Package 'globalGSA'

November 17, 2014

Type Package

Title Global Gene-Set Analysis for Association Studies.

Version 2.0

Date 2014-11-17

Author Natalia Vilor-Tejedor, M.L. Calle

Maintainer Natalia Vilor-Tejedor <nvilor@creal.cat>

Description Implementation of four different Gene set analysis (GSA) algorithms for combining the individual pvalues of a set of genetic variats (SNPs) in a gene level pvalue. The implementation includes the selection of the best inheritance model for each SNP.

License GPL (>= 2)

R topics documented:

`globalGSA-package` *Gene-set analysis for combining p-values in a joint test of association between a phenotype and a set of genetic variants (SNPs). Previously, a global test for the best inheritance model of each SNP is performed.*

Description

This package implements four different Gene-set analysis (GSA) methods for combining individual p-values of a set of SNPs. Each method provides a p-value for a joint test of association between the phenotype and the specified set of genetic variants. The four implemented methods are: [1] the globalEVT method, [2] the globalARTP method, [4] the Fisher's method [5] the Simes' method. Since the SNPs in a set may follow different modes of inheritance, previously to the GSA, a global test for the best inheritance model (dominant, recessive, log-additive and co-dominant) is performed on every SNP. The permutational p-value of the best model is obtained.

Details

Package:	globalGSA
Type:	Package
Version:	2
Date:	2013-09-22
License:	GPL (>= 2)

Author(s)

Natalia Vilor, M.Luz Calle

Maintainer: nvilor@creal.cat

References

[1] Vilor-Tejedor N, Calle ML, Gonzalez JR. Efficient and powerful testing for gene set analysis applied to Genome-Wide association studies. (under submission)

[2] Vilor-Tejedor N and Calle ML. Global adaptive rank truncated product method for gene-set analysis in association studies. Biom. J. 2014; 56:901-911. doi: 10.1002/bimj.201300192

[3] Yu, K. Li, Q. Bergen, A.W. Pfeiffer, R.M. Rosenberg, P.S. Caporaso, N. Kraft, P. and Chatterjee, N. (2009). Pathway analysis by adaptive combination of P-values. Genet, Epidemiol. December; 33(8): 700-709.

[4] Fisher, R.A. (1925). Statistical Methods for Research Workers. ISBN 0-05-002170-2.

[5] Simes, R.J. (1986). An Improved Bonferroni Procedure for Multiple Tests of Significance. Biometrika, 73, 751-754.

`globalARTP` *Global Adaptive Rank Truncated Product method.*

Description

This function provides the p-value for a joint test of association between a phenotype and a set of genetic variants (SNPs) using the Adaptive Rank Truncated Product method [1] after a global test for the best mode of inheritance of every SNP [2]. The final gene-p-value is obtained from the permutational null distribution of the test statistic.

Usage

```
globalARTP(data, B, K, gene_list, Gene = "all", addit = FALSE,
covariable = NULL, family = binomial)
```

Arguments

data	Data frame containing the variables in the model. The first column is the dependent variable which must be a binary variable defined as factor (in case-control studies, the usual codification is 1 for cases and 0 for controls). SNP values may be codified in a numerical form (0,1,2) denoting the number of minor alleles, or using a character form where the two alleles are specified, without spaces, tabs or any other symbol between the two alleles.
B	Number of permutations considered in the permutational procedure.
K	Integer that indicates the maximum truncation point.
gene_list	File that provides the name of the set (for instance, gene) where each SNP belongs. This file has two columns: the SNP-Id ("Id"), and the Gene-Id ("Gene"). The SNP-Id must have the same label as the colnames of the data file.
Gene	Name of the gene that we want to analyze. The default value is Gene= "all" that indicates that the p-values of all SNPs in the database are to be combined. In this case it is not necessary to specify the gene_list file. In other case, we need to specify the name of the gene, for instance, Gene = "Gene1", and also the gene_list file.
addit	logical to determine if only an additive inheritance model should be considered in the global Test or, conversely, if we want to consider all possible inheritance models (dominant, recessive, log-additive and co-dominant). By default, addit = FALSE.
covariable	Data frame containing the covariables in the model. Each column represents one covariable. By default, covariable=NULL.
family	This can be a character string naming a family distribution. By default, family=binomial.

Value

List with the following components:

nPerm	Number of permutations.
Gene	Considered Gene.
Trunkpoint	Considered truncation point.
Kopt	Optimal truncation point.
genevalue	gene-pvalue.

