

MetaGxPancreas: a package for pancreatic cancer gene expression analysis

Michael Zon¹, Vandana Sandhu¹, and Benjamin Haibe-Kains*^{1,2}

¹Bioinformatics and Computational Genomics Laboratory, Princess
Margaret Cancer Center, University Health Network, Toronto,
Ontario, Canada

²Department of Medical Biophysics, University of Toronto, Toronto,
Canada

May 1, 2018

Contents

1	Installing the Package	1
2	Loading Datasets	2
3	Obtaining Sample Counts in Datasets	2
4	Session Info	3

1 Installing the Package

The MetaGxPancreas package is a compendium of Pancreatic Cancer datasets. The package is publicly available and can be installed from Bioconductor into R version 3.5.0 or higher. Currently, the phenoData for the datasets is overall survival status and overall survival time. This survival information is available for 11 of the 15 datasets.

To install the MetaGxPancreas package from Bioconductor:

*benjamin.haibe.kains@utoronto.ca

```
> source("http://bioconductor.org/biocLite.R")
> biocLite("MetaGxPancreas")
```

2 Loading Datasets

First we load the MetaGxPancreas package into the workspace.

To load the packages into R, please use the following commands:

```
> library(MetaGxPancreas)
> esets = MetaGxPancreas::loadPancreasEsets()[[1]]
```

This will load 15 expression datasets. Users can modify the parameters of the function to restrict datasets that do not meet certain criteria for loading. Some example parameters are shown below:

Datasets: Retain only genes that are common across all platforms loaded (default = FALSE)

Datasets: Retain studies with a minimum sample size (default = 0)

Datasets: Retain studies with a minimum number of genes (default = 0)

Datasets: Retain studies with a minimum number of survival events (default = 0)

Datasets: Remove duplicate samples (default = TRUE)

3 Obtaining Sample Counts in Datasets

To obtain the number of samples per dataset, run the following:

```
> numSamples <- vapply(seq_along(esets), FUN=function(i, esets){
+   ncol(esets[[i]]@assayData$exprs)
+ }, numeric(1), esets=esets)
> SampleNumberSummaryAll <- data.frame(NumberOfSamples = numSamples,
+                                       row.names = names(esets))
> total <- sum(SampleNumberSummaryAll[, "NumberOfSamples"])
> SampleNumberSummaryAll <- rbind(SampleNumberSummaryAll, total)
> rownames(SampleNumberSummaryAll)[nrow(SampleNumberSummaryAll)] <- "Total"
> require(xtable)
> print(xtable(SampleNumberSummaryAll, digits = 2), floating = FALSE)
```

	NumberOfSamples
BADEA	35.00
BALAGURANATH	25.00
CHEN	63.00
COLLISON	27.00
GRUTZMANN	11.00
ICGCMICRO	265.00
KIRBY	51.00
OUH	48.00
PCSI	118.00
PEI	36.00
TCGA	146.00
UNC	125.00
WINTER	30.00
ZHANG	42.00
Total	1022.00

4 Session Info

- R version 3.5.0 RC (2018-04-16 r74618), x86_64-pc-linux-gnu
- Locale: LC_CTYPE=en_US.UTF-8, LC_NUMERIC=C, LC_TIME=en_US.UTF-8, LC_COLLATE=C, LC_MONETARY=en_US.UTF-8, LC_MESSAGES=en_US.UTF-8, LC_PAPER=en_US.UTF-8, LC_NAME=C, LC_ADDRESS=C, LC_TELEPHONE=C, LC_MEASUREMENT=en_US.UTF-8, LC_IDENTIFICATION=C
- Running under: Ubuntu 16.04.4 LTS
- Matrix products: default
- BLAS: /home/biocbuild/bbs-3.8-bioc/R/lib/libRblas.so
- LAPACK: /home/biocbuild/bbs-3.8-bioc/R/lib/libRlapack.so
- Base packages: base, datasets, grDevices, graphics, methods, parallel, stats, utils
- Other packages: AnnotationHub 2.13.0, Biobase 2.41.0, BiocGenerics 0.27.0, ExperimentHub 1.7.0, MetaGxPancreas 1.1.0, impute 1.55.0, lattice 0.20-35, xtable 1.8-2

- Loaded via a namespace (and not attached): AnnotationDbi 1.43.0, BiocInstaller 1.31.1, DBI 0.8, IRanges 2.15.0, R6 2.2.2, RSQLite 2.1.0, Rcpp 0.12.16, S4Vectors 0.19.0, bit 1.1-12, bit64 0.9-7, blob 1.1.1, compiler 3.5.0, curl 3.2, digest 0.6.15, grid 3.5.0, htmltools 0.3.6, httpuv 1.4.1, httr 1.3.1, interactiveDisplayBase 1.19.0, later 0.7.1, magrittr 1.5, memoise 1.1.0, mime 0.5, pkgconfig 2.0.1, promises 1.0.1, shiny 1.0.5, stats4 3.5.0, tools 3.5.0, yaml 2.1.18