Package 'ABAEnrichment'

October 14, 2021

Type Package

Title Gene expression enrichment in human brain regions

Version 1.22.0

Date 2019-04-02

Author Steffi Grote

Maintainer Steffi Grote <steffi_grote@eva.mpg.de>

Description The package ABAEnrichment is designed to test for
enrichment of user defined candidate genes in the set of
expressed genes in different human brain regions. The core
function 'aba_enrich' integrates the expression of the
candidate gene set (averaged across donors) and the structural
information of the brain using an ontology, both provided by
the Allen Brain Atlas project. 'aba_enrich' interfaces the
ontology enrichment software FUNC to perform the statistical
analyses. Additional functions provided in this package like
'get_expression' and 'plot_expression' facilitate exploring the
expression data, and besides the standard candidate vs. background
gene set enrichment, also three additional tests are implemented,
e.g. for cases when genes are ranked instead of divided into
candidate and background.

License GPL (>= 2)

Imports Rcpp (>= 0.11.5), gplots (>= 2.14.2), gtools (>= 3.5.0),
ABAData (>= 0.99.2), data.table (>= 1.10.4), GOfuncR (>=
1.1.2), grDevices, stats, graphics, utils

Depends R (>= 3.4)

LinkingTo Rcpp

Suggests BiocStyle, knitr, testthat

VignetteBuilder knitr

biocViews GeneSetEnrichment, GeneExpression

NeedsCompilation yes

git_url https://git.bioconductor.org/packages/ABAEnrichment

R topics documented:

---

aba_enrich   *Test genes for expression enrichment in human brain regions*

Description

Tests for enrichment of user defined candidate genes in the set of expressed protein-coding genes in different human brain regions. It integrates the expression of the candidate gene set (averaged across donors) and the structural information of the brain using an ontology, both provided by the Allen Brain Atlas project [1-4]. The statistical analysis is performed using the ontology enrichment software FUNC [5].

Usage

```
aba_enrich(genes, dataset = 'adult', test = 'hyper',
  cutoff_quantiles = seq(0.1, 0.9, 0.1), n_randsets = 1000, gene_len = FALSE,
circ_chrom = FALSE, ref_genome = 'grch37', gene_coords = NULL, silent = FALSE)
```

Arguments

genes   A dataframe with gene-identifiers (Entrez-ID, Ensembl-ID or gene-symbol) in the first column and test-dependent additional columns:

If `test='hyper'` (default) a second column with 1 for candidate genes and 0 for background genes. If no background genes are defined, all remaining protein coding genes are used as background.

If `test='wilcoxon'` a second column with the score that is associated with each gene.

If `test='binomial'` two additional columns with two gene-associated integers.

If `test='contingency'` four additional columns with four gene-associated integers.

	For `test='hyper'` the first column can also describe chromosomal regions ('chr:start-stop', e.g. '9:0-39200000').
dataset	'adult' for the microarray dataset of adult human brains; '5_stages' for RNA-seq expression data for different stages of the developing human brain, grouped into 5 developmental stages; 'dev_effect' for a developmental effect score. For details see `browseVignettes("ABAData")`.
test	'hyper' (default) for the hypergeometric test, 'wilcoxon' for the Wilcoxon rank test, 'binomial' for the binomial test and 'contingency' for the 2x2-contingency table test (fisher's exact test or chi-square).
cutoff_quantiles	
	the FUNC enrichment analyses will be performed for the sets of expressed genes at given expression quantiles defined in this vector [0,1].
n_randsets	integer defining the number of random sets created to compute the FWER.
gene_len	logical. If `test='hyper'` the probability of a background gene to be chosen as a candidate gene in a random set is dependent on the gene length.
circ_chrom	logical. When genes defines chromosomal regions, `circ_chrom=TRUE` uses background regions from the same chromosome and allows randomly chosen blocks to overlap multiple background regions. Only if `test='hyper'`.
ref_genome	'grch37' (default) or 'grch38'. Defines the reference genome used when genomic regions are provided as input or when `gene_len=TRUE`.
gene_coords	optional data.frame() for custom gene coordinates, with four columns: gene-symbols (character), chromosome (character), start (integer), end (integer). When genomic regions are provided as input or when `gene_len=TRUE`, these custom gene coordinates are used instead of the integrated ones.
silent	logical. If TRUE all output to the screen except for warnings and errors is suppressed.

Details

For details please refer to `browseVignettes("ABAEnrichment")`.

