross-population combinations.

The LD reference panels include:

- Single ancestry panels: EUR, EAS, AFR computed from UK Biobank samples of respective ancestries using ~7M imputed SNPs
- Cross-population panels: EURxEAS, EURxAFR, EASxAFR computed by combining LD information across ancestry groups
- SNP information file (7m_snp.parquet) containing SNP identifiers and allele information for harmonization

These high-density panels improve polygenic prediction accuracy compared to traditional HapMap3-based references by leveraging the full spectrum of common variation captured by modern imputation.

Usage

```
download_7MLDSC(ancestry)
```

Arguments

`ancestry`	Character string specifying which LD reference to download. Options: "EUR", "EAS", "AFR", "EURxEAS", "EURxAFR", "EASxAFR"

Details

The function downloads parquet files from a remote server and loads them directly into the global environment using assign(). The SNP information file is only downloaded once per session to avoid redundant transfers.

File sizes are substantial (~500MB-1GB each) due to the high SNP density. Ensure adequate internet connection and disk space.

Value

Invisible NULL. The function loads data directly into the global environment:

`SevenMillon_[ancestry]LDSC`
: Data frame with LD scores for the requested ancestry

`SevenMillon_SNP`
: Data frame with SNP information (loaded once per session)

Source

LD reference panels computed using SBayesRC methodology from: https://github.com/zhilizheng/SBayesRC

Based on UK Biobank imputed genotype data with ~7 million common SNPs, providing enhanced LD estimation compared to sparse reference panels.

References

Zheng, Z., et al. (2024). Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. Nature Genetics.

Examples

```
## Not run:
# Download European LD reference
download_7MLDSC("EUR")
head(SevenMillon_EURLDSC)  # LD scores now available in workspace

# Download cross-population EUR x EAS reference
download_7MLDSC("EURxEAS")
dim(SevenMillon_EURxEASLDSC)  # Cross-population LD scores

# Check SNP information (loaded automatically)
head(SevenMillon_SNP)

# Check what's in your workspace
ls(pattern = "^SevenMillon_")

## End(Not run)
```

`filter_align` *Filter and Align GWAS Data to a Reference Panel*

Description

The `filter_align` function processes a list of GWAS summary statistics data frames, harmonizes alleles according to a reference panel, removes duplicates, and aligns data to common SNPs. It's used to prepare data for further analysis such as LDSC.

Usage

```
filter_align(gwas_data_list, ref_panel, allele_match = TRUE)
```

Arguments

`gwas_data_list`	A list of data.frames where each data.frame contains GWAS summary statistics for a trait. Each data.frame should include columns for SNP identifiers, Z-scores of effect size estimates, sample sizes (N), effect allele (A1), and reference allele (A2).
`ref_panel`	A data.frame containing the reference panel data. It must include columns for SNP, A1, and A2.
`allele_match`	Logical. Whether to match alleles. Default `TRUE`.

Details

The function performs several key steps: adjusting alleles according to a reference panel, removing duplicate SNPs, and aligning all GWAS data frames to a set of common SNPs. This is often a necessary preprocessing step before performing genetic correlation and heritability analyses.

Value

A list of data.frames, each corresponding to an input GWAS summary statistics data frame, but filtered, harmonized, and aligned to the common SNPs found across all data frames.

Examples

```
## Not run:
# Assuming GWAS_List and ref_panel are already defined:
GWAS_List <- filter_align(GWAS_List, ref_panel)

## End(Not run)
```

Hapmap3 *HapMap3 SNP Information for Allele Harmonization*

Description

SNP information dataset containing approximately 1.7 million SNPs in common with either HapMap3 or the UK Biobank, used for allele harmonization in genetic analyses.

Usage

```
data("Hapmap3")
```

Format

A data frame with 1,664,852 observations and 3 variables:

`SNP` SNP identifier (rsID)

`A1` Effect allele

`A2` Reference allele

Source

Derived from bigsnpr package: https://privefl.github.io/bigsnpr/reference/download_1000G.html

Examples

```
data(Hapmap3)
head(Hapmap3)
dim(Hapmap3)
```

Hapmap3_AFRLDSC *African Population LD Score Reference Panel*

Description

LD Score reference panel for African (AFR) population, constructed using UK Biobank data. SNPs are ordered by chromosome and base pair position.

Usage

```
data("Hapmap3_AFRLDSC")
```

Format

A data frame with 1,234,911 observations and 2 variables:

`SNP` SNP identifier (rsID)

`LDSC` LD Score coefficient

Source

Constructed from UK Biobank AFR samples via PRScsx: https://github.com/getian107/PRScsx

Examples

```
data(Hapmap3_AFRLDSC)
head(Hapmap3_AFRLDSC)
dim(Hapmap3_AFRLDSC)
```

Hapmap3_AMRLDSC *Admixed American Population LD Score Reference Panel*

Description

LD Score reference panel for Admixed American (AMR) population, constructed using UK Biobank data. SNPs are ordered by chromosome and base pair position.

