

Package ‘ccLRforR’

March 3, 2026

Type Package

Title Case-Control Likelihood Ratio (ccLR)

Version 1.0

Description Implementation of case-control data analysis using likelihood ratio approaches and logistic regression for the classification of variants of uncertain significance (VUS) in breast, ovarian, or custom cancer susceptibility genes.

License GPL-2

Encoding UTF-8

LazyData true

Depends R (>= 3.5)

Imports Rcpp, dplyr, tidyr, utils, stats

LinkingTo Rcpp (>= 1.0.13)

NeedsCompilation yes

RoxygenNote 7.3.3

VUS Volume Under the ROC Surface

ccLR case-control Likelihood Ratio

Author Damianos Michaelides [aut, cre],
Maria Zanti [aut],
Christian Carrizosa [aut],
Theodora Nearchou [aut],
Kyriaki Michailidou [aut]

Maintainer Damianos Michaelides <damianosm@cing.ac.cy>

Repository CRAN

Date/Publication 2026-03-03 10:00:14 UTC

Contents

ccLR.grid	2
ccLRforR	6

incidence_data	8
ps4.ccLR	9
ps4.logistic	13
risk_rates	16

Index	19
--------------	-----------

ccLR.grid	<i>Refined Case-Control Likelihood Ratio (ccLR) Analysis for performing Grid Search by Scaling the Relative Risk</i>
-----------	--

Description

This function performs a grid-search case-control likelihood ratio analysis based on input genotype and phenotype data, optionally stratifying the results by country, ethnicity, or study. The function supports predefined or custom gene risk rates and allows for a grid scaling of the penetrance to investigate what magnitude of relative risk best fits the data.

Usage

```
ccLR.grid(cancer = c("breast", "ovarian", "custom"),
          gene = c("BRCA1", "BRCA2", "PALB2", "CHEK2", "ATM", "TP53", "custom"),
          genotypes,
          geno_notation = c("n", "n/n"),
          phenotype,
          grid = seq(0.5, 2, by=0.5),
          penetrance = c("Dorling", "Kuchenbaecker", "Antoniou", "Fortuno",
                        "Li", "Hall", "Yang", "Momozawa", "custom"),
          custom_penetrance = NULL,
          incidence_rate = c("England", "USA", "Japan", "Finland", "custom"),
          custom_incidence = NULL,
          outdir = NULL,
          output = "ccLR",
          stratifyby = NULL,
          agefilter = c(0, 80),
          exportcsv = FALSE,
          progress = FALSE
        )
```

Arguments

cancer	A character string specifying the cancer type under investigation. Options are "breast", "ovarian" or "custom" only.
gene	A character string specifying the gene of interest. Options are "BRCA1", "BRCA2", "PALB2", "CHEK2", "ATM", "TP53" or "custom" only.
genotypes	A data frame containing genotype data with the first column named "sample_ids" and subsequent columns for genotype information.

geno_notation	A character string specifying the format of the genotypes notation. Options are "n", or "n/n" only. In context, if variants take entries 0 (homozygous reference), 1 (heterozygous), 2 (homozygous alternate), and -1 (missing) then choose geno_notation="n". Alternatively, if variants take entries 0/0 (homozygous reference), 0/1 (heterozygous), 1/1 (homozygous alternate), and ./ (missing) then choose geno_notation="n/n". For other formats, please transform your dataset to one of the accepted/implemented formats.
phenotype	A data frame containing phenotype data. The required columns depend on the stratifyby parameter. If single strata is considered, i.e., if stratifyby=NULL, the data frame must include columns "sample_ids", "status", "ageInt", "AgeDiagIndex". If stratification is considered, the data frame must have an additional stratification column ("StudyCountry", "ethnicityClass", or "study") depending on the stratification variable.
grid	Optional. A vector of grid/scaling parameters that is applied to the age-specific relative risk curve. It represents how much more (or less) penetrant a specific variant may be compared to the average pathogenic variant for the same gene. For example: at 1 it assumes average gene-level pathogenicity, at 2 it assumes double the risk, and at 0.5 it assumes half the risk. Defaults to a sequence from 0.5 to 2 by 0.5 increments.
penetrance	A character string specifying the penetrance method. Options are "Dorling", "Kuchenbaecker", "Antoniou", "Fortuno", "Li", "Hall", "Yang" or "custom". Dorling contains breast rates for genes BRCA1, BRCA2, PALB2, CHEK2, and ATM. Kuchenbaecker contains breast and ovarian cancer rates for BRCA1 and BRCA2. Antoniou contains breast cancer rates for BRCA1, BRCA2, and PALB2. Fortuno and Li contain breast cancer rates for TP53. Hall contains ovarian cancer rates for ATM. Yang contains ovarian cancer rates for PALB2. If penetrance is set to "custom" the next argument "custom_penetrance" must be specified.
custom_penetrance	A data frame containing user-specified age-specific penetrance rates for variant carriers. Defaults to NULL but must be specified if penetrance = "custom". The required column structure depends on the values of cancer and gene: <ul style="list-style-type: none"> • If gene = "custom", the data frame must contain exactly two columns: "Age" and "Penetrance_Carriers". • If cancer = "custom" and gene is not "custom", the data frame must contain exactly two columns: "Age" and "Penetrance_Carriers_<gene>". • If cancer is "breast" or "ovarian" and gene is not "custom", the data frame must contain exactly two columns: "Age" and "BC_Penetrance_Carriers_<gene>" (for breast cancer) or "OC_Penetrance_Carriers_<gene>" (for ovarian cancer). <p>Column names are case-sensitive and no additional columns are permitted.</p>
incidence_rate	A character string specifying the population incidence rates to be used in the analysis. Supported options are: "England", "USA", "Japan", "Finland", or "custom". If incidence_rate is set to "custom" the next argument "custom_incidence" must be specified.

