Package 'MetaSTAAR'

November 18, 2023

Type Package

Title Meta-Analysis of STAAR (MetaSTAAR) Procedure for Dynamic Incorporation of Multiple Functional Annotations in Whole Genome Sequencing Studies

Version 0.9.6.3

Date 2023-11-18

Author Xihao Li [aut, cre], Zilin Li [aut, cre], Corbin Quick [aut]

Maintainer Xihao Li <xihaoli@unc.edu>, Zilin Li <lizl@nenu.edu.edu>

Description
An R package for performing MetaSTAAR procedure in whole genome sequencing studies.

License GPL-3

Copyright See COPYRIGHTS for details.

Imports Rcpp, STAAR, Matrix, dplyr, methods, expm, MASS

Encoding UTF-8

LazyData true

Depends R (>= 3.2.0)

LinkingTo Rcpp, RcppArmadillo, STAAR

RoxygenNote 7.2.3

Suggests knitr, rmarkdown

VignetteBuilder knitr

R topics documented:

`Burden_Effect_Size_meta` *Effect size and standard error estimates of meta-analysis of burden test for a given variant-set*

Description

The `Burden_Effect_Size_meta` function takes in the object from the merged summary statistics and covariance files of each participating study and functional annotation data to calculate the effect size and standard error estimates of meta-analysis of burden test for a given variant-set.

Usage

```
Burden_Effect_Size_meta(obj_MetaSTAAR_merge, rv_num_cutoff = 2)
```

Arguments

`obj_MetaSTAAR_merge`	an object from merging the summary statistics and covariance files from each participating study, which is the output from `MetaSTAAR_merge`.
`rv_num_cutoff`	the cutoff of minimum number of variants of analyzing a given variant-set (default = 2).

Value

a list with the following members:

`num_variant`: the number of variants with combined minor allele frequency > 0 and less than `rare_maf_cutoff` in the given variant-set that are used for performing the variant-set test using MetaSTAAR.

`cMAC`: the combined cumulative minor allele count of variants with combined minor allele frequency > 0 and less than `rare_maf_cutoff` in the given variant-set.

`Burden_Score_Stat`: the score statistic of Burden-MS(1,1) for the given variant-set.

`Burden_SE_Score`: the standard error of `Burden_Score_Stat` for the given variant-set.

`Burden_pvalue`: the Burden-MS(1,1) p-value for the given variant-set.

`Burden_Est`: the effect size estimate of Burden-MS(1,1) for the given variant-set.

`Burden_SE_Est`: the standard error estimate of `Burden_Est` for the given variant-set.

References

Li, X., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nature Genetics*, *55*(1), 154-164. (pub)

Indiv_Score_Test_Region_meta

Meta-analysis of score test for individual variants in a given variant-set

Description

The Indiv_Score_Test_Region_meta function takes in the object from the merged summary statistics and covariance files of each participating study to analyze the associations between a quantitative/dichotomous phenotype and all individual variants in a given variant-set by using the meta-analysis of score test.

Usage

```
Indiv_Score_Test_Region_meta(obj_MetaSTAAR_merge, rv_num_cutoff = 2)
```

Arguments

obj_MetaSTAAR_merge	an object from merging the summary statistics and covariance files from each participating study, which is the output from MetaSTAAR_merge.
rv_num_cutoff	the cutoff of minimum number of variants of analyzing a given variant-set (default = 2).

Value

a data frame with p rows corresponding to the p genetic variants in the given variant-set and three columns: Score (the score test statistic), SE (the standard error associated with the score test statistic), and pvalue (the score test p-value). If a variant in the given variant-set has standard error equal to 0, the p-value will be set as 1.

Indiv_Score_Test_Region_meta_cond

Meta-analysis of conditional score test for individual variants in a given variant-set

Description

The Indiv_Score_Test_Region_meta_cond function takes in the object from the merged conditional summary statistics and covariance files of each participating study to analyze the conditional associations between a quantitative/dichotomous phenotype and all individual variants in a given variant-set by using the meta-analysis of score test, adjusting for a given list of variants.

Usage

```
Indiv_Score_Test_Region_meta_cond(obj_MetaSTAAR_merge_cond, rv_num_cutoff = 2)
```

Arguments

`obj_MetaSTAAR_merge_cond`
: an object from merging the conditional summary statistics and covariance files from each participating study, adjusting for a given list of variants, which is the output from `MetaSTAAR_merge_cond`.

`rv_num_cutoff`
: the cutoff of minimum number of variants of analyzing a given variant-set (default = 2).

