# Notes in **BLUE** are comments

# Notes in **BLACK** are commands

# Notes in **RED** could or should be adjusted to suit any re-analysis

# Notes in **GREEN** are results from the program

# Data input

# load the edgeR library

library(edgeR)

library(lumi)

# set the working directory (for convenience), note reverse direction of slashes for windows

setwd("C:/Users/nicoleC/Desktop/Michelle Hill miRNA 210715")

# read in the raw data

rawdata <- read.delim("miRNA\_raw\_counts.txt")

# create a DGEList object. In this case, raw data is in columns 2 to 55, and unique identifiers are in 1 (miRNA name)

y <- DGEList(counts=rawdata[,2:55], genes=rawdata[,1])

# Filtering and Normalization

# compute the effective library size by using TMM normalization

keep <- rowSums(cpm(y)>10) >=2

y <- y[keep,]

y$samples$lib.size <- colSums(y$counts)

y <- calcNormFactors(y)

#make QC plots

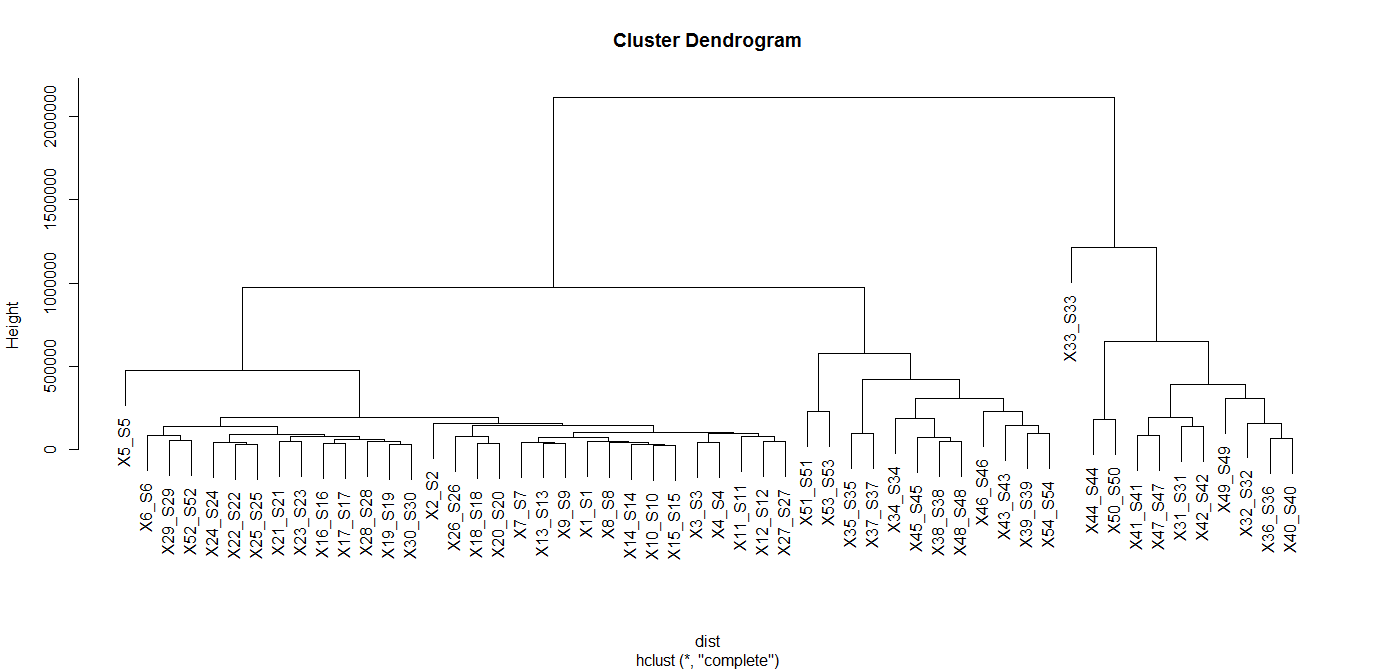
d <-cpm(y, normalized.lib.sizes=TRUE)

d<-t(d)

dist<-dist(d)

hc<-hclust(dist)

plot(hc)



# Export the graph to a PDF

pdf(file="dendrogram.pdf", paper="a4r")

plot(hc)

dev.off()

