

# Package ‘sampleClassifier’

October 16, 2023

**Type** Package

**Title** Sample Classifier

**Version** 1.24.0

**Description** The package is designed to classify microarray RNA-seq gene expression profiles.

**Depends** R (>= 4.0), MGFM, MGFR, annotate

**Imports** e1071, ggplot2, stats, utils

**Suggests** sampleClassifierData, BiocStyle, hgu133a.db, hgu133plus2.db

**biocViews** ImmunoOncology, Classification, Microarray, RNASeq, GeneExpression

**License** Artistic-2.0

**LazyData** yes

**NeedsCompilation** no

**git\_url** <https://git.bioconductor.org/packages/sampleClassifier>

**git\_branch** RELEASE\_3\_17

**git\_last\_commit** 4fdd4e8

**git\_last\_commit\_date** 2023-04-25

**Date/Publication** 2023-10-15

**Author** Khadija El Amrani [aut, cre]

**Maintainer** Khadija El Amrani <a.khadija@gmx.de>

## R topics documented:

|                                      |   |
|--------------------------------------|---|
| sampleClassifier-package . . . . .   | 2 |
| classifyProfile . . . . .            | 2 |
| classifyProfile.rnaseq . . . . .     | 4 |
| classifyProfile.rnaseq.svm . . . . . | 5 |
| classifyProfile.svm . . . . .        | 6 |
| get.heatmap . . . . .                | 7 |
| grid-internal . . . . .              | 8 |

|              |          |
|--------------|----------|
| <b>Index</b> | <b>9</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      *Expression profile classification*

---

### Description

Function to classify microarray gene expression profiles

### Usage

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

## 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=FALSE)
```

**Arguments**

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

## 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

*display classification results as heatmap*

---

**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>

**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)
```

---

grid-internal

*Internal sampleClassifier Functions*

---

**Description**

Internal sampleClassifier functions

**Details**

These are not intended to be called by the user.



# Index

- \* **RNA-seq data**
  - classifyProfile.rnaseq, 4
  - classifyProfile.rnaseq.svm, 5
- \* **SVM**
  - classifyProfile.rnaseq.svm, 5
  - classifyProfile.svm, 6
- \* **classification**
  - classifyProfile, 2
  - classifyProfile.rnaseq, 4
  - classifyProfile.rnaseq.svm, 5
  - classifyProfile.svm, 6
  - get.heatmap, 7
- \* **heatmap**
  - get.heatmap, 7
- \* **internal**
  - grid-internal, 8
- \* **microarray data**
  - classifyProfile, 2
  - classifyProfile.svm, 6
- \* **package**
  - sampleClassifier-package, 2
  - .classify\_sample (grid-internal), 8
  - .collapseRows (grid-internal), 8
  - .get\_genes (grid-internal), 8
  - .get\_mpsmat (grid-internal), 8
  - .getmarker\_ps (grid-internal), 8

classifyProfile, 2, 7  
classifyProfile.rnaseq, 4, 7  
classifyProfile.rnaseq.svm, 5  
classifyProfile.svm, 6

get.heatmap, 7  
getMarkerGenes, 3, 7  
ggplot, 7  
grid-internal, 8

mean, 3–6  
median, 3–6

sampleClassifier  
    (sampleClassifier-package), 2  
sampleClassifier-package, 2  
svm, 5, 6