To determine how western diet consumption impact the cortex of young mice

## Data

1. RNAseq samples from cortical tissue

## ---- load cleaned RNA-seq count file and design file ready  
load("~/Documents/code/short-term-diet/01\_EdgeR\_glmQLFit/data/01\_data\_cleaning.RData")  
source("~/Documents/code/01\_function/my\_edgeR.R") # tidyverse library is loaded in my\_edgeR.R   
  
out\_prefix<-"Cortex.RMoutlier"  
out\_analysis <- "./analysis/"  
out\_figure <- "./figure/"  
  
data.raw=data.raw.C  
data.design=data.design.C  
group=group.C  
  
library(tidyverse)  
#remove outlier sample CD.2weeks.Sed.C\_17989C and get rid of all running groups (pre-determined in Cortex.Rmd)  
data.raw <- data.raw %>% select(-contains("17989"),-contains("Run"))  
data.design <- data.design %>% filter(!Sample\_Name %in% c("17989C"), TERM\_3 == "Sed")  
group <- factor(data.design$Group)

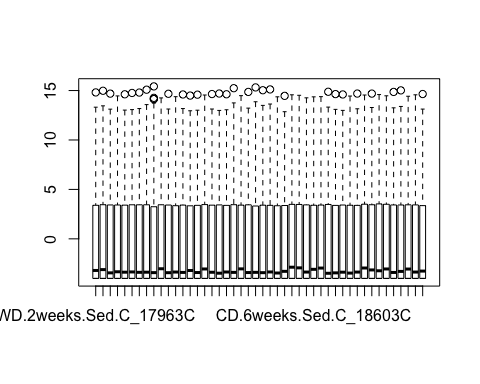
## Quality Control 1:

1. Build edgeR object
2. filtering
3. normalization
4. estimate dispersion

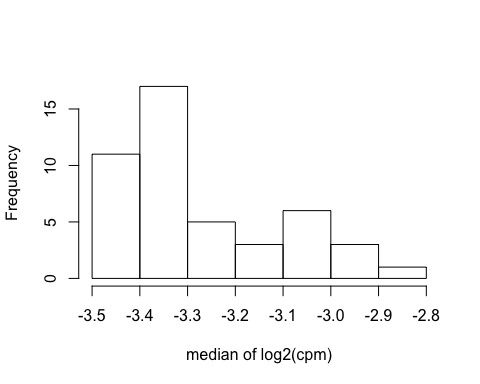
# build edgeR object  
library(edgeR)  
y <- DGEList(data.raw, group=group, genes=row.names(data.raw)) # must specify  
options(digits=3)  
y$samples

group lib.size norm.factors  
WD.2weeks.Sed.C\_17963C WD.2weeks.Sed.C 43017279 1  
WD.2weeks.Sed.C\_17964C WD.2weeks.Sed.C 37549661 1  
WD.2weeks.Sed.C\_17966C WD.2weeks.Sed.C 33696652 1  
WD.2weeks.Sed.C\_17967C WD.2weeks.Sed.C 27098273 1  
WD.4weeks.Sed.C\_17969C WD.4weeks.Sed.C 29356127 1  
WD.4weeks.Sed.C\_17970C WD.4weeks.Sed.C 27998447 1  
WD.4weeks.Sed.C\_17971C WD.4weeks.Sed.C 30152269 1  
WD.4weeks.Sed.C\_17972C WD.4weeks.Sed.C 29375786 1  
WD.4weeks.Sed.C\_17973C WD.4weeks.Sed.C 32115597 1  
WD.6weeks.Sed.C\_17974C WD.6weeks.Sed.C 33141399 1  
WD.6weeks.Sed.C\_17976C WD.6weeks.Sed.C 32511585 1  
WD.6weeks.Sed.C\_17977C WD.6weeks.Sed.C 28580449 1  
WD.6weeks.Sed.C\_17978C WD.6weeks.Sed.C 30749233 1  
WD.6weeks.Sed.C\_17979C WD.6weeks.Sed.C 43298289 1  
WD.8weeks.Sed.C\_17980C WD.8weeks.Sed.C 31456901 1  
WD.8weeks.Sed.C\_17981C WD.8weeks.Sed.C 34234419 1  
WD.8weeks.Sed.C\_17982C WD.8weeks.Sed.C 29762618 1  
WD.8weeks.Sed.C\_17985C WD.8weeks.Sed.C 36324394 1  
CD.2weeks.Sed.C\_17986C CD.2weeks.Sed.C 28313435 1  
CD.2weeks.Sed.C\_17987C CD.2weeks.Sed.C 30477586 1  
CD.2weeks.Sed.C\_17990C CD.2weeks.Sed.C 33715581 1  
CD.2weeks.Sed.C\_17991C CD.2weeks.Sed.C 31572108 1  
CD.4weeks.Sed.C\_17992C CD.4weeks.Sed.C 30289837 1  
CD.4weeks.Sed.C\_17993C CD.4weeks.Sed.C 32340022 1  
CD.4weeks.Sed.C\_17995C CD.4weeks.Sed.C 28067060 1  
CD.4weeks.Sed.C\_17996C CD.4weeks.Sed.C 34882965 1  
CD.4weeks.Sed.C\_17997C CD.4weeks.Sed.C 24995985 1  
CD.6weeks.Sed.C\_17999C CD.6weeks.Sed.C 26903295 1  
CD.6weeks.Sed.C\_18000C CD.6weeks.Sed.C 29248536 1  
CD.6weeks.Sed.C\_18602C CD.6weeks.Sed.C 28625215 1  
CD.6weeks.Sed.C\_18603C CD.6weeks.Sed.C 35993258 1  
CD.8weeks.Sed.C\_18605C CD.8weeks.Sed.C 30106799 1  
CD.8weeks.Sed.C\_18606C CD.8weeks.Sed.C 37039384 1  
WD.8weeks.Sed.C\_18624C WD.8weeks.Sed.C 33772116 1  
WD.8weeks.Sed.C\_18625C WD.8weeks.Sed.C 29289679 1  
CD.8weeks.Sed.C\_18626C CD.8weeks.Sed.C 35490496 1  
CD.8weeks.Sed.C\_18627C CD.8weeks.Sed.C 28699029 1  
WD.2weeks.Sed.C\_18673C WD.2weeks.Sed.C 29921911 1  
WD.2weeks.Sed.C\_18674C WD.2weeks.Sed.C 39569604 1  
WD.4weeks.Sed.C\_18675C WD.4weeks.Sed.C 22661282 1  
WD.6weeks.Sed.C\_18676C WD.6weeks.Sed.C 34190712 1  
CD.2weeks.Sed.C\_18677C CD.2weeks.Sed.C 30938873 1  
CD.4weeks.Sed.C\_18678C CD.4weeks.Sed.C 25295211 1  
CD.6weeks.Sed.C\_18679C CD.6weeks.Sed.C 34529870 1  
CD.6weeks.Sed.C\_18680C CD.6weeks.Sed.C 28609160 1  
CD.8weeks.Sed.C\_18681C CD.8weeks.Sed.C 47808356 1

# before filtering and normalization  
log.cpm <- cpm(y, log=TRUE, prior.count=2)   
boxplot(log.cpm)



hist(apply(log.cpm, 2, median), xlab="median of log2(cpm)", main="")



# attach gene symbols to edgeR object   
y<- symbols(y)  
# The following package has already been detached within symbol function I wrote   
#detach("package:org.Mm.eg.db", unload=TRUE) # conflict with select function, so to detach  
#detach("package:AnnotationDbi", unload=TRUE)  
  
head(y$genes)

genes Symbol Entrezid  
ENSMUSG00000000001 ENSMUSG00000000001 Gnai3 14679  
ENSMUSG00000000003 ENSMUSG00000000003 Pbsn 54192  
ENSMUSG00000000028 ENSMUSG00000000028 Cdc45 12544  
ENSMUSG00000000031 ENSMUSG00000000031 H19 14955  
ENSMUSG00000000037 ENSMUSG00000000037 Scml2 107815  
ENSMUSG00000000049 ENSMUSG00000000049 Apoh 11818  
 Genename  
ENSMUSG00000000001 guanine nucleotide binding protein (G protein), alpha inhibiting 3  
ENSMUSG00000000003 probasin  
ENSMUSG00000000028 cell division cycle 45  
ENSMUSG00000000031 H19, imprinted maternally expressed transcript  
ENSMUSG00000000037 Scm polycomb group protein like 2  
ENSMUSG00000000049 apolipoprotein H

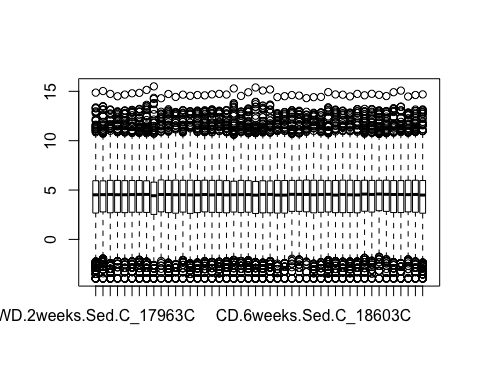
## dropNAsymbols  
y <- y[!is.na(y$genes$Symbol),]   
  
# keep gene more than 1cpm in at least 2 samples   
dim(y)

[1] 23103 46

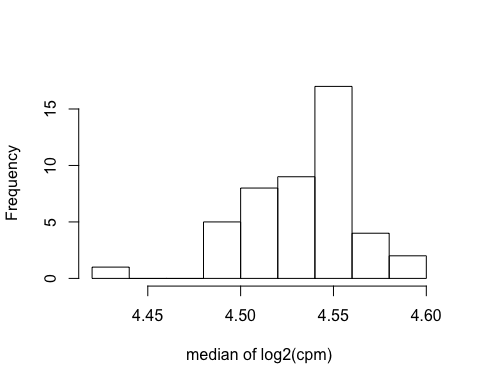
keep <- rowSums(cpm(y) > 1) >= 2 ## filter  
table(keep)

keep  
FALSE TRUE   
 8774 14329

y <- y[keep, , keep.lib.sizes=FALSE]  
  
# after filtering and before normalization  
log.cpm <- cpm(y, log=TRUE, prior.count=2)   
boxplot(log.cpm)



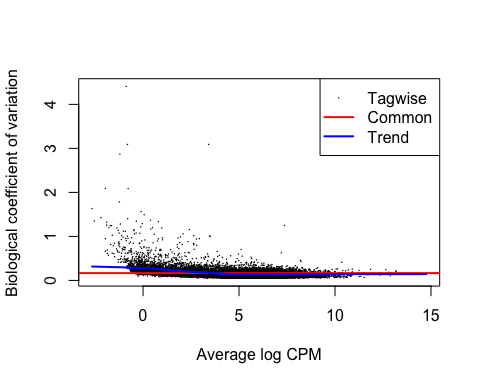
hist(apply(log.cpm, 2, median), xlab="median of log2(cpm)", main="")



# estimate dispersion  
# design  
design <- model.matrix(~0+group)  
colnames(design) <- levels(group)  
design

CD.2weeks.Sed.C CD.4weeks.Sed.C CD.6weeks.Sed.C CD.8weeks.Sed.C  
1 0 0 0 0  
2 0 0 0 0  
3 0 0 0 0  
4 0 0 0 0  
5 0 0 0 0  
6 0 0 0 0  
7 0 0 0 0  
8 0 0 0 0  
9 0 0 0 0  
10 0 0 0 0  
11 0 0 0 0  
12 0 0 0 0  
13 0 0 0 0  
14 0 0 0 0  
15 0 0 0 0  
16 0 0 0 0  
17 0 0 0 0  
18 0 0 0 0  
19 1 0 0 0  
20 1 0 0 0  
21 1 0 0 0  
22 1 0 0 0  
23 0 1 0 0  
24 0 1 0 0  
25 0 1 0 0  
26 0 1 0 0  
27 0 1 0 0  
28 0 0 1 0  
29 0 0 1 0  
30 0 0 1 0  
31 0 0 1 0  
32 0 0 0 1  
33 0 0 0 1  
34 0 0 0 0  
35 0 0 0 0  
36 0 0 0 1  
37 0 0 0 1  
38 0 0 0 0  
39 0 0 0 0  
40 0 0 0 0  
41 0 0 0 0  
42 1 0 0 0  
43 0 1 0 0  
44 0 0 1 0  
45 0 0 1 0  
46 0 0 0 1  
 WD.2weeks.Sed.C WD.4weeks.Sed.C WD.6weeks.Sed.C WD.8weeks.Sed.C  
1 1 0 0 0  
2 1 0 0 0  
3 1 0 0 0  
4 1 0 0 0  
5 0 1 0 0  
6 0 1 0 0  
7 0 1 0 0  
8 0 1 0 0  
9 0 1 0 0  
10 0 0 1 0  
11 0 0 1 0  
12 0 0 1 0  
13 0 0 1 0  
14 0 0 1 0  
15 0 0 0 1  
16 0 0 0 1  
17 0 0 0 1  
18 0 0 0 1  
19 0 0 0 0  
20 0 0 0 0  
21 0 0 0 0  
22 0 0 0 0  
23 0 0 0 0  
24 0 0 0 0  
25 0 0 0 0  
26 0 0 0 0  
27 0 0 0 0  
28 0 0 0 0  
29 0 0 0 0  
30 0 0 0 0  
31 0 0 0 0  
32 0 0 0 0  
33 0 0 0 0  
34 0 0 0 1  
35 0 0 0 1  
36 0 0 0 0  
37 0 0 0 0  
38 1 0 0 0  
39 1 0 0 0  
40 0 1 0 0  
41 0 0 1 0  
42 0 0 0 0  
43 0 0 0 0  
44 0 0 0 0  
45 0 0 0 0  
46 0 0 0 0  
attr(,"assign")  
[1] 1 1 1 1 1 1 1 1  
attr(,"contrasts")  
attr(,"contrasts")$group  
[1] "contr.treatment"

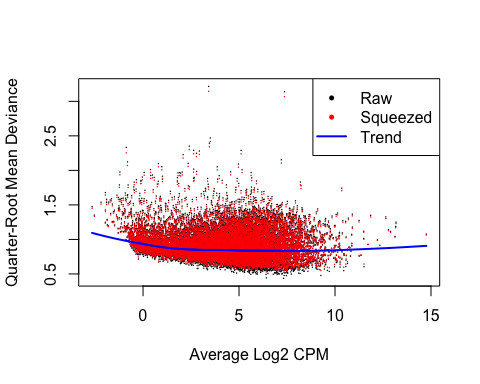
# estimateDisp--------------------------------------------------------  
y <- estimateDisp(y, design, robust=TRUE)  
# plotBCV, width="3.8in", fig.cap="Scatterplot of the biological coefficient of variation (BCV) against the average abundance of each gene. The plot shows the square-root estimates of the common, trended and tagwise NB dispersions."----  
plotBCV(y)