Usage

```
data("Hapmap3_AMRLDSC")
```

Format

A data frame with 1,183,356 observations and 2 variables:

`SNP` SNP identifier (rsID)

`LDSC` LD Score coefficient

Source

Constructed from UK Biobank AMR samples via PRScsx: https://github.com/getian107/PRScsx

Examples

```
data(Hapmap3_AMRLDSC)
head(Hapmap3_AMRLDSC)
dim(Hapmap3_AMRLDSC)
```

Hapmap3_EASLDSC *East Asian Population LD Score Reference Panel*

Description

LD Score reference panel for East Asian (EAS) population, constructed using UK Biobank data. SNPs are ordered by chromosome and base pair position.

Usage

```
data("Hapmap3_EASLDSC")
```

Format

A data frame with 1,031,412 observations and 2 variables:

`SNP` SNP identifier (rsID)

`LDSC` LD Score coefficient

Source

Constructed from UK Biobank EAS samples via PRScsx: https://github.com/getian107/PRScsx

Examples

```
data(Hapmap3_EASLDSC)
head(Hapmap3_EASLDSC)
dim(Hapmap3_EASLDSC)
```

Hapmap3_EURLDSC *European Population LD Score Reference Panel*

Description

LD Score reference panel for European (EUR) population, constructed using UK Biobank data. SNPs are ordered by chromosome and base pair position.

Usage

```
data("Hapmap3_EURLDSC")
```

Format

A data frame with 1,117,425 observations and 2 variables:

`SNP` SNP identifier (rsID)

`LDSC` LD Score coefficient

Source

Constructed from UK Biobank EUR samples via PRScsx: https://github.com/getian107/PRScsx

Examples

```
data(Hapmap3_EURLDSC)
head(Hapmap3_EURLDSC)
dim(Hapmap3_EURLDSC)
```

Hapmap3_SASLDSC *South Asian Population LD Score Reference Panel*

Description

LD Score reference panel for South Asian (SAS) population, constructed using UK Biobank data. SNPs are ordered by chromosome and base pair position.

Usage

```
data("Hapmap3_SASLDSC")
```

Format

A data frame with 1,133,674 observations and 2 variables:

`SNP` SNP identifier (rsID)

`LDSC` LD Score coefficient

Source

Constructed from UK Biobank SAS samples via PRScsx: https://github.com/getian107/PRScsx

Examples

```
data(Hapmap3_SASLDSC)
head(Hapmap3_SASLDSC)
dim(Hapmap3_SASLDSC)
```

`ldsc.bicov` *Bivariate LD Score Regression (LDSC) for Cross-Trait Genetic Covariance*

Description

'ldsc.bicov' estimates the genetic covariance between two traits using LD Score Regression (LDSC). The function harmonizes two GWAS summary-statistic tables with an LD-score table, fits a weighted least squares (WLS) regression of $z_{1j}z_{2j}$ on the LD regressor, and performs one (optional) reweighting step for efficiency. Block bootstrap can be used to obtain standard errors.

Usage

```
ldsc.bicov(gwas1, gwas2, h21, h22, LDSC, nblock = 500, sampling.time = 0)
```

Arguments

`gwas1`	A 'data.frame'/'data.table' of GWAS summary statistics for trait 1. Required columns: • 'SNP' — SNP identifier • 'Zscore' — Z-score of the SNP–trait association • 'N' — per-SNP sample size
`gwas2`	A 'data.frame'/'data.table' of GWAS summary statistics for trait 2, with the same required columns as 'gwas1'.
`h21`	Numeric. Heritability estimate of trait 1 (used in weights).
`h22`	Numeric. Heritability estimate of trait 2 (used in weights).
`LDSC`	A 'data.frame'/'data.table' containing LD scores. Required columns: • 'SNP' — SNP identifier • 'LDSC' — LD score
`nblock`	Integer. Number of blocks for block bootstrap.
`sampling.time`	Integer. Number of bootstrap replicates. If '0', only point estimates are returned.

Details

Let M be the number of SNPs after harmonization. The regressor is $\ell_j = \mathrm{LDSC}_j\sqrt{N_{1j}}\sqrt{N_{2j}}/M$. The initial WLS uses weights

$$w_j^{-1} = \big(1 + \mathrm{LDSC}_j N_{1j} h_{21}/M\big)\big(1 + \mathrm{LDSC}_j N_{2j} h_{22}/M\big),$$

followed by an efficiency reweighting with the mean term $\mu_j = \mathrm{ecov} + \mathrm{gcov}\cdot\ell_j$ via $w_j^{-1} \leftarrow w_j^{-1} + 2\mu_j^2$.

