Package 'kangar00'

May 9, 2024

Type Package

Title Kernel Approaches for Nonlinear Genetic Association Regression

Version 1.4.2

Date 2024-05-09

URL https://kangar00.manitz.org/

BugReports https://github.com/jmanitz/kangar00/issues

Description Methods to extract information on pathways, genes and various single-nucleotid polymorphisms (SNPs) from online databases. It provides functions for data preparation and evaluation of genetic influence on a binary outcome using the logistic kernel machine test (LKMT). Three different kernel functions are offered to analyze genotype information in this variance component test: A linear kernel, a size-adjusted kernel and a network-based kernel).

License GPL-2

Encoding UTF-8

Collate 'pathway.r' 'GWASdata.r' 'data.R' 'get_network_matrix.r' 'kernel.r' 'lkmt.r'

Depends R (>= 3.5.0)

Imports methods, bigmemory, sqldf, CompQuadForm, data.table, lattice, igraph

LazyData true

Suggests biomaRt, KEGGgraph, testthat

NeedsCompilation no

RoxygenNote 7.3.1

Author Juliane Manitz [aut, cre],
Benjamin Hofner [aut],
Stefanie Friedrichs [aut],
Patricia Burger [aut],
Ngoc Thuy Ha [aut],
Saskia Freytag [ctb],
Heike Bickeboeller [ctb]

Maintainer Juliane Manitz <r@manitz.org>

Repository CRAN

Date/Publication 2024-05-09 03:10:02 UTC

R topics documented:

`kangar00-package` *kangar00 package*

Description

This package includes methods to extract information on pathways, genes and SNPs from online databases and to evaluate these data using the logistic kernel machine test (LKMT) (Liu et al. 2008).

We defined SNP sets representing genes and whole pathways using knowledge on gene membership and interaction from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al. 2014). SNPs are mapped to genes via base pair positions of SNPs and transcript start and end points of genes as documented in the Ensemble database (Cunningham et al. 2015).

In the LKMT, we employed the linear kernel (Wu et al. 2010) as well as two more advanced kernels, adjusting for size bias in the number of SNPs and genes in a pathway (size-adjusted kernels), and

incorporating the network structure of genes within the pathway (pathway kernels), respectively (Freytag et al. 2012, 2014). P-values are derived in a variance component test using a moment matching method (Schaid, 2010) or Davies' algorithm (Davies, 1980).

Details

Package:	kangar00
Version:	1.1
Date:	2017-08-07
License:	GPL-2

Author(s)

Juliane Manitz [aut], Stefanie Friedrichs [aut], Patricia Burger [aut], Benjamin Hofner [aut], Ngoc Thuy Ha [aut], Saskia Freytag [ctb], Heike Bickeboeller [ctb]
Maintainer: Juliane Manitz <r@manitz.de>

References

- Cunningham F, Ridwan Amode M, Barrell D, et al. Ensembl 2015. Nucleic Acids Research 2015 43 Database issue:D662-D669

- Davies R: Algorithm as 155: the distribution of a linear combination of chi-2 random variables. J R Stat Soc Ser C 1980, 29:323-333.

- Freytag S, Bickeboeller H, Amos CI, Kneib T, Schlather M: A Novel Kernel for Correcting Size Bias in the Logistic Kernel Machine Test with an Application to Rheumatoid Arthritis. Hum Hered. 2012, 74(2):97-108.

- Freytag S, Manitz J, Schlather M, Kneib T, Amos CI, Risch A, Chang-Claude J, Heinrich J, Bickeboeller H: A network-based kernel machine test for the identification of risk pathways in genome-wide association studies. Hum Hered. 2013, 76(2):64-75.

- Friedrichs S, Manitz J, Burger P, Amos CI, Risch A, Chang-Claude JC, Wichmann HE, Kneib T, Bickeboeller H, Hofner B: Pathway-Based Kernel Boosting for the Analysis of Genome-Wide Association Studies. Computational and Mathematical Methods in Medicine. 2017(6742763), 1-17. doi:10.1155/2017/6742763.

- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 42, D199-D205 (2014).

- Liu D, Ghosh D, Lin X. Estimation and testing for the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via logistic mixed models. BMC Bioinformatics. 2008 9:292.

- Schaid DJ: Genomic similarity and kernel methods I: advancements by building on mathematical and statistical foundations. Hum Hered 2010, 70:109-131.

- Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, Lin X: Powerful SNP-Set Analysis for Case-Control Genome-Wide Association Studies. Am J Hum Genet 2010, 86:929-42

---

anno *Example annotation file for three pathways.*

Description

A dataset containing an annotation example for 4056 SNPs in three different pathways.