# glmQLFit------------------------------------------------------------  
fit <- glmQLFit(y, design, robust=TRUE)  
head(fit$coefficients)

CD.2weeks.Sed.C CD.4weeks.Sed.C CD.6weeks.Sed.C  
ENSMUSG00000000001 -10.16 -10.24 -10.09  
ENSMUSG00000000028 -13.57 -13.42 -13.71  
ENSMUSG00000000037 -13.28 -13.75 -13.06  
ENSMUSG00000000056 -9.98 -9.92 -9.99  
ENSMUSG00000000058 -10.09 -10.22 -9.87  
ENSMUSG00000000078 -9.78 -9.83 -9.68  
 CD.8weeks.Sed.C WD.2weeks.Sed.C WD.4weeks.Sed.C  
ENSMUSG00000000001 -10.06 -10.17 -10.17  
ENSMUSG00000000028 -13.40 -13.55 -13.48  
ENSMUSG00000000037 -13.22 -13.28 -12.88  
ENSMUSG00000000056 -9.85 -9.97 -10.05  
ENSMUSG00000000058 -9.95 -10.01 -10.18  
ENSMUSG00000000078 -9.73 -9.73 -9.82  
 WD.6weeks.Sed.C WD.8weeks.Sed.C  
ENSMUSG00000000001 -10.05 -10.12  
ENSMUSG00000000028 -13.58 -13.45  
ENSMUSG00000000037 -13.35 -12.98  
ENSMUSG00000000056 -9.95 -9.87  
ENSMUSG00000000058 -9.85 -9.93  
ENSMUSG00000000078 -9.61 -9.67

# QLDisp, out.width="3.8in", fig.cap="A plot of the quarter-root QL dispersion against the average abundance of each gene. Estimates are shown for the raw (before EB moderation), trended and squeezed (after EB moderation) dispersions. Note that the QL dispersions and trend shown here are relative to the NB dispersion trend shown in Figure~\ref{fig:plotBCV}."----  
plotQLDisp(fit)



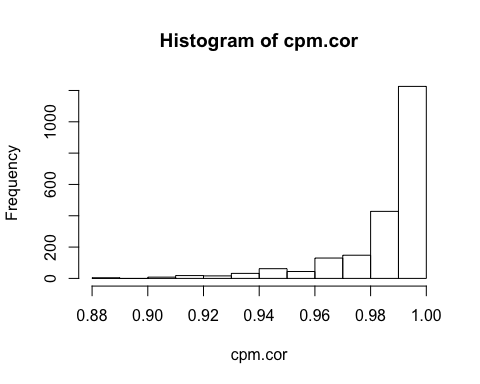
# df.prior------------------------------------------------------------  
summary(fit$df.prior)

Min. 1st Qu. Median Mean 3rd Qu. Max.   
 3.57 3.57 3.64 3.64 3.70 3.70

## Quality Control 2:

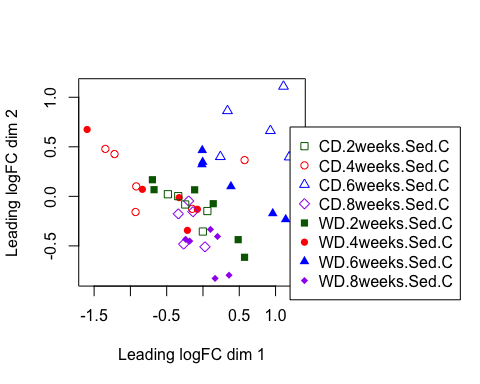
1. pearson correlation among samples
2. MDS plot
3. Principle Component Analysis

# pearson correlation  
cpm <- cpm(y, normalized.lib.size=T)   
cpm.cor <- cor(cpm, method = "pearson")  
hist(cpm.cor)

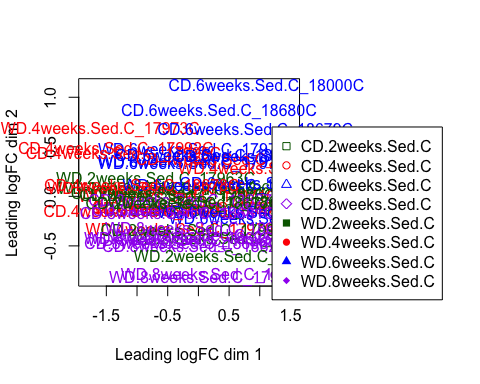


library(corrplot)  
corrplot(cpm.cor, method="square", order="hclust", cl.lim = c(0.85, 1), tl.col="purple", tl.cex = 0.75, type = "full", is.corr = FALSE)  
colorlegend(colbar = grey(1:100 / 100), 1:10, col = "red", align = "l",  
 xlim = c(0, 6), ylim = c(-0.5,-0.1), vertical = FALSE)

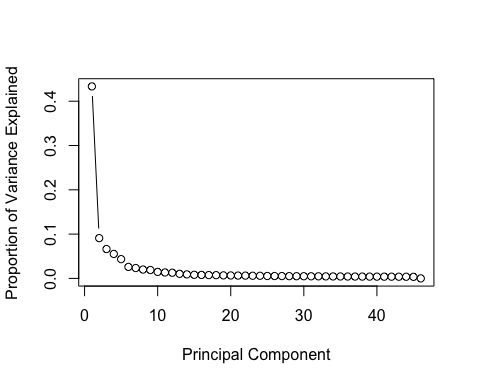
# MDS plot  
 # no label version  
par(xpd = T, mar = par()$mar + c(0,0,0,7)) # to make legends outside of the plot  
colors <- rep(c("darkgreen", "red", "blue", "purple"), 2)  
pch <- c(0,1,2,5,15,16,17,18)  
plotMDS(y, top = 500, cex = 1, pch=pch[group], dim.plot = c(1,2), ndim = 2, gene.selection = "pairwise", col=colors[group]) # col = as.numeric(group) differentiate colors between groups  
legend(1.2, 0.7,levels(group), pch=pch, col=colors)



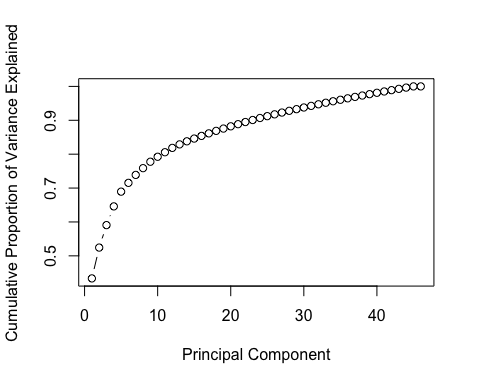
par(mar=c(5, 4, 4, 2.3) + 0.1)  
  
 # with label version, to identify outliers  
par(xpd = T, mar = par()$mar + c(0,0,0,7)) # to make legends outside of the plot  
colors <- rep(c("darkgreen", "red", "blue", "purple"), 2)  
 # pch <- c(0,1,2,5,15,16,17,18)  
plotMDS(y, top = 500, cex = 1, dim.plot = c(1,2), ndim = 2, gene.selection = "pairwise", col=colors[group]) # col = as.numeric(group) differentiate colors between groups  
legend(1.2, 0.7,levels(group), pch=pch, col=colors)



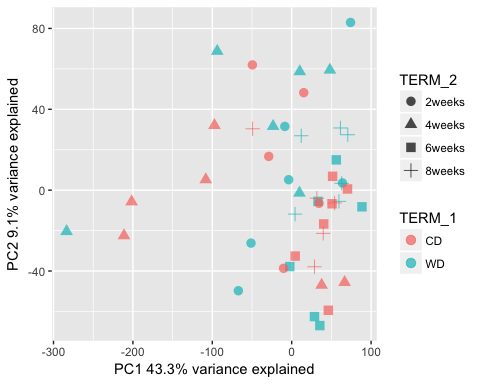
par(mar=c(5, 4, 4, 2.3) + 0.1)  
  
## ---- Pincicpal component analysis ---  
logCPM.PCA<-log.cpm   
rownames(logCPM.PCA) <- y$genes$Symbol  
#colnames(logCPM) <- paste(y$samples$group, 1:2, sep="-")  
colnames(logCPM.PCA) <- data.design$sample\_short # get it into the y project  
  
pca\_original = prcomp(t(logCPM.PCA),scale=T, center=T)  
pca\_x <- pca\_original$x  
pca\_table <- data.frame(pca\_x, data.design)  
x <- pca\_original$sdev^2/sum(pca\_original$sdev^2) # Proportion of Variance Explained for all components  
  
## Scree plot  
plot(x, xlab="Principal Component", ylab="Proportion of Variance Explained", type="b")



plot(cumsum(x), xlab="Principal Component", ylab="Cumulative Proportion of Variance Explained", type="b")



## PCA plot   
library(ggplot2)  
PCA\_plot <- function(pca\_table, PC\_x, PC\_y, color, shape){  
 #PC\_x,PC\_y are type of interger  
 #color, shape, are type of string  
 g <- ggplot(pca\_table, aes\_string(x=names(pca\_table[PC\_x]), y=names(pca\_table[PC\_y]), color=color, shape=shape))   
 g <- g + geom\_point(alpha=0.7, size=3)   
 # g <- g + labs(color = "Group", shape="Tissue")  
 g + labs(x = paste(names(pca\_table[PC\_x]), scales::percent(x[PC\_x]),"variance explained", sep=" "), y=paste(names(pca\_table[PC\_y]), scales::percent(x[PC\_y]),"variance explained", sep=" "))  
 #filename <- paste()  
 #ggsave(filename, width=7, height=7, units="in")  
}  
  
PCA\_plot(pca\_table, 1, 2, "TERM\_1", "TERM\_2")



PCA\_plot(pca\_table, 2, 3, "TERM\_1", "TERM\_2")



## Data Manipulation for GLM

# transpose cpm, observation (row) is sample, variable (column) is each gene.   
cpm <- cpm(y, normalized.lib.size=T)   
cpm.t<-t(cpm)  
colnames(cpm)

[1] "WD.2weeks.Sed.C\_17963C" "WD.2weeks.Sed.C\_17964C"  
 [3] "WD.2weeks.Sed.C\_17966C" "WD.2weeks.Sed.C\_17967C"  
 [5] "WD.4weeks.Sed.C\_17969C" "WD.4weeks.Sed.C\_17970C"  
 [7] "WD.4weeks.Sed.C\_17971C" "WD.4weeks.Sed.C\_17972C"  
 [9] "WD.4weeks.Sed.C\_17973C" "WD.6weeks.Sed.C\_17974C"  
[11] "WD.6weeks.Sed.C\_17976C" "WD.6weeks.Sed.C\_17977C"  
[13] "WD.6weeks.Sed.C\_17978C" "WD.6weeks.Sed.C\_17979C"  
[15] "WD.8weeks.Sed.C\_17980C" "WD.8weeks.Sed.C\_17981C"  
[17] "WD.8weeks.Sed.C\_17982C" "WD.8weeks.Sed.C\_17985C"  
[19] "CD.2weeks.Sed.C\_17986C" "CD.2weeks.Sed.C\_17987C"  
[21] "CD.2weeks.Sed.C\_17990C" "CD.2weeks.Sed.C\_17991C"  
[23] "CD.4weeks.Sed.C\_17992C" "CD.4weeks.Sed.C\_17993C"  
[25] "CD.4weeks.Sed.C\_17995C" "CD.4weeks.Sed.C\_17996C"  
[27] "CD.4weeks.Sed.C\_17997C" "CD.6weeks.Sed.C\_17999C"  
[29] "CD.6weeks.Sed.C\_18000C" "CD.6weeks.Sed.C\_18602C"  
[31] "CD.6weeks.Sed.C\_18603C" "CD.8weeks.Sed.C\_18605C"  
[33] "CD.8weeks.Sed.C\_18606C" "WD.8weeks.Sed.C\_18624C"  
[35] "WD.8weeks.Sed.C\_18625C" "CD.8weeks.Sed.C\_18626C"  
[37] "CD.8weeks.Sed.C\_18627C" "WD.2weeks.Sed.C\_18673C"  
[39] "WD.2weeks.Sed.C\_18674C" "WD.4weeks.Sed.C\_18675C"  
[41] "WD.6weeks.Sed.C\_18676C" "CD.2weeks.Sed.C\_18677C"  
[43] "CD.4weeks.Sed.C\_18678C" "CD.6weeks.Sed.C\_18679C"  
[45] "CD.6weeks.Sed.C\_18680C" "CD.8weeks.Sed.C\_18681C"

all(data.design$sample\_short == rownames(cpm.t))

[1] TRUE

# add design file to transposed cpm, data manipulation using filter based on design file TERMs (TERM\_1 - TERM\_4)  
cpm.t.meta <- cbind(cpm.t, data.design)  
# select all the WD and Sed group  
cpm.WD.t <-cpm.t.meta %>% filter(TERM\_1 == "WD", TERM\_3 == "Sed")   
# remove metadata and transpose table back, rows = genes, columns = samples  
gene\_number <- dim(cpm)[1]  
gene\_number

[1] 14329

cpm.WD <- cpm.WD.t[, c(1: gene\_number)] %>% t()   
# check whether metadata (which contains characters has been removed, values should all be doulbe type now)  
all(map\_lgl(cpm.WD, is.double))

[1] TRUE

# assign sample\_name\_short it to the colnames after transposition.  
colnames(cpm.WD) <- as.character(cpm.WD.t$sample\_short)  
  
# select all the CD and Sed group, and calculate the mean of each group based on weeks  
cpm.ave.CD.meta <- cpm.t.meta %>% filter(TERM\_1 == "CD") %>% group\_by(TERM\_2) %>% summarise\_all(mean)  
# check the column boundray of cpm  
cpm.ave.CD.meta[,1:2]

# A tibble: 4 x 2  
 TERM\_2 ENSMUSG00000000001  
 <fct> <dbl>  
1 2weeks 38.7  
2 4weeks 35.7  
3 6weeks 41.7  
4 8weeks 42.7

cpm.ave.CD.meta[,(gene\_number+1):(gene\_number+2)]

