

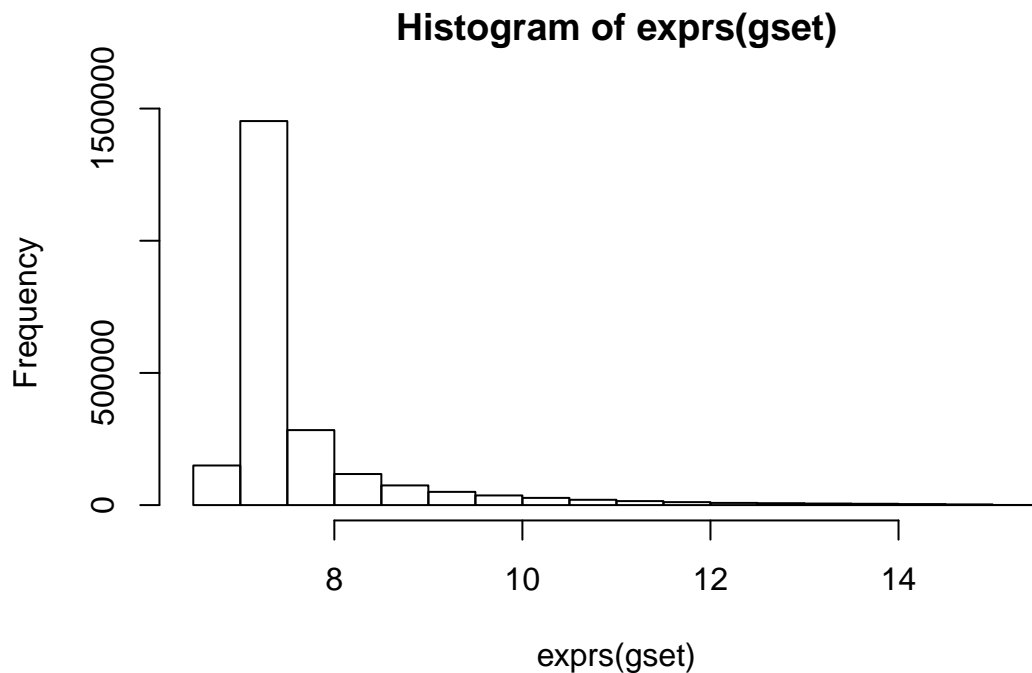
Project

Written By Chee Kit Tang (ct2819)

```
#####Group samples data such that G0-baseline,G1-4h,G2-10h,G3-24h#####
# make proper column names to match toptable
fvarLabels(gset) <- make.names(fvarLabels(gset))
# group names for all samples
gsms <- "012301230123012301230123012301230123012301230123"
sml <- c()
for (i in 1:nchar(gsms)) { sml[i] <- substr(gsms,i,i) }

#####Set up the data and proceed with analysis#####
sml <- paste("G", sml, sep="") # set group names
fl <- as.factor(sml)
gset$description <- fl
design <- model.matrix(~ description + 0, gset)
colnames(design) <- levels(fl)
fit <- lmFit(gset, design)
cont.matrix <- makeContrasts(G2-G0,G3-G0, levels=design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2)

#####Show distribution and normalization#####
par(mfrow=c(1,1))
hist(exprs(gset))
```

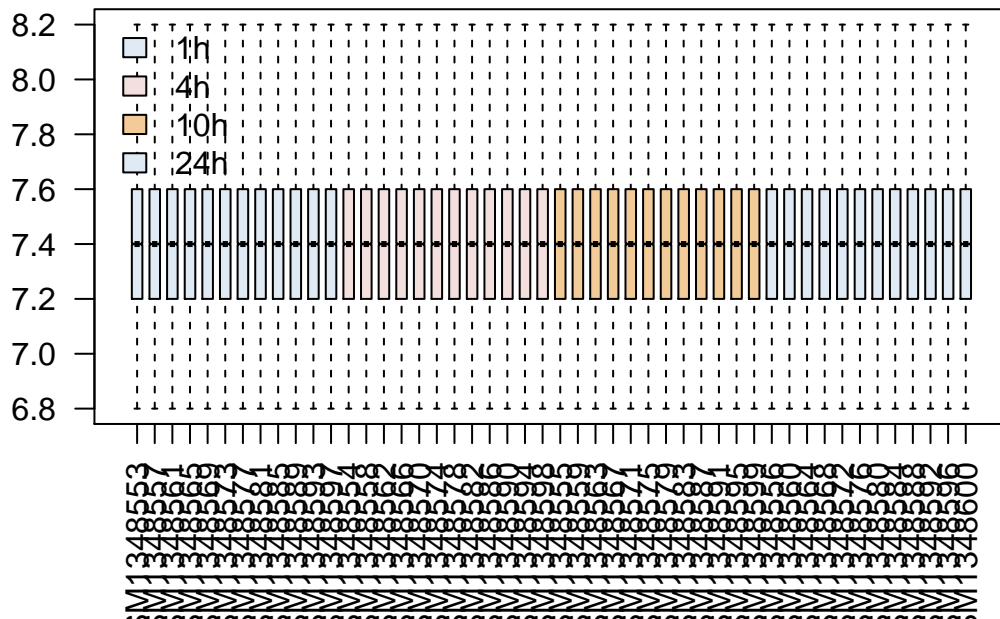


```

#Order samples by group
ex <- exprs(gset)[ , order(sml)]
sml <- sml[order(sml)]
fl <- as.factor(sml)
labels <- c("1h","4h","10h","24h")
# set parameters and draw the plot
palette(c("#dfeaf4", "#f4dfdf", "#f2cb98", "#dfeaf4", "#AABBCC"))
par(mfrow=c(1,1))
title <- paste ("GSE55924", '/', annotation(gset), "samples", sep='')
boxplot(ex, boxwex=0.6, notch=T, main=title, outline=FALSE, las=2, col=fl)
legend("topleft", labels, fill=palette(), bty="n")

```

GSE55924/GPL10558samples



```

#####Find Top 10 up and downregulated genes during 10h and 24h #####
tT <- topTable(fit2,coef=1, adjust="none", sort.by="logFC",resort.by="logFC", number=Inf,p.value=0.05,l
tT <- subset(tT, select=c("ID","P.Value","t","B","logFC","Gene.symbol","Gene.title"))
tT2 <- topTable(fit2,coef=2, adjust="none", sort.by="logFC",resort.by="logFC", number=Inf,p.value=0.05,l
tT2 <- subset(tT2, select=c("ID","P.Value","t","B","logFC","Gene.symbol","Gene.title"))
len_tT=dim(tT)[1]
tT[1:10,]

