Package ‘citbcmst’

February 19, 2015

Version 1.0.4

Date 2014-01-08

Title CIT Breast Cancer Molecular SubTypes Prediction

Description This package implements the approach to assign tumor gene expression dataset to the 6 CIT Breast Cancer Molecular Subtypes described in Guedj et al 2012.

Author Laetitia Marisa, Aurelien de Reynies, Mickael Guedj

LazyData yes

Depends R(>= 2.10.0)

Maintainer Laetitia Marisa<`MarisaL@ligue-cancer.net`>

Contact For any question or bug please contact CITR@ligue-cancer.net

License GPL (>= 2)

URL `http://cit.ligue-cancer.net`

NeedsCompilation no

Repository CRAN

Date/Publication 2014-01-10 13:13:30

R topics documented:

`cit.assignBcmst` *cit.assignBcmst*

Description

assign expression data sample(s) to CIT Breast Cancer Molecular Subtypes

Usage

```
cit.assignBcmst(  data,
data.annot,
data.colId="Probe.Set.ID",
data.colMap=c("Probe.Set.ID","Gene.Symbol","Ensembl","UniGene.ID")[1],
citbcmst.annot=NULL,citbcmst.colId="Probe.Set.ID",
citbcmst.colMap=c("Probe.Set.ID","Gene.Symbol","Ensembl","UniGene.ID")[1],
dist.method="dlda",
dist.difftopcentcutoff=NULL,
dist.disttocentcutoff=NULL,
dist.maxcutoff=NULL,
plot=FALSE)
```

Arguments

`data`	a data.frame of expression data with id as rownames
`data.annot`	a data.frame of data probes annotations
`data.colId`	name of the column in data.annot containing data probes id
`data.colMap`	name of the column in data.annot containing data probes names to map
`citbcmst.annot`	affymetrix annotation data.frame, if NULL (default) takes the embeded annotation in object citbcmst$data.annot
`citbcmst.colId`	name of the column in citbcmst.annot corresponding to rownames of citbcmst$data. Default "Probe.Set.ID"
`citbcmst.colMap`	name of the column in citbcmst.annot containing the same annotation as in data.colMap
`dist.method`	metric to compute distance to assign a sample to a subtype. Default for Affymetrix data "dlda". For other platforms "pearson".
`dist.difftopcentcutoff`	cut-off on the differences between distances to centroids. If the distance is inferior to this cut-off for n centroids the sample is assigned to the n subtypes in the output variable citbcmst.mixed. If NULL, the cut-off is defined as the 1st decile of the difference between the top 2 closest centroids on data used to compute centroids.
`dist.disttocentcutoff`	cut-off on the mad (median absolute deviation) of distances to the centroid to define a sample as outlier. If the distance to the centroid of the assigned subtype is superior to `sdisttocent`*mad(distances of centroids samples to this centroid)

`dist.maxcutoff`	samples for which nearest centroid is above this threshold are discarted (used only if `dis.meth` = "pearson" or "spearman")
`plot`	if TRUE plot an acp of citdata used to classify, and of the input data with subtype affectation and dist to centroid class

Value

a data.frame with 4 columns : "citbcmst" classification to the closest of the 6 subtypes, "citbcmst.mixed" classification to the n closest subtypes depending on dist.difftopcentcutoff, "citbcmst.core" classification without outlier and mixed samples and citbcmst.confidence a confidence assignation annotation (CORE, MIXED or OUTLIER)

Note

This is a contribution from the Tumor Identity Cards (CIT) program founded by the 'Ligue Nationale Contre le Cancer' (France): `http://cit.ligue-cancer.net`. For any question please contact `CITR@ligue-cancer.net`

Author(s)

Laetitia Marisa

Examples

```
load(list.files(system.file("extdata", package="citbcmst"), full.names=TRUE)[1])
# load exp.norm.bertheau07 object stored in /inst/extdata
exp.annot.bertheau07 <- data.frame(id=rownames(exp.norm.bertheau07), stringsAsFactors=FALSE,
row.names=rownames(exp.norm.bertheau07) )
citbcmst.bertheau07 <- cit.assignBcmst(    data=exp.norm.bertheau07,
                                           data.annot=exp.annot.bertheau07,
                                           data.colId="id",
                                           data.colMap="id" ,
                                           citbcmst.colMap="Probe.Set.ID",
                                           dist.method="dlda",
                                           plot=TRUE
                                  )
```

`citbcmst` *Coreset data used to classify data in Guedj et 2011*

Description

Breast Cancer expression dataset from CIT program projects used as a coreset to define molecular classification

Usage

`citbcmst`

Format

A list containing the expression data of the 355 coreset samples and of the selected 375 probesets to classify data, the subtypes assigned to each sample and the Affymetrix annotation of the expression data from NetAffx (version na30)

Source

CIT http://cit.ligue-cancer.net

References

Guedj et al 2011

`exp.norm.bertheau07` *Affymetrix (U133A) Breast Cancer Expression Data from Bertheau et al 2009*

Description

Normalized expression data of 46 samples, reduced to the 241 probesets (originally 22283 probesets), hybridised on Affymetrix U133A chip

Usage

```
exp.norm.bertheau07
```

Format

A data.frame of normalized expression data of 46 samples and 22283 probesets

Source

ArrayExpress E-TABM-43

References

Bertheau P, Turpin E, Rickman DS, Espie M, de Reynies A, Feugeas JP, et al. (2007). Exquisite sensitivity of TP53 mutant and basal breast cancers to a dose-dense epirubicin-cyclophosphamide regimen. PLoS Med, 4: e90.

`exp.norm.chanrion08` *Non Affymetrix Breast Cancer Expression dataset from Chanrion et al 2008*

Description

Normalized expression data of 155 samples from a ER+ cohort for Tamoxifen response study, reduced to the 68 genes (originally 4397 70-mer oligonucleotides), hybridised on Qiagen/Operon oligonucleotide microarrays

Usage

```
exp.norm.chanrion08
```

Format

A data.frame of normalized expression data of 155 samples, reduced to 68 probes with gene symbol as rownames

Source

GEO GSE_9393

References

Chanrion M, Negre V, Fontaine H, Salvetat N, Bibeau F, et al. (2008). A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer. Clin Cancer Res, 14: 1744-1752.

Index