Value

a data frame with p rows corresponding to the p genetic variants in the given variant-set and three columns: `Score_cond` (the conditional score test statistic), `SE_cond` (the standard error associated with the conditional score test statistic), and `pvalue_cond` (the conditional score test p-value). If a variant in the given variant-set has standard error equal to 0, the p-value will be set as 1.

`MetaSTAAR` *Meta-analysis of STAAR (MetaSTAAR) procedure using omnibus test*

Description

The `MetaSTAAR` function takes in the object from the merged summary statistics and covariance files of each individual study and functional annotation data to analyze the association between a quantitative/dichotomous phenotype and a variant-set by using the meta-analysis of STAAR (MetaSTAAR) procedure. For each variant-set, the MetaSTAAR-O p-value is a p-value from an omnibus test that aggregated SKAT-MS(1,25), SKAT-MS(1,1), Burden-MS(1,25), Burden-MS(1,1), ACAT-V-MS(1,25), and ACAT-V-MS(1,1) together with p-values of each test weighted by each annotation using Cauchy method.

Usage

```
MetaSTAAR(obj_MetaSTAAR_merge, annotation_phred = NULL, rv_num_cutoff = 2)
```

Arguments

`obj_MetaSTAAR_merge`
: an object from merging the summary statistics and covariance files from each participating study, which is the output from `MetaSTAAR_merge`.

`annotation_phred`
: a data frame or matrix of functional annotation data of dimension p*q (or a vector of a single annotation score with length p), where p is the number of genetic variants in the variant-set. Continuous scores should be given in PHRED score scale, where the PHRED score of j-th variant is defined to be -10*log10(rank(-score_j)/total) across the genome. (Binary) categorical scores should be taking values 0 or 1, where 1 is functional and 0 is non-functional. If not provided, MetaSTAAR will perform the SKAT-MS(1,25), SKAT-MS(1,1), Burden-MS(1,25), Burden-MS(1,1), ACAT-V-MS(1,25), ACAT-V-MS(1,1) and ACAT-O-MS tests (default = NULL).

`rv_num_cutoff`
: the cutoff of minimum number of variants of analyzing a given variant-set (default = 2).

Value

a list with the following members:

`num_variant`: the number of variants with combined minor allele frequency > 0 and less than `rare_maf_cutoff` in the given variant-set that are used for performing the variant-set test using MetaSTAAR.

`cMAC`: the combined cumulative minor allele count of variants with combined minor allele frequency > 0 and less than `rare_maf_cutoff` in the given variant-set.

`results_MetaSTAAR_O`: the MetaSTAAR-O p-value that aggregated SKAT-MS(1,25), SKAT-MS(1,1), Burden-MS(1,25), Burden-MS(1,1), ACAT-V-MS(1,25), and ACAT-V-MS(1,1) together with p-values of each test weighted by each annotation using Cauchy method.

`results_ACAT_O_MS`: the ACAT-O-MS p-value that aggregated SKAT-MS(1,25), SKAT-MS(1,1), Burden-MS(1,25), Burden-MS(1,1), ACAT-V-MS(1,25), and ACAT-V-MS(1,1) using Cauchy method.

`results_MetaSTAAR_S_1_25`: a vector of MetaSTAAR-S(1,25) p-values, including SKAT-MS(1,25) p-value weighted by MAF, the SKAT-MS(1,25) p-values weighted by each annotation, and a MetaSTAAR-S(1,25) p-value by aggregating these p-values using Cauchy method.

`results_MetaSTAAR_S_1_1`: a vector of MetaSTAAR-S(1,1) p-values, including SKAT-MS(1,1) p-value weighted by MAF, the SKAT-MS(1,1) p-values weighted by each annotation, and a MetaSTAAR-S(1,1) p-value by aggregating these p-values using Cauchy method.

`results_MetaSTAAR_B_1_25`: a vector of MetaSTAAR-B(1,25) p-values, including Burden-MS(1,25) p-value weighted by MAF, the Burden-MS(1,25) p-values weighted by each annotation, and a MetaSTAAR-B(1,25) p-value by aggregating these p-values using Cauchy method.

`results_MetaSTAAR_B_1_1`: a vector of MetaSTAAR-B(1,1) p-values, including Burden-MS(1,1) p-value weighted by MAF, the Burden-MS(1,1) p-values weighted by each annotation, and a MetaSTAAR-B(1,1) p-value by aggregating these p-values using Cauchy method.

