

## HW #2

1.

```
dir.path <- "C:\\Users\\thomas\\Documents\\Data"
cy <- read.GenePix(path=dir.path, name.Gf = "F532 Median", name.Gb = "B532 Median", name.Rf = "F635
Median", name.Rb = "B635 Median", name.W = "Flags")
```

2.

```
printtip1 <- maNorm(cy[,1], norm = "p", span=0.45)
printtip2 <- maNorm(cy[,2], norm = "p", span=0.45)
printtip3 <- maNorm(cy[,3], norm = "p", span=0.45)
printtip4 <- maNorm(cy[,4], norm = "p", span=0.45)

median1 <- maNorm(cy[,1], norm="median", span=0.45)
median2 <- maNorm(cy[,2], norm="median", span=0.45)
median3 <- maNorm(cy[,3], norm="median", span=0.45)
median4 <- maNorm(cy[,4], norm="median", span=0.45)

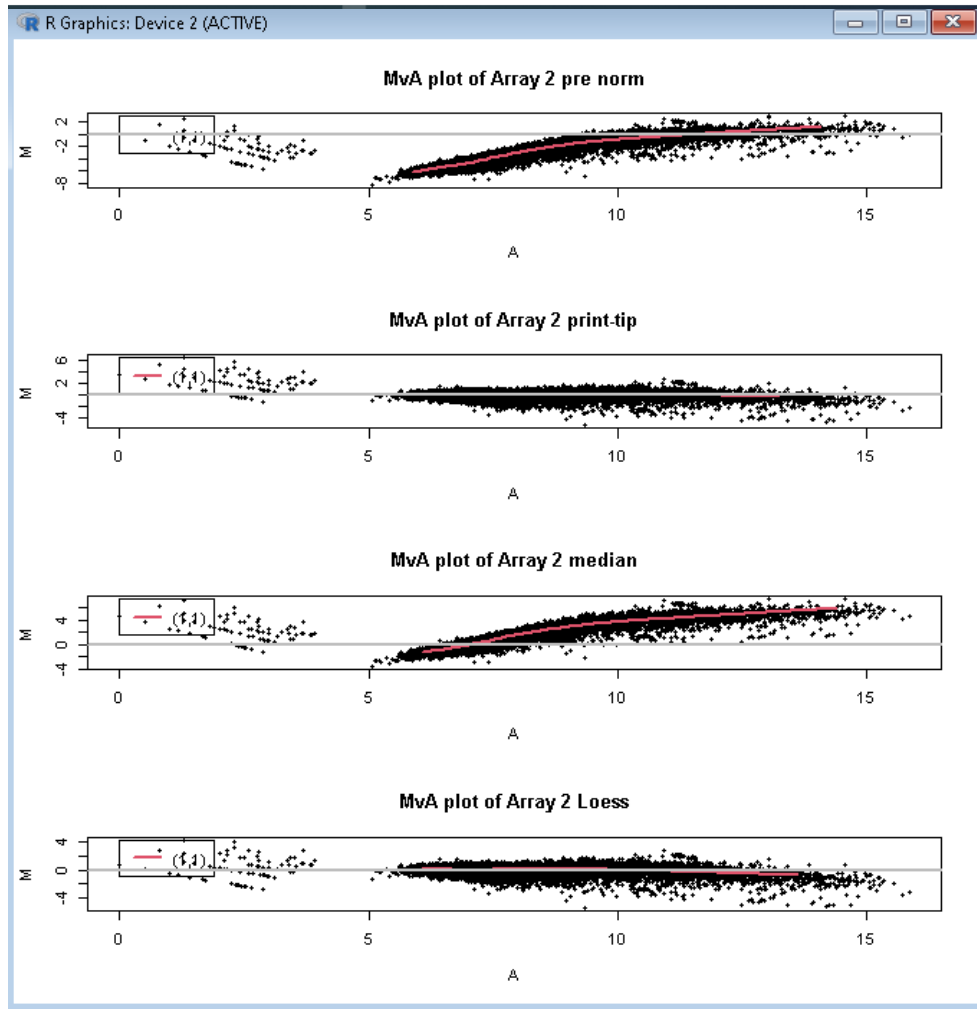
