

Package ‘ldscR’

July 29, 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),
CppMatrix

RoxygenNote 7.3.2

Config/testthat/edition 3

Contents

AFRLDSC	2
AMRLDSC	2
EASLDSC	3
EURLDSC	3
filter_align	4
hapmap3	5
ldsc.bicov	5
ldsc.univ	6
PCSK9	7
SASLDSC	7
Index	8

AFRLDSC*African Population LDSC Data*

Description

LDSC data for African populations. It includes LD scores calculated from the sample LD matrix and POET-estimated LD scores.

Usage

```
data(AFRLDSC)
```

Format

A data frame with 2 variables:

SNP SNP identifier.

LDSC LD score calculated from the sample LD matrix.

Source

LD scores derived from UK Biobank samples using PRScsx (<https://github.com/getian107/PRScsx>).

AMRLDSC*American Admixed Population LDSC Data*

Description

LDSC data for American admixed populations (Hispanic). It includes LD scores calculated from the sample LD matrix and POET-estimated LD scores.

Usage

```
data(AMRLDSC)
```

Format

A data frame with 2 variables:

SNP SNP identifier.

LDSC LD score calculated from the sample LD matrix.

Source

LD scores derived from UK Biobank samples using PRScsx (<https://github.com/getian107/PRScsx>).

EASLDSC*East Asian Population LDSC Data*

Description

LDSC data for East Asian populations. It includes LD scores calculated from the sample LD matrix and POET-estimated LD scores.

Usage

```
data(EASLDSC)
```

Format

A data frame with 2 variables:

SNP SNP identifier.

LDSC LD score calculated from the sample LD matrix.

Source

LD scores derived from UK Biobank samples using PRScsx (<https://github.com/getian107/PRScsx>).

EURLDSC*European Population LDSC Data*

Description

LDSC data for European populations. It includes LD scores calculated from the sample LD matrix and POET-estimated LD scores.

Usage

```
data(EURLDSC)
```

Format

A data frame with 2 variables:

SNP SNP identifier.

LDSC LD score calculated from the sample LD matrix.

Source

LD scores derived from UK Biobank samples using PRScsx (<https://github.com/getian107/PRScsx>).

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 and UKBB Genotype SNP Data***Description**

A data frame containing combined SNP data from 1000 Genomes Project Phase 3 and UK Biobank (UKBB) genotypes. It includes a total of 1,664,852 SNPs.

Usage

```
data(hapmap3)
```

Format

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

SNP SNP identifier.

A1 Effect allele.

A2 Reference allele.

Source

1000 Genomes Project Phase 3 and UK Biobank genotype data.

ldsc.bicov

*Single-Variate Linkage Disequilibrium Score Regression (LDSC)***Description**

The 'ldsc.bicov' function performs bi-variate genetic covariance using Linkage Disequilibrium Score Regression (LDSC) analysis. It is designed to estimate heritability from gwas1 summary statistics, accounting for linkage disequilibrium (LD) between SNPs. The function harmonizes gwas1 data with LD scores and applies non-linear optimization to estimate heritability.

Usage

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

Arguments

gwas1	A data.frame containing gwas summary statistics for the first trait. The data.frame should include columns for SNP identifiers, Z-scores of effect size estimates, sample sizes (N).
gwas2	A data.frame containing gwas summary statistics for the second trait. The data.frame should include columns for SNP identifiers, Z-scores of effect size estimates, sample sizes (N).
LDSC	A data.frame containing LD Score Regression (LDSC) estimates. It should include LDSC scores and other necessary metrics for the analysis.
nblock	The number of blocks for bootstrap-based standard error estimation.
sampling.time	The number of block bootstrap. If sampling.time=0, only genetic covariance estimate will be returned.

Details

The 'ldsc.univ' function is particularly useful for single-trait heritability analysis using LDSC. It requires well-formatted gwas1 summary statistics and an understanding of LDSC methodology. The function includes an initial estimation phase followed by a reweighting step for improved efficiency and accuracy.

Value

A data.frame containing genetic covariance estimates and their standard errors, along with the estimation error covariance estimate and its standard error.

ldsc.univ	<i>Single-Variate Linkage Disequilibrium Score Regression (LDSC)</i>
-----------	--

Description

The 'ldsc.uni' function performs single-variate Linkage Disequilibrium Score Regression (LDSC) analysis. It is designed to estimate heritability from GWAS summary statistics, accounting for linkage disequilibrium (LD) between SNPs. The function harmonizes GWAS data with LD scores and applies non-linear optimization to estimate heritability.

Usage

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

Arguments

gwas	A data.frame containing GWAS summary statistics for a single trait. The data.frame should include columns for SNP identifiers, Z-scores of effect size estimates, sample sizes (N), effect allele (A1), and reference allele (A2).
LDSC	A data.frame containing LD Score Regression (LDSC) estimates. It should include LDSC scores and other necessary metrics for the analysis.
nblock	The number of blocks for bootstrap-based standard error estimation.
sampling.time	The number of block bootstrap. If sampling.time=0, only genetic covariance estimate will be returned.

Details

The 'ldsc.univ' function is particularly useful for single-trait heritability analysis using LDSC. It requires well-formatted GWAS summary statistics and an understanding of LDSC methodology. The function includes an initial estimation phase followed by a reweighting step for improved efficiency and accuracy.

Value

A data.frame containing heritability estimates and their standard errors, along with the intercept and its standard error.

PCSK9

PCSK9

Description

A list with 12 eQTL summary data in the PCSK9 locus

Usage

```
data(PCSK9)
```

Format

A list with 12 data.frame:

SNP SNP identifier.

Zscore Zscore of eQTL data.

N Sample size of eQTL data.

A1 Effect allele.

A2 Alternative.

rownames(.) Gene ensembl or splicing identifier and tissue.

Source

eQTL summary data are provided by GTEx consortium (<https://www.gtexportal.org/home/>).

SASLDSC

South Asian LD Score Regression (LDSC) Data

Description

LDSC data for South Asian populations. It includes LD scores calculated from the sample LD matrix and POET-estimated LD scores.

Usage

```
data(SASLDSC)
```

Format

A data frame with 2 variables:

SNP SNP identifier.

LDSC LD score calculated from the sample LD matrix.

Source

LD scores derived from UK Biobank samples using PRScsx (<https://github.com/getian107/PRScsx>).

Index

* datasets

- AFRLDSC, [2](#)
- AMRLDSC, [2](#)
- EASLDSC, [3](#)
- EURLDSC, [3](#)
- hapmap3, [5](#)
- PCSK9, [7](#)
- SASLDSC, [7](#)

AFRLDSC, [2](#)

AMRLDSC, [2](#)

EASLDSC, [3](#)

EURLDSC, [3](#)

filter_align, [4](#)

hapmap3, [5](#)

ldsc.bicov, [5](#)

ldsc.univ, [6](#)

PCSK9, [7](#)

SASLDSC, [7](#)