Package 'RVPedigree'

January 20, 2016

Title Methods for Family-Based Rare-Variant Genetic Association Tests

Version 0.0.3

Author Karim Oualkacha [aut, cre],
M'Hamed Lajmi Lakhal-Chaieb [aut],
Celia M.T. Greenwood [aut],
Lennart C. Karssen [aut],
Sodbo Sharapov [aut]

Maintainer Karim Oualkacha <oualkacha.karim@uqam.ca>

Description This is a collection of the five region-based rare-variant genetic association tests. The following tests are currently implemented: ASKAT, ASKAT-Normalized, VC-C1, VC-C2 and VC-C3.

Depends R (>= 3.2.2), foreach, doParallel, utils

Imports ks, CompQuadForm, Matrix, snpStats, kinship2

Suggests GenABEL, knitr, methods

License GPL (>= 3)

LazyData true

VignetteBuilder knitr

NeedsCompilation no

RoxygenNote 5.0.1

Repository CRAN

Date/Publication 2016-01-20 09:12:22

R topics documented:

`RVPedigree-package` *RVPedigree: A package for region-based genetic association tests.*

Description

The RVPedigree package contains three methods for doing region-based association tests. It contains methods to perform autosomal rare variant association analyses with a normally-distributed or non normally distributed quantitative phenotype, specifically for the situation when there are related individuals or families in the data. The methods included are:

- ASKAT (exact test)
- Normalized ASKAT (exact test)
- VC-C1 (exact test)
- VC-C2 (permutation test of VC-C1 statistic)
- VC-C3 (permutation test of VC-C1 statistic)

RVPedigree functions

The following functions are the user-visible functions of the package:

- `RVPedigree`: main function that runs the selected test (ASKAT by default) genome-wide. If you only want to run a given test on a single genomic region, use one of the following functions.
- `ASKAT.region`: Runs the ASKAT test on a given genomic region.
- `NormalizedASKAT.region`: Runs the Normalized ASKAT test on a given genomic region.
- `VCC1.region`: Runs the VC-C1 test on a given genomic region.
- `VCC2.region`: Runs the VC-C2 test on a given genomic region.
- `VCC3.region`: Runs the VC-C3 test on a given genomic region.
- `Estim.H0.ASKAT`: Estimates the null model for the ASKAT test. The result of this function can be passed to the corresponding `ASKAT.region` function to save computation time in case multiple genomics regions are to be analyzed.
- `Estim.H0.NormalizedASKAT`: Estimates the null model for the Normalized ASKAT test. The result of this function can be passed to the corresponding `NormalizedASKAT.region` function to save computation time in case multiple genomics regions are to be analyzed.

- `Estim.H0.VCC`: Estimates the null model for the VC-C1, VC-C2 and VC-C3 tests. The result of this function can be passed to the corresponding `VCC?.region` function to save computation time in case multiple genomics regions are to be analyzed.
- `GetRelMatrix`: calculates the relationship matrix (twice the kinship matrix) based on various types of input data.
- `readMapFile`: reads a genetic map file (e.g. from Plink data) and creates the correct data frame to pass on to the various `*.region` functions.
- `Normality.test`: function to test for the normality of the phenotype data.

See Also

`GenABEL`, `snpStats`

`ASKAT.region` *Run the ASKAT method on a genomic region defined by a start and a stop base pair coordinate*

Description

Runs the ASKAT method on a given genomic region

Usage

```
ASKAT.region(y = NULL, X = NULL, Phi = NULL, type = "bed",
  filename = NULL, map = NULL, chr = 0, startpos = 0, endpos = 0,
  regionname = NULL, U = NULL, S = NULL, RH.Null = NULL,
  weights = NULL)
```

Arguments

`y`	vector of phenotype data (one entry per individual), of length n.
`X`	matrix of covariates including intercept (dimension: $n \times p$, with p the number of covariates)
`Phi`	Relationship matrix (i.e. twice the kinship matrix); an $n \times n$ square symmetric positive-definite matrix.
`type`	character, `'ped'`, `'bed'` (default) or `'shapeit-haps'` format of input file containing haplotype data
`filename`	character, path to input file containing haplotype data
`map`	object, data.frame contains 3 columns: rsID, chromosome, position in bp as output by e.g. `readMapFile`.
`chr`	character, chromosome number (basically from 1 to 22 as used by Plink), on which the region of interest is located
`startpos`	numeric, start position (in bp, base pairs) of the region of interest (default: 0)
`endpos`	numeric, end position (in bp, base pairs) of the region of interest (default: 0)

`regionname` (character) Name of the region/gene on which you are running the association test. This name is used in the output of this function and can be used to distinguish different regions if this function is run multiple times.