# A tibble: 4 x 2  
 ENSMUSG00000099250 Sample\_ID  
 <dbl> <dbl>  
1 9.72 NA  
2 15.4 NA  
3 8.57 NA  
4 9.44 NA

# remove meta column and transform  
cpm.ave.CD <- cpm.ave.CD.meta[, 2:(gene\_number+1)] %>% t()   
colnames(cpm.ave.CD) <- cpm.ave.CD.meta[,1] %>% unlist() %>% paste("CD.ave.", ., sep="")  
all(map\_lgl(cpm.ave.CD, is.double))

[1] TRUE

# The following is to subtract the mean of CD from WD of each week, and glm model the difference between WD and CD group  
dim(cpm.WD)

[1] 14329 24

dim(cpm.ave.CD)

[1] 14329 4

cpm.clean <- data.frame(cpm.WD, cpm.ave.CD) ## column 1:14395 are WD count read, 14396:14399 are CD count read.  
dim(cpm.clean)

[1] 14329 28

head(cpm.clean)

WD.2weeks.Sed.C\_17963C WD.2weeks.Sed.C\_17964C  
ENSMUSG00000000001 31.56 37.03  
ENSMUSG00000000028 1.11 1.41  
ENSMUSG00000000037 1.11 1.24  
ENSMUSG00000000056 45.38 42.03  
ENSMUSG00000000058 38.66 44.21  
ENSMUSG00000000078 56.03 64.47  
 WD.2weeks.Sed.C\_17966C WD.2weeks.Sed.C\_17967C  
ENSMUSG00000000001 37.76 42.38  
ENSMUSG00000000028 1.41 1.22  
ENSMUSG00000000037 1.04 2.24  
ENSMUSG00000000056 51.79 48.12  
ENSMUSG00000000058 36.53 50.71  
ENSMUSG00000000078 56.58 55.47  
 WD.4weeks.Sed.C\_17969C WD.4weeks.Sed.C\_17970C  
ENSMUSG00000000001 39.97 33.24  
ENSMUSG00000000028 1.02 1.48  
ENSMUSG00000000037 1.34 3.43  
ENSMUSG00000000056 41.06 42.87  
ENSMUSG00000000058 43.20 46.52  
ENSMUSG00000000078 54.32 59.66  
 WD.4weeks.Sed.C\_17971C WD.4weeks.Sed.C\_17972C  
ENSMUSG00000000001 39.96 40.34  
ENSMUSG00000000028 1.06 1.66  
ENSMUSG00000000037 3.36 2.48  
ENSMUSG00000000056 46.23 43.38  
ENSMUSG00000000058 39.58 32.55  
ENSMUSG00000000078 49.59 53.02  
 WD.4weeks.Sed.C\_17973C WD.6weeks.Sed.C\_17974C  
ENSMUSG00000000001 33.121 46.56  
ENSMUSG00000000028 2.008 1.40  
ENSMUSG00000000037 0.428 2.15  
ENSMUSG00000000056 41.022 48.43  
ENSMUSG00000000058 24.528 61.54  
ENSMUSG00000000078 46.520 69.75  
 WD.6weeks.Sed.C\_17976C WD.6weeks.Sed.C\_17977C  
ENSMUSG00000000001 38.88 45.77  
ENSMUSG00000000028 1.02 1.42  
ENSMUSG00000000037 1.06 1.60  
ENSMUSG00000000056 44.25 51.57  
ENSMUSG00000000058 55.70 56.58  
ENSMUSG00000000078 63.50 71.94  
 WD.6weeks.Sed.C\_17978C WD.6weeks.Sed.C\_17979C  
ENSMUSG00000000001 42.90 42.62  
ENSMUSG00000000028 1.49 1.10  
ENSMUSG00000000037 1.29 1.39  
ENSMUSG00000000056 46.01 48.90  
ENSMUSG00000000058 41.85 56.76  
ENSMUSG00000000078 62.30 69.67  
 WD.8weeks.Sed.C\_17980C WD.8weeks.Sed.C\_17981C  
ENSMUSG00000000001 39.13 42.16  
ENSMUSG00000000028 1.41 1.14  
ENSMUSG00000000037 1.70 3.07  
ENSMUSG00000000056 51.14 50.76  
ENSMUSG00000000058 49.60 50.21  
ENSMUSG00000000078 63.22 64.92  
 WD.8weeks.Sed.C\_17982C WD.8weeks.Sed.C\_17985C  
ENSMUSG00000000001 40.37 40.65  
ENSMUSG00000000028 1.18 1.45  
ENSMUSG00000000037 3.05 2.87  
ENSMUSG00000000056 51.16 52.17  
ENSMUSG00000000058 50.46 47.18  
ENSMUSG00000000078 66.42 57.19  
 WD.8weeks.Sed.C\_18624C WD.8weeks.Sed.C\_18625C  
ENSMUSG00000000001 39.33 40.75  
ENSMUSG00000000028 1.62 1.84  
ENSMUSG00000000037 1.37 1.73  
ENSMUSG00000000056 51.66 54.41  
ENSMUSG00000000058 50.53 45.60  
ENSMUSG00000000078 65.87 62.94  
 WD.2weeks.Sed.C\_18673C WD.2weeks.Sed.C\_18674C  
ENSMUSG00000000001 36.45 44.58  
ENSMUSG00000000028 1.35 1.32  
ENSMUSG00000000037 3.53 1.21  
ENSMUSG00000000056 45.54 47.92  
ENSMUSG00000000058 52.77 47.61  
ENSMUSG00000000078 65.15 57.52  
 WD.4weeks.Sed.C\_18675C WD.6weeks.Sed.C\_18676C  
ENSMUSG00000000001 42.78 42.94  
ENSMUSG00000000028 1.09 1.15  
ENSMUSG00000000037 4.33 2.09  
ENSMUSG00000000056 45.06 46.64  
ENSMUSG00000000058 40.72 44.85  
ENSMUSG00000000078 62.80 63.44  
 CD.ave.2weeks CD.ave.4weeks CD.ave.6weeks CD.ave.8weeks  
ENSMUSG00000000001 38.66 35.66 41.66 42.73  
ENSMUSG00000000028 1.27 1.48 1.11 1.51  
ENSMUSG00000000037 1.69 1.05 2.12 1.80  
ENSMUSG00000000056 46.12 49.12 46.06 52.49  
ENSMUSG00000000058 41.34 36.42 51.65 47.58  
ENSMUSG00000000078 56.63 53.89 62.75 59.49

# WD-CD.ave, result will be combined into D\_all value  
D\_2wk <- cpm.clean %>% select(contains("2weeks")) %>% mutate\_all(funs(Dif=.- CD.ave.2weeks)) %>% select(contains("C\_Dif"))  
D\_4wk <- cpm.clean %>% select(contains("4weeks")) %>% mutate\_all(funs(Dif=.- CD.ave.4weeks)) %>% select(contains("C\_Dif"))  
D\_6wk <- cpm.clean %>% select(contains("6weeks")) %>% mutate\_all(funs(Dif=.- CD.ave.6weeks)) %>% select(contains("C\_Dif"))  
D\_8wk <- cpm.clean %>% select(contains("8weeks")) %>% mutate\_all(funs(Dif=.- CD.ave.8weeks)) %>% select(contains("C\_Dif"))  
  
D\_all <- data.frame(D\_2wk, D\_4wk, D\_6wk, D\_8wk)   
rownames(D\_all) <- rownames(cpm.ave.CD)  
head(D\_all)

WD.2weeks.Sed.C\_17963C\_Dif WD.2weeks.Sed.C\_17964C\_Dif  
ENSMUSG00000000001 -7.101 -1.629  
ENSMUSG00000000028 -0.163 0.134  
ENSMUSG00000000037 -0.579 -0.447  
ENSMUSG00000000056 -0.737 -4.084  
ENSMUSG00000000058 -2.682 2.871  
ENSMUSG00000000078 -0.592 7.844  
 WD.2weeks.Sed.C\_17966C\_Dif WD.2weeks.Sed.C\_17967C\_Dif  
ENSMUSG00000000001 -0.8959 3.7212  
ENSMUSG00000000028 0.1374 -0.0575  
ENSMUSG00000000037 -0.6462 0.5545  
ENSMUSG00000000056 5.6774 2.0086  
ENSMUSG00000000058 -4.8085 9.3680  
ENSMUSG00000000078 -0.0438 -1.1597  
 WD.2weeks.Sed.C\_18673C\_Dif WD.2weeks.Sed.C\_18674C\_Dif  
ENSMUSG00000000001 -2.2120 5.9208  
ENSMUSG00000000028 0.0737 0.0425  
ENSMUSG00000000037 1.8370 -0.4781  
ENSMUSG00000000056 -0.5749 1.8098  
ENSMUSG00000000058 11.4246 6.2661  
ENSMUSG00000000078 8.5208 0.8895  
 WD.4weeks.Sed.C\_17969C\_Dif WD.4weeks.Sed.C\_17970C\_Dif  
ENSMUSG00000000001 4.304 -2.41975  
ENSMUSG00000000028 -0.463 -0.00716  
ENSMUSG00000000037 0.291 2.38502  
ENSMUSG00000000056 -8.069 -6.25391  
ENSMUSG00000000058 6.780 10.10137  
ENSMUSG00000000078 0.429 5.76706  
 WD.4weeks.Sed.C\_17971C\_Dif WD.4weeks.Sed.C\_17972C\_Dif  
ENSMUSG00000000001 4.30 4.678  
ENSMUSG00000000028 -0.42 0.182  
ENSMUSG00000000037 2.32 1.433  
ENSMUSG00000000056 -2.89 -5.740  
ENSMUSG00000000058 3.16 -3.874  
ENSMUSG00000000078 -4.30 -0.872  
 WD.4weeks.Sed.C\_17973C\_Dif WD.4weeks.Sed.C\_18675C\_Dif  
ENSMUSG00000000001 -2.541 7.114  
ENSMUSG00000000028 0.525 -0.388  
ENSMUSG00000000037 -0.618 3.286  
ENSMUSG00000000056 -8.103 -4.069  
ENSMUSG00000000058 -11.894 4.302  
ENSMUSG00000000078 -7.370 8.905  
 WD.6weeks.Sed.C\_17974C\_Dif WD.6weeks.Sed.C\_17976C\_Dif  
ENSMUSG00000000001 4.9082 -2.7752  
ENSMUSG00000000028 0.2872 -0.0902  
ENSMUSG00000000037 0.0288 -1.0638  
ENSMUSG00000000056 2.3676 -1.8117  
ENSMUSG00000000058 9.8813 4.0449  
ENSMUSG00000000078 7.0075 0.7556  
 WD.6weeks.Sed.C\_17977C\_Dif WD.6weeks.Sed.C\_17978C\_Dif  
ENSMUSG00000000001 4.112 1.2428  
ENSMUSG00000000028 0.302 0.3764  
ENSMUSG00000000037 -0.522 -0.8323  
ENSMUSG00000000056 5.511 -0.0504  
ENSMUSG00000000058 4.929 -9.8054  
ENSMUSG00000000078 9.191 -0.4458  
 WD.6weeks.Sed.C\_17979C\_Dif WD.6weeks.Sed.C\_18676C\_Dif  
ENSMUSG00000000001 0.9652 1.2835  
ENSMUSG00000000028 -0.0114 0.0389  
ENSMUSG00000000037 -0.7294 -0.0265  
ENSMUSG00000000056 2.8340 0.5750  
ENSMUSG00000000058 5.1019 -6.8045  
ENSMUSG00000000078 6.9243 0.6934  
 WD.8weeks.Sed.C\_17980C\_Dif WD.8weeks.Sed.C\_17981C\_Dif  
ENSMUSG00000000001 -3.609 -0.579  
ENSMUSG00000000028 -0.107 -0.372  
ENSMUSG00000000037 -0.101 1.263  
ENSMUSG00000000056 -1.353 -1.739  
ENSMUSG00000000058 2.019 2.631  
ENSMUSG00000000078 3.729 5.424  
 WD.8weeks.Sed.C\_17982C\_Dif WD.8weeks.Sed.C\_17985C\_Dif  
ENSMUSG00000000001 -2.365 -2.0835  
ENSMUSG00000000028 -0.335 -0.0678  
ENSMUSG00000000037 1.248 1.0612  
ENSMUSG00000000056 -1.338 -0.3252  
ENSMUSG00000000058 2.879 -0.4073  
ENSMUSG00000000078 6.923 -2.3037  
 WD.8weeks.Sed.C\_18624C\_Dif WD.8weeks.Sed.C\_18625C\_Dif  
ENSMUSG00000000001 -3.406 -1.9892  
ENSMUSG00000000028 0.103 0.3262  
ENSMUSG00000000037 -0.431 -0.0694  
ENSMUSG00000000056 -0.839 1.9162  
ENSMUSG00000000058 2.943 -1.9878  
ENSMUSG00000000078 6.380 3.4480

## Generalized linear mGLM

### z ~ u + duration, z= y(WD)-y(CD)ave

1. Model the difference of WD and CD by duration on diet
2. The difference of WD and CD is calculated by subtracting the average of CD from each WD

* duration: 2wks, 4wks, 6wks, 8wks

## ---- GLM, z ~ u + duration, z= y(WD)-y(CD)  
# use durataion as predictor, duration matches the column of D\_all  
duration <- c(rep("2weeks", 6), rep("4weeks", 6), rep("6weeks", 6), rep("8weeks", 6))  
  
# for each gene, generate GLM with duration as prediction, collect the summary of the model for each gene  
glm.fit <- D\_all %>% t() %>% as.data.frame() %>%  
 map(~ lm(. ~ duration)) %>%   
 map(summary)  
  
colnames(D\_all)

[1] "WD.2weeks.Sed.C\_17963C\_Dif" "WD.2weeks.Sed.C\_17964C\_Dif"  
 [3] "WD.2weeks.Sed.C\_17966C\_Dif" "WD.2weeks.Sed.C\_17967C\_Dif"  
 [5] "WD.2weeks.Sed.C\_18673C\_Dif" "WD.2weeks.Sed.C\_18674C\_Dif"  
 [7] "WD.4weeks.Sed.C\_17969C\_Dif" "WD.4weeks.Sed.C\_17970C\_Dif"  
 [9] "WD.4weeks.Sed.C\_17971C\_Dif" "WD.4weeks.Sed.C\_17972C\_Dif"  
[11] "WD.4weeks.Sed.C\_17973C\_Dif" "WD.4weeks.Sed.C\_18675C\_Dif"  
[13] "WD.6weeks.Sed.C\_17974C\_Dif" "WD.6weeks.Sed.C\_17976C\_Dif"  
[15] "WD.6weeks.Sed.C\_17977C\_Dif" "WD.6weeks.Sed.C\_17978C\_Dif"  
[17] "WD.6weeks.Sed.C\_17979C\_Dif" "WD.6weeks.Sed.C\_18676C\_Dif"  
[19] "WD.8weeks.Sed.C\_17980C\_Dif" "WD.8weeks.Sed.C\_17981C\_Dif"  
[21] "WD.8weeks.Sed.C\_17982C\_Dif" "WD.8weeks.Sed.C\_17985C\_Dif"  
[23] "WD.8weeks.Sed.C\_18624C\_Dif" "WD.8weeks.Sed.C\_18625C\_Dif"

glm.fit %>% names() %>% head()