custom_incidence	<p>A data frame containing user-specified age-specific incidence rates. Defaults to NULL but must be specified if <code>incidence_rate = "custom"</code>.</p> <p>The data frame must contain exactly two columns:</p> <ul style="list-style-type: none"> • "Age": Age (in years). • "Incidence_rates": Population incidence rate at the corresponding age. <p>Column names are case-sensitive and no additional columns are permitted.</p>
outdir	Optional. A character string specifying the output directory. The default is set to NULL and in this case the output file containing the results is stored to a temporary file. To specify a permanent location this argument needs be specified.
output	Optional. A character string specifying the output file name. Defaults to "ccLR".
stratifyby	Optional. A character string specifying the stratification variable. Options are "country", "ethnicity", or "study", or NULL for single strata. The default entry is NULL.
agefilter	A numeric vector of length 2 specifying the age range to include in the analysis. Defaults to ages 0 to 80.
exportcsv	Optional. A logical value indicating whether to export the results as a CSV file (on top of printing the results in R). Defaults to FALSE.
progress	Optional. If TRUE, it returns the progress of the variants analysed. The default entry is FALSE.

Details

The function implements a grid-search case-control likelihood ratio methodology for different genetic variants and optionally stratifies results by the specified variable. The grid search ccLR approach makes use of the grid scaling values (which are subject to choice) to scale the gene-specific average relative risk and identify what risk best fits the case-control data. Only samples diagnosed or interviewed between the ages of 21 and 80 are included in the analysis. The likelihood ratios derived are evaluated against the ACMG/AMP thresholds.

Value

A data frame containing the results of the case-control likelihood ratio analysis. If `exportcsv = TRUE`, the results are saved as a CSV file.

Author(s)

Damianos Michaelides <damianosm@cing.ac.cy>, Maria Zanti, Christian Carrizosa, Theodora Nearchou, Kyriaki Michailidou

References

Antoniou, A., Pharoah, P. D. P., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., et al. (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet.* 72, 1117–1130.

Antoniou, A. C., Casadei, S., Heikkinen, T., Barrowdale, D., Pylkas, K., Roberts, J., ... and Tischkowitz, M. (2014). Breast-cancer risk in families with mutations in PALB2. *New England Journal of Medicine*, 371(6), 497-506.

Dorling, L. et al. (2021). Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med* 384, 428-439.

Fortuno, C., Feng, B. J., Carroll, C., Innella, G., Kohlmann, W., Lázaro, C., ..., and Spurdle, A. B. (2024). Cancer risks associated with TP53 pathogenic variants: Maximum likelihood analysis of extended pedigrees for diagnosis of first cancers beyond the Li-Fraumeni syndrome spectrum. *JCO Precision Oncology*, 8, e2300453.

Hall, M. J., Bernhisel, R., Hughes, E., Larson, K., Rosenthal, E. T., Singh, N. A., ... & Kurian, A. W. (2021). Germline pathogenic variants in the ataxia telangiectasia mutated (ATM) gene are associated with high and moderate risks for multiple cancers. *Cancer Prevention Research*, 14(4), 433-440.

Kuchenbaecker, K. B. J. L. Hopper, D. R. Barnes et al. (2017). Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA*, vol. 317, no. 23, pp. 2402–2416.

Li, S., MacInnis, R. J., Lee, A., Nguyen-Dumont, T., Dorling, L., Carvalho, S., ..., and Antoniou, A. C. (2022). Segregation analysis of 17,425 population-based breast cancer families: evidence for genetic susceptibility and risk prediction. *The American Journal of Human Genetics*, 109(10), 1777-1788.

Yang, X., Leslie, G., Doroszuk, A., Schneider, S., Allen, J., Decker, B., ... & Tischkowitz, M. (2020). Cancer risks associated with germline PALB2 pathogenic variants: an international study of 524 families. *Journal of clinical oncology*, 38(7), 674-685.

