Homework #2 - Stefan Zdraljevic Normalization of Affymetrix data set

This dataset was obtained from the GEO database - Reference series **GSE21784**.

The data consists of *C.elegans* expression data at three timepoints during the lifecycle (L4, adult day 6, adult day 15). Each timepoint is replicated in triplicate.

Data was downloaded, saved, and unzipped in my homeworke dirtectory

Load required packages

```
# source("http://bioconductor.org/biocLite.R")
# biocLite("affy")
library(affy)


setwd("~/Northwestern/Courses/Stats_for_Bioinformatics/Homework/HW2/GSE21784_RAW/")
```

Read in .CEL files and rename

```
expd <- ReadAffy()
sampleNames(expd) <- c("L4_1","L4_2","L4_3", "Day6_1", "Day6_2","Day6_3","Day15_1","Day15_2
","Day15_3")
```
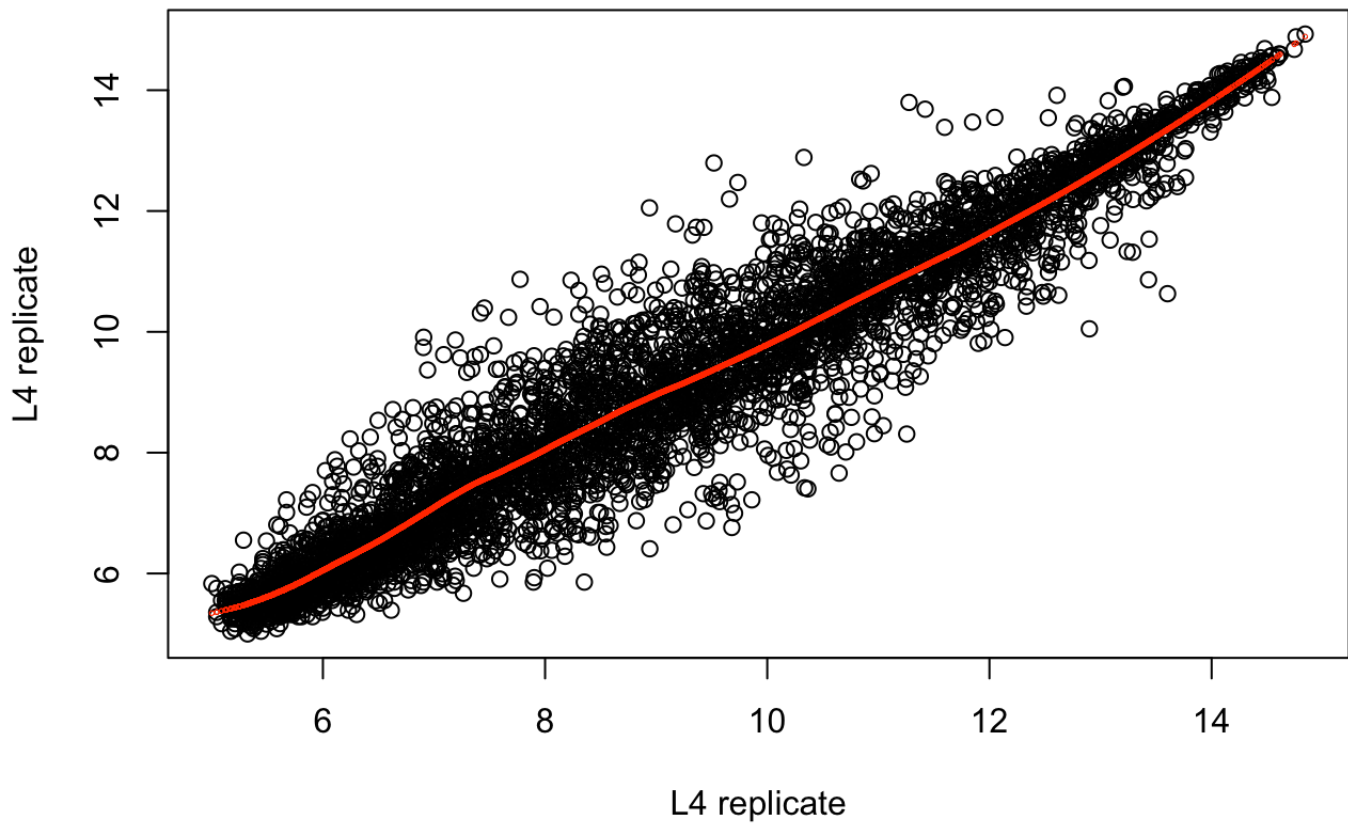
Plot two replicate experiments against each other

```
y <- log2(pm(expd)[,1])[1:5000]
```

```
##
```

```
x <- log2(pm(expd)[,2])[1:5000]
LO <- loess(y~x,span=0.3)

plot(x,y,main="L4 vs L4 replicates",xlab="L4 replicate",ylab="L4 replicate")
points(x,predict(LO),col=2,lwd=.5,cex=.3)
```
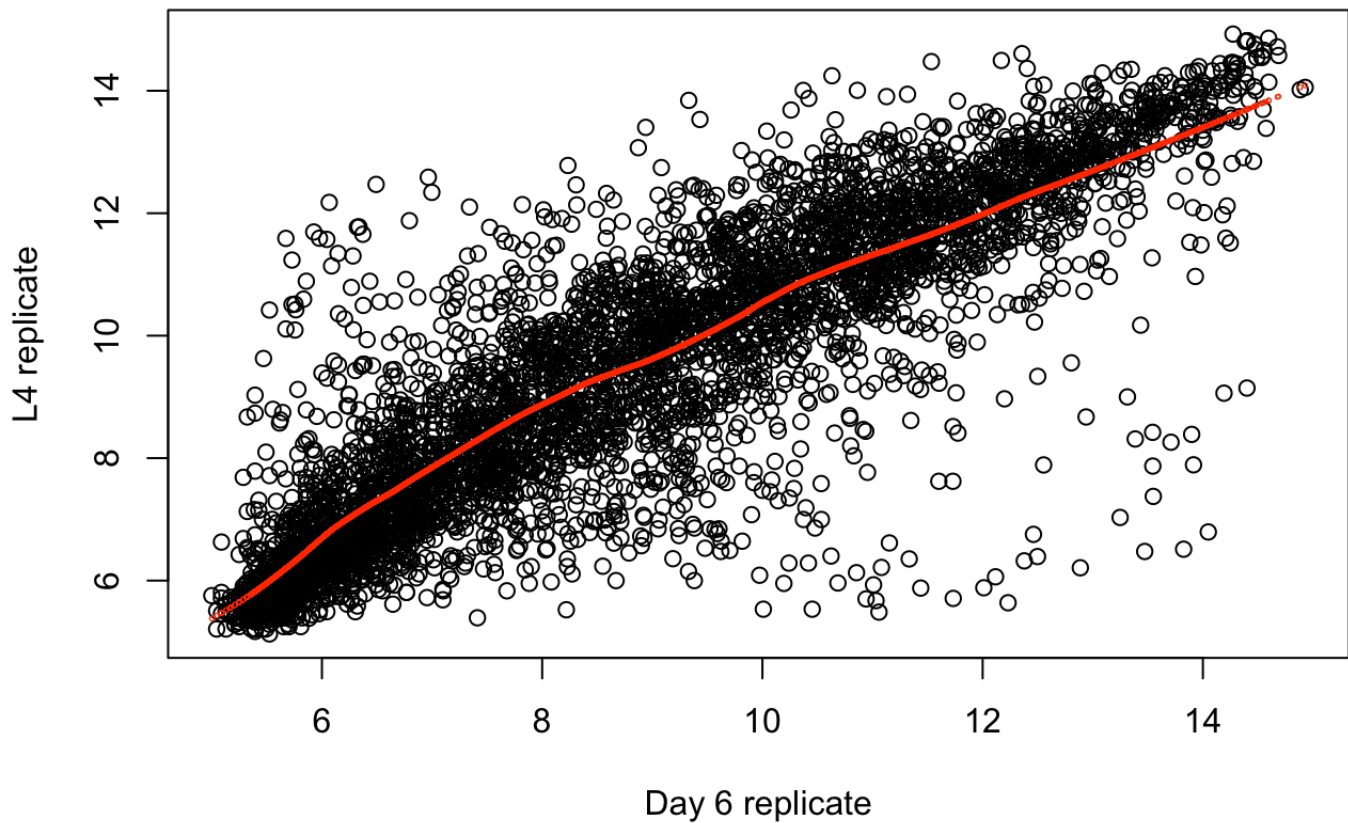
L4 vs L4 replicates

