entrega1

Asier Ràfales Vila

15/5/2019

Series GSE64896

Title: Gene expression of distinct lung dendritic cell subsets

Organism: Mus musculus

Experiment type: Expression profiling by array

Summary:

Pulmonary dendritic cells are heterogenous cells comprise four distinct subsets including two conventional dendritic cell subsets, CD103+ and CD11bhiCD14lo cells, and two monocyte-derived dendritic cell subsets. Their functions in terms of migration and T cell activation are distinct, but genes regulating their features are to be determined. We used microarrays to identify a select set of genes that are expressed in conventinal dendritic cells and in monocyte-derived dendriti cells.

Overall design:

Four distinct lung DC subsets were purified by flow cytometry-based sorting after inhalation of lipopolusaccharide and ovalbumin. Each subset has three replicates.

Contributor(s):

Nakano H, Cook D ## Citation(s)
Nakano H, Moran TP, Nakano K, Gerrish KE et al. Complement receptor C5aR1/CD88 and dipeptidyl peptidase-4/CD26 define distinct hematopoietic lineages of dendritic cells. J Immunol 2015 Apr 15;194(8):3808-19. PMID: 25769922

Load Matrix into R environment

Contains tha samples as columns and the transcripts as rows so we can see the expression of the transcripts in the different samples.

```
express <- read.delim("~/Documentos/uni/2n/3r_trimestre/omics techniques/2nd_part/Rafales_Asier_OmicsTe
```

Load targets

Contains the sample_nam, target, replicate, gender and age.

```
targets <- read.delim("~/Documentos/uni/2n/3r_trimestre/omics techniques/2nd_part/Rafales_Asier_OmicsTe
```

Data analysis

Sumary of matrix expressions.

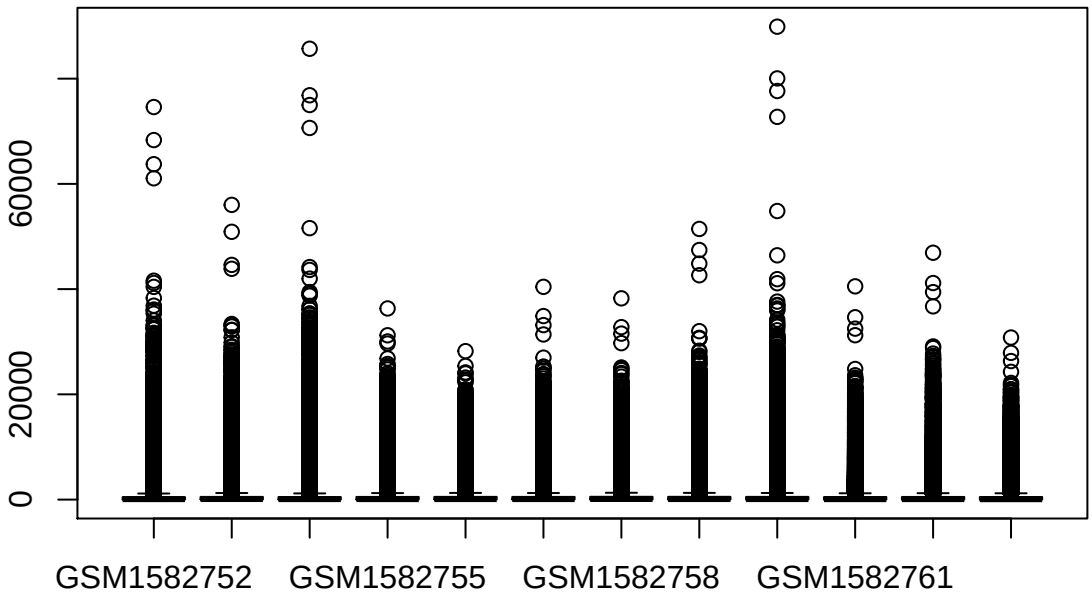
```
summary(express)
```

```
##    GSM1582752         GSM1582753         GSM1582754
##  Min.   :    0.06   Min.   :    0.04   Min.   :    0.09
##  1st Qu.:   16.73   1st Qu.:   19.41   1st Qu.:   19.90
##  Median :   82.78   Median :  103.82   Median :   91.39
##  Mean   :  776.37   Mean   :  738.99   Mean   :  785.13
##  3rd Qu.:  472.02   3rd Qu.:  509.50   3rd Qu.:  481.36
##  Max.   :74600.30   Max.   :56012.80   Max.   :85679.30
##    GSM1582755         GSM1582756          GSM1582757
##  Min.   :    0.10   Min.   :    0.029   Min.   :    0.03
##  1st Qu.:   17.36   1st Qu.:   17.572   1st Qu.:   18.61
##  Median :   88.27   Median :   95.221   Median :   94.10
##  Mean   :  719.51   Mean   :  693.145   Mean   :  708.26
##  3rd Qu.:  504.46   3rd Qu.:  516.469   3rd Qu.:  509.85
##  Max.   :36325.40   Max.   :28213.500   Max.   :40433.20
##    GSM1582758         GSM1582759         GSM1582760
##  Min.   :    0.04   Min.   :    0.03   Min.   :    0.13
##  1st Qu.:   19.63   1st Qu.:   20.41   1st Qu.:   20.53
##  Median :  108.72   Median :  101.82   Median :   96.72
##  Mean   :  706.03   Mean   :  733.64   Mean   :  776.47
##  3rd Qu.:  529.77   3rd Qu.:  523.15   3rd Qu.:  514.17
##  Max.   :38245.60   Max.   :51433.90   Max.   :89875.00
##    GSM1582761         GSM1582762         GSM1582763
##  Min.   :    0.02   Min.   :    0.08   Min.   :    0.035
##  1st Qu.:   17.01   1st Qu.:   18.28   1st Qu.:   14.394
##  Median :   83.23   Median :   87.71   Median :   72.412
##  Mean   :  711.59   Mean   :  733.97   Mean   :  697.236
##  3rd Qu.:  490.29   3rd Qu.:  499.35   3rd Qu.:  488.632
##  Max.   :40529.50   Max.   :46915.90   Max.   :30798.200
```

