

Package ‘kangaroo’

January 9, 2017

Type Package

Title Kernel Approaches for Nonlinear Genetic Association Regression

Version 0.5

Date 2016-12-12

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

Maintainer Juliane Manitz <jmanitz@uni-goettingen.de>

Description This package includes methods to extract information on pathways, genes and 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

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

Depends R (>= 3.1.0)

Imports methods,
igraph,
KEGGgraph,
biomaRt,
sqldf,
CompQuadForm,
lattice

LazyData true

Suggests knitr,
rmarkdown

VignetteBuilder knitr

NeedsCompilation no

RoxygenNote 5.0.1

R topics documented:

kangar00-package	2
anno	3
calc_kernel,GWASdata-method	4
geno	6
get_anno,snp_info,pathway_info-method	6
get_network_matrix,pathway-method	7
gwas	8
GWASdata	8
hsa04020	10
hsa04022_info	10
kernel-class	11
lkmt-class	12
lkmt.net.kernel.hsa04020	13
lkmt_test	13
lowrank_kernel-class	15
make_psd,matrix-method	16
net.kernel.hsa04020	16
pathway	17
pathway_info	20
pheno	21
read_geno,character-method	21
rewire_network	22
rs10243170_info	23
snp_info	23
Index	25

kangar00-package	<i>kangar00 package</i>
------------------	-------------------------

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

Details

Package:	kangar00
Type:	Package
Version:	1.0
Date:	2015-02-04

License: GPL-2
Depends: methods, igraph, biomaRt, KEGGgraph, sqldf

Author(s)

Juliane Manitz, Stefanie Friedrichs, Patricia Burger, Benjamin Hofner
Maintainer: Juliane Manitz <jmanitz@uni-goettingen.de>

References

- Schaid DJ: Genomic similarity and kernel methods I: advancements by building on mathematical and statistical foundations. Hum Hered 2010, 70:109-131.
- 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.
- 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
- Fiona Cunningham, M. Ridwan Amode, Daniel Barrell et al. Ensembl 2015. Nucleic Acids Research 2015 43 Database issue:D662-D669
- 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).
- 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.
- 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.

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 chromosom

snp includes rs-numbers of example SNPs

position gives positions of example SNPs

Source

simulated data

```
calc_kernel,GWASdata-method
```

Calculates the kernel-matrix for a pathway

Description

Uses individuals' genotypes to calculate a 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 kernelfunction. 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 kernelmatrix. 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 low-rank kernel will be computed.
type	character indicating the kernel type: Use "lin" for linear kernel, "sia" for size-adjusted or "net" for network-based kernel.
calculation	character specifying if the kernel matrix is computed in 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 lowrank_kernel will be returned, which is not necessarily quadratic and symmetric.

Methods (by class)

- GWASdata: Calculates a linear kernel
- GWASdata: Calculates a size adjusted-kernel
- GWASdata: Calculates a network-based kernel

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)
calc_kernel(gwas, hsa04020, knots = NULL, type=net, calculation=cpu)
```

geno	<i>Example genotypes for 50 individuals.</i>
------	--

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

get_anno,snp_info,pathway_info-method	<i>Annotates SNPs via genes to pathways</i>
---------------------------------------	---

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 "rsnumber".
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 arguments can be added.

Value

A data.frame including 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)
data(rs10243170_info)
get_anno(rs10243170_info, hsa04022_info)
```

get_network_matrix,pathway-method

Function to calculate the network matrix

Description

This function creates the networkmatrix representing the gene-gene interaction structure within a particular pathway. In this process a KEGG kgml file is downloaded and saved in the working directory.

Usage

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

Arguments

object	A pathway object identifying the pathway for which gene interaction information should be extracted. Here, KEGG IDs of format 'hsa00100' are used and information is downloaded from the KEGG database.
directed	A logic argument, stating whether the network matrix should be returned directed (TRUE) or undirected (FALSE).