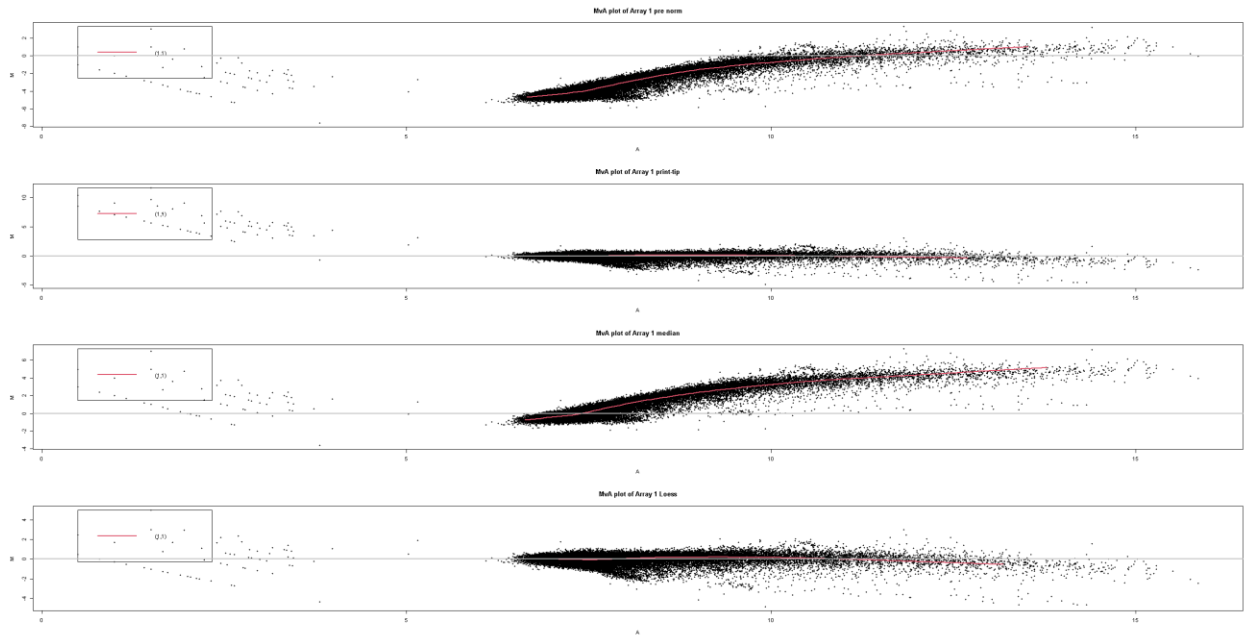
loess1 <- maNorm(cy[,1], norm="loess", span=0.45)
loess2 <- maNorm(cy[,2], norm="loess", span=0.45)
loess3 <- maNorm(cy[,3], norm="loess", span=0.45)
loess4 <- maNorm(cy[,4], norm="loess", span=0.45)

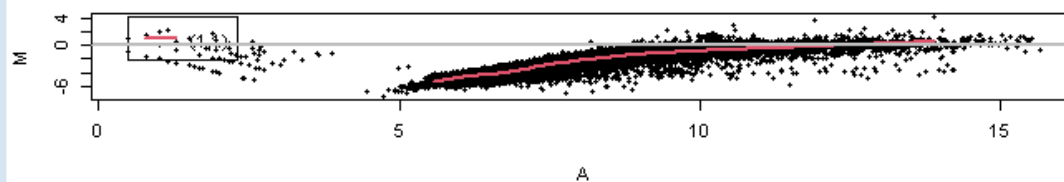
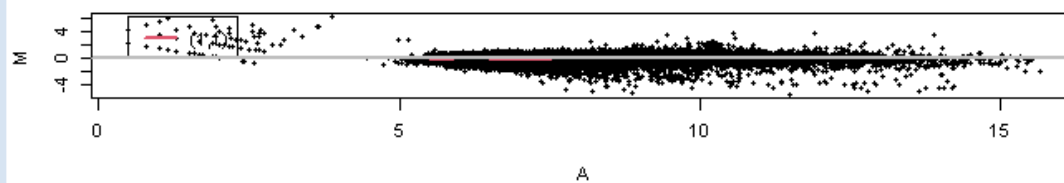
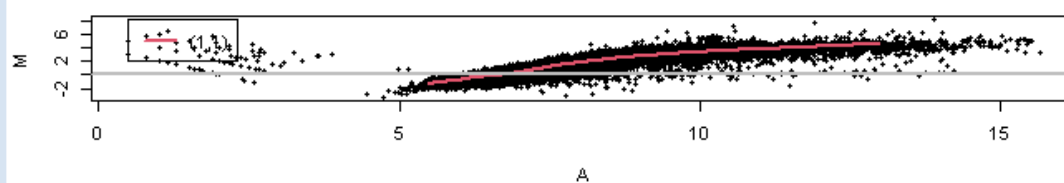
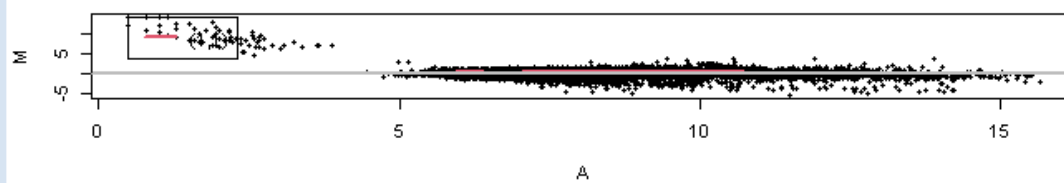
par(mfrow=c(4,1))
maPlot(cy[,1], main = "MvA plot of Array 1 pre norm")
maPlot(printtip1, main = "MvA plot of Array 1 print-tip")
maPlot(median1, main = "MvA plot of Array 1 median")
maPlot(loess1, main = "MvA plot of Array 1 Loess")

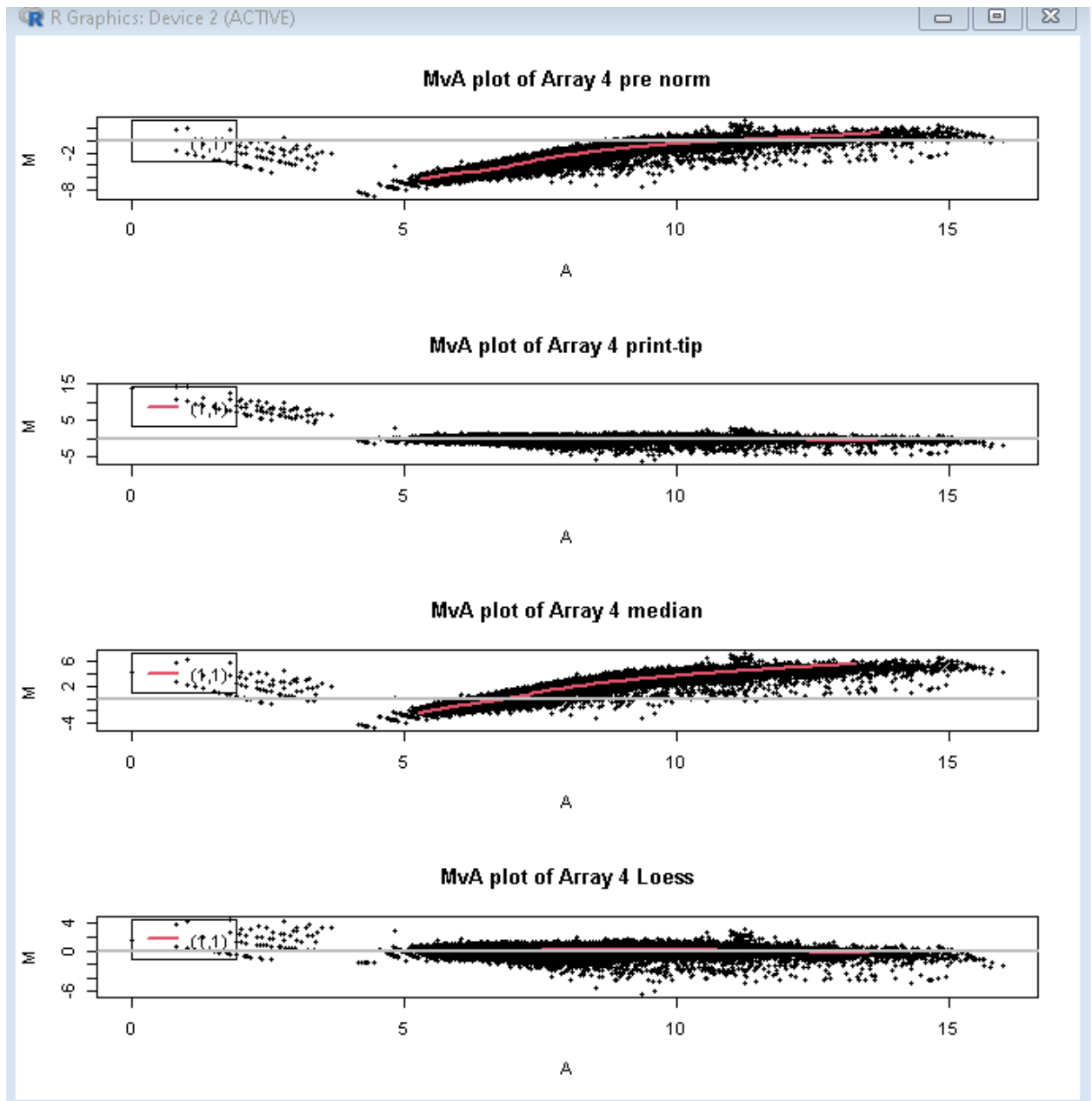
par(mfrow=c(4,1))
maPlot(cy[,2], main = "MvA plot of Array 2 pre norm")
maPlot(printtip2, main = "MvA plot of Array 2 print-tip")
maPlot(median2, main = "MvA plot of Array 2 median")
maPlot(loess2, main = "MvA plot of Array 2 Loess")

par(mfrow=c(4,1))
maPlot(cy[,3], main = "MvA plot of Array 3 pre norm")
maPlot(printtip3, main = "MvA plot of Array 3 print-tip")
maPlot(median3, main = "MvA plot of Array 3 median")
maPlot(loess3, main = "MvA plot of Array 3 Loess")

par(mfrow=c(4,1))
maPlot(cy[,4], main = "MvA plot of Array 4 pre norm")
maPlot(printtip4, main = "MvA plot of Array 4 print-tip")
maPlot(median4, main = "MvA plot of Array 4 median")
maPlot(loess4, main = "MvA plot of Array 4 Loess")
```



