

Package ‘MetaSTAARlite’

February 27, 2025

Type Package

Title MetaSTAARlite Pipeline for Meta-Analysis of Whole-Genome/Whole-Exome Sequencing Data

Version 0.9.6

Date 2025-02-27

Author Xihao Li [aut, cre], Yohhan Kumarasinghe [aut, cre], Zilin Li [aut, cre]

Maintainer Xihao Li <xihao.li@unc.edu>, Yohhan Kumarasinghe <yohhan@unc.edu>, Zilin Li <lizl@nenu.edu.cn>

Description A lightweight R package for using the MetaSTAARlite pipeline in meta-analysis of whole-genome/whole-exome sequencing data.

License GPL-3

Copyright See COPYRIGHTS for details.

Imports Rcpp, STAAR, MetaSTAAR, STAARpipeline, STAARpipelineSummary, dplyr, SeqArray, SeqVarTools, GenomicFeatures, TxDb.Hsapiens.UCSC.hg38.knownGene, Matrix, methods, expm, MASS

Encoding UTF-8

LazyData true

Depends R (>= 3.2.0)

RoxygenNote 7.3.2

Suggests knitr, rmarkdown

VignetteBuilder knitr

Contents

coding_MetaSTAARlite	2
coding_MetaSTAARlite_cond	4
coding_MetaSTAARlite_worker	5
custom_MetaSTAARlite	7
custom_MetaSTAARlite_cond	9
custom_MetaSTAARlite_worker	11
Gene_Centric_Coding_Results_Summary_meta	13
Gene_Centric_Noncoding_Results_Summary_meta	14
individual_analysis_MetaSTAARlite	15
individual_analysis_MetaSTAARlite_cond	16
individual_analysis_MetaSTAARlite_worker	17
Individual_Analysis_Results_Summary_meta	18

MetaSTAARlite_merge	19
MetaSTAARlite_merge_cond	20
MetaSTAARlite_worker_cov	22
MetaSTAARlite_worker_cov_cond	23
MetaSTAARlite_worker_sumstat	24
ncRNA_MetaSTAARlite	26
ncRNA_MetaSTAARlite_cond	27
ncRNA_MetaSTAARlite_worker	29
noncoding_MetaSTAARlite	31
noncoding_MetaSTAARlite_cond	33
noncoding_MetaSTAARlite_worker	34

Index	37
--------------	-----------

coding_MetaSTAARlite	<i>Performs meta-analysis of coding functional categories using MetaSTAARlite</i>
----------------------	---

Description

This function performs meta-analysis to detect associations between a quantitative/dichotomous phenotype and coding functional categories of a gene by using the MetaSTAARlite pipeline. For each coding functional category, 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

```
coding_MetaSTAARlite(
  chr,
  gene_name,
  genes,
  sample.sizes,
  coding_sumstat_gene_list,
  coding_cov_gene_list,
  cov_maf_cutoff,
  rare_maf_cutoff = 0.01,
  rv_num_cutoff = 2,
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)
```

Arguments

chr	an integer which specifies the chromosome number.
gene_name	a character which specifies the name of the gene to be meta-analyzed using MetaSTAARlite.
genes	the genes_info object from the STAARpipeline package.

<code>sample.sizes</code>	a numeric vector with the length of <code>study.names</code> indicating the sample size of each study.
<code>coding_sumstat_gene_list</code>	a list containing study-specific summary statistics corresponding to the specified gene.
<code>coding_cov_gene_list</code>	a list containing study-specific sparse weighted covariance matrices corresponding to the specified gene.
<code>cov_maf_cutoff</code>	a numeric vector with the length of <code>study.names</code> indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.
<code>rare_maf_cutoff</code>	a numeric value specifying the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
<code>rv_num_cutoff</code>	an integer specifying the cutoff of minimum number of variants of meta-analyzing a given variant-set (default = 2).
<code>check_qc_label</code>	a logical value indicating whether variants need to be dropped according to <code>qc_label</code> specified in <code>coding_MetaSTAARlite_worker</code> (default = FALSE).
<code>variant_type</code>	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
<code>Use_annotation_weights</code>	a logical value which determines if annotations will be used as weights or not (default = TRUE).
<code>Annotation_name</code>	a character vector of annotation names used in MetaSTAARlite (default = NULL).
<code>silent</code>	a logical value which determines if the report of error messages will be suppressed (default = FALSE).

Value

a list of data frames containing the MetaSTAAR p-values (including MetaSTAAR-O) corresponding to each coding functional category of the given gene.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

coding_MetaSTAARlite_cond

Performs conditional meta-analysis of coding functional categories using MetaSTAARlite

Description

This function performs meta-analysis to detect conditional associations between a quantitative/dichotomous phenotype and coding functional categories of a gene adjusting for set of known variants by using the MetaSTAARlite pipeline. For each coding functional category, the conditional 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

```
coding_MetaSTAARlite_cond(
  chr,
  gene_name,
  genes,
  sample.sizes,
  coding_sumstat_gene_list,
  coding_cov_gene_list,
  coding_cov_cond_gene_list,
  cov_maf_cutoff,
  rare_maf_cutoff = 0.01,
  rv_num_cutoff = 2,
  effect.cond = c("homogeneous", "heterogeneous"),
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)
```

Arguments

chr	an integer which specifies the chromosome number.
gene_name	a character which specifies the name of the gene to be meta-analyzed using MetaSTAARlite.
genes	the genes_info object from the STAARpipeline package.
sample.sizes	a numeric vector with the length of study.names indicating the sample size of each study.
coding_sumstat_gene_list	a list containing study-specific summary statistics corresponding to the specified gene.
coding_cov_gene_list	a list containing study-specific sparse weighted covariance matrices corresponding to the specified gene.

coding_cov_cond_gene_list	a list containing study-specific summary statistics and covariance matrices corresponding to the specified gene for variants to be conditioned on.
cov_maf_cutoff	a numeric vector with the length of <code>study.names</code> indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.
rare_maf_cutoff	a numeric value specifying the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
rv_num_cutoff	an integer specifying the cutoff of minimum number of variants of meta-analyzing a given variant-set (default = 2).
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 <code>qc_label</code> specified in <code>coding_MetaSTAARlite_worker</code> (default = FALSE).
variant_type	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
Use_annotation_weights	a logical value which determines if annotations will be used as weights or not (default = TRUE).
Annotation_name	a character vector of annotation names used in MetaSTAARlite (default = NULL).
silent	a logical value which determines if the report of error messages will be suppressed (default = FALSE).