Plot L4 vs Day 6

```
x <- log2(pm(expd)[,1])[1:5000]
y <- log2(pm(expd)[,4])[1:5000]
LO <- loess(y~x,span=0.3)

plot(x,y,main="L4 vs Day 6",xlab="Day 6 replicate",ylab="L4 replicate")
points(x,predict(LO),col=2,lwd=.5,cex=.3)
```
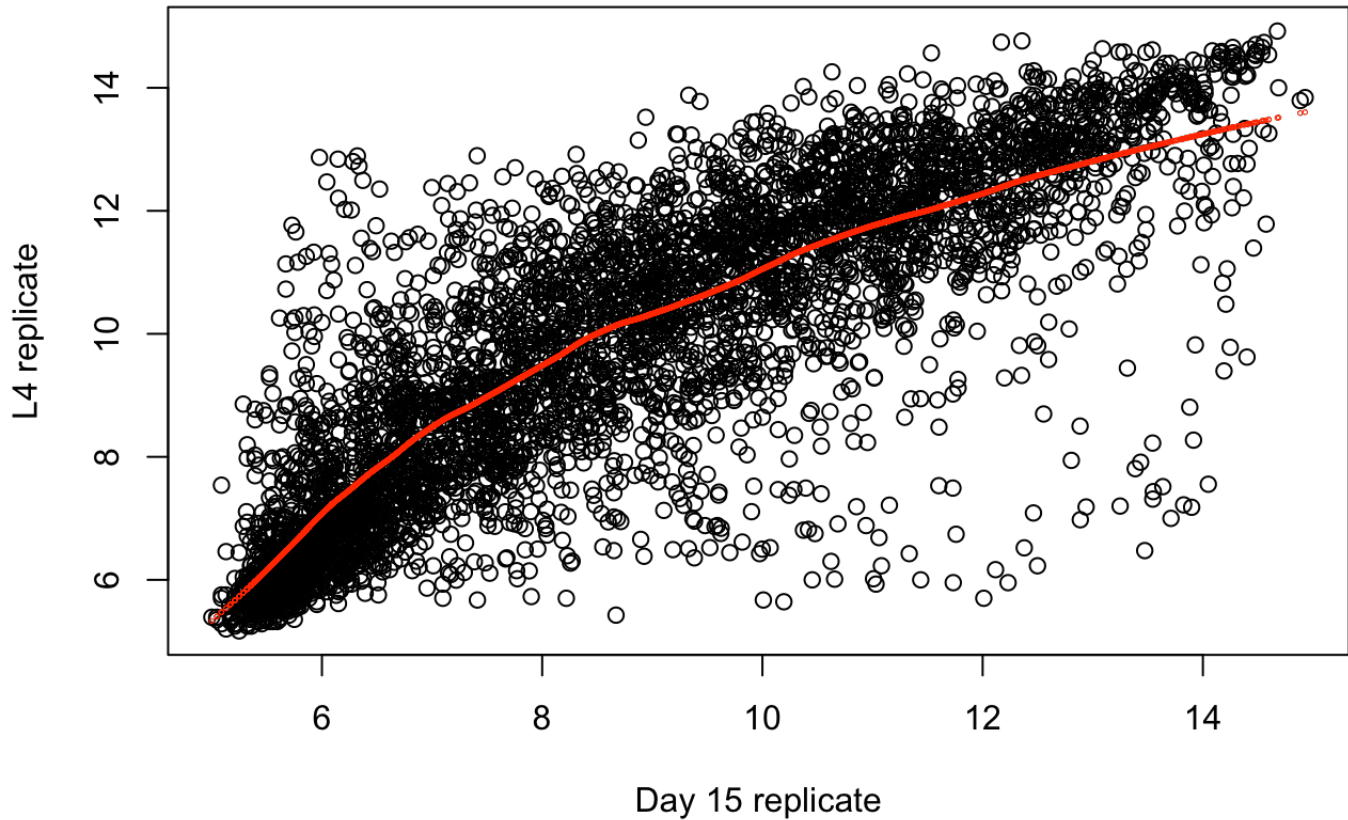
L4 vs Day 6

