

Analysis of Gene Expression

Transcriptional signature of prion-induced neurotoxicity in a *Drosophila* model of transmissible mammalian prion disease

Niek Scholten

11-03-21

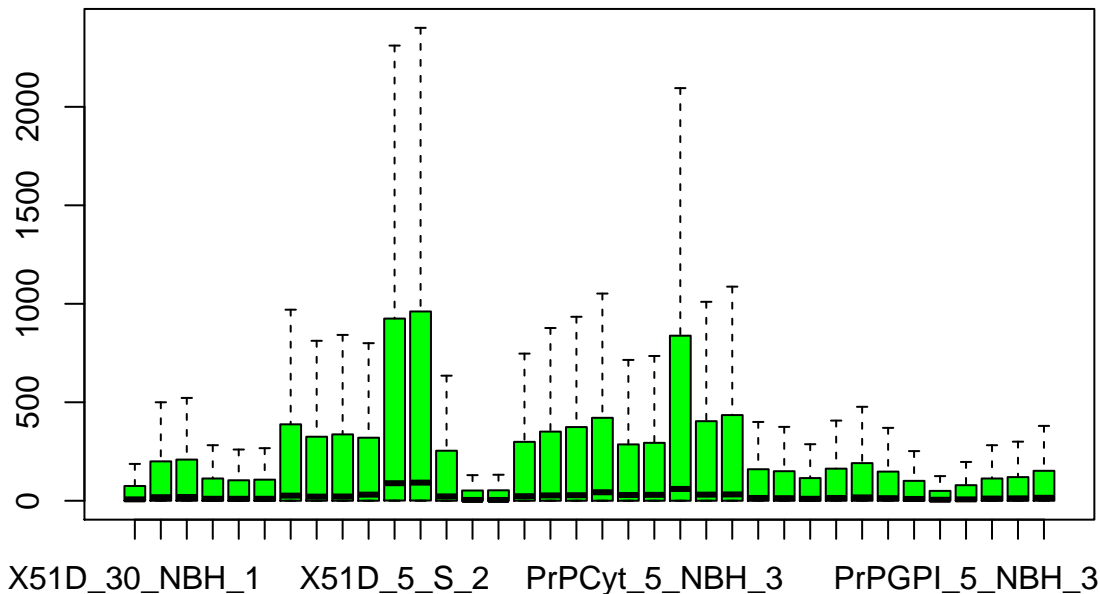
```
setwd("/home/nieks/Documents/Bio-Informatica/Thema 07 - Analyse van Genexpressie/Praktijkopdracht")
data <- read.table("Data/GSE144028.txt")
```

```
summary(data)
```

```
## X51D_30_NBH_1      X51D_30_NBH_2      X51D_30_NBH_3      X51D_30_S_1
## Min.   :      0    Min.   :      0    Min.   :      0    Min.   :      0
## 1st Qu.:      0    1st Qu.:      0    1st Qu.:      0    1st Qu.:      0
## Median :      7    Median :     18    Median :     19    Median :     10
## Mean   :    377    Mean   :    972    Mean   :   1030    Mean   :    568
## 3rd Qu.:     75    3rd Qu.:    200    3rd Qu.:    209    3rd Qu.:    113
## Max.   :3445037    Max.   :8342368    Max.   :8875291    Max.   :5560520
## X51D_30_S_2      X51D_30_S_3      X51D_5_NBH_1      X51D_5_NBH_2
## Min.   :      0    Min.   :      0    Min.   :      0    Min.   :      0.0
## 1st Qu.:      0    1st Qu.:      0    1st Qu.:      0    1st Qu.:      0.0
## Median :     10    Median :     10    Median :     26    Median :     21.0
## Mean   :    480    Mean   :    509    Mean   :    869    Mean   :    688.3
## 3rd Qu.:    104    3rd Qu.:    107    3rd Qu.:    388    3rd Qu.:    325.0
## Max.   :4122340    Max.   :4386825    Max.   :3832490    Max.   :2415360.0
## X51D_5_NBH_3      X51D_5_S_1      X51D_5_S_2      X51D_5_S_3
## Min.   :      0.0    Min.   :      0.0    Min.   :      0    Min.   :      0
## 1st Qu.:      0.0    1st Qu.:      0.0    1st Qu.:      1    1st Qu.:      1
## Median :     21.5    Median :     31.0    Median :     89    Median :     92
## Mean   :    718.8    Mean   :    722.4    Mean   :   1999    Mean   :   2092
## 3rd Qu.:    337.0    3rd Qu.:    320.0    3rd Qu.:    925    3rd Qu.:    961
## Max.   :2533918.0    Max.   :3111359.0    Max.   :7272134    Max.   :7625567
## PrPCyt_30_NBH_1    PrPCyt_30_NBH_2    PrPCyt_30_NBH_3    PrPCyt_30_S_1
## Min.   :      0    Min.   :      0.0    Min.   :      0.0    Min.   :      0
## 1st Qu.:      0    1st Qu.:      0.0    1st Qu.:      0.0    1st Qu.:      0
## Median :     22    Median :      4.0    Median :      4.0    Median :     23
## Mean   :    855    Mean   :    176.5    Mean   :    181.1    Mean   :    793
## 3rd Qu.:    254    3rd Qu.:    52.0    3rd Qu.:    53.0    3rd Qu.:    299
## Max.   :5261726    Max.   :1059586.0    Max.   :1096115.0    Max.   :4058764
## PrPCyt_30_S_2      PrPCyt_30_S_3      PrPCyt_5_NBH_1      PrPCyt_5_NBH_2
## Min.   :      0    Min.   :      0    Min.   :      0    Min.   :      0.0
## 1st Qu.:      0    1st Qu.:      0    1st Qu.:      0    1st Qu.:      0.0
## Median :     27    Median :     28    Median :     43    Median :     29.0
## Mean   :    857    Mean   :     914    Mean   :     828    Mean   :    591.5
## 3rd Qu.:    351    3rd Qu.:    374    3rd Qu.:    421    3rd Qu.:    286.0
## Max.   :3769299    Max.   :4079216    Max.   :3163765    Max.   :2692026.0
```