`results_MetaSTAAR_A_1_25`: a vector of MetaSTAAR-A(1,25) p-values, including ACAT-V-MS(1,25) p-value weighted by MAF, the ACAT-V-MS(1,25) p-values weighted by each annotation, and a MetaSTAAR-A(1,25) p-value by aggregating these p-values using Cauchy method.

`results_MetaSTAAR_A_1_1`: a vector of MetaSTAAR-A(1,1) p-values, including ACAT-V-MS(1,1) p-value weighted by MAF, the ACAT-V-MS(1,1) p-values weighted by each annotation, and a MetaSTAAR-A(1,1) p-value by aggregating these p-values using Cauchy method.

References

Li, X., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nature Genetics*, *55*(1), 154-164. (pub)

Li, X., Li, Z., et al. (2020). Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nature Genetics 52*(9), 969-983. (pub)

Liu, Y., et al. (2019). Acat: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. *The American Journal of Human Genetics 104*(3), 410-421. (pub)

MetaSTAAR_cond *Meta-analysis of STAAR (MetaSTAAR) procedure for conditional analysis using omnibus test*

Description

The MetaSTAAR_cond function takes in the object from the merged conditional summary statistics and covariance files of each participating study and functional annotation data to analyze the conditional association between a quantitative/dichotomous phenotype and a variant-set by using the meta-analysis of STAAR (MetaSTAAR) procedure, adjusting for a given list of variants. For each variant-set, the MetaSTAAR-O p-value is a p-value from an omnibus test that aggregated conditional SKAT-MS(1,25), SKAT-MS(1,1), Burden-MS(1,25), Burden-MS(1,1), ACAT-V-MS(1,25), and ACAT-V-MS(1,1) together with conditional p-values of each test weighted by each annotation using Cauchy method.

Usage

```
MetaSTAAR_cond(
  obj_MetaSTAAR_merge_cond,
  annotation_phred = NULL,
  rv_num_cutoff = 2
)
```

Arguments

obj_MetaSTAAR_merge_cond
: an object from merging the conditional summary statistics and covariance files from each participating study, adjusting for a given list of variants, which is the output from MetaSTAAR_merge_cond.

annotation_phred
: a data frame or matrix of functional annotation data of dimension p*q (or a vector of a single annotation score with length p), where p is the number of genetic variants in the variant-set. Continuous scores should be given in PHRED score scale, where the PHRED score of j-th variant is defined to be -10*log10(rank(-score_j)/total) across the genome. (Binary) categorical scores should be taking values 0 or 1, where 1 is functional and 0 is non-functional. If not provided, MetaSTAAR will perform the SKAT-MS(1,25), SKAT-MS(1,1), Burden-MS(1,25), Burden-MS(1,1), ACAT-V-MS(1,25), ACAT-V-MS(1,1) and ACAT-O-MS tests (default = NULL).

rv_num_cutoff
: the cutoff of minimum number of variants of analyzing a given variant-set (default = 2).

Value

a list with the following members:

num_variant: the number of variants with combined minor allele frequency > 0 and less than rare_maf_cutoff in the given variant-set that are used for performing the variant-set test using MetaSTAAR.

cMAC: the combined cumulative minor allele count of variants with combined minor allele frequency > 0 and less than rare_maf_cutoff in the given variant-set.

`results_MetaSTAAR_O_cond`: the conditional MetaSTAAR-O p-value that aggregated conditional SKAT-MS(1,25), SKAT-MS(1,1), Burden-MS(1,25), Burden-MS(1,1), ACAT-V-MS(1,25), and ACAT-V-MS(1,1) together with conditional p-values of each test weighted by each annotation using Cauchy method.

`results_ACAT_O_MS_cond`: the conditional ACAT-O-MS p-value that aggregated conditional SKAT-MS(1,25), SKAT-MS(1,1), Burden-MS(1,25), Burden-MS(1,1), ACAT-V-MS(1,25), and ACAT-V-MS(1,1) using Cauchy method.

`results_MetaSTAAR_S_1_25_cond`: a vector of conditional MetaSTAAR-S(1,25) p-values, including conditional SKAT-MS(1,25) p-value weighted by MAF, the conditional SKAT-MS(1,25) p-values weighted by each annotation, and a conditional MetaSTAAR-S(1,25) p-value by aggregating these p-values using Cauchy method.

`results_MetaSTAAR_S_1_1_cond`: a vector of conditional MetaSTAAR-S(1,1) p-values, including conditional SKAT-MS(1,1) p-value weighted by MAF, the conditional SKAT-MS(1,1) p-values weighted by each annotation, and a conditional MetaSTAAR-S(1,1) p-value by aggregating these p-values using Cauchy method.