[1] "ENSMUSG00000000001" "ENSMUSG00000000028" "ENSMUSG00000000037"  
[4] "ENSMUSG00000000056" "ENSMUSG00000000058" "ENSMUSG00000000078"

# look at two examples "Apoe", "Pecam1" in glm model  
y$genes[y$genes$Symbol=="Apoe",]

genes Symbol Entrezid Genename  
ENSMUSG00000002985 ENSMUSG00000002985 Apoe 11816 apolipoprotein E

glm.fit[["ENSMUSG00000002985"]]

Call:  
lm(formula = . ~ duration)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-139.56 -84.47 7.05 49.36 284.98   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)  
(Intercept) -18.7 45.4 -0.41 0.68  
duration4weeks 34.1 64.1 0.53 0.60  
duration6weeks 62.7 64.1 0.98 0.34  
duration8weeks 66.7 64.1 1.04 0.31  
  
Residual standard error: 111 on 20 degrees of freedom  
Multiple R-squared: 0.0649, Adjusted R-squared: -0.0753   
F-statistic: 0.463 on 3 and 20 DF, p-value: 0.711

y$genes[y$genes$Symbol=="Pecam1",]

genes Symbol Entrezid  
ENSMUSG00000020717 ENSMUSG00000020717 Pecam1 18613  
 Genename  
ENSMUSG00000020717 platelet/endothelial cell adhesion molecule 1

glm.fit[["ENSMUSG00000020717"]]

Call:  
lm(formula = . ~ duration)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-5.461 -1.967 -0.402 2.049 6.151   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)  
(Intercept) 1.2814 1.3058 0.98 0.34  
duration4weeks 0.0896 1.8466 0.05 0.96  
duration6weeks -0.3399 1.8466 -0.18 0.86  
duration8weeks 0.0125 1.8466 0.01 0.99  
  
Residual standard error: 3.2 on 20 degrees of freedom  
Multiple R-squared: 0.0032, Adjusted R-squared: -0.146   
F-statistic: 0.0214 on 3 and 20 DF, p-value: 0.996

## Obtain the statistics from GLM

# Use home-made pipeline stat\_table\_all to:   
source("~/Documents/code/01\_function/my\_GLM.R")   
# 1. retrieve the R-square(r2) and the corrected p value (q) from the fit summary of each gene  
# 2. filter genes based on qval<0.05, r2>0.5, arrange based on q (ascending).  
# 3. from filtered genes, retrieve coefficient of estimate and p val for each gene for down stream analysis.  
stat\_table\_all <- stat\_table(glm.fit)  
dim(stat\_table\_all)

[1] 906 12

head(stat\_table\_all)

ENSMUSG\_ID r2 pval qval  
ENSMUSG00000020178 ENSMUSG00000020178 0.972 1.01e-15 1.45e-11  
ENSMUSG00000038805 ENSMUSG00000038805 0.960 3.87e-14 2.77e-10  
ENSMUSG00000068696 ENSMUSG00000068696 0.957 7.22e-14 3.45e-10  
ENSMUSG00000071234 ENSMUSG00000071234 0.953 1.71e-13 6.14e-10  
ENSMUSG00000023274 ENSMUSG00000023274 0.942 1.45e-12 4.17e-09  
ENSMUSG00000021919 ENSMUSG00000021919 0.936 4.40e-12 7.88e-09  
 X.Intercept.\_est duration4weeks\_est duration6weeks\_est  
ENSMUSG00000020178 0.3734 -0.7145 -11.219  
ENSMUSG00000038805 0.3852 -0.3049 -3.326  
ENSMUSG00000068696 7.4938 -12.7602 -47.431  
ENSMUSG00000071234 0.5603 -0.9227 -7.543  
ENSMUSG00000023274 0.2246 -0.1137 -2.774  
ENSMUSG00000021919 -0.0182 -0.0847 -0.644  
 duration8weeks\_est X.Intercept.\_pval  
ENSMUSG00000020178 -0.56308 0.304251  
ENSMUSG00000038805 -0.30381 0.005591  
ENSMUSG00000068696 -8.22708 0.000285  
ENSMUSG00000071234 -0.52783 0.079856  
ENSMUSG00000023274 -0.36346 0.090060  
ENSMUSG00000021919 0.00116 0.570175  
 duration4weeks\_pval duration6weeks\_pval  
ENSMUSG00000020178 0.169043 1.22e-15  
ENSMUSG00000038805 0.097636 3.00e-14  
ENSMUSG00000068696 0.000036 1.53e-14  
ENSMUSG00000071234 0.044100 1.27e-13  
ENSMUSG00000023274 0.531068 1.22e-12  
ENSMUSG00000021919 0.072068 4.93e-12  
 duration8weeks\_pval  
ENSMUSG00000020178 0.2741  
ENSMUSG00000038805 0.0988  
ENSMUSG00000068696 0.0028  
ENSMUSG00000071234 0.2333  
ENSMUSG00000023274 0.0549  
ENSMUSG00000021919 0.9794

# add the gene names to stat\_table\_all   
stat\_table\_all<-left\_join(stat\_table\_all, y$genes, by=c("ENSMUSG\_ID"="genes"))  
dim(stat\_table\_all)

[1] 906 15

head(stat\_table\_all)

ENSMUSG\_ID r2 pval qval X.Intercept.\_est  
1 ENSMUSG00000020178 0.972 1.01e-15 1.45e-11 0.3734  
2 ENSMUSG00000038805 0.960 3.87e-14 2.77e-10 0.3852  
3 ENSMUSG00000068696 0.957 7.22e-14 3.45e-10 7.4938  
4 ENSMUSG00000071234 0.953 1.71e-13 6.14e-10 0.5603  
5 ENSMUSG00000023274 0.942 1.45e-12 4.17e-09 0.2246  
6 ENSMUSG00000021919 0.936 4.40e-12 7.88e-09 -0.0182  
 duration4weeks\_est duration6weeks\_est duration8weeks\_est  
1 -0.7145 -11.219 -0.56308  
2 -0.3049 -3.326 -0.30381  
3 -12.7602 -47.431 -8.22708  
4 -0.9227 -7.543 -0.52783  
5 -0.1137 -2.774 -0.36346  
6 -0.0847 -0.644 0.00116  
 X.Intercept.\_pval duration4weeks\_pval duration6weeks\_pval  
1 0.304251 0.169043 1.22e-15  
2 0.005591 0.097636 3.00e-14  
3 0.000285 0.000036 1.53e-14  
4 0.079856 0.044100 1.27e-13  
5 0.090060 0.531068 1.22e-12  
6 0.570175 0.072068 4.93e-12  
 duration8weeks\_pval Symbol Entrezid  
1 0.2741 Adora2a 11540  
2 0.0988 Six3 20473  
3 0.0028 Gpr88 64378  
4 0.2333 Syndig1l 627191  
5 0.0549 Cd4 12504  
6 0.9794 Chat 12647  
 Genename  
1 adenosine A2a receptor  
2 sine oculis-related homeobox 3  
3 G-protein coupled receptor 88  
4 synapse differentiation inducing 1 like  
5 CD4 antigen  
6 choline acetyltransferase

# export the gene list table for Kegg and GO search  
# write.table(stat\_table\_all, paste(out\_analysis, out\_prefix, "\_gene\_list.txt", sep=""), sep="\t", row=F, quote=F)  
  
# export the background gene list for Kegg and GO search, subtract genes of stat\_table\_all from y$genes  
bg\_gene\_list <- y$genes   
dim(bg\_gene\_list)

[1] 14329 4

#write.table(bg\_gene\_list, paste(out\_analysis, out\_prefix, "\_bg\_gene.txt", sep=""), sep="\t", row=F, quote=F)

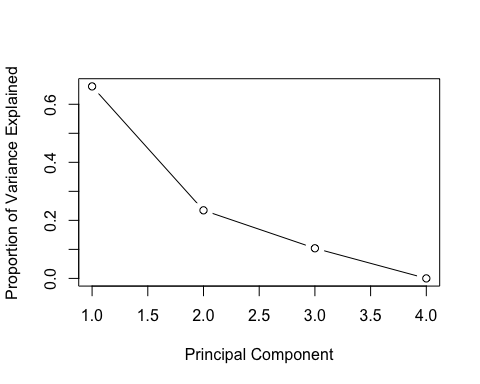
## Explore the GLM model

1. PCA of the coefficient matrix (row: genes; colums: coefficients)
2. heatmap of the coefficient matrix
3. line plot of all the coefficients based on each gene
4. VennDiagram plotting of genes and overlap of genes that are significant changed under each coefficient catagory
5. Tile plot showing the pattern of coeffiecint

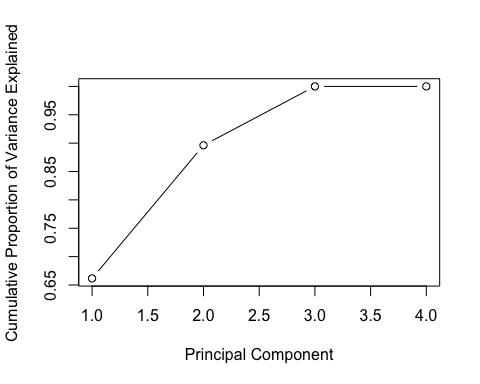
# principal component analysis of estimate of the coefficient  
stat\_est <- stat\_table\_all %>% select(contains("est"))  
head(stat\_est)

X.Intercept.\_est duration4weeks\_est duration6weeks\_est  
1 0.3734 -0.7145 -11.219  
2 0.3852 -0.3049 -3.326  
3 7.4938 -12.7602 -47.431  
4 0.5603 -0.9227 -7.543  
5 0.2246 -0.1137 -2.774  
6 -0.0182 -0.0847 -0.644  
 duration8weeks\_est  
1 -0.56308  
2 -0.30381  
3 -8.22708  
4 -0.52783  
5 -0.36346  
6 0.00116

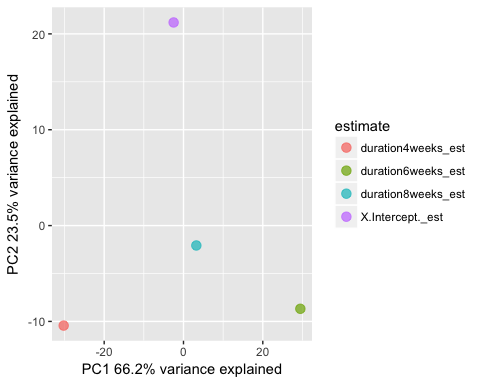
pca\_original = prcomp(t(stat\_est),scale=T, center=T)  
pca\_x <- data.frame(pca\_original$x, estimate = rownames(pca\_original$x))  
pca\_table <- pca\_x  
x <- pca\_original$sdev^2/sum(pca\_original$sdev^2) # Proportion of Variance Explained for all components  
  
# Scree plot  
plot(x, xlab="Principal Component", ylab="Proportion of Variance Explained", type="b")



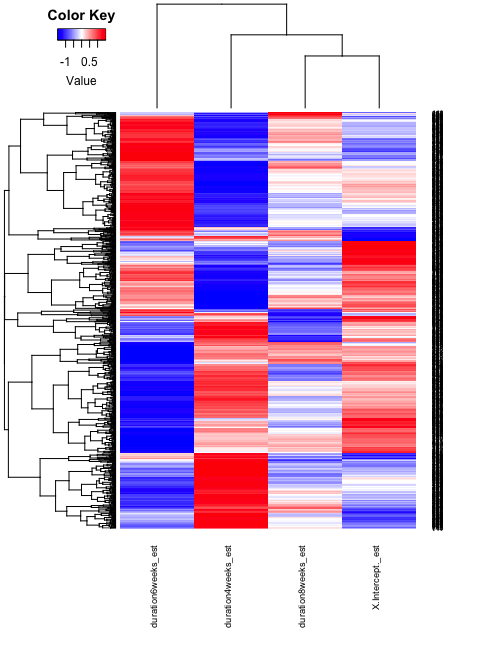
plot(cumsum(x), xlab="Principal Component", ylab="Cumulative Proportion of Variance Explained", type="b")



# PCA plot  
g <- ggplot(pca\_table, aes(x=pca\_table[1], y=pca\_table[2], color= estimate))  
 g <- g + geom\_point(alpha=0.7, size=3)   
 # g <- g + labs(color = "Group", shape="Tissue")  
 g + labs(x = paste(names(pca\_table[1]), scales::percent(x[1]),"variance explained", sep=" "), y=paste(names(pca\_table[2]), scales::percent(x[2]),"variance explained", sep=" "))



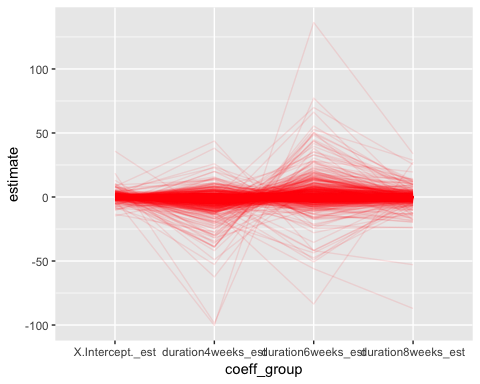
# hirachical analysis of estimate of the coefficient  
stat\_est\_h<- t(scale(t(stat\_est))) %>% as.data.frame() %>% as.matrix()  
   
library(gplots)  
col.pan <- colorpanel(100, "blue", "white", "red")  
heatmap.2(stat\_est\_h, col=col.pan, Rowv=TRUE, scale="none",   
 trace="none", dendrogram="both", cexRow=0.5, cexCol=0.7, density.info="none",  
 margin=c(9,4), lhei=c(2,10), lwid=c(2,6))



