# Package 'RICE'

October 21, 2024

**Type** Package

**Title** Integrating Common and Rare Variants Improves Polygenic Risk Prediction Across Diverse Populations

**Version** 0.1.0

**Author** Jacob Williams [aut, cre]

**Maintainer** Jacob Williams <jacob.williams@nih.gov>

**Description** Creates a rare variant polygenic risk score using burden scores extracted from a ADGS file. Burden scores are extracted for specific rare variant sets using the Burden_Score function. A rare variant PRS is constructed using the RareVariantPRS function.

**License** GPL-3

**Imports** Rcpp, SCANG, dplyr, SeqArray, SeqVarTools, GenomicFeatures, TxDb.Hsapiens.UCSC.hg38.knownGene, GMMAT, GENESIS, Matrix, methods, caret, glmnet, SuperLearner, STAAR, STAARpipeline

**Encoding** UTF-8

**LazyData** true

**Depends** R (>= 4.0.0)

**RoxygenNote** 7.3.2

**Suggests** knitr, rmarkdown

**VignetteBuilder** knitr

## Contents



---

`Burden_Scores` *Extract Burden Scores from a aGDS*

---

### Description

Extracts a non-weighted Burden Score; b = G x 1, for a given chromosome, protein-coding gene, and functional category for rare variants in either the coding region or noncoding region.

## Usage

```
Burden_Scores(
  region = c("Coding", "Noncoding"),
  chr,
  gene_name,
 category = c("plof", "plof_ds", "missense", "disruptive_missense", "synonymous",
  "downstream", "upstream", "UTR", "promoter_CAGE", "promoter_DHS", "enhancer_CAGE",
    "enhancer_DHS", "ncRNA"),
  genofile,
  IDs,
  rare_maf_cutoff = 0.01,
  rv_num_cutoff = 2,
  QC_label = "annotation/filter",
  variant_type = c("SNV", "Indel", "variant"),
  geno_missing_imputation = c("mean", "minor"),
  Annotation_dir = "annotation/info/FunctionalAnnotation",
  Annotation_name_catalog,
  silent = FALSE
)
```

## Arguments

| | |
|---|---|
| region | A character value specifying the region of rare variants, either Coding or Noncoding. Coding region contains 5 functional categories plof, plof_ds, missense, disruptive_missense, or synonymous and noncoding region contains 8 functional categories downstream, upstream, UTR, promoter_CAGE, promoter_DHS, enhancer_CAGE, enhancer_DHS, or ncRNA. |
| chr | A numeric value specifying chromosome in which the protein-coding gene lies in. |
| gene_name | A character value specifying the protein-coding gene. |
| category | A character value specifying the functional category in the protein-coding gene for which the burden score will be constructed from. When region is specified as Coding, choices include plof, plof_ds, missense, disruptive_missense, or synonymous and when region is specified as Noncoding choices include downstream, upstream, UTR, promoter_CAGE, promoter_DHS, enhancer_CAGE, enhancer_DHS, or ncRNA. |
| genofile | The opened AGDS file using SeqOpen in SeqArray. |
| IDs | Subject IDs of individuals to construct a burden score with. Should either be a character or numeric vector matching the type from SeqGetData(genofile,"sample.id"). |
| rare_maf_cutoff | The maximum allele frequency for defining rare variants to be included in the construction of the burden score (default = 0.01). |
| rv_num_cutoff | A numeric value specifying the cutoff of minimum number of variants to include in the construction of a burden score (default = 2). |
| QC_label | Character value specifying the channel name of the QC label in the GDS/aGDS file (default = "annotation/filter"). |
| variant_type | Character value specifying the type of variant included in the analysis. Choices include "SNV", "Indel", or "variant" (default = "SNV"). |

`geno_missing_imputation`
: Character value specifying the method of handling missing genotypes. Either "mean" or "minor" (default = "mean").

`Annotation_dir`
: Character value specifying the channel name of the annotations in the aGDS file (default = "annotation/info/FunctionalAnnotation").

`Annotation_name_catalog`
: A data frame containing the name and the corresponding channel name in the aGDS file.

`silent`
: Logical: should the report of error messages be suppressed (default = FALSE).

## Value

A column vector containing the burden score for a given chromosome, protein-coding gene, and functional category for rare variants in either the coding or noncoding region.

# Index