Value

The altered pathway object, in which the slots 'adj' and 'sign' have been changed according to the downloaded information on the pathway.

Author(s)

Stefanie Friedrichs, Patricia Burger

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 genotypes for the individuals listed in geno

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

Source

simulated data

GWASdata

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

Description

S4 class for an object representing a Genome-wide Association 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.

Slots

geno An object of any type, including genotype information. The format needs to be one line per individual and on column 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 GWAS 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

```

data(pheno)
data(geno)
data(anno)
gwas <- new(GWASdata, pheno=pheno, geno=geno, anno=anno, desc="some study")
# show method
data(gwas)
gwas

```

```
# summary method
data(gwas)
summary(gwas)
# SNPs and genes in pathway
data(gwas)
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 dimension 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

kernel-class	<i>An S4 class representing the kernel of a pathway</i>
--------------	---

Description

An S4 class representing the kernel of 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 kernels, "sia" for size-adjusted or "net" for network-based kernels.

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

pathway A pathway object

Author(s)

Juliane Manitz

Examples

```
data(net.kernel.hsa04020)
show(net.kernel.hsa04020)
summary(net.kernel.hsa04020)
plot(net.kernel.hsa04020)
```

lkmt-class

An S4 class to represent the variance component test.

Description

An S4 class to represent the variance component test.

show Shows basic information on lkmt object

summary Summarizes information on lkmt object

Usage

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

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

Arguments

object	An object of class lkmt.
...	Further arguments can be added to the function

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 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

```
# show method
data(lkmt.net.kernel.hsa04020)
lkmt.net.kernel.hsa04020
# 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 the 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

```
lkmt_test
```

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

Description

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

For parameter "satt" a pathways influence on the probability of beeing a case is evaluated in the logistic kernel machine test and p-values are determined using a Satterthwaite 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 <code>kernel</code> including the pathway representing kernel-matrix based on which the test statistic will be calculated.
GWASdata	A <code>GWASdata</code> object stating the data used in analysis.
method	A character specifying which method will be used for p-value calculation. Available are <code>satt</code> for the Satterthwaite approximation and <code>davies</code> for Davies' algorithm. For more details see the references.
...	Further arguments can be given to the function.
x1	A <code>matrix</code> which is the similarity matrix calculated for the pathway to be tested.
x2	An <code>lm</code> or <code>glm</code> 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(net.kernel.hsa04020)
data(gwas)
lkmt_test(pheno ~ sex + age, net.kernel.hsa04020, gwas, method=satt)
```

lowrank_kernel-class	<i>An S4 class to represent a low-rank kernel for a SNPset at specified knots</i>
----------------------	---

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 linear kernels, "sia" for size-adjusted or "net" for network-based kernels.

kernel kernel matrix of dimension equal to individuals

pathway pathway object

Author(s)

Juliane Manitz

Examples

```
data(gwas)
K.lin.knots <- calc_kernel(gwas, hsa04020, knots=gwas, type=lin, calculation=cpu)
```

```
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 tolerance 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 λ 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.

```
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

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 of character containing gene names

`plot` plots pathway as `igraph` object

`sample_genes` function randomly selects effect genes in pathway and returns a vector of length `no` with vertex id's of sampled genes

Usage

```
pathway(object, ...)
```

```
## S4 method for signature ANY
```

```
pathway(id, adj, sign)
```

```
## 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("legend", "nodes", NA), 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

