Package ‘BrainGEP’

January 8, 2015

Type Package

Title Prioritises genes using adult and developing Allen Human Brain Atlas (AHBA) normalised microarray data

Version 1.0.4

Date 2015-1-8

Author Vesna Lukic

Maintainer Vesna Lukic <lukic@wehi.edu.au>

Depends R (>= 2.10)

Imports gtools, qgraph, corrplot, MASS, reshape2

Description Prioritises genes using Allen Human Brain Atlas data. The adult normalised data can be downloaded from http://human.brain-map.org/static/download (under 'Complete normalized microarray datasets'). The developing normalised data can be downloaded from http://www.brainspan.org/static/download.html (under 'Prenatal LMD Microarray Dataset').

License GPL-2

LazyData yes

Suggests RUnit, BiocGenerics

R topics documented:

BrainGEP-package *BrainGEP*

Description

Prioritises genes using Allen Human Brain Atlas data, for both adult and developing brains. Includes reference and candidate Epileptic Encephalopathy genes and the Celsius data set for this collection of genes

Details

Package:	BrainGEP
Type:	Package
Version:	1.0
Date:	2014-04-30
License:	GPL-2

Author(s)

Vesna Lukic

Maintainer: Vesna Lukic (lukic@wehi.edu.au)

References

http://human.brain-map.org/ http://www.brainspan.org/ http://genome.ucla.edu/projects/UGET

AHBA_corrplot *AHBA_corrplot*

Description

Visualisation of a correlation matrix for a set of input genes, the the weighted results of the adult human brains if T="AHB", and for the developing human brains if T="DHB"

Usage

```
AHBA_corrplot(T, Y, C, R)
```

Arguments

T	"AHB" or "DHB"
Y	Reference genes
C	"pearson" or "spearman"
R	Regional analysis option for cortex (CX), cerebellum (CB) and brainstem (BS), for adult AHBA only

Author(s)

Vesna Lukic

References

http://human.brain-map.org/ http://www.brainspan.org/

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Original_package_data/"

Load_AHBA_data(path)

Ref_genes<-c("PTEN","MEF2C","SCN2A","SCN5A","SCN8A")

AHBA_corrplot("DHB",Ref_genes,"pearson","R")
```

AHBA_ECDF *AHBA_ECDF*

Description

Draws an ECDF curve for pairwise combinations of reference genes against the pairwise combinations of 1000 randomly chosen genes representing the null, for the weighted adult human brains if T="AHB" and for the weighted developing human brains if T="DHB"

Usage

```
AHBA_ECDF(T, Y, C, R, Re)
```

Arguments

T	"DHB" or "AHB"
Y	Reference genes
C	Correlation method "pearson" or "spearman"
R	Vector of a sufficient number of random genes to represent the null
Re	Regional analysis option for CX (cortex), BS (brainstem) and CB (cerebellum)

Author(s)

Vesna Lukic

References

http://human.brain-map.org/ http://www.brainspan.org/

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Original_package_data"

Load_AHBA_data(path)

Ref_genes<-c(" SCN8A "," PTEN ", " SCN2A ")

AHBA_ECDF("AHB",Ref_genes,"spearman",unlist(random_gene_symbol_1000),"R")
```

AHBA_eFDR *AHBA_eFDR*

Description

This function prioritises the true candidate genes as well as N sets of randomly chosen candidates genes (of the same size as the true candidate set), and plots the average of the randomly chosen candidate sets' connectivity to the reference genes. This number is divided by the true candidate genes connectivity to the reference genes. The result is an eFDR graph, plotted with the connectivity on the x-axis.