**MvA plot of Array 3 pre norm****MvA plot of Array 3 print-tip****MvA plot of Array 3 median****MvA plot of Array 3 Loess**



```
3. logratio4 <- maM(cy[,4])
   ar4<-na.omit(logratio4)
   median4log <- na.omit(maM(median4))
   printtip4log <- na.omit(maM(printtip4))
   loess4log <- na.omit(maM(loess4))
```

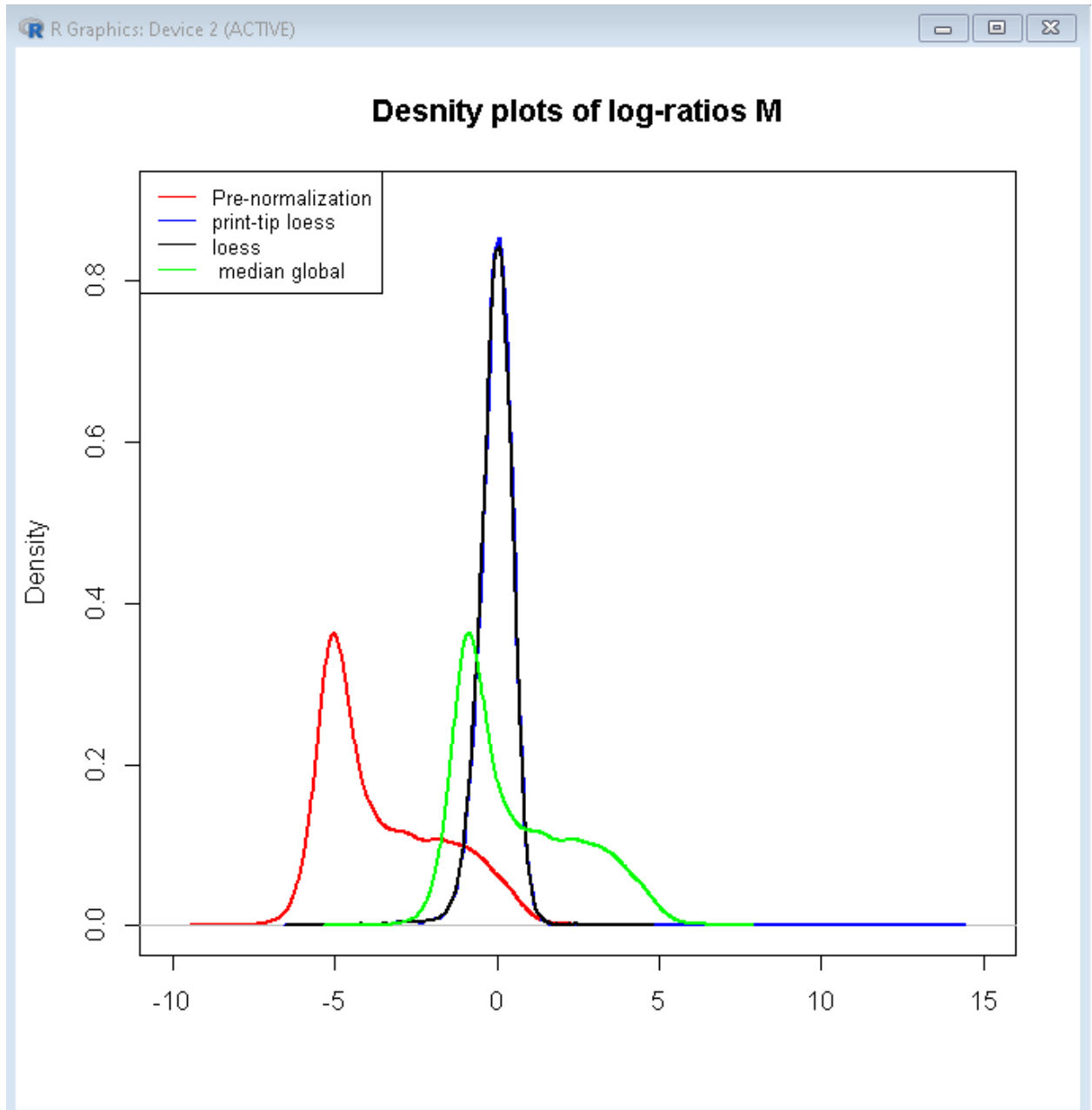
```
plot(density(ar4), lwd = 2, xlab = "", col = "red", xlim = c(-10, 15), ylim = c(0, 0.9), main = "Density plots of
log-ratios M")
```

```
lines(density(printtip4log), col = "blue", lwd = 2)
```

```
lines(density(loess4log), col = "black", lwd = 2)
```

```
lines(density(median4log), col = "green", lwd = 2)
```

```
legend("topleft", legend = c("Pre-normalization", "print-tip loess", "loess", "median global"),
col=c("red", "blue", "black", "green"), lty=1, cex=0.8)
```



4. Looking through all the plots it seems like print-tip loess is the most preferred for this data set. Loess and print tip loess seem to produce almost similar normalized results, but I believe print-tip loess shows to be the better out of the two.

5.
 

```
for(i in 1:4){
  name <- paste("sample", i, sep = ".")
  bg <- maRb(cy[,i])
  fg <- maRf(cy[,i])
  diff <- fg - bg
  diff[diff < 0] <- NA
  assign(name, log2(diff))
}
```

```

#Data separated for each array:
data.median1 <- apply(sample.1, 2, median, na.rm = T)
data.median2 <- apply(sample.2, 2, median, na.rm = T)
data.median3 <- apply(sample.3, 2, median, na.rm = T)
data.median4 <- apply(sample.4, 2, median, na.rm = T)

data.norm1 <- sweep(sample.1, 2, data.median1)
data.norm2 <- sweep(sample.2, 2, data.median2)
data.norm3 <- sweep(sample.3, 2, data.median3)
data.norm4 <- sweep(sample.4, 2, data.median4)

#matrix for all 4 arrays together:
data.prenorm <- cbind(sample.1, sample.2, sample.3, sample.4)
data.median <- apply(data.prenorm, 2, median, na.rm = T)
data.norm <- sweep(data.prenorm, 2, data.median)

colnames(data.norm) <- c("Array 1", "Array 2", "Array 3", "Array 4")

