Package 'FFBSKAT'

March 30, 2014

Type Package

Title FFBSKAT: Fast Family-Based SKAT

Version 1.0

Date 2014-30-03

Author Gulnara R. Svishcheva and Nadezhda M. Belonogova

Contact G.R. Svishcheva <gulsvi@mail.ru>, N.M. Belonogova <belon@bionet.nsc.ru>

Maintainer Anatoly V. Kirichenko <kianvl@mail.ru>, Sodbo Sharapov <sharapovsodbo@gmail.com>

Depends R (>= 2.10.0), GenABEL(>= 1.7-4)

Imports CompQuadForm, survey, parallel, foreach, doSNOW, SKAT

Description a package for fast family-based sequence kernel association analysis of quantitative traits.

License GPL-3

URL http //mga.bionet.nsc.ru/soft/FFBSKAT/

LazyLoad yes

R topics documented:

FFBSKAT-package *FFBSKAT: Fast Family-Based Sequence Kernel Association Test*

Description

The FFBSKAT package supplies an efficient kernel machine-based regression approach to the region-based association analysis aimed at identification of rare genetic variants for family-based or genetically related samples. The FFBSKAT package is the fast implementation of the method proposed by Schifano et al. [2012], Chen et al. [2013] and Oualkacha et al. [2013]. Method provides a score-based variance component test for regional association of a set of SNPs with a continuous phenotype in the presence of additional covariates and within-family correlations.

Details

Package: FFBSKAT
Type: Package
Version: 1.0
Date: 2014-07-02
License: GPLv3

Author(s)

Gulnara Svishcheva <gulsvi@mail.ru>
Nadezhda Belonogova <belon@bionet.nsc.ru>

References

Schifano E.D. et al. (2012) SNP set association analysis for familial data. Genet Epidemiol doi: 10.1002/gepi.21676. [Epub ahead of print].
Chen H. et al. (2013) Sequence kernel association test for quantitative traits in family samples. Genet. Epidemiol., 37, 196-204.
Oualkacha K. et al. (2013) Adjusted sequence kernel association test for rare variants controlling for cryptic and family relatedness. Genet. Epidemiol., 37, 366-76.

See Also

FFBSKAT
GenABEL

example.data *A small example data set for regional association analysis*

Description

`genodata` A matrix containing genotypes of 50 genetic variants (given in columns) in 66 individuals (given in rows). Three genotypes are coded as 0, 1 and 2.

`phenodata` A data.frame containing "trait", "sex" and "age" columns: a quantitative trait to be analyzed and its covariates.

`snpdata` A data.frame with descriptive information on 50 genetic variants in `genodata`. The important column is "gene": it assigns each variant to a certain gene region.

`kin` A kinship matrix for the 66 individuals.

Usage

```
data(example.data)
```

Examples

```
library(FFBSKAT)
data(example.data)

## Run FFBSKAT with sliding window (default):
out1 <- FFBSKAT(trait ~ age + sex, phenodata, genodata, kin,
return.nullmod = TRUE)

## Run FFBSKAT with regions defined in snpdata$gene and with
## null model obtained in first run:
out2 <- FFBSKAT(nullmod = out1$nullmod, phenodata = phenodata,
genodata = genodata, reg = snpdata$gene)
```

FFBSKAT *Fast Family-Based SKAT*

Description

A regional association analysis in related samples

Usage

```
FFBSKAT(formula, phenodata, genodata, kin, nullmod,
return.nullmod = FALSE, reg = NULL, sliding.window = c(20, 10),
kernel = "linear.weighted", beta.par = c(1, 25), weights = NULL,
mode = "add", method = "Kuonen", acc = 1e-6, ncores = 1)
```

Arguments

formula	referring to the column(s) in phenodata to be analysed as outcome and, if needed, covariates.
phenodata	a data frame containing columns mentioned in formula: trait to analyse and, if needed, covariates.
genodata	an object of gwaa.data-class or snp.data-class. It can be also a data frame or matrix (with ids given in rows) containing genotypes coded as AA = 0, Aa = 1 and aa = 2, where a is a minor allele. NAs in genodata will be imputed by the mean values. Monomorphic and duplicated variants will be omitted.
kin	a kinship matrix to evaluate the null model. It can be calculated from a pedigree (e.g. using kinship2 package) or from genotypic data (with ibs function of GenABEL).
nullmod	an object of polygenic data class. Setting nullmod allows to avoid re-estimation of the null model that does not depend on genotypes and can be calculated once for a trait. nullmod object in proper format can be obtained by running either FFBSKAT with return.nullmod = TRUE or polygenic function. Entries used in nullmod are $residualY and $InvSigma with

non-empty names. `nullmod$call$formula` will be compared to the `formula` given in input. Note that, if covariates are mentioned in `formula` or `nullmod$call$formula`, they should be given in `phenodata` to be used in kernel matrix calculation.

