Package ‘MGFR’

April 14, 2017

Type Package
Title Marker Gene Finder in RNA-seq data
Version 1.0.0
Author Khadija El Amrani
Maintainer Khadija El Amrani <khadija.el-amrani@charite.de>
Description The package is designed to detect marker genes from RNA-seq data.
Depends R (>= 3.3)
Imports biomaRt, annotate
biocViews Genetics, GeneExpression, RNASeq
License GPL-3
LazyData yes
NeedsCompilation no

R topics documented:

MGFR-package .............................................................. 1
getMarkerGenes.rnaseq .................................................... 2
getMarkerGenes.rnaseq.html ............................................. 3
ref.mat ................................................................. 4

Index 5
getMarkerGenes.rnaseq

Author(s)
Khadija El Amrani Maintainer: Khadija El Amrani <khadija.el-amrani@charite.de>

Examples

data(ref.mat)
res.list <- getMarkerGenes.rnaseq(ref.mat, samples2compare="all", annotate=TRUE, gene.ids.type="ensembl", score.cutoff=1)
names(res.list)
## show the first 20 markers of liver
res.list[["liver_markers"]][1:20]

getMarkerGenes.rnaseq  Marker Gene Detection

Description
Function to detect marker genes using normalized RNA-seq data

Usage
getMarkerGenes.rnaseq(data.mat, samples2compare="all", annotate=TRUE, gene.ids.type="ensembl", score.cutoff=1)

Arguments
data.mat  RNA-seq gene expression matrix with genes corresponding to rows and samples corresponding to columns.
samples2compare  A character vector with the sample names to be compared (e.g. c("liver", "lung", "brain")). By default all samples in the reference matrix are used.
annotate  A boolean value. If TRUE the gene symbol and the entrez gene id are shown.
gene.ids.type  Type of the used gene identifiers, the following gene identifiers are supported: ensembl, refseq and ucsc gene ids. default is ensembl.
score.cutoff  A value in the interval [0,1] to filter the markers according to the specificity score. The default value is 1 (no filtering).

Details
For each marker in the output list, the gene id and the corresponding score are shown. If annotate is TRUE, the gene symbol and the entrez gene id are shown. The score is used to rank the markers according to their specificity. A lower value means a higher specificity.

Value
A list with marker genes associated with each sample type.

Author(s)
Khadija El Amrani <khadija.el-amrani@charite.de>
Function to detect marker genes using normalized RNA-seq data and show the marker genes in HTML tables with links to various online annotation sources (Ensembl, GenBank and EntrezGene repositories).

**Usage**

```r
getMarkerGenes.rnaseq.html(data.mat, samples2compare="all", gene.ids.type="ensembl", score.cutoff=1, directory = getwd())
```

**Arguments**

- `data.mat`  
  RNA-seq gene expression matrix with genes corresponding to rows and samples corresponding to columns.

- `samples2compare`  
  A character vector with the sample names to be compared (e.g. `c("liver", "lung", "brain")`). By default all samples in the reference matrix are used.

- `gene.ids.type`  
  Type of the used gene identifiers, the following gene identifiers are supported: ensembl, refseq and ucsc gene ids. default is ensembl.

- `score.cutoff`  
  A value in the interval [0,1] to filter the markers according to the specificity score. The default value is 1 (no filtering).

- `directory`  
  Path to the directory where to save the html pages, default is the current working directory.

**Details**

This function is based on the function `htmlpage` from the R-package `annotate`.

**Value**

This function is used only for the side effect of creating HTML tables.

**Author(s)**

Khadija El Amrani <khadija.el-amrani@charite.de>

**Examples**

```r
data(ref.mat)
res.list <- getMarkerGenes.rnaseq.html(ref.mat, samples2compare="all", annotate=TRUE, gene.ids.type="ensembl", score.cutoff=1, directory = getwd())
names(res.list)
## show the first 20 markers of liver
res.list[["liver_markers"]][1:20]
```
RNA-seq gene expression data set

Description
RNA-seq gene expression data set derived from 5 tissue types (lung, liver, heart, kidney, and brain) from the ArrayExpress database (E-MTAB-1733). Each tissue type is represented by 3 replicates.

Usage
data(ref.mat)

Format
A matrix with 32431 genes and 15 samples.

Value
RNA-seq data matrix

Examples
data(ref.mat)
Index

*Topic RNA-seq data
  getMarkerGenes.rnaseq, 2
  getMarkerGenes.rnaseq.html, 3
*Topic dataset
  ref.mat, 4
*Topic marker genes
  getMarkerGenes.rnaseq, 2
  getMarkerGenes.rnaseq.html, 3
*Topic package
  MGFR-package, 1
  getMarkerGenes.rnaseq, 2
  getMarkerGenes.rnaseq.html, 3
htmlpage, 3
matrix, 4
MGFR (MGFR-package), 1
MGFR-package, 1
ref.mat, 4