References

[1] Vilor-Tejedor N and Calle ML. Global adaptive rank truncated product method for gene-set analysis in association studies. Biom. J. 2014; 56:901-911. doi: 10.1002/bimj.201300192

[2] Yu, K. Li, Q. Bergen, A.W. Pfeiffer, R.M. Rosenberg, P.S. Caporaso, N. Kraft, P. and Chatterjee, N. (2009). Pathway analysis by adaptive combination of P-values. Genet, Epidemiol. December; 33(8): 700-709.

Examples

```
# load the included example dataset.
# This is a simulated case/control study data set
# with 2000 patients (1000 cases / 1000 controls)
# and 10 SNPs, where all of them have
```

```
# a direct association with the outcome:
data(data)
#globalARTP(data, B=1000, K=10, Gene="all", addit = FALSE)

# it may take some time,
# hence the result of this example is included:
data(ans11)

# You can test:
globalARTP(data, B=1, K=10, Gene="all", addit = FALSE)

# We consider that the first four SNPs
# are included in "Gene1",
# and the other six SNPs
# are included in "Gene2":
data(gene_list)
#globalARTP(data, B=1000, K=10, gene_list=gene_list, Gene="Gene1", addit = FALSE)

# it may take some time,
# hence the result of this example is included:
data(ans1)

# You can test:
globalARTP(data, B=1, K=10, gene_list=gene_list, Gene="Gene1", addit = FALSE)
```

`globalEVT` *Global Adaptive Extreme Value Distribution method.*

Description

This function provides the p-value for a joint test of association between a phenotype and a set of genetic variants (SNPs) using an Adaptive Extreme Value Distribution after a global test for the best mode of inheritance of every SNP. The final gene-p-value is obtained from

Usage

```
globalEVT(data, K, gene_list, Gene = "all", addit = FALSE,
covariable = NULL, family = binomial, LDinfo = NULL)
```

Arguments

`data`	Data frame containing the variables in the model. The first column is the dependent variable which must be a binary variable defined as factor (in case-control studies, the usual codification is 1 for cases and 0 for controls). SNP values may be codified in a numerical form (0,1,2) denoting the number of minor alleles, or using a character form where the two alleles are specified, without spaces, tabs or any other symbol between the two alleles.
`K`	Integer that indicates the maximum truncation point.
`gene_list`	File that provides the name of the set (for instance, gene) where each SNP belongs. This file has two columns: the SNP-Id ("Id"), and the Gene-Id ("Gene"). The SNP-Id must have the same label as the colnames of the data file.

Gene	Name of the gene that we want to analyze. The default value is Gene= "all" that indicates that the p-values of all SNPs in the database are to be combined. In this case it is not necessary to specify the gene_list file. In other case, we need to specify the name of the gene, for instance, Gene = "Gene1", and also the gene_list file.
addit	logical to determine if only an additive inheritance model should be considered in the global Test or, conversely, if we want to consider all possible inheritance models (dominant, recessive, log-additive and co-dominant). By default, addit = FALSE.
covariable	Data frame containing the covariables in the model. Each column represents one covariable. By default, covariable=NULL.
family	This can be a character string naming a family distribution. By default, family=binomial.
LDinfo	Data frame containing the linkage disequilibrium between SNPs. By default, LDinfo=NULL.

Value

List with the following components:

Gene	Considered Gene.
Trunkpoint	Considered truncation point.
genevalue	gene-pvalue.

References

[1] Vilor-Tejedor N, Calle ML, Gonzalez JR. Efficient and powerful testing for gene set analysis applied to Genome-Wide association studies. (under submission)

Examples

```
# load the included example dataset.
# This is a simulated case/control study data set
# with 2000 patients (1000 cases / 1000 controls)
# and 10 SNPs, where all of them have
# a direct association with the outcome:
data(data)
globalEVT(data, K=10)
```

`globalFisher` *Global Fisher combination method.*

Description

This function provides the p-value for a joint test of association between a phenotype and a set of genetic variants (SNPs) using the Fisher method [1] after a global test for the best mode of inheritance of every SNP. The final gene-p-value is obtained from the permutational null distribution of the test statistic

Usage

```
globalFisher(data, B, gene_list, Gene = "all", addit = FALSE,
covariable = NULL, family = binomial)
```

Arguments

data	Data frame containing the variables in the model. The first column is the dependent variable which must be a binary variable defined as factor (in case-control studies, the usual codification is 1 for cases and 0 for controls). SNP values may be codified in a numerical form (0,1,2) denoting the number of minor alleles, or using a character form where the two alleles are specified, without spaces, tabs or any other symbol between the two alleles.
B	Number of permutations considered in the permutational procedure.
gene_list	File that provides the name of the set (for instance, gene) where each SNP belongs. This file has two columns: the SNP-Id ("Id"), and the Gene-Id ("Gene"). The SNP-Id must have the same label as the colnames of the data file.
Gene	Name of the gene that we want to analyze. The default value is Gene= "all" that indicates that the p-values of all SNPs in the database are to be combined. In this case it is not necessary to specify the gene_list file. In other case, we need to specify the name of the gene, for instance, Gene = "Gene1", and also the gene_list file.
addit	logical to determine if only an additive inheritance model should be considered in the global Test or, conversely, if we want to consider all possible inheritance models (dominant, recessive, log-additive and co-dominant). By default, addit = FALSE.
covariable	Data frame containing the covariables in the model. Each column represents one covariable. By default, covariable=NULL.
family	This can be a character string naming a family distribution. By default, family=binomial.