`U` (optional) Matrix of Eigenvectors of the relationship matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$. If this parameter is not given, it will be computed, so when running this function for many regions time can be saved by specifying not only `Phi`, but also `S` and `U`.

`S` (optional) Matrix of Eigenvalues of the relationship matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$. If this parameter is not given, it will be computed, so when running this function for many regions, time can be saved by specifying not only `Phi`, but also `S` and `U`.

`RH.Null` (optional) output of `Estim.H0.ASKAT` function. In analyses of many regions, it is not necessary to calculate the the null hypothesis for each region. One estimation per trait is enough.

`weights` optional numeric vector of genotype weights. If this option is not specified, the beta distribution is used for weighting the variants, with each weight given by $w_i = dbeta(f_i, 1, 25)^2$, with f_i the minor allele frequency (MAF) of variant i. This default is the same as used by the SKAT package. This vector is used as the diagonal of the $m \times m$ matrix W, with m the number of variants.

Value

A data frame containing the results of the association test. The data frame contains the following columns:

- `Score.Test`: the score of the given association test
- `P.value`: the p-value of the association test
- `N.Markers`: the number of markers in the region
- `regionname`: Name of the region/gene on which you are running the association test

Author(s)

Lennart C. Karssen, Sodbo Sharapov

`Estim.H0.ASKAT` *Estimation of the variance components under the null model using the ASKAT method*

Description

Estimation of the variance components under the null model using the ASKAT method

Usage

```
Estim.H0.ASKAT(y, X, S, U)
```

Arguments

y	Vector of phenotype values
X	A matrix of covariates, including intercept.
S	Matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$.
U	Matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$.

Value

A vector with the following values:

- `s.e`: variance component due to the error term (residuals)
- `s.g`: variance component due to the polygenic residual (polygenic background)
- `beta`: the estimate of the effects of the covariate (if there are some in the data)

Author(s)

Karim Oualkacha

M'Hamed Lajmi Lakhal-Chaieb

Estim.H0.NormalizedASKAT

Estimation of the variance components under the null model using the normalized ASKAT method

---

Description

Estimation of the variance components under the null model using the normalized ASKAT method

Usage

```
Estim.H0.NormalizedASKAT(y, X, S, U)
```

Arguments

y	Vector of phenotype values
X	A matrix of covariates, including intercept.
S	Matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$.
U	Matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$.

Value

A vector with the following values:

- `s.e`: variance component due to the error term (residuals)
- `s.g`: variance component due to the polygenic residual (polygenic background)
- `beta`: the estimate of the effects of the covariate (if there are some in the data)

Author(s)

Karim Oualkacha

M'Hamed Lajmi Lakhal-Chaieb

`Estim.H0.VCC` *Estimate the model parameters under the null model*

Description

This function estimates the model parameters under the null model when there is no region genotypes effect, for the VCC methods

Usage

```
Estim.H0.VCC(y, X, S, U)
```

Arguments

`y`	Vector of phenotype values
`X`	A matrix of covariates, including intercept.
`S`	Matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$.
`U`	Matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$.

Value

a list of:

- `h.hat`: estimate of the (narrow sense) heritability
- `gam.hat`: a vector of estimates of fixed effects of the covariates (if there are any).
- `residuals`: residuals from adjusting the null model with only covariates in the model.

Author(s)

Karim Oualkacha

M'Hamed Lajmi Lakhal-Chaieb

GetRelMatrix *Estimate relationship matrix based on pedigree or genomic data*

Description

Estimate relationship matrix based on pedigree or genomic data. This function can use either Plink files as input (for both pedigree-based and genomic relationship matrix calculation), or genetic data in GenABEL format (for genomic relationships).

