

MM2Sdata: Companion data package for Medulloblastoma Subtype Predictions

Deena M.A. Gendoo^{*1} and Benjamin Haibe-Kains^{†1}

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

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

June 10, 2015

Contents

1	Introduction	1
2	Loading package for case studies	1
3	Explore the Mouse Gene Expression Object	2
4	Explore the Human Gene Expression Object	2
5	Session Info	4

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>

2 Loading package for case studies

Both packages are publicly available and can be installed from Bioconductor version 2.8 or higher in R version 2.13.0 or higher.

The package contains the following ExpressionSets:

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

GSE37418Expr: Gene expression for 76 primary Medulloblastoma human samples

Please consult the manual of the MM2Sdata package for more details.

^{*}deena.gendoo@utoronto.ca

[†]benjamin.haibe.kains@utoronto.ca

```

#source("http://www.bioconductor.org/biocLite.R")
#biocLite("MM2Sdata")
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 object is masked from 'package:stats':
##
##   xtabs
##
## The following objects are masked from 'package:base':
##
##   anyDuplicated, append, as.data.frame, as.vector, cbind,
##   colnames, do.call, duplicated, eval, evalq, Filter, Find, get,
##   intersect, is.unsorted, lapply, Map, mapply, match, mget,
##   order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##   rbind, Reduce, rep.int, rownames, sapply, setdiff, sort,
##   table, tapply, union, unique, unlist, unsplit
##
## 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 33
```

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
## GSM918578_mbt003-u133v2      1     GSM918578      subgroup: G4
## GSM918579_mbt004-u133v2      2     GSM918579      subgroup: G4
## GSM918580_mbt006-u133v2      3     GSM918580      subgroup: WNT
## GSM918581_mbt008-u133v2      4     GSM918581      subgroup: G4
## GSM918582_mbt009-u133v2      5     GSM918582      subgroup: SHH
## GSM918583_mbt010-u133v2      6     GSM918583      subgroup: G4

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

##
##           subgroup: G3           subgroup: G4           subgroup: SHH
##                16                39                10
## subgroup: SHH OUTLIER           subgroup: U           subgroup: WNT
##                1                2                8
```

5 Session Info

```
## \begin{itemize}\raggedright
##   \item R version 3.2.0 Patched (2015-05-20 r68389), \verb|x86_64-apple-darwin10.8.0|
##   \item Locale: \verb|en_CA.UTF-8/en_CA.UTF-8/en_CA.UTF-8/C/en_CA.UTF-8/en_CA.UTF-8|
##   \item Base packages: base, datasets, graphics, grDevices,
##     methods, parallel, stats, utils
##   \item Other packages: Biobase~2.28.0, BiocGenerics~0.14.0,
##     knitr~1.10.5, MM2Sdata~1.0.0
##   \item Loaded via a namespace (and not attached): evaluate~0.7,
##     formatR~1.2, highr~0.5, magrittr~1.5, stringi~0.4-1,
##     stringr~1.0.0, tools~3.2.0
## \end{itemize}
```