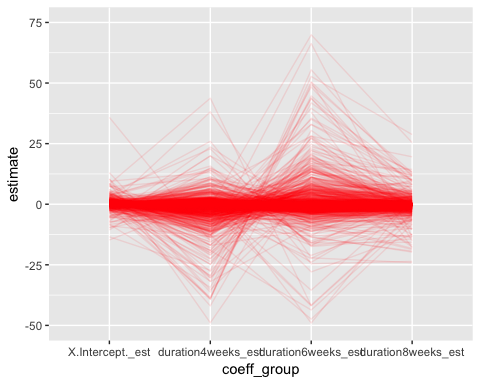
# plot the estimate of coefficient of each gene  
stat\_table\_est.gather <- stat\_table\_all %>%   
 select("ENSMUSG\_ID", contains("\_est")) %>%   
 gather(key=coeff\_group, value=estimate, -ENSMUSG\_ID)  
str(stat\_table\_est.gather)

'data.frame': 3624 obs. of 3 variables:  
 $ ENSMUSG\_ID : chr "ENSMUSG00000020178" "ENSMUSG00000038805" "ENSMUSG00000068696" "ENSMUSG00000071234" ...  
 $ coeff\_group: chr "X.Intercept.\_est" "X.Intercept.\_est" "X.Intercept.\_est" "X.Intercept.\_est" ...  
 $ estimate : num 0.373 0.385 7.494 0.56 0.225 ...

stat\_table\_est.gather$coeff\_group <- factor(stat\_table\_est.gather$coeff\_group,  
 levels =c("X.Intercept.\_est", "duration4weeks\_est", "duration6weeks\_est","duration8weeks\_est"),  
 ordered = TRUE)  
ggplot(stat\_table\_est.gather, aes(x=coeff\_group,y=estimate, group=ENSMUSG\_ID)) +   
 geom\_line(alpha=0.1, color="red")



ggplot(stat\_table\_est.gather, aes(x=coeff\_group,y=estimate, group=ENSMUSG\_ID))+   
 geom\_line(alpha=0.1, color="red") + scale\_y\_continuous(limits= c(-50, 75))



# To identify each gene’s critical predictor. binarized the estimate and significance matrices by hard cutoffs.  
stat\_pval <- stat\_table\_all %>% select(contains("\_pval"))  
head(stat\_pval)

X.Intercept.\_pval duration4weeks\_pval duration6weeks\_pval  
1 0.304251 0.169043 1.22e-15  
2 0.005591 0.097636 3.00e-14  
3 0.000285 0.000036 1.53e-14  
4 0.079856 0.044100 1.27e-13  
5 0.090060 0.531068 1.22e-12  
6 0.570175 0.072068 4.93e-12  
 duration8weeks\_pval  
1 0.2741  
2 0.0988  
3 0.0028  
4 0.2333  
5 0.0549  
6 0.9794

head(stat\_est)

X.Intercept.\_est duration4weeks\_est duration6weeks\_est  
1 0.3734 -0.7145 -11.219  
2 0.3852 -0.3049 -3.326  
3 7.4938 -12.7602 -47.431  
4 0.5603 -0.9227 -7.543  
5 0.2246 -0.1137 -2.774  
6 -0.0182 -0.0847 -0.644  
 duration8weeks\_est  
1 -0.56308  
2 -0.30381  
3 -8.22708  
4 -0.52783  
5 -0.36346  
6 0.00116

stat\_binary <- (apply(stat\_pval, 2, function(x) x<0.05) & apply(stat\_est, 2, function(x) abs(x)>0.2)) %>% as.data.frame()  
rowSums(stat\_binary) %>% table()

.  
 0 1 2 3 4   
 1 435 334 96 40

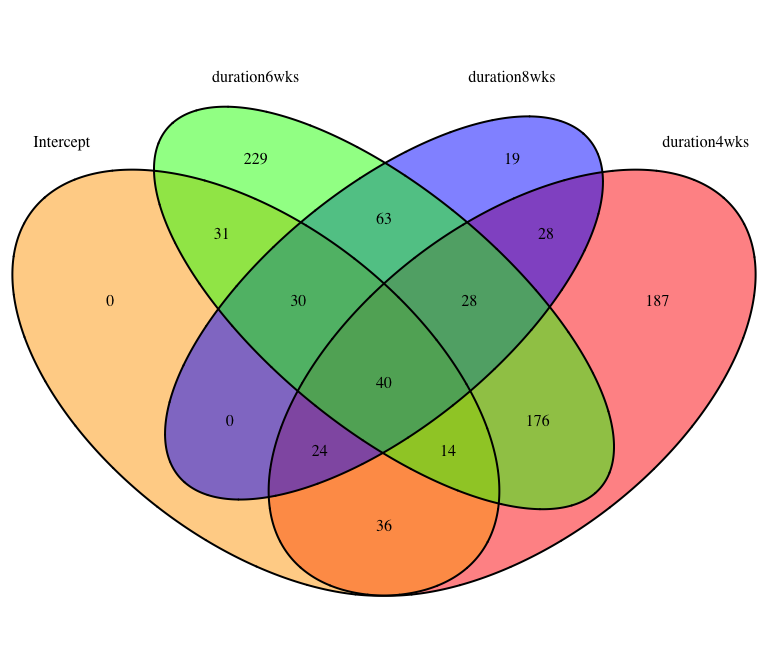
# summary of the numbers of genes passes the filter  
apply(stat\_binary, 2, table)

X.Intercept.\_pval duration4weeks\_pval duration6weeks\_pval  
FALSE 731 373 295  
TRUE 175 533 611  
 duration8weeks\_pval  
FALSE 674  
TRUE 232

# sepertate out genes by their coefficients  
col <- as.list(names(stat\_binary))  
genes\_by\_coeff <- map(col, ~ stat\_table\_all[["Symbol"]][stat\_binary[[.]]])  
names(genes\_by\_coeff) <- c("Intercept", "duration4wks", "duration6wks", "duration8wks")  
glimpse(genes\_by\_coeff)

List of 4  
 $ Intercept : chr [1:175] "Six3" "Gpr88" "Drd2" "Gpr6" ...  
 $ duration4wks: chr [1:533] "Gpr88" "Syndig1l" "Tacr1" "Ptpn9" ...  
 $ duration6wks: chr [1:611] "Adora2a" "Six3" "Gpr88" "Syndig1l" ...  
 $ duration8wks: chr [1:232] "Gpr88" "Ido1" "Adck2" "Fos" ...

# to check the overlap of genes in different coefficient catagories  
library(VennDiagram)  
grid.newpage()  
within(genes\_by\_coeff,  
 draw.quad.venn(area1 = length(Intercept),   
 area2 = length(duration4wks),   
 area3 = length(duration6wks),   
 area4 = length(duration8wks),  
 n12 = length(intersect(Intercept, duration4wks)),   
 n13 = length(intersect(Intercept, duration6wks)),  
 n14 = length(intersect(Intercept, duration8wks)),  
 n23 = length(intersect(duration4wks, duration6wks)),  
 n24 = length(intersect(duration4wks, duration8wks)),  
 n34 = length(intersect(duration6wks, duration8wks)),  
 n123 = Intercept %>% intersect(duration4wks) %>% intersect(duration6wks) %>% length(),  
 n124 = Intercept %>% intersect(duration4wks) %>% intersect(duration8wks) %>% length(),  
 n134 = Intercept %>% intersect(duration6wks) %>% intersect(duration8wks) %>% length(),  
 n234 = duration4wks %>% intersect(duration6wks) %>% intersect(duration8wks) %>% length(),  
 n1234 = Intercept %>% intersect(duration4wks) %>% intersect(duration6wks) %>% intersect(duration8wks) %>% length(),  
 category = names(genes\_by\_coeff),  
 fill = c("orange", "red", "green", "blue"))  
 )

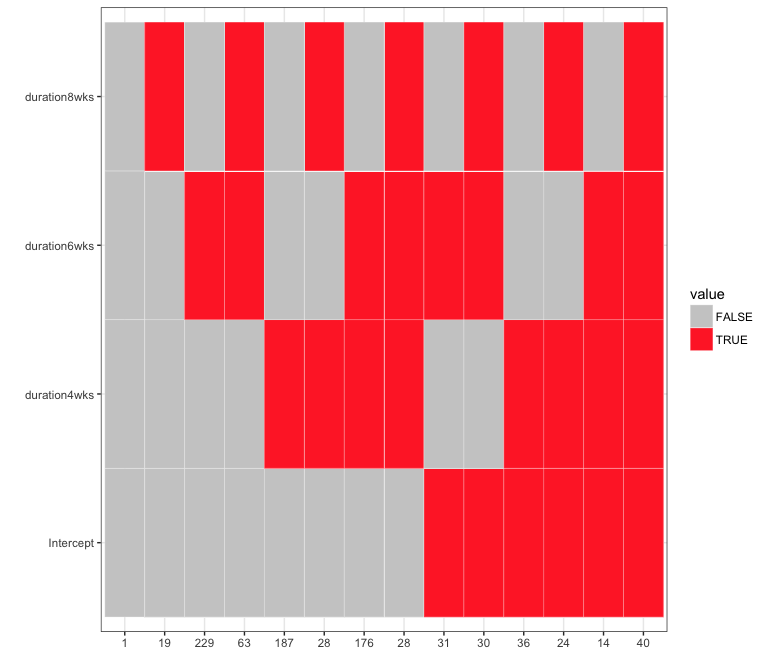


$Intercept  
 [1] "Six3" "Gpr88" "Drd2" "Gpr6"   
 [5] "Slc35d3" "Klhl13" "Ldlr" "Nexn"   
 [9] "4930432K21Rik" "Rasgrp2" "Tmem200a" "Hrh1"   
 [13] "Shq1" "Scn5a" "Ofd1" "Ybey"   
 [17] "3632451O06Rik" "Fam177a" "Lpcat2" "Gpatch11"   
 [21] "Fam111a" "Mtrf1" "Nectin4" "Phf7"   
 [25] "Lrrc49" "Diexf" "Micall2" "Arhgap22"   
 [29] "Insig1" "Mme" "Arl5c" "C77080"   
 [33] "Bcl6b" "Ppfibp2" "Zfp804a" "Nr1d1"   
 [37] "Ptcd1" "Anapc4" "Ddx27" "Slc22a3"   
 [41] "Dtwd2" "Pik3c3" "Vipr1" "Gm5083"   
 [45] "Rars2" "Tcerg1l" "E130102H24Rik" "Zfp930"   
 [49] "Ankra2" "Nanos2" "Gtf2h1" "B3gnt9"   
 [53] "Ttc34" "Tada1" "Ccdc183" "Pex13"   
 [57] "Samd3" "Fam53b" "Tigd2" "Herpud2"   
 [61] "Rasd1" "Trim17" "Rbm43" "Pcsk9"   
 [65] "Mta2" "Dbr1" "Dhx32" "Rac2"   
 [69] "4930513N10Rik" "Gpt" "Cog6" "Pum3"   
 [73] "Runx2" "Nrap" "Nub1" "Gart"   
 [77] "Wrb" "Sdhaf3" "Zfp791" "Zwilch"   
 [81] "Mustn1" "Ubxn4" "Bbs9" "Smad7"   
 [85] "Zkscan6" "Oxnad1" "Glyctk" "Rassf6"   
 [89] "Arhgap27" "Nat8f5" "Dusp7" "St7l"   
 [93] "Pced1b" "Cdkal1" "2700046G09Rik" "Ufd1"   
 [97] "Il21r" "Card19" "Eif4e2" "Gemin8"   
[101] "Gm11696" "6430550D23Rik" "Fra10ac1" "Pdhx"   
[105] "Cbx8" "Iqcc" "Kcna5" "Znrf1"   
[109] "P4ha3" "Hspa9" "Rabgef1" "Slc52a3"   
[113] "Zdhhc15" "2310057M21Rik" "Cys1" "Hvcn1"   
[117] "Cyb5d2" "Iba57" "Arntl" "Tpm1"   
[121] "Top1" "Otx1" "Trim26" "Agk"   
[125] "Meltf" "Sh3yl1" "Rerg" "Zbed4"   
[129] "Rasip1" "Kcne1l" "Lfng" "Met"   
[133] "Hpdl" "Klhdc1" "Mettl25" "Tecpr1"   
[137] "Snx2" "Ccdc181" "Ccl6" "Cmtm3"   
[141] "Thtpa" "Adprm" "BC030500" "Zc3h8"   
[145] "Sox18" "Slc9a9" "Mob3c" "Atpaf1"   
[149] "Slc35g2" "Wdr92" "Cbwd1" "Hic2"   
[153] "Isg15" "6330409D20Rik" "Ucp3" "Mettl2"   
[157] "Atg14" "C530005A16Rik" "1810010H24Rik" "Sema5b"   
[161] "Aggf1" "Cutc" "LOC100861615" "Ephb3"   
[165] "Ubc" "Araf" "Plrg1" "Aspa"   
[169] "Zfp7" "Zfp202" "Cluap1" "Inip"   
[173] "Lhx2" "Tgfa" "Ttc26"   
  