Value

a list of data frames containing the conditional MetaSTAAR p-values (including MetaSTAAR-O) corresponding to each coding functional category of the given gene.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

coding_MetaSTAARlite_worker

Generates summary statistics of coding functional categories using MetaSTAARlite

Description

This function uses MetaSTAARlite to generate variant-level summary statistics and sparse covariance matrices for coding functional categories of a gene.

Usage

```
coding_MetaSTAARlite_worker(
  chr,
  gene_name,
  genofile,
  obj_nullmodel,
  genes,
  known_loci = NULL,
  cov_maf_cutoff = 0.05,
  signif.digits = NULL,
  QC_label = "annotation/filter",
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Annotation_dir = "annotation/info/FunctionalAnnotation",
  Annotation_name_catalog,
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)
```

Arguments

chr	an integer which specifies the chromosome number.
gene_name	a character which specifies the name of the gene to be meta-analyzed using MetaSTAARlite.
genofile	an object of opened annotated GDS (aGDS) file.
obj_nullmodel	an object from fitting the null model, which is either the output from <code>fit_nullmodel</code> function in the STAARpipeline package, or the output from <code>fitNullModel</code> function in the GENESIS package and transformed using the <code>genesis2staar_nullmodel</code> function in the STAARpipeline package.
genes	the <code>genes_info</code> object from the STAARpipeline package.
known_loci	the data frame of variants to be adjusted for in conditional analysis. Should contain four columns in the following order: chromosome (CHR), position (POS), reference allele (REF), and alternative allele (ALT) (default = NULL).
cov_maf_cutoff	a numeric value indicating the maximum minor allele frequency cutoff under which the sparse weighted covariance file between variants is stored (default = 0.05).
signif.digits	an integer indicating the number of significant digits to be used for storing the sparse weighted covariance file. If <code>signif.digits</code> is NULL, it is assumed that no rounding will be performed (default = NULL).
QC_label	a character specifying the channel name of the QC label in the GDS/aGDS file (default = "annotation/filter").
check_qc_label	a logical value indicating whether variants need to be dropped according to <code>qc_label</code> . If <code>check_qc_label</code> is FALSE, then the summary statistics will be stored for PASS variants from the study. If <code>check_qc_label</code> is TRUE, then the summary statistics will be stored for all variants from the study, together with an additional column of <code>qc_label</code> (default = FALSE).
variant_type	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").

Annotation_dir a character specifying the channel name of the annotations in the aGDS file (default = "annotation/info/FunctionalAnnotation").

Annotation_name_catalog a data frame containing the annotation name and the corresponding channel name in the aGDS file.

Use_annotation_weights a logical value which specifies if annotations will be used as weights or not (default = TRUE).

Annotation_name a character vector of annotation names used in MetaSTAARlite (default = NULL).

silent a logical value which determines if the report of error messages will be suppressed (default = FALSE).

Value

a list of the following objects corresponding to each coding functional category of the given gene: (1) the data frame of all variants in the variant-set (the variant-level 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), variant annotations specified in Annotation_name, and the low-rank decomposed component of the covariance file; (2) the sparse matrix of all variants in the variant-set whose minor allele frequency is below cov_maf_cutoff (the sparse weighted covariance file); (3) the summary statistics and covariance matrices corresponding to the specified gene for variants to be conditioned on in known_loci.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

custom_MetaSTAARlite	<i>Performs meta-analysis of custom variant-sets (masks) using MetaSTAARlite</i>
----------------------	--

Description

This function performs meta-analysis to detect associations between a quantitative/dichotomous phenotype and a user-specified custom mask by using the MetaSTAARlite pipeline. For each custom mask, 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

```

custom_MetaSTAARlite(
  chr,
  mask_name,
  sample.sizes,
  custom_sumstat_mask_list,
  custom_cov_mask_list,
  cov_maf_cutoff,
  rare_maf_cutoff = 0.01,
  rv_num_cutoff = 2,
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)

```

Arguments

<code>chr</code>	an integer which specifies the chromosome number.
<code>mask_name</code>	a character which specifies the name of the mask to be meta-analyzed using MetaSTAARlite.
<code>sample.sizes</code>	a numeric vector with the length of <code>study.names</code> indicating the sample size of each study.
<code>custom_sumstat_mask_list</code>	a list containing study-specific summary statistics corresponding to the custom mask.
<code>custom_cov_mask_list</code>	a list containing study-specific sparse weighted covariance matrices corresponding to the custom mask.
<code>cov_maf_cutoff</code>	a numeric vector with the length of <code>study.names</code> indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.
<code>rare_maf_cutoff</code>	a numeric value specifying the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
<code>rv_num_cutoff</code>	an integer specifying the cutoff of minimum number of variants of meta-analyzing a given variant-set (default = 2).
<code>check_qc_label</code>	a logical value indicating whether variants need to be dropped according to <code>qc_label</code> specified in custom_MetaSTAARlite_worker (default = FALSE).
<code>variant_type</code>	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
<code>Use_annotation_weights</code>	a logical value which determines if annotations will be used as weights or not (default = TRUE).
<code>Annotation_name</code>	a character vector of annotation names used in MetaSTAARlite (default = NULL).
<code>silent</code>	a logical value which determines if the report of error messages will be suppressed (default = FALSE).

Value

a list of data frames containing the MetaSTAAR p-values (including MetaSTAAR-O) corresponding to the custom mask.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

custom_MetaSTAARlite_cond

Performs conditional meta-analysis of custom variant-sets (masks) using MetaSTAARlite

Description

This function performs meta-analysis to detect conditional associations between a quantitative/dichotomous phenotype and a user-specified custom mask adjusting for set of known variants by using the MetaSTAARlite pipeline. For each custom mask, the conditional 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

```
custom_MetaSTAARlite_cond(
  chr,
  mask_name,
  sample.sizes,
  custom_sumstat_mask_list,
  custom_cov_mask_list,
  custom_cov_cond_mask_list,
  cov_maf_cutoff,
  rare_maf_cutoff = 0.01,
  rv_num_cutoff = 2,
  effect.cond = c("homogeneous", "heterogeneous"),
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)
```

