

# Package ‘sampleClassifier’

October 18, 2017

**Type** Package  
**Title** Sample Classifier  
**Version** 1.0.0  
**Date** 2017-01-23  
**Author** Khadija El-Amrani  
**Maintainer** Khadija El-Amrani <khadija.el-amrani@charite.de>  
**Description** The package is designed to classify gene expression profiles.  
**Depends** R (>= 3.4), MGFM, MGFR, annotate  
**Imports** e1071, ggplot2, stats, utils  
**Suggests** sampleClassifierData, BiocStyle, hgu133a.db, hgu133plus2.db  
**biocViews** Classification, Microarray, RNASeq, GeneExpression  
**License** Artistic-2.0  
**LazyData** yes  
**NeedsCompilation** no

## R topics documented:

sampleClassifier-package . . . . .	1
classifyProfile . . . . .	2
classifyProfile.rnaseq . . . . .	3
classifyProfile.rnaseq.svm . . . . .	4
classifyProfile.svm . . . . .	5
get.heatmap . . . . .	7
<b>Index</b>	<b>8</b>

---

sampleClassifier-package  
*Sample Classifier*

---

## Description

The package is designed to classify samples from microarray and RNA-seq gene expression datasets.

## Details

Package: sampleClassifier  
 Type: Package  
 Version: 1.0.0  
 License: GPL-3

### Author(s)

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

### Examples

```
## Not run:
library(sampleClassifierData)
data("se_micro_refmat")
micro_refmat <- assay(se_micro_refmat)
data("se_micro_testmat")
micro_testmat <- assay(se_micro_testmat)
res1.list <- classifyProfile(ref_matrix=micro_refmat, query_mat=micro_testmat,
chip1="hgu133plus2", chip2="hgu133a", write2File=FALSE)
res1.list

## End(Not run)
```

---

classifyProfile	<i>Expression profile classification</i>
-----------------	--

---

### Description

Function to classify microarray gene expression profiles

### Usage

```
classifyProfile(ref_matrix, query_mat, chip1 = "hgu133plus2", chip2 = "hgu133a", fun1 = median, fun2 = mean)
```

### Arguments

ref_matrix	Normalized microarray data matrix to be used as reference, with probe sets corresponding to rows and samples corresponding to columns.
query_mat	Normalized microarray query matrix to be classified, with probe sets corresponding to rows and samples corresponding to columns.
chip1	Chip name of the reference matrix.
chip2	Chip name of the query matrix. This parameter can be ignored if the reference and query matrix are from the same chip.
fun1	<a href="#">mean</a> or <a href="#">median</a> . This will specify the number of marker genes that will be used for classification. Default is <a href="#">median</a> .
fun2	<a href="#">mean</a> or <a href="#">median</a> . This will be used to summarize the expression values of probe sets that belong to the same gene. This parameter can be ignored if the reference and query matrix are from the same chip. Default is <a href="#">mean</a> .

`write2File` If TRUE, the classification results for each query profile will be written to a file.  
`out.dir` Path to a directory to write the classification results, default is the current working directory.

### Details

Each query profile is compared to all sample types in the reference matrix and a similarity score is calculated. The similarity score is based on the number of marker genes that are shared between the query and the reference. These marker genes are given in a file if `write2File` is TRUE.

### Value

A list with top hits for each query profile, sorted according to a similarity score.

### Author(s)

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

### See Also

see also [getMarkerGenes](#).

### Examples

```
library(sampleClassifierData)
data("se_micro_refmat")
micro_refmat <- assay(se_micro_refmat)
data("se_micro_testmat")
micro_testmat <- assay(se_micro_testmat)
res1.list <- classifyProfile(ref_matrix=micro_refmat, query_mat=micro_testmat,
chip1="hgu133plus2", chip2="hgu133a", write2File=FALSE)
res1.list
```

---

classifyProfile.rnaseq

*Expression profile classification*

---

### Description

Function to classify RNA-seq gene expression profiles

### Usage

```
classifyProfile.rnaseq(ref_matrix, query_mat, gene.ids.type="ensembl", fun1 = median, write2File=
```

### Arguments

ref_matrix	RNA-seq data matrix to be used as reference, with genes corresponding to rows and samples corresponding to columns.
query_mat	RNA-seq query matrix to be classified, with genes corresponding to rows and samples corresponding to columns.
gene.ids.type	Type of the used gene identifiers, the following gene identifiers are supported: ensembl, refseq and ucsc gene ids. Default is ensembl.
fun1	<a href="#">mean</a> or <a href="#">median</a> . This will specify the number of marker genes that will be used for classification. Default is <a href="#">median</a> .
write2File	A logical value. If TRUE the classification results will be written to a file.
out.dir	Path to the directory, in which to write the results. Default is the actual working directory.

### Details

Each query profile is compared to all sample types in the reference matrix and a similarity score is calculated. The similarity score is based on the number of marker genes that are shared between the query and the reference. These marker genes are given in a file if write2File is TRUE.

### Value

A list with top hits for each query profile, sorted according to a similarity score.