Usage

```
data(anno)
```

Format

A `data frame` with 4056 rows and 5 variables:

pathway includes KEGG identifiers of three example pathways

gene names of genes in the pathways

chr specifies the chromosome

snp includes rs-numbers of example SNPs

position gives positions of example SNPs

Source

simulated data

Examples

```
data(anno)
head(anno)
# create gwas object
data(pheno)
data(geno)
gwas <- new('GWASdata', pheno=pheno, geno=geno, anno=anno, desc="some study")
```

calc_kernel *Calculate the kernel-matrix for a pathway*

Description

Uses individuals' genotypes to create a kernel object including the calculated kernel matrix for a specific pathway. Each numeric value within this matrix is calculated from two individuals' genotypevectors of the SNPs within the pathway by a kernel function. It can be interpreted as the genetic similiarity of the individuals. Association between the pathway and a binary phenotype (case-control status) can be evaluated in the logistic kernel machine test, based on the kernel object. Three kernel functions are available.

Usage

```
## S4 method for signature 'GWASdata'
calc_kernel(
  object,
  pathway,
  knots = NULL,
  type = c("lin", "sia", "net"),
  calculation = c("cpu", "gpu"),
  ...
)

## S4 method for signature 'GWASdata'
lin_kernel(object, pathway, knots = NULL, calculation = c("cpu", "gpu"), ...)

## S4 method for signature 'GWASdata'
sia_kernel(object, pathway, knots = NULL, calculation = c("cpu", "gpu"), ...)

## S4 method for signature 'GWASdata'
net_kernel(object, pathway, knots = NULL, calculation = c("cpu", "gpu"), ...)
```

Arguments

object	GWASdata object containing the genotypes of the individuals for which a kernel will be calculated.
pathway	object of the class pathway specifying the SNP set for which a kernel will be calculated.
knots	GWASdata object, if specified a kernel will be computed.
type	character indicating the kernel type: Use 'lin' to specify the linear kernel, 'sia' for the size-adjusted or 'net' for the network-based kernel.
calculation	character specifying if the kernel matrix is computed on CPU or GPU.
...	further arguments to be passed to kernel computations.

Details

Different types of kernels can be constructed:

- `type='lin'` creates the linear kernel assuming additive SNP effects to be evaluated in the logistic kernel machine test.
- `type='sia'` calculates the size-adjusted kernel which takes into consideration the numbers of SNPs and genes in a `pathway` to correct for size bias.
- `type='net'` calculates the network-based kernel. Here not only information on gene membership and gene/pathway size in number of SNPs is incorporated, but also the interaction structure of genes in the `pathway`.

For more details, check the references.

Value

Returns an object of class `kernel`, including the similarity `matrix` of the `pathway` for the considered individuals.
If `knots` are specified low-rank kernel of class a `lowrank_kernel` will be returned, which is not necessarily quadratic and symmetric.

Methods (by class)

- `lin_kernel(GWASdata)`:
- `sia_kernel(GWASdata)`:
- `net_kernel(GWASdata)`:

Author(s)

Stefanie Friedrichs, Juliane Manitz

References

- Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, Lin X Powerful SNP-Set Analysis for Case-Control Genome-Wide Association Studies. Am J Hum Genet 2010, 86:929-42
- Freytag S, Bickeboeller H, Amos CI, Kneib T, Schlather M: A Novel Kernel for Correcting Size Bias in the Logistic Kernel Machine Test with an Application to Rheumatoid Arthritis. Hum Hered. 2012, 74(2):97-108.
- Freytag S, Manitz J, Schlather M, Kneib T, Amos CI, Risch A, Chang-Claude J, Heinrich J, Bickeboeller H: A network-based kernel machine test for the identification of risk pathways in genome-wide association studies. Hum Hered. 2013, 76(2):64-75.

See Also

`kernel-class`,`pathway`

Examples

```
data(gwas)
data(hsa04020)
lin_kernel <- calc_kernel(gwas, hsa04020, knots=NULL, type='lin', calculation='cpu')
summary(lin_kernel)
net_kernel <- calc_kernel(gwas, hsa04020, knots=NULL, type='net', calculation='cpu')
summary(net_kernel)
```

---

geno *Example genotypes for 50 individuals.*

Description

A matrix containing example genotypes for 4056 SNPs of 50 individuals. Column names give the rs-numbers of 4056 example SNPs, row names the identifiers of 50 example individuals.

Usage

```
data(geno)
```

Format

A `matrix` with 5 rows and 4056 columns:

each entry in the matrix represents a simulated minor allele count for the corresponding SNP and individual.