```
## PrPCyt_5_NBH_3      PrPCyt_5_S_1      PrPCyt_5_S_2      PrPCyt_5_S_3
## Min.   :    0.0      Min.   :    0      Min.   :    0.0      Min.   :    0.0
## 1st Qu.:    0.0      1st Qu.:    0      1st Qu.:    0.0      1st Qu.:    0.0
## Median :   30.0      Median :   60      Median :   31.0      Median :   32.0
## Mean   :  603.4      Mean   :  1537      Mean   :   764.7      Mean   :   821.5
## 3rd Qu.:  294.0      3rd Qu.:   838      3rd Qu.:  403.8      3rd Qu.:  435.0
## Max.   :2734069.0      Max.   :4603176      Max.   :2386987.0      Max.   :2556960.0
## PrPGPI_40_NBH_1      PrPGPI_40_NBH_2      PrPGPI_40_NBH_3      PrPGPI_40_S_1
## Min.   :    0      Min.   :    0      Min.   :    0      Min.   :    0
## 1st Qu.:    0      1st Qu.:    0      1st Qu.:    0      1st Qu.:    0
## Median :   14      Median :   13      Median :   10      Median :   14
## Mean   :  1556      Mean   :  1521      Mean   :  1116      Mean   :  1235
## 3rd Qu.:   160      3rd Qu.:   150      3rd Qu.:   115      3rd Qu.:   163
## Max.   :18885278      Max.   :18935887      Max.   :13407360      Max.   :14289546
## PrPGPI_40_S_2      PrPGPI_40_S_3      PrPGPI_5_NBH_1      PrPGPI_5_NBH_2
## Min.   :    0      Min.   :    0      Min.   :    0      Min.   :    0
## 1st Qu.:    0      1st Qu.:    0      1st Qu.:    0      1st Qu.:    0
## Median :   17      Median :   13      Median :    9      Median :    5
## Mean   :  1318      Mean   :   979      Mean   :  1077      Mean   :   629
## 3rd Qu.:   191      3rd Qu.:   148      3rd Qu.:   101      3rd Qu.:   50
## Max.   :14709751      Max.   :9635362      Max.   :11252267      Max.   :6579166
## PrPGPI_5_NBH_3      PrPGPI_5_S_1      PrPGPI_5_S_2      PrPGPI_5_S_3
## Min.   :    0      Min.   :    0      Min.   :    0      Min.   :    0
## 1st Qu.:    0      1st Qu.:    0      1st Qu.:    0      1st Qu.:    0
## Median :    7      Median :   11      Median :   12      Median :   15
## Mean   :   968      Mean   :   793      Mean   :   782      Mean   :  1018
## 3rd Qu.:    79      3rd Qu.:   113      3rd Qu.:   120      3rd Qu.:   152
## Max.   :10120453      Max.   :6111197      Max.   :6215874      Max.   :7851434
```

```
boxplot((data), outline = FALSE, col = 'green')
```



```
library(affy)
```

```
## 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':
##
##   anyDuplicated, append, as.data.frame, basename, cbind, colnames,
##   dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##   grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,
##   order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##   rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
##   union, unique, unsplit, which.max, which.min

## Loading required package: Biobase