$duration4wks  
 [1] "Gpr88" "Syndig1l" "Tacr1" "Ptpn9"   
 [5] "Adck2" "Inhba" "Rxrg" "Trib2"   
 [9] "Banp" "Oprk1" "Fos" "Sh2d3c"   
 [13] "Ldlr" "4930432K21Rik" "Nuak1" "Spred2"   
 [17] "Gfod1" "Smad3" "Cln6" "Cwc25"   
 [21] "Cdyl2" "Naglu" "Lrrc61" "Tmem200a"   
 [25] "Casp7" "Dll4" "Rilpl1" "Dusp4"   
 [29] "Junb" "Mccc2" "Schip1" "Dnajc21"   
 [33] "Ier2" "Spata5" "Hrh1" "Shq1"   
 [37] "Erf" "Rrad" "Scn5a" "Ofd1"   
 [41] "Cbfa2t3" "Ybey" "St8sia5" "Heyl"   
 [45] "Mical2" "Ccdc112" "Mamld1" "3632451O06Rik"  
 [49] "Sgsm1" "Myadml2" "Zfp459" "Fam177a"   
 [53] "Fmnl1" "Tdp1" "Pfkfb3" "Bhlhe40"   
 [57] "Hyi" "Zfand3" "Sco1" "Cc2d1b"   
 [61] "Lrfn2" "Lig4" "Lpcat2" "Mfap3"   
 [65] "Pdp1" "Cd180" "Cnot4" "Gpatch11"   
 [69] "Fam111a" "Slc10a7" "Plcl2" "S1pr2"   
 [73] "Smg9" "Bdnf" "B4galt4" "Mtrf1"   
 [77] "Heatr3" "Trpm4" "Mark3" "Nfatc2"   
 [81] "Ngfr" "Nectin4" "Htr1b" "Phf7"   
 [85] "Lrrc49" "Crem" "Ipmk" "Zfyve1"   
 [89] "Glrx2" "4930453N24Rik" "Ccnf" "Trim68"   
 [93] "Micall2" "Arhgap22" "Blzf1" "Insig1"   
 [97] "Dusp16" "Arl5c" "Zfand4" "Tefm"   
[101] "Jun" "Bcl6b" "Slc25a35" "Gtf2ird1"   
[105] "Fam212b" "Mex3d" "A930017K11Rik" "Ccdc6"   
[109] "Plekha3" "Ppfibp2" "Phf13" "Grip1"   
[113] "Ilvbl" "Armc5" "E2f6" "Fam222a"   
[117] "Synpo" "Htr6" "Ciart" "Ptcd1"   
[121] "Frmd3" "Anapc4" "Zbtb12" "Lhfp"   
[125] "Actr5" "Zmynd8" "Slc22a3" "Hunk"   
[129] "Lmcd1" "Gpr83" "Zscan22" "Pms1"   
[133] "Dusp8" "Chaf1a" "Hrh2" "Plcd1"   
[137] "Mmp24" "Mycn" "Gpr1" "Pik3c3"   
[141] "Thap1" "Gm20878" "Ankrd33b" "Snap29"   
[145] "Layn" "Gm5083" "Ddx50" "Pla2g4e"   
[149] "Zfp366" "Prkag2" "Abcb1b" "Tcerg1l"   
[153] "Zfp930" "Ppm1d" "Nanos2" "Rassf1"   
[157] "Gtf2h1" "Tfdp1" "Pou6f2" "Cacng4"   
[161] "Cables1" "Gpr68" "Uhrf2" "B3gnt9"   
[165] "A430108G06Rik" "Map3k21" "LOC100862446" "Gpr137b"   
[169] "Homer1" "Tada1" "Fam117a" "Rasl11b"   
[173] "Cbln2" "Ccdc183" "Hist1h2ao" "Narf"   
[177] "P4ha1" "Cnep1r1" "Pex13" "Abhd6"   
[181] "Ubox5" "Irs2" "4931440F15Rik" "Nck1"   
[185] "Kbtbd2" "Bend3" "2410002F23Rik" "Tigd2"   
[189] "Mapk14" "Zbtb40" "Camk1g" "Rhobtb2"   
[193] "Lsm11" "Adam19" "Ephb2" "St7"   
[197] "Rhob" "1700029J07Rik" "Rnd3" "Tgfb3"   
[201] "Lrrc29" "Pop1" "Herpud2" "Arhgef7"   
[205] "Rasd1" "Rtel1" "Prdm4" "Mettl3"   
[209] "Kcnf1" "Wisp1" "Tiparp" "Lss"   
[213] "Dvl2" "Rbm43" "Wdcp" "Cd300c2"   
[217] "Jade2" "Prickle1" "Dhx32" "Abcc10"   
[221] "Slc35f3" "4930513N10Rik" "Dusp14" "Sgms1"   
[225] "Pafah2" "Plekho2" "Serinc2" "Cog6"   
[229] "Pum3" "Sema4c" "Chadl" "Hmox1"   
[233] "Brix1" "Apold1" "Dars" "Sf3a3"   
[237] "Cpne2" "Fbn1" "Usp18" "Scpep1"   
[241] "Dnttip2" "Cd302" "Adamts1" "Zdhhc5"   
[245] "Foxred1" "Ubxn10" "Xbp1" "Cited2"   
[249] "Ltv1" "Bcor" "Fam171a1" "Nrap"   
[253] "Nub1" "Rad9b" "Gart" "Zfp438"   
[257] "2210408I21Rik" "Wrb" "Nr4a1" "Zfp180"   
[261] "Mtmr4" "Ppp2r5a" "Nop2" "Sdhaf3"   
[265] "Sh2d5" "Galns" "Zfp791" "Arhgef3"   
[269] "Hps1" "Calhm2" "D630045J12Rik" "Zwilch"   
[273] "Mustn1" "Zswim4" "Hspa12b" "Egr2"   
[277] "Rfwd2" "Zfp316" "Slfn2" "Rnf4"   
[281] "Ubxn4" "Usp43" "Hexim2" "Ppp1r13l"   
[285] "Erp44" "Man2b2" "Ccl9" "Smad7"   
[289] "Vegfc" "Ric8b" "Cfap70" "Tmem220"   
[293] "Aagab" "F2rl2" "Exosc3" "Gspt2"   
[297] "Sh3rf3" "Etv5" "Xylb" "Ldlrad3"   
[301] "Glyctk" "Srp68" "Snhg5" "Actr8"   
[305] "Alox12b" "Arhgap27" "Nat8f5" "Nol10"   
[309] "Crnkl1" "Smu1" "Nkrf" "Nxpe4"   
[313] "Tmem100" "St7l" "Pgm1" "Mak16"   
[317] "Atp8b2" "Cep131" "Wdsub1" "Nacad"   
[321] "Sord" "Golt1b" "Pced1b" "Tmem121b"   
[325] "Cdkal1" "Zfp790" "Tiam2" "Adcyap1"   
[329] "Etv6" "Creb5" "Fstl3" "2700046G09Rik"  
[333] "Ufd1" "Adra1d" "Hnrnpf" "Naa35"   
[337] "Smg8" "Naa25" "Rundc1" "Ndufaf4"   
[341] "Ints5" "Tlr9" "Pdyn" "Tox4"   
[345] "Card19" "Eif4e2" "Srpk3" "Ythdf2"   
[349] "Gldn" "Mycl" "6430550D23Rik" "Ido2"   
[353] "Mybpc1" "Rgs11" "Irak1" "Rrs1"   
[357] "Iqgap2" "Il17ra" "E2f4" "Pdhx"   
[361] "Cbx8" "Cyr61" "Zcwpw1" "Atp8b1"   
[365] "Kcna5" "Mus81" "Plpbp" "Fam46a"   
[369] "Olig2" "Slc2a9" "P4ha3" "Hspa9"   
[373] "Rabgef1" "Ggnbp2" "Slc52a3" "Zdhhc15"   
[377] "2310057M21Rik" "Stx3" "Tspyl2" "Asb4"   
[381] "Cys1" "Hvcn1" "Top3a" "Smad2"   
[385] "Rhbdd1" "Cercam" "Zfp687" "Axin1"   
[389] "Uck2" "Krcc1" "Ddx10" "Idua"   
[393] "Trim16" "Mis12" "Lrrtm2" "Iba57"   
[397] "Arntl" "Fancc" "Slc9a8" "Ldlrap1"   
[401] "Neurl2" "Tpm1" "Egr3" "Pdzrn4"   
[405] "Rtn4rl2" "Top1" "Otx1" "Rassf5"   
[409] "Agk" "Rad17" "Meltf" "Sh3yl1"   
[413] "1700034H15Rik" "Rin1" "Surf6" "Srarp"   
[417] "Kcnip3" "Rasip1" "Lfng" "Loxl3"   
[421] "Ddr1" "Zufsp" "Plaa" "Dhx16"   
[425] "Met" "Fam84a" "Inafm2" "Jdp2"   
[429] "Ercc8" "Hlcs" "Sgsh" "4732491K20Rik"  
[433] "Nfe2l3" "Mpp3" "Eed" "Mettl25"   
[437] "Oas1c" "Wdr91" "Tecpr1" "Mphosph10"   
[441] "Mthfr" "Zfp874b" "Snx2" "Tob2"   
[445] "Ulk3" "Med26" "Gtpbp4" "Gramd1b"   
[449] "Zfp90" "Akap7" "Dhx8" "Sema6c"   
[453] "Serinc3" "Ccl6" "Prpf3" "Fkbp5"   
[457] "Thtpa" "Cdyl" "Trim9" "Fam110b"   
[461] "BC030500" "Ints8" "Mob3c" "Lrrc57"   
[465] "Atpaf1" "Slc35g2" "B230206H07Rik" "Klhl8"   
[469] "Wdr92" "Cbwd1" "Rpusd2" "Zranb1"   
[473] "Isg15" "Unc45a" "Epha2" "Irf8"   
[477] "Moap1" "6330409D20Rik" "Tmem109" "Tbx2"   
[481] "Ucp3" "Clec4a2" "Cdkn1a" "B4galt5"   
[485] "Atg14" "Fam155a" "Kansl2" "1810010H24Rik"  
[489] "Rars" "D17H6S53E" "Cyb5rl" "Sema5b"   
[493] "Mrpl50" "Arf6" "Aggf1" "Gimap1"   
[497] "Mboat1" "Cnot9" "Rrm2" "Gdpd5"   
[501] "BC034090" "Rbm19" "Srfbp1" "Sowahb"   
[505] "Araf" "Plrg1" "Mcam" "Aspa"   
[509] "Zfp7" "Synj2" "Pced1a" "Sec23ip"   
[513] "Cluap1" "Zbtb17" "Scg2" "Gm960"   
[517] "Gm11549" "Fndc8" "Inip" "Arsg"   
[521] "Chsy1" "Tbcd" "Tmem8" "Gimap6"   
[525] "Tmem200b" "Spidr" "Tgfa" "Ttc26"   
[529] "Cactin" "Nup43" "Pcsk1" "Rnpepl1"   
[533] "Car14"   
  