Here we can see that the means are similar in every sample, also that the min is a very low number and the max a very big number that difere so much of the first and third quartile so we can supose that there are some large outliers.

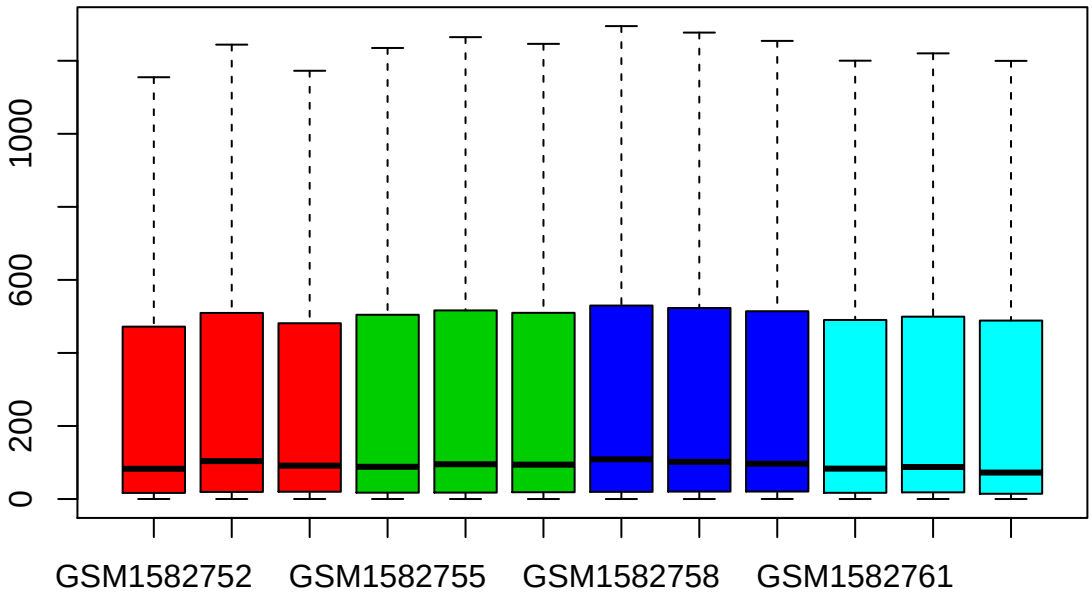
Boxplot

```
boxplot(express, col = c(2,2,2,3,3,3,4,4,4,5,5,5))
```



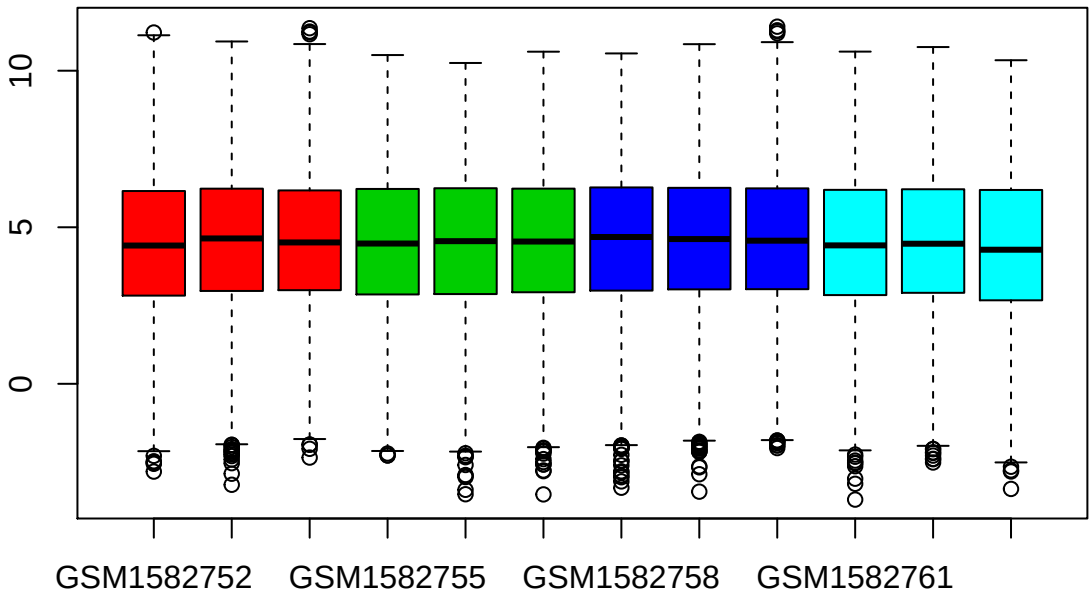
In this first box plot we can't see anythig exept that the means are low compared with the outliers, and that there are very much ouliers ans someones are very large. This confirem our Asumption on the sumary of the matrix expressions.

```
boxplot(express, col = c(2,2,2,3,3,3,4,4,4,5,5,5), outline = FALSE)
```