# An MDS plot shows the biological coefficient of variation between the samples. The two

# dimensions are the biggest and second biggest sources of variation within the data. Again, this plot

# seems to suggest that the biggest differentiators of these samples are not related to the properties of

# the genes tested.

d <-t(d)

plotMDS(d, col=c(rep(1,15), rep(2, 15), rep(1, 12), rep(2, 12)))

legend("topright", legend = c("Exosome", "Pellet"), col = 1:2, pch = 15)plotMDS(d, col=c(rep(1,30), rep(2,24)))

legend("topright", legend = c("HEK", "PC3"), col = 1:2, pch = 15)

# Export the graph to a PDF

pdf(file="MDS.pdf", paper="a4r")

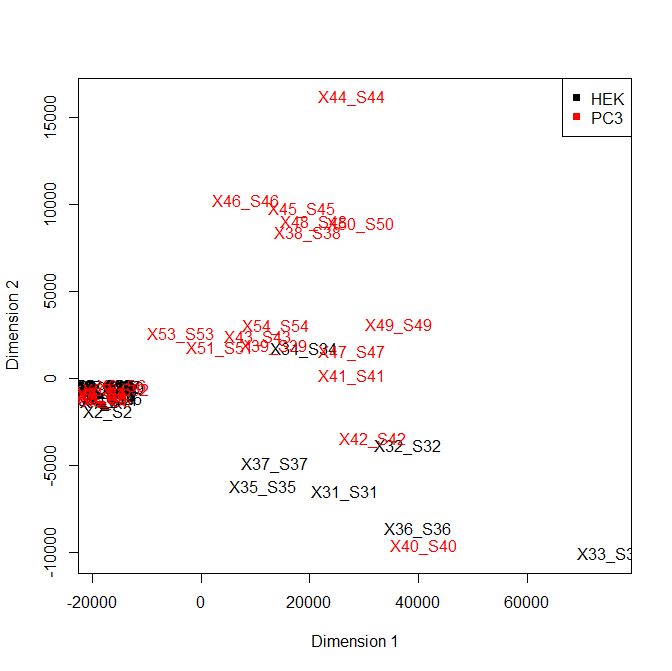
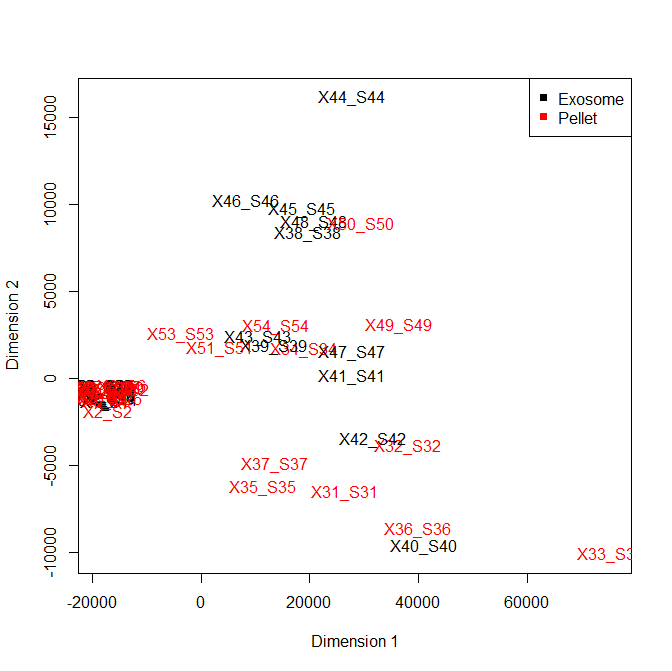
plotMDS(d, col=c(rep(1,15), rep(2, 15), rep(1, 12), rep(2, 12)))

legend("topright", legend = c("Exosome", "Pellet"), col = 1:2, pch = 15)

plotMDS(d, col=c(rep(1,30), rep(2,24)))

legend("topright", legend = c("HEK", "PC3"), col = 1:2, pch = 15)

dev.off()



#First questions. Start with the cellular miRNA, hopefully can find the answers. A. What are the miRNAs altered by

#expression of each cavin? (compared to GFP control) B. Can they explain the observed reduction in migration? C.

#Common versus different miRNAs/action pathways between the 3 cavins.