`results_MetaSTAAR_B_1_25_cond`: a vector of conditional MetaSTAAR-B(1,25) p-values, including conditional Burden-MS(1,25) p-value weighted by MAF, the conditional Burden-MS(1,25) p-values weighted by each annotation, and a conditional MetaSTAAR-B(1,25) p-value by aggregating these p-values using Cauchy method.

`results_MetaSTAAR_B_1_1_cond`: a vector of conditional MetaSTAAR-B(1,1) p-values, including conditional Burden-MS(1,1) p-value weighted by MAF, the conditional Burden-MS(1,1) p-values weighted by each annotation, and a conditional MetaSTAAR-B(1,1) p-value by aggregating these p-values using Cauchy method.

`results_MetaSTAAR_A_1_25_cond`: a vector of conditional MetaSTAAR-A(1,25) p-values, including conditional ACAT-V-MS(1,25) p-value weighted by MAF, the conditional ACAT-V-MS(1,25) p-values weighted by each annotation, and a conditional MetaSTAAR-A(1,25) p-value by aggregating these p-values using Cauchy method.

`results_MetaSTAAR_A_1_1_cond`: a vector of conditional MetaSTAAR-A(1,1) p-values, including conditional ACAT-V-MS(1,1) p-value weighted by MAF, the conditional ACAT-V-MS(1,1) p-values weighted by each annotation, and a conditional MetaSTAAR-A(1,1) p-value by aggregating these p-values using Cauchy method.

References

Li, X., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nature Genetics*, *55*(1), 154-164. (pub)

Li, X., Li, Z., et al. (2020). Dynamic incorporation of multiple in silico functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nature Genetics 52*(9), 969-983. (pub)

Liu, Y., et al. (2019). Acat: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. *The American Journal of Human Genetics 104*(3), 410-421. (pub)

`MetaSTAAR_individual_analysis`

Meta-analysis of individual variants using `MetaSTAAR`

Description

The MetaSTAAR_individual_analysis function takes in the summary statistics file from each participating study (the output from MetaSTAAR_worker_sumstat) to analyze the associations between a quantitative/dichotomous phenotype and all individual variants in the merged variant list by using the meta-analysis of score test.

Usage

```
MetaSTAAR_individual_analysis(
  chr,
  start.loc,
  end.loc,
  study.names,
  sample.sizes,
  sumstat.dir,
  common_mac_cutoff,
  trait,
  segment.size = 5e+05,
  check_qc_label = FALSE
)
```

Arguments

chr	a numeric value indicating the chromosome of the genetic region of interest.
start.loc	a numeric value indicating the starting location (position) of the genetic region of interest.
end.loc	a numeric value indicating the ending location (position) of the genetic region of interest.
study.names	a character vector containing the name of each participating study in the meta-analysis.
sample.sizes	a numeric vector with the length of study.names indicating the sample size of each study.
sumstat.dir	a character vector containing the directories of the study-specific summary statistics file folders.
common_mac_cutoff	the cutoff of minimum combined minor allele count (inclusive) in defining "common" variants.
trait	a character value indicating the underlying trait of interest for the meta-analysis.
segment.size	a numeric value indicating the length of each segment of which the summary statistics and sparse weighted covariance files are stored. Note that the input value should be aligned with the input values of MetaSTAAR_worker_cov (default = 5e+05).
check_qc_label	a logical value indicating whether variants need to be dropped according to qc_label specified in MetaSTAAR_worker_sumstat and MetaSTAAR_worker_cov. If check_qc_label is FALSE, it is assumed that no variant will be dropped (default = FALSE).

Value

a data frame with p rows corresponding to the p genetic variants in the merged variant list and the following columns: chromosome (chr), position (pos), reference allele (ref), alternative allele (alt), combined alternative allele count (alt_AC), combined minor allele count (MAC), combined minor allele frequency (MAF), combined sample size (N), the score test p-value (p), the log score test p-value (logp), the score test statistic (Score), the standard error associated with the score test statistic (Score_SE), the estimated effect size of the minor allele (Est), the standard error associated with the estimated effect size (Est_se). If a variant in the merged variant list has standard error equal to 0, the p-value will be set as 1.

References

Li, X., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nature Genetics*, *55*(1), 154-164. (pub)

`MetaSTAAR_merge` *The preliminary data manipulation step for* `MetaSTAAR`

Description

The `MetaSTAAR_merge` function takes in the summary statistics file and the sparse weighted covariance file (the output from `MetaSTAAR_worker_sumstat` and `MetaSTAAR_worker_cov`) from each participating study and performs the preliminary data manipulation step by merging them into a single unified summary statistics file and a covariance file, respectively.