Value

A ‘data.frame‘ with:

- ‘ecov‘ — estimated intercept (environmental covariance)
- ‘ecov.se‘ — standard error of ‘ecov‘ (bootstrap; ‘NA‘ if no bootstrap)
- ‘gcov‘ — estimated genetic covariance (slope)
- ‘gcov.se‘ — standard error of ‘gcov‘ (bootstrap; ‘NA‘ if no bootstrap)
- ‘M‘ — number of SNPs analyzed

`ldsc.trans` *Trans-Ancestry Bivariate Linkage Disequilibrium Score Regression (LDSC)*

Description

The ‘ldsc.trans‘ function performs bivariate genetic covariance estimation using Linkage Disequilibrium Score Regression (LDSC) in a trans-ancestry setting. It estimates the genetic covariance between two traits from different ancestries, using a cross-ancestry LD score for the regression regressor and ancestry-specific LD scores for the variance weights. The function also supports block bootstrap for standard error estimation.

Usage

```
ldsc.trans(
  gwas1,
  gwas2,
  h21,
  h22,
  LDSC1,
  LDSC2,
  LDSC_Tran,
  nblock = 500,
  sampling.time = 0
)
```

Arguments

`gwas1`	A data.frame or data.table containing GWAS summary statistics for the first trait (ancestry 1). Must include: • `SNP`: SNP identifier (string) • `Zscore`: Z-score for the SNP-trait association • `N`: Sample size for the SNP
`gwas2`	A data.frame or data.table containing GWAS summary statistics for the second trait (ancestry 2), with the same column requirements as `gwas1`.
`h21`	Numeric. Heritability estimate of the first trait.
`h22`	Numeric. Heritability estimate of the second trait.
`LDSC1`	A data.frame or data.table containing ancestry-1 LD scores for variance weighting. Must include:

- SNP: SNP identifier
- LDSC: LD score value for ancestry 1

LDSC2 — A data.frame or data.table containing ancestry-2 LD scores for variance weighting. Must include:

- SNP: SNP identifier
- LDSC: LD score value for ancestry 2

LDSC_Tran — A data.frame or data.table containing cross-ancestry LD scores for the regression regressor. Must include:

- SNP: SNP identifier
- LDSC: Cross-ancestry LD score

nblock — Integer. Number of blocks used for block bootstrap standard error estimation.

sampling.time — Integer. Number of bootstrap replicates. If set to 0, only point estimates will be returned.

Value

A data.frame with the following columns:

- ecov: Estimated intercept (environmental covariance)
- ecov.se: Standard error of the intercept
- gcov: Estimated genetic covariance
- gcov.se: Standard error of the genetic covariance
- M: Number of SNPs used in the analysis

---

ldsc.univ *Univariate LD Score Regression (LDSC) for SNP-Heritability*

Description

'ldsc.univ' estimates SNP-heritability for a single trait using Linkage Disequilibrium Score Regression (LDSC). The function harmonizes the GWAS summary statistics with an LD-score table, fits a weighted least squares (WLS) regression of Z_j^2 on the LD regressor, and performs an optional block bootstrap for standard errors.

Usage

```
ldsc.univ(gwas, LDSC, nblock = 500, sampling.time = 500)
```

Arguments

gwas — A 'data.frame' or 'data.table' of GWAS summary statistics for one trait. Required columns:

- 'SNP' — SNP identifier
- 'Zscore' — Z-score of the SNP–trait association
- 'N' — per-SNP sample size
- 'A1' — effect allele
- 'A2' — reference allele

	Only ‘SNP‘, ‘Zscore‘, and ‘N‘ are used in the computation; alleles are kept for consistency checks and downstream use.
LDSC	A ‘data.frame‘ or ‘data.table‘ containing LD scores. Required columns: • ‘SNP‘ — SNP identifier • ‘LDSC‘ — LD score
nblock	Integer. Number of blocks for block bootstrap.
sampling.time	Integer. Number of bootstrap replicates. If ‘0‘, only point estimates are returned.

Details

Let M be the number of SNPs after harmonization. The regressor is $\ell_j = \mathrm{LDSC}_j\, N_j/M$, and we regress Z_j^2 on $(1, \ell_j)$ with WLS. Initial weights use a small ridge factor $(1 + 0.1\,\ell_j)^{-2}$, followed by an efficiency reweighting that plugs in the estimated heritability $\hat{h}^2$: $w_j = (1 + \ell_j\,\hat{h}^2)^{-2}$.

Value

A ‘data.frame‘ with:

- ‘intercept‘ — estimated LDSC intercept
- ‘intercept.se‘ — standard error of ‘intercept‘ (bootstrap; ‘NA‘ if no bootstrap)
- ‘h2‘ — estimated SNP-heritability (slope)
- ‘h2.se‘ — standard error of ‘h2‘ (bootstrap; ‘NA‘ if no bootstrap)
- ‘M‘ — number of SNPs analyzed

Index