Package 'iC10'

September 23, 2015

Type Package

Title A Copy Number and Expression-Based Classifier for Breast Tumours

Version 1.1.3

Date 2015-09-22

Author Oscar M Rueda

Maintainer Oscar M. Rueda <Oscar.Rueda@cruk.cam.ac.uk>

Description Implementation of the classifier described in the paper 'Genome-driven integrated classification of breast cancer validated in over 7,500 samples' (Ali HR et al., Genome Biology 2014). It uses copy number and/or expression form breast cancer data, trains a pamr classifier (Tibshirani et al.) with the features available and predicts the iC10 group.

License GPL-3

Depends pamr, iC10TrainingData

NeedsCompilation no

Repository CRAN

Date/Publication 2015-09-23 08:05:08

R topics documented:

iC10-package *A copy number and expression-based classifier for breast tumours.*

Description

iC10 implements the classifier described in the paper 'Genome-driven integrated classification of breast cancer validated in over 7,500 samples' (Ali HR et al., Genome Biology 2014). It uses copy number and/or expression form breast cancer data, trains a pamr classifier (Tibshirani et al.) with the features available and predicts the iC10 group.

Details

Package:	iC10
Type:	Package
Version:	1.0
Date:	2014-03-22
License:	What license is it under?

data(train.CN) data(train.Exp) features <- matchFeatures(Exp=train.Exp, Exp.by.feat="probe") features <- normalizeFeatures(features, "scale") res <- iC10(features) summary(res) goodnessOfFit(res, newdata=features) compare(res, ic10=1:5, newdata=features) compare(res, ic10=6:10, newdata=features)

Author(s)

Oscar M Rueda

Maintainer: <Oscar.Rueda@cruk.cam.ac.uk>

References

Ali HR et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biology 2014; 15:431. Curtis et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346-352. Tibshirani et al. Diagnosis of multiple cancer types by shrunken centroids of gene expression. PNAS 2002; 99(10):6567-6572.

See Also

pamr, CONOR

Examples

```
require(iC10TrainingData)
data(train.CN)
data(train.Exp)
features <- matchFeatures(Exp=train.Exp, Exp.by.feat="probe")
features <- normalizeFeatures(features, "scale")
```

```
res <- iC10(features)
summary(res)
goodnessOfFit(res, newdata=features)
compare(res, iC10=1:2, newdata=features)
compare(res, iC10=2:4, newdata=features)
```

`compare` *Compare results of the iC10 classifier*

Description

This function plots the centroids of the training set versus the average profiles of the new data classified in each group.

Usage

```
compare(obj, iC10=1:10, newdata, name.test="Test",...)
## S3 method for class 'iC10'
compare(obj, iC10=1:10, newdata, name.test="Test",...)
```

Arguments

`obj`	An object of class `iC10`, a result of a call to `iC10()`
`iC10`	Groups to plot
`newdata`	Set of features of the new data to compare. They must be the same samples classified and contained in `x`. A result of a call to `matchFeatures()` or `normalizeFeatures()`
`name.test`	Name of the new data set to appear in the text of the plot
`...`	Additional arguments passed to `plot()`

Value

A plot is returned with two plots per groups requested.

Author(s)

Oscar M. Rueda

References

Ali HR et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biology 2014; 15:431. Curtis et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346-352.

See Also

`iC10`, `plot.iC10`, `matchFeatures`, `normalizeFeatures`

Examples

```
require(iC10TrainingData)
data(train.CN)
data(train.Exp)
features <- matchFeatures(Exp=train.Exp, Exp.by.feat="probe")
features <- normalizeFeatures(features, "scale")
res <- iC10(features)
compare(res, 1:3, newdata=features)
```

---

getCNfeatures *Internal function for matching copy number features.*

Description

This function should not be called directly

Usage

```
getCNfeatures(CN, Probes, Map, by.feat, ref, Synonyms)
```

Arguments

CN	CN features matrix
Probes	Vector with the probes to match
Map	data.frame with the genomic description of the features to match
by.feat	"probe" or "gene", indicating if match should be done by probe position or gene name.
ref	hg18 or hg19 (only relevant if matching is done by probe position).
Synonyms	data.frame with available synonym gene names to match (only relevant if matching is done by gene name).

Value

A matrix with the copy number features

Author(s)

Oscar M Rueda

getExpfeatures *Internal function for matching expression features.*

Description

Internal function for matching expression features.

Usage

```
getExpfeatures(Exp, Probes, Synonyms, by.feat)
```

Arguments

Exp	Matrix of expression features
Probes	Vector of probes to match
Synonyms	vector of synonyms fo gene names
by.feat	either "probe" or "gene"

Value

A matrix with the Probes in Exp.

Note

This function is not supposed to be called directly. use matchFeatures instead.

Author(s)

Oscar M Rueda

References

Curtis et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346-352.

See Also

matchFeatures

goodnessOfFit *Goodness of fit results of the iC10 classifier*

Description

Goodness of fit results of the iC10 classifier: this function computes correlations between the signatures of the training dataset and the classified features.