Here we can se better tha plot, we remove the outliers so we confirm that the means and quartiles are similar.

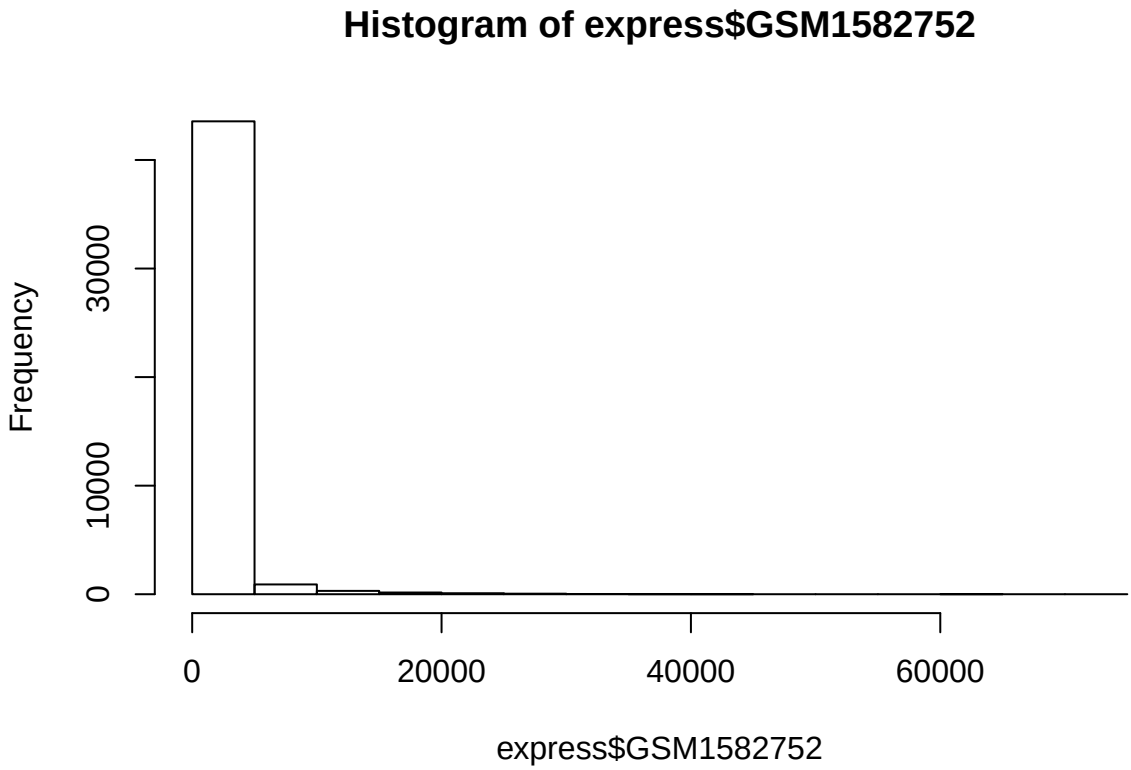
```
boxplot(log(express), col = c(2,2,2,3,3,3,4,4,4,5,5,5))
```



