Shank3 sleep deprivation differential expression

October 03, 2018

Contents

Description]
Importing data	2
Plot PCA of log unnormalized data	2
Control Genes Negative control genes	
Normalizations TMM Normalization	
edgeR Differential Expression Analysis Shank3 Home Cage control VS Wild Type Home Cage Controls	1: 1:
DE TABLE + Positive Controls table	13
Heatmaps Heatmap gene by bene	
Group gene profiles by genotype	22 22 24

Description

This is the report of the analysis made for the paper Shank3 Modulates Sleep and Circadian Activity by Ashley M. Ingiosi, Taylor Wintler, Hannah Schoch, Kristan G. Singletary, Dario Righelli, Leandro G. Roser, Davide Risso, Marcos G. Frank and Lucia Peixoto.

Autism Spectrum Disorder (ASD) is the most prevalent neurodevelopmental disorder in the US that often co-presents with sleep problems. Sleep impairments in ASD predict the severity of ASD core diagnostic symptoms and have a considerable impact on the quality of life of caregivers. However, little is known about the underlying molecular mechanism(s) of sleep impairments in ASD. In this study we investigated the role of Shank3, a high confidence ASD gene candidate, in the regulation of sleep. We show that Shank3 mutant mice have problems falling asleep despite accumulating sleep pressure. Using RNA-seq we show that sleep deprivation doubles the differences in gene expression between mutants and wild types and downregulates circadian transcription factors Per3, Dec2, and Rev-erb α . Shank3 mutants also have trouble regulating

locomotor activity in the absence of light. Overall, our study shows that Shank3 is an important modulator of sleep and circadian activity.

Importing data

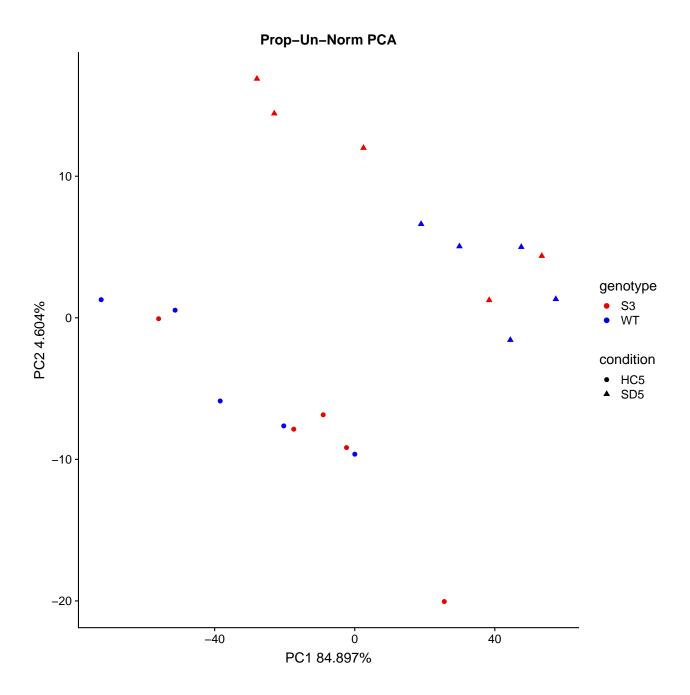
Importing data and filtering out those genes with cpm lesser than 1. We use the filtered.data method in NOISeq package.

Plot PCA of log unnormalized data

PCA Plot of filtered not-normalized data.

```
PlotPCAPlotlyFunction(counts.data.frame=log1p(filteredCountsProp),
    design.matrix=designMatrix,
    shapeColname="condition", colorColname="genotype", xPCA="PC1", yPCA="PC2",
    plotly.flag=FALSE, show.plot.flag=TRUE, prefix.plot="Prop-Un-Norm")
```