```

##	ID	P.Value	t	B	logFC
##	ILMN_1684982	ILMN_1684982	3.975001e-07	5.854644	4.987895 1.900000
##	ILMN_2186061	ILMN_2186061	5.020987e-05	4.447661	1.348401 1.575000
##	ILMN_2340259	ILMN_2340259	5.073660e-03	2.935040	-2.166336 1.466667
##	ILMN_1707727	ILMN_1707727	6.756506e-06	5.042258	2.869256 1.383333
##	ILMN_2052208	ILMN_2052208	3.569539e-03	3.062259	-1.900907 1.341667
##	ILMN_1697448	ILMN_1697448	6.144829e-11	8.333691	11.221868 1.333333
##	ILMN_1788874	ILMN_1788874	3.533399e-02	2.164699	-3.602817 1.241667
##	ILMN_1694075	ILMN_1694075	6.306159e-03	2.854837	-2.329953 1.158333

```
## ILMN_1707312 ILMN_1707312 7.365838e-05 4.331213 1.056457 1.125000
## ILMN_1663092 ILMN_1663092 5.208012e-06 5.117968 3.065537 1.100000
## Gene.symbol
## ILMN_1684982 PDK4
## ILMN_2186061 PFKFB3
## ILMN_2340259 PDE4B
## ILMN_1707727 ANGPTL4
## ILMN_2052208 GADD45A
## ILMN_1697448 TXNIP
## ILMN_1788874 SERPINA3
## ILMN_1694075 GADD45A
## ILMN_1707312 NFIL3
## ILMN_1663092 CITED2
## Gene.title
## ILMN_1684982 pyruvate dehydrogenase kinase 4
## ILMN_2186061 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3
## ILMN_2340259 phosphodiesterase 4B
## ILMN_1707727 angiotensin II type 1 receptor
## ILMN_2052208 growth arrest and DNA damage inducible alpha
## ILMN_1697448 thioredoxin interacting protein
## ILMN_1788874 serpin family A member 3
## ILMN_1694075 growth arrest and DNA damage inducible alpha
## ILMN_1707312 nuclear factor, interleukin 3 regulated
## ILMN_1663092 Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2
tT[(len_tT-9):len_tT,]
```

```
## ID P.Value t B logFC
## ILMN_1800626 ILMN_1800626 4.914741e-04 -3.735230 -0.3922977 -0.6000000
## ILMN_1708041 ILMN_1708041 6.072029e-05 -4.390043 1.2036485 -0.6750000
## ILMN_1683133 ILMN_1683133 2.558834e-09 -7.274689 8.6415941 -0.6750000
## ILMN_1672004 ILMN_1672004 4.461120e-04 -3.766571 -0.3183589 -0.7416667
## ILMN_1715555 ILMN_1715555 1.990130e-07 -6.050261 5.4986659 -0.7666667
## ILMN_1738095 ILMN_1738095 2.088167e-05 -4.710502 2.0153192 -0.7833333
## ILMN_1671703 ILMN_1671703 1.150093e-02 -2.626303 -2.7793438 -0.8000000
## ILMN_1781285 ILMN_1781285 2.195766e-03 -3.233573 -1.5326945 -0.8083333
## ILMN_1793543 ILMN_1793543 3.384129e-11 -8.505260 11.6241063 -1.2833333
## ILMN_1791728 ILMN_1791728 3.458452e-13 -9.852886 14.6111091 -1.5916667
## Gene.symbol Gene.title
## ILMN_1800626 SESN1 sestrin 1
## ILMN_1708041 PLEKHF1 pleckstrin homology and FYVE domain containing 1
## ILMN_1683133 KLF15 Kruppel like factor 15
## ILMN_1672004 TOB1 transducer of ERBB2, 1
## ILMN_1715555 DBP D-box binding PAR bZIP transcription factor
## ILMN_1738095 PER2 period circadian clock 2
## ILMN_1671703 ACTA2 actin, alpha 2, smooth muscle, aorta
## ILMN_1781285 DUSP1 dual specificity phosphatase 1
## ILMN_1793543 CIART circadian associated repressor of transcription
## ILMN_1791728 SLC25A25 solute carrier family 25 member 25
```

```
len_tT2=dim(tT2)[1]
tT2[1:10,]
```

```
## ID P.Value t B logFC
## ILMN_2186061 ILMN_2186061 9.868659e-09 6.895051 9.6636768 2.441667
```

ILMN_ID	ILMN_ID	1.219157e-05	4.869498	3.1702369	2.433333
ILMN_1788874	ILMN_1788874	2.387851e-04	3.966193	0.4644561	2.275000
ILMN_1782050	ILMN_1782050	3.558649e-07	5.885979	6.3984000	2.150000
ILMN_2052208	ILMN_2052208	1.796797e-05	4.755060	2.8163322	2.083333
ILMN_1800512	ILMN_1800512	2.417822e-09	7.290657	10.9383317	2.050000
ILMN_1686664	ILMN_1686664	1.333135e-05	4.843196	3.0886668	1.925000
ILMN_1694075	ILMN_1694075	3.019519e-05	4.600600	2.3430329	1.866667
ILMN_2186137	ILMN_2186137	2.815915e-04	3.913952	0.3153459	1.816667
ILMN_1684982	ILMN_1684982	1.174762e-06	5.546505	5.3075336	1.800000
ILMN_2186061	Gene.symbol				
ILMN_2340259	PFKFB3				
ILMN_1788874	PDE4B				
ILMN_1782050	SERPINA3				
ILMN_2052208	CEBPD				
ILMN_1800512	GADD45A				
ILMN_1686664	HMOX1				
ILMN_1694075	MT2A				
ILMN_2186137	GADD45A				
ILMN_1684982	RRAD				
	PDK4				
	Gene.title				
ILMN_2186061	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3				
ILMN_2340259	phosphodiesterase 4B				
ILMN_1788874	serpin family A member 3				
ILMN_1782050	CCAAT/enhancer binding protein delta				
ILMN_2052208	growth arrest and DNA damage inducible alpha				
ILMN_1800512	heme oxygenase 1				
ILMN_1686664	metallothionein 2A				
ILMN_1694075	growth arrest and DNA damage inducible alpha				
ILMN_2186137	RRAD, Ras related glycolysis inhibitor and calcium channel regulator				
ILMN_1684982	pyruvate dehydrogenase kinase 4				

tT2[(len_tT2-9):len_tT2,]