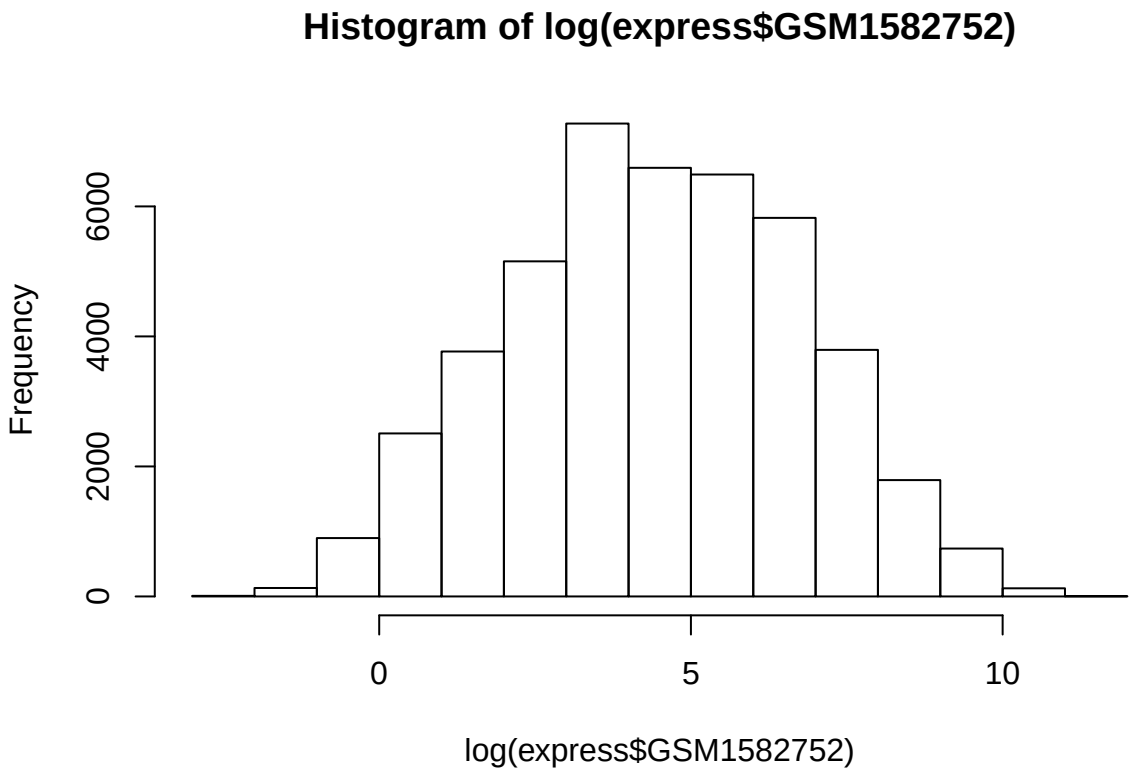
Here perform a log of the data to be able to see better the information.

Histograms

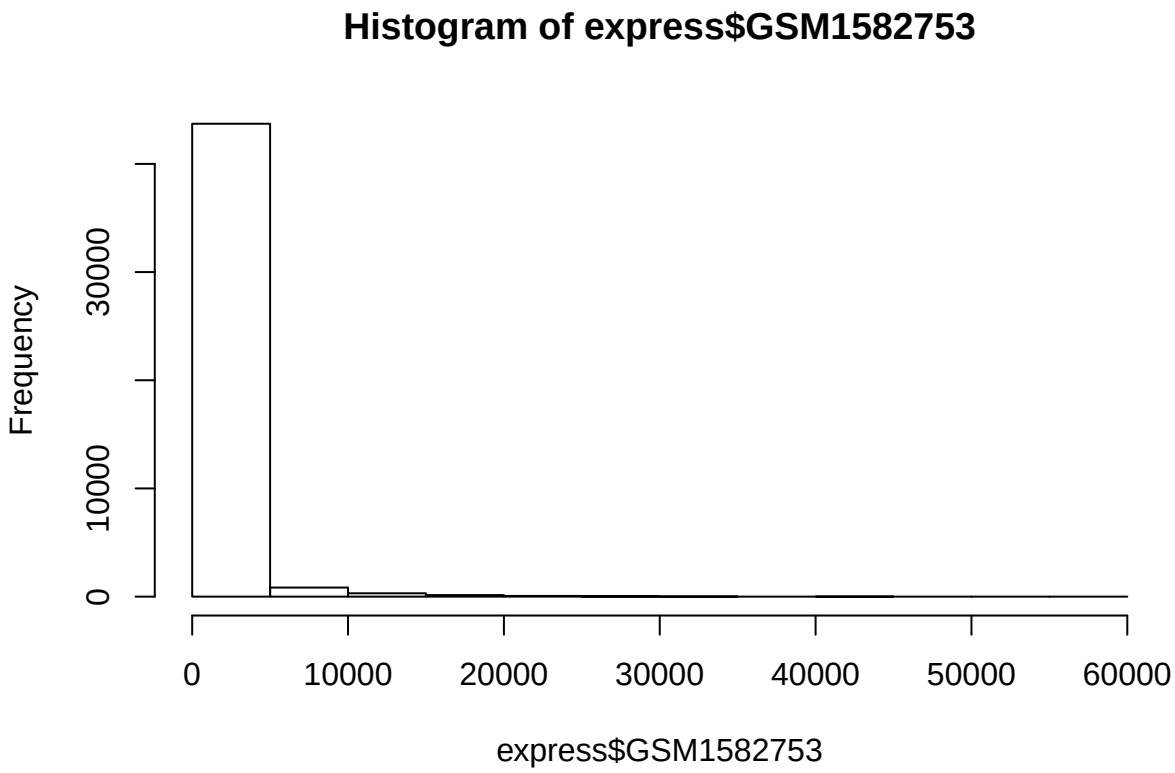
```
#1
##GSM1582752
hist(express$GSM1582752)
```



