MM2Sdata: Companion data package for Medulloblastoma Subtype Predictions

Deena M.A. Gendoo[*1] and Benjamin Haibe-Kains[†1]

[1]Bioinformatics and Computational Genomics Laboratory, Princess Margaret Cancer Center, University Health Network, Toronto, Ontario, Canada
[2]Medical Biophysics Department, University of Toronto, Toronto, Ontario, Canada

March 15, 2019

Contents

1 Introduction

The MM2Sdata package is providing companion gene expression datasets for the MM2S package. The package contains normalized expression data for Human Medulloblastoma (GSE37418) as well as Mouse Medulloblastoma models (GSE36594).

Please refer to the manuscript URL: http://www.sciencedirect.com/science/article/pii/S0888754315000774

PLEASE NOTE: The MM2Sdata package contains a subset of samples from GSE36594 and GSE37418. Should you wish to test the full processed GSE36594 and GSE37418 datasets, these can be accessed from Github. (Please refer to the References section below for additional details).

2 Loading package for case studies

MM2Sdata is publicly available and can be installed in R version 2.13.0 or higher. To load the MM2Sdata package:

---

[*]deena.gendoo@utoronto.ca

[†]benjamin.haibe.kains@utoronto.ca

```
install.packages("MM2Sdata")
```

The MM2Sdata package under CRAN contains the following ExpressionSets:

GSE36594expr: Gene expression for 20 Medulloblastoma mouse samples that are sample replicates for the GTML mouse model.

GSE37418Expr: Gene expression for 10 primary Medulloblastoma human samples

```
library(MM2Sdata)

## Loading required package:  Biobase
## Loading required package:  BiocGenerics
## Loading required package:  parallel
##
## Attaching package:  'BiocGenerics'
## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, parApply, parCapply, parLapply,
##     parLapplyLB, parRapply, parSapply, parSapplyLB
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
## The following objects are masked from 'package:base':
##
##     Filter, Find, Map, Position, Reduce, anyDuplicated, append,
##     as.data.frame, basename, cbind, colMeans, colSums, colnames,
##     dirname, do.call, duplicated, eval, evalq, get, grep, grepl,
##     intersect, is.unsorted, lapply, lengths, mapply, match, mget,
##     order, paste, pmax, pmax.int, pmin, pmin.int, rank, rbind,
##     rowMeans, rowSums, rownames, sapply, setdiff, sort, table,
##     tapply, union, unique, unsplit, which, which.max, which.min
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'.  To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.
```

3 Explore the Mouse Gene Expression Object

We first load GSE36594expr, which contains mouse Medulloblastoma gene expression data. We can then search for a specific mouse genotype (GTML) for future analyses.

```
data(GSE36594Expr)
ExprMat<-exprs(GSE36594Expr)
GTML<-ExprMat[,grep("GTML_MB",(colnames(exprs(GSE36594Expr))))]

#Change mouse sample names for clarity
for(sample in 1:ncol(GTML))
{
  newnames<-strsplit(x=(colnames(GTML)[sample]),split="_")[[1]][1]
```

```
  colnames(GTML)[sample]<-newnames
}

# Determine Final Size of the Object
dim(GTML)

## [1] 23783    20
```

4 Explore the Human Gene Expression Object

We first load the gene expression data of 76 primary human patient tumours from GSE37418. We can determine the subtypes attributed to each sample.

```
data(GSE37418Expr)
head(pData(GSE37418Expr))

##                         sample geo_accession characteristics_ch1
## GSM918580_mbt006-u133v2      3     GSM918580        subgroup: WNT
## GSM918593_mbt035-u133v2     16     GSM918593        subgroup: WNT
## GSM918582_mbt009-u133v2      5     GSM918582        subgroup: SHH
## GSM918606_mbt075-u133v2     29     GSM918606        subgroup: SHH
## GSM918611_mbt085-u133v2     34     GSM918611         subgroup: G3
## GSM918589_mbt031-u133v2     12     GSM918589         subgroup: G3

# Determine exact sample counts for each MB subtype across the 76 samples
table(pData(GSE37418Expr)$characteristics_ch1)

##
##          subgroup: G3         subgroup: G4        subgroup: SHH
##                     3                    3                    2
## subgroup: SHH OUTLIER          subgroup: U          subgroup: WNT
##                     0                    0                    2
```

5 References and Extra Notes

The MM2Sdata package contains companion datasets for use in the MM2S package. Both MM2S and MM2Sdata are publicly available and can be installed in R version 2.13.0 or higher.

The current MM2Sdata package on CRAN is TRUNCATED for example purposes only and submission into CRAN. The full processed GSE36594 and GSE37418 datasets can be accessed from Github, as well as the Haibe-Kains (BHK) lab Website.

The MM2Sdata package under Github contains the following ExpressionSets:

GSE36594expr: Gene expression for 56 Medulloblastoma mouse samples, on which 32 are sample replicates for the GTML mouse model.

GSE37418Expr: Gene expression for 76 primary Medulloblastoma human samples.

Please refer to the following data repositories and websites for additional information, as necessary:

MM2Sdata on Github: https://github.com/DGendoo OR https://github.com/bhklab

BHK Lab Website: http://www.pmgenomics.ca/bhklab/software/mm2s

The following code snippet is an example installation of the data repositories from Github.

```
# library(Biobase)
# library(devtools)
# install_github(repo="DGendoo/MM2S")
# install_github(repo="DGendoo/MM2Sdata")
```

6 License

The MM2Sdata package is released under the Artistic-2.0 License.

The MM2Sdata package is provided "AS-IS" and without any warranty of any kind. In no event shall the University Health Network (UHN) or the authors be liable for any consequential damage of any kind, or any damages resulting from the use of MM2Sdata.

7 Session Info

```
## \begin{itemize}\raggedright
##   \item R version 3.5.3 (2019-03-11), \verb|x86_64-apple-darwin15.6.0|
##   \item Locale: \verb|C/en_CA.UTF-8/en_CA.UTF-8/C/en_CA.UTF-8/en_CA.UTF-8|
##   \item Running under: \verb|macOS High Sierra 10.13.6|
##   \item Matrix products: default
##   \item BLAS: \verb|/Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib|
##   \item LAPACK: \verb|/Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib|
##   \item Base packages: base, datasets, grDevices, graphics,
##     methods, parallel, stats, utils
##   \item Other packages: Biobase~2.42.0, BiocGenerics~0.28.0,
##     MM2Sdata~1.0.3
##   \item Loaded via a namespace (and not attached): compiler~3.5.3,
##     evaluate~0.12, highr~0.7, knitr~1.21, magrittr~1.5,
##     stringi~1.2.4, stringr~1.3.1, tools~3.5.3, xfun~0.4
## \end{itemize}
```