ID	P.Value	t	B	logFC	
ILMN_1736670	ILMN_1736670	1.718949e-05	-4.768169	2.8567380	-1.016667
ILMN_1751084	ILMN_1751084	1.748401e-06	-5.432743	4.9442006	-1.058333
ILMN_1750062	ILMN_1750062	2.853939e-07	-5.948416	6.5998684	-1.091667
ILMN_1778668	ILMN_1778668	1.354042e-04	-4.143820	0.9783827	-1.116667
ILMN_1691846	ILMN_1691846	3.236746e-04	-3.869596	0.1895018	-1.125000
ILMN_1728298	ILMN_1728298	8.230547e-06	-4.984671	3.5289294	-1.183333
ILMN_2364022	ILMN_2364022	8.457978e-06	-4.976703	3.5040394	-1.316667
ILMN_1671703	ILMN_1671703	1.011524e-05	-4.924318	3.3406663	-1.500000
ILMN_1674243	ILMN_1674243	3.302111e-05	-4.573816	2.2615112	-1.525000
ILMN_1791728	ILMN_1791728	1.048543e-13	-10.213987	19.8668739	-1.650000
ILMN_1736670	Gene.symbol				Gene.title
ILMN_1751084	PPP1R3C				protein phosphatase 1 regulatory subunit 3C
ILMN_1750062	FSD2				fibronectin type III and SPRY domain containing 2
ILMN_1778668	PPARGC1A				PPARG coactivator 1 alpha
ILMN_1691846	TAGLN				transgelin
ILMN_1728298	GOS2				G0/G1 switch 2
ILMN_2364022	SBK1				SH3 domain binding kinase 1
ILMN_1671703	SLC16A3				solute carrier family 16 member 3
ILMN_1674243	ACTA2				actin, alpha 2, smooth muscle, aorta
	TFRC				transferrin receptor

```
## ILMN_1791728      SLC25A25      solute carrier family 25 member 25
```

```
#####Comparing validation of microarray result#####
```

```
tTFull <- topTable(fit2,coef=1, adjust="none", sort.by="logFC",resort.by="logFC", number=Inf)
tTFull <- subset(tTFull,select=c("ID","P.Value","t","B","logFC","Gene.symbol","Gene.title"))
tT2Full <- topTable(fit2,coef=2, adjust="none", sort.by="logFC",resort.by="logFC", number=Inf)
tT2Full <- subset(tT2Full,select=c("ID","P.Value","t","B","logFC","Gene.symbol","Gene.title"))
```

```
genes_ID<-c('ILMN_1707727','ILMN_1663092','ILMN_2052208','ILMN_1684982','ILMN_1660847','ILMN_1697448',
```

```
tTFull[genes_ID,]
```

##	ID	P.Value	t	B	logFC
##	ILMN_1707727	ILMN_1707727	6.756506e-06	5.042258	2.8692563
##	ILMN_1663092	ILMN_1663092	5.208012e-06	5.117968	3.0655369
##	ILMN_2052208	ILMN_2052208	3.569539e-03	3.062259	-1.9009072
##	ILMN_1684982	ILMN_1684982	3.975001e-07	5.854644	4.9878951
##	ILMN_1660847	ILMN_1660847	7.520336e-04	3.596117	-0.7169428
##	ILMN_1697448	ILMN_1697448	6.144829e-11	8.333691	11.2218683
##	ILMN_1750521	ILMN_1750521	1.184073e-03	3.444822	-1.0629943
##	ILMN_1724162	ILMN_1724162	1.800751e-05	-4.754409	2.1276521
##	ILMN_1704629	ILMN_1704629	3.078196e-07	-5.927023	5.1769925

```
## Gene.symbol
```

```
## ILMN_1707727 ANGPTL4
## ILMN_1663092 CITED2
## ILMN_2052208 GADD45A
## ILMN_1684982 PDK4
## ILMN_1660847 PFKFB3
## ILMN_1697448 TXNIP
## ILMN_1750521 UCP3
## ILMN_1724162 ARX
## ILMN_1704629 SLC1A7
```