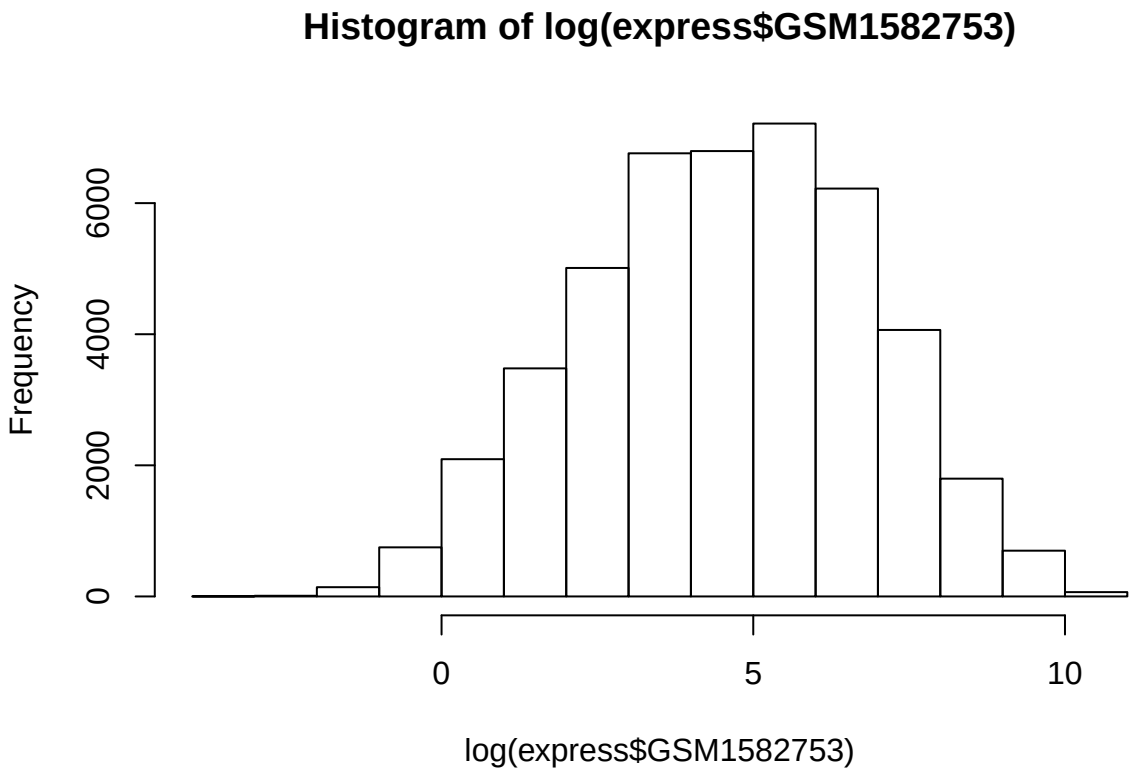
```r
hist(log(express$GSM1582752))
```




```r
##GSM1582753
hist(express$GSM1582753)
```