#To make life easier, load just a subset of the data.

rawdata <- read.delim("miRNA\_raw\_counts\_PC3\_Pellet\_Subset.txt")

# create a DGEList object. In this case, raw data is in columns 2 to 55, and unique identifiers are in 1 (miRNA name)

y <- DGEList(counts=rawdata[,2:13], genes=rawdata[,1])

# Filtering and Normalization

# compute the effective library size by using TMM normalization

keep <- rowSums(cpm(y)>10) >=2

y <- y[keep,]

y$samples$lib.size <- colSums(y$counts)

y <- calcNormFactors(y)

#make QC plots

d <-cpm(y, normalized.lib.sizes=TRUE)

d<-t(d)

dist<-dist(d)

hc<-hclust(dist)

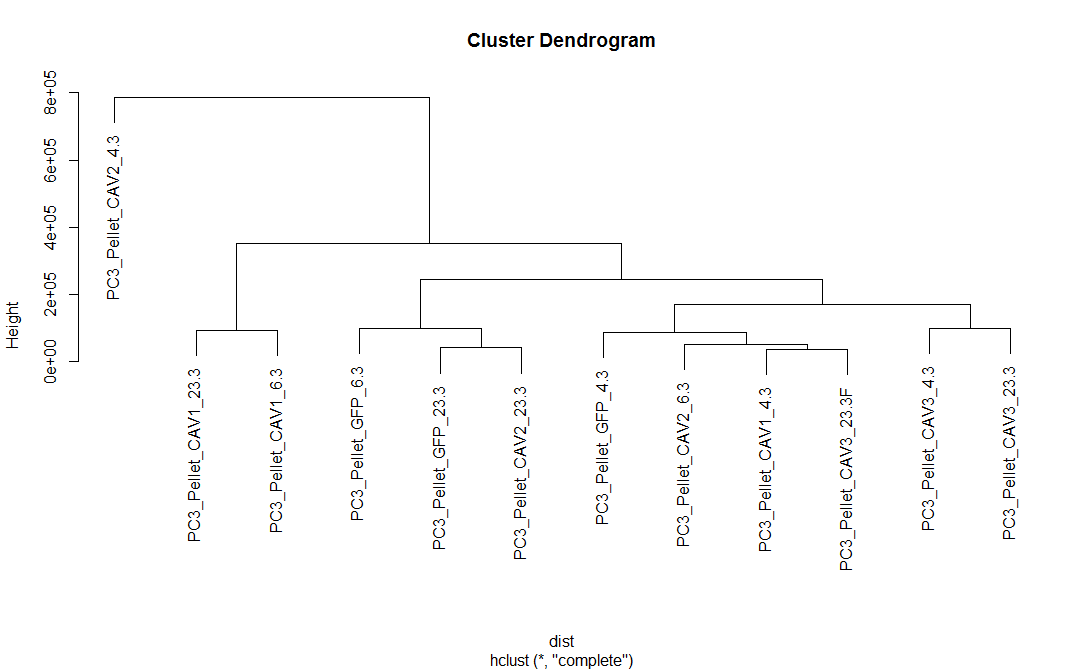
plot(hc)

#export normalized data into tab delimited text for exploration in genespring

cpm<- cpm(y, normalized.lib.sizes=TRUE)