Zanti, M. et al. (2023). A likelihood ratio approach for utilizing case-control data in the clinical classification of rare sequence variants: application to BRCA1 and BRCA2. *Hum Mutat*.

Zanti M et al. (2025). Analysis of more than 400,000 women provides case-control evidence for BRCA1 and BRCA2 variant classification. *Nature Communications*.

Momozawa, Y., Sasai, R., Usui, Y., Shiraishi, K., Iwasaki, Y., Taniyama, Y., ... & Kubo, M. (2022). Expansion of cancer risk profile for BRCA1 and BRCA2 pathogenic variants. *JAMA oncology*, 8(6), 871-878.

Examples

```
## Define simulated inputs - genotypes and phenotype

genotypes <- data.frame(
  sample_ids = 1:100,
  variant1 = rbinom(100, 2, 0.3),
  variant2 = rbinom(100, 2, 0.2)
)

phenotype <- data.frame(
  sample_ids = 1:100,
  status = rbinom(100, 1, 0.5),
  ageInt = floor(runif(100, 21, 80)),
  AgeDiagIndex = floor(runif(100, 21, 80)),
  StudyCountry = sample(c("USA", "UK", "Canada"), 100, replace = TRUE)
```

```
)  
  
# Run the function  
ccLR.grid(  
  cancer = "breast",  
  gene = "PALB2",  
  genotypes = genotypes,  
  geno_notation="n",  
  phenotype = phenotype,  
  penetrance = "Antoniou",  
  incidence_rate = "England",  
  stratifyby = "country",  
  exportcsv = TRUE,  
  progress = FALSE  
)
```

ccLRforR

Case-Control Likelihood Ratio (ccLR)

Description

This package provides tools for implementing the case-control likelihood ratio (ccLR) analyses and logistic regression applying the PS4 criteria for genetic data, supporting optional stratification by country, ethnicity, or study. It includes functionality for built-in or custom cancer gene risk rates.

The package is designed for researchers analysing genetic breast, ovarian, or custom cancer data using the standard and grid-search ccLR approaches or a logistic regression applying the PS4 criteria.

The ccLR method compares the likelihood of the distribution of the variant of interest among cases and controls, under the hypothesis that the variant is associated with similar risks of the disease in question, as the "average" pathogenic variant, compared to the likelihood under the hypothesis that it is a benign variant not associated with increased risk.

The grid search ccLR approach makes use of the grid scaling values (which are subject to choice) to scale the gene-specific average relative risk and identify what risk best fits the case-control data.

The package includes functions for:

- ccLR analysis for built-in gene-specific and age-specific penetrances ([ps4.ccLR](#)).
- Grid search ccLR analysis that uses grid parameters to scale the penetrance levels and find which best fits the data ([ccLR.grid](#)).
- Likelihood Ratio Test via Logistic Regression to evaluate variants' pathogenicity against the PS4 criteria ([ps4.logistic](#)).

Key features:

- Optional stratification by country, ethnicity, or study.
- Built-in penetrances by Dorling et. al., 2021, Kuchenbaecker et. al., 2017, Antoniou et. al., 2003, Fortuno et. al., 2024, Li et. al., 2022, Hall et. al., 2021, and Yang et. al., 2020.
- Support for custom cancer types, genes penetrances, and incidence rates.
- Exclusion of samples outside the age range 21-80.

Author(s)

Damianos Michaelides [aut, cre], Maria Zanti [aut], Christian Carrizosa [aut], Theodora Nearchou [aut], Kyriaki Michailidou [aut]

Maintainer: Damianos Michaelides <damianosm@cing.ac.cy>

References

Antoniou, A., Pharoah, P. D. P., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., et al. (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet.* 72, 1117–1130.

Antoniou, A. C., Casadei, S., Heikkinen, T., Barrowdale, D., Pylkas, K., Roberts, J., ... and Tischkowitz, M. (2014). Breast-cancer risk in families with mutations in PALB2. *New England Journal of Medicine*, 371(6), 497-506.

Dorling, L. et al. (2021). Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med* 384, 428-439.

Fortuno, C., Feng, B. J., Carroll, C., Innella, G., Kohlmann, W., Lázaro, C., ..., and Spurdle, A. B. (2024). Cancer risks associated with TP53 pathogenic variants: Maximum likelihood analysis of extended pedigrees for diagnosis of first cancers beyond the Li-Fraumeni syndrome spectrum. *JCO Precision Oncology*, 8, e2300453.

Kuchenbaecker, K. B. J. L. Hopper, D. R. Barnes et al. (2017). Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA*, vol. 317, no. 23, pp. 2402–2416.

Li, S., MacInnis, R. J., Lee, A., Nguyen-Dumont, T., Dorling, L., Carvalho, S., ..., and Antoniou, A. C. (2022). Segregation analysis of 17,425 population-based breast cancer families: evidence for genetic susceptibility and risk prediction. *The American Journal of Human Genetics*, 109(10), 1777-1788.