Value

A list with components

results	a dataframe with the FWERs from the enrichment analyses per brain region and age category, ordered by 'n_significant','min_FWER' and 'mean_FWER', 'age_category' and 'structure_id'. 'min_FWER' for example denotes the minimum FWER for expression enrichment of the candidate genes in this brain region across all expression cutoffs. 'n_significant' reports the number of cutoffs at which the FWER was below 0.05. 'FWERs' is a semicolon separated string with the single FWERs for all cutoffs. 'equivalent_structures' is a semicolon separated string that lists structures with identical expression data due to lack of independent expression measurements in all regions.
genes	a dataframe of the input genes, excluding those genes for which no expression data is available and which therefore were not included in the enrichment analysis.

cutoffs   a dataframe with the expression values that correspond to the requested cutoff
                     quantiles.

Author(s)

Steffi Grote

References

[1] Hawrylycz, M.J. et al. (2012) An anatomically comprehensive atlas of the adult human brain
transcriptome, Nature 489: 391-399. doi: 10.1038/nature11405
[2] Miller, J.A. et al. (2014) Transcriptional landscape of the prenatal human brain, Nature 508:
199-206. doi: 10.1038/nature13185
[3] Allen Institute for Brain Science. Allen Human Brain Atlas. Available from: http://human.
brain-map.org/
[4] Allen Institute for Brain Science. BrainSpan Atlas of the Developing Human Brain. Available
from: http://brainspan.org/
[5] Pruefer, K. et al. (2007) FUNC: A package for detecting significant associations between gene
sets and ontological, BMC Bioinformatics 8: 41. doi: 10.1186/14712105841

See Also

```
browseVignettes("ABAEnrichment")
browseVignettes("ABAData")
```
get_expression
plot_expression
get_name
get_id
get_sampled_substructures
get_superstructures
get_annotated_genes

Examples

```
#### Note that arguments 'cutoff_quantiles' and 'n_randsets' are reduced
#### to lower computational time in the examples.
#### Using the default values is recommended.

#### Perform an enrichment analysis for the developing brain
#### with defined background genes
#### and with random sets dependent on gene length
gene_ids = c('PENK', 'COCH', 'PDYN', 'CA12', 'SYNDIG1L', 'MME',
    'ANO3', 'KCNJ6', 'ELAVL4', 'BEAN1', 'PVALB', 'EPN3', 'PAX2', 'FAB12')
is_candidate = rep(c(1,0),each=7)
genes = data.frame(gene_ids, is_candidate)
res = aba_enrich(genes, dataset='5_stages', cutoff_quantiles=c(0.5,0.9),
    n_randsets=100, gene_len=TRUE)
## see results for the brain regions with highest enrichment
## for children (age_category 3)
fwers = res[[1]]
```

```
head(fwers[fwers$age_category==3,])
## see the input genes dataframe (only genes with expression data available)
res[2]
## see the expression values that correspond to the requested cutoff quantiles
res[3]


# For more examples please refer to the package vignette.
```

---

get_annotated_genes *Get genes that are expressed in enriched or user-defined brain regions*

Description

Uses an object returned from [aba_enrich](#) as input and returns the brain regions that are significantly (given a FWER-threshold) enriched, together with the genes that are expressed in those brain regions (i.e. are 'annotated' to the brain regions). Alternatively, also user-defined brain regions, dataset and expression cutoffs can be used as input.

Usage

```
get_annotated_genes(res, fwer_threshold = 0.05, background = FALSE,
  structure_ids = NULL, dataset = NULL, cutoff_quantiles = NULL, genes = NULL)
```

Arguments

res   an object returned from [aba_enrich](#) (list of 3 elements). If not defined, `structure_ids`, `dataset` and `cutoff_quantiles` have to be specified.

fwer_threshold numeric defining the FWER-threshold. Given `res` as input, `get_annotated_genes` extracts all brain-region/expression-cutoff combinations from `res` that have a FWER < `fwer_threshold` and adds the (candidate) genes that are annotated to those brain regions at the given expression cutoffs.

background   logical indicating whether background genes should be included. Only used when `res` is defined and contains the results from a hypergeometric test (which is the default in [aba_enrich](#)).

structure_ids   vector of brain structure IDs, e.g. 'Allen:10208'. If `res` is not defined, `structure_ids` specifies the brain regions for which annotated genes at the given `cutoff_quantiles` will be returned.

dataset   'adult' for the microarray dataset of adult human brains; '5_stages' for RNA-seq expression data of the developing human brain, grouped into 5 developmental stages; 'dev_effect' for a developmental effect score. Only used when `res` is not defined.

cutoff_quantiles

                vector of numeric values between 0 and 1. They define the expression quantiles (across all genes) which are used as cutoffs to decide whether a gene counts as expressed (and gets annotated to a brain region) or not. Only used when `res` is not defined.

genes   optional vector of gene identifiers, either Entrez-ID, Ensembl-ID or gene-symbol.
                         If defined, only annotations of those genes are returned. If not defined, all ex-
                         pressed genes from Allen Brain Atlas are returned. Only used when res is not
                         defined.