Value

List with the following components:

nPerm	Number of permutations.
Gene	Considered Gene.
genevalue	gene-pvalue.

References

[1] Fisher, R.A. (1925). Statistical Methods for Research Workers. ISBN 0-05-002170-2.

Examples

```
# load the included example dataset.
# This is a simulated case/control study data set
# with 2000 patients (1000 cases / 1000 controls)
# and 10 SNPs, where all of them have
# a direct association with the outcome:
data(data)
#globalFisher(data, B=1000,  Gene="all", addit=FALSE)
```

```
# it may take some time,
# hence the result of this example is included:
data(ans21)

# You can test:
globalFisher(data, B=1,  Gene="all", addit=FALSE)

# We consider that the first four SNPs
# are included in "Gene1",
# and the other six SNPs
# are included in "Gene2":
data(gene_list)
#globalFisher(data, B=1000, gene_list=gene_list, Gene="Gene1", addit=FALSE)

# it may take some time,
# hence the result of this example is included:
data(ans2)

# You can test:
globalFisher(data, B=1, gene_list=gene_list, Gene="Gene1", addit=FALSE)
```

`globalSimes` *Global Simes' combination method.*

Description

This function provides the p-value for a joint test of association between a phenotype and a set of genetic variants (SNPs) using the Simes method [1] after a global test for the best mode of inheritance of every SNP. The final gene-p-value is obtained from the permutational null distribution of the test statistic

Usage

```
globalSimes(data, B, gene_list, Gene = "all", addit = FALSE,
covariable = NULL, family = binomial)
```

Arguments

`data`	Data frame containing the variables in the model. The first column is the dependent variable which must be a binary variable defined as factor (in case-control studies, the usual codification is 1 for cases and 0 for controls). SNP values may be codified in a numerical form (0,1,2) denoting the number of minor alleles, or using a character form where the two alleles are specified, without spaces, tabs or any other symbol between the two alleles.
`B`	Number of permutations considered in the permutational procedure.
`gene_list`	File that provides the name of the set (for instance, gene) where each SNP belongs. This file has two columns: the SNP-Id ("Id"), and the Gene-Id ("Gene"). The SNP-Id must have the same label as the colnames of the data file.
`Gene`	Name of the gene that we want to analyze. The default value is Gene= "all" that indicates that the p-values of all SNPs in the database are to be combined. In this case it is not necessary to specify the gene_list file. In other case, we

	need to specify the name of the gene, for instance, Gene = "Gene1", and also the gene_list file.
addit	logical to determine if only an additive inheritance model should be considered in the global Test or, conversely, if we want to consider all possible inheritance models (dominant, recessive, log-additive and co-dominant). By default, addit = FALSE.
covariable	Data frame containing the covariables in the model. Each column represents one covariable. By default, covariable=NULL.
family	This can be a character string naming a family distribution. By default, family=binomial.

Value

List with the following components:

nPerm	Number of permutations.
Gene	Considered Gene.
genevalue	gene-pvalue.

References

[1] Simes, R.J. (1986). An Improved Bonferroni Procedure for Multiple Tests of Significance. Biometrika, 73, 751-754.

Examples

```
# load the included example dataset.
# This is a simulated case/control study data set
# with 2000 patients (1000 cases / 1000 controls)
# and 10 SNPs, where all of them have
# a direct association with the outcome:
data(data)
#globalSimes(data, B=1000,  Gene="all", addit=FALSE)

# it may take some time,
# hence the result of this example is included:
data(ans31)

# You can test:
globalSimes(data, B=1,  Gene="all", addit=FALSE)

# We consider that the first four SNPs
# are included in "Gene1",
# and the other six SNPs
# are included in "Gene2":
data(gene_list)
#globalSimes(data, B=1000, gene_list=gene_list, Gene="Gene1", addit=FALSE)

# it may take some time,
# hence the result of this example is included:
data(ans3)

# You can test:
globalSimes(data, B=1, gene_list=gene_list, Gene="Gene1", addit=FALSE)
```

Index