[1] FALSE



Control Genes

Negative control genes

Loading Negative Control Genes to normalize data

```
library(readxl)
sd.ctrls <- read_excel(path="./data/controls/Additional File 4 full list of BMC genomics SD&RS2.xlsx",
sd.ctrls <- sd.ctrls[order(sd.ctrls$adj.P.Val),]
sd.neg.ctrls <- sd.ctrls[sd.ctrls$adj.P.Val > 0.9, ]
```

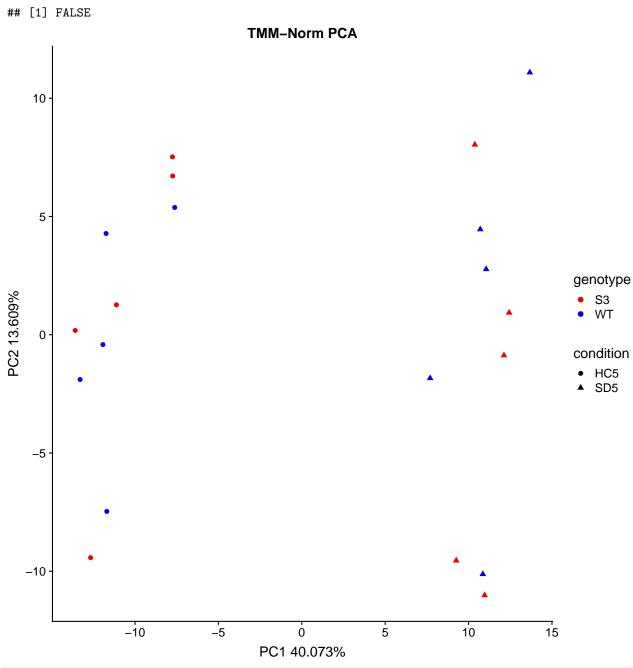
positive control genes

Loading Positive Control Genes to detect them during the differential expression step.

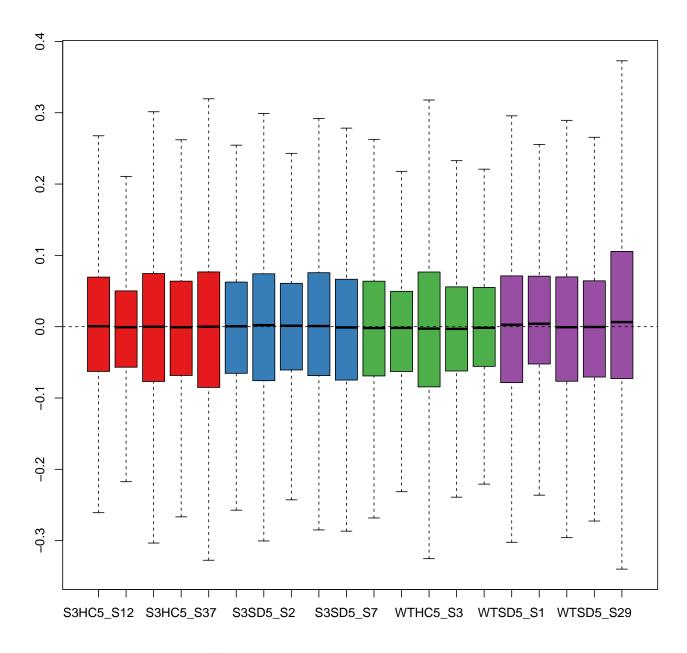
Normalizations

TMM Normalization

Normalizing data with TMM, as implemented in edgeR package, and plotting a PCA and an RLE plot of them.



pal <- RColorBrewer::brewer.pal(9, "Set1")
plotRLE(as.matrix(normPropCountsUqua), outline=FALSE, col=pal[designMatrix\$gcondition])</pre>

