Package 'iC10TrainingData'

July 16, 2024

Type Package

Title Training Datasets for iC10 Package

Version 2.0.1

Date 2024-07-16

Author Oscar M Rueda and Jose Antonio Seoane Fernandez

Maintainer Oscar M. Rueda <Oscar.Rueda@mrc-bsu.cam.ac.uk>

Description Training datasets for iC10; which 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. Genomic annotation for the training dataset has been obtained from Mark Dunning's lluminaHumanv3.db package.

License GPL-3

NeedsCompilation no

Repository CRAN

Date/Publication 2024-07-16 08:00:02 UTC

Depends R (>= 3.5.0)

Contents

iC10TrainingData-package *Training Datasets for iC10 Package*

Description

Training datasets for iC10; which 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. Genomic annotation for the training dataset has been obtained from Mark Dunning's lluminaHumanv3.db package. Training datasets for iC10; which implements the classifier described in the METABRIC paper 'The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups' (Curtis et al., Nature 2012). 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

The DESCRIPTION file:

Package:	iC10TrainingData
Type:	Package
Title:	Training Datasets for iC10 Package
Version:	2.0.1
Date:	2024-07-16
Author:	Oscar M Rueda and Jose Antonio Seoane Fernandez
Maintainer:	Oscar M. Rueda <Oscar.Rueda@mrc-bsu.cam.ac.uk>
Description:	Training datasets for iC10; which implements the classifier described in the paper 'Genome-driven integ
License:	GPL-3
Packaged:	2014-09-15 12:22:07 UTC; rueda01
NeedsCompilation:	no
Repository:	CRAN
Date/Publication:	2014-09-26 12:13:01

Index of help topics:

Training datasets for iC10; which implements the classifier described in the METABRIC paper 'The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups' (Curtis et al., Nature 2012). 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.

Author(s)

Oscar M Rueda and Jose Antonio Seoane Fernandez

Maintainer: Oscar M. Rueda <Oscar.Rueda@mrc-bsu.cam.ac.uk>

References

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

`iC10`

Examples

```
data(train.CN)
data(train.Exp)
```

IntClustMemb *Class Membership for the training set*

Description

iC10 assignment for the Metabric training dataset (997 samples).

Usage

```
data(IntClustMemb)
```

Format

The format is: Factor w/ 10 levels "1","2","3","4",..: 2 9 3 3 8 6 7 7 7 3 ... - attr(*, "names")= chr [1:997] "MB.0135" "MB.0167" "MB.0136" "MB.3403" ...

Source

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

Examples

```
data(IntClustMemb)
barplot(table(IntClustMemb))
```

Map.All *Probe mapping of the complete set of features of the training set*

Description

Probe mapping of the complete set of features of the training set

Usage

```
data(Map.All)
```

Format

A data frame with 714 observations on the following 10 variables:

`Probe_ID` a character vector with the Illumina probe ids that flank the features

`Gene_symbol` a factor with the hugo gene names

`Ensembl_ID` a factor with the ensemble ids

`Cytoband` a factor with the cytobands (on hg18)

`Genomic_location_hg18` a factor with the genomic locations on hg18

`chromosome_name_hg18` a numeric vector with the chromosome on hg18

`start_position_hg18` a numeric vector with the start position on hg18

`end_position_hg18` a numeric vector with the end position on hg18

`Synonyms_0` a character vector with the gene name synonyms of the feature

`Gene.Chosen` a character vector (YES or NO) specifiying the probe chosen for gene-based selection

`Genomic_location_hg19` a factor with the genomic locations on hg19

`chromosome_name_hg19` a numeric vector with the chromosome on hg19

`start_position_hg19` a numeric vector with the start position on hg19

`end_position_hg19` a numeric vector with the end position on hg19

`chromosome_name_hg38` a numeric vector with the chromosome on hg38

`start_position_hg38` a numeric vector with the start position on hg38

`end_position_hg38` a numeric vector with the end position on hg38

Source

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

Examples

```
data(Map.All)
head(Map.All)
```

Map.CN *Probe mapping of the copy number features of the training set.*

Description

Probe mapping of the copy number features of the training set.

Usage

```
data(Map.CN)
```

Format

A data frame with 38 observations on the following 8 variables.

`Probe_ID` a character vector with the Illumina probe ids that flank the features

`Gene_symbol` a factor with the hugo gene names

`Ensembl_ID` a factor with the ensemble ids

`Cytoband` a factor with the cytobands (on hg18)

`Genomic_location_hg18` a factor with the genomic locations on hg18

`chromosome_name_hg18` a numeric vector with the chromosome on hg18

`start_position_hg18` a numeric vector with the start position on hg18

`end_position_hg18` a numeric vector with the end position on hg18

`Genomic_location_hg19` a factor with the genomic locations on hg19

`chromosome_name_hg19` a numeric vector with the chromosome on hg19

`start_position_hg19` a numeric vector with the start position on hg19

`end_position_hg19` a numeric vector with the end position on hg19

`chromosome_name_hg38` a numeric vector with the chromosome on hg38

`start_position_hg38` a numeric vector with the start position on hg38

`end_position_hg38` a numeric vector with the end position on hg38

Source

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

Examples

```
data(Map.CN)
head(Map.CN)
```

Map.Exp *Probe mapping of the Expression features of the training set*

Description

Probe mapping of the Expression features of the training set

Usage

```
data(Map.Exp)
```

Format

A data frame with 711 observations on the following 10 variables.

`Probe_ID` a character vector with the Illumina probe ids that flank the features

`Gene_symbol` a factor with the hugo gene names

`Ensembl_ID` a factor with the ensemble ids

`Cytoband` a factor with the cytobands (on hg18)

`Genomic_location_hg18` a factor with the genomic locations on hg18

`chromosome_name_hg18` a numeric vector with the chromosome on hg18

`start_position_hg18` a numeric vector with the start position on hg18

`end_position_hg18` a numeric vector with the end position on hg18

`Synonyms_0` a character vector with the gene name synonyms of the feature

`Gene.Chosen` a character vector (YES or NO) specifiying the probe chosen for gene-based selection

`Genomic_location_hg19` a factor with the genomic locations on hg19

`chromosome_name_hg19` a numeric vector with the chromosome on hg19

`start_position_hg19` a numeric vector with the start position on hg19

`end_position_hg19` a numeric vector with the end position on hg19

Source

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

Examples

```
data(Map.Exp)
head(Map.Exp)
```

train.CN *Copy number data for the training set*

Description

Copy number data for the training set

Usage

```
data(train.CN)
```

Format

A matrix with 714 rows and 997 columns. Rows are features and columns are training samples.

Details

Each row corresponds to one copy number feature for all samples in the training set. Note that it includes all features in the classifier. Note also that, depending on the data available and the type of matching (gene or probe) only some of the features will be used.

Source

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

Examples

```
data(train.CN)
summary(train.CN)
```

train.Exp *Expression data for the training set.*

Description

Expression data for the training set.

Usage

```
data(train.Exp)
```

Format

A matrix with 714 rows and 997 columns. Rows are features and columns are training samples.

Details

Each row corresponds to one expression feature for all samples in the training set. Note that it includes all features in the classifier. Note that, depending on the data available and the type of matching (gene or probe) only some of the features will be used.

Source

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

Examples

```
data(train.Exp)
summary(train.Exp)
```

Index