Plot L4 vs Day 15

```
x <- log2(pm(expd)[,1])[1:5000]
y <- log2(pm(expd)[,7])[1:5000]
LO <- loess(y~x,span=0.3)

plot(x,y,main="L4 vs Day 15",xlab="Day 15 replicate",ylab="L4 replicate")
points(x,predict(LO),col=2,lwd=.5,cex=.3)
```
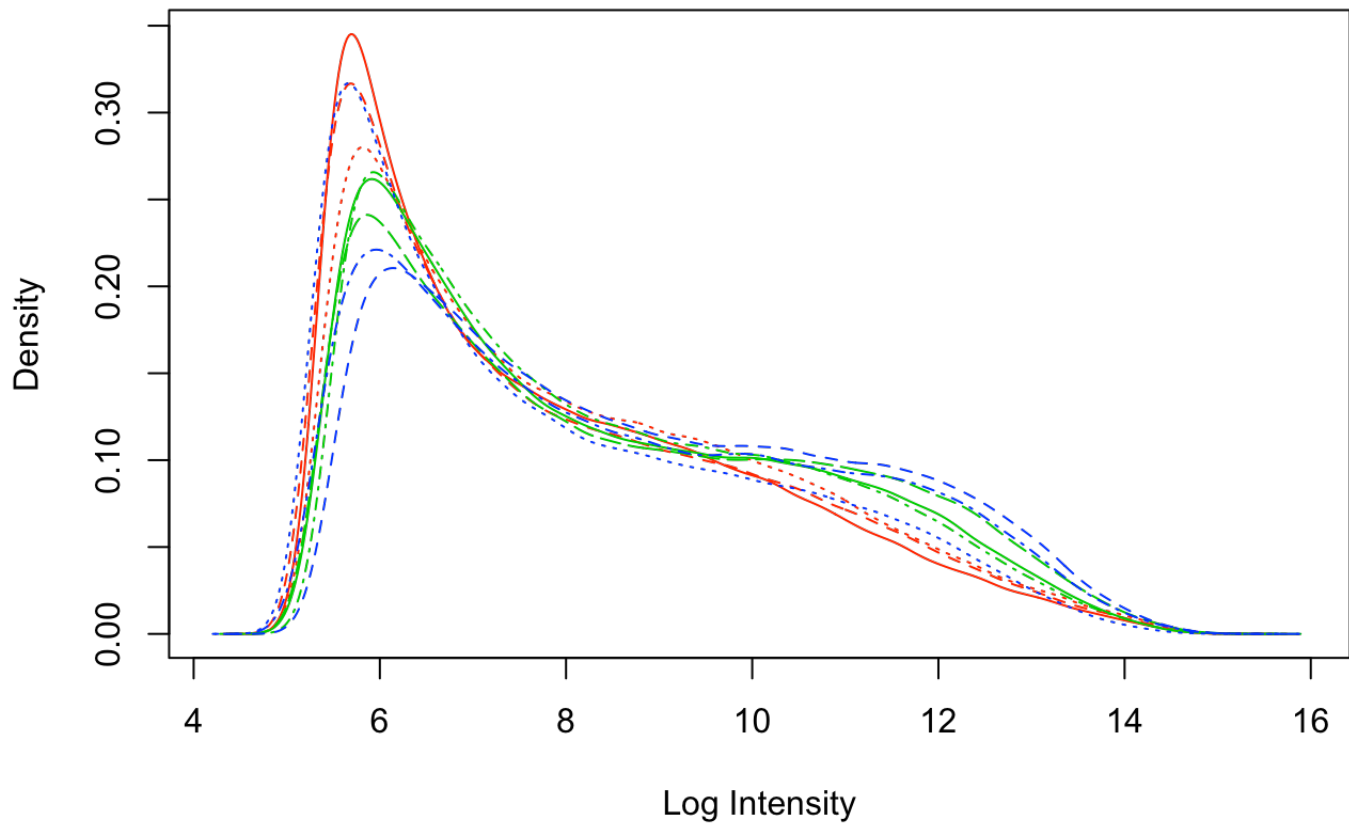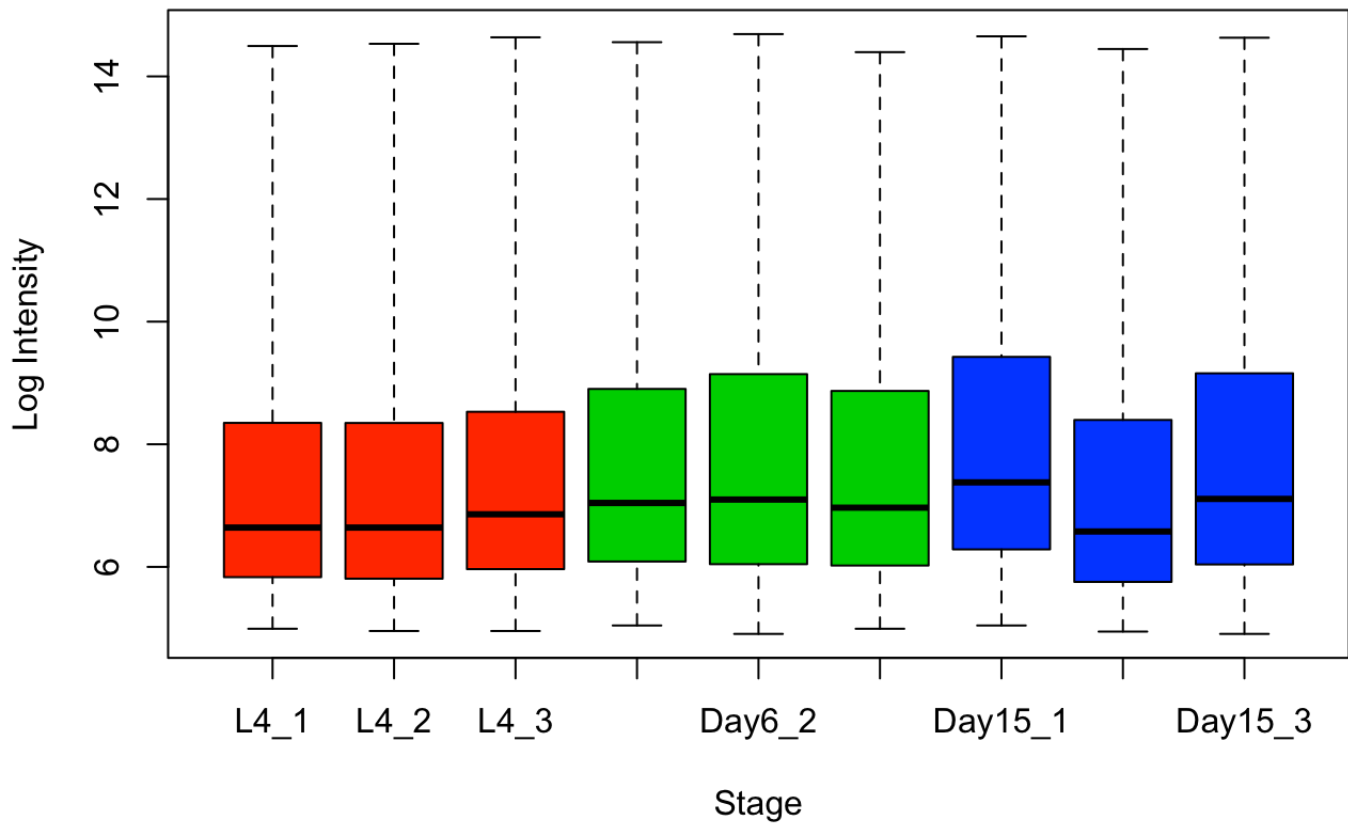
L4 vs Day 15