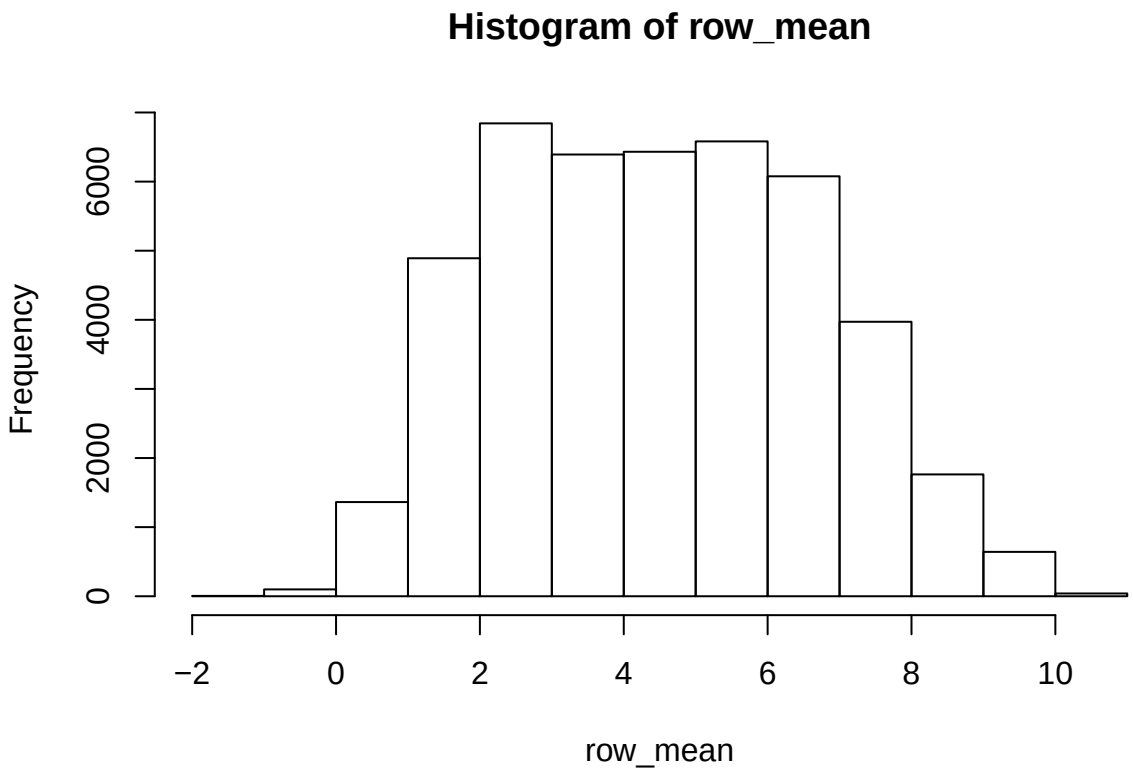
```r
hist(log(express$GSM1582753))
```



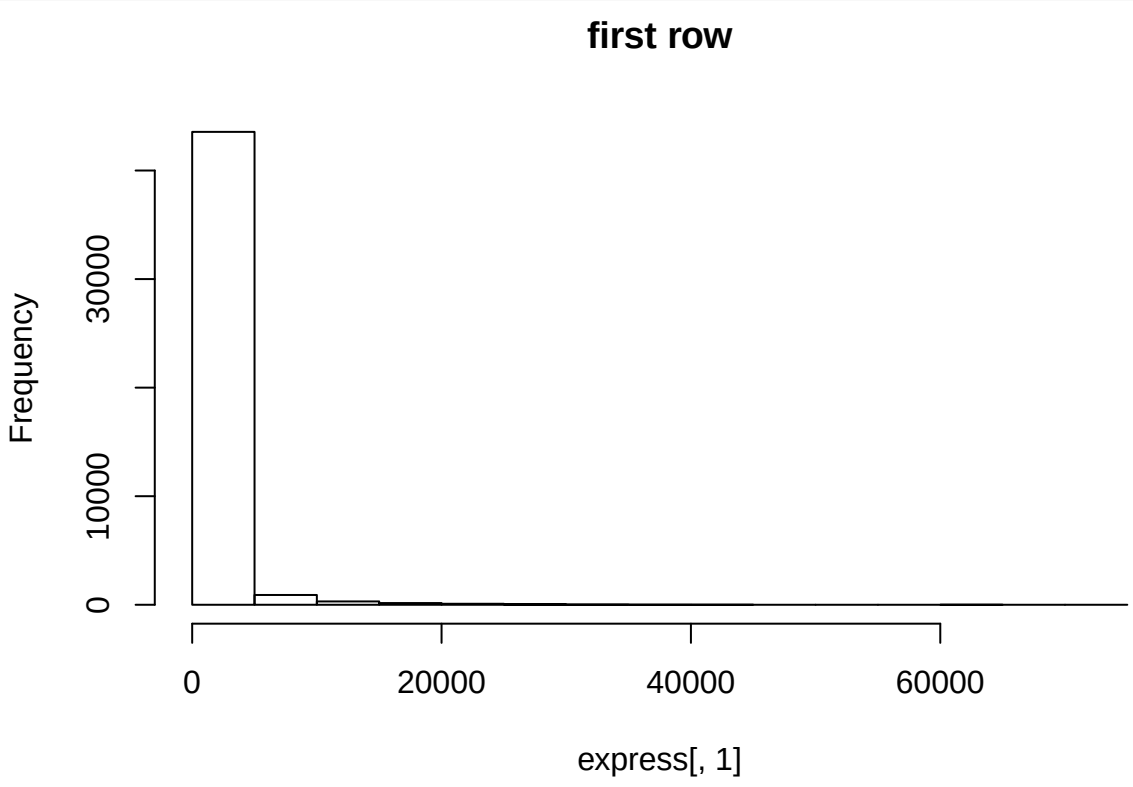
With these 4 histograms of the two of the samples we can assume that the res will be similar. The first histograms of each sample shows us that the majority of the expression is between 0 and 5000 The second ones, I have used those to confirm what we observe in the log boxplots.

Hist means

```r
row_mean <- rowMeans(log(express))
hist(row_mean)
```




```
## or
hist(express[,1],main="first row")
```



```
means <- apply(express,1,mean)
```

Install biocelite

```
if (!require(GEOquery)) {
  source("http://bioconductor.org/biocLite.R")
  biocLite("GEOquery")
  }
```

```
## Loading required package: GEOquery
```

```
## Setting options('download.file.method.GEOquery'='auto')
```

```
## Setting options('GEOquery.inmemory.gpl'=FALSE)
```

Expression sets

Now we are going to make a dataset with all information.