Usage

```
MetaSTAAR_merge(
  chr,
  start.loc,
  end.loc,
  study.names,
  sample.sizes,
  sumstat.dir,
  cov.dir,
  rare_maf_cutoff = 0.01,
  cov_maf_cutoff,
  trait,
  segment.size = 5e+05,
  check_qc_label = FALSE
)
```

Arguments

`chr`	a numeric value indicating the chromosome of the genetic region of interest.
`start.loc`	a numeric value indicating the starting location (position) of the genetic region of interest.
`end.loc`	a numeric value indicating the ending location (position) of the genetic region of interest.

study.names	a character vector containing the name of each participating study in the meta-analysis.
sample.sizes	a numeric vector with the length of study.names indicating the sample size of each study.
sumstat.dir	a character vector containing the directories of the study-specific summary statistics file folders.
cov.dir	a character vector containing the directories of the study-specific sparse weighted covariance file folders.
rare_maf_cutoff	the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
cov_maf_cutoff	a numeric vector with the length of study.names indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.
trait	a character value indicating the underlying trait of interest for the meta-analysis.
segment.size	a numeric value indicating the length of each segment of which the summary statistics and sparse weighted covariance files are stored. Note that the input value should be aligned with the input values of MetaSTAAR_worker_cov (default = 5e+05).
check_qc_label	a logical value indicating whether variants need to be dropped according to qc_label specified in MetaSTAAR_worker_sumstat and MetaSTAAR_worker_cov. If check_qc_label is FALSE, it is assumed that no variant will be dropped (default = FALSE).

Value

a list with the following members:

info: the merged data frame of all variants in the genetic region of interest whose combined minor allele frequency is below rare_maf_cutoff, including the following information (listed in the same order as U and the rows/columns of cov): chromosome (chr), position (pos), reference allele (ref), alternative allele (alt), combined minor allele count (MAC), and combined minor allele frequency (MAF).

U the merged score statistics vector of all variants in the genetic region of interest whose combined minor allele frequency is below rare_maf_cutoff.

cov the merged covariance matrix of all variants in the genetic region of interest whose combined minor allele frequency is below rare_maf_cutoff.

References

Li, X., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nature Genetics*, *55*(1), 154-164. (pub)

MetaSTAAR_merge_cond *The preliminary data manipulation step for* MetaSTAAR_cond

Description

The MetaSTAAR_merge_cond function takes in the summary statistics file and the sparse weighted covariance file (the output from MetaSTAAR_worker_sumstat and MetaSTAAR_worker_cov) as well as the covariance file for conditional analysis (the output from MetaSTAAR_worker_cov_cond) from each participating study and performs the preliminary data manipulation step by merging them into a single unified conditional summary statistics file and a conditional covariance file, respectively, adjusting for a given list of variants.

Usage

```
MetaSTAAR_merge_cond(
  chr,
  start.loc,
  end.loc,
  study.names,
  sample.sizes,
  sumstat.dir,
  cov.dir,
  covcond.dir,
  rare_maf_cutoff = 0.01,
  cov_maf_cutoff,
  trait,
  region,
  segment.size = 5e+05,
  effect.cond = c("homogeneous", "heterogeneous"),
  check_qc_label = FALSE
)
```

Arguments

chr	a numeric value indicating the chromosome of the genetic region of interest.
start.loc	a numeric value indicating the starting location (position) of the genetic region of interest.
end.loc	a numeric value indicating the ending location (position) of the genetic region of interest.
study.names	a character vector containing the name of each participating study in the meta-analysis.
sample.sizes	a numeric vector with the length of study.names indicating the sample size of each study.
sumstat.dir	a character vector containing the directories of the study-specific summary statistics file folders.
cov.dir	a character vector containing the directories of the study-specific sparse weighted covariance file folders.
covcond.dir	a character vector containing the directories of the study-specific covariance file folders for conditional analysis.

`rare_maf_cutoff`
: the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).

`cov_maf_cutoff`
: a numeric vector with the length of `study.names` indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.

`trait`
: a character value indicating the underlying trait of interest for the meta-analysis.

`region`
: a character value indicating the underlying region of a given list of variants adjusted for conditional analysis of the meta-analysis.

`segment.size`
: a numeric value indicating the length of each segment of which the summary statistics and sparse weighted covariance files are stored. Note that the input value should be aligned with the input values of `MetaSTAAR_worker_cov` (default = 5e+05).