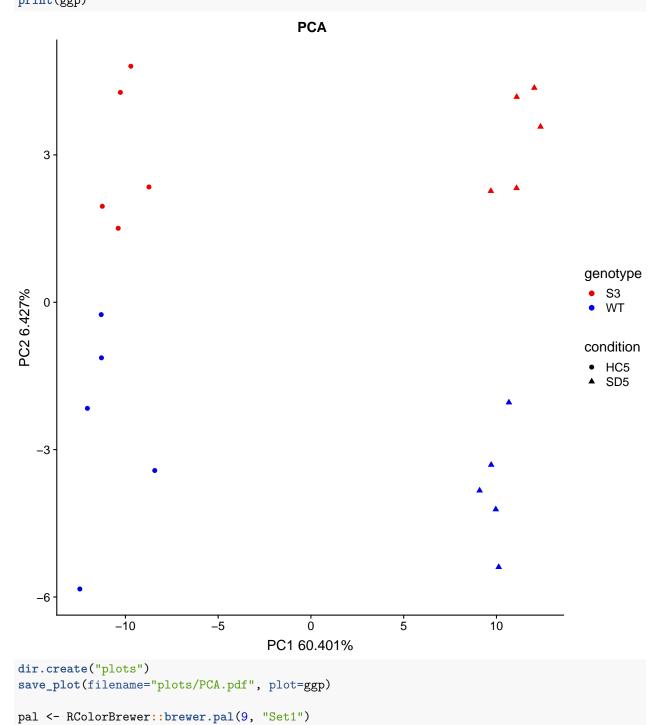


TMM + RUVs Normalization

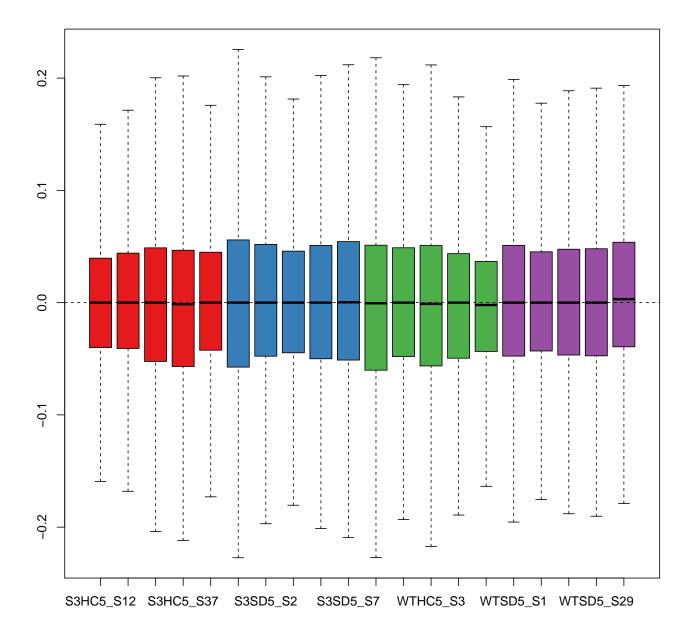
Applying a RUVs method of RUVSeq package on normalized data, in order to adjust the counts for the unwanted variation. And of corse we plot a PCA and an RLE plot on these data.

```
plotly.flag=FALSE, show.plot.flag=FALSE, save.plot=FALSE,
prefix.plot=NULL)
```

[1] FALSE
print(ggp)



plotRLE(normExprData, outline=FALSE, col=pal[designMatrix\$gcondition])



edgeR Differential Expression Analysis

Making differential expression analysis with edgeR package on four different contrasts.

Here is a brief legend:

- WTSD5: Wild Type Sleep Deprivation 5 days.
- KOHC5: Knock Out Home Cage Control 5 days.
- KOSD5: Knock Out Sleep Deprivation 5 days.