Arguments

<code>chr</code>	an integer which specifies the chromosome number.
<code>mask_name</code>	a character which specifies the name of the mask to be meta-analyzed using MetaSTAARlite.
<code>sample.sizes</code>	a numeric vector with the length of <code>study.names</code> indicating the sample size of each study.
<code>custom_sumstat_mask_list</code>	a list containing study-specific summary statistics corresponding to the custom mask.
<code>custom_cov_cond_mask_list</code>	a list containing study-specific summary statistics and covariance matrices corresponding to the custom mask for variants to be conditioned on.
<code>cov_maf_cutoff</code>	a numeric vector with the length of <code>study.names</code> indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.
<code>rare_maf_cutoff</code>	a numeric value specifying the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
<code>rv_num_cutoff</code>	an integer specifying the cutoff of minimum number of variants of meta-analyzing a given variant-set (default = 2).
<code>effect.cond</code>	a character value indicating the effects of variants to be adjusted for in conditional analysis are "homogeneous" or "heterogeneous" (default = "homogeneous").
<code>check_qc_label</code>	a logical value indicating whether variants need to be dropped according to <code>qc_label</code> specified in custom_MetaSTAARlite_worker (default = FALSE).
<code>variant_type</code>	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
<code>Use_annotation_weights</code>	a logical value which determines if annotations will be used as weights or not (default = TRUE).
<code>Annotation_name</code>	a character vector of annotation names used in MetaSTAARlite (default = NULL).
<code>silent</code>	a logical value which determines if the report of error messages will be suppressed (default = FALSE).
<code>coding_cov_mask_list</code>	a list containing study-specific sparse weighted covariance matrices corresponding to the custom mask.

Value

a list of data frames containing the conditional MetaSTAAR p-values (including MetaSTAAR-O) corresponding to the custom mask.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

custom_MetaSTAARlite_worker

Generates summary statistics of custom variant-sets (masks) using MetaSTAARlite

Description

This function uses MetaSTAARlite to generate variant-level summary statistics and sparse covariance matrices for a user-specified custom mask.

Usage

```
custom_MetaSTAARlite_worker(
  chr,
  variant_list,
  agds_variant_list = NULL,
  genofile,
  obj_nullmodel,
  known_loci = NULL,
  cov_maf_cutoff = 0.05,
  signif.digits = NULL,
  QC_label = "annotation/filter",
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Annotation_dir = "annotation/info/FunctionalAnnotation",
  Annotation_name_catalog,
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)
```

Arguments

- | | |
|-------------------|--|
| chr | an integer which specifies the chromosome number. |
| variant_list | the data frame of variants in the custom mask to be included. Should contain five columns in the following order: chromosome (CHR), position (POS), reference allele (REF), alternative allele (ALT), and mask name (MaskName). An example is given (here) |
| agds_variant_list | the pre-loaded data frame of all variants in the aGDS for faster computation. Should contain four columns in the following order: chromosome (CHR), position (POS), reference allele (REF), and alternative allele (ALT), |
| genofile | an object of opened annotated GDS (aGDS) file. |
| obj_nullmodel | an object from fitting the null model, which is either the output from <code>fit_nullmodel</code> function in the <code>STAARpipeline</code> package, or the output from <code>fitNullModel</code> function in the <code>GENESIS</code> package and transformed using the <code>genesis2staar_nullmodel</code> function in the <code>STAARpipeline</code> package. |

known_loci	the data frame of variants to be adjusted for in conditional analysis. Should contain four columns in the following order: chromosome (CHR), position (POS), reference allele (REF), and alternative allele (ALT) (default = NULL).
cov_maf_cutoff	a numeric value indicating the maximum minor allele frequency cutoff under which the sparse weighted covariance file between variants is stored (default = 0.05).
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 = NULL).
QC_label	a character specifying the channel name of the QC label in the GDS/aGDS file (default = "annotation/filter").
check_qc_label	a logical value indicating whether variants need to be dropped according to qc_label. If check_qc_label is FALSE, then the summary statistics will be stored for PASS variants from the study. If check_qc_label is TRUE, then the summary statistics will be stored for all variants from the study, together with an additional column of qc_label (default = FALSE).
variant_type	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
Annotation_dir	a character specifying the channel name of the annotations in the aGDS file (default = "annotation/info/FunctionalAnnotation").
Annotation_name_catalog	a data frame containing the annotation name and the corresponding channel name in the aGDS file.
Use_annotation_weights	a logical value which specifies if annotations will be used as weights or not (default = TRUE).
Annotation_name	a character vector of annotation names used in MetaSTAARlite (default = NULL).
silent	a logical value which determines if the report of error messages will be suppressed (default = FALSE).

Value

a list of the following objects corresponding to the custom mask: (1) the data frame of all variants in the variant-set (the variant-level 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), variant annotations specified in Annotation_name, and the low-rank decomposed component of the covariance file; (2) the sparse matrix of all variants in the variant-set whose minor allele frequency is below cov_maf_cutoff (the sparse weighted covariance file); (3) the summary statistics and covariance matrices corresponding to the specified gene for variants to be conditioned on in known_loci.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

Gene_Centric_Coding_Results_Summary_meta

Generates summary table and visualization for the meta-analysis of coding functional categories that was conducted using MetaSTAARlite

Description

This function takes in objects of gene-centric coding meta-analysis results from MetaSTAARlite and generates a summary table, Manhattan plot, and QQ plot for the meta-analysis of coding functional categories that was conducted based on the parameters provided by the user.

Usage

```
Gene_Centric_Coding_Results_Summary_meta(
  gene_centric_coding_jobs_num,
  input_path,
  output_path,
  gene_centric_results_name,
  alpha = 2.5e-06,
  manhattan_plot = FALSE,
  QQ_plot = FALSE
)
```

Arguments

gene_centric_coding_jobs_num	an integer which specifies the number of jobs done in the gene-centric coding meta-analysis.
input_path	a character which specifies the file path to the gene-centric coding meta-analysis results files.
output_path	a character which specifies the file path to the desired location of the produced summary table and visualizations for gene-centric coding meta-analysis.
gene_centric_results_name	a character which specifies the name (excluding the jobs number) of the gene-centric coding meta-analysis results files.
alpha	a numeric value which specifies the desired significance threshold for the gene-centric coding meta-analysis (default = 2.5E-06).
manhattan_plot	a logical value which determines if a Manhattan plot is generated (default = FALSE).
QQ_plot	a logical value which determines if a QQ plot is generated (default = FALSE).

Gene_Centric_Noncoding_Results_Summary_meta

Generates summary table and visualization for the meta-analysis of noncoding functional categories that was conducted using MetaS-TAARlite

Description

This function takes in objects of gene-centric noncoding meta-analysis results from MetaSTAARlite and generates a summary table, Manhattan plot, and QQ plot for the meta-analysis of noncoding functional categories that was conducted based on the parameters provided by the user.