<code>object</code>	An object of class <code>pathway-class</code>
<code>...</code>	further arguments specifying plotting options in plot.igraph
<code>id</code>	A character representing the pathway id.
<code>adj</code>	A matrix representing the network adjacency matrix of dimension equaling the number of genes (1 interaction, 0 otherwise)
<code>sign</code>	A numeric vector indicating the interaction type for each link (1 activation, -1 inhibition) in the interaction network for the pathway.
<code>x</code>	pathway object
<code>y</code>	missing (placeholder)
<code>highlight.genes</code>	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
<code>gene.names</code>	character indicating whether the genes names should appear in a legend (legend), as vertex label (nodes), or should be omitted (NA)
<code>main</code>	optional overall main title, default is NULL, which uses the pathway id
<code>asp</code>	a numeric constant, which gives the aspect ratio parameter for plot, default is 0.95
<code>vertex.size</code>	a numeric constant specifying the vertex size, default is 11
<code>vertex.color</code>	a character or numeric constant specifying the vertex color, default is 'khaki1'
<code>vertex.label.cex</code>	a numeric constant specifying the the vertex label size, default is 0.8,
<code>edge.width</code>	a numeric constant specifying the edge width, default is 2
<code>edge.color</code>	a character or numeric constant specifying the edge color, default is 'olivedrab4'
<code>no</code>	a numeric constant specifying the number of genes to be sampled, default is 3

Value

`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 diameter,
trans transitivity, and
s_trans signed transitivity (Kunegis et al., 2009).

Methods (by generic)

- `get_genes`:

Slots

id A character representing the pathway id, e.g. hsa00100 as used in the KEGG database.
adj A matrix representing 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

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 edges. In Proceedings of the 18th international conference on World wide web, pp. 741-750. ACM Press.

Examples

```
pathway(id="hsa04022", adj=matrix(0), sign=as.vector(matrix(0)[matrix(0)!=0]))
#show method
data(hsa04020)
hsa04020
#summary method
data(hsa04020)
summary(hsa04020)
# convert to igraph object
data(hsa04020)
str(hsa04020)
g <- pathway2igraph(hsa04020)
str(g)
# analyse pathway network properties
data(hsa04020)
summary(hsa04020)
analyze(hsa04020)
# extract gene names from pathway
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

An S4 class for an object assigning genes to pathways

This function lists all genes forming 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.

show Shows basic information on pathway_info object

summary Summarizes information on pathway_info object

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 information should be extracted. Here KEGG IDs (format: 'hsa00100') are used.
object	An object of class <code>pathway_info</code> .

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

Examples

```
pathway_info("hsa04022")
# show method
data(hsa04022_info)
hsa04022_info
# summary method
data(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

read_genotype, character-method

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

Description

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

Usage

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

Arguments

file.path	character, which contains the path to the data file to be read
save.path	character, which contains the path for the backingfile
sep	character. A field delimiter. See <code>bigmemory::read.big.matrix</code> for details.
header	logical. Does the data set contain column names?
use.fread	logical. Should the dataset be read using the package <code>fread</code> ?
row.names	logical. Does the dataset include rownames?
...	further arguments to be passed to <code>bigmemory::read.big.matrix</code> .

Details

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

rewire_network	<i>Rewires interactions in a pathway, which go through a gene not represented by any SNPs in the considered GWAS dataset.</i>
----------------	---

Description

Rewires interactions in a pathway, which go through a gene not represented by any SNPs in the considered GWAS 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 and will be removed

Value

A pathway object including the rewired network

Author(s)

Stefanie Friedrichs, Juliane Manitz

rs10243170_info	<i>Example snp_info object for SNP rs10243170.</i>
-----------------	--

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

snp_info	<i>An S4 class for an object assigning SNP positions to rs-numbers (for internal use)</i>
----------	---

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 Ensembl 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 <code>snp_info</code> .

Value

A `data.frame` including the SNP positions with columns 'chromosome', 'position' and 'rsnumber'. 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

Examples

```
# snp_info
data(rs10243170_info)
snp_info(c("rs234"))
# show
data(rs10243170_info)
rs10243170_info
# summary
data(rs10243170_info)
summary(rs10243170_info)
```


Index

*Topic **datasets**

- anno, [3](#)
- geno, [6](#)
- gwas, [8](#)
- hsa04020, [10](#)
- hsa04022_info, [10](#)
- lkmt.net.kernel.hsa04020, [13](#)
- net.kernel.hsa04020, [16](#)
- pheno, [21](#)
- rs10243170_info, [23](#)