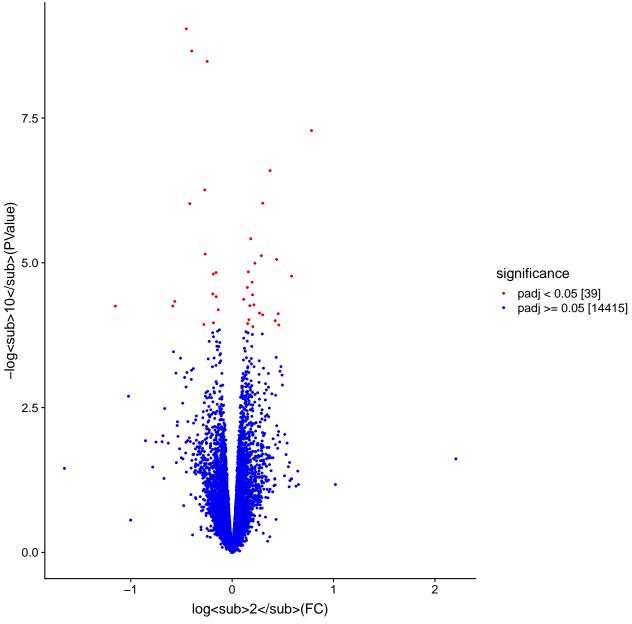
```
padj.thr <- 0.05
venn.padgj.thr <- 0.1
desMat <- cbind(designMatrix, ruvedSExprData$W)
colnames(desMat) <- c(colnames(designMatrix), colnames(ruvedSExprData$W))</pre>
```

Shank3 Home Cage control VS Wild Type Home Cage Controls

volcano plot

A volcano plot of differential expressed genes.

S3HC5 - WTHC5 Volcano Plot



```
de <- sum(res.o$FDR < padj.thr)
nde <- sum(res.o$FDR >= padj.thr)
detable <- cbind(de,nde)
rownames(detable) <- names(rescList1)[1]
ddetable <- detable

tot.ctrls <- dim(sd.pos.ctrls)[1]
idx.pc <- which(tolower(res.o$gene) %in% tolower(sd.pos.ctrls[,1]))
tot.pc.de <- sum(res.o$FDR[idx.pc] < padj.thr)
tot.pc.nde <- length(idx.pc) - tot.pc.de

wt <- res.o[which(res.o$FDR < padj.thr),]
wt.sign.genes.entrez <- rownames(res.o)[which(res.o$FDR < venn.padgj.thr)]</pre>
```

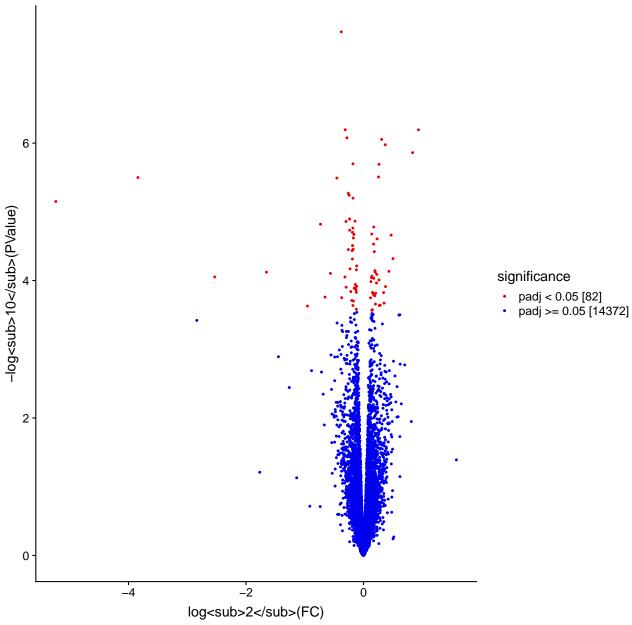
```
kowthc5 <- res.o[which(res.o$FDR < padj.thr),]
kowthc5.sign.genes.entrez <- rownames(res.o)[which(res.o$FDR < venn.padgj.thr)]</pre>
```

Shank3 Sleed Deprivation VS Wild Type Sleep Deprivation

volcano plot

A volcano plot of differential expressed genes.





