Package 'MetaGxPancreas'

March 20, 2025

Title Transcriptomic Pancreatic Cancer Datasets

Version 1.26.0

Description A collection of pancreatic Cancer transcriptomic datasets that are part of the MetaGxData package compendium. This package contains multiple pancreas cancer datasets that have been downloaded from various resources and turned into SummarizedExperiment objects. The details of how the authors normalized the data can be found in the experiment data section of the objects. Additionally, the location the data was obtained from can be found in the url variables of the experiment data portion of each SE.

License Artistic-2.0

Encoding UTF-8

Depends SummarizedExperiment, ExperimentHub, R (>= 3.6.0)

Imports stats, impute, S4Vectors, AnnotationHub

Suggests testthat, knitr, BiocStyle, rmarkdown, markdown

VignetteBuilder knitr

biocViews ExpressionData, ExperimentHub, CancerData, Homo_sapiens_Data, ArrayExpress, GEO, NCI, MicroarrayData, ExperimentData, SequencingData

LazyData yes

RoxygenNote 7.1.1

git_url https://git.bioconductor.org/packages/MetaGxPancreas

git_branch RELEASE_3_20

git_last_commit 9b0a357

git_last_commit_date 2024-10-29

Repository Bioconductor 3.20

Date/Publication 2025-03-20

Author Michael Zon [aut],
Vandana Sandhu [aut],
Christopher Eeles [ctb],
Benjamin Haibe-Kains [aut, cre]

Maintainer Benjamin Haibe-Kains <benjamin.haibe.kains@utoronto.ca>

Contents

loadPancreasDatasets *Function to load pancreas cancer expression profiles from the Experiment Hub*

Description

This function returns pancreas cancer patient cohorts in SummarizedExperiment object from the hub and a vector of patients from the datasets that are duplicates

Usage

```
loadPancreasDatasets(
  removeDuplicates = TRUE,
  quantileCutoff = 0,
  rescale = FALSE,
  minNumberGenes = NA,
  minSampleSize = NA,
  minNumberEvents = NA,
  removeSeqSubset = FALSE,
  keepCommonOnly = FALSE,
  imputeMissing = FALSE
)
```

Arguments

removeDuplicates

remove patients with a Spearman correlation greater than or equal to 0.98 with other patient expression profiles (default TRUE)

quantileCutoff A numeric between 0 and 1 specifying to remove genes with standard deviation below the required quantile (default 0)

rescale apply centering and scaling to the expression sets (default FALSE)

minNumberGenes an integer specifying to remove expression sets with less genes than this number (default 0)

minSampleSize an integer specifying the minimum number of patients required in an SE (default 0)

minNumberEvents

an integer specifying how man survival events must be in the dataset to keep the dataset (default 0)

removeSeqSubset

currently only removes the ICGSSEQ dataset as it contains the same patients as the ICGS microarray dataset (default TRUE, currently just ICGSSEQ)

keepCommonOnly remove probes not common to all datasets (default FALSE)

imputeMissing impute missing expression value via knn

Value

a list with two elements. The First element named SummarizedExperiments contains the datasets as Bioconductor SummarizedExperiment objects. The second element named duplicates contains a vector with patient IDs for the duplicate patients (those with Spearman correlation greater than or equal to 0.98 with other patient expression profiles).

Examples

```
sumExptsAndDuplicates <- loadPancreasDatasets()
```

Index