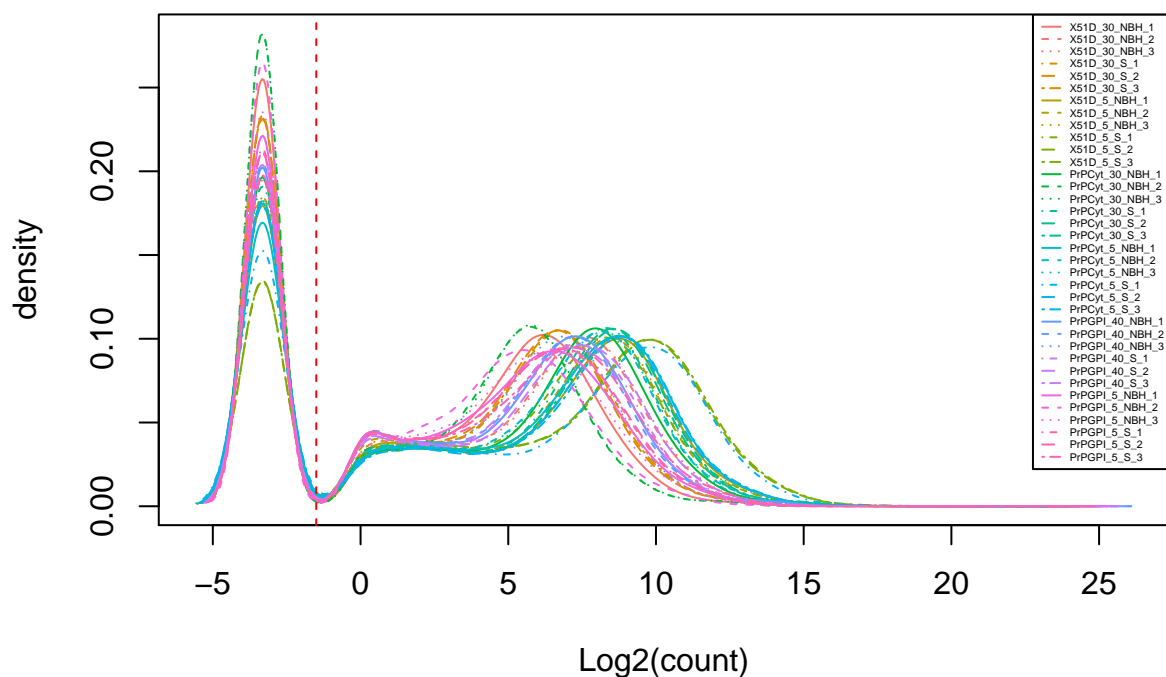
## Welcome to Bioconductor
##
##   Vignettes contain introductory material; view with
##   'browseVignettes()'. To cite Bioconductor, see
##   'citation("Biobase)"', and for packages 'citation("pkgname)".

library(scales)
myColors <- hue_pal()(12)

plotDensity(log2(data + 0.1), col=rep(myColors, each=3),
            lty=c(1:ncol(data)), xlab="Log2(count)",
            main="Expression Distribution")

legend('topright', names(data), lty=c(1:ncol(data)),
       col=rep(myColors, each=3), cex=0.32)
abline(v=-1.5, lwd=1, col='red', lty=2)
```

Expression Distribution



```
library(DESeq2)
```

```
## Loading required package: S4Vectors
## Loading required package: stats4
##
## Attaching package: 'S4Vectors'
## The following object is masked from 'package:base':
##
##     expand.grid
## Loading required package: IRanges
## Loading required package: GenomicRanges
## Loading required package: GenomeInfoDb
## Loading required package: SummarizedExperiment
## Loading required package: MatrixGenerics
## Loading required package: matrixStats
##
## Attaching package: 'matrixStats'
## The following objects are masked from 'package:Biobase':
##
##     anyMissing, rowMedians
##
## Attaching package: 'MatrixGenerics'
## The following objects are masked from 'package:matrixStats':
```

