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 many 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)
loadPancreasDatasets

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

loadPancreasDatasets, 2