Source

simulated data

Examples

```
data(geno)
head(geno)
# create gwas object
data(pheno)
data(anno)
gwas <- new('GWASdata', pheno=pheno, geno=geno, anno=anno, desc="some study")
```

`get_anno,snp_info,pathway_info-method`

Annotates SNPs via genes to pathways

Description

A function to create the annotation for a `GWASdata` object. It combines a `snp_info` and a `pathway_info` object into an annotation `data.frame` used for `pathway` analysis on GWAS. SNPs are assigned to pathways via gene membership.

Usage

```
## S4 method for signature 'snp_info,pathway_info'
get_anno(object1, object2, ...)
```

Arguments

`object1`	A `snp_info` object with SNP information as returned by the `snp_info` function. The included `data frame` contains the columns 'chr', 'position' and 'snp'.
`object2`	A `pathway_info` object with information on genes contained in pathways. It is created by the `pathway_info` function and contains a `data frame` with columns 'pathway', 'gene_start', 'gene_end', 'chr', 'gene'.
`...`	further argdata(hsa04020)

Value

A `data.frame` mapping SNPs to genes and genes to pathways. It includes the columns 'pathway', 'gene', 'chr', 'snp' and 'position'.

Author(s)

Stefanie Friedrichs, Saskia Freytag, Ngoc-Thuy Ha

See Also

`snp_info`, `pathway_info`

Examples

```
data(hsa04022_info)  # pathway_info('hsa04020')
data(rs10243170_info)# snp_info("rs10243170")
get_anno(rs10243170_info, hsa04022_info)
```

get_network_matrix *Function to calculate the network* matrix *for a* pathway *object*

Description

get_network_matrix creates the adjacency matrix representing the gene-gene interaction structure within a particular pathway. Note that a KEGG kgml file is downloaded and saved in the working directory.

Usage

```
## S4 method for signature 'pathway'
get_network_matrix(object, directed = TRUE, method = "auto")
```

Arguments

object	A pathway object identifying the pathway for which gene interaction infomation should be extracted. Here, KEGG IDs of format 'hsa00100' are used and information is downloaded from the KEGG database.
directed	A logical argument, stating whether the network matrix should return directed (TRUE) or undirected (FALSE) links.
method	Download method to be used for downloading files, passed to via KEGGgraph::retrieveKGML to utils::download.file function. Currently supports 'auto' (default), 'internal', 'wininet' (Windows only), 'libcurl', 'wget' and 'curl'.

Value

get_network_matrix returns the modified pathway object, where the slots adj and sign are altered according to the downloaded information in the KEGG kgml file.

Author(s)

Stefanie Friedrichs, Patricia Burger, Juliane Manitz

gwas *Example* GWASdata *object.*

Description

An object of type GWASdata containing the example files for annotation, phenotypes and genotypes.

Usage

```
data(gwas)
```

Format

An object of class GWASdata:

geno contains example genotypes

anno example annotation for three pathways

pheno exemplary phenotypes for all 'genotyped' individuals

desc a description of the GWAS study, here 'example study'

Source

simulated data

Examples

```
# create gwas object
data(pheno)
data(geno)
data(anno)
gwas <- new('GWASdata', pheno=pheno, geno=geno, anno=anno, desc="some study")
```

GWASdata *S4 class for an object representing a Genome-wide Assocaition Study.*

Description

S4 class for an object representing a Genome-wide Assocaition Study.

'GWASdata' is a GWASdata object constructor.

show displays basic information on GWASdata object

summary summarizes the content of a GWASdata object and gives an overview about the information included in a GWASdata object. Summary statistics for phenotype and genotype data are calculated.

GeneSNPsize creates a data.frame of pathway names with numbers of snps and genes in each pathway.

Usage

```
GWASdata(object, ...)

## S4 method for signature 'ANY'
GWASdata(geno, anno, pheno = NULL, desc = "")

## S4 method for signature 'GWASdata'
show(object)

## S4 method for signature 'GWASdata'
summary(object)

## S4 method for signature 'GWASdata'
GeneSNPsize(object)
```

Arguments

object	A GWASdata object.
...	Further arguments can be added to the function.
geno	An object of any type, including the genotype information.
anno	A data.frame containing the annotation file for the GWASdata object.
pheno	A data.frame specifying individual IDs, phenotypes and covariates to be included in the regression model.
desc	A character giving the GWAS description, e.g. name of study.

Methods (by generic)

- GeneSNPsize(GWASdata): creates a data.frame of pathway names with numbers of snps and genes in each pathway.