Usage

```
Gene_Centric_Noncoding_Results_Summary_meta(
  gene_centric_noncoding_jobs_num,
  input_path,
  output_path,
  gene_centric_results_name,
  ncRNA_jobs_num,
  ncRNA_input_path,
  ncRNA_output_path,
  ncRNA_results_name,
  alpha = 2.5e-06,
  alpha_ncRNA = 2.5e-06,
  ncRNA_pos = NULL,
  manhattan_plot = FALSE,
  QQ_plot = FALSE
)
```

Arguments

gene_centric_noncoding_jobs_num	an integer which specifies the number of jobs done in the gene-centric noncoding meta-analysis.
input_path	a character which specifies the file path to the gene-centric noncoding meta-analysis results files.
output_path	a character which specifies the file path to the desired location of the produced summary table and visualizations for gene-centric noncoding meta-analysis.
gene_centric_results_name	a character which specifies the name (excluding the jobs number) of the gene-centric noncoding meta-analysis results files.
ncRNA_jobs_num	an integer which specifies the number of jobs done in the ncRNA meta-analysis.
ncRNA_input_path	a character which specifies the file path to the ncRNA meta-analysis results files.
ncRNA_output_path	a character which specifies the file path to the desired location of the produced summary table and visualizations for the ncRNA meta-analysis.

ncRNA_results_name	a character which specifies the name (excluding the jobs number) of the ncRNA meta-analysis results files.
alpha	a numeric value which specifies the desired significance threshold for the gene-centric noncoding meta-analysis (default = 2.5E-06).
alpha_ncRNA	a numeric value which specifies the desired significance threshold for the ncRNA meta-analysis (default is 2.5E-06).
manhattan_plot	a logical value which determines if a Manhattan plot is generated (default = FALSE).
QQ_plot	a logical value which determines if a QQ plot is generated (default = FALSE).

individual_analysis_MetaSTAARlite

Performs meta-analysis of individual variants by using the MetaSTAARlite pipeline.

Description

This function performs meta-analysis to detect associations between a quantitative/dichotomous phenotype and each individual variant in a genetic region by using score test.

Usage

```
individual_analysis_MetaSTAARlite(
  sample.sizes,
  sumstat.list,
  mac_cutoff = 20,
  check_qc_label = FALSE
)
```

Arguments

sample.sizes	a numeric vector with the length of study.names indicating the sample size of each study.
sumstat.list	a character vector containing the directories of the study-specific summary statistics file folders.
mac_cutoff	an integer specifying the cutoff of minimum minor allele count in defining individual variants. Default is 20.
check_qc_label	a logical value indicating whether variants need to be dropped according to qc_label. Default is FALSE. specified in generate_MetaSTAAR_sumstat and generate_MetaSTAAR_cov . If check_qc_label is FALSE, it is assumed that no variant will be dropped (default = FALSE).

Value

a data frame containing the score test p-value and the estimated effect size of the alternative allele for each individual variant in the given genetic region.

individual_analysis_MetaSTAARlite_cond

Performs conditional meta-analysis of individual variants by using the MetaSTAARlite pipeline.

Description

This function performs meta-analysis to detect associations between a quantitative/dichotomous phenotype and each individual variant in a genetic region by using score test.

Usage

```
individual_analysis_MetaSTAARlite_cond(
  individual_results,
  sample.sizes,
  sumstat.list,
  covcond.list,
  mac_cutoff = 20,
  effect.cond = c("homogeneous", "heterogeneous"),
  check_qc_label = FALSE
)
```

Arguments

individual_results	a dataframe containing the unconditional results of the meta-analysis.
sample.sizes	a numeric vector with the length of study.names indicating the sample size of each study.
sumstat.list	a character vector containing the directories of the study-specific summary statistics file folders.
covcond.list	a character vector containing the names of variants to be conditioned on.
mac_cutoff	the cutoff of minimum minor allele count in defining individual variants. Default is 20.
effect.cond	a character vector specifying whether the effects of variants to be adjusted for in conditional analysis are "homogeneous" or "heterogeneous". Default is "homogeneous".
check_qc_label	a logical value indicating whether variants need to be dropped according to qc_label. Default is FALSE. specified in generate_MetaSTAAR_sumstat and generate_MetaSTAAR_cov . If check_qc_label is FALSE, it is assumed that no variant will be dropped (default = FALSE).

Value

a data frame containing the score test p-value and the estimated effect size of the alternative allele for each individual variant in the given genetic region.

individual_analysis_MetaSTAARlite_worker

Performs the worker step of MetaSTAARlite for individual variants

Description

This function uses MetaSTAARliteWorker to generate variant summary statistics and sparse LD matrices for gene-centric coding analysis.

Usage

```
individual_analysis_MetaSTAARlite_worker(
  chr,
  start_loc,
  end_loc,
  genofile,
  obj_nullmodel,
  known_loci = NULL,
  subsegment.size = 50000,
  QC_label = "annotation/filter",
  check_qc_label = FALSE,
  variant_type = c("variant", "SNV", "Indel"),
  Annotation_dir = "annotation/info/FunctionalAnnotation",
  Annotation_name_catalog,
  Use_annotation_weights = c(FALSE, TRUE),
  Annotation_name = NULL,
  silent = FALSE
)
```

Arguments

chr	an integer which specifies the chromosome number.
start_loc	an integer which specifies the starting location of variants to be analyzed.
end_loc	an integer which specifies the end location of variants to be analyzed.
genofile	an object of opened annotated GDS (aGDS) file with variant annotation information (and without genotype).
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.
known_loci	the data frame of variants to be adjusted for in conditional analysis. Should contain four columns in the following order: chromosome (CHR), position (POS), reference allele (REF), and alternative allele (ALT). Default is NULL.
subsegment.size	a numeric value which specifies the size of subsegments. Default is 5e4.
QC_label	a character specifying the channel name of the QC label in the GDS/aGDS file. Default is "annotation/filter".
check_qc_label	a logical value indicating whether variants need to be dropped according to qc_label specified in generate_MetaSTAAR_sumstat and generate_MetaSTAAR_cov . Default is FALSE.

variant_type	a character specifying the types of variants to be considered. Choices include "SNV", "Indel", or "variant" (default = "SNV").
Annotation_dir	a character specifying the channel name of the annotations in the aGDS file. Default is "annotation/info/FunctionalAnnotation"
Annotation_name_catalog	a data frame containing the annotation name and the corresponding channel name in the aGDS file.
Use_annotation_weights	a logical value which specifies if annotations will be used as weights or not. Default is TRUE.
Annotation_name	a vector of annotation names used in MetaSTAAR. Default is NULL.
silent	a logical value which determines if the report of error messages will be suppressed. Default is FALSE.

