

MM2S: a package for Medulloblastoma Subtype Predictions

Deena M.A. Gendoo*^{1,2} and Benjamin Haibe-Kains^{†1,2}

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

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

February 24, 2016

Contents

1	Introduction	1
2	Obtaining Microarray Gene Expression Data	1
3	References and Extra Notes	4
4	License	4
5	Session Info	4

1 Introduction

The *MM2S* package is providing relevant functions for subtype prediction of Medulloblastoma primary samples, mouse models, and cell lines.

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

Please also refer to the References section for additional information on downloading the *MM2S* package from Github, or running the *MM2S* server from the Lab website.

This vignette focuses on downloading and processing a gene expression dataset, and formatting it for use in *MM2S*.

2 Obtaining Microarray Gene Expression Data

We describe how to obtain and process Gene Expression Microarray Data Human Medulloblastoma Patients from the Gene Expression Omnibus (GEO). These data are downloaded using the GSE Series Number of the Dataset of Interest. For this example, we are downloading GSE37418 dataset, which contains 76 samples of Human MB patients.

First, we need to install several Bioconductor packages and the *MM2S* package, if they are not already available. To properly map probes from gene expression data against their corresponding gene symbols, we rely on BrainArray CDF. This requires an additional download the of CDF package and installation, as well as a modified Affy package. To do so, please run the following commands:

*deena.gendoo@utoronto.ca

†benjamin.haibe.kains@utoronto.ca

```

#Bioconductor package installation
source("http://bioconductor.org/biocLite.R")
biocLite(c("GEOquery", "Biobase"))
install.packages("MM2S", repos="http://cran.r-project.org")

#CDF installation
download.file(
  url = "http://mbni.org/customcdf/20.0.0/entrezg.download/hgu133plus2hsentrezgcdf_20.0.0.tar.gz",
  method = "auto", destfile = "hgu133plus2hsentrezgcdf_20.0.0.tar.gz")
install.packages("hgu133plus2hsentrezgcdf_20.0.0.tar.gz", type = "source", repos=NULL)

#Modified Affy package
download.file(
  url = "http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/20.0.0/affy_1.48.0.tar.gz",
  method = "auto", destfile = "affy_1.48.0.tar.gz")
install.packages("affy_1.48.0.tar.gz", type = "source", repos=NULL)

```

We now load all the libraries:

```

suppressPackageStartupMessages(library(MM2S))
suppressPackageStartupMessages(library(affy))
suppressPackageStartupMessages(library(Biobase))
suppressPackageStartupMessages(library(GEOquery))
suppressPackageStartupMessages(library(hgu133plus2hsentrezgcdf))

```

Next we download the cel files containing raw expression data into a folder.

```

gse<-getGEOSuppFiles(GEO = "GSE37418")
untar(tarfile = "./GSE37418/GSE37418_RAW.tar", exdir = "CelFiles")

```

We process the downloaded cell files and normalize the gene expression values using BrainArray CDF.

```

# Generate the Affy Expression Object
affyRaw <- ReadAffy(cefile.path = "CelFiles", verbose = F,
  cdfname="hgu133plus2hsentrezgcdf", compress = T)

# View object
affyRaw

#Perform Data Background Correction and Normalization
eset <- expresso(affyRaw, bgcorrect.method="rma", normalize.method="quantiles",
  pmcorrect.method="pmonly", summary.method="medianpolish", verbose = FALSE)
#Obtain the Microarray Expression Dataset
datamatrix<-exprs(eset)

# Polish the rownames (remove the _at from the Entrez IDs)
rownames(datamatrix)<-gsub(rownames(datamatrix), pattern="_at", replacement="")

# Create a new variable representing the cleaned microarray data that will be used in MM2S
ExprMatrix<-datamatrix

```

Now we can use this data and run several MM2S functions to determine the MB subtypes of given samples. We first perform MM2S predictions on a subset of the samples for demonstration.

```
# Conduct Subtype Predictions the samples, save results in a XLS file  
HumanPreds<-MM2S.human(InputMatrix=ExprMatrix[,1:10],xls_output=FALSE,parallelize=4)
```

Now these predictions can be viewed using a variety of MM2S functions. Here we generate a heatmap of MM2S confidence predictions for the 10 samples we have tested.

```
# Now generate a heatmap of the predictions and save the results in a PDF file.  
# This indicates MM2S confidence perdictions for each sample .  
# We view the samples here.  
PredictionsHeatmap(InputMatrix=HumanPreds$Predictions,pdf_output=TRUE,pdfheight=12,pdfwidth=10)
```

3 References and Extra Notes

Both MM2S and MM2Sdata are publicly available and can be installed in R version 2.13.0 or higher. Both packages are also available on Github. Companion datasets are also available on the Haibe-Kains (BHK) Lab website.

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

MM2S and 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")
```

A good tutorial on microarray data processing is also available on Biostars: <https://www.biostars.org/p/53870/>

4 License

The MM2S package is released under the GPL-3.0 License.

The MM2S 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 MM2S.

5 Session Info