Usage

```
AHBA_eFDR(T, N, X, Y, C, P, Rval, S)
```

Arguments

T	"DHB" or "AHB"
N	The number of randomly chosen candidate sets to use
X	Reference genes
Y	Candidate genes
C	"pearson" or "spearman"
P	Percentile value (threshold) above which significance is declared.
Rval	Allows user to input their own value of R based on calculated percentile
S	"sig" for significant connections (recommended)

Author(s)

Vesna Lukic

References

Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing Author(s): Yoav Benjamini and Yosef Hochberg

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Package_data1/"

Load_AHBA_data(path)

Ref_genes<-c("SCN8A "," PTEN "," SCN2A ")

Cand_genes<-c(" SCN5A "," MEF2C ")

AHBA_eFDR("DHB",5,Ref_genes,Cand_genes,"spearman","NA",0.1,"sig")
```

AHBA_heatmap *AHBA_heatmap*

Description

Draws a heatmap for a set of input reference genes, for the 6 adult human brains if T="AHB" and the 4 developing human brains if T="DHB"

Usage

```
AHBA_heatmap(T, Y, R)
```

Arguments

T	"DHB" or "AHB"
Y	Reference genes
R	If "R" is given as an argument, a heatmap will be drawn for each of the cortex, cerebellum and brainstem regions

Author(s)

Vesna Lukic

References

http://human.brain-map.org/ http://www.brainspan.org/

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Original_package_data"

Load_AHBA_data(path)

Ref_genes<-c(" SCN8A "," PTEN ", " SCN2A ")

AHBA_heatmap("DHB",Ref_genes,"NA")
```

AHBA_network *AHBA_network*

Description

Plots a network of the input genes, for a given correlation measure from the weighted correlations of the adult human brains if T="AHB", and for the developing human brains if T="DHB". The 'P' argument is the percentile above which connections are declared significant. The significant network will only be plotted if there are more than 2 significant connections between genes. Needs at least 4 input reference genes

Usage

```
AHBA_network(T,Y,C,P,R,Rval,Col,Lay)
```

Arguments

T	"DHB" or "AHB"
Y	Set of reference, or input genes
C	Correlation measure, either "pearson" or "spearman"
P	Percentile value (threshold) above which significance is declared.
R	Regional analysis option for cortex (CX), cerebellum (CB) and brainstem (BS), for adult AHBA only
Rval	Allows user to input their own value of R based on calculated percentile
Col	Specify vector of colours for the genes
Lay	Layout of network "spring" or "circular"

Value

A network with all the input genes, a network with the significant connections between the input genes, a csv file containing the pairwise input genes and corresponding R values

Author(s)

Vesna Lukic

References

http://human.brain-map.org/ http://www.brainspan.org/

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Original_package_data/"

Load_AHBA_data(path)

Ref_genes<-c(" SCN8A "," PTEN "," SCN2A "," SCN5A "," MEF2C ")

AHBA_network("AHB",Ref_genes,"pearson",0.5,"NA","NA","NA","spring")
```

AHBA_pairwise_gene_correlations

AHBA_pairwise_gene_correlations

Description

Plots one gene expression against another, for all pairwise combinations of input genes, for the 4 developing and 6 adult brains. Points are coloured according to region of origin within the brain for the adult brains only, as the spatial brain sample information is available.

Usage

```
AHBA_pairwise_gene_correlations(T, Y, R)
```

Arguments

T	"DHB" or "AHB"
Y	Reference genes
R	Regional analysis option for cortex (CX), cerebellum (CB) and brainstem (BS), for adult AHBA only

Value

Returns a pdf file of the pairwise gene expression patterns for the 4 "DHB" brains or the 6 "AHB" brains

Author(s)

Vesna Lukic

References

http://human.brain-map.org/ http://www.brainspan.org/

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Original_package_data/"

Load_AHBA_data(path)

Ref_genes<-c(" SCN8A "," PTEN "," SCN2A "," SCN5A "," MEF2C ")

AHBA_pairwise_gene_correlations("AHB",Ref_genes,"NA")
```

AHBA_percentile *AHBA_percentile*

Description

Takes correlation measure and desired percentile as input, and outputs the corresponding correlation value.