Usage

```
GetRelMatrix(datatype = NULL, plinkbasefile = NULL, is.binary = FALSE,
  transpose = FALSE, header = FALSE, path2Plink = "plink",
  weight = "freq", gwaa.data = NULL, pedigreefile = NULL)
```

Arguments

`datatype`	character, `"pedigree"` or `"genomic"`. Estimate the relationship matrix using either pedigree information or genotype data
`plinkbasefile`	character, path to files in plink format. E.g. if you have files `test.ped` and `test.map`, plinkbasefile should be `test`. More details are also given in the vignette.
`is.binary`	logical, indicate whether the plink files are in binary format (`.bed`/`.bim`/`.fam`)
`transpose`	logical, indicate whether the plink text files are transposed or not (`.tped` and `.tfam` files)
`header`	logical, indicate whether the input text files have header or not
`path2Plink`	character, path to the binary (executable file) of plink_1.90 or later (the Plink_1.90 binary is used for efficient computation of the relationship matrix). More details are also given in the vignette.
`weight`	character, either `"no"` or `"freq"`. We suggest to use `weight="freq"`, which weighs by allelic frequency assuming HWE. See help for the `ibs()` function of the GenABEL package
`gwaa.data`	object, name of object gwaa.data-class, which is GenABEL genotype/phenotype data format
`pedigreefile`	reserved for future development

Details

Note that, with respect to the input parameters it is important to distinguish the two options of `datatype`. If `datatype` is equal to `"pedigree"` the user should specify the options `plinkbasefile`, `is.binary`, `transpose`, `header` and `path2Plink`. In that case, the kinship will be estimated based on the pedigree data in the Plink files. If, however, `datatype` is equal to `"genomic"`, there are two options: either one uses the parameter `gwaa.data` to tell this function to compute the genomics relationship matrix from a previously stored GenABEL data object. Or, if `gwaa.data` is empty, the Plink-related parameters have to be specified and the genomics relationship matrix will be computed based on the genetic data in the Plink file.

Value

matrix object which contains the estimated relationship matrix

Author(s)

Sodbo Sharapov

Examples

```
system.file("extdata", "2012.csv", package = "testdat")
## Not run:
pedRel <- GetRelMatrix(datatype="pedigree",
                       plinkbasefile=system.file("extdata",
                                                 "data",
                                          package="RVPedigree"),
                       transpose=FALSE)
pedRel <- GetRelMatrix(datatype="pedigree",
                       plinkbasefile=system.file("extdata",
                                                 "dataT",
                                          package="RVPedigree"),
                       transpose=TRUE)
pedRel <- GetRelMatrix(datatype="pedigree",
                       plinkbasefile=system.file("extdata",
                                                 "dataB",
                                          package="RVPedigree"),
                       is.binary=TRUE)
pedRel <- GetRelMatrix(datatype="pedigree",
                       plinkbasefile=system.file("extdata",
                                                 "OneFamilyExample",
                                          package="RVPedigree"),
                       transpose=FALSE, header=TRUE)
pedRel <- GetRelMatrix(datatype="pedigree",
                       plinkbasefile=system.file("extdata",
                                                 "TwoFamilyExample",
                                          package="RVPedigree"),
                       transpose=FALSE, header=TRUE)

load(system.file("extdata", "gwaa.data.RData")
genRel <- GetRelMatrix(datatype="genomic", gwaa.data=data1, weight="no")
genRel <- GetRelMatrix(datatype="genomic", gwaa.data=data1, weight="freq")
genRelError <- GetRelMatrix(datatype="genomic", gwaa.data=pedRel, weight="freq")

genPlinkRel <- GetRelMatrix(data="genomic",
                            path2Plink="plink_1.90",
                            system.file("extdata",
                                        "data",
                                         package="RVPedigree"),
genPlinkRel[1:10, 1:10]

genPlinkRel <- GetRelMatrix(data="genomic",
                            path2Plink="plink_1.90",
                            system.file("extdata",
```

```
                                        "dataT",
                                         package="RVPedigree"),
                            transpose=TRUE)
genPlinkRel[1:10, 1:10]

genPlinkRel <- GetRelMatrix(data="genomic",
                            path2Plink="plink_1.90",
                            system.file("extdata",
                                        "dataB",
                                         package="RVPedigree"),
                            is.binary=TRUE)
genPlinkRel[1:10, 1:10]

# GetRelMatrix(file="OneFamilyExample.ped", datatype="genomic")
# GetRelMatrix(file="OneFamilyExample.ped", datatype="pedasdfa")
# GetRelMatrix(file="OneFamilyExaample", datatype="pedigree")

## End(Not run)
```

