Package ‘citccmst’

February 19, 2015

Version 1.0.2

Date 2014-01-07

Title CIT Colon Cancer Molecular SubTypes Prediction

Description This package implements the approach to assign tumor gene expression dataset to the 6 CIT Colon Cancer Molecular Subtypes described in Marisa et al 2013.

Author Laetitia Marisa

LazyData yes

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 13:13:32

R topics documented:

`cit.assignCcmst` *cit.assignCcmst*

Description

assign expression data sample(s) to CIT Colon Cancer Molecular Subtype(s)

Usage

```
cit.assignCcmst(  data,
data.annot,
data.colId="Probe.Set.ID",
data.colMap=c("Probe.Set.ID","Gene.Symbol","Ensembl","UniGene.ID")[1],
citccmst.annot=NULL,citccmst.colId="Probe.Set.ID",
citccmst.colMap=c("Probe.Set.ID","Gene.Symbol","Ensembl","UniGene.ID")[1],
dist.method="dqda",
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
`citccmst.annot`	affymetrix annotation data.frame, if NULL (default) take the embedded annotation in object citccmst$data.annot
`citccmst.colId`	name of the column in citccmst.annot corresponding to rownames of citccmst$data. Default "Probe.Set.ID"
`citccmst.colMap`	name of the column in citccmst.annot containing the same annotation as in data.colMap
`dist.method`	metric to compute distance to assign a sample to a subtype ("pearson", "dlda", "dqda","euclidian"). Default "dqda".
`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 citccmst.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 discarded (used only if `dis.meth` = "pearson" or "spearman")
`plot`	if TRUE plot an acp of cit data used to classify, and of the input data with subtype affectation and dist to centroid class

Value

a data.frame with 4 columns : "citccmst" assignments to the closest of the 6 subtype centroids, "citccmst.mixed" assignment to the n closest subtypes depending on dist.difftopcentcutoff, "citccmst.core" assignments without outlier and mixed samples and citccmst.confidence a confidence assignment 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="citccmst"), full.names=TRUE))
##load citvalid.exp.norm object
#citvalid.exp.annot <- data.frame(id=rownames(citvalid.exp.norm), stringsAsFactors=FALSE,
#                                 row.names=rownames(citvalid.exp.norm) )
#citccmst <- cit.assignCcmst(     data=citvalid.exp.norm,
#                                 data.annot=citvalid.exp.annot,
#                                 data.colId="id",
#                                 data.colMap="id" ,
#                                 citccmst.colMap="Probe.Set.ID",
#                                 plot=TRUE
#                           )
#head(citccmst)
```

`citccmst` *CIT colon expresson dataset used to classify data in Marisa et al 2013*

Description

Colon Cancer expression dataset from CIT program used to define colon cancer molecular classification

Usage

```
citccmst
```

Format

A list containing the expression data of the 443 samples and of the selected 57 probesets to classify data, the subtypes assigned to each sample and the Affymetrix annotation of the expression data from NetAffx (version na31)

Source

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

References

Marisa et al 2013

Index