Value

two objects. First, 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. Second, 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.

Individual_Analysis_Results_Summary_meta

Generates summary table and/or visualization for the meta-analysis of individual variants that was conducted.

Description

This function generates a summary table, Manhattan plot, and QQ plot for the meta-analysis of individual variants that was conducted based on the parameters provided by the user.

Usage

```
Individual_Analysis_Results_Summary_meta(
  jobs_num,
  input_path,
  output_path,
  individual_results_name,
  alpha = 5e-09,
  manhattan_plot = FALSE,
  QQ_plot = FALSE
)
```

Arguments

jobs_num	an integer which specifies the number of jobs done in the meta-analysis.
input_path	a character which specifies the file path to the meta-analysis results files.
output_path	a character which specifies the file path to the desired location of the produced summary table and visualizations.
alpha	a numeric value which specifies the desired significance threshold. Default is 5E-09.
manhattan_plot	a logical value which determines if a Manhattan plot is generated. Default is FALSE.
QQ_plot	a logical value which determines if a QQ plot is generated. Default is FALSE.
individual_results_name	a character which specifies the name (excluding the jobs number) of the meta-analysis results files.

MetaSTAARlite_merge	<i>Merges the generated summary statistics and sparse covariance matrices of different studies</i>
---------------------	--

Description

This function merges the generated variant-level summary statistics and sparse covariance matrices of all participating studies in preparation for the meta-analysis step of MetaSTAARlite.

Usage

```
MetaSTAARlite_merge(
  chr,
  sample.sizes,
  sumstat.list,
  cov.list,
  rare_maf_cutoff = 0.01,
  cov_maf_cutoff,
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Use_annotation_weights = TRUE,
  Annotation_name = NULL
)
```

Arguments

chr	an integer which specifies the chromosome number.
sample.sizes	a numeric vector with the length of study.names indicating the sample size of each study.
sumstat.list	a list containing study-specific summary statistics from all participating studies.
cov.list	a list containing study-specific sparse weighted covariance matrices from all participating studies.

<code>rare_maf_cutoff</code>	a numeric value specifying the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
<code>cov_maf_cutoff</code>	a numeric vector with the length of <code>study.names</code> indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.
<code>check_qc_label</code>	a logical value indicating whether variants need to be dropped according to <code>qc_label</code> (default = FALSE).
<code>variant_type</code>	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
<code>Use_annotation_weights</code>	a logical value which determines if annotations will be used as weights or not (default = TRUE).
<code>Annotation_name</code>	a character vector of annotation names used in MetaSTAARlite (default = NULL).

Value

a list with the following members:

`info`: the merged data frame of all variants in the variant-set 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 variant-set of interest whose combined minor allele frequency is below `rare_maf_cutoff`.

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

`annotation_phred` the merged functional annotation data in PHRED score scale of all variants in the variant-set 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](#))

MetaSTAARlite_merge_cond

Merges the generated summary statistics and covariance matrices of different studies for conditional analysis

Description

This function merges the generated variant-level summary statistics and covariance matrices of all participating studies in preparation for the conditional meta-analysis step of MetaSTAARlite.

Usage

```
MetaSTAARlite_merge_cond(
  chr,
  sample.sizes,
  sumstat.list,
  cov.list,
  covcond.list,
  rare_maf_cutoff = 0.01,
  cov_maf_cutoff,
  effect.cond = c("homogeneous", "heterogeneous"),
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Use_annotation_weights = TRUE,
  Annotation_name = NULL
)
```

Arguments

<code>chr</code>	an integer which specifies the chromosome number.
<code>sample.sizes</code>	a numeric vector with the length of <code>study.names</code> indicating the sample size of each study.
<code>sumstat.list</code>	a list containing study-specific summary statistics from all participating studies.
<code>cov.list</code>	a list containing study-specific sparse weighted covariance matrices from all participating studies.
<code>covcond.list</code>	a list containing study-specific summary statistics and covariance matrices for variants to be conditioned on from all participating studies.
<code>rare_maf_cutoff</code>	the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
<code>cov_maf_cutoff</code>	a numeric vector with the length of <code>study.names</code> indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.
<code>effect.cond</code>	a character value indicating the effects of variants to be adjusted for in conditional analysis are "homogeneous" or "heterogeneous" (default = "homogeneous").
<code>check_qc_label</code>	a logical value indicating whether variants need to be dropped according to <code>qc_label</code> (default = FALSE).
<code>variant_type</code>	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
<code>Use_annotation_weights</code>	a logical value which determines if annotations will be used as weights or not (default = TRUE).
<code>Annotation_name</code>	a character vector of annotation names used in MetaSTAARlite (default = NULL).

Value

a list with the following members:

info: the merged data frame of all variants in the variant-set 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 variant-set 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 variant-set of interest whose combined minor allele frequency is below `rare_maf_cutoff`, adjusting for a given list of variants.

annotation_phred the merged functional annotation data in PHRED score scale of all variants in the variant-set 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](#))

MetaSTAARlite_worker_cov

Generating sparse weighted covariance file using MetaSTAARlite (the "worker" step)

Description

The MetaSTAARlite_worker_cov function takes in genotype and the object from fitting the null model to generate the sparse weighted covariance file for the given variant-set.

Usage

```
MetaSTAARlite_worker_cov(
  genotype,
  obj_nullmodel,
  cov_maf_cutoff = 0.05,
  qc_label = NULL,
  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. dgCMatrx 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 either the output from <code>fit_nullmodel</code> function in the STAARpipeline package, or the output from <code>fitNullModel</code> function in the GENESIS package and transformed using the <code>genesis2staar_nullmodel</code> function in the STAARpipeline 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 (default = 0.05).
qc_label	a vector of quality control status for each variant in genotype, 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).
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).

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](#))

MetaSTAARlite_worker_cov_cond

Generating summary statistics file for conditional analysis using MetaSTAARlite (the "worker" step)

Description

The MetaSTAARlite_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 summary statistics file for the given variant-set, adjusting for a given list of variants.

Usage

```
MetaSTAARlite_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 either the output from `fit_nullmodel` function in the `STAARpipeline` package, or the output from `fitNullModel` function in the `GENESIS` package and transformed using the `genesis2staar_nullmodel` function in the `STAARpipeline` 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](#))