$duration6wks  
 [1] "Adora2a" "Six3" "Gpr88" "Syndig1l"   
 [5] "Cd4" "Chat" "Isl1" "Penk"   
 [9] "Trh" "Drd2" "Gpr6" "Sh3rf2"   
 [13] "Rgs9" "Tacr1" "Htr1d" "Slc35d3"   
 [17] "Lrrc10b" "Drd1" "Slc5a7" "Slc10a4"   
 [21] "Ido1" "Ptpn9" "Ppp1r1b" "Klhl13"   
 [25] "Adck2" "Inhba" "Pde1b" "Rxrg"   
 [29] "Trib2" "Banp" "Oprk1" "Fos"   
 [33] "Gng7" "Sh2d3c" "Rarb" "Ldlr"   
 [37] "Nexn" "Serpina9" "Pde10a" "Six3os1"   
 [41] "4930432K21Rik" "Nuak1" "Spred2" "Dach1"   
 [45] "Gfod1" "Numb" "Rasgrp2" "Smad3"   
 [49] "Adcy5" "Cln6" "Gnb4" "Cwc25"   
 [53] "Cdyl2" "Asic4" "Naglu" "Lrrc61"   
 [57] "Ecel1" "Per2" "Casp7" "Dll4"   
 [61] "Rilpl1" "Dusp4" "Junb" "Mccc2"   
 [65] "Schip1" "Dnajc21" "Ier2" "Sik2"   
 [69] "Spata5" "Hrh1" "Shq1" "Slco5a1"   
 [73] "Scn5a" "Ofd1" "Cbfa2t3" "Rtn4ip1"   
 [77] "Ybey" "St8sia5" "Heyl" "Mical2"   
 [81] "Ccdc112" "Mamld1" "3632451O06Rik" "Prag1"   
 [85] "Baz1a" "Gpr149" "Sgsm1" "Rbm11"   
 [89] "Myadml2" "B3galt6" "Zfp459" "Fam177a"   
 [93] "Fmnl1" "Tdp1" "Pfkfb3" "Bhlhe40"   
 [97] "Igf2bp2" "Osbpl3" "Zfand3" "Sco1"   
[101] "Ttc22" "Slc37a1" "Cc2d1b" "Lrfn2"   
[105] "Lig4" "Mfap3" "Pdp1" "Cd180"   
[109] "Ptgs2" "Cnot4" "Gpatch11" "Bace2"   
[113] "Slc10a7" "Plcl2" "Ninl" "Irf2"   
[117] "Bdnf" "B4galt4" "Mtrf1" "Trim45"   
[121] "Trpm4" "Mark3" "Nfatc2" "Ngfr"   
[125] "Nectin4" "Lrrc49" "4931415C17Rik" "Crem"   
[129] "Hspbap1" "Csrnp2" "Itga9" "Ipmk"   
[133] "Zfyve1" "Diexf" "Dlk1" "Magel2"   
[137] "Glrx2" "Tipin" "Phf21b" "Ccnf"   
[141] "Trim68" "Pim1" "Arhgap39" "Gramd4"   
[145] "Mme" "Dusp16" "Arl5c" "Zfand4"   
[149] "Zc3h6" "Tefm" "C77080" "Jun"   
[153] "Slc25a35" "Fbxl4" "Oacyl" "Fam212b"   
[157] "Dnajc1" "Ccdc6" "Plekha3" "Plppr1"   
[161] "Pxdn" "Phf13" "Grip1" "Zfp804a"   
[165] "B3gat2" "2810029C07Rik" "Nr1d1" "Ilvbl"   
[169] "Armc5" "E2f6" "Dgkk" "Fam222a"   
[173] "P4ha2" "Synpo" "Ptcd1" "Frmd3"   
[177] "Ahdc1" "Parn" "Mcf2l" "Kdm4d"   
[181] "Lhfp" "Ddx27" "Zmynd8" "Gpr21"   
[185] "Slc22a3" "Ehhadh" "Hunk" "Zfp941"   
[189] "Gpr83" "Pms1" "Dtwd2" "Myo5b"   
[193] "Prr5" "Chaf1a" "Slc35f4" "Plin4"   
[197] "Mmp24" "Lzts1" "Gpr1" "Pik3c3"   
[201] "Gm20878" "Ankrd33b" "Snap29" "Rnf165"   
[205] "Rin3" "Vipr1" "Gm5083" "Ddx50"   
[209] "Pla2g4e" "Zfp366" "Prkag2" "Abcb1b"   
[213] "Sp9" "Lrrc56" "Rars2" "Ccdc103"   
[217] "Tcerg1l" "E130102H24Rik" "Mapk4" "Ankra2"   
[221] "Fzd1" "Ppm1d" "Rassf1" "Cacng4"   
[225] "Cables1" "Gpr68" "Uhrf2" "Mall"   
[229] "Epm2a" "Map3k21" "Ttc34" "LOC100862446"   
[233] "Gpr137b" "Homer1" "Fam117a" "Cbln2"   
[237] "Ccdc183" "P4ha1" "Usb1" "Plekhg5"   
[241] "Mkl1" "Abhd6" "Ubox5" "Prkab1"   
[245] "Irs2" "Dusp1" "Mak" "Kbtbd2"   
[249] "Bend3" "Samd3" "Fam53b" "Zmym1"   
[253] "Tigd2" "Mapk14" "Zbtb40" "Camk1g"   
[257] "Rhobtb2" "Lsm11" "Ephb2" "St7"   
[261] "Rhob" "1700029J07Rik" "6430571L13Rik" "Rnd3"   
[265] "Pop1" "Dio2" "Arhgef7" "Cd276"   
[269] "Rasd1" "Blm" "Trim62" "Asap2"   
[273] "Cftr" "Omg" "Wisp1" "Ift52"   
[277] "Ptpn1" "Gabrq" "Trim17" "Synpr"   
[281] "Lss" "Actn2" "Meis1" "Galnt7"   
[285] "Rbm43" "Pcsk9" "Wdcp" "Ammecr1"   
[289] "Mta2" "Jade2" "Dbr1" "Prickle1"   
[293] "Dhx32" "Rac2" "4930513N10Rik" "Haus3"   
[297] "Gpt" "Dusp14" "Sgms1" "Strip2"   
[301] "Pafah2" "Serinc2" "Cdh9" "Sema4c"   
[305] "Adamts3" "Hmox1" "Gk5" "Apold1"   
[309] "Morc2b" "Dars" "Nadk" "Cpne2"   
[313] "Runx2" "Fbn1" "Cd302" "Adamts1"   
[317] "Ubxn10" "Xbp1" "Pcp4l1" "Bcor"   
[321] "Fam171a1" "Nub1" "Anxa3" "Bbs2"   
[325] "Gart" "Lrsam1" "Sdr42e1" "Zfp438"   
[329] "Spry2" "Wrb" "Traf3" "Nr4a1"   
[333] "Nrde2" "Mfhas1" "Gjb6" "Zfp180"   
[337] "Mtmr4" "Ppp2r5a" "Sdhaf3" "Sh2d5"   
[341] "Zfp791" "Rfesd" "Arhgef3" "Pcdh20"   
[345] "Tmcc1" "Zswim4" "Nipal2" "Tbx1"   
[349] "Rfwd2" "Hcfc2" "Nprl3" "Zfp316"   
[353] "Slfn2" "Usp43" "Bbs9" "Ppp1r13l"   
[357] "Ccdc134" "Pxmp4" "Man2b2" "Cfap70"   
[361] "Zkscan6" "Sgcd" "Zmat1" "Zc3h3"   
[365] "Lgals12" "Sh3rf3" "Etv5" "Oxnad1"   
[369] "Pgbd5" "Zfp277" "Ddx59" "Rassf6"   
[373] "Nol10" "Dusp7" "Pid1" "Nxpe4"   
[377] "Acin1" "Mmachc" "Ptprr" "Rimkla"   
[381] "Wdsub1" "P2ry2" "Fam221a" "Pipox"   
[385] "Sord" "Furin" "Cdkal1" "Etv6"   
[389] "Creb5" "Kirrel3" "Fstl3" "Nos3"   
[393] "Cdh4" "Ufd1" "Mocs1" "Rundc1"   
[397] "Col11a1" "Ndufaf4" "Il21r" "Tesk2"   
[401] "Pdyn" "Nt5dc1" "Gemin8" "Cd36"   
[405] "Gm16299" "Srpk3" "Gm11696" "Gldn"   
[409] "4933407L21Rik" "Spred3" "Pex2" "Trak1"   
[413] "Ifit1bl1" "Mms22l" "Clspn" "Fra10ac1"   
[417] "A130010J15Rik" "Rrs1" "Il17ra" "E2f4"   
[421] "Dnajb5" "Eif4b" "Cbx8" "Cyr61"   
[425] "Iqcc" "Kcna5" "Mus81" "Ptpn12"   
[429] "Plpbp" "Fam46a" "Znrf1" "1110032F04Rik"  
[433] "Ryk" "P4ha3" "Scube3" "Gtf2ird2"   
[437] "Klf11" "Kifc1" "Zdhhc15" "2310057M21Rik"  
[441] "Etfrf1" "Mpp6" "Cyb5d2" "Socs3"   
[445] "Heatr1" "Bend7" "D10Wsu102e" "Gm9776"   
[449] "Ccdc14" "Ttc38" "Gm14204" "Tbx3"   
[453] "Tpbgl" "Mis12" "Fxr1" "Ccno"   
[457] "BC048403" "Zfp810" "Fam196a" "Prkd2"   
[461] "Papd7" "Slc9a8" "Tpm1" "Filip1"   
[465] "Dnd1" "Rtn4rl2" "Top1" "Slc9a5"   
[469] "Trim26" "Rassf5" "Tex10" "Agk"   
[473] "Casp9" "Wnt9a" "Meltf" "Sh3yl1"   
[477] "Rerg" "Scn4b" "Ak4" "Srarp"   
[481] "Zbed4" "Rasip1" "Cacna2d2" "Evpl"   
[485] "Ccl28" "Kcne1l" "Adnp2" "Tmem241"   
[489] "Lfng" "Trpc7" "Max" "Zbtb25"   
[493] "Mxi1" "Fam241a" "Nfam1" "H1f0"   
[497] "Met" "Hpdl" "Inafm2" "Rwdd3"   
[501] "Qtrt1" "Klhdc1" "Mtcl1" "Prkch"   
[505] "Spred1" "1110008L16Rik" "Kcna4" "Dpf2"   
[509] "Ppp6r1" "1700003D09Rik" "N6amt1" "Snx2"   
[513] "Tob2" "Irs4" "Csmd2" "Dph6"   
[517] "Tyw3" "Sp2" "Gramd1b" "Ccdc181"   
[521] "Gfpt2" "Slc16a6" "Unc13c" "Ccl6"   
[525] "Ccnd2" "Prpf3" "Cmtm3" "Zhx2"   
[529] "Trim9" "Adprm" "Rflnb" "Aco1"   
[533] "BC030500" "Mfsd4a" "Zc3h8" "Sox18"   
[537] "Rps6ka5" "Slc9a9" "Sapcd2" "Atpaf1"   
[541] "Vgf" "Brca2" "Hic1" "Rft1"   
[545] "Slc35g2" "Cbwd1" "Lin9" "Hic2"   
[549] "Tmem267" "Bcl2" "Btg2" "Unc45a"   
[553] "Noxred1" "Dclk2" "Flcn" "E130317F20Rik"  
[557] "Moap1" "Mettl2" "Cdkn1a" "B4galt5"   
[561] "C530005A16Rik" "Rin2" "Kansl2" "Ap4b1"   
[565] "Cyb5rl" "1700019D03Rik" "Nmb" "Utp18"   
[569] "B3gntl1" "Gimap1" "Mamstr" "Cutc"   
[573] "LOC100861615" "Ephb3" "Zfp119a" "Fbxo16"   
[577] "Zscan18" "Ubc" "Nr6a1" "Araf"   
[581] "Mlip" "Arhgap31" "Cdc42ep3" "Smim3"   
[585] "Uspl1" "Sept9" "Zbtb16" "Ece1"   
[589] "Bcl6" "Zfp202" "Tspan14" "Bloc1s5"   
[593] "Arpin" "Wdr25" "Zbtb17" "Riox2"   
[597] "Lrrc8d" "9630001P10Rik" "Rabl3" "Spry4"   
[601] "Fndc8" "Cbx4" "Lhx2" "Zfp512"   
[605] "Npbwr1" "Pdia4" "Baiap3" "Nlrx1"   
[609] "Ttc26" "Uvssa" "Socs2"   
  
$duration8wks  
 [1] "Gpr88" "Ido1" "Adck2" "Fos"   
 [5] "Sh2d3c" "Ldlr" "Pde10a" "4930432K21Rik"  
 [9] "Smad3" "Cwc25" "Lrrc61" "Per2"   
 [13] "Junb" "Ier2" "Ofd1" "Rtn4ip1"   
 [17] "Ybey" "Ccdc112" "3632451O06Rik" "Rbm11"   
 [21] "Myadml2" "B3galt6" "Zfp459" "Fam177a"   
 [25] "Igf2bp2" "Osbpl3" "Sco1" "Slc37a1"   
 [29] "Lpcat2" "Gpatch11" "Bace2" "Fam111a"   
 [33] "Slc10a7" "S1pr2" "Ninl" "Mtrf1"   
 [37] "Trim45" "Lrrc49" "Csrnp2" "Pdzd9"   
 [41] "Diexf" "Glrx2" "Tipin" "Phf21b"   
 [45] "Arhgap39" "Insig1" "Arl5c" "Tefm"   
 [49] "Fbxl4" "Mex3d" "A930017K11Rik" "Plekha3"   
 [53] "Plppr1" "Phf13" "2810029C07Rik" "Htr6"   
 [57] "Ciart" "Frmd3" "Parn" "Anapc4"   
 [61] "Ddx27" "Pms1" "Dtwd2" "Prr5"   
 [65] "Chaf1a" "Slc35f4" "Plin4" "Pik3c3"   
 [69] "Rin3" "Gm5083" "Zfp366" "Rars2"   
 [73] "Ifit1" "Tcerg1l" "E130102H24Rik" "Zfp930"   
 [77] "Fzd1" "Ppm1d" "Gtf2h1" "Pou6f2"   
 [81] "Epm2a" "A430108G06Rik" "Ttc34" "Ccdc183"   
 [85] "Usb1" "Dusp1" "Mak" "Samd3"   
 [89] "Tigd2" "Rnd3" "Cd276" "Trim62"   
 [93] "Prdm4" "Omg" "Ift52" "Tiparp"   
 [97] "Rbm43" "Mta2" "Dhx32" "Rac2"   
[101] "4930513N10Rik" "Gpt" "Sgms1" "Plekho2"   
[105] "Cog6" "Pum3" "Morc2b" "Dars"   
[109] "Scpep1" "Dnttip2" "Prss57" "Nrap"   
[113] "Nub1" "Anxa3" "Gart" "Sdr42e1"   
[117] "Wrb" "Gm14440" "Traf3" "Nr4a1"   
[121] "Gjb6" "Sdhaf3" "Tmcc1" "Nipal2"   
[125] "Egr2" "Pak7" "Bbs9" "Pxmp4"   
[129] "Zkscan6" "Exosc3" "Gspt2" "Oxnad1"   
[133] "Zfp277" "Actr8" "Arhgap27" "Serpine1"   
[137] "Smu1" "St7l" "P2ry2" "Golt1b"   
[141] "Cdkal1" "Adcyap1" "Il21r" "Pdyn"   
[145] "Eif4e2" "6430550D23Rik" "Spred3" "Mybpc1"   
[149] "Ifit1bl1" "Fra10ac1" "Iqgap2" "Eif4b"   
[153] "Pdhx" "Cbx8" "Kcna5" "Znrf1"   
[157] "1110032F04Rik" "Hspa9" "Rabgef1" "Ggnbp2"   
[161] "Zdhhc15" "Hvcn1" "Cyb5d2" "Ddx10"   
[165] "Tbx3" "Rbm4" "Notch4" "Arntl"   
[169] "Tpm1" "Xlr3a" "Trit1" "Agk"   
[173] "Rad17" "Casp9" "Wnt9a" "Sh3yl1"   
[177] "Hnrnpa1" "Zbed4" "Rasip1" "Adnp2"   
[181] "Gm12522" "Lfng" "Loxl3" "Ell"   
[185] "Nfam1" "H1f0" "Met" "Klhdc1"   
[189] "Eed" "Zfp874b" "Snx2" "Capn11"   
[193] "Ccdc181" "Ccl6" "Nt5c1a" "Cmtm3"   
[197] "BC030500" "Zc3h8" "Sox18" "Sapcd2"   
[201] "Atpaf1" "Slc35g2" "Wdr92" "Cbwd1"   
[205] "Hic2" "Noxred1" "Moap1" "Fmo2"   
[209] "Mettl2" "Atg14" "C530005A16Rik" "Ap4b1"   
[213] "Vps37b" "Ing2" "Sema5b" "Cutc"   
[217] "LOC100861615" "Zscan18" "Ubc" "Nop14"   
[221] "Araf" "Mlip" "Aspa" "Zfp7"   
[225] "Bloc1s5" "Cluap1" "Arpin" "Riox2"   
[229] "Rabl3" "Lhx2" "Chsy1" "Zfp512"

# To classify genes in more detail, we collapsed the binarized matrix gene by gene and grouped genes from the unique collapsed patterns.  
head(stat\_binary)

X.Intercept.\_pval duration4weeks\_pval duration6weeks\_pval  
1 FALSE FALSE TRUE  
2 TRUE FALSE TRUE  
3 TRUE TRUE TRUE  
4 FALSE TRUE TRUE  
5 FALSE FALSE TRUE  
6 FALSE FALSE TRUE  
 duration8weeks\_pval  
1 FALSE  
2 FALSE  
3 TRUE  
4 FALSE  
5 FALSE  
6 FALSE