Slots

geno An object of any type, including genotype information. The format needs to be one line per individual and on colum per SNP in minor-allele coding (0,1,2). Other values between 0 and 2, as from impute dosages, are allowed. Missing values must be imputed prior to creation of a GWASdata object.

anno A data.frame mapping SNPs to genes and genes to pathways. Needs to include the columns 'pathway' (pathway ID, e.g. hsa number from KEGG database), 'gene' (gene name (hgnc_symbol)), 'chr' (chromosome), 'snp' (rsnumber) and 'position' (base pair position of SNP).

pheno A data.frame specifying individual IDs, phenotypes and covariates to be included in the regression model e.g. ID, pheno, sex, pack.years. Note: IDs have to be in the first column!

desc A character giving the GWAS description, e.g. name of study.

Author(s)

Juliane Manitz, Stefanie Friedrichs

Examples

```
# create gwas data object
data(pheno)
data(geno)
data(anno)
gwas <- new('GWASdata', pheno=pheno, geno=geno, anno=anno, desc="some study")
# show and summary methods
gwas
summary(gwas)
# SNPs and genes in pathway
GeneSNPsize(gwas)
```

hsa04020 *Example pathway object for pathway hsa04020.*

Description

An object of class pathway for the pathway with KEGG identifier hsa04020.

Usage

```
data(hsa04020)
```

Format

A pathway object including 180 genes.

id KEGG identifier of the example pathways

adj gives the quadratic adjacency matrix for the pathway and with that the network topology. Matrix dimensions equal the number of genes in the pathway

sign includes a vector of signs to distinguish activations and inhibitions in the adjacency matrix

Source

simulated data and Ensembl extract

hsa04022_info *Example pathway_info object for pathway hsa04022.*

Description

An object of class pathway_info for the pathway with KEGG identifier hsa04020.

Usage

```
data(hsa04022_info)
```

Format

A pathway_info object including information on 163 genes.

info a data frame including information on genes included in pathway. Has columns 'pathway', 'gene_start', 'gene_end', 'chr', and 'gene'

Source

Ensembl extract

Examples

```
## Not run:
pathway_info('hsa04020')

## End(Not run)
```

`kernel-class` *An S4 class representing a kernel matrix calculated for a pathway*

Description

An S4 class representing a kernel matrix calculated for a pathway

`show` displays the `kernel` object briefly

`summary` generates a `kernel` object summary including the number of individuals and genes for the `pathway`

`plot` creates an image plot of a `kernel` object

Usage

```
## S4 method for signature 'kernel'
show(object)

## S4 method for signature 'kernel'
summary(object)

## S4 method for signature 'kernel,missing'
plot(x, y = NA, hclust = FALSE, ...)
```

Arguments

`object`	An object of class `kernel`
`x`	the `kernel` object to be plotted.
`y`	missing (placeholder).
`hclust`	`logical`, indicating whether a dendrogram should be added.
`...`	further arguments to be passed to the function.

Slots

`type` A `character` representing the kernel type: Use `'lin'` for linear kernel, `'sia'` for the size-adjusted or `'net'` for the network-based kernel.

`kernel` A kernel `matrix` of dimension equal to the number of individuals

`pathway` A `pathway` object

Author(s)

Juliane Manitz

Examples

```
data(gwas)
data(hsa04020)
net_kernel <- calc_kernel(gwas, hsa04020, knots=NULL, type='net', calculation='cpu')

show(net_kernel)
summary(net_kernel)
plot(net_kernel, hclust=TRUE)
```

---

lkmt-class *An S4 class to represent the variance component test.*

Description

An S4 class to represent the variance component test.

`show` displays basic information on `lkmt` object

`summary` generates a `lkmt` object summary including the used kernel, pathway and the test result

Usage

```
## S4 method for signature 'lkmt'
show(object)

## S4 method for signature 'lkmt'
summary(object)
```

Arguments

`object`	An object of class `lkmt`.

Value

`show` Basic information on `lkmt` object.

`summary` Summarized information on `lkmt` object.

Slots

`formula` A formula stating the regression nullmodel that will be used in the variance component test.

`kernel` An object of class `kernel` representing the similarity matrix of the individuals based on which the pathways influence is evaluated.

`GWASdata` An object of class `GWASdata` including the data on which the test is conducted.

`statistic` A `vector` giving the value of the variance component test statistic.

`df` A `vector` containing the number of degrees of freedom.

`p.value` A `vector` giving the p-value calculated for the `pathway` object considered in the variance component test.

For details on the variance component test see the references.