`Normality.test` *Test for normality of the trait/phenotype*

Description

Test for normality of the trait

Usage

```
Normality.test(y = NULL, X = NULL, pedigree = NULL, plot = FALSE)
```

Arguments

`y`	vector of phenotype data (one entry per individual), of length n.
`X`	matrix of covariates including intercept (dimension: $n \times p$, with p the number of covariates)
`pedigree`	a pedigree as output by `read.pedigree`. This is a data frame consisting of four columns (family ID, individual ID, father ID and mother ID) as use in the traditional linkage format and e.g. Plink files.
`plot`	(logical) If set to TRUE a histogram will be plotted of the phenotype residuals after adjusting for covariates (default: FALSE).

Details

This function is used to test whether the phenotype is distributed normally, based on the Shapiro-Wilk test.

Value

A list with the following elements:

- `SW.pvalue`: Shapiro-Wilk p-value that indicates whether the phenotype is distributed normally.
- `resid.0`: the residuals after regressing the phenotype onto X.

Author(s)

Karim Oualkacha

`NormalizedASKAT.region`

Run the normalized ASKAT method on a genomic region defined by a start and a stop base pair coordinate

---

Description

Runs the normalized ASKAT method on a given genomic region. Rank-based normalization is applied to the phenotype residauls under the null model, after adjusting for covariate effects

Usage

```
NormalizedASKAT.region(y = NULL, X = NULL, Phi = NULL, type = "bed",
  filename = NULL, map = NULL, chr = 0, startpos = 0, endpos = 0,
  regionname = NULL, U = NULL, S = NULL, RH.Null = NULL,
  weights = NULL)
```

Arguments

`y`	vector of phenotype data (one entry per individual), of length n.
`X`	matrix of covariates including intercept (dimension: $n \times p$, with p the number of covariates)
`Phi`	Relationship matrix (i.e. twice the kinship matrix); an $n \times n$ square symmetric positive-definite matrix.
`type`	character, `'ped'`, `'bed'` (default) or `'shapeit-haps'` format of input file containing haplotype data
`filename`	character, path to input file containing haplotype data
`map`	object, data.frame contains 3 columns: rsID, chromosome, position in bp as output by e.g. `readMapFile`.
`chr`	character, chromosome number (basically from 1 to 22 as used by Plink), on which the region of interest is located
`startpos`	numeric, start position (in bp, base pairs) of the region of interest (default: 0)
`endpos`	numeric, end position (in bp, base pairs) of the region of interest (default: 0)

`regionname` (character) Name of the region/gene on which you are running the association test. This name is used in the output of this function and can be used to distinguish different regions if this function is run multiple times.

`U` (optional) Matrix of Eigenvectors of the relationship matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$. If this parameter is not given, it will be computed, so when running this function for many regions time can be saved by specifying not only `Phi`, but also `S` and `U`.

`S` (optional) Matrix of Eigenvalues of the relationship matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$. If this parameter is not given, it will be computed, so when running this function for many regions, time can be saved by specifying not only `Phi`, but also `S` and `U`.

`RH.Null` (optional) output of `Estim.H0.NormalizedASKAT` function. Practically, you don't need to calculate th enull hypothesis for every region. One estimation per trait is enough.

`weights` optional numeric vector of genotype weights. If this option is not specified, the beta distribution is used for weighting the variants, with each weight given by $w_i = dbeta(f_i, 1, 25)^2$, with f_i the minor allele frequency (MAF) of variant i. This default is the same as used by the SKAT package. This vector is used as the diagonal of the $m \times m$ matrix W, with m the number of variants.