Usage

```
goodnessOfFit(obj, iC10=1:10, newdata=NULL,...)
## S3 method for class 'iC10'
goodnessOfFit(obj, iC10=1:10, newdata=NULL,...)
```

Arguments

obj	An object of iC10 class.
iC10	Groups to compute goodness of fit.
newdata	The feature data to compute the goodness of fit. Must be the samples classified in obj. It can be a call to matchFeatures or normalizeFeatures. If NULL, obj$fitted is used.
...	Additional arguments passed to cor (like method; Default is pearson)

Value

It prints the correlation for each iC10.

Author(s)

Oscar M Rueda

References

Ali HR et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biology 2014; 15:431. Curtis et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346-352.

See Also

iC10

Examples

```
require(iC10TrainingData)
data(train.CN)
data(train.Exp)
features <- matchFeatures(Exp=train.Exp, Exp.by.feat="probe")
features <- normalizeFeatures(features, "scale")
res <- iC10(features)
goodnessOfFit(res, newdata=features)
```

`iC10` *A copy number and expression-based classfier for breast cancers*

Description

iC10 implements the classifier described in the paper 'Genome-driven integrated classification of breast cancer validated in over 7,500 samples' (Ali HR et al., Genome Biology 2014). It uses copy number and/or expression form breast cancer data, trains a pamr classifier (Tibshirani et al.) with the features available and predicts the iC10 group.

Usage

```
iC10(x, seed=25435)
```

Arguments

`x`	An object with class `iC10features`: A list with elements 'train.CN', 'train.Exp', 'train.iC10', 'CN', 'Exp', 'map.cn', 'map.exp'
`seed`	seed to initialize random number generator. It is passed to `set.seed()`. See details.

Details

This function trains a `pamr` classifier and predicts the set of samples. The shrinkage parameter is obtained with crossvalidation, therefore different runs can give different results (unless a seed is specified).

Value

An object of class `iC10`. A list with the following elements:

`class`	Prediction classes for the samples
`posterior`	Probablitites for each sample to belong to each of the 10 groups
`centroids`	Shrunken Centroids for each of the 10 groups.
`fitted`	Normalized features for the samples classified.
`map.cn`	Annotation data for the copy number features
`map.exp`	Annotation data for the expression features

Author(s)

Oscar M. Rueda

References

Ali HR et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biology 2014; 15:431. Curtis et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346-352. Tibshirani et al. Diagnosis of multiple cancer types by shrunken centroids of gene expression. PNAS 2002; 99(10):6567-6572.

See Also

See `pamr.train`, `pamr.cv` and `pamr.predict` in package `pamr`.

Examples

```
require(iC10TrainingData)
data(train.CN)
data(train.Exp)
features <- matchFeatures(Exp=train.Exp, Exp.by.feat="probe")
features <- normalizeFeatures(features, "scale")
res <- iC10(features)
```

---

matchFeatures *Matching features from the classifier to the test data.*

Description

This function matches available copy number and/or expression data features to the training signatures; using either genomic position or HUGO gene name for copy number features and either Illumina probe names or HUGO gene name for expression features.

Usage

```
matchFeatures(CN = NULL, Exp = NULL,
CN.by.feat = c("gene", "probe"),
Exp.by.feat = c("gene", "probe"),
ref="hg19")
```

Arguments

`CN`	Data must be log2 copy number ratios. Two formats are allowed: - a matrix where each row represents a gene and each column a sample. In this case `CN.by.feat` must be "gene" and the rownames must be the hgnc gene names. - a data.frame with segmented data. The following columns must exist: 'ID' for the sample name, 'chromosome_name' for the chromosome (must be numeric),

	'loc.start' for the start position of the region, 'loc.end' for the end position of the region, 'seg.mean' for the log2ratio of the segment. If NULL, copy number is not used in the classifier.
`Exp`	Matrix with the expression data to classify. Each row must be a gene or an Illumina probe, and each column must correspond to a sample. Rownames must be either Illumina probes, in which case `Exp.by.feat` must be "probe"; or hgnc gene names, in which case `Exp.by.feat` must be "gene". If NULL, expression is not used in the classifier.
`CN.by.feat`	Either "probe" or "gene", Default is "probe".
`Exp.by.feat`	Either "probe" or "gene", Default is "gene".
`ref`	Etiher "hg18" or "hg19". It is used to match the copy number probes if `CN.by.feat` is "probe"

Details

One of `CN` or `Exp` must be not NULL. If matching is done by gene, hgnc gene name is used to match the rownames of the features. A list of synonym gene names is used (see `Map.All`). For copy number features matched by probe, the maximum log ratio in absolute value inside the limits of the feature is used. If there is no copy number in that region, the value of the probe before it is used.

Value

A list with the following elements is returned:

`CN`	copy number data to classify
`train.CN`	copy number training data
`Exp`	expression data to classify
`train.Exp`	expression training data
`train.iC10`	iC10 assignments for the training data
`map.cn`	annotation data for the copy number features
`map.exp`	annotation data for the expression features

Note

Note that the training set will be different, depending on the features matched. Genomic annotation for the training dataset has been obtained from Mark Dunning's lluminaHumanv3.db package.