My info

```
##
myInfo=list(myName="Asier Rafales", myLab="Bioinformatics Lab",
          myContact="asier.rafales@alum.esci.upf.edu", myTitle="Practical Exercise on ExpressionSets")
show(myInfo)
```

```
## $myName
## [1] "Asier Rafales"
##
## $myLab
## [1] "Bioinformatics Lab"
##
## $myContact
## [1] "asier.rafales@alum.esci.upf.edu"
##
## $myTitle
## [1] "Practical Exercise on ExpressionSets"
```

```
myGenes <- colnames(express)
head(myGenes)
```

```
## [1] "GSM1582752" "GSM1582753" "GSM1582754" "GSM1582755" "GSM1582756"
## [6] "GSM1582757"
```

```
express2 <- data.matrix(express)
```

```
head(express)
```

```
##              GSM1582752 GSM1582753 GSM1582754 GSM1582755 GSM1582756
## 1415670_at     3240.650   2988.160   2861.250   3015.620   3105.240
## 1415671_at     3826.060   3329.820   3434.740   3515.940   3724.770
## 1415672_at     1713.440   2981.820   2539.800   2095.020   2564.900
```

```
## 1415673_at      174.158    415.275    281.054    198.522    285.155
## 1415674_a_at   2269.720   1793.640   1390.290   1651.190   1284.620
## 1415675_at     1076.620    838.742    788.479   1059.740   1169.250
##              GSM1582757 GSM1582758 GSM1582759 GSM1582760 GSM1582761
## 1415670_at     2751.230   2762.700   2900.190   2787.280   2780.850
## 1415671_at     3479.620   3567.860   3424.690   3964.880   4899.290
## 1415672_at     2345.180   2869.780   2185.000   2118.560   3800.830
## 1415673_at      277.386    315.861    279.232    365.565    233.091
## 1415674_a_at   1485.200   1508.270   1573.610   1531.340   1357.980
## 1415675_at      883.466    920.747   1003.530   1160.910   1016.120
##              GSM1582762 GSM1582763
## 1415670_at     2258.090   2742.190
## 1415671_at     4491.040   4759.990
## 1415672_at     2892.650   2863.870
## 1415673_at      245.356    207.875
## 1415674_a_at   1259.730   1157.930
## 1415675_at     1451.670   1046.010
```

```
myEset <- ExpressionSet(express2)
class(myEset)
```

```
## [1] "ExpressionSet"
## attr(,"package")
## [1] "Biobase"
```

```
show(myEset)
```

```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 45101 features, 12 samples
##   element names: exprs
## protocolData: none
## phenoData: none
## featureData: none
## experimentData: use 'experimentData(object)'
## Annotation:
```

```
columnDesc <-  data.frame(labelDescription= c("Sample Names", "target","replicate","colors", "Age at di
columnDesc
```

```
##              labelDescription
## 1                Sample Names
## 2                      target
## 3                   replicate
## 4                      colors
## 5 Age at disease onset(weeks)
## 6 Sex of patient (male/female
```

```
myAnnotDF <- new("AnnotatedDataFrame", data=targets, varMetadata= columnDesc)
show(myAnnotDF)
```

```
## An object of class 'AnnotatedDataFrame'
##   rowNames: 1 2 ... 12 (12 total)
##   varLabels: Sample_name Target ... age (6 total)
##   varMetadata: labelDescription
```

```
rownames(pData(myAnnotDF))<-pData(myAnnotDF)$Sample_name
myEset <- ExpressionSet(assayData=express2, phenoData=myAnnotDF)
```

```
show(myEset)
```

```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 45101 features, 12 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: GSM1582752 GSM1582753 ... GSM1582763 (12 total)
##   varLabels: Sample_name Target ... age (6 total)
##   varMetadata: labelDescription
## featureData: none
## experimentData: use 'experimentData(object)'
## Annotation:
```

```
myDesc <- new("MIAME", name= myInfo[["myName"]],
            lab= myInfo[["myLab"]],
            contact= myInfo[["myContact"]] ,
            title=myInfo[["myTitle"]])
print(myDesc)
```

```
## Experiment data
##   Experimenter name: Asier Rafales
##   Laboratory: Bioinformatics Lab
##   Contact information: asier.rafales@alum.esci.upf.edu
##   Title: Practical Exercise on ExpressionSets
##   URL:
##   PMIDs:
##   No abstract available.
```

Finished set

```
myEset <- ExpressionSet(assayData=express2,
                        phenoData=myAnnotDF,
                        fetureNames =myGenes,
                        experimentData = myDesc)
show(myEset)
```

```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 45101 features, 12 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: GSM1582752 GSM1582753 ... GSM1582763 (12 total)
##   varLabels: Sample_name Target ... age (6 total)
##   varMetadata: labelDescription
## featureData: none
## experimentData: use 'experimentData(object)'
## Annotation:
```

```
dim(exprs(myEset))
```

```
## [1] 45101    12
```

```
class(phenoData(myEset))
```

```
## [1] "AnnotatedDataFrame"
## attr(,"package")
## [1] "Biobase"
```

```
class(pData(phenoData(myEset)))
```

```
## [1] "data.frame"
```

```
head(pData(phenoData(myEset)))
```

```
##            Sample_name                  Target Replicate colors gender age
## GSM1582752  GSM1582752          LungCD103+cDCs       rep1      2   male  10
## GSM1582753  GSM1582753          LungCD103+cDCs       rep2      2   male  10
## GSM1582754  GSM1582754          LungCD103+cDCs       rep3      2   male  10
## GSM1582755  GSM1582755 LungCD11bhi_CD14lo_cDCs       rep1      3   male  10
## GSM1582756  GSM1582756 LungCD11bhi_CD14lo_cDCs       rep2      3   male  10
## GSM1582757  GSM1582757 LungCD11bhi_CD14lo_cDCs       rep3      3   male  10
```

```
head(pData(myEset))
```

```
##            Sample_name                  Target Replicate colors gender age
## GSM1582752  GSM1582752          LungCD103+cDCs       rep1      2   male  10
## GSM1582753  GSM1582753          LungCD103+cDCs       rep2      2   male  10
## GSM1582754  GSM1582754          LungCD103+cDCs       rep3      2   male  10
## GSM1582755  GSM1582755 LungCD11bhi_CD14lo_cDCs       rep1      3   male  10
## GSM1582756  GSM1582756 LungCD11bhi_CD14lo_cDCs       rep2      3   male  10
## GSM1582757  GSM1582757 LungCD11bhi_CD14lo_cDCs       rep3      3   male  10
```