`effect.cond`
: a character value indicating the effects of variants to be adjusted for in conditional analysis are "homogeneous" or "heterogeneous" (default = "homogeneous").

`check_qc_label`
: a logical value indicating whether variants need to be dropped according to `qc_label` specified in `MetaSTAAR_worker_sumstat` and `MetaSTAAR_worker_cov`. If `check_qc_label` is FALSE, it is assumed that no variant will be dropped (default = FALSE).

Value

a list with the following members:

`info`: the merged data frame of all variants in the genetic region of interest whose combined minor allele frequency is below `rare_maf_cutoff`, including the following information (listed in the same order as `U` and the rows/columns of `cov`): chromosome (chr), position (pos), reference allele (ref), alternative allele (alt), combined minor allele count (MAC), and combined minor allele frequency (MAF).

`U_cond`: the merged conditional score statistics vector of all variants in the genetic region of interest whose combined minor allele frequency is below `rare_maf_cutoff`, adjusting for a given list of variants.

`cov_cond`: the merged conditional covariance matrix of all variants in the genetic region of interest whose combined minor allele frequency is below `rare_maf_cutoff`, adjusting for a given list of variants.

References

Li, X., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nature Genetics*, *55*(1), 154-164. (pub)

`MetaSTAAR_merge_varlist` *The preliminary data manipulation step for* `MetaSTAAR` *given a variant list*

Description

The MetaSTAAR_merge_varlist function takes in the summary statistics file and the sparse weighted covariance file (the output from MetaSTAAR_worker_sumstat and MetaSTAAR_worker_cov) from each participating study and performs the preliminary data manipulation step by merging them into a single unified summary statistics file and a covariance file, respectively.

Usage

```
MetaSTAAR_merge_varlist(
  chr,
  variant_pos,
  study.names,
  sample.sizes,
  sumstat.dir,
  cov.dir,
  rare_maf_cutoff = 0.01,
  cov_maf_cutoff,
  trait,
  segment.size = 5e+05,
  check_qc_label = FALSE
)
```

Arguments

chr	a numeric value indicating the chromosome of the genetic region of interest.
variant_pos	a numeric vector indicating all possible positions of the variants to be included in the variant-set.
study.names	a character vector containing the name of each participating study in the meta-analysis.
sample.sizes	a numeric vector with the length of study.names indicating the sample size of each study.
sumstat.dir	a character vector containing the directories of the study-specific summary statistics file folders.
cov.dir	a character vector containing the directories of the study-specific sparse weighted covariance file folders.
rare_maf_cutoff	the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
cov_maf_cutoff	a numeric vector with the length of study.names indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.
trait	a character value indicating the underlying trait of interest for the meta-analysis.
segment.size	a numeric value indicating the length of each segment of which the summary statistics and sparse weighted covariance files are stored. Note that the input value should be aligned with the input values of MetaSTAAR_worker_cov (default = 5e+05).
check_qc_label	a logical value indicating whether variants need to be dropped according to qc_label specified in MetaSTAAR_worker_sumstat and MetaSTAAR_worker_cov. If check_qc_label is FALSE, it is assumed that no variant will be dropped (default = FALSE).

Value

a list with the following members:

`info`: the merged data frame of all variants in the given variant position list of interest whose combined minor allele frequency is below `rare_maf_cutoff`, including the following information (listed in the same order as `U` and the rows/columns of `cov`): chromosome (chr), position (pos), reference allele (ref), alternative allele (alt), combined minor allele count (MAC), and combined minor allele frequency (MAF).

`U` the merged score statistics vector of all variants in the given variant position list of interest whose combined minor allele frequency is below `rare_maf_cutoff`.

`cov` the merged covariance matrix of all variants in the given variant position list of interest whose combined minor allele frequency is below `rare_maf_cutoff`.

References

Li, X., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nature Genetics*, *55*(1), 154-164. (pub)

`MetaSTAAR_merge_varlist_cond` *The preliminary data manipulation step for* `MetaSTAAR_cond` *given a variant list*

Description

The `MetaSTAAR_merge_varlist_cond` function takes in the summary statistics file and the sparse weighted of the covariance file (the output from `MetaSTAAR_worker_sumstat` and `MetaSTAAR_worker_cov`) as well as the covariance file for conditional analysis (the output from `MetaSTAAR_worker_cov_cond`) from each participating study and performs the preliminary data manipulation step by merging them into a single unified conditional summary statistics file and a conditional covariance file, respectively, adjusting for a given list of variants.