Author(s)

Juliane Manitz, Stefanie Friedrichs

References

- Liu D, Lin X, Ghosh D: Semiparametric regression of multidimensional genetic pathway data: least-squares kernel machines and linear mixed models. Biometrics 2007, 63(4):1079-88.
- Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, Lin X: Powerful SNP-Set Analysis for Case-Control Genome-Wide Association Studies. Am J Hum Genet 2010, 86:929-42

Examples

```
data(hsa04020)
data(gwas)
# compute kernel
net_kernel <- calc_kernel(gwas, hsa04020, knots=NULL, type='net', calculation='cpu')
# perform LKMT
res <- lkmt_test(pheno ~ sex + age, net_kernel, gwas, method='satt')
# show and summary methods
show(res)
summary(res)
# summary method
summary(lkmt.net.kernel.hsa04020)
```

---

lkmt.net.kernel.hsa04020

Example test result for the network-based `kernel` for `pathway` hsa04020.

Description

An object of class `lkmt` containing exemplary test results for an application of the logistic kernel machine test, derived from the example data.

Usage

```
data(lkmt.net.kernel.hsa04020)
```

Format

An object of class `lkmt` for the network-based `kernel` and the `pathway` hsa04020.

formular gives a formular defining the nullmodel used in the logistic kernel machine test

kernel includes the `kernel` object of the `pathway` to be evaluated

GWASdata gives the `GWASdata` object including the study data considered in testing

statistic gives the value of the test statistic

df specifies the degrees of freedom

p.value includes teh p-value resulting from the test

Source

simulated data and Ensembl extract

Examples

```
data(hsa04020)
data(gwas)
net_kernel <- calc_kernel(gwas, hsa04020, knots=NULL, type='net', calculation='cpu')
lkmt_test(pheno ~ sex + age, net_kernel, gwas, method='satt')
```

---

lkmt_test *A function to calculate the p-values for kernel matrices.*

Description

For parameter `'satt'` a pathway's influence on the probability of beeing a case is evaluated in the logistic kernel machine test and p-values are determined using a Sattherthwaite approximation as described by Dan Schaid.

For parameter `'davies'` a pathways influence on the probability of beeing a case is evaluated using the p-value calculation method described by Davies. Here the function `davies` from package **CompQuadForm** is used.

Usage

```
lkmt_test(formula, kernel, GWASdata, method = c("satt", "davies"), ...)

## S4 method for signature 'matrix'
score_test(x1, x2)

## S4 method for signature 'matrix'
davies_test(x1, x2)
```

Arguments

formula	The formula to be used for the regression nullmodel.
kernel	An object of class kernel including the pathway representing kernel-matrix based on which the test statistic will be calculated.
GWASdata	A GWASdata object stating the data used in analysis.
method	A character specifying which method will be used for p-value calculation. Available are 'satt' for the Satterthwaite approximation and 'davies' for Davies' algorithm. For more details see the references.
...	Further arguments can be given to the function.
x1	A matrix which is the similarity matrix calculated for the pathway to be tested.
x2	An lm or glm object of the nullmodel with fixed effects covariates included, but no genetic random effects.

Value

An lkmt object including the following test results

- The formula of the regression nullmodel used in the variance component test.
- An object of class kernel including the similarity matrix of the individuals based on which the pathways influence is evaluated.
- An object of class GWASdata stating the data on which the test was conducted.
- statistic A vector giving the value of the variance component test statistic.
- df A vector giving the number of degrees of freedom.
- p.value A vector giving the p-value calculated for the pathway in the variance component test.

Author(s)

Stefanie Friedrichs, Juliane Manitz

References

For details on the variance component test

- Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, Lin X: Powerful SNP-Set Analysis for Case-Control Genome-Wide Association Studies. Am J Hum Genet 2010, 86:929-42
- Liu D, Lin X, Ghosh D: Semiparametric regression of multidimensional genetic pathway data: least-squares kernel machines and linear mixed models. Biometrics 2007, 63(4):1079-88.

For details on the p-value calculation see

- Schaid DJ: Genomic Similarity and Kernel Methods I: Advancements by Building on Mathematical and Statistical Foundations. Hum Hered 2010, 70:109-31
- Davies R: Algorithm as 155: the distribution of a linear combination of chi-2 random variables. J R Stat Soc Ser C 1980, 29:323-333.

Examples

```
data(hsa04020)
data(gwas)
net_kernel <- calc_kernel(gwas, hsa04020, knots=NULL, type='net', calculation='cpu')
lkmt_test(pheno ~ sex + age, net_kernel, gwas, method='satt')
```

`lowrank_kernel-class` *An S4 class to represent a low-rank kernel for a SNPset at specified knots*

Description

An S4 class to represent a low-rank kernel for a SNPset at specified knots

Details

This kernel is used for predictions. If observations and knots are equal, better construct a full-rank kernel of class `kernel`.