```

6.

```

yL1<- na.omit(data.loess2[,1])
yL2<- na.omit(data.loess2[,2])
yL3<- na.omit(data.loess2[,3])
yL4<- na.omit(data.loess2[,4])

xL1<- na.omit(data.loess2[,1])
xL2<- na.omit(data.loess2[,2])
xL3<- na.omit(data.loess2[,3])
xL4<- na.omit(data.loess2[,4])

corr1 <- cor.test(x=xL1, y=yL2, method ="spearman",exact=FALSE)
corr2 <- cor.test(x=xL1, y=yL3, method ="spearman",exact=FALSE)
corr3 <- cor.test(x=xL1, y=yL4, method ="spearman",exact=FALSE)
corr4<- cor.test(x=xL2, y=yL3, method ="spearman",exact=FALSE)
corr5 <- cor.test(x=xL2, y=yL4, method ="spearman",exact=FALSE)
corr6 <- cor.test(x=xL3, y=yL4, method ="spearman",exact=FALSE)

Spearman's rank correlation rho

data:  xL1 and yL2
S = 4.3032e+12, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.6952404

Spearman's rank correlation rho

data:  xL1 and yL3
S = 3.3213e+12, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to
sample estimates:
      rho
0.7647799

```

Spearman's rank correlation rho

```
data:  xL1 and yL4
S = 4.1802e+12, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.7039508
```

Spearman's rank correlation rho

```
data:  xL2 and yL3
S = 3.9045e+12, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.723478
```

Spearman's rank correlation rho

```
data:  xL2 and yL4
S = 4.0887e+12, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.7104355
```

Spearman's rank correlation rho

```
data:  xL3 and yL4
S = 3.4476e+12, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.7558382
```

```
y1 <- data.norm[,1]
y2 <- data.norm[,2]
y3 <- data.norm[,3]
y4 <- data.norm[,4]
```

```
x1 <- data.norm[,1]
x2 <- data.norm[,2]
x3 <- data.norm[,3]
x4 <- data.norm[,4]
```

```
corr <- cor.test(x=x2, y=y2, method="spearman",exact=FALSE)
```

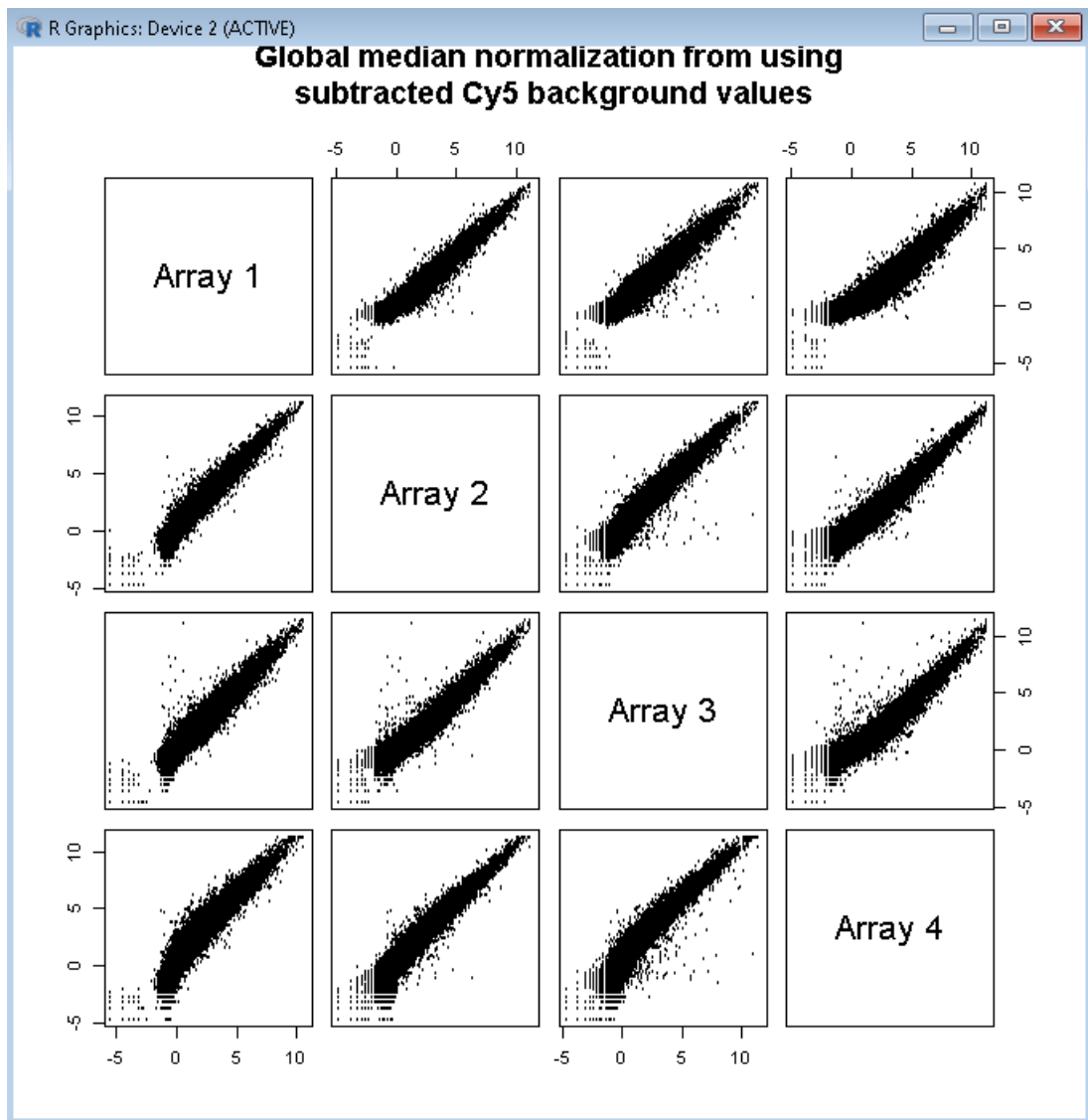
```
cor(data.frame(na.omit(data.norm)), method = "spearman")
      Array.1   Array.2   Array.3   Array.4
Array.1 1.0000000 0.8957946 0.8784760 0.8985979
Array.2 0.8957946 1.0000000 0.8758774 0.9075121
Array.3 0.8784760 0.8758774 1.0000000 0.8848059
Array.4 0.8985979 0.9075121 0.8848059 1.0000000
```

```
data.loess <- cbind(loess1, loess2, loess3, loess4)
data.loess2 <- data.matrix(data.loess)
colnames(data.loess2) <- c("Array 1", "Array 2", "Array 3", "Array 4")
cor(data.frame(data.loess2), method = "spearman")
```

