Package ‘kangar00’

February 17, 2020

Type Package
Title Kernel Approaches for Nonlinear Genetic Association Regression
Version 1.4
Date 2020-02-12
Author 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.org>
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 (Friedrichs et al., 2017, <doi:10.1155/2017/6742763>).
License GPL-2
Collate 'pathway.r' 'GWASdata.r' 'data.R' 'get_network_matrix.r' 'kernel.r' 'lkmt.r'
Depends R (>= 3.1.0)
Imports methods, bigmemory, sqldf, biomaRt, KEGGgraph, CompQuadForm, data.table, lattice, igraph
LazyData true
Suggests testthat
NeedsCompilation no
RoxygenNote 6.1.1
Repository CRAN
Date/Publication 2020-02-17 09:10:08 UTC

R topics documented:

kangar00-package

1
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


---

**anno**  
*Example annotation file for three pathways.*

---

**Description**

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

**Usage**

```r
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**

```r
data(anno)
head(anno)
# create gwas object
data(pheno)
data(geno)
gwas <- new('GWASdata', pheno=pheno, geno=geno, anno=anno, desc="some study")
```
**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’ genotype vectors of the SNPs within the pathway by a kernel function. It can be interpreted as the genetic similarity 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**

```r
## 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 `lowrank_kernel` will be returned, which is not necessarily quadratic and symmetric.

Methods (by class)

- `GWASdata`

Author(s)

Stefanie Friedrichs, Juliane Manitz

References


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(phen0)
data(anno)
gwas <- new('GWASdata', pheno=phen0, geno=geno, anno=anno, desc="some study")
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

```r
## 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 arguments

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

```r
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 information 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

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

Description

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

'S4data' 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, ...)

```r
## 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**: 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 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 `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

```r
# 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
kernel-class

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

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

A vector containing the number of degrees of freedom. 

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


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)

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 the 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')

Description

For parameter 'satt' a pathway's influence on the probability of being 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 being 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"), ...)
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:


For details on the p-value calculation see:

Examples

```r
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

```r
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 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 $\lambda$ 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


Examples

```r
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) neighbors

Usage

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

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

```r
## 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 representing the `pathway` id.
- `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`. 
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,

eedge.width  
  a numeric constant specifying the edge width, default is 2

dedge.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 diameter,

trans  transitivity, and

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

gene_names 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:
- get_genes:
- sample_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, Stefanie Friedrichs, Patricia Burger

References

Details to the computation and interpretation can be found in:


Examples

```r
# 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
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)
```
**Description**

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**.

**Usage**

```r
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 **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**

```r
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**

```r
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 person's age

**Source**

simulated data

**Examples**

```r
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**

```r
## 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 delimiter. 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` 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**

```r
## Not run:
path <- system.file("extdata", "geno.txt", package = "kangaroo")
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

```r
## 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

```r
## 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 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
snp_info

Usage

snp_info(x, 
## S4 method for signature 'character'
  x
## S4 method for signature 'snp_info'
  object
## S4 method for signature 'snp_info'
  object

Arguments

x A character vector of SNP rsnumbers for which positions will be extracted.
...
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

*Topic **datasets**
  anno, 4
  geno, 7
  gwas, 9
  hsa04020, 12
  hsa04022_info, 12
  lkmt.net.kernel.hsa04020, 15
  net.kernel.hsa04020, 20
  pheno, 25
  rs10243170_info, 28

*Topic **package**
  kangar00-package, 2
  analyze (pathway), 20
  analyze, pathway-method (pathway), 20
  anno, 4
  ANY-method (pathway), 20
  calc_kernel, 5
  calc_kernel, GWASdata-method
    (calc_kernel), 5
  character (read_geno, character-method), 26
  davies, 16
  davies_test, matrix-method (lkmt_test), 16
  fread, 26

GeneSNPsize (GWASdata), 10
GeneSNPsize, GWASdata-method (GWASdata), 10
get_genes, pathway-method (pathway), 20
get_network_matrix, 9
get_network_matrix, pathway-method
  (get_network_matrix), 9
gwas, 9
GWASdata, 8–10, 10, 11, 15–17, 26, 27
GWASdata, ANY-method (GWASdata), 10
hsa04020, 12
hsa04022_info, 12
igraph, 20–22
kangar00 (kangar00-package), 2
kangar00-package, 2
kernel, 5, 6, 14–18, 20
kernel (kernel-class), 13
kernel-class, 13
lin_kernel (calc_kernel), 5
lin_kernel, GWASdata-method
  (calc_kernel), 5
lkmt, 15, 16
lkmt (lkmt-class), 14
lkmt-class, 14
lkmt.net.kernel.hsa04020, 15
lkmt_test, 16
lowrank_kernel (lowrank_kernel-class), 18
lowrank_kernel-class, 18
make_psd (make_psd, matrix-method), 19
make_psd, matrix-method, 19
matrix, 17
matrix (make_psd, matrix-method), 19
net.kernel.hsa04020, 20
net_kernel (calc_kernel), 5
net_kernel, GWASdata-method
  (calc_kernel), 5
INDEX

pathway, 5, 6, 8–13, 15, 16, 18, 20, 20, 21–24, 27
pathway, ANY-method (pathway), 20
pathway2igraph (pathway), 20
pathway2igraph, pathway-method (pathway), 20
pathway_info, 8, 12, 24, 24, 29
pathway_info, character-method (pathway_info), 24
pheno, 25
plot, kernel, ANY-method (kernel-class), 13
plot, kernel, missing-method (kernel-class), 13
plot, pathway, ANY-method (pathway), 20
plot, pathway, missing-method (pathway), 20
read.big.matrix, 26
read_geno (read_geno, character-method), 26
read_geno, character-method, 26
rewire_network, 27
rewire_network, pathway-method (rewire_network), 27
rs10243170_info, 28
sample_genes (pathway), 20
sample_genes, pathway-method (pathway), 20
score_test, matrix-method (lkmt_test), 16
show, GWASdata, ANY-method (lkmt-class), 14
show, GWASdata-method (GWASdata), 10
show, kernel, ANY-method (kernel-class), 13
show, lkmt-method (lkmt-class), 14
show, pathway, ANY-method (pathway), 20
show, pathway-method (pathway), 20
show, pathway_info, ANY-method (pathway_info), 24
show, pathway_info-method (pathway_info), 24
show, snp_info-method (snp_info), 28
sia_kernel (calc_kernel), 5
sia_kernel, GWASdata-method (calc_kernel), 5
snp_info, 8, 24, 28, 28, 29
snp_info, character-method (snp_info), 28
summary, GWASdata, ANY-method (lkmt-class), 14
summary, GWASdata-method (GWASdata), 10
summary, kernel, ANY-method (kernel-class), 13
summary, lkmt-method (lkmt-class), 14
summary, pathway, ANY-method (pathway), 20
summary, pathway-method (pathway), 20
summary, pathway_info, ANY-method (pathway_info), 24
summary, pathway_info-method (pathway_info), 24
summary, snp_info-method (snp_info), 28