Value

A data frame containing the results of the association test. The data frame contains the following columns:

- `Score.Test`: the score of the given association test
- `P.value`: the p-value of the association test
- `N.Markers`: the number of markers in the region
- `regionname`: Name of the region/gene on which you are running the association test

Author(s)

Lennart C. Karssen, Sodbo Sharapov

`readMapFile` *Read file with information about SNPs chromosome and position, for example from regular PLINK .map OR .bim file*

Description

These files have (at least) the following columns (separated by white space):

- column 1: chromosome
- column 2: variant name
- column 3: genetic distance in morgans (optional, see the `morgans` option)
- column 4: base-pair position (bp units)

Usage

```
readMapFile(filename = "NULL", morgans = TRUE)
```

Arguments

filename	character, path to input file containing genomic map data, e.g. a Plink .map, .bim or .tped file.
morgans	logical, indicate whether input file contains column with genetic distance between SNPs (in which case the bp positions are in the fourth column)

Value

matrix object with columns contains chromosome, rsID and position in bp

Author(s)

Sodbo Sharapov

See Also

read.haplo, read.haplo.pedfile, read.haplo.bedfile, read.haplo.shapeit_haps

---

RVPedigree *RVPedigree main function*

Description

Main function of the RVPedigree package

Usage

```
RVPedigree(method = "ASKAT", y = NULL, X = NULL, Phi = NULL,
  filename = NULL, type = "bed", regions = NULL, weights = NULL,
  Nperm = 100, pvalThreshold = 0.1, VCC3afterVCC1 = FALSE, Ncores = 1)
```

Arguments

method	character, selects the method to use for the association testing. Can be one of the following: • "ASKAT" (default) • "NASKAT", normalized ASKAT • "VCC1", VC-C1 • "VCC2", VC-C2 • "VCC3", VC-C3
y	vector of phenotype data (one entry per individual), of length n.

`X` matrix of covariates including intercept (dimension: $n \times p$, with p the number of covariates)

`Phi` Relationship matrix (i.e. twice the kinship matrix); an $n \times n$ square symmetric positive-definite matrix.

`filename` character, path to input file containing haplotype data

`type` character, `'ped'`, `'bed'` (default) or `'shapeit-haps'` format of input file containing haplotype data

`regions` a data frame with details of the genomic regions in which the association test specified by the `method` parameter should be run. The data frame should have one row per region and (at least) four columns with the following names:

- `Name`: Name of the region (e.g. `Gene 01`)
- `Chr`: Chromosome on which the region is located.
- `StartPos`: The base pair coordinate at which the region starts
- `EndPos`: The base pair coordinate at which the region ends.

Any other columns will be ignored.

`weights` optional numeric vector of genotype weights. If this option is not specified, the beta distribution is used for weighting the variants, with each weight given by $w_i = dbeta(f_i, 1, 25)^2$, with f_i the minor allele frequency (MAF) of variant i. This default is the same as used by the SKAT package. This vector is used as the diagonal of the $m \times m$ matrix W, with m the number of variants.

`Nperm` (integer) The number of permutations to be done to calculate the empirical p-value if the VCC2 or VCC3 method is used. For other methods this parameter is ignored (default: 100).

`pvalThreshold` (numeric) Threshold for the association p-value. Regions with a p-value below this threshold will not be present in the output data frame (default: 0.1).

`VCC3afterVCC1` (logical) Boolean value that indicates whether the VC-C3 method should automatically be run on the variants passing the p-value threshold set using the `pvalThreshold` parameter (default: FALSE).

`Ncores` (integer) Number of processor (CPU) cores to be used in parallel when doing running the association analysis. If the number of regions is larger than the number of cores, then each region gets to use maximum one core. If the number of cores is larger than the number of regions and the VCC2 or VCC3 methods are selected, the remaining cores are distributed among the regions to parallelize the permutations used to determine the p-value (default: 1).

Details

The RVPedigree function is the main function of the RVPedigree used package.

Under the hood this function calls `ASKAT.region`, `NormalizedASKAT.region`, `VCC1.region`, `VCC2.region` or `VCC3.region`, depending on the `method` parameter specified by the user.

Value

A data frame containing results of the association test specified by the `method` parameter for each region in the data frame specified by the `regions` parameter. The output data frame contains the following columns:

- `Score.Test`: the score of the given association test
- `P.value`: the p-value of the association test
- `N.Markers`: the number of markers in the region
- `regionname`: Name of the regions/genes on which you are running the association tests

Note that regions that do not contain any genetic variants will be removed from the output.