Usage

```
MetaSTAAR_merge_varlist_cond(
  chr,
  variant_pos,
  study.names,
  sample.sizes,
  sumstat.dir,
  cov.dir,
  covcond.dir,
  rare_maf_cutoff = 0.01,
  cov_maf_cutoff,
  trait,
  region,
  segment.size = 5e+05,
  effect.cond = c("homogeneous", "heterogeneous"),
  check_qc_label = FALSE
)
```

Arguments

chr	a numeric value indicating the chromosome of the genetic region of interest.
variant_pos	a numeric vector indicating all possible positions of the variants to be included in the variant-set.
study.names	a character vector containing the name of each participating study in the meta-analysis.
sample.sizes	a numeric vector with the length of study.names indicating the sample size of each study.
sumstat.dir	a character vector containing the directories of the study-specific summary statistics file folders.
cov.dir	a character vector containing the directories of the study-specific sparse weighted covariance file folders.
covcond.dir	a character vector containing the directories of the study-specific covariance file folders for conditional analysis.
rare_maf_cutoff	the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
cov_maf_cutoff	a numeric vector with the length of study.names indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.
trait	a character value indicating the underlying trait of interest for the meta-analysis.
region	a character value indicating the underlying region of a given list of variants adjusted for conditional analysis of the meta-analysis.
segment.size	a numeric value indicating the length of each segment of which the summary statistics and sparse weighted covariance files are stored. Note that the input value should be aligned with the input values of MetaSTAAR_worker_cov (default = 5e+05).
effect.cond	a character value indicating the effects of variants to be adjusted for in conditional analysis are "homogeneous" or "heterogeneous" (default = "homogeneous").
check_qc_label	a logical value indicating whether variants need to be dropped according to qc_label specified in MetaSTAAR_worker_sumstat and MetaSTAAR_worker_cov. If check_qc_label is FALSE, it is assumed that no variant will be dropped (default = FALSE).

Value

a list with the following members:

info: the merged data frame of all variants in the given variant position list of interest whose combined minor allele frequency is below rare_maf_cutoff, including the following information (listed in the same order as U and the rows/columns of cov): chromosome (chr), position (pos), reference allele (ref), alternative allele (alt), combined minor allele count (MAC), and combined minor allele frequency (MAF).

U_cond: the merged conditional score statistics vector of all variants in the given variant position list of interest whose combined minor allele frequency is below rare_maf_cutoff, adjusting for a given list of variants.

cov_cond: the merged conditional covariance matrix of all variants in the given variant position list of interest whose combined minor allele frequency is below rare_maf_cutoff, adjusting for a given list of variants.

References

Li, X., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nature Genetics*, *55*(1), 154-164. (pub)

`MetaSTAAR_worker_cov` *Generating sparse weighted covariance file using MetaSTAARWorker*

Description

The `MetaSTAAR_worker_cov` function takes in genotype, the object from fitting the null model, and variant position to generate the sparse weighted covariance file for the given variant-set as a rectangle format.

Usage

```
MetaSTAAR_worker_cov(
  genotype,
  obj_nullmodel,
  cov_maf_cutoff,
  variant_pos,
  region_midpos,
  qc_label = NULL,
  segment.size = 5e+05,
  signif.digits = 3
)
```

Arguments

`genotype`	an n*p genotype matrix (dosage matrix) of the target sequence, where n is the sample size and p is the number of genetic variants. If the input genotype matrix is sparse (e.g. `dgCMatrix` format), it is assumed that it has been flipped to represent minor allele coding.
`obj_nullmodel`	an object from fitting the null model, which is the output from either `fit_null_glm` function for unrelated samples or `fit_null_glmmkin` function for related samples in the `STAAR` package.
`cov_maf_cutoff`	a numeric value indicating the maximum minor allele frequency cutoff under which the sparse weighted covariance file between variants is stored.
`variant_pos`	a numeric vector of length p (listed in the same order as the columns of `genotype`) indicating the position of the variants in the variant-set.
`region_midpos`	a numeric value indicating the middle position of variant-set by which the shorter edge of the rectangle is defined.
`qc_label`	a vector of quality control status for each variant in `variant_pos`, where a pass variant is labeled as "PASS". If `qc_label` is NULL, it is assumed that all variants are pass variants in the study (default = NULL).
`segment.size`	a numeric value indicating the length of each segment of which the sparse weighted covariance file is stored (default = 5e+05).
`signif.digits`	an integer indicating the number of significant digits to be used for storing the sparse weighted covariance file. If `signif.digits` is NULL, it is assumed that no rounding will be performed (default = 3).

