

Package ‘p.exact’

February 11, 2015

Type Package

Title Exact P-values for Genome-Wide Association Analyses in Inbred Populations

Version 0.1-0

Date 2015-02-04

Author Xia Shen

Maintainer Xia Shen <xia.shen@ki.se>

Description This package implements for calculating exact dataset-specific p-values for genome-wide association (GWA) analyses in inbred populations. The module is compatible with existing GenABEL/gwaa.data data formats.

Depends R (>= 2.10), GenABEL, svMisc

License GPL (>= 2)

LazyLoad yes

R topics documented:

arab	1
p.exact	2
p.exact.binary	2
p.exact.gaussian	4

arab	<i>Arabidopsis thaliana</i> data set from Atwell et al. 2010 Nature
------	---

Description

The ‘arab’ data contains 107 phenotypes of *Arabidopsis* and genotypes from a 250K SNP chip as GenABEL gwaa.data format.

Format

The ‘Arabidopsis’ data set contains a gwaa.data object. See package GenABEL.)

Source

Atwell, S., Y. S. Huang, B. J. Vilhjalmsen, G. Willems, M. Horton, et al., 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465: 627-631.

References

Atwell, S., Y. S. Huang, B. J. Vilhjalmsón, G. Willems, M. Horton, et al., 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465: 627-631.

<code>p.exact</code>	<i>Exact p-values</i>
----------------------	-----------------------

Description

`p.exact`: Exact P-values for Genome-Wide Association Analyses in Inbred Populations

Details

The `p.exact` package implements for calculating exact dataset-specific p-values for genome-wide association (GWA) analyses in inbred populations. The module is compatible with existing GenABEL/gwaa.data data formats.

For converting data from other formats, see

`convert.snp.illumina` (Illumina/Affymetrix-like format), `convert.snp.text` (conversion from human-readable GenABEL format), `convert.snp.ped` (Linkage, Merlin, Mach, and similar files), `convert.snp.mach` (Mach-format), `convert.snp.tped` (from PLINK TPED format), `convert.snp.affymetrix` (BRML-style files).

For converting of GenABEL's data to other formats, see `export.merlin` (MERLIN and MACH formats), `export.impute` (IMPUTE, SNPTEST and CHIAMO formats), `export.plink` (PLINK format, also exports phenotypic data).

To load the data, see `load.gwaa.data`.

For data management and manipulations see `merge.gwaa.data`, `merge.snp.data`, `gwaa.data-class`, `snp.data-class`, `snp.names`, `snp.subset`.

Author(s)

Xia Shen

References

If you use `p.exact` package in your analysis, please cite the following work:

Xia Shen (2015). Beyond permutation test: calculating exact dataset-specific p-values for genome-wide association studies in inbred populations. *Submitted*.

See Also

GenABEL

p.exact.binary	<i>Exact p-values for genome-wide association analysis of case-control data in inbred populations</i>
----------------	---

Description

The function imports GenABEL (gwaa.data class) data format and calculates the exact dataset-specific p-values of a case-control phenotype for each variant or a given odds ratio and allele frequency.

Usage

```
p.exact.binary(pheno, gwaa.object, or = NULL, or.maf = NULL,
               low.maf = 0.05, high.ld = 0.9, method = "logOR", type = "two-sided",
               con.table = NULL)
```

Arguments

pheno	A string that gives the binary phenotype name in gwaa.object or a vector that gives phenotypic values match the order in gwaa.object.
gwaa.object	An object of gwaa.data-class to be analyzed.
or	An (optional) vector gives the odds ratios to be tested, and or.maf must also be given if not testing for the whole genome.
or.maf	An (optional) vector gives the minor allele frequencies in accordance with or, must be used together with or. If or.maf or or is missing, exact p-values will be calculated for the whole genome.
low.maf	A numeric value that gives the cut-off of the lowest minor allele frequency allowed in the analysis.
high.ld	A numeric value that gives the cut-off of the highest linkage disequilibrium R-square allowed in the analysis, i.e. LD pruning.
method	A string tells the method used, currently only "logOR" is available.
type	A string tells the statistical test type, can be 'one-sided' or 'two-sided'.
con.table	Genome-wide contingency tables. If NULL, to be calculated based on data. If a string tells the file name, load from the file or to be calculated and saved into the file. If an R matrix, regard as genome-wide contingency tables (see the saved R object for the format).

Value

The function returns a data frame of exact p-values (\$p.exact) and corresponding odds ratios \$or and MAFs \$MAF.

Note

None.

Author(s)

Xia Shen

References

Xia Shen (2015). Beyond permutation test: calculating exact dataset-specific p-values for genome-wide association studies in inbred populations. *Submitted*.

See Also

ccfast, glm

Examples

```
## Not run:
## loading example gwaa.data of data from Atwell et al. (2010) Nature
data(arab)

## running a regular GWA analysis for AvrRPM1
ccl <- ccfast('X33_i.avrRpm1..i.', data = arab)

## check the top finding using the exact p-value
top <- which.min(ccl[, 'Pldf'])
or <- ccl[top, 'effB']
f <- summary(arab[, top])$Q.2
maf <- min(f, 1 - f)
exact <- p.exact.binary(pheno = 'X33_i.avrRpm1..i.', gwaa.object = arab, or = or, or.maf = maf)

## End(Not run)
```

p.exact.gaussian	<i>Exact p-values for genome-wide association analysis in inbred populations</i>
------------------	--

Description

The function imports GenABEL (gwaa.data class) data format and calculates the exact dataset-specific p-values for each variant or a given effect size and allele frequency.

Usage

```
p.exact.gaussian(gwaa.object = NULL, n = NULL, maf = NULL, beta,
  beta.maf = NULL, low.maf = 0.05, type = "two-sided")
```

Arguments

gwaa.object	An (optional) object of gwaa.data-class.
n	An (optional) integer gives the sample size, only used when gwaa.object = NULL.
maf	An (optional) vector gives minor allele frequencies across the genome, only used when gwaa.object = NULL.
beta	An vector gives the effect sizes to be tested, and beta.maf must also be given if not testing for the whole genome.
beta.maf	An (optional) vector gives the minor allele frequencies in accordance with beta, must be used together with beta. If beta.maf is missing, exact p-values will be calculated for the whole genome.

<code>low.maf</code>	A numeric value that gives the cut-off of the lowest minor allele frequency allowed in the analysis.
<code>type</code>	A string tells the statistical test type, can be 'one-sided' or 'two-sided'.

Value

The function returns a data frame of exact p-values (`$p.exact`) and corresponding effect sizes `$beta` and MAFs `$MAF`.

Note

None.

Author(s)

Xia Shen

References

Xia Shen (2015). Beyond permutation test: calculating exact dataset-specific p-values for genome-wide association studies in inbred populations. *Submitted*.

See Also

`qtscore`, `t.test`

Examples

```
## Not run:
## loading example gwaa.data in GenABEL
data(ge03d2ex.clean)

## running a regular GWA analysis for height
qt1 <- qtscore(height, data = ge03d2ex.clean)

## running the multivariate GWAS again
exact <- p.exact.gaussian(ge03d2ex.clean, beta = qt1[, 'effB'])

## End(Not run)
```