Author(s)

Lennart C. Karssen

`VCC1.region` *Run the VC-C1 method on a genomic region defined by a start and a stop base pair coordinate*

Description

Runs the VC-C1 method on a given genomic region

Usage

```
VCC1.region(y = NULL, X = NULL, Phi = NULL, type = "bed",
  filename = NULL, map = NULL, chr = 0, startpos = 0, endpos = 0,
  regionname = NULL, U = NULL, S = NULL, RH.Null = NULL,
  weights = NULL)
```

Arguments

`y`	vector of phenotype data (one entry per individual), of length n.
`X`	matrix of covariates including intercept (dimension: $n \times p$, with p the number of covariates)
`Phi`	Relationship matrix (i.e. twice the kinship matrix); an $n \times n$ square symmetric positive-definite matrix.
`type`	character, `'ped'`, `'bed'` (default) or `'shapeit-haps'` format of input file containing haplotype data
`filename`	character, path to input file containing haplotype data
`map`	object, data.frame contains 3 columns: rsID, chromosome, position in bp as output by e.g. `readMapFile`.
`chr`	character, chromosome number (basically from 1 to 22 as used by Plink), on which the region of interest is located
`startpos`	numeric, start position (in bp, base pairs) of the region of interest (default: 0)
`endpos`	numeric, end position (in bp, base pairs) of the region of interest (default: 0)
`regionname`	(character) Name of the region/gene on which you are running the association test. This name is used in the output of this function and can be used to distinguish different regions if this function is run multiple times.

U — (optional) Matrix of Eigenvectors of the relationship matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$. If this parameter is not given, it will be computed, so when running this function for many regions time can be saved by specifying not only Phi, but also S and U.

S — (optional) Matrix of Eigenvalues of the relationship matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$. If this parameter is not given, it will be computed, so when running this function for many regions, time can be saved by specifying not only Phi, but also S and U.

RH.Null — (optional) output of Estim.H0.VCC function. Practically, you don't need to calculate the null hypothesis for every region. One estimation per trait is enough.

weights — optional numeric vector of genotype weights. If this option is not specified, the beta distribution is used for weighting the variants, with each weight given by $w_i = dbeta(f_i, 1, 25)^2$, with f_i the minor allele frequency (MAF) of variant i. This default is the same as used by the SKAT package. This vector is used as the diagonal of the $m \times m$ matrix W, with m the number of variants.

Value

A data frame containing the results of the association test. The data frame contains the following columns:

- Score.Test: the score of the given association test
- P.value: the p-value of the association test
- N.Markers: the number of markers in the region
- regionname: Name of the region/gene on which you are running the association test

Author(s)

Sodbo Sharapov, Lennart C. Karssen

VCC2.region — *Run the VC-C2 method on a genomic region defined by a start and a stop base pair coordinate*

Description

Runs the VC-C2 method on a given genomic region

Usage

```
VCC2.region(y = NULL, X = NULL, Phi = NULL, type = "bed",
  filename = NULL, map = NULL, chr = 0, startpos = 0, endpos = 0,
  regionname = NULL, U = NULL, S = NULL, RH.Null = NULL,
  weights = NULL, Nperm = 100, Ncores = 1)
```