Value

GTSinvG_rare: the sparse matrix of all variants in the variant-set whose minor allele frequency is below cov_maf_cutoff (the sparse weighted covariance file), stored as a rectangle format.

References

Li, X., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nature Genetics*, *55*(1), 154-164. (pub)

MetaSTAAR_worker_cov_cond

Generating covariance file for conditional analysis using MetaSTAAR-Worker

---

Description

The MetaSTAAR_worker_cov_cond function takes in genotype, the genotype of variants to be adjusted for in conditional analysis, the object from fitting the null model, variant information and adjusted variant information (unique identifier) to generate the conditional covariance file for the given variant-set, adjusting for a given list of variants.

Usage

```
MetaSTAAR_worker_cov_cond(
  genotype,
  genotype_adj,
  obj_nullmodel,
  variant_info,
  variant_adj_info
)
```

Arguments

genotype	an n*p genotype matrix (dosage matrix) of the target sequence, where n is the sample size and p is the number of genetic variants. If the input genotype matrix is sparse (e.g. dgCMatrix format), it is assumed that it has been flipped to represent minor allele coding.
genotype_adj	an n*p_adj genotype matrix (dosage matrix) of the target sequence, where n is the sample size and p_adj is the number of genetic variants to be adjusted for in conditional analysis (or a vector of a single variant with length n if p_adj is 1).
obj_nullmodel	an object from fitting the null model, which is the output from either fit_null_glm function for unrelated samples or fit_null_glmmkin function for related samples in the STAAR package.
variant_info	a data frame or matrix of variant information (unique identifier) with p rows (listed in the same order as the columns of genotype) and should contain the following 4 columns: chromosome (chr), position (pos), reference allele (ref), and alternative allele (alt).
variant_adj_info	a data frame or matrix of adjusted variant information (unique identifier) with p_adj rows (listed in the same order as the rows of genotype_adj) and should contain the following 4 columns: chromosome (chr), position (pos), reference allele (ref), and alternative allele (alt).

Value

a list with the following members:

GTPG_cond: the covariance matrix between all variants in the variant-set (rows) and all variants in the conditional variant-set (columns) (the covariance file for conditional analysis).

variant_info: the data frame or matrix of variant information (unique identifier) with p rows (listed in the same order as the rows of GTPG_cond) and 4 columns: chromosome (chr), position (pos), reference allele (ref), and alternative allele (alt).

variant_adj_info: the data frame or matrix of adjusted variant information (unique identifier) with p_adj rows (listed in the same order as the columns of GTPG_cond) and 4 columns: chromosome (chr), position (pos), reference allele (ref), alternative allele (alt), score statistic (U), and variance (V).

References

Li, X., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nature Genetics*, *55*(1), 154-164. (pub)

MetaSTAAR_worker_sumstat

Generating summary statistics file using MetaSTAARWorker

Description

The MetaSTAAR_worker_sumstat function takes in genotype, the object from fitting the null model, and variant information (unique identifier) to generate the summary statistics file for the given variant-set.

Usage

```
MetaSTAAR_worker_sumstat(
  genotype,
  obj_nullmodel,
  variant_info,
  qc_label = NULL
)
```

Arguments

genotype	an n*p genotype matrix (dosage matrix) of the target sequence, where n is the sample size and p is the number of genetic variants.
obj_nullmodel	an object from fitting the null model, which is the output from either fit_null_glm function for unrelated samples or fit_null_glmmkin function for related samples in the STAAR package.
variant_info	a data frame or matrix of variant information (unique identifier) with p rows (listed in the same order as the columns of genotype) and should contain the following 4 columns: chromosome (chr), position (pos), reference allele (ref), and alternative allele (alt).
qc_label	a vector of quality control status for each variant in variant_info, where a pass variant is labeled as "PASS". If qc_label is NULL, it is assumed that all variants are pass variants in the study (default = NULL).

Value

`sumstat`: the data frame of all variants in the variant-set (the summary statistics file), including the following information: chromosome (chr), position (pos), reference allele (ref), alternative allele (alt), quality control status (qc_label, optional), alternative allele count (alt_AC), minor allele count (MAC), minor allele frequency (MAF), study sample size (N), score statistic (U), variance (V), and the (low-rank decomposed) dense component of the covariance file.

References

Li, X., et al. (2023). Powerful, scalable and resource-efficient meta-analysis of rare variant associations in large whole genome sequencing studies. *Nature Genetics*, *55*(1), 154-164. (pub)

Index