Details

Genes get annotated to a brain region when their expression value in that brain region, which is
provided by the Allen Brain Atlas, exceeds a certain cutoff. Multiple cutoffs can be used. They are
defined as quantiles of gene expression across all genes and brain regions. An expression cutoff of
e.g. 0.8 means that only genes with expression levels higher than 80% of all measured values count
as 'expressed' and get annotated to the corresponding brain region. Note that those annotations are
inherited by all superstructures (parent nodes).
When the 'dev_effect' dataset is used as dataset-argument, the expression value is replaced by
a measure of how much a gene's expression changes during the development (prenatal to adult).
Genes with a developmental score higher than the cutoff then get annotated to the brain regions.

Value

A data frame with the age category, the brain region ID, the expression cutoff quantile and the anno-
tated genes for significantly enriched or user-defined brain-region/expression-cutoff combinations.
If res was given as input, two additional columns are added: the FWER and the score which was
used for the genes in the [aba_enrich](#) input (1/0 for candidate and background genes for the hyper-
geometric test or scores for the wilcoxon rank sum test). Given res as input, the output is ordered
by the FWER, else it is ordered by the expression cutoff.

Author(s)

Steffi Grote

References

[1] Hawrylycz, M.J. et al. (2012) An anatomically comprehensive atlas of the adult human brain
transcriptome, Nature 489: 391-399. doi: [10.1038/nature11405](#)
[2] Miller, J.A. et al. (2014) Transcriptional landscape of the prenatal human brain, Nature 508:
199-206. doi: [10.1038/nature13185](#)
[3] Allen Institute for Brain Science. Allen Human Brain Atlas. Available from: http://human.brain-map.org/
[4] Allen Institute for Brain Science. BrainSpan Atlas of the Developing Human Brain. Available
from: http://brainspan.org/

See Also

```
browseVignettes("ABAEnrichment")
browseVignettes("ABAData")
```
[aba_enrich](#)
[get_name](#)
[get_id](#)

Examples

```
#### Note that arguments 'cutoff_quantiles' and 'n_randsets' are reduced
#### to lower computational time in the examples.

## perform an enrichment analysis
## for expression of 7 candidate and 7 background genes
## and get candidate genes annotated to brain regions that have a FWER < 0.05
set.seed(123)
gene_ids = c('PENK', 'COCH', 'PDYN', 'CA12', 'SYNDIG1L', 'MME',
    'ANO3', 'KCNJ6', 'ELAVL4', 'BEAN1', 'PVALB', 'EPN3', 'PAX2', 'FAB12')
is_candidate = rep(c(1,0), each=7)
genes = data.frame(gene_ids, is_candidate)
res = aba_enrich(genes, dataset='5_stages', cutoff_quantiles=c(0.3,0.5,0.7,0.9),
    n_randset=100)
anno = get_annotated_genes(res, fwer_threshold=0.05)
head(anno)

## find out which of the above genes have expression above
## the 50% and 70% expression-cutoff, respectively,
## in the Cerebellar Cortex of the developing human brain (Allen:10657)
get_annotated_genes(structure_ids="Allen:10657", dataset="5_stages",
    cutoff_quantiles=c(0.5,0.7), genes=gene_ids)
```

---

get_expression	*Get expression data for given genes and brain structure IDs*

Description

Expression data obtained from the Allen Brain Atlas project [1-4].

Usage

```
get_expression(structure_ids, gene_ids, dataset = "adult")
```

Arguments

structure_ids   vector of brain structure IDs, e.g. 'Allen:10208'.

gene_ids   vector of gene identifiers, either Entrez-ID, Ensembl-ID or gene-symbol.

dataset   'adult' (default) for the microarray dataset of adult human brains; '5_stages' for RNA-seq expression data of the developing human brain, grouped into 5 developmental stages; 'dev_effect' for a developmental effect score.

Details

Get gene expression in defined brain regions from adult or developing humans, or a developmental effect score for the developing human brain. Expression data is obtained from the Allen Brain Atlas project [1-4], averaged across donors, and for the developing human brain divided into five major age categories. The developmental effect score is based on expression data of the developing human

brain. If a requested brain region has no expression data annotated, data from sampled substructures of this region is returned.

Please refer to the `ABAData` package vignette for details on the datasets.

Value

A matrix with expression values or developmental effect scores per brain region (rows) and gene (columns).