Author(s)

Oscar M Rueda

References

Ali HR et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biology 2014; 15:431. Curtis et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346-352.

See Also

normalizeFeatures, iC10

Examples

```
require(iC10TrainingData)
data(train.CN)
data(train.Exp)
features <- matchFeatures(Exp=train.Exp,Exp.by.feat="probe", ref="hg18")
str(features)
```

---

normalizeFeatures *Normalization of expression features*

Description

Normalization of expression features. Several methods available in the package CONOR can be used.

Usage

```
normalizeFeatures(x, method=c("none", "scale"))
```

Arguments

x	An object result of a call to matchFeatures
method	Several methods are available: "none": No normalization is done "scale": Each expression feature is scaled to have zero mean and standard deviation 1

Details

No further normalization is needed on the copy number, as log2 ratios are comparable between platforms.

Value

A list of the same format as matchFeatures, but with train.Exp anfd Exp normalized.

Note

As CONOR package is no longer maintained, the methods are not available temporarily. We will include more normalization methods in the next version of this package.

Author(s)

Oscar M Rueda

References

Ali HR et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biology 2014; 15:431. Curtis et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346-352.

Examples

```
require(iC10TrainingData)
data(train.CN)
data(train.Exp)
features <- matchFeatures(Exp=train.Exp,
Exp.by.feat="probe", ref="hg18")
features <- normalizeFeatures(features, "scale")
```

---

plot.iC10 *Plot results of the iC10 classifier*

Description

Plot results of the iC10 classifier, in two different formats: either the signatures of the training set or the signatures of the new data classified.

Usage

```
## S3 method for class 'iC10'
plot(x, sample.name=1, newdata = NULL,...)
```

Arguments

x	An object of iC10 class:
sample.name	Number of sample to plot (if newdata is NULL). It can be either a number or a character with the sample name.
newdata	An object result to call to matchFeatures or normalizeFeatures containing the features of the samples to plot.
...	Additional arguments passed to plot.

Details

Two types of plots can be produced. If newdata is NULL, a panel 6x2 is drawn with the 10 profiles of the signatures of the training set and the profile of the features of sample.name and the distribution of the probabilities of classification to each iC10 for that sample. If newdata is not nutll, a panel 6x2 (with the 11th panel empty) is drawn with the 10 profiles of newdata samples and their distribution into the clusters. The features are sorted by type: copy number (if available) are drawn in grey, and then expression, each of them are sorted by genomic position.

Value

A 6x2 plot is produced.

Author(s)

Oscar M Rueda

References

Ali HR et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biology 2014; 15:431. Curtis et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346-352.

See Also

`iC10`

Examples

```
require(iC10TrainingData)
data(train.CN)
data(train.Exp)
features <- matchFeatures(Exp=train.Exp, Exp.by.feat="probe")
features <- normalizeFeatures(features, "scale")
res <- iC10(features)
plot(res, sample.name=10)
plot(res, newdata=features)
```

---

print.iC10 *Print results of the iC10 classifier*

Description

Print results of the iC10 classifier

Usage

```
## S3 method for class 'iC10'
print(x, ...)
```

Arguments

`x`	An object of `iC10` class:
`...`	Additional arguments passed to `print`.

Value

It returns a call to `str`.

Author(s)

Oscar M Rueda

References

Ali HR et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biology 2014; 15:431. Curtis et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346-352.

See Also

`iC10`

Examples

```
require(iC10TrainingData)
data(train.CN)
data(train.Exp)
features <- matchFeatures(Exp=train.Exp, Exp.by.feat="probe")
features <- normalizeFeatures(features, "scale")
res <- iC10(features)
res
```

---

summary.iC10 *Summary results of the iC10 classifier*

Description

Summary results of the iC10 classifier: shows the distribution of samples classified into each iC10 group and a summary of the maximum posterior probablity for each sample. Small values pinpoint samples with no clear group assigned.

Usage

```
## S3 method for class 'iC10'
summary(object, ...)
```

Arguments

`object`	An object of `iC10` class.
`...`	Additional arguments passed to `summary`.

Value

The function prints a table of the classification ad a summary of the maximum posterior probability for each sample.

Author(s)

Oscar M Rueda

References

Ali HR et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome Biology 2014; 15:431. Curtis et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012; 486:346-352. Tibshirani et al. Diagnosis of multiple cancer types by shrunken centroids of gene expression. PNAS 2002; 99(10):6567-6572.

See Also

See `iC10` and `pamr.train`, `pamr.cv` and `pamr.predict` in package `pamr`.

Examples

```
require(iC10TrainingData)
data(train.CN)
data(train.Exp)
features <- matchFeatures(Exp=train.Exp,
Exp.by.feat="probe", ref="hg18")
features <- normalizeFeatures(features, "scale")
res <- iC10(features)
summary(res)
```

Index