MetaSTAARlite_worker_sumstat

*Generating variant-level summary statistics file using MetaSTAARlite
(the "worker" step)*

Description

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

Usage

```
MetaSTAARlite_worker_sumstat(
  genotype,
  obj_nullmodel,
  variant_info,
  qc_label = NULL,
```



```

    annotation_phred = NULL,
    for_individual_analysis = FALSE
  )

```

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 either the output from <code>fit_nullmodel</code> function in the <code>STAARpipeline</code> package, or the output from <code>fitNullModel</code> function in the <code>GENESIS</code> package and transformed using the <code>genesis2staar_nullmodel</code> function in the <code>STAARpipeline</code> 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 <code>variant_info</code> , where a PASS variant is labeled as "PASS". If <code>qc_label</code> is NULL, it is assumed that all variants are PASS variants in the study (default = NULL).
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). Continuous scores should be given in PHRED score scale, where the PHRED score of j-th variant is defined to be $-10 \cdot \log_{10}(\text{rank}(-\text{score}_j)/\text{total})$ 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, STAAR will perform the SKAT(1,25), SKAT(1,1), Burden(1,25), Burden(1,1), ACAT-V(1,25), ACAT-V(1,1) and ACAT-O tests (default = NULL).
for_individual_analysis	a logical value indicating whether it is used for individual (single-variant) meta-analysis (default = FALSE).

Value

sumstat: the data frame of all variants in the variant-set or the list of individual variants (the variant-level 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), variant annotations provided in `annotation_phred`, and the low-rank decomposed 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](#))

ncRNA_MetaSTAARlite	<i>Performs meta-analysis of long noncoding RNA (ncRNA) category using MetaSTAARlite</i>
---------------------	--

Description

This function performs meta-analysis to detect associations between a quantitative/dichotomous phenotype and the exonic and splicing category of an ncRNA gene by using the MetaSTAARlite pipeline. For each coding functional category, 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

```
ncRNA_MetaSTAARlite(
  chr,
  gene_name,
  sample.sizes,
  ncRNA_sumstat_gene_list,
  ncRNA_cov_gene_list,
  cov_maf_cutoff,
  rare_maf_cutoff = 0.01,
  rv_num_cutoff = 2,
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)
```

Arguments

chr	an integer which specifies the chromosome number.
gene_name	a character which specifies the name of the ncRNA gene to be meta-analyzed using MetaSTAARlite.
sample.sizes	a numeric vector with the length of study.names indicating the sample size of each study.
ncRNA_sumstat_gene_list	a list containing study-specific summary statistics corresponding to the specified gene.
ncRNA_cov_gene_list	a list containing study-specific sparse weighted covariance matrices corresponding to the specified gene.
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.
rare_maf_cutoff	the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).

rv_num_cutoff	the cutoff of minimum number of variants of meta-analyzing a given variant-set (default = 2).
check_qc_label	a logical value indicating whether variants need to be dropped according to qc_label specified in ncRNA_MetaSTAARlite_worker (default = FALSE).
variant_type	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
Use_annotation_weights	a logical value which determines if annotations will be used as weights or not (default = TRUE).
Annotation_name	a character vector of annotation names used in MetaSTAARlite (default = NULL).
silent	a logical value which determines if the report of error messages will be suppressed (default = FALSE).

Value

a list of data frames containing the MetaSTAAR p-values (including MetaSTAAR-O) corresponding to the exonic and splicing category of the given ncRNA gene.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

ncRNA_MetaSTAARlite_cond

Performs conditional meta-analysis of long noncoding RNA (ncRNA) category using MetaSTAARlite

Description

This function performs meta-analysis to detect conditional associations between a quantitative/dichotomous phenotype and the exonic and splicing category of an ncRNA gene adjusting for set of known variants by using the MetaSTAARlite pipeline. For each coding functional category, the conditional 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

```
ncRNA_MetaSTAARlite_cond(
  chr,
  gene_name,
  sample.sizes,
  ncRNA_sumstat_gene_list,
  ncRNA_cov_gene_list,
  ncRNA_cov_cond_gene_list,
  cov_maf_cutoff,
  rare_maf_cutoff = 0.01,
  rv_num_cutoff = 2,
  effect.cond = c("homogeneous", "heterogeneous"),
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)
```

Arguments

<code>chr</code>	an integer which specifies the chromosome number.
<code>gene_name</code>	a character which specifies the name of the ncRNA gene to be meta-analyzed using MetaSTAARlite.
<code>sample.sizes</code>	a numeric vector with the length of <code>study.names</code> indicating the sample size of each study.
<code>ncRNA_sumstat_gene_list</code>	a list containing study-specific summary statistics corresponding to the specified gene.
<code>ncRNA_cov_gene_list</code>	a list containing study-specific sparse weighted covariance matrices corresponding to the specified gene.
<code>cov_maf_cutoff</code>	a numeric vector with the length of <code>study.names</code> indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.
<code>rare_maf_cutoff</code>	the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
<code>rv_num_cutoff</code>	the cutoff of minimum number of variants of meta-analyzing a given variant-set (default = 2).
<code>effect.cond</code>	a character value indicating the effects of variants to be adjusted for in conditional analysis are "homogeneous" or "heterogeneous" (default = "homogeneous").
<code>check_qc_label</code>	a logical value indicating whether variants need to be dropped according to <code>qc_label</code> specified in ncRNA_MetaSTAARlite_worker (default = FALSE).
<code>variant_type</code>	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
<code>Use_annotation_weights</code>	a logical value which determines if annotations will be used as weights or not (default = TRUE).

Annotation_name
a character vector of annotation names used in MetaSTAARlite (default = NULL).

silent
a logical value which determines if the report of error messages will be suppressed (default = FALSE).

coding_cov_cond_gene_list
a list containing study-specific summary statistics and covariance matrices corresponding to the specified gene for variants to be conditioned on.

Value

a list of data frames containing the conditional MetaSTAAR p-values (including MetaSTAAR-O) corresponding to the exonic and splicing category of the given ncRNA gene.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

ncRNA_MetaSTAARlite_worker

Generates summary statistics of long noncoding RNA (ncRNA) category using MetaSTAARlite

Description

This function uses MetaSTAARlite to generate variant-level summary statistics and sparse covariance matrices the exonic and splicing category of an ncRNA gene.