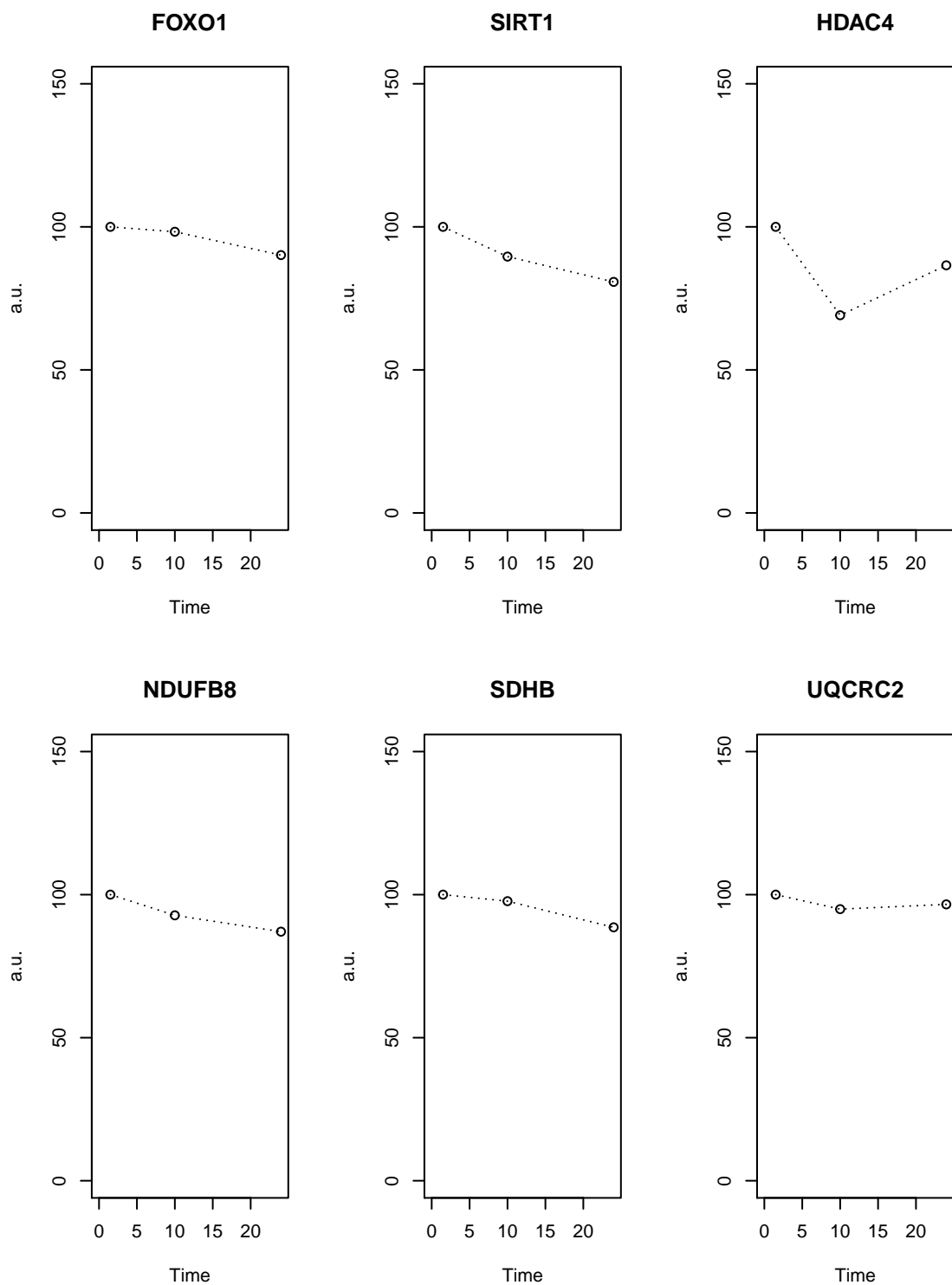
```
## Gene.title
## ILMN_1707727 angiopoietin like 4
## ILMN_1663092 Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2
## ILMN_2052208 growth arrest and DNA damage inducible alpha
## ILMN_1684982 pyruvate dehydrogenase kinase 4
## ILMN_1660847 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3
## ILMN_1697448 thioredoxin interacting protein
## ILMN_1750521 uncoupling protein 3
## ILMN_1724162 aristaless related homeobox
## ILMN_1704629 solute carrier family 1 member 7
```

```
tT2Full[genes_ID,]
```

##	ID	P.Value	t	B	logFC
##	ILMN_1707727	ILMN_1707727	1.917527e-03	3.280505	-1.4068261
##	ILMN_1663092	ILMN_1663092	1.176240e-04	4.187428	1.1061093
##	ILMN_2052208	ILMN_2052208	1.796797e-05	4.755060	2.8163322
##	ILMN_1684982	ILMN_1684982	1.174762e-06	5.546505	5.3075336
##	ILMN_1660847	ILMN_1660847	1.306502e-07	6.168948	7.3128070
##	ILMN_1697448	ILMN_1697448	7.953463e-10	7.604493	11.9426226
##	ILMN_1750521	ILMN_1750521	3.088146e-04	3.884587	0.2319541
##	ILMN_1724162	ILMN_1724162	7.383249e-07	-5.678877	5.7318487
##	ILMN_1704629	ILMN_1704629	2.852002e-09	-7.244139	10.7888917

```
##          Gene.symbol
## ILMN_1707727      ANGPTL4
## ILMN_1663092      CITED2
## ILMN_2052208      GADD45A
## ILMN_1684982      PDK4
## ILMN_1660847      PFKFB3
## ILMN_1697448      TXNIP
## ILMN_1750521      UCP3
## ILMN_1724162      ARX
## ILMN_1704629      SLC1A7
##
##                                     Gene.title
## ILMN_1707727                      angiopoietin like 4
## ILMN_1663092 Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2
## ILMN_2052208                      growth arrest and DNA damage inducible alpha
## ILMN_1684982                      pyruvate dehydrogenase kinase 4
## ILMN_1660847                      6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3
## ILMN_1697448                      thioredoxin interacting protein
## ILMN_1750521                      uncoupling protein 3
## ILMN_1724162                      aristaless related homeobox
## ILMN_1704629                      solute carrier family 1 member 7

#####Comparing some of the genes result#####
genes_symbol<-c('FOXO1','SIRT1','HDAC4','NDUFB8','SDHB','UQCRC2')
genes_ID<-c('ILMN_1738816','ILMN_1739083','ILMN_1764396','ILMN_1661170','ILMN_1667257','ILMN_1718853')
x<-c(1.5,10,24)
par(mfrow=c(1,3))
for(i in 1:6){
  if(i==4){
    par(mfrow=c(1,3))
  }
  y<-c(100,100*2^-abs(tTFull[genes_ID[i], 'logFC']),100*2^-abs(tT2Full[genes_ID[i], 'logFC']))
  plot(x,y, xlim=c(0,24),ylim=c(0,150),main=genes_symbol[i],xlab='Time',ylab='a.u.')
  lines(x,y,lty='dotted')
}
```



```
# calculate a distance matrix between each sample (each array)
dst <- dist(t(exprs(gset)))
# Hierarchical cluster analysis on above distance matrix
titleList<-c('Hierarchical Clustering-Average','Hierarchical Clustering-Maximum','Hierarchical Clustering-Single')
methodList<-c("average","complete","single")
```