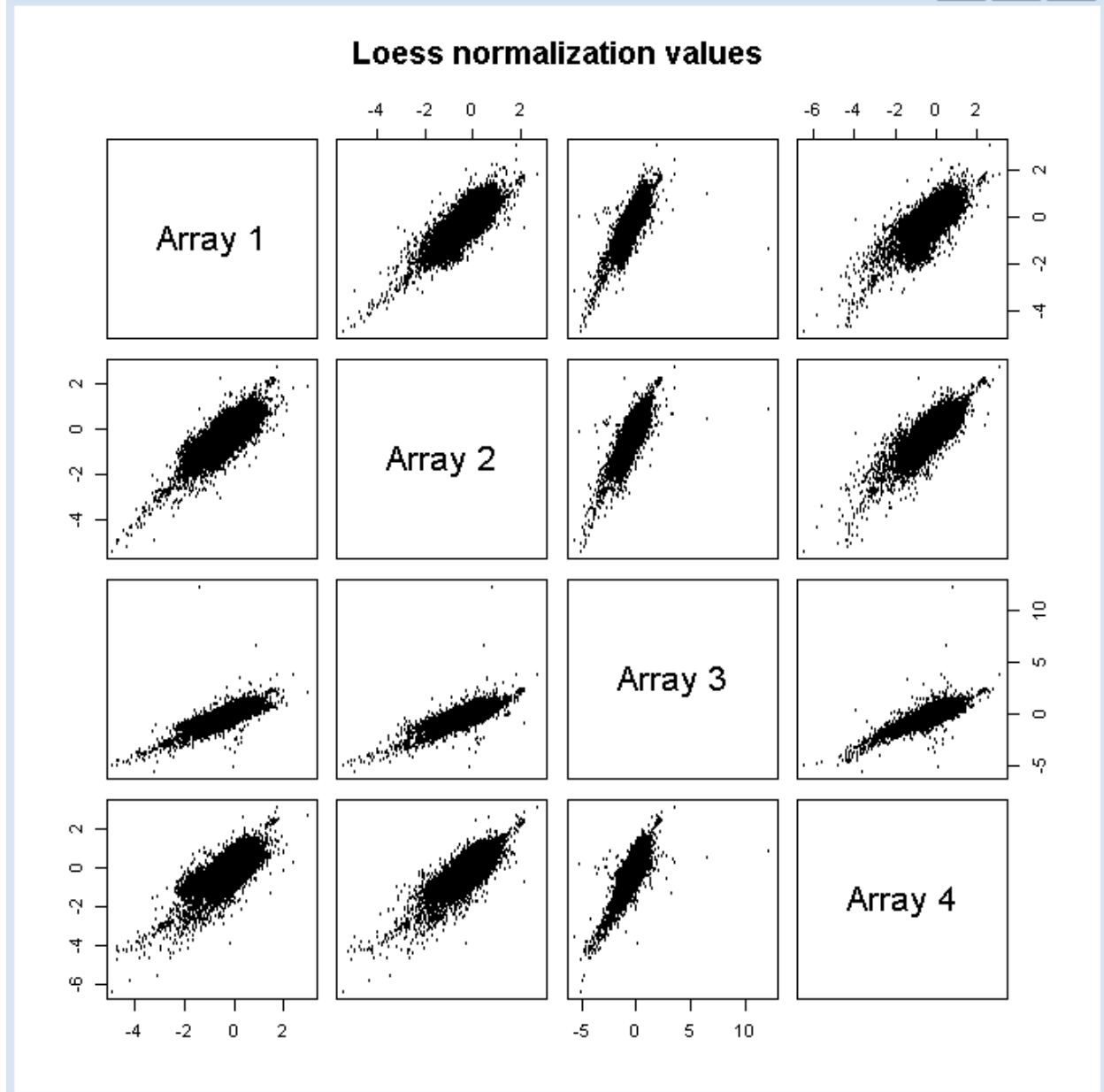
```
      Array.1   Array.2   Array.3   Array.4
Array.1 1.0000000 0.6952404 0.7647799 0.7039508
Array.2 0.6952404 1.0000000 0.7234780 0.7104355
Array.3 0.7647799 0.7234780 1.0000000 0.7558382
Array.4 0.7039508 0.7104355 0.7558382 1.0000000
```

```
pairs(data.norm, pch=21,col=1,main="Global median normalization from using\n subtracted Cy5 background values",cex=0.4)
```





```
pairs(data.loess2, pch=21,col=1,main="Loess normalization values",cex=0.4)
```