Histogram of log intensity

```
hist(expd[,1:9],col=c(rep(2,3),rep(3,3),rep(4,3)), ylab = "Density", xlab = "Log Intensity"
)
```
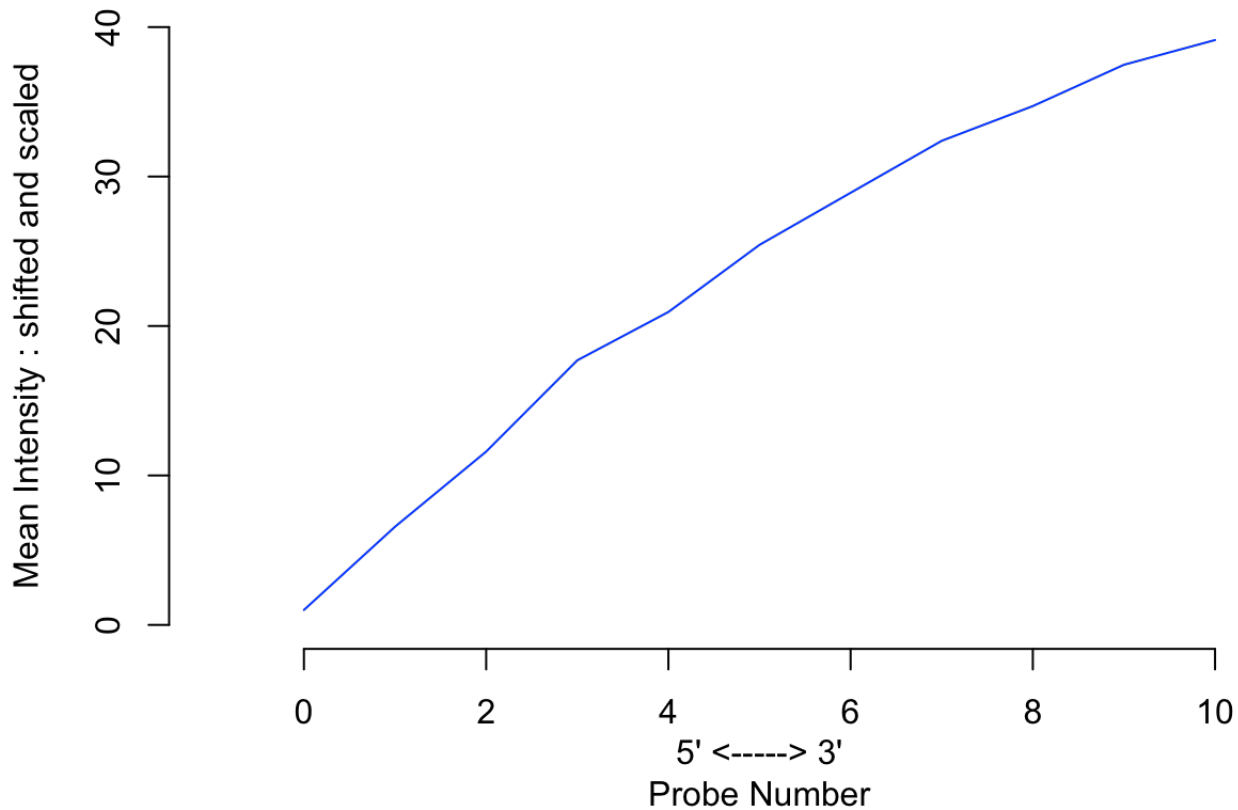
Boxplots of replicates

```
boxplot(expd,col=c(rep(2,3),rep(3,3),rep(4,3)), xlab = "Stage", ylab = "Log Intensity")
```