Usage

```
AHBA_percentile(T, Y, C, P, Rval)
```

Arguments

T	"DHB" or "AHB"
Y	Set of genes
C	Correlation measure "pearson" or "spearman"
P	The percentile desired
Rval	Allows user to specifically input their own value of R based on a calculated percentile

Value

Correlation value corresponding to correlation measure and percentile inputted

Author(s)

Vesna Lukic

References

http://human.brain-map.org/ http://www.brainspan.org/

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Package_data1/"

Load_AHBA_data(path)

AHBA_percentile("AHB",unlist(random_gene_symbol_1000),"pearson","NA",0.43642)
```

AHBA_Prioritise *AHBA_Prioritise*

Description

Prioritises candidate genes against a set of reference genes, for a given correlation measure and threshold cutoff, for the adult human brains if T="AHB", and for the developing human brains if T="DHB"

Usage

```
AHBA_Prioritise(T, X, Y, C, P, Rval, R)
```

Arguments

T	"DHB" or "AHB"
X	Reference genes
Y	Candidate genes
C	"pearson" or "spearman"
P	Percentile cut-off
Rval	Allows user to input their own value of R based on calculated percentile
R	Regional analysis option for cortex (CX), cerebellum (CB) and brainstem (BS), for adult AHBA only

Value

Returns a list of prioritised genes for using both the percentile value (P) as the threshold, as well as without any threshold

Author(s)

Vesna Lukic

References

http://human.brain-map.org/ http://www.brainspan.org/

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Original_package_data/"

Load_AHBA_data(path)

AHBA_Prioritise("DHB",EE_ref_genes,EE_cand_genes,"spearman",0.95,"NA","R")
```

AHBA_refgenes *AHBA_refgenes*

Description

Calculates pairwise correlations between the reference genes using Pearsons or Spearmans correlation measure, individually for four developing human brains if "DHB" is chosen or six adult human brains if "AHB" is chosen, as well as across the combined DHB or AHB brains using the Weighted Mean method

Usage

```
AHBA_refgenes(T, Y, C, R)
```

Arguments

T	"DHB" or "AHB"
Y	Reference genes
C	Either "pearson" or "spearman". Other correlation measures are yet to be implemented
R	Regional analysis option for cortex (CX), cerebellum (CB) and brainstem (BS), for adult AHBA only

Value

Pairwise correlations for individual and combined developing human brains

Author(s)

Vesna Lukic

References

http://human.brain-map.org/ http://www.brainspan.org/ http://en.wikipedia.org/wiki/Weighted_arithmetic_mean

---

AHB_Pearson_top_5p *AHB_Pearson_top_5p*

Description

Cut-off R value for top 5

Usage

```
data(AHB_Pearson_top_5p)
```

AHB_Spearman_top_5p *AHB_Spearman_top_5p*

Description

Cut-off R value for top 5

Usage

```
data(AHB_Spearman_top_5p)
```

CADD_with_genes *CADD_with_genes*

Description

CADD scores for EE candidate genes

Usage

```
data(CADD_with_genes)
```

Format

A data frame with 180 observations on the following 9 variables.

`CHROM`	a factor with levels 1 10 11 ...
`POS`	a numeric vector
`REF`	a factor with levels `A` `C` `G` `T`
`ALT`	a factor with levels `A` `C` `G` `T`
`RawScore`	a numeric vector
`PHRED`	a numeric vector
`Deleterious...1..else...0`	a numeric vector
`Gene`	a factor with levels `ABCB9` `ACOT4` `AGPAT3` ...
`val`	a numeric vector

Source

http://cadd.gs.washington.edu/info

References

http://cadd.gs.washington.edu/info

Examples

```
data(CADD_with_genes)
```

---

candidatelist *candidatelist*

Description

List of 182 candidate genes with extra information

Usage

```
data(candidatelist)
```

Format

A data frame with 182 observations on the following 6 variables.