Small set.

```
smallEset <- myEset[1:15,c(1:3,6:8)]
dim(exprs(smallEset))
```

```
## [1] 15  6
```

```
dim(pData(smallEset))
```

```
## [1] 6 6
```

```
head(pData(smallEset))
```

```
##            Sample_name                   Target Replicate colors gender
## GSM1582752  GSM1582752           LungCD103+cDCs       rep1      2   male
## GSM1582753  GSM1582753           LungCD103+cDCs       rep2      2   male
## GSM1582754  GSM1582754           LungCD103+cDCs       rep3      2   male
## GSM1582757  GSM1582757  LungCD11bhi_CD14lo_cDCs       rep3      3   male
## GSM1582758  GSM1582758 LungCD11bhi_CD14hi_moDCs       rep1      4   male
## GSM1582759  GSM1582759 LungCD11bhi_CD14hi_moDCs       rep2      4   male
##            age
## GSM1582752  10
## GSM1582753  10
## GSM1582754  10
## GSM1582757  10
## GSM1582758  10
```

```
## GSM1582759  10
```

```
all(colnames(exprs(smallEset))==rownames(pData(smallEset)))
```

```
## [1] TRUE
```

Young set.

```
youngEset <- myEset[,pData(myEset)$age<30]
dim(exprs(youngEset))
```

```
## [1] 45101    12
```

```
head(pData(youngEset))
```

```
##             Sample_name                  Target Replicate colors gender age
## GSM1582752  GSM1582752          LungCD103+cDCs      rep1      2   male  10
## GSM1582753  GSM1582753          LungCD103+cDCs      rep2      2   male  10
## GSM1582754  GSM1582754          LungCD103+cDCs      rep3      2   male  10
## GSM1582755  GSM1582755 LungCD11bhi_CD14lo_cDCs      rep1      3   male  10
## GSM1582756  GSM1582756 LungCD11bhi_CD14lo_cDCs      rep2      3   male  10
## GSM1582757  GSM1582757 LungCD11bhi_CD14lo_cDCs      rep3      3   male  10
```

Second method to performa a dataset

```
if (!require(GEOquery)) {
  BiocManager::install("GEOquery")
}
require(GEOquery)
gse <- getGEO("GSE64896")
```

```
## Found 1 file(s)
```

```
## GSE64896_series_matrix.txt.gz
```

```
## Parsed with column specification:
## cols(
##   ID_REF = col_character(),
##   GSM1582752 = col_double(),
##   GSM1582753 = col_double(),
##   GSM1582754 = col_double(),
##   GSM1582755 = col_double(),
##   GSM1582756 = col_double(),
##   GSM1582757 = col_double(),
##   GSM1582758 = col_double(),
##   GSM1582759 = col_double(),
##   GSM1582760 = col_double(),
##   GSM1582761 = col_double(),
##   GSM1582762 = col_double(),
##   GSM1582763 = col_double()
## )
```

```
## File stored at:
```

```
## /tmp/RtmpaVc0dp/GPL1261.soft
```

```
## Warning: 64 parsing failures.
##   row     col           expected    actual         file
## 45038 SPOT_ID 1/0/T/F/TRUE/FALSE --Control literal data
## 45039 SPOT_ID 1/0/T/F/TRUE/FALSE --Control literal data
## 45040 SPOT_ID 1/0/T/F/TRUE/FALSE --Control literal data
## 45041 SPOT_ID 1/0/T/F/TRUE/FALSE --Control literal data
## 45042 SPOT_ID 1/0/T/F/TRUE/FALSE --Control literal data
## ..... ....... .................. ......... .....------
## See problems(...) for more details.
```

```
class(gse)
```

```
## [1] "list"
```

```
names(gse)
```

```
## [1] "GSE64896_series_matrix.txt.gz"
```

```
gse[[1]]
```

```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 45101 features, 12 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: GSM1582752 GSM1582753 ... GSM1582763 (12 total)
##   varLabels: title geo_accession ... tissue:ch1 (44 total)
##   varMetadata: labelDescription
## featureData
##   featureNames: 1415670_at 1415671_at ... AFFX-TrpnX-M_at (45101
##     total)
##   fvarLabels: ID GB_ACC ... Gene Ontology Molecular Function (16
##     total)
##   fvarMetadata: Column Description labelDescription
## experimentData: use 'experimentData(object)'
## Annotation: GPL1261
```

```
esetFromGEO <- gse[[1]]

show(myEset)
```

```
## ExpressionSet (storageMode: lockedEnvironment)
## assayData: 45101 features, 12 samples
##   element names: exprs
## protocolData: none
## phenoData
##   sampleNames: GSM1582752 GSM1582753 ... GSM1582763 (12 total)
##   varLabels: Sample_name Target ... age (6 total)
##   varMetadata: labelDescription
## featureData: none
## experimentData: use 'experimentData(object)'
## Annotation:
```