Package 'kdps'

July 22, 2025

Type Package

Title Kinship Decouple and Phenotype Selection (KDPS)

Version 1.0.0

Description A phenotype-aware algorithm for resolving cryptic relatedness in genetic studies. It removes related individuals based on kinship or identity-by-descent (IBD) scores while prioritizing subjects with phenotypes of interest. This approach helps maximize the retention of informative subjects, particularly for rare or valuable traits, and improves statistical power in genetic and epidemiological studies. KDPS supports both categorical and quantitative phenotypes, composite scoring, and customizable pruning strategies using a fuzziness parameter. Benchmark results show improved phenotype retention and high computational efficiency on large-scale datasets like the UK Biobank. Methods used include Manichaikul et al. (2010) <doi:10.1093/bioinformatics/btq559> for kinship estimation, Purcell et al. (2007) <doi:10.1086/519795> for IBD estimation, and Bycroft et al. (2018) <doi:10.1038/s41586-018-0579-z> for UK Biobank data reference.

URL https://github.com/UCSD-Salem-Lab/kdps

BugReports https://github.com/UCSD-Salem-Lab/kdps/issues

License MIT + file LICENSE

Encoding UTF-8

Imports data.table, dplyr, progress, tibble

Depends R (>= 4.1.0)

RoxygenNote 7.3.2

Suggests devtools, knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Wanjun Gu [aut, cre] (ORCID: <https://orcid.org/0000-0002-7342-7000>)

Maintainer Wanjun Gu <wanjun.gu@ucsf.edu>

Repository CRAN

Date/Publication 2025-07-22 10:30:27 UTC

1

Contents

---

kdps   *Kinship Decouple and Phenotype Selection (KDPS)*

Description

The 'kdps' function identifies subjects to be removed from a study based on kinship and pheno-type information. It uses kinship matrices and phenotype data to evaluate and prioritize subjects according to their phenotype scores, taking into account their relatedness. The algorithm can prior-itize subjects with high or low phenotype values and filter out subjects based on kinship thresholds and phenotype rankings. It aims to refine the study population by removing subjects that do not meet specific genetic and phenotypic criteria, thus enhancing the robustness of genetic association studies.

Usage

```
kdps(
  phenotype_file = system.file("extdata", "simple_pheno.txt", package = "kdps"),
  kinship_file = system.file("extdata", "simple_kinship.txt", package = "kdps"),
  fuzziness = 0,
  phenotype_name = "pheno2",
  prioritize_high = FALSE,
  prioritize_low = FALSE,
  phenotype_rank = c("DISEASED1", "DISEASED2", "HEALTHY"),
  fid_name = "FID",
  iid_name = "IID",
  fid1_name = "FID1",
  iid1_name = "IID1",
  fid2_name = "FID2",
  iid2_name = "IID2",
  kinship_name = "KINSHIP",
  kinship_threshold = 0.0442,
  phenotypic_naive = FALSE
)
```

Arguments

phenotype_file	A string specifying the path to the phenotype data file.
kinship_file	A string specifying the path to the kinship matrix file.
fuzziness	An integer representing the level of fuzziness allowed in removing related sub-jects, with a default of 0 (no fuzziness).
phenotype_name	The name of the phenotype column in the phenotype file.

prioritize_high

    A logical indicating whether to prioritize subjects with high phenotype values for removal.

prioritize_low   A logical indicating whether to prioritize subjects with low phenotype values for removal.

phenotype_rank   A character vector specifying the ranking of phenotypes from highest priority (first) to lowest.

fid_name   The column name for family IDs in the phenotype file.

iid_name   The column name for individual IDs in the phenotype file.

fid1_name   The column name for the first individual's family ID in the kinship file.

iid1_name   The column name for the first individual's ID in the kinship file.

fid2_name   The column name for the second individual's family ID in the kinship file.

iid2_name   The column name for the second individual's ID in the kinship file.

kinship_name   The name of the kinship score column in the kinship file.

kinship_threshold

    A numeric threshold for the kinship score, above which individuals are considered related.

phenotypic_naive

    A logical indicating whether to ignore phenotype information when resolving conflicts between related individuals.

Details

The function first processes phenotype and kinship data from the specified files, then evaluates subjects based on the provided parameters. It calculates weights for each subject based on their phenotype and uses these weights along with the kinship information to identify subjects that should be removed to minimize relatedness in the study population. The function offers flexibility in handling phenotypes through ranking and prioritization options and can adjust the stringency of relatedness filtering through the kinship threshold and fuzziness parameter.

Value

A data frame with two columns, 'FID' and 'IID', representing the family and individual IDs of subjects suggested for removal. This output can be used to refine the study population by excluding these subjects in subsequent analyses.

Examples

```
kdps(
  phenotype_file = system.file("extdata", "simple_pheno.txt", package = "kdps"),
  kinship_file = system.file("extdata", "simple_kinship.txt", package = "kdps"),
  fuzziness = 0,
  phenotype_name = "pheno2",
  prioritize_high = FALSE,
  prioritize_low = FALSE,
  phenotype_rank = c("DISEASED1", "DISEASED2", "HEALTHY"),
  fid_name = "FID",
```

```
    iid_name = "IID",
    fid1_name = "FID1",
    iid1_name = "IID1",
    fid2_name = "FID2",
    iid2_name = "IID2",
    kinship_name = "KINSHIP",
    kinship_threshold = 0.0442,
    phenotypic_naive = FALSE
)
```

Index