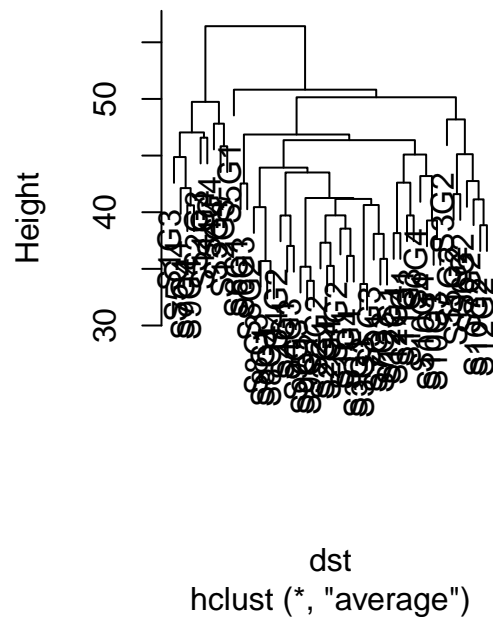
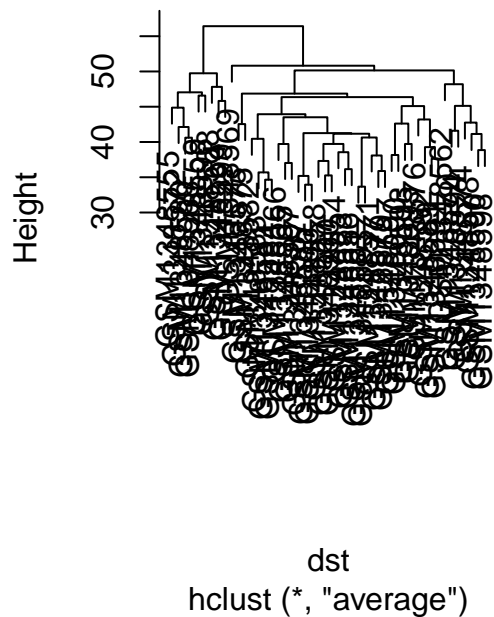
```

hhLabel<-NULL
for(i in 1:12){
  for(j in 1:4){
    hhLabel<-c(hhLabel,paste(c('S',i,'G',j), collapse = ""))
  }
}
for(i in 1:3){
  hh <- hclust(dst, method=methodList[i])
  # We will plot both of them on the same plot
  par(mfrow=c(1,2))
  par(mfrow=c(1,2))
  # plot default is by sample name
  plot(hh,main=titleList[i])

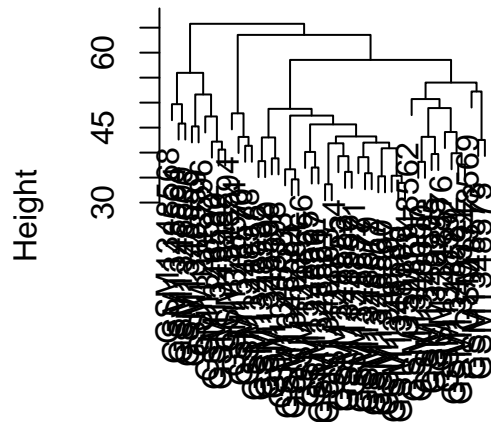
  # label sample by group
  plot(hh, label=hhLabel,main=titleList[i])
}

```

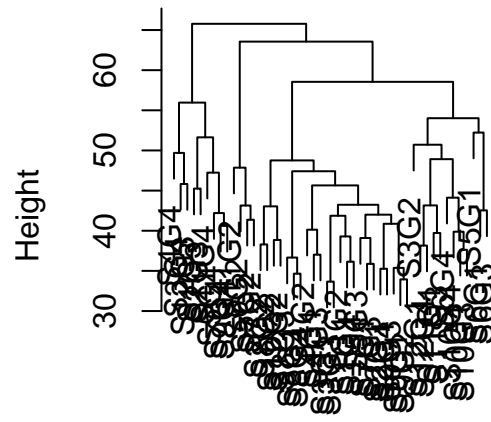
Hierarchical Clustering–Average Hierarchical Clustering–Average



Hierarchical Clustering–Maximum

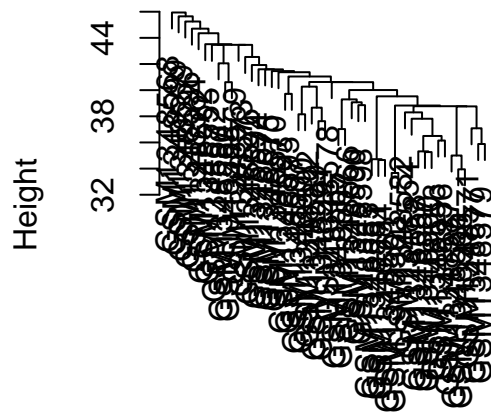


dst
hclust (*, "complete")

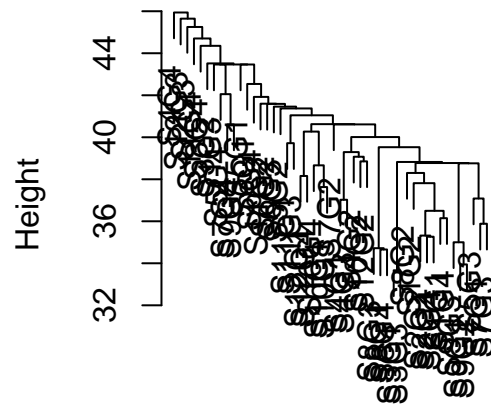


dst
hclust (*, "complete")