Parsons, M. T. et al. (2024). Evidence-based recommendations for gene-specific ACMG/AMP variant classification from the ClinGen ENIGMA BRCA1 and BRCA2 Variant Curation Expert Panel. *Am J Hum Genet*.

Zanti, M. et al. (2023). A likelihood ratio approach for utilizing case-control data in the clinical classification of rare sequence variants: application to BRCA1 and BRCA2. *Hum Mutat*.

Zanti M et al. (2025). Analysis of more than 400,000 women provides case-control evidence for BRCA1 and BRCA2 variant classification. *Nature Communications*.

See Also

[ps4.cclR](#), [ps4.logistic](#), [cclR.grid](#)

incidence_data	<i>Age-specific population incidence rates for breast and ovarian cancer</i>
----------------	--

Description

Age-specific population incidence rates used to compute baseline hazards in the likelihood ratio model.

The incidence rates data are stored with one row per age (from 0 to 80 years) and separate columns for each cancer type and population.

The first column is:

- "Age": Age in years.

The remaining columns follow the naming convention: "<Cancer>_<Population>", where:

- <Cancer> is "BC" (breast cancer) or "OC" (ovarian cancer)
- <Population> is one of "England", "USA", "Finland", or "Japan"

Each cell contains the annual incidence rate for the corresponding age, cancer type, and population. These rates are used to construct cumulative baseline hazards in the likelihood calculations.

Usage

```
incidence_data
```

Format

A data frame with 81 rows and 9 columns containing age-specific annual population incidence rates.

Age Age in years (integer, ranging from 0 to 80).

BC_England Breast cancer incidence rate for England at the given age.

OC_England Ovarian cancer incidence rate for England at the given age.

BC_Finland Breast cancer incidence rate for Finland at the given age.

OC_Finland Ovarian cancer incidence rate for Finland at the given age.

BC_USA Breast cancer incidence rate for the USA at the given age.

OC_USA Ovarian cancer incidence rate for the USA at the given age.

BC_Japan Breast cancer incidence rate for Japan at the given age.

OC_Japan Ovarian cancer incidence rate for Japan at the given age.

Examples

```
## Load the incidence data
data(incidence_data)
head(incidence_data)
```

Description

This function performs the case-control likelihood ratio analysis based on input genotype and phenotype data, optionally stratifying the results by country, ethnicity, or study. The function supports predefined or custom gene risk rates.

Usage

```
ps4.ccLR(cancer = c("breast", "ovarian", "custom"),
         gene = c("BRCA1", "BRCA2", "PALB2", "CHEK2", "ATM", "TP53", "custom"),
         genotypes,
         geno_notation = c("n", "n/n"),
         phenotype,
         penetrance = c("Dorling", "Kuchenbaecker", "Antoniou", "Fortuno",
                       "Li", "Hall", "Yang", "Momozawa", "custom"),
         custom_penetrance = NULL,
         incidence_rate = c("England", "USA", "Japan", "Finland", "custom"),
         custom_incidence = NULL,
         outdir = NULL,
         output = "ccLR",
         stratifyby = NULL,
         agefilter = c(0, 80),
         exportcsv = FALSE,
         progress = FALSE
        )
```

Arguments

- | | |
|---------------|---|
| cancer | A character string specifying the cancer type under investigation. Options are "breast", "ovarian" or "custom" only. |
| gene | A character string specifying the gene of interest. Options are "BRCA1", "BRCA2", "PALB2", "CHEK2", "ATM", "TP53" or "custom" only. |
| genotypes | A data frame containing genotype data with the first column named "sample_ids" and subsequent columns for genotype information. |
| geno_notation | A character string specifying the format of the genotypes notation. Options are "n", or "n/n" only. In context, if variants take entries 0 (homozygous reference), 1 (heterozygous), 2 (homozygous alternate), and -1 (missing) then choose geno_notation="n". Alternatively, if variants take entries 0/0 (homozygous reference), 0/1 (heterozygous), 1/1 (homozygous alternate), and ./ (missing) then choose geno_notation="n/n". For other formats, please transform your dataset to one of the accepted/implemented formats. |