DE TABLE + Positive Controls table

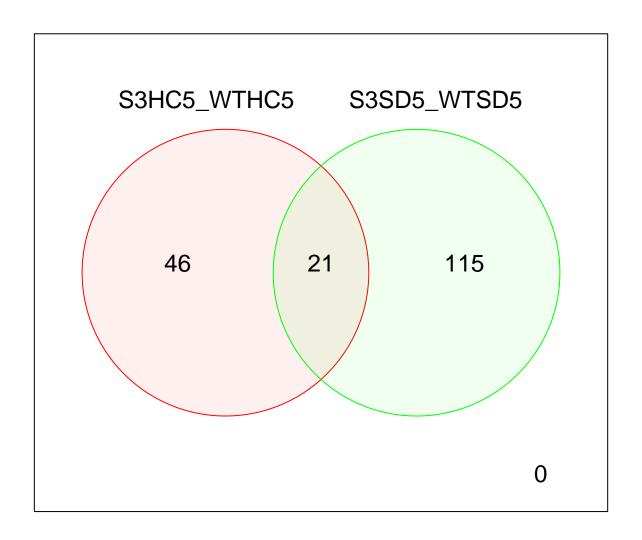
We present a summarization of the results. The first table is a summarization on how many genes are Differentially Expressed. The second table explains on the first column how many positive controls we have, on the second column how many positive controls have been identified over the differentially expressed genes, and, finally, on the third column how many positive controls have beed identified on the NOT differentially expressed genes.

Venn Diagram

KOHC5-WTHC5 vs KOSD5-WTSD5

We take the results of the two contrasts. Knock Out Sleed Deprivation VS Wild Type Sleep Deprivation and Knock Out Home Cage control VS Wild Type Home Cage Controls . And plot the results in a Venn Diagram

S3HC5_WTHC5 venn S3SD5_WTSD5



Heatmaps

Setting up the data structures for the heatmps.

```
de.genes.symb <- attachGeneColumnToDf(as.data.frame(de.genes.entr,</pre>
                                                      row.names=de.genes.entr),
                                     genesMap=gene.map,
                                     rowNamesIdentifier="ENTREZID",
                                     mapFromIdentifier="ENTREZID",
                                     mapToIdentifier="SYMBOL")
# de.genes.symb[which(is.na(de.genes.symb$gene)),]
de.genes.symb$gene[which(de.genes.symb$de.genes.entr=="100039826")] <- "Gm2444" ## not annotated in no
de.genes.symb$gene[which(de.genes.symb$de.genes.entr=="210541")] <- "Entrez:210541" ## not annotated in
de.genes.counts <- normExprData[match(de.genes.symb$de.genes.entr, rownames(normExprData)),]</pre>
rownames(de.genes.counts) <- de.genes.symb$gene</pre>
de.gene.means <- computeGeneMeansOverGroups(counts=de.genes.counts,</pre>
                             design=designMatrix, groupColumn="gcondition")
library(gplots)
library(clusterExperiment)
color.palette = clusterExperiment::seqPal3#c("black", "yellow")
pal <- colorRampPalette(color.palette)(n = 1000)</pre>
library(pheatmap)
filter2 <- rowMeans(de.gene.means)>0
filter <- apply(de.gene.means, 1, function(x) log(x[4]/x[3]) * log(x[2]/x[1]) < 0)
filter[is.na(filter)] <- FALSE</pre>
```

Heatmap gene by bene

```
SD5
                                                                                              0.1
                                                                                                   genotype
                                                                                              0.05
                                                                                                      S3
                                                                                                      WT
                                                                                              -0.05
                                                                             Cluster: 2 Size: 42
                                                                                              -0.1
                                                                             Cluster: 1 Size: 75
                                                                             Cluster: 3 Size: 61
                      S3HC5_S8
                                 WTHC5_S3
                                    WTHC5_S3
                                       WTHC5_S36
                                                        WTSD5_S22
                                                            S3SD5_S2
                          WTHC5_S20
                                           WTSD5_S1
                                              WTSD5_S38
                                                  WTSD5_S29
                                                     WTSD5_S13
                                                               S3SD5_S7
                                                                  S3SD5_S1
                             WTHC5_S16
clusterized.genes <- as.data.frame(ph1$kmeans$cluster)</pre>
gene.map <- convertGenesViaMouseDb(gene.list=rownames(clusterized.genes), fromType="SYMBOL")</pre>
converted.clusterized.gens <- attachGeneColumnToDf(mainDf=clusterized.genes, genesMap=gene.map,</pre>
                       rowNamesIdentifier="SYMBOL", mapFromIdentifier="SYMBOL", mapToIdentifier="ENTREZID"
converted.clusterized.gens$gene[which(rownames(converted.clusterized.gens)=="Gm2444")] <- "100039826"
converted.clusterized.gens$gene[which(rownames(converted.clusterized.gens)=="Entrez:210541")] <- "2105"
converted.clusterized.gens <- converted.clusterized.gens[order(converted.clusterized.gens$`ph1$kmeans$c</pre>
save_pheatmap_pdf(filename="plots/heatmap_kmeans_k3.pdf", plot=ph1, width=20, height=20)
## pdf
##
      2
```

