Package 'OAPRS'

May 25, 2023

Type Package

Title Sample overlap in Polygenic Risk Score

Version 0.1.0

Author Who wrote it

Maintainer Seokho Jeong <seokho92@snu.ac.kr>

Description OAPRS is designed to help and guide adjusting sample overlap bias in building PRS without overfitting.
OAPRS consists of four main steps 1.summary information preparation, 2.sample overlap adjustment, 3.PRS construction, and 4.validation using visual diagnostics.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Imports data.table, dplyr, ggplot2, pROC, RcppArmadillo, Rcpp (>= 1.0.9)

LinkingTo Rcpp, RcppArmadillo

Archs x64

R topics documented:

Check_Sums *Format Summary*

Description

Reformat GWAS summary statistics file before overlap_adjustment

Usage

```
Check_Sums(
  input_file,
  Genome_Build,
  Pop,
 cols = c(BETA = "BETA", SE = "SE", Z = NULL, MAF = "MAF", Pval = "P", CHR = "CHR", POS
   = "POS", REF = "Allele1", ALT = "Allele2", SNP = "SNPID", n = "N", n_eff = NULL,
    n_case = NULL, n_ctrl = NULL),
  Spcf_n = NULL,
  Spcf_n_eff = NULL,
  Spcf_n_case = NULL,
  Spcf_n_ctrl = NULL,
  phenotype = "binary",
  allele_flip = TRUE,
  filter_by_hapmap3 = TRUE,
  fill_missing = TRUE,
  hapmap3_only = TRUE,
  minimum.P = 9.88131291682493e-324,
  save_path = NULL
)
```

Arguments

input_file	Path of Summary statistics to be formatted
Genome_Build	Genome build. hg37 or hg38
Pop	Population group (EUR or EAS)
cols	Column names of Summary statistics
Spcf_n	Specify sample size if not stated in column
Spcf_n_eff	Specify effective sample size if not stated in column
Spcf_n_case	Size of overlapped samples if not stated in column
Spcf_n_ctrl	Effective Sample size of summary to be adjusted
phenotype	Effective Sample Size of the overlapped
allele_flip	Sample Size of the all cases
filter_by_hapmap3	Sample Size of the all controls
fill_missing	Imputation missing columns with given Summary statistics
hapmap3_only	Scope variants into hapmap3 markers only
minimum.P	Minimum pvalue supported : default=9.88e-324
save_path	Path for saving formatted summary statistics (optional)

Value

Formatted Summary Statistics

Examples

```
cs = Check_Sums(system.file('extdata/example/consortium.ss',package = 'OAPRS'),
Genome_Build = "hg37", Pop = "eas",
cols = c(BETA="beta",Pval="pval",CHR="chrom",POS="pos",REF="ref",ALT="alt",SNP="rsids"),
Spcf_n=249625,Spcf_n_case = 50466, Spcf_n_ctrl = 199159)
```

---

diagnostic_plt *Score Evaluation*

Description

After applying polygenic risk score estimation tools, score evaluation for generating diagnostic plots

Usage

```
diagnostic_plt(
  scores,
  title,
  Output_Plot_path = NULL,
  keep.ind = NULL,
  ref.ind = NULL,
  ref_name = "",
  pheno_col = NULL,
  covar_cols = NULL,
  method_names = NULL
)
```

Arguments

scores	Evaluated personalized scores
title	Plot title
Output_Plot_path	Output Path for generated diagnostic plot
keep.ind	Sample indices to be evaluated.
ref.ind	Sample indices for Reference group
ref_name	Reference group name
pheno_col	Column name for phenotype (Optional)
covar_cols	Specify PRS method platform (prscs, ldpred2)
method_names	New labels for methods (Optional)

Value

ggplot2 object of visual diagnostic

`exclude_overlap` *Overlap Adjustment*

Description

Overlap Adjustment from a GWAS summary with summary of Overlapped samples

Usage

```
exclude_overlap(file_all, file_ov, save_path = NULL, phenotype, dropna = T)
```

Arguments

`file_all`	object/Path of Summary statistics to be adjusted
`file_ov`	object/Path of Summary statistics of Overlapped samples
`save_path`	Output path of adjusted summary statistics
`phenotype`	Phenotype if binary : "binary", continuous : "continuous"
`dropna`	Drop markers that has standard error NA's?

Value

Adjusted summary

Examples

```
print('adj_ss = exclude_overlap(cs,ts,"adj.txt",phenotype="binary")')
```

`Marker_select_ld` *Variant filtering for generating diagnostic plots*

Description

Variant filtering by Independent LD blocks

Usage

```
Marker_select_ld(
  ss,
  pt = 1e-04,
  probs = c(0, 0.01, 0.05, 0.1, 0.2, 0.5),
  Genome_Build,
  Pop,
  methods = c("all", "IVW", "RZ")
)
```

Arguments

ss	Summary statistic object or path for formatted summary statistics
pt	Threshold for pruning
probs	Multiple P-value cutoffs for diagnostic plot
Genome_Build	Genome build. hg37 or hg38
Pop	Population group (EUR or EAS)
methods	Specify sample size if not stated in column

Value

Dataframe of logical filter of markers, threshold, methods

PtoZ *P-value to z-score*

Description

P-value to z-score transformation

Usage

```
PtoZ(P, bet)
```

Arguments

P	P-values to convert
bet	Signed vector of coefficients

---

score_eval *Score Evaluation*

Description

After applying polygenic risk score estimation tools, score evaluation for generating diagnostic plots

Usage

```
score_eval(
  prs_res_paths,
  lds,
  target_path,
  output_score_path = NULL,
  platform = "prscs",
  bim_path = NULL,
  fam_path = NULL,
  pheno_path = NULL,
  ID_col,
  pheno_col,
  cl = NULL
)
```

Arguments

prs_res_paths	directory path of PRS result(e.g. prscs)
lds	Variant Filter results
target_path	Plink binary file header of target/validation genotypes
output_score_path	Path for storing output scores.(Optional)
platform	Specify PRS method platform (prscs, ldpred2)
bim_path	Optional target bim file path
fam_path	Optional target fam file path
pheno_path	Covariate file of target/validation samples for additional
ID_col	Column name for Sample ID in pheno_file
pheno_col	Column name for phenotypes. If covariate file is not given. phenotype column from fam will be taken.
cl	Cluster given from parallel

Value

Formatted Summary Statistics

Index