- `Gene` a factor with levels `ABCB9` `ACOT4` `AGPAT3` ...
- `Study.identified` a factor with levels `Epi4K` `Epi4K, MIPS`
- `Mutation` a factor with levels `missense` `missense (benign pp2)` `missense (unknown)` `nonsense` `splice site`
- `Why.not.definite.` a factor with levels `n=1` `n=1 (Epi4k), all inherited in MIPs` `n=1, two confirmed mutations withi` `n=2, but no more than chance (large gene)`
- `Gene.intolerance` a factor with levels ` intolerant gene (Epi4K)`
- `Associations.with.other.Neurological.conditions` a factor with levels ` ASD, ID` `hemimegalencephaly` `ID` `ID, ?spasms` `MCD - lissencephaly` `MCD - PVNH` `Migraine, ataxia` `neonatal seizures` `photosensitive epilepsy`

References

Epi4K Consortium & Epilepsy Phenome/Genome Project (http://www.epgp.org/epi4k/)

Examples

```
data(candidatelist)
```

DHB_Pearson_top_5p *DHB_Pearson_top_5p*

Description

Cut-off R value for top 5

Usage

```
data(DHB_Pearson_top_5p)
```

DHB_Spearman_top_5p *DHB_Spearman_top_5p*

Description

Cut-off R value for top 5

Usage

```
data(DHB_Spearman_top_5p)
```

EE_cand_genes *EE_cand_genes*

Description

EE candidate genes

Format

The format is: chr "EE_cand_genes.rda"

Source

Epi4K

EE_ref_genes *EE_ref_genes*

Description

29 reference EE genes

Format

The format is: chr "EE_ref_genes.rda"

Source

Genes collected from thorough literature searches

Genes_in_ABA *Genes_in_ABA*

Description

This function checks for the presence of genes in the Allen Brain Atlas, returns genes that are present or returns an alternative name if found

Usage

```
Genes_in_ABA(T,x)
```

Arguments

T	"AHB" or "DHB"
x	Set of reference or candidate genes

Value

Returns list of genes present, and any alternative names found if the input gene name is not present

Note

Use this function on the set of reference and candidate genes

Author(s)

Vesna Lukic

References

http://genome.ucsc.edu/

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Package_data1"
Load_AHBA_data(path)

Genes_in_ABA("AHB","SCN8A")
```

---

gene_aliases *gene_aliases*

Description

This data is used to search for an alternative gene name if a gene is not found in the Allen Brain Atlas data

Usage

```
data(gene_aliases)
```

Format

A data frame with 640312 observations on 2 variables.

Source

http://genome.ucsc.edu/

References

http://genome.ucsc.edu/

Examples

```
data(gene_aliases)
```

---

list2df *list2df*

Description

Converts a list to a data frame

Usage

```
list2df(x)
```

Arguments

x	A list, such as pairwise_names

Value

Returns a data frame out of the list supplied

Author(s)

Vesna Lukic

References

http://www.inside-r.org/packages/cran/Kmisc/docs/list2df

Examples

```
example_list<-vector(mode="list")
k1<-0

for (i in 1:3){
for (j in 4:6) {
k1<-k1+1
  tmp<-c(i,j)
  example_list[[k1]]<-tmp
}
}

example_df<-list2df(example_list)
```

`Load_AHBA_data` *Load_AHBA_data*

Description

Loads processed and saved data

Usage

```
Load_AHBA_data(save_path)
```

Arguments

`save_path`	Path to saved data

Author(s)

Vesna Lukic

References

http://human.brain-map.org/ http://www.brainspan.org/

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Original_package_data"

Load_AHBA_data(path)
```

Mann_whitney_wilcoxon

Mann_whitney_wilcoxon

Description

Performs one-sided Mann-Whitney-Wilcoxon rank-sum tests by prioritising the EE reference genes against the EE candidate genes and using added information about whether a mutation in the candidate genes is predicted to be deleterious by sources such as the Gene Intolerance Score (GIT), PolyPhen2 predicted damaging, prior neuro gene, or has a CADD score > 25. Outputs p-values obtained from each source, including the p-values obtained using 1000 permutations of the labels with respect to the genes (to take into account the lack of independence between genes)

Usage

```
Mann_whitney_wilcoxon(T, X, Y, C, P, Rval, S)
```

Arguments

T	"DHB" or "AHB"
X	Reference genes
Y	Candidate genes
C	"pearson" or "spearman"
P	Percentile cut-off
Rval	Allows user to input their own value of R based on calculated percentile
S	Put "sig" if significant correlations are required

Author(s)

Vesna Lukic

References

David F. Bauer (1972), Constructing confidence sets using rank statistics. _Journal of the American Statistical Association_ *67*, 687-690.

