

GSE104665

Angela Rizzo-Campos Guillermo Ayala

3 de abril de 2024

Índice general

1 Data	1
2 Packages	1
3 Downloading the data	1
4 Making the ExpressionSet	2

1 Data

This is a study of colorectal cancer. The data set can be downloaded from https://www.agilent.com/?gclid=EAIaIQobChMI2M2N_-SA6AIVWofVCh0vZAeOEAAAYASAAEgKSR_D_BwE&gclidsrc=aw.ds We will made the ExpressionSet.

We define the variable to control the execution of the code chunks.
The directory where the data sets have to be saved.

2 Packages

The needed packages are loaded.

```
pacman::p_load(Biobase, GEOquery, limma, hgug4112a.db)
```

3 Downloading the data

```
wd = getwd()
setwd(dirTamiData)
GEOquery::getGEOSuppFiles("GSE104645")
setwd("GSE104645")
system("tar xvf GSE104645_RAW.tar")
system("gzip -d *.gz")
x = read.maimages(dir(".", "txt"), "agilent", green.only = TRUE,
                  other.columns="gIsWellAboveBG")
GSE104645raw = x
save(GSE104645raw, file=paste0(dirTamiData, "GSE104645raw.rda"))
setwd(wd)
```

The previously save data set is loaded.

```
load(paste0(dirTamiData,"GSE104645raw.rda"))
```

The probes without ENTREZID identifier are removed.

```
a = AnnotationDbi::select(hgug4112a.db, keys=GSE104645raw$genes[, "ProbeName"],
                           column=c("ENTREZID", "ENSEMBL"), keytype="PROBEID")
a = a[!is.na(a[, "ENTREZID"]),]
c1 = match(unique(a[, 1]), a[, 1])
a1 = a[c1,]
c2 = match(unique(a1[, 2]), a1[, 2])
a2 = a1[c2,]
a2 = na.omit(a2)
GSE104645raw2 = GSE104645raw[match(a2[, 1], GSE104645raw$genes[, "ProbeName"]),]
dim(GSE104645raw2)
rownames(GSE104645raw2) = a2[, 1]
save(GSE104645raw2, file=paste0(dirTamiData, "GSE104645raw2.rda"))
```

Hacemos la corrección de fondo.

```
GSE104645e1 = backgroundCorrect(GSE104645raw2, method="normexp")
```

Ahora normalizamos aplicando una normalización de cuantiles.

```
GSE104645e1 = normalizeBetweenArrays(GSE104645e1, method="quantile")
```

Guardamos los datos.

```
save(GSE104645e1, file=paste0(dirTamiData, "GSE104645e1.rda"))
```

Loading the previously saved file.

```
load(paste0(dirTamiData, "GSE104645e1.rda"))
```

4 Making the ExpressionSet

Now we are going to make the ExpressionSet.

```
pd0 = data.frame(case = 1:ncol(GSE104645e1))
rownames(pd0) = colnames(GSE104645e1$E)
metadata = data.frame(labelDescription = c("case"),
                       row.names=colnames(pd0))
phenotypedata = new("AnnotatedDataFrame", data = pd0,
                    varMetadata = metadata)
experimentdata = new('MIAME', name="Okita A, Takahashi S, Ouchi K, Inoue M
et al.", lab="Tohoku University Hospital", contact = "akira.okita.d8@tohoku.ac.jp",
title = "Consensus molecular subtypes classification
of colorectal cancer as a predictive factor for chemotherapeutic
efficacy against metastatic colorectal cancer",
abstract = "Gene expression profile from 193 formalin-fixed and
paraffin embedded primary tumor samples.",
url = "https://www.ncbi.nlm.nih.gov/pubmed/29721154",
other = list(notes = "An example of Agilent microarray"))
```

```
GSE104645 = new("ExpressionSet",exprs=GSE104645e1$E,phenoData = phenotypedata,  
               experimentData = experimentdata,annotation = "hgug4112a.db")
```

```
fData(GSE104645) = a2[,1:2]  
save(GSE104645,file=paste0(dirTamiData,"GSE104645.rda"))
```