Slots

`type` character, kernel type: Use `'lin'` for the linear kernel, `'sia'` for the size-adjusted or `'net'` for the network-based kernel.

`kernel` kernel `matrix` of dimension equal to individuals

`pathway` `pathway` object

Author(s)

Juliane Manitz

Examples

```
data(gwas)
data(hsa04020)
square <- calc_kernel(gwas, hsa04020, knots=gwas, type='lin', calculation='cpu')
dim(square@kernel)

gwas2 <- new('GWASdata', pheno=pheno[1:10,], geno=geno[1:10,], anno=anno, desc="study 2")
low_rank <- calc_kernel(gwas, hsa04020, knots = gwas2, type='net', calculation='cpu')
dim(low_rank@kernel)
```

make_psd,matrix-method

Adjust network matrix *to be positive semi-definite*

Description

Adjust network matrix to be positive semi-definite

Usage

```
## S4 method for signature 'matrix'
make_psd(x, eps = sqrt(.Machine$double.eps))
```

Arguments

x	A matrix specifying the network adjacency matrix.
eps	A numeric value, setting the tolance for smallest eigenvalue adjustment

Details

For a matrix N, the closest positive semi-definite matrix is calculated as N* = rho*N + (1+rho)*I, where I is the identity matrix and rho = 1/(1 - lambda) with lambda the smallest eigenvalue of N. For more details check the references.

Value

The matrix x, if it is positive definite and the closest positive semi-definite matrix if x is not positive semi-definite.

Author(s)

Juliane Manitz, Saskia Freytag, Stefanie Friedrichs

References

- Freytag S, Manitz J, Schlather M, Kneib T, Amos CI, Risch A, Chang-Claude J, Heinrich J, Bickeboeller H: A network-based kernel machine test for the identification of risk pathways in genome-wide association studies. Hum Hered. 2013, 76(2):64-75.

Examples

```
set.seed(2345)
m <- matrix(data=sample(size=25, c(0,0,1), replace=TRUE),5,5)
m <- m + t(m)
min(eigen(m, only.values = TRUE, symmetric = TRUE)$values)
round(make_psd(m),2)
```

net.kernel.hsa04020 *Example network-based kernel matrix for pathway hsa04020.*

Description

An example of a kernel object.

Usage

```
data(net.kernel.hsa04020)
```

Format

An object of class kernel and type 'network' for the pathway hsa04020.

type specifies which kernel function was used to calculate the kernel

kernel includes the kernel matrix calculated for the pathway

pathway includes the pathway object of the pathway, for which the kernel matrix was calculated

Source

simulated data and Ensembl extract

Examples

```
data(net.kernel.hsa04020)
# derivation
data(gwas)
data(hsa04020)
net_kernel <- calc_kernel(gwas, hsa04020, knots=NULL, type='net', calculation='cpu')
# are the value differences smaller than machine epsilon?
all(abs(net.kernel.hsa04020@kernel - net_kernel@kernel) < sqrt(.Machine$double.eps))
```

pathway *An S4 class to represent a gene-gene interaction network*

Description

An S4 class to represent a gene-gene interaction network

pathway is the pathway object constructor.

show displays the pathway object briefly

summary generates a pathway object summary including basic network properties.

pathway2igraph converts a pathway object into an igraph object with edge attribute sign

analyze pathway network properties

get_genes is a helper function that extracts the gene names in a pathway and returns a vector containing character elements of gene names

plot visualizes the pathway as igraph object

sample_genes randomly selects effect gene in a pathway according the betweenness centrality and (no -1) neighors

Usage

```
pathway(object, ...)

## S4 method for signature 'ANY'
pathway(id, adj = matrix(0), sign = NULL)

## S4 method for signature 'pathway'
show(object)

## S4 method for signature 'pathway'
summary(object)

## S4 method for signature 'pathway'
pathway2igraph(object)

## S4 method for signature 'pathway'
analyze(object, ...)

## S4 method for signature 'pathway'
get_genes(object)

## S4 method for signature 'pathway,missing'
plot(
  x,
  y = NA,
  highlight.genes = NULL,
  gene.names = c(NULL, "legend", "nodes"),
  main = NULL,
  asp = 0.95,
  vertex.size = 11,
  vertex.color = "khaki1",
  vertex.label.cex = 0.8,
  edge.width = 2,
  edge.color = "olivedrab4",
  ...
)

## S4 method for signature 'pathway'
sample_genes(object, no = 3)
```