Myles Hollander and Douglas A. Wolfe (1973), _Nonparametric Statistical Methods._ New York: John Wiley & Sons. Pages 27-33 (one-sample), 68-75 (two-sample). Or second edition (1999).

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Package_data1/"

Load_AHBA_data(path)

Mann_whitney_wilcoxon("DHB", EE_ref_genes, EE_cand_genes, "pearson", "NA", 0.45, "sig")
```

`merge_with_order` *merge_with_order*

Description

Merges two data frames while preserving the original row order in one data frame

Usage

```
merge_with_order(x, y, ..., sort = T, keep_order)
```

Arguments

`x`	First data frame
`y`	Second data frame
...	Additional arguments to pass to the function
`sort`	Should the resulting data frame be sorted?
`keep_order`	Put '1' if the original order is to be preserved

Author(s)

Tal Galili

References

http://www.r-statistics.com/2012/01/merging-two-data-frame-objects-while-preserving-the-rows-order/

`Process_AHBA_data` *Process_AHBA_data*

Description

Processes downloaded AHBA data for developing and adult brains, makes it suitable for analysis using all other functions in package, and saves the data. This function only needs to be run once.

Usage

```
Process_AHBA_data(DHB_path,AHB_path,save_path)
```

Arguments

`DHB_path`	Path to data set of developing AHBA data (character vector)
`AHB_path`	Path to data set of adult AHBA data (character vector)
`save_path`	Path to saved processed data (character vector)

Value

RData objects get saved that can be read in later and used for further analysis

Author(s)

Vesna Lukic

References

http://human.brain-map.org/ http://www.brainspan.org/

---

random_gene_symbol_1000 *random_gene_symbol_1000*

Description

One thousand randomly chosen genes representing the null of the DHB and AHB datasets

Usage

```
data(random_gene_symbol_1000)
```

Format

The format is: List of 1000 $: chr " AANAT " $: chr " ABCC4 " $: chr " ABCE1 " $: chr " ABHD4 " $: chr " ABR " $: chr " ABTB2 " [list output truncated]

Source

http://human.brain-map.org/

References

http://human.brain-map.org/

Examples

```
data(random_gene_symbol_1000)
```

Tally_genes *Tally_genes*

Description

Tallies gene scores for use in eFDR

Usage

```
Tally_genes(T, X, Y, C, P, Rval, S)
```

Arguments

T	"DHB" or "AHB"
X	Reference genes
Y	Candidate genes
C	"pearson" or "spearman"
P	Percentile cut-off
Rval	Allows user to input their own value of R based on calculated percentile
S	Put "sig" if significant correlations are required

Value

Produces histograms of connectivity

Author(s)

Vesna Lukic

References

http://human.brain-map.org/ http://www.brainspan.org/

Examples

```
path<-"/home/users/lab0605/Bioinformatics/SNPchipdata/AllenBrainAtlas/Package_data1/"

Load_AHBA_data(path)

Ref_genes<-c(" SCN8A "," PTEN ", " SCN2A ")
Cand_genes<-c(" SCN5A "," BRCA1 ")

Tally_genes("AHB",Ref_genes,Cand_genes,"pearson",0.95,"NA","sig")
```

---

w_ap *w_ap*

Description

Adult AHBA weightings derived from Pearson's correlation

---

w_as *w_as*

Description

Adult AHBA weightings derived from Spearman's correlation

---

w_dp *w_dp*

Description

Developing AHBA weightings derived from Pearson’s correlation

---

w_ds *w_ds*

Description

Developing AHBA weightings derived from Spearman’s correlation

Index