Arguments

y	vector of phenotype data (one entry per individual), of length n.
X	matrix of covariates including intercept (dimension: $n \times p$, with p the number of covariates)
Phi	Relationship matrix (i.e. twice the kinship matrix); an $n \times n$ square symmetric positive-definite matrix.
type	character, 'ped', 'bed' (default) or 'shapeit-haps' format of input file containing haplotype data
filename	character, path to input file containing haplotype data
map	object, data.frame contains 3 columns: rsID, chromosome, position in bp as output by e.g. readMapFile.
chr	character, chromosome number (basically from 1 to 22 as used by Plink), on which the region of interest is located
startpos	numeric, start position (in bp, base pairs) of the region of interest (default: 0)
endpos	numeric, end position (in bp, base pairs) of the region of interest (default: 0)
regionname	(character) Name of the region/gene on which you are running the association test. This name is used in the output of this function and can be used to distinguish different regions if this function is run multiple times.
U	(optional) Matrix of Eigenvectors of the relationship matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$. If this parameter is not given, it will be computed, so when running this function for many regions time can be saved by specifying not only Phi, but also S and U.
S	(optional) Matrix of Eigenvalues of the relationship matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$. If this parameter is not given, it will be computed, so when running this function for many regions, time can be saved by specifying not only Phi, but also S and U.
RH.Null	(optional) output of Estim.H0.VCC function. Practically, you don't need to calculate the null hypothesis for every region. One estimation per trait is enough.
weights	optional numeric vector of genotype weights. If this option is not specified, the beta distribution is used for weighting the variants, with each weight given by $w_i = dbeta(f_i, 1, 25)^2$, with f_i the minor allele frequency (MAF) of variant i. This default is the same as used by the SKAT package. This vector is used as the diagonal of the $m \times m$ matrix W, with m the number of variants.
Nperm	Integer, number of permutations to use for empirical p-value estimation (default: 100).
Ncores	(integer) Number of processor (CPU) cores to be used in parallel when doing the permutations to determine the p-value (default: 1).

Value

A data frame containing the results of the association test. The data frame contains the following columns:

- Score.Test: the score of the given association test

- P.value: the p-value of the association test
- N.Markers: the number of markers in the region
- regionname: Name of the region/gene on which you are running the association test

Author(s)

Lennart C. Karssen

VCC3.region *Run the VC-C3 method on a genomic region defined by a start and a stop base pair coordinate*

Description

Runs the VC-C3 method on a given genomic region

Usage

```
VCC3.region(y = NULL, X = NULL, Phi = NULL, type = "bed",
  filename = NULL, map = NULL, chr = 0, startpos = 0, endpos = 0,
  regionname = NULL, U = NULL, S = NULL, RH.Null = NULL,
  weights = NULL, Nperm = 100, Ncores = 1)
```

Arguments

y	vector of phenotype data (one entry per individual), of length n.
X	matrix of covariates including intercept (dimension: $n \times p$, with p the number of covariates)
Phi	Relationship matrix (i.e. twice the kinship matrix); an $n \times n$ square symmetric positive-definite matrix.
type	character, 'ped', 'bed' (default) or 'shapeit-haps' format of input file containing haplotype data
filename	character, path to input file containing haplotype data
map	object, data.frame contains 3 columns: rsID, chromosome, position in bp as output by e.g. readMapFile.
chr	character, chromosome number (basically from 1 to 22 as used by Plink), on which the region of interest is located
startpos	numeric, start position (in bp, base pairs) of the region of interest (default: 0)
endpos	numeric, end position (in bp, base pairs) of the region of interest (default: 0)
regionname	(character) Name of the region/gene on which you are running the association test. This name is used in the output of this function and can be used to distinguish different regions if this function is run multiple times.

U	(optional) Matrix of Eigenvectors of the relationship matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$. If this parameter is not given, it will be computed, so when running this function for many regions time can be saved by specifying not only Phi, but also S and U.
S	(optional) Matrix of Eigenvalues of the relationship matrix obtained from spectral decomposition of the relationship matrix: $\Phi = USU^T$. If this parameter is not given, it will be computed, so when running this function for many regions, time can be saved by specifying not only Phi, but also S and U.
RH.Null	(optional) output of Estim.H0.VCC function. Practically, you don't need to calculate the null hypothesis for every region. One estimation per trait is enough.
weights	optional numeric vector of genotype weights. If this option is not specified, the beta distribution is used for weighting the variants, with each weight given by $w_i = dbeta(f_i, 1, 25)^2$, with f_i the minor allele frequency (MAF) of variant i. This default is the same as used by the SKAT package. This vector is used as the diagonal of the $m \times m$ matrix W, with m the number of variants.
Nperm	Integer, number of permutations to use for empirical p-value estimation (default: 100).
Ncores	(integer) Number of processor (CPU) cores to be used in parallel when doing the permutations to determine the p-value (default: 1).

Value

A data frame containing the results of the association test. The data frame contains the following columns:

- Score.Test: the score of the given association test
- P.value: the p-value of the association test
- N.Markers: the number of markers in the region
- regionname: Name of the region/gene on which you are running the association test

Author(s)

Lennart C. Karssen

Index