condition

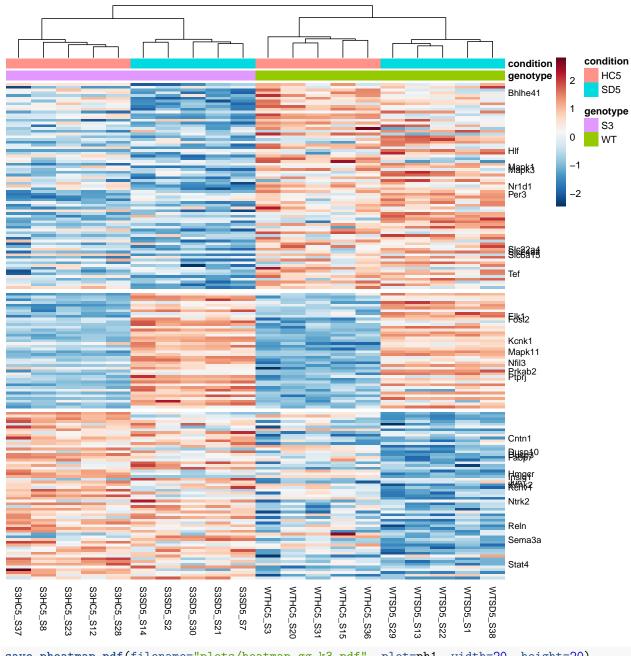
HC5

0.15

condition

genotype

```
WriteDataFrameAsTsv(data.frame.to.save=converted.clusterized.gens, file.name.path="plots/clustered_gene"
ord.de.genes.counts <- de.heatmap[match(rownames(converted.clusterized.gens), rownames(de.heatmap)),]
idx <- which(!(rownames(ord.de.genes.counts) %in% gene_names))</pre>
rownames(ord.de.genes.counts)[idx] <- ""</pre>
gaps.row <- c()</pre>
for(i in c(1:3))
    li <- length(which(converted.clusterized.gens$`ph1$kmeans$cluster`==i))</pre>
    1 <- ifelse(i!=1, gaps.row[i-1]+li, li)</pre>
    gaps.row <- c(gaps.row, 1)</pre>
}
heatmap_data_scaled <- t(scale(t(log(ord.de.genes.counts+1)), center = TRUE, scale = TRUE))
library(dendextend)
column_dend <- as.dendrogram(hclust(dist(t(heatmap_data_scaled))))</pre>
ord <- labels(column_dend)</pre>
ord[11:15] <- labels(column_dend)[16:20]
ord[16:20] <- labels(column_dend)[11:15]
column_dend <- rotate(column_dend, ord)</pre>
ph1 <- pheatmap(heatmap_data_scaled, cluster_cols=as.hclust(column_dend), scale="none",</pre>
             color=pal, border_color=NA, fontsize_row=9, fontsize_col=9, cluster_rows=FALSE,
             annotation_col=ann.col, gaps_row=gaps.row)
```



save_pheatmap_pdf(filename="plots/heatmap_gg_k3.pdf", plot=ph1, width=20, height=20)