Hierarchical Clustering–Minimum

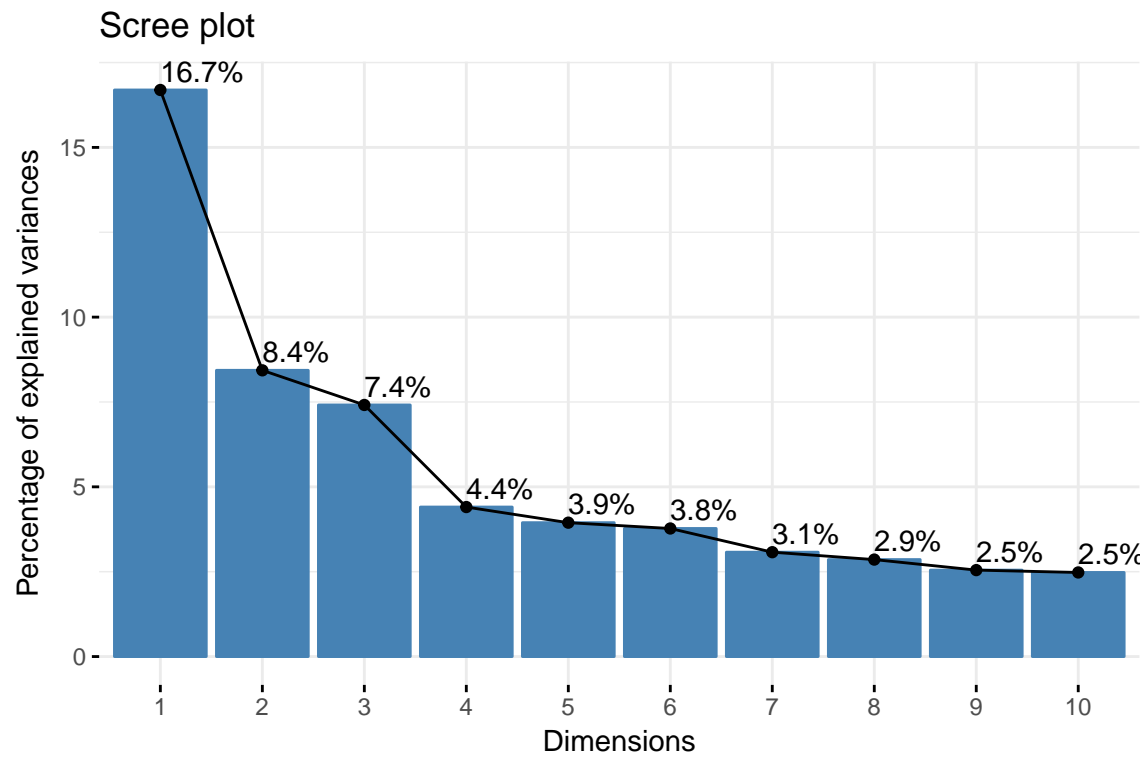


dst
hclust (*, "single")

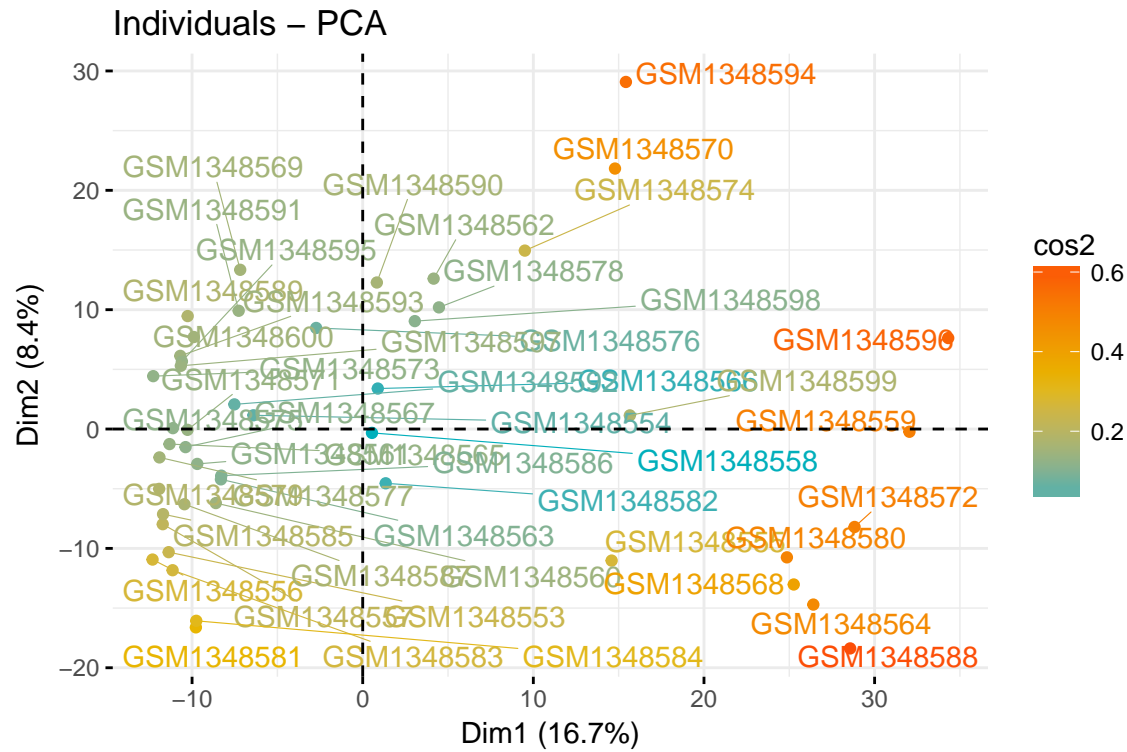


dst
hclust (*, "single")

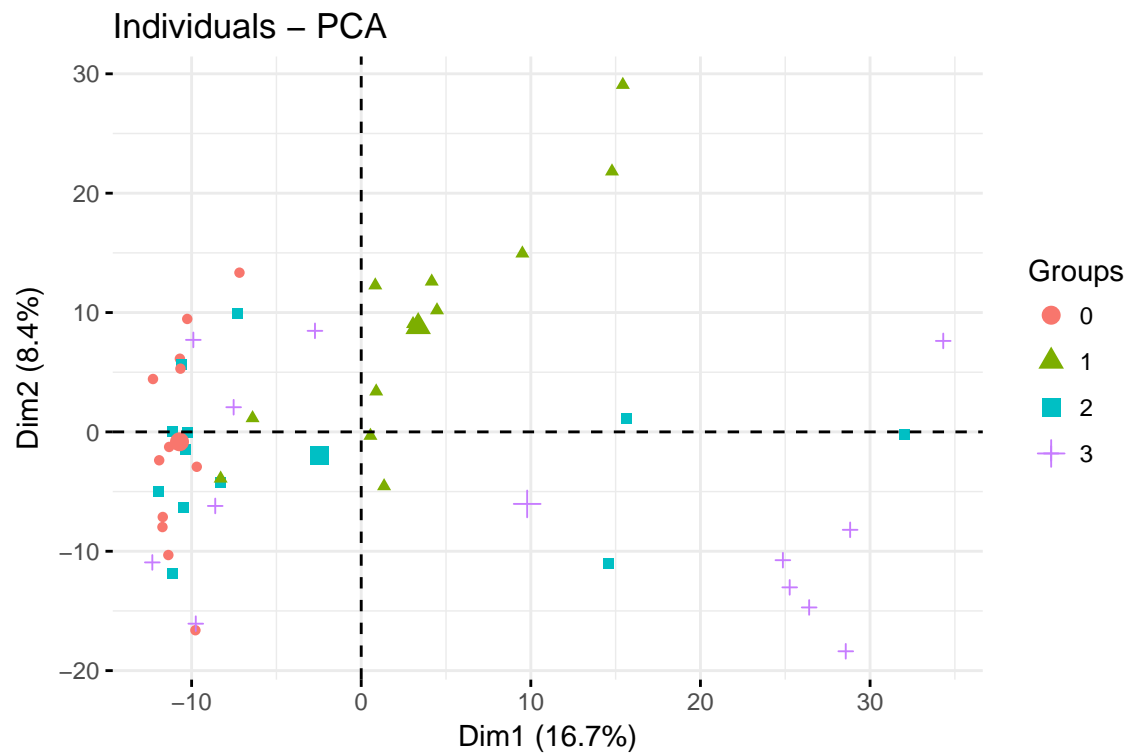
```
for (i in 1:nchar(gsms)) { sml[i] <- substr(gsms,i,i) }
PC=prcomp(t(exprs(gset)))
fviz_eig(PC, addlabels=TRUE)
```