Identification of possible RNA degradation at different ends of the mRNA transcript

```
deg <- AffyRNAdeg(expd[,1])
plotAffyRNAdeg(deg)
```
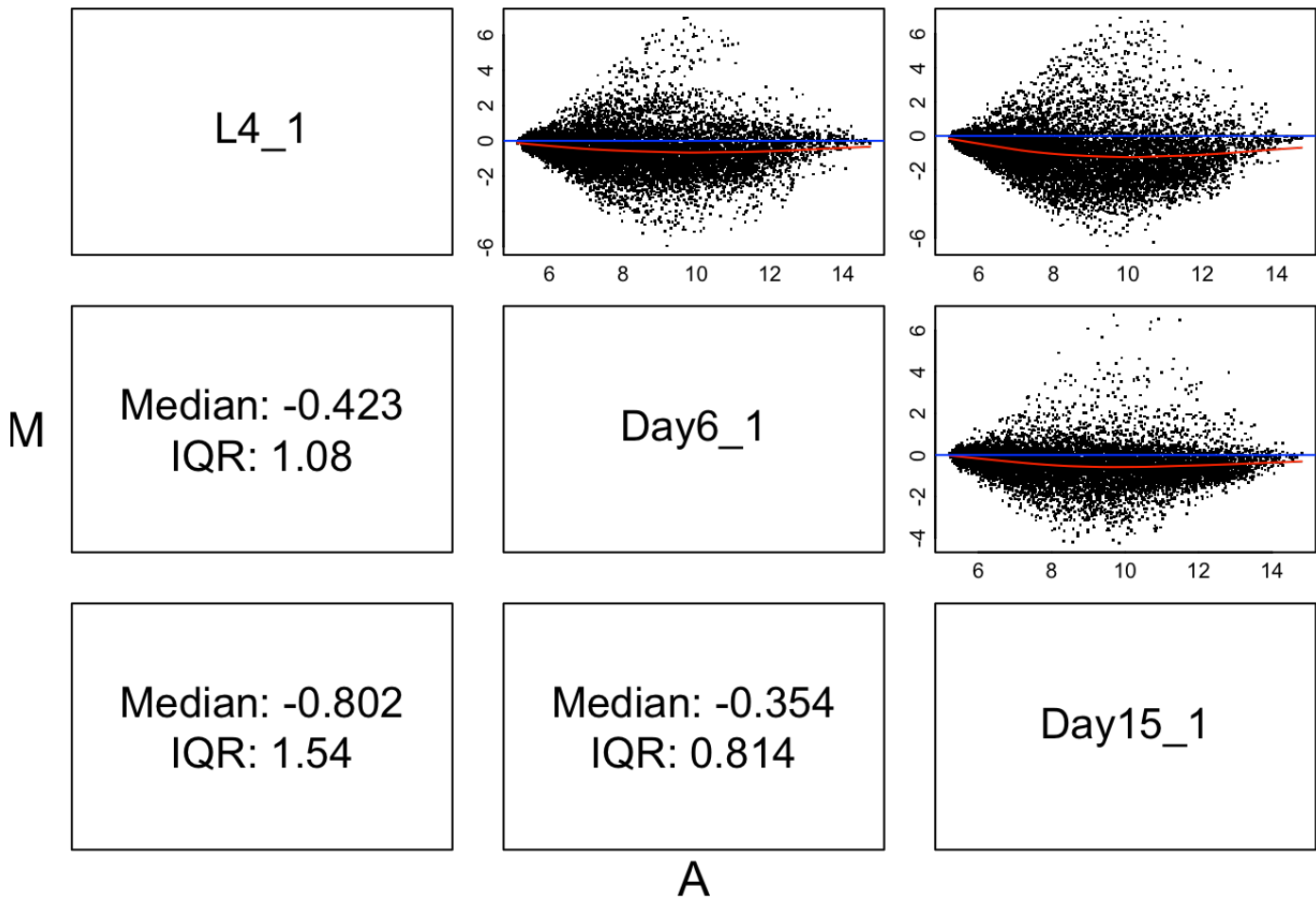
RNA degradation plot