phenotype	A data frame containing phenotype data. The required columns depend on the stratifyby parameter. If single strata is considered, i.e., if stratifyby=NULL, the data frame must include columns "sample_ids", "status", "ageInt", "AgeDiagIndex". If stratification is considered, the data frame must have an additional stratification column ("StudyCountry", "ethnicityClass", or "study") depending on the stratification variable.
penetrance	A character string specifying the penetrance method. Options are "Dorling", "Kuchenbaecker", "Antoniou", "Fortuno", "Li", "Hall", "Yang" or "custom". Dorling contains breast rates for genes BRCA1, BRCA2, PALB2, CHEK2, and ATM. Kuchenbaecker contains breast and ovarian cancer rates for BRCA1 and BRCA2. Antoniou contains breast cancer rates for BRCA1, BRCA2, and PALB2. Fortuno and Li contain breast cancer rates for TP53. Hall contains ovarian cancer rates for ATM. Yang contains ovarian cancer rates for PALB2. If penetrance is set to "custom" the next argument "custom_penetrance" must be specified.
custom_penetrance	A data frame containing user-specified age-specific penetrance rates for variant carriers. Defaults to NULL but must be specified if penetrance = "custom". The required column structure depends on the values of cancer and gene: <ul style="list-style-type: none"> • If gene = "custom", the data frame must contain exactly two columns: "Age" and "Penetrance_Carriers". • If cancer = "custom" and gene is not "custom", the data frame must contain exactly two columns: "Age" and "Penetrance_Carriers_<gene>". • If cancer is "breast" or "ovarian" and gene is not "custom", the data frame must contain exactly two columns: "Age" and "BC_Penetrance_Carriers_<gene>" (for breast cancer) or "OC_Penetrance_Carriers_<gene>" (for ovarian cancer). <p>Column names are case-sensitive and no additional columns are permitted.</p>
incidence_rate	A character string specifying the population incidence rates to be used in the analysis. Supported options are: "England", "USA", "Japan", "Finland", or "custom". If incidence_rate is set to "custom" the next argument "custom_incidence" must be specified.
custom_incidence	A data frame containing user-specified age-specific incidence rates. Defaults to NULL but must be specified if incidence_rate = "custom". The data frame must contain exactly two columns: <ul style="list-style-type: none"> • "Age": Age (in years). • "Incidence_rates": Population incidence rate at the corresponding age. <p>Column names are case-sensitive and no additional columns are permitted.</p>
outdir	Optional. A character string specifying the output directory. The default is set to NULL and in this case the output file containing the results is stored to a temporary file. To specify a permanent location this argument needs be specified.
output	Optional. A character string specifying the output file name. Defaults to "ccLR".
stratifyby	Optional. A character string specifying the stratification variable. Options are "country", "ethnicity", or "study", or NULL for single strata. The default entry is NULL.

agefilter	A numeric vector of length 2 specifying the age range to include in the analysis. Defaults to ages 0 to 80.
exportcsv	Optional. A logical value indicating whether to export the results as a CSV file (on top of printing the results in R). Defaults to FALSE.
progress	Optional. If TRUE, it returns the progress of the variants analysed. The default entry is FALSE.

Details

The function implements the case-control likelihood ratio methodology for different genetic variants and stratifies results by the specified variable. It validates inputs, applies the calculations based on the chosen method, and generates a summary of the results. Only samples diagnosed or interviewed between the ages of 21 and 80 are included in the analysis. The likelihood ratios derived are evaluated against the ACMG/AMP thresholds. For the grid search ccLR approach, see [ccLR.grid](#).

Value

A data frame containing the results of the case-control likelihood ratio analysis. If `exportcsv = TRUE`, the results are saved as a CSV file in the directory set by `outdir`.

Author(s)

Damianos Michaelides <damianosm@cing.ac.cy>, Maria Zanti, Christian Carrizosa, Theodora Nearchou, Kyriaki Michailidou

References

- Antoniou, A., Pharoah, P. D. P., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., et al. (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet.* 72, 1117–1130.
- Antoniou, A. C., Casadei, S., Heikkinen, T., Barrowdale, D., Pylkas, K., Roberts, J., ... and Tischkowitz, M. (2014). Breast-cancer risk in families with mutations in PALB2. *New England Journal of Medicine*, 371(6), 497-506.
- Dorling, L. et al. (2021). Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med* 384, 428-439.
- Fortuno, C., Feng, B. J., Carroll, C., Innella, G., Kohlmann, W., Lázaro, C., ..., and Spurdle, A. B. (2024). Cancer risks associated with TP53 pathogenic variants: Maximum likelihood analysis of extended pedigrees for diagnosis of first cancers beyond the Li-Fraumeni syndrome spectrum. *JCO Precision Oncology*, 8, e2300453.
- Hall, M. J., Bernhisel, R., Hughes, E., Larson, K., Rosenthal, E. T., Singh, N. A., ... & Kurian, A. W. (2021). Germline pathogenic variants in the ataxia telangiectasia mutated (ATM) gene are associated with high and moderate risks for multiple cancers. *Cancer Prevention Research*, 14(4), 433-440.
- Kuchenbaecker, K. B. J. L. Hopper, D. R. Barnes et al. (2017). Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA*, vol. 317, no. 23, pp. 2402–2416.

Li, S., MacInnis, R. J., Lee, A., Nguyen-Dumont, T., Dorling, L., Carvalho, S., ..., and Antoniou, A. C. (2022). Segregation analysis of 17,425 population-based breast cancer families: evidence for genetic susceptibility and risk prediction. *The American Journal of Human Genetics*, 109(10), 1777-1788.