For expression data from the developing human brain ('5_stages') it is a list with an expression matrix for each of the 5 developmental stages.

Author(s)

Steffi Grote

References

[1] Hawrylycz, M.J. et al. (2012) An anatomically comprehensive atlas of the adult human brain transcriptome, Nature 489: 391-399. doi: 10.1038/nature11405

[2] Miller, J.A. et al. (2014) Transcriptional landscape of the prenatal human brain, Nature 508: 199-206. doi: 10.1038/nature13185

[3] Allen Institute for Brain Science. Allen Human Brain Atlas. Available from: http://human.brain-map.org/

[4] Allen Institute for Brain Science. BrainSpan Atlas of the Developing Human Brain. Available from: http://brainspan.org/

See Also

```
browseVignettes("ABAEnrichment")
browseVignettes("ABAData")
```
plot_expression
aba_enrich
get_name
get_id
get_sampled_substructures
get_annotated_genes

Examples

```
## get expression data of six genes in two brain regions
## from the developing human brain,
## each of the five list elements corresponds to an age category
get_expression(structure_ids=c('Allen:10657','Allen:10208'),
    gene_ids=c('ENSG00000168036','ENSG00000157764','ENSG00000182158',
    'ENSG00000147889'), dataset='5_stages')

## see the package vignette for more examples
```

---

get_id   *Get the structure ID of a brain region given its name*

Description

Returns brain regions given (part of) their name, together with their structure IDs from the ontologies for the adult and for the developing brain (e.g. 'Allen:10657' as used throughout the ABAEnrichment package).

Usage

```
get_id(structure_name)
```

Arguments

structure_name (partial) name of a brain structure, e.g. 'telencephalon'

Value

a data frame with the full names of the brain structures that contain structure_name; together with the ontology ('developmental' or 'adult') and the structure IDs.

Author(s)

Steffi Grote

References

[1] Allen Institute for Brain Science. Allen Human Brain Atlas. Available from: http://human.brain-map.org/
[2] Allen Institute for Brain Science. BrainSpan Atlas of the Developing Human Brain. Available from: http://brainspan.org/

See Also

get_name
get_sampled_substructures
get_superstructures
browseVignettes("ABAData")

Examples

```
## get structure IDs of brain regions that contain 'accumbens' in their names
get_id('accumbens')
## get structure IDs of brain regions that contain 'telencephalon' in their name
get_id('telencephalon')
## get all brain regions that have direct or indirect expression data
```

```
all_regions = get_id('')
head(all_regions)
```

---

get_name   *Get the full name of a brain region given structure IDs*

Description

Returns the full name of brain regions given the structure IDs, e.g. 'Allen:10657' as used throughout the ABAEnrichment package. The full name is composed of an acronym and the name as used by the Allen Brain Atlas project [1-2].

Usage

```
get_name(structure_ids)
```

Arguments

structure_ids   a vector of brain structure IDs, e.g. c('Allen:10657','Allen:10173') or c(10657,10173)

Value

vector of the full names of the brain structures; composed of acronym, underscore and name.

Note

The acronym is added because the names alone are not unique.

Author(s)

Steffi Grote

References

[1] Allen Institute for Brain Science. Allen Human Brain Atlas. Available from: http://human.brain-map.org/
[2] Allen Institute for Brain Science. BrainSpan Atlas of the Developing Human Brain. Available from: http://brainspan.org/

See Also

get_id
get_sampled_substructures
get_superstructures

Examples

```
## get the full names of the brain structures 'Allen:10657' and 'Allen:10225'
get_name(c('Allen:10657','Allen:10225'))
```

---

get_sampled_substructures

Return sampled substructures of a given brain region

Description

The function returns for a given brain structure ID all its substructures with available expression data, potentially including the structure itself.

Usage

```
get_sampled_substructures(structure_id)
```

Arguments

structure_id   a brain structure ID, e.g. 'Allen:10657' or '10657'

Details

The ontology enrichment analysis in aba_enrich tests all brain regions for which data is available, although the region might not have been sampled directly. In this case the region inherits the expression data from its substructures with available expression data. The function get_sampled_substructures helps to explore where the expression data for a brain region came from.

Value

vector of brain structure IDs that contains all substructures of the requested brain region that were sampled.