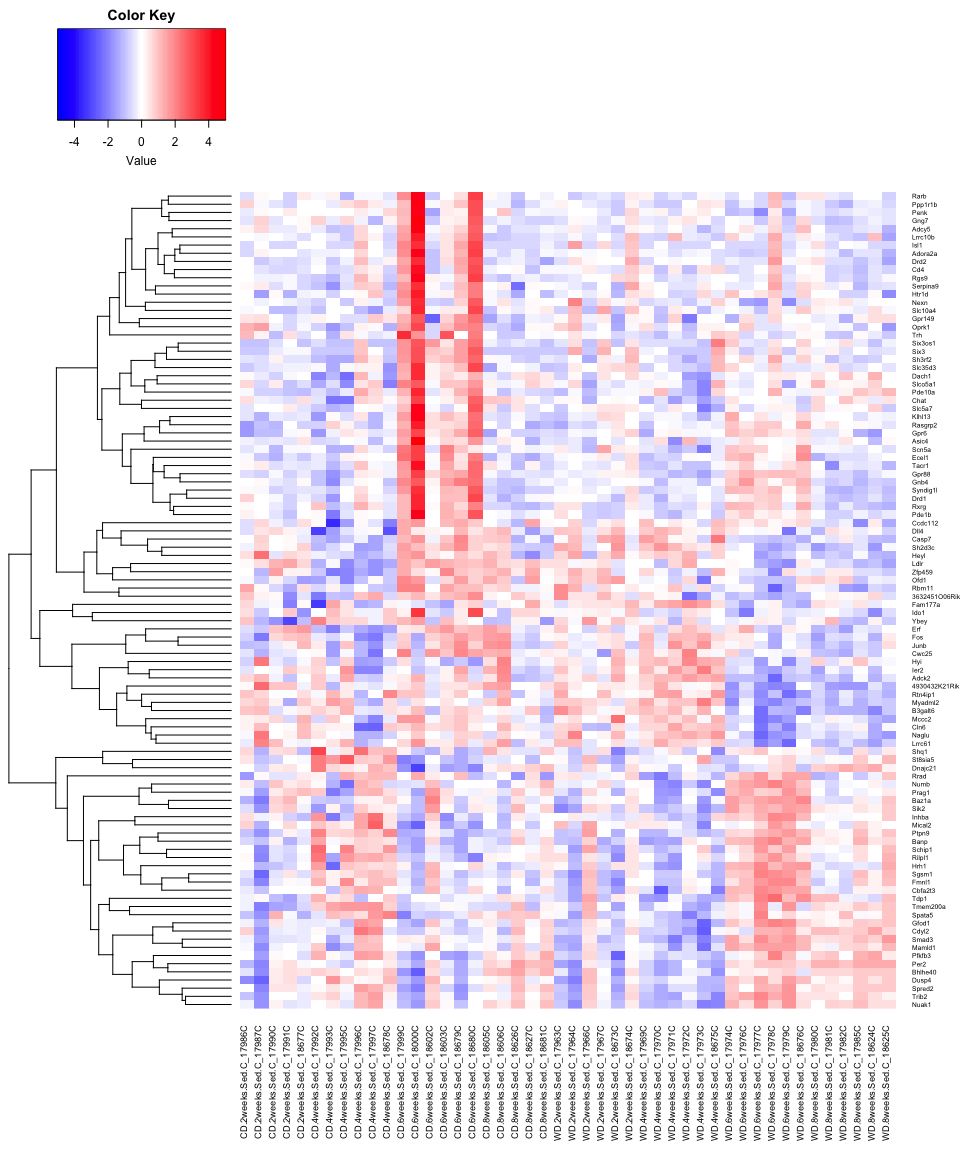
colnames(stat\_binary) <- c("Intercept", "duration4wks", "duration6wks", "duration8wks")  
#stat\_binary <- stat\_binary %>% add\_column(Symbol=stat\_table\_all$Symbol, .before=1)  
stat\_binary <- stat\_binary %>% mutate(pattern= paste(Intercept,duration4wks,duration6wks, duration8wks ,sep = "\_"))  
pattern\_TF <- table(stat\_binary$pattern) # %>% as.data.frame()  
stat\_binary.gather <- stat\_binary %>% gather(key=coeff\_group, value=value, -pattern)  
stat\_binary.gather$coeff\_group <- factor(stat\_binary.gather$coeff\_group,   
 levels = c("Intercept","duration4wks", "duration6wks", "duration8wks"),  
 ordered = TRUE)  
  
ggplot(stat\_binary.gather, aes(x = coeff\_group, y = pattern, fill = value)) + geom\_tile(colour = "white") +  
 theme\_bw() + xlab("") + ylab("") + coord\_flip() +  
 # scale\_x\_discrete(labels = c("Intercept", "duration4wks", "duration6wks", "duration8wks")) +  
 scale\_y\_discrete(labels = pattern\_TF) +  
 scale\_fill\_manual(values = c("grey80", "firebrick1"))



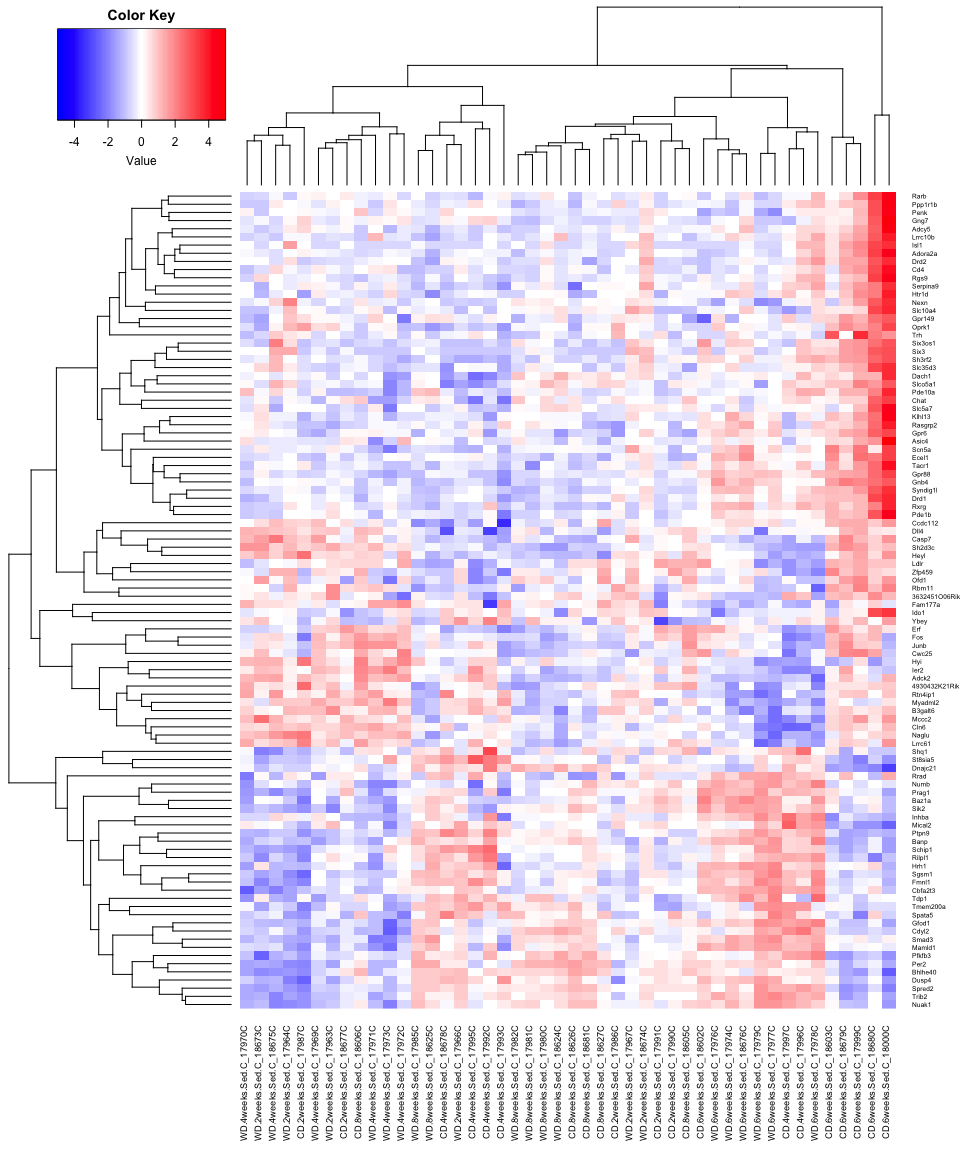
## Hierarchical Clustering of all the samples based on Top 100 genes from GLM

1. heatmap with ordering the sample by their treatment
2. heatmap without ordering the sample (column)

# heatmap   
logCPM <- log.cpm  
rownames(logCPM) <- y$genes$Symbol  
colnames(logCPM) <- data.design$sample\_short # get it into the y project  
logCPM<- logCPM[stat\_table\_all$Symbol[1:100],]  
 # get rid of the running group  
logCPM<- t(scale(t(logCPM))) %>% as.data.frame() %>% as.matrix()  
logCPM.order <- logCPM[, colnames(logCPM) %>% order()]  
  
library(gplots)  
#heatmap, columns ordered by their group/treatment  
col.pan <- colorpanel(100, "blue", "white", "red")  
heatmap.2(logCPM.order, col=col.pan, Rowv=TRUE, Colv= FALSE, scale="none",   
 trace="none", dendrogram="row", cexRow=0.5, cexCol=0.7, density.info="none",  
 margin=c(9,4), lhei=c(2,10), lwid=c(2,6))



#heatmap, columns ordered by the clustering  
col.pan <- colorpanel(100, "blue", "white", "red")  
heatmap.2(logCPM.order, col=col.pan, Rowv=TRUE, scale="none",   
 trace="none", dendrogram="both", cexRow=0.5, cexCol=0.7, density.info="none",  
 margin=c(9,4), lhei=c(2,10), lwid=c(2,6))



## GO and Kegg pathway analysis

## ---- GO and Kegg pathway analysis  
# Xulong's function  
source("~/Documents/code/01\_function/Kegg\_function.R")  
gk.diet <- myGK(stat\_table\_all$Symbol)  
head(gk.diet$BP[, c("Term", "Pvalue")])

Term Pvalue  
1 cellular macromolecule metabolic process 3.52e-23  
2 cellular metabolic process 2.62e-22  
3 metabolic process 4.53e-22  
4 nitrogen compound metabolic process 3.96e-21  
5 organic substance metabolic process 9.71e-20  
6 primary metabolic process 1.04e-19

head(gk.diet$MF[, c("Term", "Pvalue")])

Term Pvalue  
1 binding 7.46e-20  
2 heterocyclic compound binding 2.63e-15  
3 organic cyclic compound binding 9.34e-15  
4 ion binding 3.10e-13  
5 cation binding 2.55e-11  
6 nucleic acid binding 7.99e-11

head(gk.diet$CC[, c("Term", "Pvalue")])

Term Pvalue  
1 cell part 5.98e-29  
2 cell 7.11e-29  
3 intracellular 9.85e-29  
4 intracellular part 5.85e-28  
5 intracellular organelle 6.70e-23  
6 organelle 1.23e-22

gk.diet$KEGG[, c("Term", "Pvalue")]

Term Pvalue  
1 Colorectal cancer 0.000256  
2 Pathways in cancer 0.000269  
3 Pancreatic cancer 0.001943  
4 MAPK signaling pathway 0.002663  
5 Non-small cell lung cancer 0.005815  
6 Renal cell carcinoma 0.007715  
7 Glycosaminoglycan degradation 0.009215  
8 RNA degradation 0.009849

# output the analysis result  
#file\_name<- list("./analysis/Cortex\_BP.txt", "./analysis/Cortex\_MF.txt", "./analysis/Cortext\_CC.txt", "./analysis/Cortext\_KEGG.txt")  
#walk2(gk.diet, file\_name, write.table, sep="\t", row=F, quote=F)  
#write.table(stat\_table\_all, "./analysis/diet\_gene.txt", sep="\t", row=F, quote=F)

sessionInfo()

R version 3.4.3 (2017-11-30)  
Platform: x86\_64-apple-darwin15.6.0 (64-bit)  
Running under: OS X El Capitan 10.11.6  
  
Matrix products: default  
BLAS: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRblas.0.dylib  
LAPACK: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRlapack.dylib  
  
locale:  
[1] en\_US.UTF-8/en\_US.UTF-8/en\_US.UTF-8/C/en\_US.UTF-8/en\_US.UTF-8  
  
attached base packages:  
 [1] grid parallel stats4 stats graphics grDevices utils   
 [8] datasets methods base   
  
other attached packages:  
 [1] GO.db\_3.5.0 pathview\_1.18.2 org.Hs.eg.db\_3.5.0   
 [4] org.Mm.eg.db\_3.5.0 KEGG.db\_3.2.3 GOstats\_2.44.0   
 [7] graph\_1.56.0 Category\_2.44.0 Matrix\_1.2-12   
[10] AnnotationDbi\_1.40.0 VennDiagram\_1.6.19 futile.logger\_1.4.3   
[13] gplots\_3.0.1 corrplot\_0.84 IRanges\_2.12.0   
[16] S4Vectors\_0.16.0 Biobase\_2.38.0 BiocGenerics\_0.24.0   
[19] edgeR\_3.20.9 limma\_3.34.9 bindrcpp\_0.2   
[22] forcats\_0.3.0 stringr\_1.3.0 dplyr\_0.7.4   
[25] purrr\_0.2.4 readr\_1.1.1 tidyr\_0.8.0   
[28] tibble\_1.4.2 ggplot2\_2.2.1 tidyverse\_1.2.1   
[31] knitr\_1.20 BiocStyle\_2.6.1   
  
loaded via a namespace (and not attached):  
 [1] nlme\_3.1-131 bitops\_1.0-6 lubridate\_1.7.3   
 [4] bit64\_0.9-7 httr\_1.3.1 rprojroot\_1.3-2   
 [7] Rgraphviz\_2.22.0 tools\_3.4.3 backports\_1.1.2   
[10] utf8\_1.1.3 R6\_2.2.2 KernSmooth\_2.23-15   
[13] DBI\_0.8 lazyeval\_0.2.1 colorspace\_1.3-2   
[16] tidyselect\_0.2.4 mnormt\_1.5-5 bit\_1.1-12   
[19] compiler\_3.4.3 cli\_1.0.0 rvest\_0.3.2   
[22] xml2\_1.2.0 labeling\_0.3 KEGGgraph\_1.38.1   
[25] caTools\_1.17.1 scales\_0.5.0 psych\_1.7.8   
[28] genefilter\_1.60.0 RBGL\_1.54.0 digest\_0.6.15   
[31] foreign\_0.8-69 rmarkdown\_1.9 XVector\_0.18.0   
[34] AnnotationForge\_1.20.0 pkgconfig\_2.0.1 htmltools\_0.3.6   
[37] rlang\_0.2.0 readxl\_1.0.0 rstudioapi\_0.7   
[40] RSQLite\_2.0 bindr\_0.1.1 jsonlite\_1.5   
[43] gtools\_3.5.0 RCurl\_1.95-4.10 magrittr\_1.5   
[46] Rcpp\_0.12.16 munsell\_0.4.3 stringi\_1.1.7   
[49] yaml\_2.1.18 zlibbioc\_1.24.0 plyr\_1.8.4   
[52] blob\_1.1.0 gdata\_2.18.0 crayon\_1.3.4   
[55] lattice\_0.20-35 Biostrings\_2.46.0 haven\_1.1.1   
[58] splines\_3.4.3 annotate\_1.56.1 KEGGREST\_1.18.1   
[61] hms\_0.4.2 locfit\_1.5-9.1 pillar\_1.2.1   
[64] reshape2\_1.4.3 codetools\_0.2-15 futile.options\_1.0.0   
[67] XML\_3.98-1.10 glue\_1.2.0 evaluate\_0.10.1   
[70] lambda.r\_1.2 modelr\_0.1.1 png\_0.1-7   
[73] cellranger\_1.1.0 gtable\_0.2.0 assertthat\_0.2.0   
[76] xtable\_1.8-2 broom\_0.4.3 survival\_2.41-3   
[79] memoise\_1.1.0 statmod\_1.4.30 GSEABase\_1.40.1