Yang, X., Leslie, G., Doroszuk, A., Schneider, S., Allen, J., Decker, B., ... & Tischkowitz, M. (2020). Cancer risks associated with germline PALB2 pathogenic variants: an international study of 524 families. *Journal of clinical oncology*, 38(7), 674-685.

Zanti, M. et al. (2023). A likelihood ratio approach for utilizing case-control data in the clinical classification of rare sequence variants: application to BRCA1 and BRCA2. *Hum Mutat*.

Zanti M et al. (2025). Analysis of more than 400,000 women provides case-control evidence for BRCA1 and BRCA2 variant classification. *Nature Communications*.

Momozawa, Y., Sasai, R., Usui, Y., Shiraishi, K., Iwasaki, Y., Taniyama, Y., ... & Kubo, M. (2022). Expansion of cancer risk profile for BRCA1 and BRCA2 pathogenic variants. *JAMA oncology*, 8(6), 871-878.

Examples

```
## Define simulated inputs - genotypes and phenotype

genotypes <- data.frame(
  sample_ids = 1:100,
  variant1 = rbinom(100, 2, 0.3),
  variant2 = rbinom(100, 2, 0.2)
)

phenotype <- data.frame(
  sample_ids = 1:100,
  status = rbinom(100, 1, 0.5),
  ageInt = floor(runif(100, 21, 80)),
  AgeDiagIndex = floor(runif(100, 21, 80)),
  StudyCountry = sample(c("USA", "UK", "Canada"), 100, replace = TRUE)
)

# Run the function
ps4.ccLR(
  cancer = "breast",
  gene = "BRCA1",
  genotypes = genotypes,
  geno_notation="n",
  phenotype = phenotype,
  penetrance = "Dorling",
  incidence_rate = "England",
  stratifyby = "country",
  exportcsv = TRUE,
  progress = TRUE
)
```

ps4.logistic

*Logistic Regression PS4 Criterion***Description**

This function performs logistic regression, calculates the likelihood ratio test, the odds ratio, and confidence intervals around it to compare against gene-specific PS4 criteria. The function performs based on input genotype and phenotype data. The factors assessed in the model are the ages and an optional stratification factor which is either country or ethnicity.

Usage

```
ps4.logistic(
  gene = c("BRCA1", "BRCA2", "PALB2", "CHEK2", "ATM", "TP53", "custom"),
  genotypes,
  geno_notation = c("n", "n/n"),
  phenotype,
  custom_rules = NULL,
  outdir = NULL,
  output = "PS4",
  stratifyby = NULL,
  agefilter = c(0, 80),
  exportcsv = FALSE,
  progress = FALSE
)
```

Arguments

gene	A character string specifying the gene of interest. Options are "BRCA1", "BRCA2", "PALB2", "CHEK2", "ATM", "TP53" and custom.
genotypes	A data frame containing genotype data with the first column named "sample_ids" and subsequent columns for genotype information.
geno_notation	A character string specifying the format of the genotypes notation. Options are "n", or "n/n" only. In context, if variants take entries 0 (homozygous reference), 1 (heterozygous), 2 (homozygous alternate), and -1 (missing) then choose geno_notation="n". Alternatively, if variants take entries 0/0 (homozygous reference), 0/1 (heterozygous), 1/1 (homozygous alternate), and ./ (missing) then choose geno_notation="n/n". For other formats, please transform your dataset to one of the accepted/implemented formats.
phenotype	A data frame containing phenotype data. The required columns depend on the stratifyby parameter. If single strata is considered, i.e., if stratifyby=NULL, the data frame must include columns "sample_ids", "status", "ageInt", "AgeDiagIndex". If stratification is considered, the data frame must have an additional stratification column ("StudyCountry", "ethnicityClass", or "study") depending on the stratification variable.

custom_rules	Optional. A named list of functions that define user-specified PS4 decision rules for one or more genes. Each function must return "Yes" or "No" when evaluated, and will be passed the arguments 'OR', 'LCI', 'UCI', and 'pval'. By default, hard-coded thresholds for BRCA1, BRCA2, ATM, CHEK2, PALB2, and TP53 are applied (see Details). Supplying a 'custom_rules' list allows users to: (a) Override the default criteria for one or more of these genes, and (b) Define thresholds for "custom" genes. Check the Examples section for an example.
outdir	Optional. A character string specifying the output directory. The default is set to NULL and in this case the output file containing the results is stored to a temporary file. To specify a permanent location this argument needs to be specified.
output	Optional. A character string specifying the output file name. Defaults to "PS4".
stratifyby	A character string specifying the stratification variable. Options are "country", "ethnicity", or "study", or NULL for single strata. The default entry is NULL.
agefilter	A numeric vector of length 2 specifying the age range to include in the analysis. Defaults to ages 0 to 80.
exportcsv	Optional. A logical value indicating whether to export the results as a CSV file (on top of printing the results in R). Defaults to FALSE.
progress	Optional. If TRUE, it returns the progress of the variants analysed. The default entry is FALSE.