7.

```

for(i in 1:4){
  name <- paste("samp", i, sep = ".")
  bg <- maRb(cy[,i])
  fg <- maRf(cy[,i])
  diff <- fg - bg
  assign(name, diff)
}
data.prelog <- cbind(samp.1, samp.2, samp.3, samp.4)
colnames(data.prelog) <- c("Array 1", "Array 2", "Array 3", "Array 4")

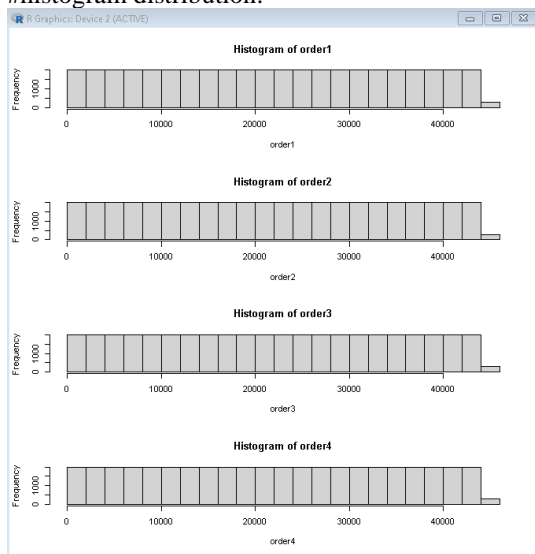
sorted <- apply(data.prelog, 2, sort)
meanValues1 <- rowMeans(data.frame(sorted[,1]),na.rm=T)
meanValues2 <- rowMeans(data.frame(sorted[,2]),na.rm=T)

```

```
meanValues3 <- rowMeans(data.frame(sorted[,3]),na.rm=T)
meanValues4 <- rowMeans(data.frame(sorted[,4]),na.rm=T)
data.meanValues <- cbind(meanValues1, meanValues2, meanValues3, meanValues4)
```

```
prelog1<- rank(data.prelog[,1], ties="first")
prelog2<- rank(data.prelog[,2], ties="first")
prelog3<- rank(data.prelog[,3], ties="first")
prelog4<- rank(data.prelog[,4], ties="first")
```

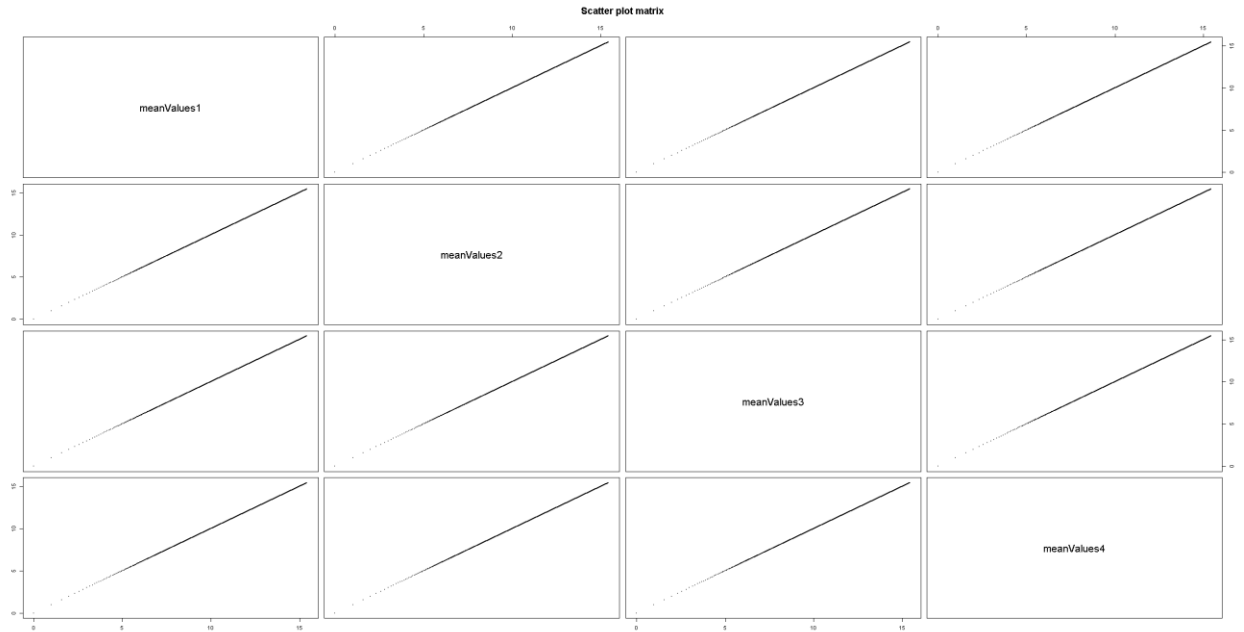
```
order1<- rank(data.meanValues[,1], ties="first")
order2 <- rank(data.meanValues[,2], ties="first")
order3 <- rank(data.meanValues[,3], ties="first")
order4 <- rank(data.meanValues[,4], ties="first")
data.order <- cbind(order1, order2, order3, order4)
#histogram distribution:
```



```
colnames(data.order) <- c("order1", "order2", "order 3", "order 4")
```

```
8. > orderLog <- log2(data.order)
> cor(orderLog, method="spearman")
      order1 order2 order 3 order 4
order1      1      1      1      1
order2      1      1      1      1
order 3      1      1      1      1
order 4      1      1      1      1
```

```
pairs(orderLog, pch=21,col=1,main="Scatter plot matrix",cex=0.4)
```