MVA plot of different day measurements

```
gn <- sample(geneNames(expd),1000)
pms <- pm(expd[,c(1,4,7)],gn)
mva.pairs(pms)
```

file:///Users/stefanzdraljevic/Northwestern/Courses/Stats_for_Bioinformatics/Homework/HW2/GSE21784_RAW/HW2_Stefan_Zdraljevic.html

Page 7 of 11

MVA plot

Normalization of data

```
normalized.expd <- merge(normalize(expd[,1:3],method="loess"),
                         normalize(expd[,4:6],method="loess"))
```

```
## Done with 1 vs 2 in iteration 1
## Done with 1 vs 3 in iteration 1
## Done with 2 vs 3 in iteration 1
## 1 0.009579
## Done with 1 vs 2 in iteration 1
## Done with 1 vs 3 in iteration 1
## Done with 2 vs 3 in iteration 1
## 1 0.01199
```

```
normalized.expd <- merge(normalized.expd,
                         normalize(expd[,7:9],method="loess"))
```

```
## Done with 1 vs 2 in iteration 1
## Done with 1 vs 3 in iteration 1
## Done with 2 vs 3 in iteration 1
## 1 0.1492
```

```
boxplot(normalized.expd,col=c(rep(2,3),rep(3,3),rep(4,3)), main = "Normalized Expression Da
ta", xlab = "Stage", ylab = "Log Expression")
```
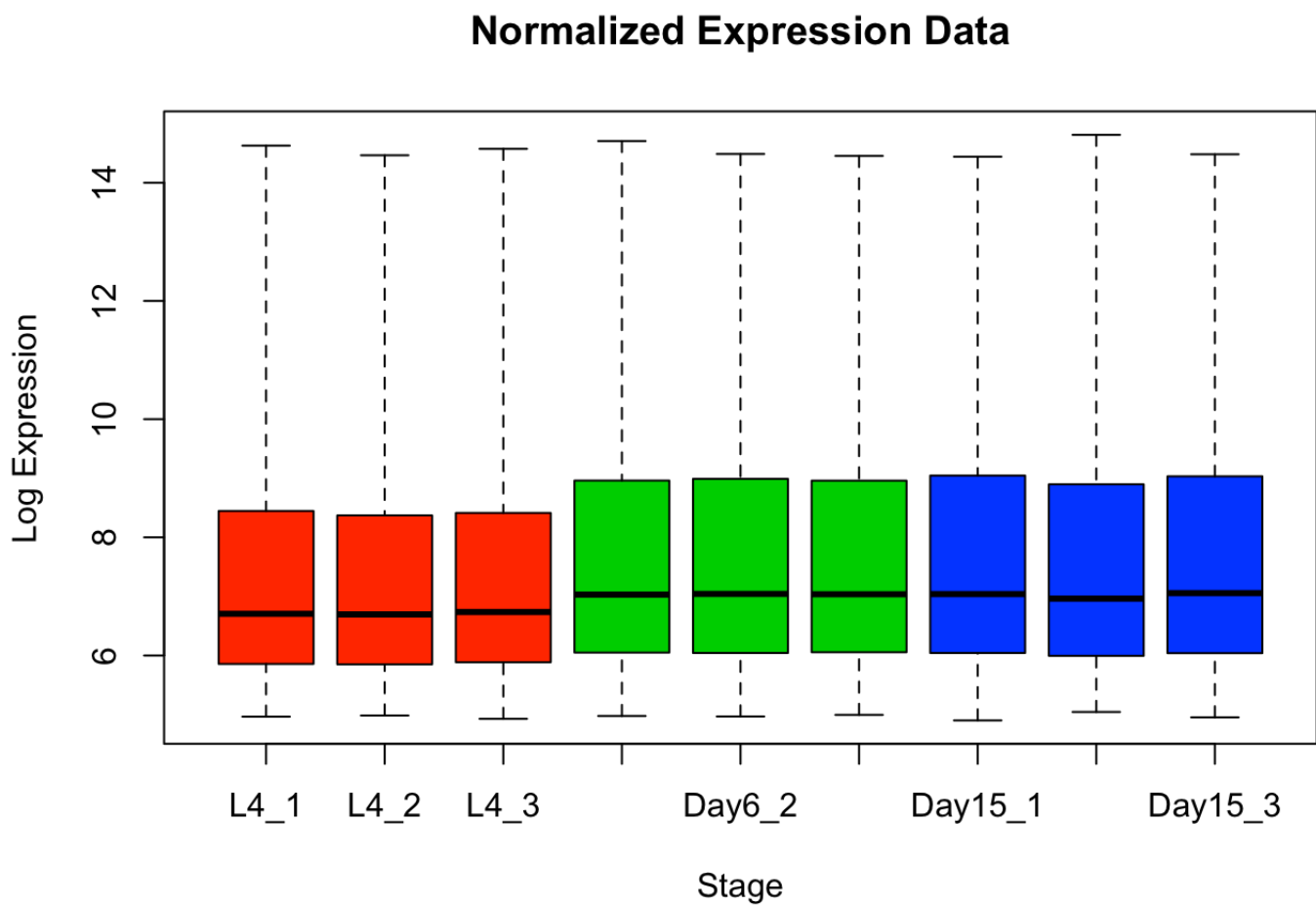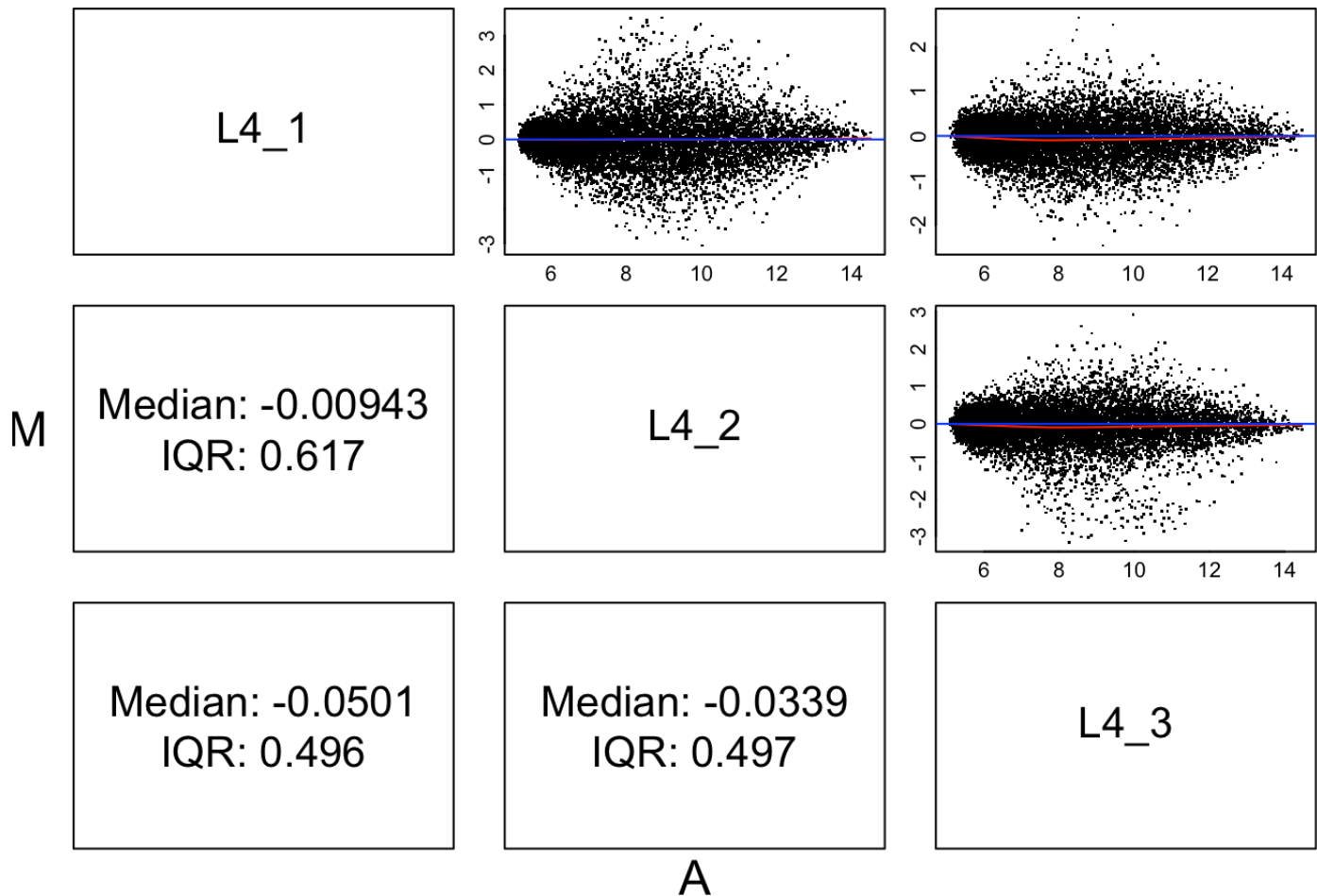

Normalized MVA compared to un-normalized

```
gn <- sample(geneNames(normalized.expd),1000)


pms <- pm(normalized.expd[,c(1,2,3)],gn)
mva.pairs(pms, main = "Normalzed MVA Plot")
```

Normalzed MVA Plot


```
pms <- pm(expd[,c(1,2,3)],gn)
mva.pairs(pms, main = "unNormalized MVA Plot")
```

unNormalized MVA Plot

M

Median: -0.0418
IQR: 0.629

L4_2

Median: -0.207
IQR: 0.52

Median: -0.146
IQR: 0.511

L4_3

A