Details

The function implements the case-control likelihood ratio methodology for different genetic variants and stratifies results by the specified variable. It validates inputs, applies the calculations based on the chosen method, and generates a summary of the results. Only samples diagnosed or interviewed between the ages of 21 and 80 are included in the analysis.

The function implements ClinGen-specified, gene-specific criteria for applying the ACMG/AMP rule PS4 (case-control evidence of pathogenicity). It evaluates each variant using the odds ratio (OR), relative risk (RR), Wald confidence interval (CI), and association test p-value from logistic regression, and then applies thresholds that differ by gene. For BRCA1/2, PS4 is assigned when $p \leq 0.05$, $OR \geq 4$, and the 95

Value

A data frame containing the results of the logistic regression and likelihood ratio test analysis, evaluated against the PS4 criteria. If `exportcsv = TRUE`, the results are saved as a CSV file.

Author(s)

Damianos Michaelides <damianosm@cing.ac.cy>, Maria Zanti, Christian Carrizosa, Theodora Nearchou, Kyriaki Michailidou

References

Parsons, M. T. et al. Evidence-based recommendations for gene-specific ACMG/AMP variant classification from the ClinGen ENIGMA BRCA1 and BRCA2 Variant Curation Expert Panel. *Am J Hum Genet* (2024).

Zanti, M. et al. (2023). A likelihood ratio approach for utilizing case-control data in the clinical classification of rare sequence variants: application to BRCA1 and BRCA2. Hum Mutat.

Zanti M et al. (2025). Analysis of more than 400,000 women provides case-control evidence for BRCA1 and BRCA2 variant classification. Nature Communications.

Examples

```
## Example 1:
## Define simulated inputs - genotypes and phenotype

genotypes <- data.frame(
  sample_ids = 1:100,
  variant1 = rbinom(100, 2, 0.3),
  variant2 = rbinom(100, 2, 0.2)
)

phenotype <- data.frame(
  sample_ids = 1:100,
  status = rbinom(100, 1, 0.5),
  ageInt = floor(runif(100, 21, 80)),
  AgeDiagIndex = floor(runif(100, 21, 80)),
  StudyCountry = sample(c("USA", "UK", "Canada"), 100, replace = TRUE)
)

# Run the function

ps4.logistic(
  gene = "CHEK2",
  genotypes = genotypes,
  geno_notation="n",
  phenotype = phenotype,
  stratifyby = "country",
  exportcsv = TRUE,
  progress = FALSE
)

## Example 2:
## Define simulated inputs - genotypes and phenotype

genotypes <- data.frame(
  sample_ids = 1:100,
  variantX = rbinom(100, 2, 0.1)
)

phenotype <- data.frame(
  sample_ids = 1:100,
  status = rbinom(100, 1, 0.5),
  ageInt = floor(runif(100, 21, 80)),
  AgeDiagIndex = floor(runif(100, 21, 80)),
  ethnicityClass = sample(c("European", "Asian", "African"), 100, replace = TRUE)
)
```

```

## Define a custom rule for a "custom" gene:
### Flag "Yes" if OR >= 2.5 and CI lower bound >= 1.2

custom_rules <- list(
  CUSTOM = function() ifelse(OR >= 2.5 && LCI >= 1.2, "Yes", "No")
)

## Run the function
ps4.logistic(
  gene = "custom",
  genotypes = genotypes,
  geno_notation = "n",
  phenotype = phenotype,
  custom_rules = custom_rules,
  stratifyby = "ethnicity",
  exportcsv = FALSE,
  progress = TRUE,
)

```

risk_rates

Breast and Ovarian Cancer Risk Rates: Dorling et al. (2021), Kuchenbaecker et al. (2017), Antoniou et al. (2003), Fortuno et al. (2024), Li et al. (2022), Hall et al. (2021), Yang et al. (2020), and Momozawa et al. (2022)

Description

These datasets provide age-specific disease penetrances (relative risks) for breast and ovarian cancers. The datasets are derived from Dorling et al. (2021), Kuchenbaecker et al. (2017), Antoniou et al. (2003), Antoniou et al. (2014), Fortuno et al. (2024), Li et al. (2022), Hall et al. (2021), Yang et al. (2020), and Momozawa et al. (2022). The datasets are used for the calculation of the case-control likelihood ratio (ccLR) analyses of the BRCA1, BRCA2, PALB2, CHEK2, ATM, and TP53 genetic variants. Dorling contains breast cancer rates for genes BRCA1, BRCA2, PALB2, CHEK2, and ATM. Kuchenbaecker contains breast and ovarian cancer rates for BRCA1 and BRCA2. Antoniou contains breast cancer rates for BRCA1, BRCA2, and PALB2. Fortuno and Li contain breast cancer rates for TP53. Hall contains ovarian cancer rates for ATM. Yang contains ovarian cancer rates for PALB2. Momozawa contains breast and ovarian cancer rates for BRCA1 and BRCA2.