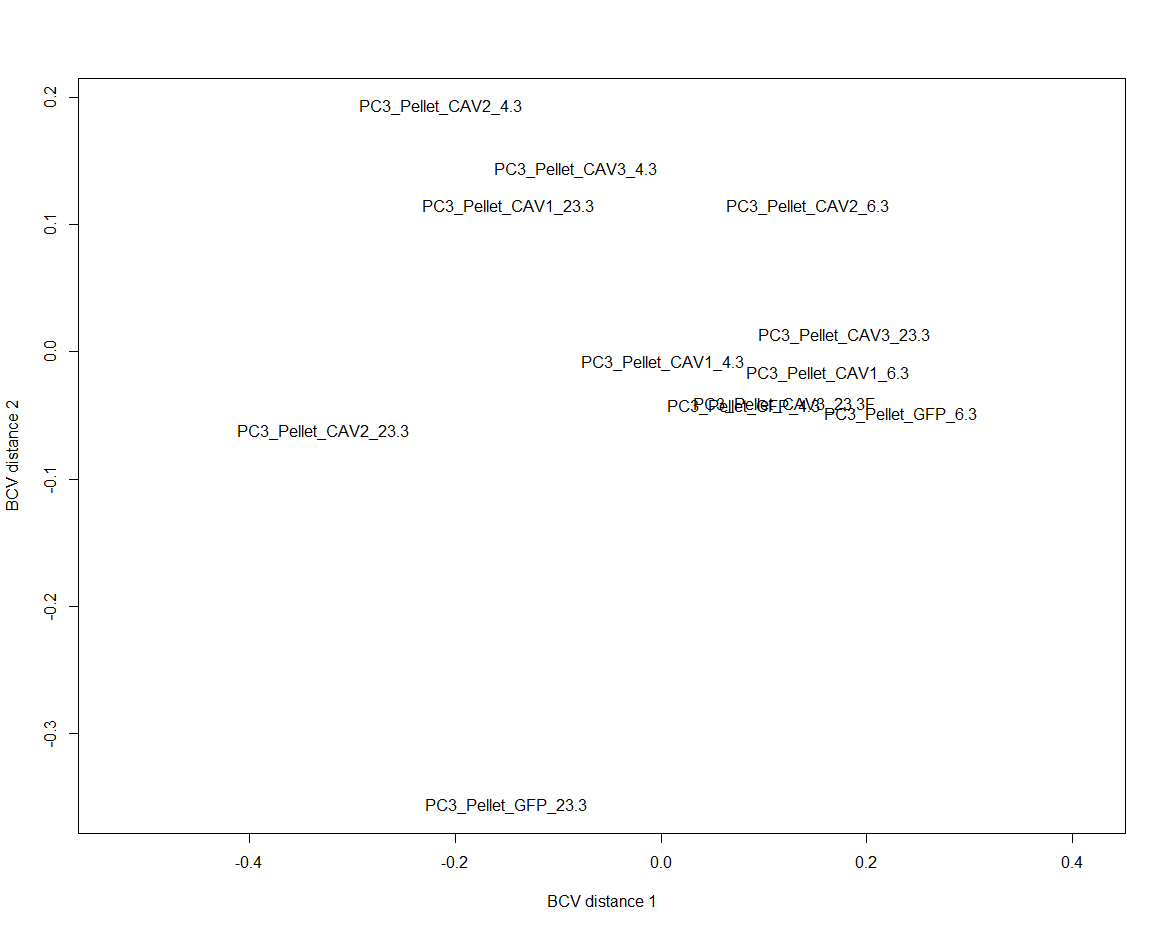
rownames(cpm)<-y$genes[,1]

write.table(cpm, "cpm\_miRNA\_PC3\_subset.txt", sep="\t")



#plot the MDS

plotMDS(y, method="bcv")