*Topic **package**

- kangar00-package, [2](#)

- analyze,pathway,ANY-method (pathway), [17](#)
- analyze,pathway-method (pathway), [17](#)
- anno, [3](#)

- calc_kernel,GWASdata-method, [4](#)

- davies_test,matrix-method (lkmt_test), [13](#)

- GeneSNPsize,GWASdata-method (GWASdata), [8](#)

- geno, [6](#)
- get_anno,snp_info,pathway_info-method, [6](#)
- get_genes,pathway,ANY-method (pathway), [17](#)

- get_genes,pathway-method (pathway), [17](#)
- get_network_matrix,pathway-method, [7](#)
- gwas, [8](#)

- GWASdata, [8](#), [8](#), [12](#), [14](#)
- GWASdata,ANY-method (GWASdata), [8](#)

- hsa04020, [10](#)
- hsa04022_info, [10](#)

- igraph, [17](#)

- kangar00 (kangar00-package), [2](#)
- kangar00-package, [2](#)
- kernel, [12](#), [14](#), [15](#)
- kernel (kernel-class), [11](#)
- kernel-class, [11](#)

- lin_kernel,GWASdata-method
(calc_kernel,GWASdata-method), [4](#)

- lkmt (lkmt-class), [12](#)

- lkmt-class, [12](#)

- lkmt.net.kernel.hsa04020, [13](#)

- lkmt_test, [13](#)

- lowrank_kernel (lowrank_kernel-class), [15](#)

- lowrank_kernel-class, [15](#)

- make_psd,matrix-method, [16](#)
- matrix, [14](#)

- net.kernel.hsa04020, [16](#)
- net_kernel,GWASdata-method
(calc_kernel,GWASdata-method), [4](#)

- pathway, [6](#), [17](#), [17](#)

- pathway,ANY-method (pathway), [17](#)

- pathway2igraph,pathway,ANY-method
(pathway), [17](#)

- pathway2igraph,pathway-method
(pathway), [17](#)

- pathway_info, [7](#), [20](#), [20](#)

- pathway_info,character-method
(pathway_info), [20](#)

- pheno, [21](#)

- plot,kernel,ANY-method (kernel-class), [11](#)

- plot,kernel,missing-method
(kernel-class), [11](#)

- plot,pathway,ANY-method (pathway), [17](#)

- plot,pathway,missing-method (pathway), [17](#)

- plot.igraph, [18](#)

- read_geno,character-method, [21](#)

- rewire_network, [22](#)

- rewire_network,pathway-method
(rewire_network), [22](#)

- rs10243170_info, [23](#)

sample_genes, pathway-method (pathway),
 17
 score_test, matrix-method (lkmt_test), 13
 show, GWASdata, ANY-method (lkmt-class),
 12
 show, GWASdata-method (GWASdata), 8
 show, kernel-method (kernel-class), 11
 show, lkmt-method (lkmt-class), 12
 show, pathway, ANY-method (pathway), 17
 show, pathway-method (pathway), 17
 show, pathway_info, ANY-method
 (pathway_info), 20
 show, pathway_info-method
 (pathway_info), 20
 show, snp_info-method (snp_info), 23
 sia_kernel, GWASdata-method
 (calc_kernel, GWASdata-method),
 4
 snp_info, 7, 23, 24
 snp_info, character-method (snp_info), 23
 summary, GWASdata, ANY-method
 (lkmt-class), 12
 summary, GWASdata-method (GWASdata), 8
 summary, kernel, ANY-method
 (kernel-class), 11
 summary, kernel-method (kernel-class), 11
 summary, lkmt-method (lkmt-class), 12
 summary, pathway, ANY-method (pathway), 17
 summary, pathway-method (pathway), 17
 summary, pathway_info, ANY-method
 (pathway_info), 20
 summary, pathway_info-method
 (pathway_info), 20
 summary, snp_info-method (snp_info), 23