Usage

Dorling
 Kuchenbaecker
 Antoniou
 Fortuno
 Li
 Hall

Yang
Momozawa

Details

Each dataset is a data frame that contains:

Age Numeric. The age range or specific ages.

Relative_risk Numeric. The relative risk for carriers of the pre-mentioned gene variants.

References

- Antoniou, A., Pharoah, P. D. P., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., et al. (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am. J. Hum. Genet.* 72, 1117–1130.
- Antoniou, A. C., Casadei, S., Heikkinen, T., Barrowdale, D., Pylkas, K., Roberts, J., ... and Tischkowitz, M. (2014). Breast-cancer risk in families with mutations in PALB2. *New England Journal of Medicine*, 371(6), 497-506.
- Dorling, L. et al. (2021). Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med* 384, 428-439.
- Fortuno, C., Feng, B. J., Carroll, C., Innella, G., Kohlmann, W., Lázaro, C., ..., and Spurdle, A. B. (2024). Cancer risks associated with TP53 pathogenic variants: Maximum likelihood analysis of extended pedigrees for diagnosis of first cancers beyond the Li-Fraumeni syndrome spectrum. *JCO Precision Oncology*, 8, e2300453.
- Hall, M. J., Bernhisel, R., Hughes, E., Larson, K., Rosenthal, E. T., Singh, N. A., ... & Kurian, A. W. (2021). Germline pathogenic variants in the ataxia telangiectasia mutated (ATM) gene are associated with high and moderate risks for multiple cancers. *Cancer Prevention Research*, 14(4), 433-440.
- Kuchenbaecker, K. B. J. L. Hopper, D. R. Barnes et al. (2017). Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA*, vol. 317, no. 23, pp. 2402–2416.
- Li, S., MacInnis, R. J., Lee, A., Nguyen-Dumont, T., Dorling, L., Carvalho, S., ..., and Antoniou, A. C. (2022). Segregation analysis of 17,425 population-based breast cancer families: evidence for genetic susceptibility and risk prediction. *The American Journal of Human Genetics*, 109(10), 1777-1788.
- Yang, X., Leslie, G., Doroszuk, A., Schneider, S., Allen, J., Decker, B., ... & Tischkowitz, M. (2020). Cancer risks associated with germline PALB2 pathogenic variants: an international study of 524 families. *Journal of clinical oncology*, 38(7), 674-685.
- Momozawa, Y., Sasai, R., Usui, Y., Shiraishi, K., Iwasaki, Y., Taniyama, Y., ... & Kubo, M. (2022). Expansion of cancer risk profile for BRCA1 and BRCA2 pathogenic variants. *JAMA oncology*, 8(6), 871-878.

Examples

```
## Load the Dorling dataset
data(Dorling)
head(Dorling)

## Load the Kuchenbaecker dataset
data(Kuchenbaecker)
head(Kuchenbaecker)

## Load the Antoniou dataset
data(Antoniou)
head(Antoniou)

## Load the Fortuno dataset
data(Fortuno)
head(Fortuno)

## Load the Li dataset
data(Li)
head(Li)

## Load the Hall dataset
data(Hall)
head(Hall)

## Load the Yang dataset
data(Yang)
head(Yang)

## Load the Momozawa dataset
data(Momozawa)
head(Momozawa)
```

Index

- * **Breast Cancer research**

- ccLR.grid, 2
 - ccLRforR, 6
 - ps4.ccLR, 9
 - ps4.logistic, 13

- * **breast cancer research**

- incidence_data, 8
 - risk_rates, 16

- * **case-control likelihood ratio**

- ccLR.grid, 2
 - ccLRforR, 6
 - ps4.ccLR, 9
 - ps4.logistic, 13

- * **case-control**

- ccLR.grid, 2
 - ccLRforR, 6
 - ps4.ccLR, 9
 - ps4.logistic, 13

- * **genetic analysis**

- ccLRforR, 6

- * **genetic variants**

- ccLR.grid, 2
 - ps4.ccLR, 9
 - ps4.logistic, 13

- * **likelihood ratio**

- ccLR.grid, 2
 - ccLRforR, 6
 - ps4.ccLR, 9
 - ps4.logistic, 13

- * **ovarian cancer research**

- incidence_data, 8
 - risk_rates, 16

Antoniou (risk_rates), 16

ccLR.grid, 2, 6, 7, 11
ccLRforR, 6
ccLRforR-package (ccLRforR), 6

Dorling (risk_rates), 16

Fortuno (risk_rates), 16

Hall (risk_rates), 16

incidence_data, 8

Kuchenbaecker (risk_rates), 16

Li (risk_rates), 16

Momozawa (risk_rates), 16

ps4.ccLR, 6, 7, 9

ps4.logistic, 6, 7, 13

risk_rates, 16

Yang (risk_rates), 16