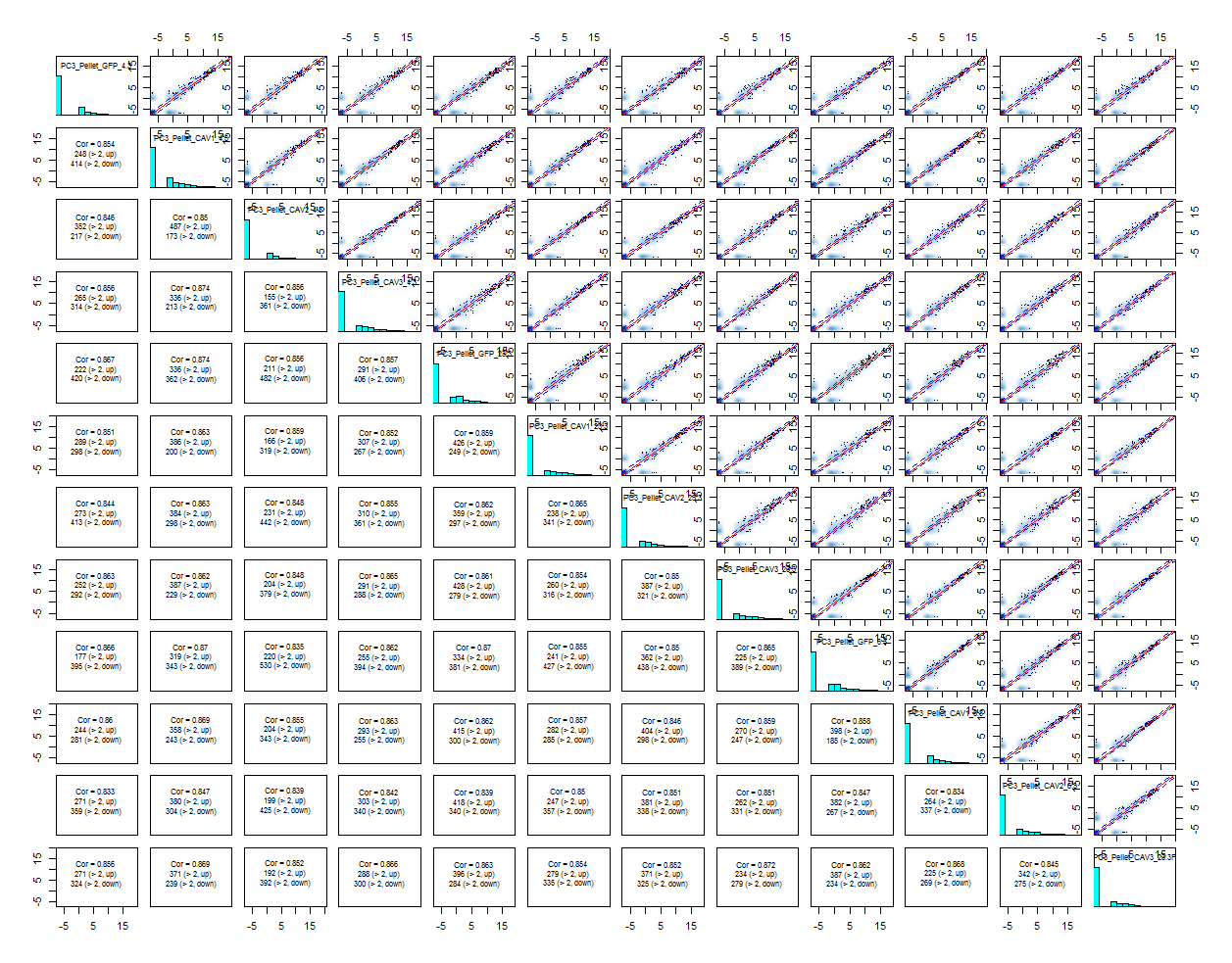


#make a scatterplot

cpm<-replace(cpm, cpm==0, 0.01)

min<-ExpressionSet(assayData=cpm)

pairs(min, smoothScatter=TRUE)



#NOTE: Correlations are generally higher between replicates, but not that much higher than between the other genes.

# annotate groups

Groups <-factor(c(rep(c("GFP", "CAV1", "CAV2","CAV3"),3)))

design <- model.matrix(~Groups)

rownames(design) <- colnames(y)

#Perform the statistical analysis

# estimate the overall level of biological variability

y <- estimateGLMCommonDisp(y, design, verbose=TRUE)

# estimate genewise and tagwise dispersions

y <- estimateGLMTrendedDisp(y, design)

y <- estimateGLMTagwiseDisp(y, design)

# fit the linear model

fit <- glmFit(y, design)

# Identify differentially expressed genes first with individual comparisons

# define groups

y$samples$group <- c(rep(c("GFP", "CAV1", "CAV2","CAV3"),3))

# calculate differential expression

GFP\_vs\_CAV1 <- exactTest(y, pair=c("GFP","CAV1"))

# summarize results, -1 means down regulated, 1 means up regulated.

# The following results mean that there was 1 down-regulated gene, 1 up-regulated gene, and 364

# non-changing genes

summary(de <- decideTestsDGE(GFP\_vs\_CAV1, p=0.05, adjust="BH"))

-1 1

0 364

1 1

topTags(GFP\_vs\_CAV1, n=10)

Comparison of groups: CAV1-GFP

genes logFC logCPM PValue FDR

1116 hsa-miR-363-3p -4.121184 4.078700 1.277446e-11 4.675454e-09

464 hsa-miR-574-5p 2.267544 6.350468 2.906779e-05 5.319405e-03

1963 hsa-miR-146a-5p -1.783144 4.821864 1.935443e-03 2.361240e-01

1496 hsa-miR-203a-3p -1.406424 9.404166 4.361309e-03 3.972078e-01

313 hsa-miR-629-5p 1.269437 5.520661 6.502367e-03 3.972078e-01

440 hsa-miR-10b-3p -1.741761 2.505876 6.871165e-03 3.972078e-01

1873 hsa-miR-20b-5p -1.646172 3.548263 9.908283e-03 3.972078e-01

1890 hsa-miR-18a-5p -1.172839 5.262917 1.119472e-02 3.972078e-01

84 hsa-miR-20a-5p -1.267616 11.483039 1.161578e-02 3.972078e-01

1951 hsa-miR-221-3p 1.118144 12.487931 1.311914e-02 3.972078e-01

# calculate differential expression

GFP\_vs\_CAV2 <- exactTest(y, pair=c("GFP","CAV2"))