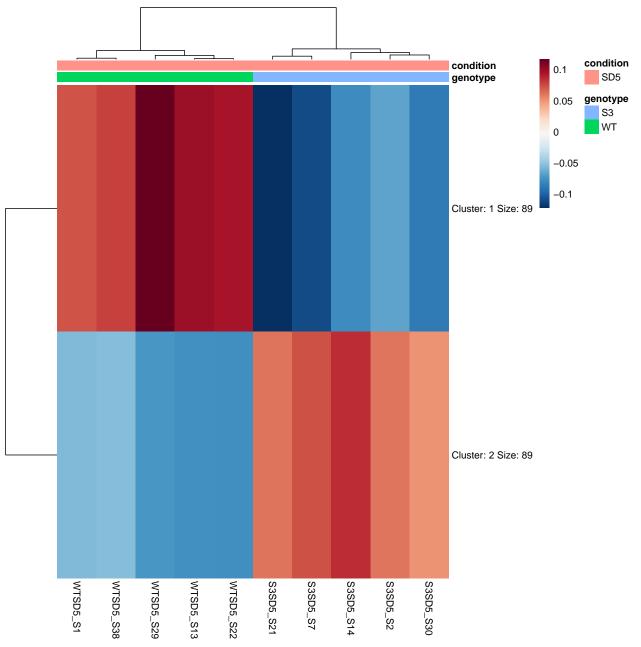
```
## pdf
##
```

other heatmaps

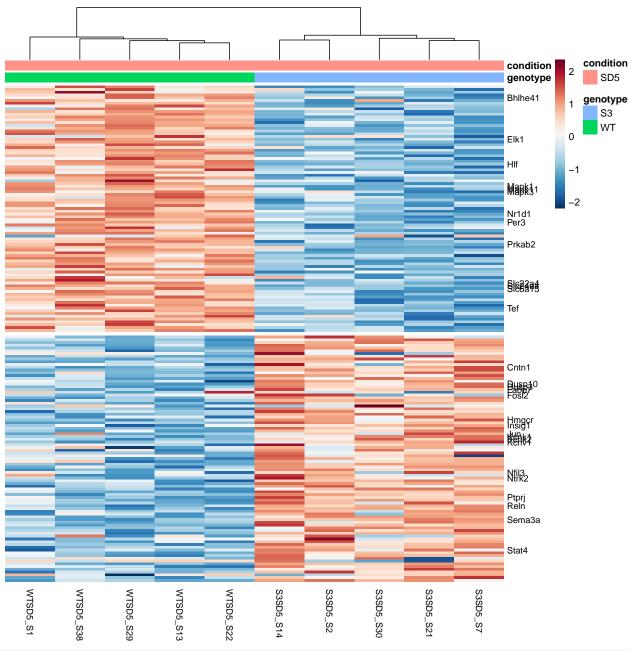
```
heatmap_data <- t(scale(t(log(ord.de.genes.counts+1)), center = TRUE, scale = FALSE))</pre>
ph1 <- pheatmap(heatmap_data, cluster_cols=as.hclust(column_dend), scale="none",</pre>
            color=pal, border_color=NA, fontsize_row=9, fontsize_col=9, cluster_rows=FALSE,
            annotation_col=ann.col, gaps_row=gaps.row,
```

condition condition HC5 genotype SD5 0.5 Bhlhe41 genotype S3 WT HIf -1 Mapk3 -1.5 Nr1d1 Per3 -2 \$168394 Tef Fbs12 Kcnk1 Mapk11 Nfil3 Brkab2 Ptprj Cntn1 Pusp₂0 Higher Kenka Ntrk2 Reln Sema3a Stat4 WTSD5_S38 S3HC5_S37 S3HC5_S8 S3HC5_S12 S3HC5_S28 S3SD5_S14 S3SD5_S30 S3SD5_S21 S3SD5_S7 WTHC5_S3 WTHC5_S31 WTHC5_S15 WTHC5_S36 WTSD5_S29 WTSD5_S13 WTSD5_S22 WTSD5_S1 S3HC5_S23 S3SD5_S2 WTHC5_S20 save_pheatmap_pdf(filename="plots/heatmap_gg_k3_no_scale.pdf", plot=ph1, width=20, height=20) ## pdf ## ## Only SD samples heatmap_data <- t(scale(t(log(de.heatmap[, grep("^SD", desMat[,2])]+1)), center = TRUE, scale = FALSE))</pre> ph1 <- pheatmap(heatmap_data, cluster_cols=TRUE, scale="none",</pre> color=pal, border_color=NA, fontsize_row=10, kmeans_k=2, annotation_col=ann.col)