### Author(s)

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

### Examples

```
library(sampleClassifierData)
data("se_rnaseq_refmat")
rnaseq_refmat <- assay(se_rnaseq_refmat)
data("se_rnaseq_testmat")
rnaseq_testmat <- assay(se_rnaseq_testmat)
res2.list <- classifyProfile.rnaseq(ref_matrix=rnaseq_refmat, query_mat=rnaseq_testmat,
gene.ids.type="ensembl",write2File=FALSE)
res2.list
```

---

classifyProfile.rnaseq.svm

*Expression profile classification*

---

### Description

Function to classify RNA-seq gene expression profiles using support vector machines (SVM)

**Usage**

```
classifyProfile.rnaseq.svm(ref_matrix, query_mat, gene.ids.type="ensembl", fun1 = median)
```

**Arguments**

ref_matrix	RNA-seq data matrix to be used as reference, with genes corresponding to rows and samples corresponding to columns.
query_mat	RNA-seq query matrix to be classified, with genes corresponding to rows and samples corresponding to columns.
gene.ids.type	Type of the used gene identifiers, the following gene identifiers are supported: ensembl, refseq and ucsc gene ids. Default is ensembl.
fun1	<a href="#">mean</a> or <a href="#">median</a> . This will specify the number of marker genes that will be used for classification. Default is <a href="#">median</a> .

**Details**

This function is based on the function [svm](#) from the R-package 'e1071'.

**Value**

A data frame with the predicted classes for each query profile.

**Author(s)**

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

**Examples**

```
library(sampleClassifierData)
data("se_rnaseq_refmat")
rnaseq_refmat <- assay(se_rnaseq_refmat)
data("se_rnaseq_testmat")
rnaseq_testmat <- assay(se_rnaseq_testmat)
res2.svm.df <- classifyProfile.rnaseq.svm(ref_matrix=rnaseq_refmat, query_mat=rnaseq_testmat,
gene.ids.type="ensembl")
res2.svm.df
```

---

classifyProfile.svm    *Expression profile classification*

---

**Description**

Function to classify microarray gene expression profiles using support vector machines (SVM)

**Usage**

```
classifyProfile.svm(ref_matrix, query_mat, chip1 = "hgu133plus2", chip2 = "hgu133a", fun1 = median)
```

## Arguments

ref_matrix	Normalized microarray data matrix to be used as reference, with probe sets corresponding to rows and samples corresponding to columns.
query_mat	Normalized microarray query matrix to be classified, with probe sets corresponding to rows and samples corresponding to columns.
chip1	Chip name of the reference matrix.
chip2	Chip name of the query matrix. This parameter can be ignored if the reference and query matrix are from the same chip.
fun1	<a href="#">mean</a> or <a href="#">median</a> . This will specify the number of marker genes that will be used for classification. Default is <a href="#">median</a> .
fun2	<a href="#">mean</a> or <a href="#">median</a> . This will be used to summarize the expression values of probe sets that belong to the same gene. This parameter can be ignored if the reference and query matrix are from the same chip. Default is <a href="#">mean</a> .

## Details

This function is based on the function [svm](#) from the R-package 'e1071'.

## Value

A data frame with the predicted classes for each query profile.

## Author(s)

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

## See Also

see also [getMarkerGenes](#).

## Examples

```
library(sampleClassifierData)
data("se_micro_refmat")
micro_refmat <- assay(se_micro_refmat)
data("se_micro_testmat")
micro_testmat <- assay(se_micro_testmat)
res1.svm.df <- classifyProfile.svm(ref_matrix=micro_refmat, query_mat=micro_testmat,
chip1="hgu133plus2", chip2="hgu133a")
res1.svm.df
```

---

get.heatmap	<i>display classification results as heatmap</i>
-------------	--

---

**Description**

Function to display the classification predictions as a heatmap

**Usage**

```
get.heatmap(res.list)
```

**Arguments**

res.list            the result list returned by the function [classifyProfile](#) or [classifyProfile.rnaseq](#)

**Details**

This function is based on the function [ggplot](#) from the R-package 'ggplot2'.

**Value**

This function is used only for the side effect of creating a heatmap.

**Author(s)**

Khadija El Amrani <[khadija.el-amrani@charite.de](mailto:khadija.el-amrani@charite.de)>

**Examples**

```
library(sampleClassifierData)
data("se_micro_refmat")
micro_refmat <- assay(se_micro_refmat)
data("se_micro_testmat")
micro_testmat <- assay(se_micro_testmat)
res1.list <- classifyProfile(ref_matrix=micro_refmat, query_mat=micro_testmat,
chip1="hgu133plus2",chip2="hgu133a", write2File=FALSE)
get.heatmap(res1.list)
```

# Index

- \*Topic **RNA-seq data**
  - classifyProfile.rnaseq, 3
  - classifyProfile.rnaseq.svm, 4
- \*Topic **SVM**
  - classifyProfile.rnaseq.svm, 4
  - classifyProfile.svm, 5
- \*Topic **classification**
  - classifyProfile, 2
  - classifyProfile.rnaseq, 3
  - classifyProfile.rnaseq.svm, 4
  - classifyProfile.svm, 5
  - get.heatmap, 7
- \*Topic **heatmap**
  - get.heatmap, 7
- \*Topic **microarray data**
  - classifyProfile, 2
  - classifyProfile.svm, 5
- \*Topic **package**
  - sampleClassifier-package, 1

classifyProfile, 2, 7  
classifyProfile.rnaseq, 3, 7  
classifyProfile.rnaseq.svm, 4  
classifyProfile.svm, 5

get.heatmap, 7  
getMarkerGenes, 3, 6  
ggplot, 7

mean, 2, 4–6  
median, 2, 4–6

sampleClassifier  
(sampleClassifier-package), 1  
sampleClassifier-package, 1  
svm, 5, 6