# summarize results, -1 means down regulated, 1 means up regulated.

# The following results mean that there was 10 down-regulated genes, 5 up-regulated genes, and 351

# non-changing genes

summary(de <- decideTestsDGE(GFP\_vs\_CAV2, p=0.05, adjust="BH"))

[,1]

-1 10

0 351

1 5

topTags(GFP\_vs\_CAV2, n=20)

Comparison of groups: CAV2-GFP

genes logFC logCPM PValue FDR

1116 hsa-miR-363-3p -5.318171 4.078700 8.685486e-14 3.178888e-11

2377 hsa-miR-10b-5p -2.783009 11.185039 1.518303e-07 2.778494e-05

1963 hsa-miR-146a-5p -2.712775 4.821864 1.224634e-05 1.494054e-03

464 hsa-miR-574-5p 2.273898 6.350468 2.896513e-05 2.650309e-03

1951 hsa-miR-221-3p 1.865365 12.487931 4.980827e-05 3.645965e-03

440 hsa-miR-10b-3p -2.935282 2.505876 6.542387e-05 3.990856e-03

1496 hsa-miR-203a-3p -1.982541 9.404166 7.785224e-05 4.070560e-03

2333 hsa-miR-125b-2-3p -2.535897 4.502000 1.585208e-04 7.252325e-03

1624 hsa-miR-221-5p 2.008512 9.118028 3.010612e-04 1.224316e-02

1328 hsa-miR-934-5p 2.731264 5.510174 3.504928e-04 1.282804e-02

2178 hsa-miR-424-3p -1.899921 8.159926 4.703532e-04 1.564993e-02

1441 hsa-miR-222-3p 1.876378 12.525898 6.209947e-04 1.894034e-02

1740 hsa-miR-139-5p -1.625772 4.259091 8.198907e-04 2.308308e-02

687 hsa-miR-522-3p -3.286971 2.915056 1.647720e-03 4.307611e-02

1541 hsa-miR-148a-5p -1.697725 7.128815 1.977391e-03 4.824834e-02

1114 hsa-miR-4454-5p 1.533263 6.838510 2.513040e-03 5.605243e-02

518 hsa-miR-183-5p 1.345068 11.390613 2.603528e-03 5.605243e-02

473 hsa-miR-148a-3p -1.518469 13.873040 2.798630e-03 5.690547e-02

1873 hsa-miR-20b-5p -1.794395 3.548263 4.519800e-03 8.706562e-02

558 hsa-miR-222-5p 1.534013 4.560775 4.828104e-03 8.835430e-02

# calculate differential expression

GFP\_vs\_CAV3 <- exactTest(y, pair=c("GFP","CAV3"))

# summarize results, -1 means down regulated, 1 means up regulated.

# The following results mean that there was 1 down-regulated gene, 0 up-regulated genes, and 364

# non-changing genes

summary(de <- decideTestsDGE(GFP\_vs\_CAV3, p=0.05, adjust="BH"))

[,1]

-1 1

0 365

1 0

topTags(GFP\_vs\_CAV3, n=10)

Comparison of groups: CAV3-GFP

genes logFC logCPM PValue FDR

1116 hsa-miR-363-3p -3.358461 4.078700 2.050113e-09 7.503413e-07

1873 hsa-miR-20b-5p -2.328636 3.548263 4.643175e-04 7.067893e-02

464 hsa-miR-574-5p 1.853919 6.350468 5.793355e-04 7.067893e-02

235 hsa-miR-106a-5p -1.702122 6.240982 1.187664e-03 9.370946e-02

2184 hsa-miR-17-5p -1.666540 9.741503 1.280184e-03 9.370946e-02

84 hsa-miR-20a-5p -1.584213 11.483039 1.757741e-03 1.072222e-01

1566 hsa-miR-210-3p 1.582657 8.690336 2.959346e-03 1.547315e-01

2220 hsa-miR-3195-5p -1.819565 4.509448 3.732009e-03 1.586891e-01

444 hsa-miR-18a-3p -1.560324 3.382714 4.435852e-03 1.586891e-01

1496 hsa-miR-203a-3p -1.389626 9.404166 4.820068e-03 1.586891e-01

#Conclusions for question 1: 1 miRNA are in common between all 3 CAVs (miR-363-3p). 1 miRNA is statistically

#significant in CAV1 and CAV2, but fails to make significance in CAV3 (miR-574-5p). CAV2 seems to have more

#biological impact on the cells than the other 2 CAVs.