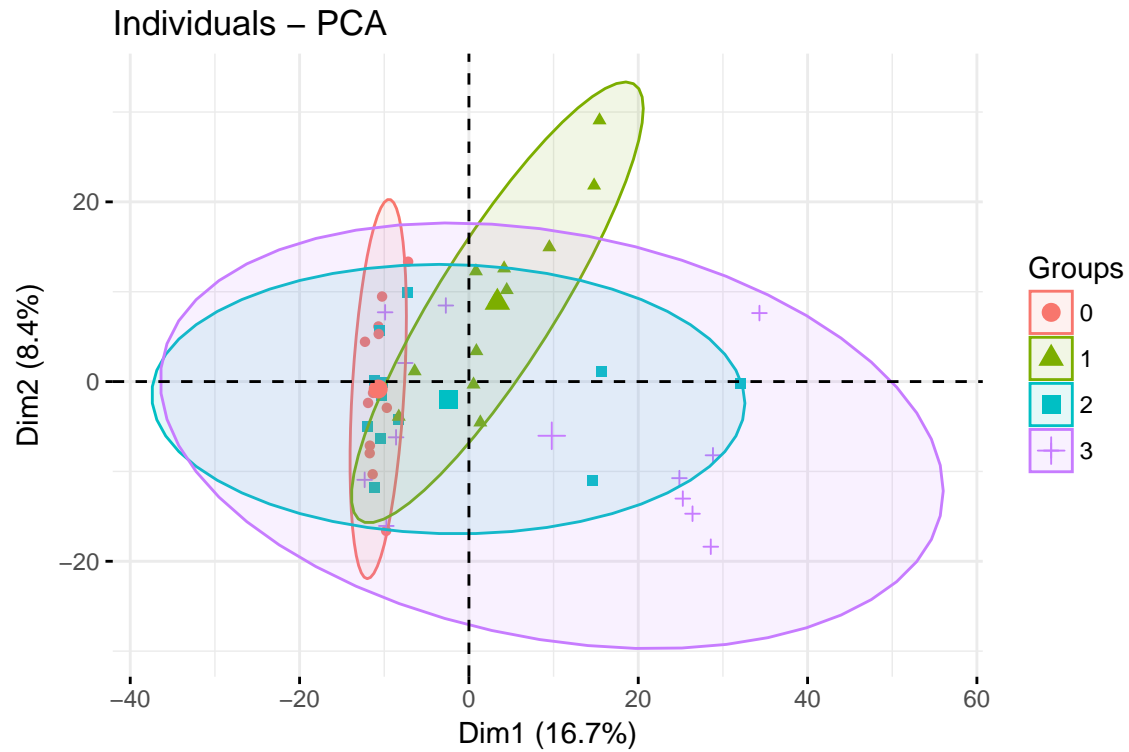
```
fviz_pca_ind(PC,  
  col.ind = "cos2", # Color by the quality of representation  
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
  repel = TRUE      # Avoid text overlapping  
)
```



```
fviz_pca_ind(PC, label="none", habillage=sml)
```

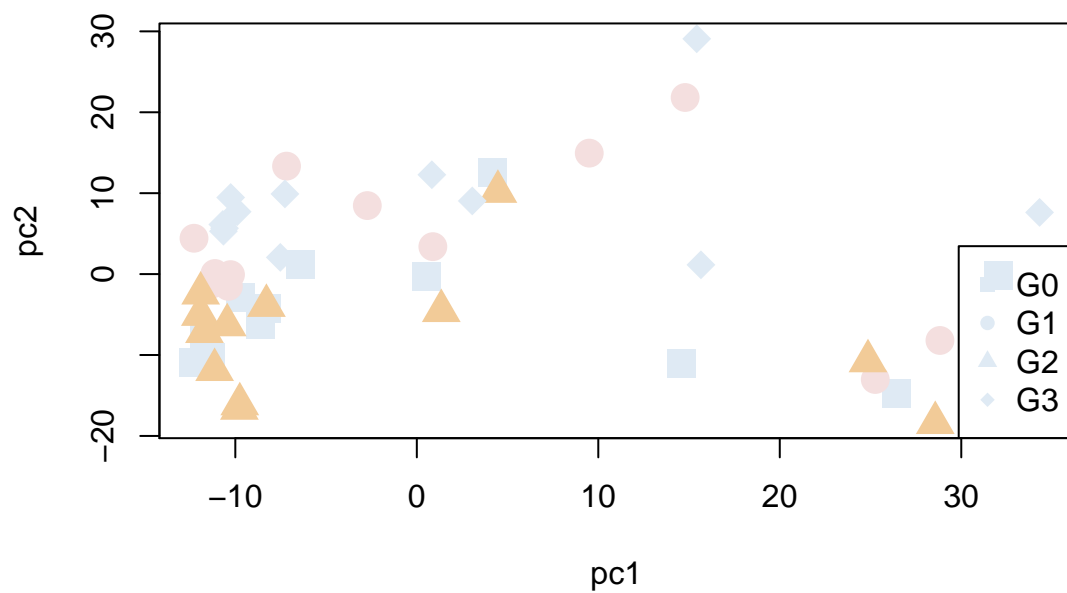


```
fviz_pca_ind(PC, label="none", habillage=sml,
  addEllipses=TRUE, ellipse.level=0.9)
```