Author(s)

Steffi Grote

References

[1] Allen Institute for Brain Science. Allen Human Brain Atlas. Available from: http://human.brain-map.org/
[2] Allen Institute for Brain Science. BrainSpan Atlas of the Developing Human Brain. Available from: http://brainspan.org/

See Also

```
browseVignettes("ABAEnrichment")
browseVignettes("ABAData")
aba_enrich
get_name
get_superstructures
```

Examples

```
## get the brain structures from which the brain structures
## 'Allen:4010' and 'Allen:10208' inherit their expression data
get_sampled_substructures('Allen:4010')
get_sampled_substructures('Allen:10208')
```

---

get_superstructures   *Returns all superstructures of a brain region using the Allen Brain Atlas ontology*

Description

Returns all superstructures of a brain region and the brain region itself given a structure ID, e.g. 'Allen:10657' as used throughout the ABAEnrichment package. The output vector contains the superstructures according to the hierarchy provided by the Allen Brain Atlas ontology [1,2] beginning with the root ('brain' or 'neural plate') and ending with the requested brain region.

Usage

```
get_superstructures(structure_id)
```

Arguments

structure_id   a brain structure ID, e.g. 'Allen:10657' or '10657'

Value

vector of brain structure IDs that contains all superstructures of the requested brain region and the brain region itself. The order of the brain regions follows the hierarchical organization of the brain.

Note

The ontologies for the adult and the developing human brain are different.

Author(s)

Steffi Grote

References

[1] Allen Institute for Brain Science. Allen Human Brain Atlas. Available from: http://human.brain-map.org/
[2] Allen Institute for Brain Science. BrainSpan Atlas of the Developing Human Brain. Available from: http://brainspan.org/

See Also

get_name
get_id
get_sampled_substructures

Examples

```
## Get the IDs of the superstructures of the precentral gyrus
## (adult brain ontology)
get_superstructures('Allen:4010')

## Get the IDs and the names of the superstructures
## of the dorsolateral prefrontal cortex
## (developing brain ontology)
data.frame(superstructure=get_name(get_superstructures("Allen:10173")))
```

---

plot_expression *Plot a heatmap given expression data*

Description

The function produces a (heatmap.2 from package gplots). To be used with gene-expression matrices obtained from get_expression.

Usage

```
plot_expression(expr_mat, dendro = TRUE, gene_vars = NULL, main = "")
```

Arguments

expr_mat	matrix of expression values with genes in columns and brain regions in rows, like returned from get_expression (rownames are expected to be brain region identifiers, and their acronym is added automatically)
dendro	logical indicating whether rows and columns should be rearranged with a dendrogram based on row/column means (using hclust).
gene_vars	optional data.frame of gene-associated variables to be plotted as a colored sidebar. This can be the same input as for an enrichment analysis, i.e. a data.frame with genes in the first columns and with additional columns depending on the test used in aba_enrich. For a hypergeometric test the genes are grouped into candidate (red) and background genes (black); for the Wilcoxon rank-sum test, the side bar indicates the user-defined scores; for the binomial test the side bar shows *A/(A+B)* and for the 2x2 contingency table test *((A+1)/(B+1)) / ((C+1)/(D+1))* (+1 added to prevent division by 0).
main	optional character to use as a title for the heatmap.

Value

Invisibly, a list with components

`rowInd`	row index permutation vector as returned by `order.dendrogram`
`colInd`	column index permutation vector.
`call`	the matched call
`carpet`	reordered 'x' values used to generate the main 'carpet'
`rowDendrogram`	row dendrogram, if present
`colDendrogram`	column dendrogram, if present
`breaks`	values used for color break points
`col`	colors used
`colorTable`	A three-column data frame providing the lower and upper bound and color for each bin

Author(s)

Steffi Grote

References

[1] Hawrylycz, M.J. et al. (2012) An anatomically comprehensive atlas of the adult human brain transcriptome, Nature 489: 391-399. doi: 10.1038/nature11405
[2] Miller, J.A. et al. (2014) Transcriptional landscape of the prenatal human brain, Nature 508: 199-206. doi: 10.1038/nature13185
[3] Allen Institute for Brain Science. Allen Human Brain Atlas. Available from: http://human.brain-map.org/
[4] Allen Institute for Brain Science. BrainSpan Atlas of the Developing Human Brain. Available from: http://brainspan.org/

See Also

```
browseVignettes("ABAEnrichment")
browseVignettes("ABAData")
```
get_expression
aba_enrich
get_name
get_id
get_sampled_substructures
get_annotated_genes
heatmap.2
hclust

Examples

```
## plot expression data of six genes in two brain regions
## from children (age_category 3)
expr = get_expression(structure_ids=c("Allen:10657","Allen:10208"),
    gene_ids=c('RFFL', 'NTS', 'LIPE', 'GALNT6', 'GSN', 'BTBD16', 'CERS2'),
    dataset="5_stages")
expr_age_3 = expr[[3]]
plot_expression(expr_age_3)

## see the package vignette for more examples
```

Index