breaks = c(min(heatmap_data), seq(quantile(as.vector(heatmap_data), .01), quantile(as.vector



idx <- which(!(rownames(ord.de.genes.counts) %in% gene_names))</pre>

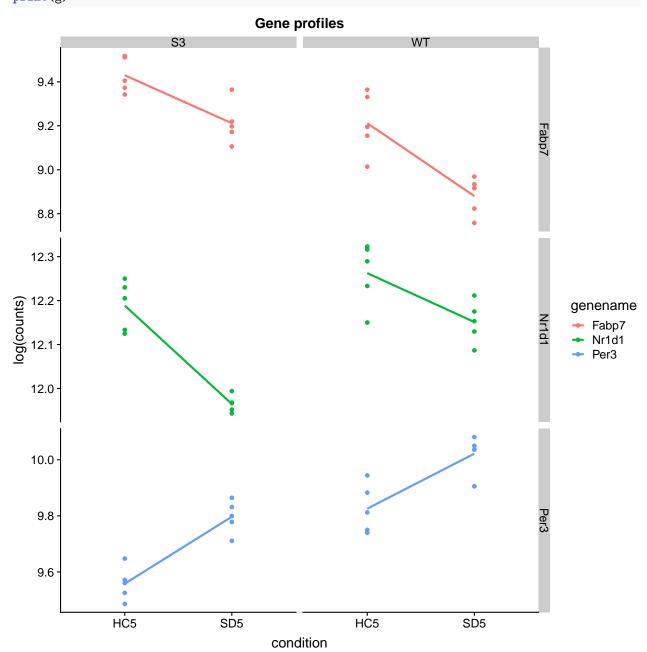


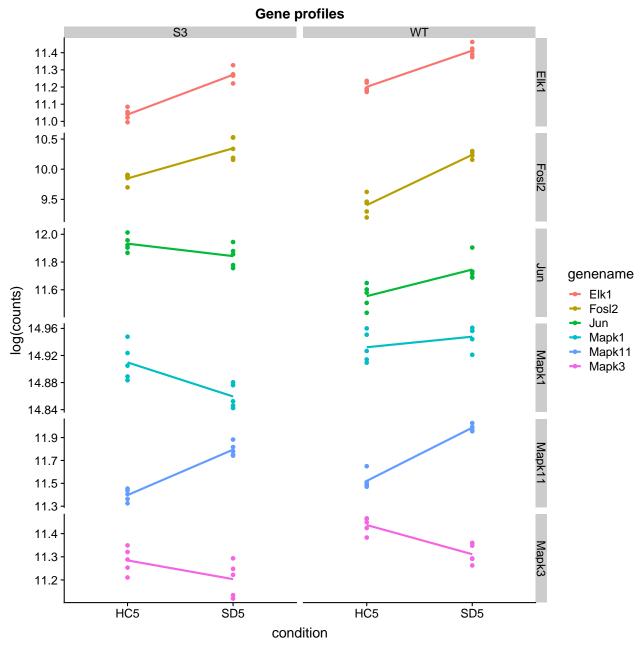
save_pheatmap_pdf(filename="plots/heatmap_gg_sd_only.pdf", plot=ph1, width=20, height=20)

```
## pdf
## 2
```

Group gene profiles

Group gene profiles by genotype





Group gene profiles by condition