Arguments

object	An object of class pathway-class
...	Further arguments can be added to the function.
id	A character repesenting the pathway id.
adj	A matrix respresenting the network adjacency matrix of dimension equaling the number of genes (1 interaction, 0 otherwise)
sign	A numeric vector indicating the interaction type for each link (1 activation, -1 inhibition) in the interaction network for the pathway.
x	pathway object
y	missing (placeholder)
highlight.genes	vector of gene names or node id's, which should be highlighted in a different color, default is NULL so that no genes are highlighted
gene.names	character indicating whether the genes names should appear in a legend ('legend'), as vertex label ('nodes'), or should be omitted (NA)
main	optional overall main title, default is NULL, which uses the pathway id
asp	a numeric constant, which gives the aspect ratio parameter for plot, default is 0.95
vertex.size	a numeric constant specifying the vertex size, default is 11
vertex.color	a character or numeric constant specifying the vertex color, default is 'khaki1'
vertex.label.cex	a numeric constant specifying the the vertex label size, default is 0.8,
edge.width	a numeric constant specifying the edge width, default is 2
edge.color	a character or numeric constant specifying the edge color, default is 'olive-drab4'
no	a numeric constant specifying the number of genes to be sampled, default is 3

Value

pathway2igraph returns an unweighted igraph object with edge attribute sign

analyze returns a data.frame consisting of

id pathway id,

vcount number of genes,

ecount number of links,

inh_ecount number of inhibition links,

density network density,

av_deg average degree,

inh_deg average degree of inhibition links,

diam network diamter,

trans transitivity, and

s_trans signed transitivity (Kunegis et al., 2009).

get_genes returns a character vector of gene names extracted from adjacency matrix rownames.

sample_genes returns a vector of length no with vertex id's of sampled genes

Methods (by generic)

- analyze(pathway):
- get_genes(pathway):
- sample_genes(pathway):

Slots

id A character repesenting the pathway id, e.g. hsa00100 as used in the KEGG database.

adj A matrix respresenting the network adjacency matrix of dimension equaling the number of genes (1 interaction, 0 otherwise)

sign A numeric vector indicating the interaction type for each link (1 activation, -1 inhibition) in the interaction network for the pathway.

Author(s)

Juliane Manitz, Stefanie Friedrichs, Patricia Burger

References

Details to the computation and interpretation can be found in:

- Kolaczyk, E. D. (2009). Statistical analysis of network data: methods and models. Springer series in statistics. Springer.
- Kunegis, J., A. Lommatzsch, and C. Bauckhage (2009). The slashdot zoo: Mining a social network with negative egdes. In Proceedings of the 18th international conference on World wide web, pp. 741-750. ACM Press.

Examples

```
# pathway object constructor
pathway(id="hsa04022")

# convert to igraph object
data(hsa04020)
str(hsa04020)
g <- pathway2igraph(hsa04020)
str(g)

# analyze pathway network properties
data(hsa04020)
summary(hsa04020)
analyze(hsa04020)

# extract gene names from pathway object
get_genes(hsa04020)

# plot pathway as igraph object
plot(hsa04020)
sample3 <- sample_genes(hsa04020, no = 3)
```

```
plot(hsa04020, highlight.genes = sample3)

# sample effect genes
sample3 <- sample_genes(hsa04020, no = 3)
plot(hsa04020, highlight.genes = sample3)
sample5 <- sample_genes(hsa04020, no = 5)
plot(hsa04020, highlight.genes = sample5)
```

`pathway_info` *An S4 class for an object assigning genes to pathways*

Description

This function lists all genes formig a particular `pathway`. Start and end positions of these genes are extracted from the Ensemble database. The database is accessed via the R-package **biomaRt**.

Usage

```
pathway_info(x)

## S4 method for signature 'character'
pathway_info(x)

## S4 method for signature 'pathway_info'
show(object)

## S4 method for signature 'pathway_info'
summary(object)
```

Arguments

`x`	A `character` identifying the pathway for which gene infomation should be extracted. Here KEGG IDs (format: 'hsa00100') are used.
`object`	An object of class `pathway_info`.

Value

A `data.frame` including as many rows as genes appear in the `pathway`. for each gene its name, the start and end point and the chromosome it lies on are given.

`show` Basic information on `pathway_info` object.

`summary` Summarized information on `pathway_info` object.

Slots

`info` A `data.frame` including information on genes contained in pathways with columns 'pathway', 'gene_start', 'gene_end', 'chr' and 'gene'.

Author(s)

Stefanie Friedrichs, Juliane Manitz

See Also

snp_info, get_anno

Examples

```
data(hsa04022_info) # pathway_info('hsa04020')
show(hsa04022_info)
summary(hsa04022_info)
```

---

pheno *Example phenotype file for 50 individuals.*

Description

A dataset containing simulated example phenotypes for 50 individuals row names include the identifiers of 50 example individuals.

