Package 'ldscR'

December 26, 2023

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)

RoxygenNote 7.2.3

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

R topics documented:

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).

---

filterss   *Filter and Align GWAS Data to a Reference Panel*

Description

The `filterss` 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

```
filterss(gwas_data_list, ref_panel)
```

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.

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

```
# Assuming GWAS_List and ref_panel are already defined:
filtered_data <- filterss(GWAS_List, ref_panel)
```

---

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.

---

ldscR	*Estimate Heritability and Genetic Correlation Matrix Using LDSC estimated from sample LD matrix estimate*

Description

The ldscR function estimates heritability and genetic correlation matrices using Linkage Disequilibrium Score Regression (LDSC). It processes GWAS summary statistics, harmonizes alleles with a reference panel, and computes genetic covariance and error covariance matrices.

Usage

```
ldscR(GWAS_List, LDSC)
```

Arguments

GWAS_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).
LDSC	A data.frame containing LD Score Regression (LDSC) estimates. It should include LDSC scores and other necessary metrics for the analysis.

Details

The ldscR function is designed for advanced genetic statistics and requires a good understanding of GWAS summary statistics, LDSC methodology, and statistical genetics. Users should ensure that input data is correctly formatted and that they understand the implications of the estimates produced by the function.

Value

A list containing the following elements:

- GCovEstEstimated genetic covariance matrix.
- GCovSEStandard errors of the estimated genetic covariance matrix.
- ECovEstEstimated error covariance matrix.
- ECovSEStandard errors of the estimated error covariance matrix.

Examples

```
data(hapmap3)
data(EURLDSC)
ref_panel <- hapmap3
LDSC <- EURLDSC
results <- ldscR(GWAS_List, ref_panel, LDSC)
```

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