`return.nullmod`
: logical value indicating whether a `nullmod` object should be returned.

`reg`
: a vector assigning a region for each genetic marker in `genodata`. If NULL, `sliding.window` parameters are used.

`sliding.window`
: the sliding window size and step. Has no effect if `reg` is defined.

`kernel`
: one of "linear.weighted" (default), "quadratic", "IBS", "IBS.weighted", "2wayIX". See Details for "linear.weighted" kernel description and [Wu, 2011] for other kernel types. "2wayIX" kernel considers SNP-SNP interaction terms along with main effects. For "linear.weighted" and "IBS.weighted" kernels, weights can be varied by defining `weights` or `beta.par`.

`beta.par`
: two positive numeric shape parameters in the beta distribution to assign weights for each SNP in weighted kernels (see Details). Default = c(1, 25) is recommended for analysis of rare variants. Has no effect for unweighted kernels or if `weights` are defined.

`weights`
: a numeric vector or a function of MAF to assign weights for each SNP in the weighted kernels. Has no effect if one of unweighted kernels was chosen. If NULL, the weights will be calculated using beta distribution (see Details)

`mode`
: the mode of inheritance: "add", "dom" or "rec" for additive, dominant or recessive mode, respectively. For dominant (recessive) mode genotypes will be recoded as AA = 0, Aa = 1 and aa = 1 (AA = 0, Aa = 0 and aa = 1), where a is a minor allele. Default mode is additive.

`method`
: either "Kuonen" or "Davies". Method of computing the p-value. Default = "Kuonen".

`acc`
: accuracy parameter for "Davies" method.

`ncores`
: number of CPUs for parallel calculations. Default = 1

Details

By default, FFBSKAT uses the linear weighted kernel function, $K = GWWG^T$, where G is the $n \times p$ genotype matrix for n individuals and p SNPs in the region, and W is the $p \times p$ diagonal weight matrix. Given shape parameters of the beta function, `beta.par = c(a, b)`, the weights are defined using probability density function of the beta distribution:

$$W_i = (B(a,b))^{-1} MAF_i^{a-1}(1-MAF_i)^{b-1},$$

where MAF_i is a minor allelic frequency for i^{th} SNP in region, which is estimated from genotypes, and $B(a,b)$ is a beta function. This way of defining weights is the same as in original SKAT (see [Wu, 2011] for details). The formula:

$$Q = 0.5\tilde{y}^T\Omega^{-1}K\Omega^{-1}\tilde{y}$$

is used to calculate score statistic, where $\tilde{y}$ and Ω are environmental residuals and covariance matrix obtained under the null hypothesis, respectively. Depending on the method option chosen, either Kuonen or Davies method is used to calculate p-values from the score statistic Q. Both an Applied Statistics algorithm that inverts the characteristic function of the mixture chisq [Davies, 1980] and

a saddlepoint approximation [Kuonen, 1999] are nearly exact, with the latter usually being a bit faster. For other kernel types, see [Wu, 2011].

Value

A list with values:

`results`	a data frame containing p-values, numbers of variants and informative polymorphic variants for each of analyzed regions.
`sample.size`	after omitting NA's in trait and, if used, covariates.
`nullmod`	a null model object of class `polygenic`, returned if `return.nullmod = TRUE`.

Author(s)

Gulnara Svishcheva <gulsvi@mail.ru>
Nadezhda Belonogova <belon@bionet.nsc.ru>

References

Davies R.B. (1980) Algorithm AS 155: The Distribution of a Linear Combination of chi-2 Random Variables, Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 29, N 3, P. 323-333.
Kuonen D. (1999) Saddlepoint Approximations for Distributions of Quadratic Forms in Normal Variables. Biometrika, Vol. 86, No. 4, P. 929-935.
Wu M.C. et al. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet., Vol. 89, P. 82-93.

See Also

FFBSKAT-package
GenABEL

Examples

```
library(FFBSKAT)
data(example.data)

## Run FFBSKAT with sliding window (default):
out1 <- FFBSKAT(trait ~ age + sex, phenodata, genodata, kin,
return.nullmod = TRUE)

## Run FFBSKAT with regions defined in snpdata$gene and with
## null model obtained in first run:
out2 <- FFBSKAT(nullmod = out1$nullmod, phenodata = phenodata,
genodata = genodata, reg = snpdata$gene)
```

Index