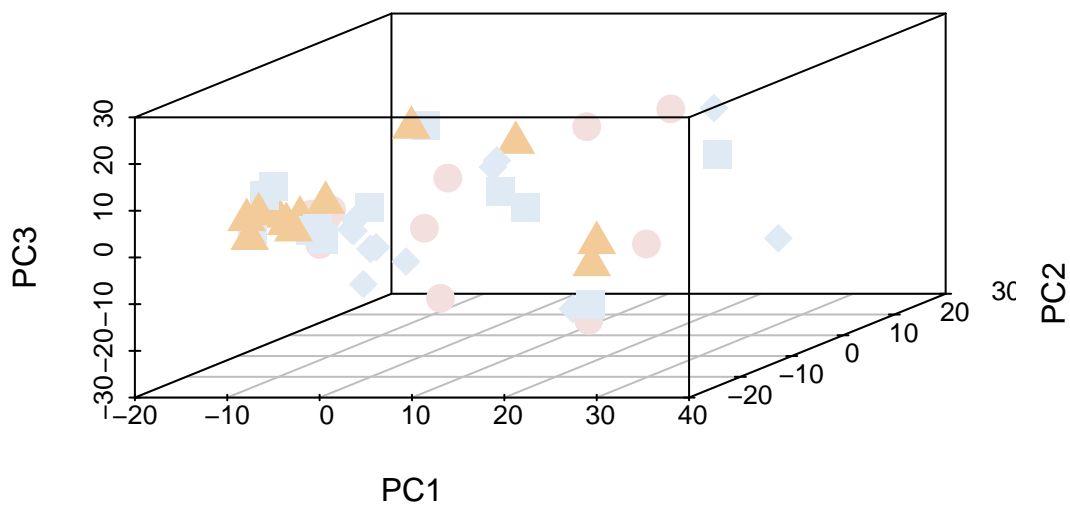
```
scores = predict(PC)

# extract PC1, PC2 and PC3
pc1 <- scores[,1]
pc2 <- scores[,2]
pc3 <- scores[,3]
shape <- as.numeric(f1) + 14 #add 14 so that it will has fill
par(mfrow=c(1,1))
plot(pc1, pc2, col=f1, pch=shape, cex=2)
legend("bottomright",col=f1, pch=unique(shape), paste(unique(f1)))
```



```
par(mfrow=c(1,1))
scatterplot3d(pc1, pc2, pc3, xlab="PC1", ylab="PC2", zlab="PC3", pch=shape, color=as.numeric(f1), main=
```

3D PCA



```
# Define p-value cut-off
p_cutoff <- 0.05
```

```

# Calculate the number of genes
numGenes <- nrow(exprs(gset))

#Plot MA for D3-D0
completeTopTable <- topTable(fit2,coef=2, adjust="none", number=numGenes)
completeTopTable<-completeTopTable[order(completeTopTable$ID,method='radix'),]

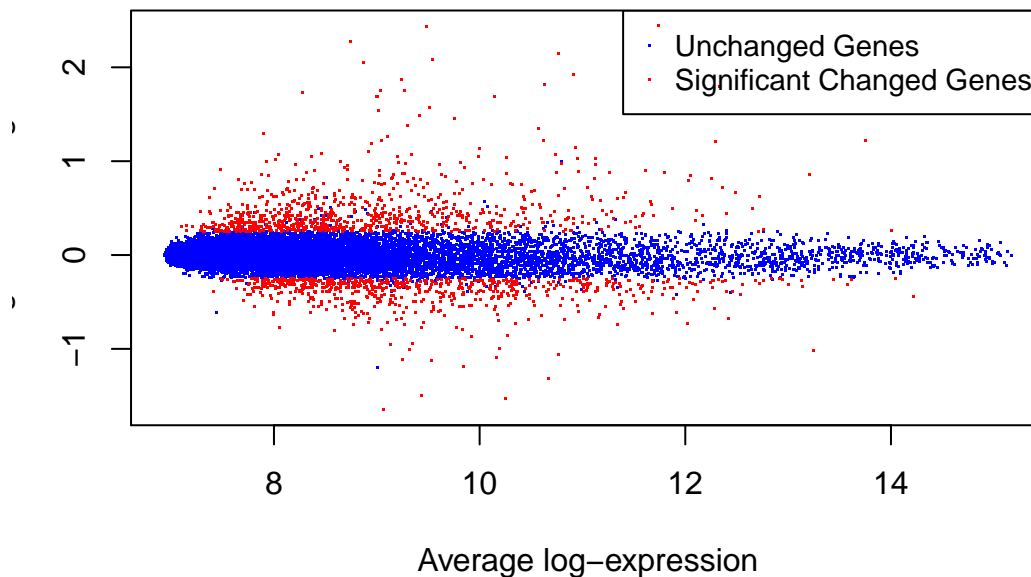
selected <- completeTopTable$adj.P.Val <= p_cutoff & (completeTopTable$logFC>=0.25 | completeTopTable$logFC<=-0.25)

status <- character (length=numGenes)
status <- rep ("Unchanged Genes", numGenes)
names (status) <- seq (1, numGenes, 1)
status [selected] <- "Significant Changed Genes"

limma::plotMA(fit2[,2], status=status, values=c("Unchanged Genes", "Significant Changed Genes"), col=c("blue", "red"),
text(x=12, y=9, labels=paste("P = ", p_cutoff), col="black", font=2)

```

G3 – G0



```

#Plot MA for D2-D0
completeTopTable <- topTable(fit2,coef=1, adjust="none", number=numGenes)
completeTopTable<-completeTopTable[order(completeTopTable$ID,method='radix'),]

selected <- completeTopTable$adj.P.Val <= p_cutoff & (completeTopTable$logFC>=0.25 | completeTopTable$logFC<=-0.25)

status <- character (length=numGenes)
status <- rep ("Unchanged Genes", numGenes)
names (status) <- seq (1, numGenes, 1)
status [selected] <- "Significant Changed Genes"

```

```
limma::plotMA(fit2[,1], status=status, values=c("Unchanged Genes", "Significant Changed Genes"), col=c(
text(x=12, y=9, labels=paste("P = ", p_cutoff), col="black", font=2)
```