Usage

```
ncRNA_MetaSTAARlite_worker(
  chr,
  gene_name,
  genofile,
  obj_nullmodel,
  known_loci = NULL,
  cov_maf_cutoff = 0.05,
  signif.digits = NULL,
  QC_label = "annotation/filter",
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Annotation_dir = "annotation/info/FunctionalAnnotation",
  Annotation_name_catalog,
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)
```

Arguments

<code>chr</code>	an integer which specifies the chromosome number.
<code>gene_name</code>	a character which specifies the name of the gene to be meta-analyzed using MetaSTAARlite.
<code>genofile</code>	an object of opened annotated GDS (aGDS) file.
<code>obj_nullmodel</code>	an object from fitting the null model, which is either the output from <code>fit_nullmodel</code> function in the <code>STAARpipeline</code> package, or the output from <code>fitNullModel</code> function in the <code>GENESIS</code> package and transformed using the <code>genesis2staar_nullmodel</code> function in the <code>STAARpipeline</code> package.
<code>known_loci</code>	the data frame of variants to be adjusted for in conditional analysis. Should contain four columns in the following order: chromosome (CHR), position (POS), reference allele (REF), and alternative allele (ALT) (default = NULL).
<code>cov_maf_cutoff</code>	a numeric value indicating the maximum minor allele frequency cutoff under which the sparse weighted covariance file between variants is stored (default = 0.05).
<code>signif.digits</code>	an integer indicating the number of significant digits to be used for storing the sparse weighted covariance file. If <code>signif.digits</code> is NULL, it is assumed that no rounding will be performed (default = NULL).
<code>QC_label</code>	a character specifying the channel name of the QC label in the GDS/aGDS file (default = "annotation/filter").
<code>check_qc_label</code>	a logical value indicating whether variants need to be dropped according to <code>qc_label</code> . If <code>check_qc_label</code> is FALSE, then the summary statistics will be stored for PASS variants from the study. If <code>check_qc_label</code> is TRUE, then the summary statistics will be stored for all variants from the study, together with an additional column of <code>qc_label</code> (default = FALSE).
<code>variant_type</code>	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
<code>Annotation_dir</code>	a character specifying the channel name of the annotations in the aGDS file (default = "annotation/info/FunctionalAnnotation").
<code>Annotation_name_catalog</code>	a data frame containing the annotation name and the corresponding channel name in the aGDS file.
<code>Use_annotation_weights</code>	a logical value which specifies if annotations will be used as weights or not (default = TRUE).
<code>Annotation_name</code>	a character vector of annotation names used in MetaSTAARlite (default = NULL).
<code>silent</code>	a logical value which determines if the report of error messages will be suppressed (default = FALSE).

Value

a list of the following objects corresponding to the exonic and splicing category of the given ncRNA gene: (1) the data frame of all variants in the variant-set (the variant-level 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`), variant annotations specified in `Annotation_name`, and the low-rank decomposed component of the covariance file; (2) the sparse matrix of all variants in the variant-set whose minor

allele frequency is below `cov_maf_cutoff` (the sparse weighted covariance file); (3) the summary statistics and covariance matrices corresponding to the specified gene for variants to be conditioned on in `known_loci`.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

noncoding_MetaSTAARlite

Performs meta-analysis of noncoding functional categories using MetaSTAARlite

Description

This function performs meta-analysis to detect associations between a quantitative/dichotomous phenotype and noncoding functional categories of a gene by using the MetaSTAARlite pipeline. For each coding functional category, 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

```
noncoding_MetaSTAARlite(
  chr,
  gene_name,
  sample.sizes,
  noncoding_sumstat_gene_list,
  noncoding_cov_gene_list,
  cov_maf_cutoff,
  rare_maf_cutoff = 0.01,
  rv_num_cutoff = 2,
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)
```

Arguments

<code>chr</code>	an integer which specifies the chromosome number.
<code>gene_name</code>	a character which specifies the name of the gene to be meta-analyzed using MetaSTAARlite.

<code>sample.sizes</code>	a numeric vector with the length of <code>study.names</code> indicating the sample size of each study.
<code>noncoding_sumstat_gene_list</code>	a list containing study-specific summary statistics corresponding to the specified gene.
<code>noncoding_cov_gene_list</code>	a list containing study-specific sparse weighted covariance matrices corresponding to the specified gene.
<code>cov_maf_cutoff</code>	a numeric vector with the length of <code>study.names</code> indicating the maximum minor allele frequency cutoffs under which the sparse weighted covariance files between variants are stored.
<code>rare_maf_cutoff</code>	a numeric value specifying the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
<code>rv_num_cutoff</code>	an integer specifying the cutoff of minimum number of variants of meta-analyzing a given variant-set (default = 2).
<code>check_qc_label</code>	a logical value indicating whether variants need to be dropped according to <code>qc_label</code> specified in <code>noncoding_MetaSTAARlite_worker</code> (default = FALSE).
<code>variant_type</code>	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
<code>Use_annotation_weights</code>	a logical value which determines if annotations will be used as weights or not (default = TRUE).
<code>Annotation_name</code>	a character vector of annotation names used in MetaSTAARlite (default = NULL).
<code>silent</code>	a logical value which determines if the report of error messages will be suppressed (default = FALSE).

Value

a list of data frames containing the MetaSTAAR p-values (including MetaSTAAR-O) corresponding to each noncoding functional category of the given gene.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

noncoding_MetaSTAARlite_cond

Performs conditional meta-analysis of noncoding functional categories using MetaSTAARlite

Description

This function performs meta-analysis to detect conditional associations between a quantitative/dichotomous phenotype and noncoding functional categories of a gene adjusting for set of known variants by using the MetaSTAARlite pipeline. For each coding functional category, the conditional 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

```
noncoding_MetaSTAARlite_cond(
  chr,
  gene_name,
  sample.sizes,
  noncoding_sumstat_gene_list,
  noncoding_cov_gene_list,
  noncoding_cov_cond_gene_list,
  cov_maf_cutoff,
  rare_maf_cutoff = 0.01,
  rv_num_cutoff = 2,
  effect.cond = c("homogeneous", "heterogeneous"),
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)
```

Arguments

chr	an integer which specifies the chromosome number.
gene_name	a character which specifies the name of the gene to be meta-analyzed using MetaSTAARlite.
sample.sizes	a numeric vector with the length of study.names indicating the sample size of each study.
noncoding_sumstat_gene_list	a list containing study-specific summary statistics corresponding to the specified gene.
noncoding_cov_gene_list	a list containing study-specific sparse weighted covariance matrices corresponding to the specified gene.
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.

rare_maf_cutoff	a numeric value specifying the cutoff of maximum minor allele frequency in defining rare variants (default = 0.01).
rv_num_cutoff	an integer specifying the cutoff of minimum number of variants of meta-analyzing a given variant-set (default = 2).
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 noncoding_MetaSTAARlite_worker (default = FALSE).
variant_type	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
Use_annotation_weights	a logical value which determines if annotations will be used as weights or not (default = TRUE).
Annotation_name	a character vector of annotation names used in MetaSTAARlite (default = NULL).
silent	a logical value which determines if the report of error messages will be suppressed (default = FALSE).
coding_cov_cond_gene_list	a list containing study-specific summary statistics and covariance matrices corresponding to the specified gene for variants to be conditioned on.