Usage

```
data(pheno)
```

Format

A data frame with 50 rows and 3 variables:

pheno includes the case-control status for each individual, coded as 1(case) or 0 (control)

sex includes gender information for the 50 individuals, coded as 1 (male) or 0 (female)

age numerical value giving the persons age

Source

simulated data

Examples

```
data(pheno)
head(pheno)
# create gwas object
data(geno)
data(anno)
gwas <- new('GWASdata', pheno=pheno, geno=geno, anno=anno, desc="some study")
```

`read_geno,character-method`
read genotype data from file to one of several available objects, which can be passed to a GWASdata object `GWASdata`.

Description

read genotype data from file to one of several available objects, which can be passed to a GWASdata object `GWASdata`.

Usage

```
## S4 method for signature 'character'
read_geno(
  file.path,
  save.path = NULL,
  sep = " ",
  header = TRUE,
  use.fread = TRUE,
  use.big = FALSE,
  row.names = FALSE,
  ...
)
```

Arguments

`file.path`	`character` giving the path to the data file to be read
`save.path`	`character` containing the path for the backingfile
`sep`	`character`. A field delimeter. See `read.big.matrix` for details.
`header`	`logical`. Does the data set contain column names?
`use.fread`	`logical`. Should the dataset be read using the function `fread` `fread` from package **data.table**?
`use.big`	`logical`. Should the dataset be read using the function `read.big.matrix` from package **bigmemory**?
`row.names`	`logical`. Does the dataset include rownames?
`...`	further arguments to be passed to `read_geno`.

Details

If the data set contains rownames specified, set option `has.row.names = TRUE`.

Examples

```
## Not run:
path <- system.file("extdata", "geno.txt", package = "kangar00")
geno <- read_geno(path, save.path = getwd(), sep = " ", use.fread = FALSE, row.names = FALSE)

## End(Not run)
```

`rewire_network` *Rewires interactions in a* `pathway`*, which go through a gene not represented by any SNPs in the considered* `GWASdata` *dataset.*

Description

Rewires interactions in a `pathway`, which go through a gene not represented by any SNPs in the considered `GWASdata` dataset.

Usage

```
## S4 method for signature 'pathway'
rewire_network(object, x)
```

Arguments

`object`	`pathway` object which's network `matrix` will be rewired
`x`	A `vector` of gene names, indicating which genes are not represented by SNPs in the considered `GWASdata` object and will be removed

Value

A `pathway` object including the rewired network `matrix`

Author(s)

Juliane Manitz, Stefanie Friedrichs

Examples

```
## Not run:
data(hsa04020)
summary(hsa04020)
hsa04020_rewired <- rewire_network(hsa04020, x=c('ADCY3', 'CALML3','GNAQ'))
summary(hsa04020_rewired)

## End(Not run)
```

rs10243170_info *Example* snp_info *object for SNP rs10243170.*

Description

An object of class snp_info for rs10243170.

Usage

```
data(rs10243170_info)
```

Format

A snp_info object including information on the SNP as extracted from the Ensembl database.

info a data frame including the extracted information on the SNP. Columns given are 'chr', 'position', and 'rsnumber'

Source

Ensembl extract

Examples

```
## Not run:
snp_info("rs10243170")

## End(Not run)
```

snp_info *An S4 class for an object assigning SNP positions to rs-numbers (for internal use)*

Description

An S4 class for an object assigning SNP positions to rs-numbers (for internal use)

This function gives for a vector of SNP identifiers the position of each SNP as extracted from the Ensemble database. The database is accessed via the R-package **biomaRt**.

show Shows basic information on snp_info object

summary Summarizes information on snp_info object

Usage

```
snp_info(x, ...)

## S4 method for signature 'character'
snp_info(x)

## S4 method for signature 'snp_info'
show(object)

## S4 method for signature 'snp_info'
summary(object)
```

Arguments

x	A character vector of SNP rsnumbers for which positions will be extracted.
...	further arguments can be added.
object	An object of class snp_info.

Value

A data.frame including the SNP positions with columns 'chromosome', 'position' and 'snp'. SNPs not found in the Ensemble database will not be listed in the returned snp_info object, SNPs with multiple positions will appear several times.

show Basic information on snp_info object.

summary Summarized information on snp_info object.

Slots

info A data.frame including information on SNP positions

Author(s)

Stefanie Friedrichs

See Also

pathway_info, get_anno

Examples

```
data(rs10243170_info) # snp_info("rs10243170")

rs10243170_info
summary(rs10243170_info)
```

Index