9. From looking at the data the normalization method that was the best is the quantile normalization method because when you see the spearman correlation numbers the datas are all normalized to each other. They are all at the value of 1. It means a perfect association between the ranks.

10.

```
f.parse <- function(path=getwd(), file="Inflammation_qRT-PCR.csv",out=out.fi) {
  d <- read.table(paste(path,file,sep=""),skip=11,sep="," ,header=T)
  u <- as.character(unique(d$Name))
  u <- u[u!=""]; u <- u[!is.na(u)];
  ref <- unique(as.character(d$Name[d$Type=="Reference"]))
  u <- unique(c(ref,u))
  p <- unique(toupper(as.character(d$Name.1)))
  p <- sort(setdiff(p,c("")))
  mat <- matrix(0,nrow=length(u),ncol=length(p))
  dimnames(mat) <- list(u,p)
  for (i in 1:length(u)) {
    print(paste(i," ",u[i],sep=""))
    tmp <- d[d$Name %in% u[i],c(1:3,6,9)]
    g <- toupper(unique(as.character(tmp$Name.1)))
    g <- sort(setdiff(g,c(" ",hg)))
    for (j in 1:length(g)) {
      v <- tmp[toupper(as.character(tmp$Name.1)) %in% g[j],5]
      v <- v[v!=999]
      v <- v[(v/mean(v))<1.5 & ((v/mean(v))>0.67)] #gene j vector
      hv3 <- NULL
      for (k in 1:length(hg)) { #housekeeping gene vector (each filtered by reps)
        hv <- tmp[toupper(as.character(tmp$Name.1)) %in% hg[k],5]
        hv <- hv[hv!=999]
        hv3 <- c(hv3,hv[(hv/mean(hv))<1.5 & ((hv/mean(hv))>0.67)])
      }
      sv <- mean(as.numeric(v)) - mean(as.numeric(hv3)) #scaled value for gene j
      if(i==1) { #reference sample only
        mat[u[i],g[j]] <- sv
        next
      }
      mat[u[i],g[j]] <- sv - mat[u[1],g[j]]
    }
  }
  mat[1,][!is.na(mat[1,])] <- 0
  fc <- 2^(-1 * mat)
  write.table(t(c("Subject",dimnames(mat)[[2]])),paste(path,out,sep=""),quote=F,sep="t",col.names=F,row.names
  =F)
  write.table(round(fc,3),paste(path,out,sep=""),quote=F,sep="t",append=T,col.names=F)
}
# run function
pa <- "C:/Users/thomas/Documents/Inflammation_"
```

```
patients<-read.delim("Inflammation_fold_chg_matrix.txt", header=T, row.names=1)
```

data matrix of fold changes created in file Inflammation\_fold\_chg\_matrix.txt:

```
spearman<-cor(data.frame(patients),method="spearman")
```

```
11. x1<-patients[1,]  
    x3<-patients[3,]  
    x2<-patients[2,]  
    x4<-patients[4,]  
    x5<-patients[5,]
```

```

x6<-patients[6,]
x7<-patients[7,]
x8<-patients[8,]
x9<-patients[9,]
x10<-patients[10,]
x11<-patients[11,]
x12<-patients[12,]
x13<-patients[13,]
x14<-patients[14,]
x15<-patients[15,]
cor.test(x=as.numeric(x5), y=as.numeric(x15), method="spearman")
> cor.test(x=as.numeric(x5), y=as.numeric(x15), method="spearman")

```

Spearman's rank correlation rho

```

data:  as.numeric(x5) and as.numeric(x15)
S = 124, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.9694581

```

So the patients 434\_3 and patient 434\_8 were shown to be the most correlated.  
 plot(patient, main="Scatter plot for 434\_3 vs 434\_8")

**Scatter plot for 434\_3 vs 434\_8**