```

##
##   colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,
##   colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
##   colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
##   colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
##   colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
##   colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
##   colWeightedMeans, colWeightedMedians, colWeightedSds,
##   colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,
##   rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
##   rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
##   rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##   rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##   rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
##   rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##   rowWeightedSds, rowWeightedVars

## The following object is masked from 'package:Biobase':
##
##   rowMedians

(ddsMat <- DESeqDataSetFromMatrix(countData = data,
                                  colData = data.frame(samples = names(data)),
                                  design = ~ 1))

## class: DESeqDataSet
## dim: 17742 36
## metadata(1): version
## assays(1): counts
## rownames(17742): FBgn0000003 FBgn0000008 ... __not_aligned
##   __too_low_aQual
## rowData names(0):
## colnames(36): X51D_30_NBH_1 X51D_30_NBH_2 ... PrPGPI_5_S_2 PrPGPI_5_S_3
## colData names(1): samples

rld.dds <- vst(ddsMat)
rld <- assay(rld.dds)

sampledists <- dist( t( rld ))

library(pheatmap)

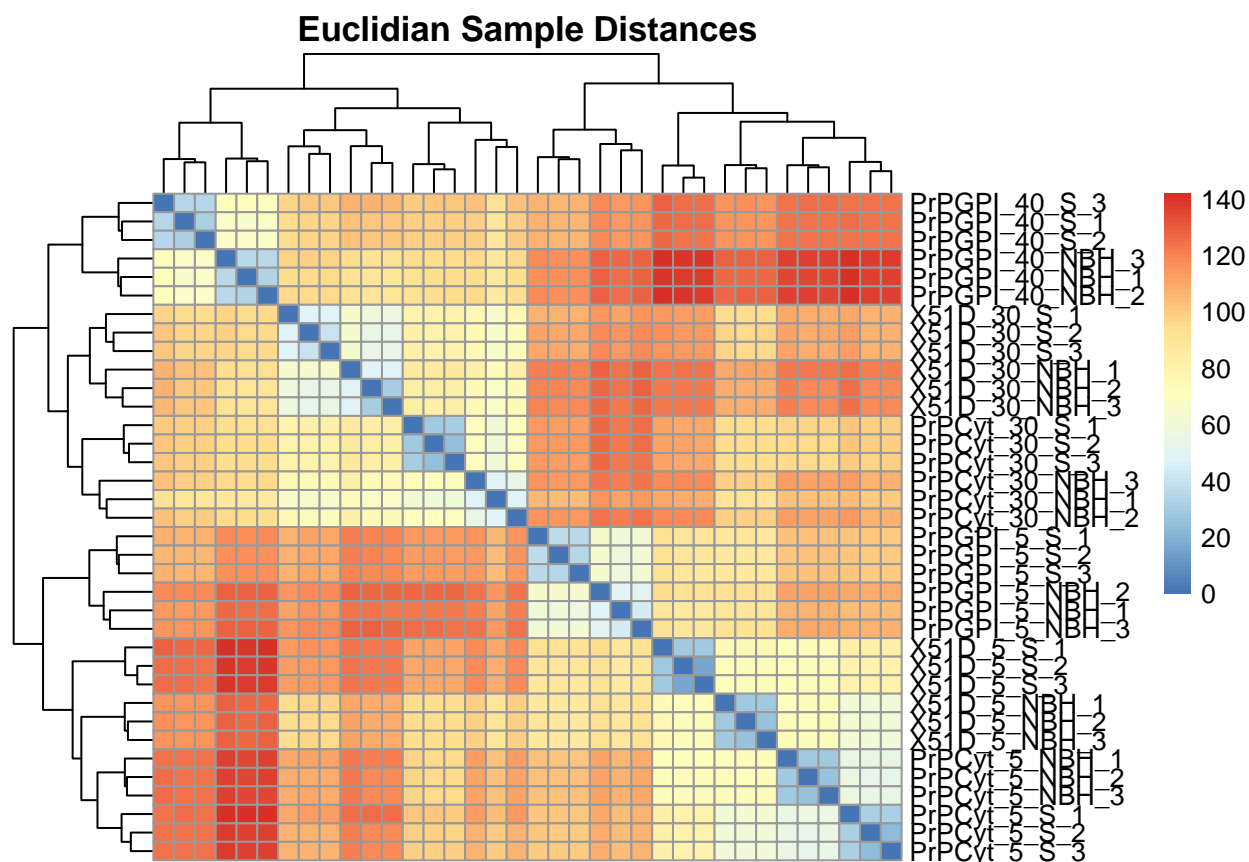
sampleDistMatrix <- as.matrix(sampledists)

annotation <- data.frame(Type = factor(rep(rep(1:2, each = 3), each = 6),
                                     labels = c("Normal Brain Homogenate", "Scrapie")))

rownames(annotation) <- names(counts)

pheatmap(sampleDistMatrix, show_colnames = FALSE,
          # annotation_col = annotation, # Gives an error
          clustering_distance_rows = sampledists,
          clustering_distance_cols = sampledists,
          main = "Euclidian Sample Distances")

```



```
library(PoiClaClu)
library(ggplot2)

dds <- assay(ddsMat)
poisd <- PoissonDistance( t(dds) )

samplePoisDistMatrix <- as.matrix(poisd$dd)

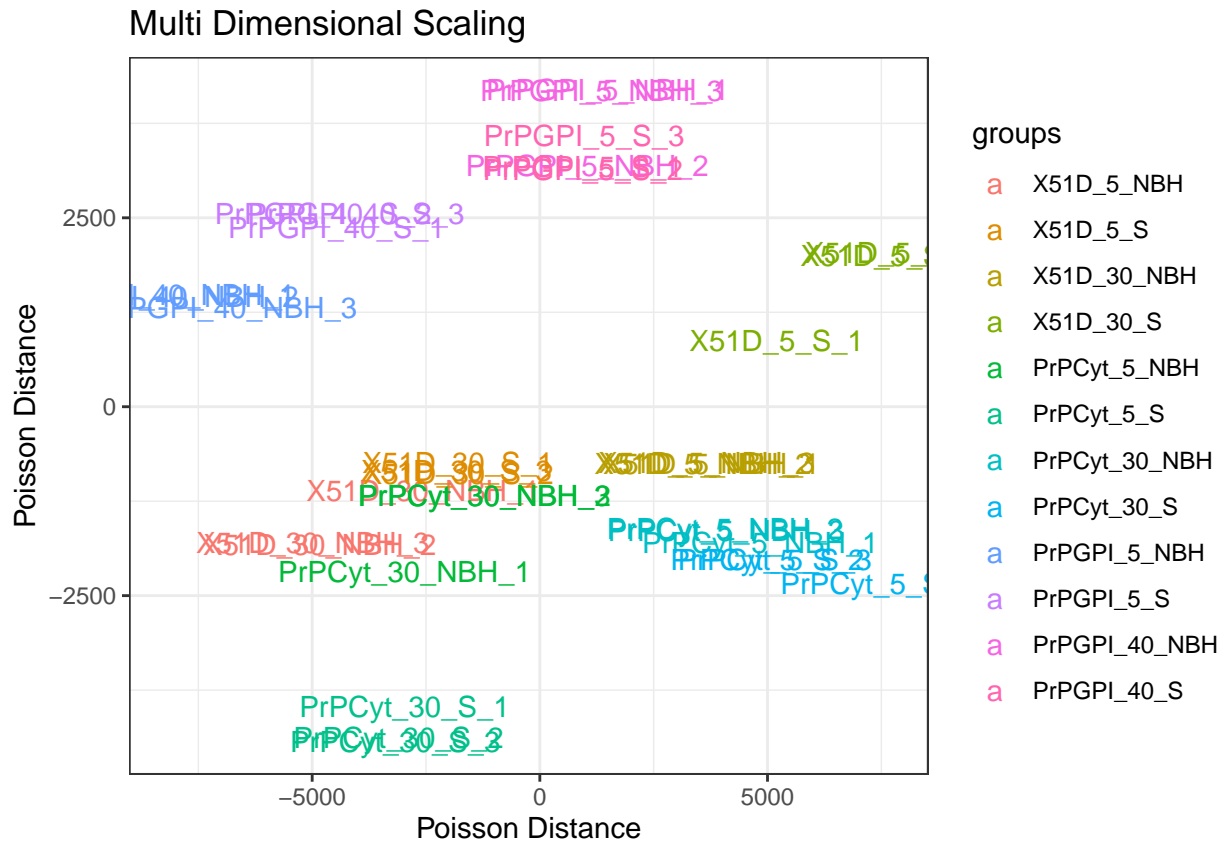
mdsPoisData <- data.frame( cmdscale(samplePoisDistMatrix) )

names(mdsPoisData) <- c('x_coord', 'y_coord')

groups <- factor(rep(1:12, each=3),
                 labels = c("X51D_5_NBH",
                           "X51D_5_S",
                           "X51D_30_NBH",
                           "X51D_30_S",
                           "PrPCyt_5_NBH",
                           "PrPCyt_5_S",
                           "PrPCyt_30_NBH",
                           "PrPCyt_30_S",
                           "PrPGPI_5_NBH",
                           "PrPGPI_5_S",
                           "PrPGPI_40_NBH",
                           "PrPGPI_40_S"))

coldata <- names(data)
```

```
ggplot(mdsPoisData, aes(x_coord, y_coord, color = groups, label = coldata)) +
  geom_text(size = 4) +
  ggtitle('Multi Dimensional Scaling') +
  labs(x = "Poisson Distance", y = "Poisson Distance") +
  theme_bw()
```



Some samples clearly deviate from the other 2 in the group. This is especially clear with X51D_5_S, PrPCyt_30_NBH, X51D_30_NBH & PrPCyt_5_S. Strangely, these samples are all the first one in their respective group. This could indicate that the first tests were less accurate. Since 3 samples must remain in each group, no data will be removed from the set.