Value

a list of data frames containing the conditional MetaSTAAR p-values (including MetaSTAAR-O) corresponding to each noncoding functional category of the given gene.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

noncoding_MetaSTAARlite_worker

Generates summary statistics of noncoding functional categories using MetaSTAARlite

Description

This function uses MetaSTAARlite to generate variant-level summary statistics and sparse covariance matrices for noncoding functional categories of a gene.

Usage

```

noncoding_MetaSTAARlite_worker(
  chr,
  gene_name,
  genofile,
  obj_nullmodel,
  known_loci = NULL,
  cov_maf_cutoff = 0.05,
  signif.digits = NULL,
  QC_label = "annotation/filter",
  check_qc_label = FALSE,
  variant_type = c("SNV", "Indel", "variant"),
  Annotation_dir = "annotation/info/FunctionalAnnotation",
  Annotation_name_catalog,
  Use_annotation_weights = TRUE,
  Annotation_name = NULL,
  silent = FALSE
)

```

Arguments

<code>chr</code>	an integer which specifies the chromosome number.
<code>gene_name</code>	a character which specifies the name of the gene to be meta-analyzed using MetaSTAARlite.
<code>genofile</code>	an object of opened annotated GDS (aGDS) file.
<code>obj_nullmodel</code>	an object from fitting the null model, which is either the output from <code>fit_nullmodel</code> function in the STAARpipeline package, or the output from <code>fitNullModel</code> function in the GENESIS package and transformed using the <code>genesis2staar_nullmodel</code> function in the STAARpipeline package.
<code>known_loci</code>	the data frame of variants to be adjusted for in conditional analysis. Should contain four columns in the following order: chromosome (CHR), position (POS), reference allele (REF), and alternative allele (ALT) (default = NULL).
<code>cov_maf_cutoff</code>	a numeric value indicating the maximum minor allele frequency cutoff under which the sparse weighted covariance file between variants is stored (default = 0.05).
<code>signif.digits</code>	an integer indicating the number of significant digits to be used for storing the sparse weighted covariance file. If <code>signif.digits</code> is NULL, it is assumed that no rounding will be performed (default = NULL).
<code>QC_label</code>	a character specifying the channel name of the QC label in the GDS/aGDS file (default = "annotation/filter").
<code>check_qc_label</code>	a logical value indicating whether variants need to be dropped according to <code>qc_label</code> . If <code>check_qc_label</code> is FALSE, then the summary statistics will be stored for PASS variants from the study. If <code>check_qc_label</code> is TRUE, then the summary statistics will be stored for all variants from the study, together with an additional column of <code>qc_label</code> (default = FALSE).
<code>variant_type</code>	a character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV").
<code>Annotation_dir</code>	a character specifying the channel name of the annotations in the aGDS file (default = "annotation/info/FunctionalAnnotation").

Annotation_name_catalog	a data frame containing the annotation name and the corresponding channel name in the aGDS file.
Use_annotation_weights	a logical value which specifies if annotations will be used as weights or not (default = TRUE).
Annotation_name	a character vector of annotation names used in MetaSTAARlite (default = NULL).
silent	a logical value which determines if the report of error messages will be suppressed (default = FALSE).

Value

a list of the following objects corresponding to each noncoding functional category of the given gene: (1) the data frame of all variants in the variant-set (the variant-level 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), variant annotations specified in Annotation_name, and the low-rank decomposed component of the covariance file; (2) the sparse matrix of all variants in the variant-set whose minor allele frequency is below cov_maf_cutoff (the sparse weighted covariance file); (3) the summary statistics and covariance matrices corresponding to the specified gene for variants to be conditioned on in known_loci.

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, Z., Li, X., et al. (2022). A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies. *Nature Methods*. ([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](#))

Index

`coding_MetaSTAARlite`, [2](#)
`coding_MetaSTAARlite_cond`, [4](#)
`coding_MetaSTAARlite_worker`, [3](#), [5](#), [5](#)
`custom_MetaSTAARlite`, [7](#)
`custom_MetaSTAARlite_cond`, [9](#)
`custom_MetaSTAARlite_worker`, [8](#), [10](#), [11](#)

`fit_null_glm`, [17](#)
`fit_null_glmkin`, [17](#)
`fit_nullmodel`, [6](#), [11](#), [22](#), [24](#), [25](#), [30](#), [35](#)

`Gene_Centric_Coding_Results_Summary_meta`,
[13](#)
`Gene_Centric_Noncoding_Results_Summary_meta`,
[14](#)
`generate_MetaSTAAR_cov`, [15–17](#)
`generate_MetaSTAAR_sumstat`, [15–17](#)
`genesis2staar_nullmodel`, [6](#), [11](#), [22](#), [24](#), [25](#),
[30](#), [35](#)

`individual_analysis_MetaSTAARlite`, [15](#)
`individual_analysis_MetaSTAARlite_cond`,
[16](#)
`individual_analysis_MetaSTAARlite_worker`,
[17](#)
`Individual_Analysis_Results_Summary_meta`,
[18](#)

`MetaSTAARlite_merge`, [19](#)
`MetaSTAARlite_merge_cond`, [20](#)
`MetaSTAARlite_worker_cov`, [22](#)
`MetaSTAARlite_worker_cov_cond`, [23](#)
`MetaSTAARlite_worker_sumstat`, [24](#)

`ncRNA_MetaSTAARlite`, [26](#)
`ncRNA_MetaSTAARlite_cond`, [27](#)
`ncRNA_MetaSTAARlite_worker`, [27](#), [28](#), [29](#)
`noncoding_MetaSTAARlite`, [31](#)
`noncoding_MetaSTAARlite_cond`, [33](#)
`noncoding_MetaSTAARlite_worker`, [32](#), [34](#),
[34](#)

`STAAR`, [17](#)
`STAARpipeline`, [2](#), [4](#), [6](#), [11](#), [22](#), [24](#), [25](#), [30](#), [35](#)