RNAseq: CD8+ OT-I Stimulation Timecourse in SH2B3 mice

Protocol:

1. Use salmon to quantify transcripts

2. Use tximeta to get count matrix

3. use with DEseq for expression

import data table & create table with file paths to sample quant files

```
## importing quantifications
```

```
## reading in files with read_tsv
```

```
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
## found matching transcriptome:
## [ Ensembl - Mus musculus - release 99 ]
## loading existing EnsDb created: 2022-04-14 17:58:52
## loading existing transcript ranges created: 2022-04-14 17:58:54
## loading existing EnsDb created: 2022-04-14 17:58:52
## obtaining transcript-to-gene mapping from database
## loading existing gene ranges created: 2022-04-14 18:01:41
## summarizing abundance
## summarizing counts
## summarizing length
```

Using DESeq to get read counts

```
#differential expression with DESeq2
dds = DESeqDataSet(gse, design = ~ SampleGenotype + SampleTime)
```

```
## using counts and average transcript lengths from tximeta
```

```
## Warning in DESeqDataSet(gse, design = ~SampleGenotype + SampleTime): some
## variables in design formula are characters, converting to factors
```

```
#filter out empty rows

#get empty rows
nonzero = rowSums(counts(dds)) > 1
dds = dds[nonzero, ]

#factor so genotypes are grouped together
dds$SampleGenotype = factor(dds$SampleGenotype, levels = c("WT", "KO")) %>%
  relevel(dds$SampleGenotype, ref = "WT")
```

```
dds = DESeq(dds)
```

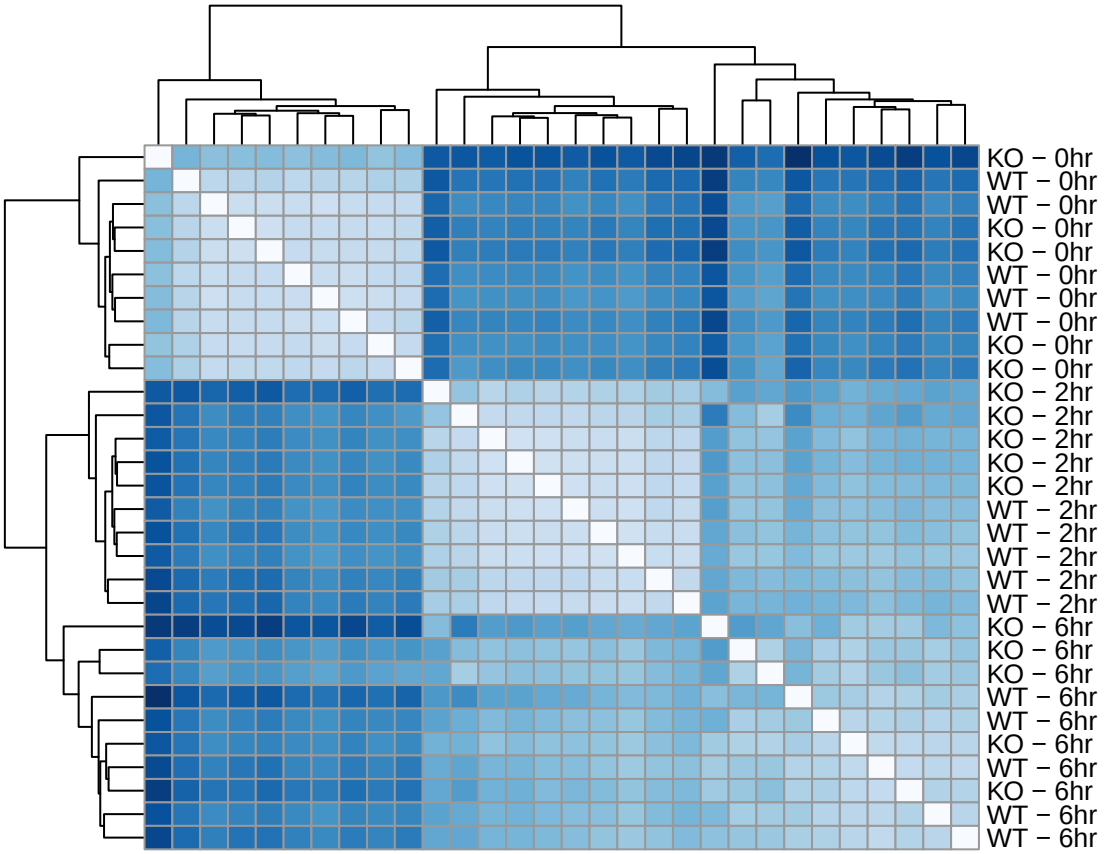
```
## estimating size factors
```

```
## using 'avgTxLength' from assays(dds), correcting for library size
```

```
## estimating dispersions
## gene-wise dispersion estimates
## mean-dispersion relationship
## final dispersion estimates
## fitting model and testing
```

```
#log2 fold changes and pvalues for WT vs KO
dds.res = results(dds, contrast = c("SampleGenotype", "WT", "KO"))
```

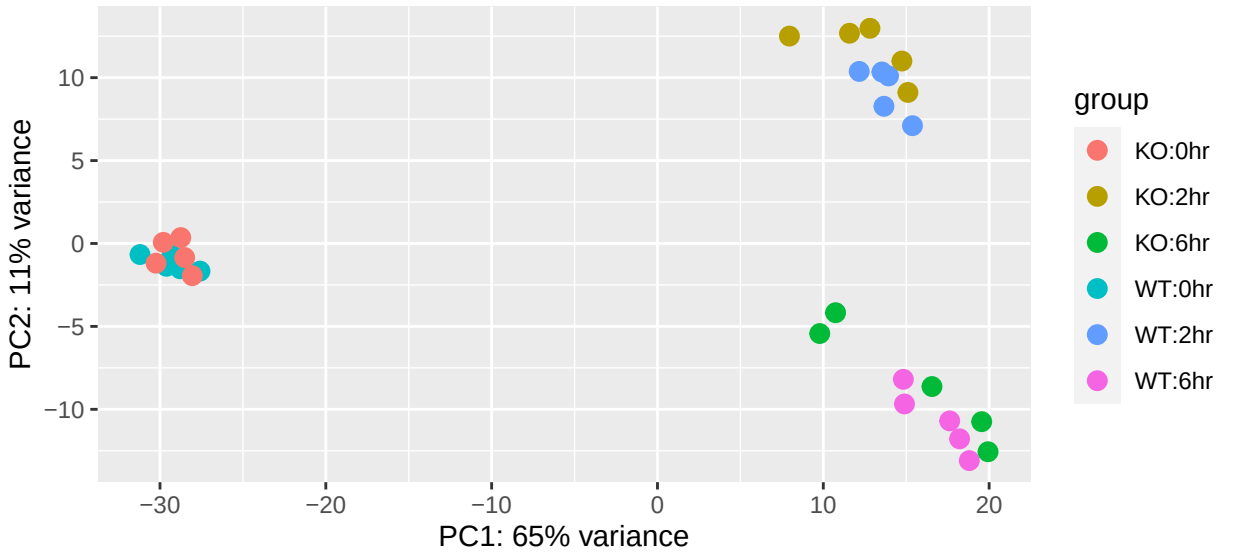
Annotating dataframe with gene symbols



Visualizing sample distances

PCA plot (all time points)

```
library(vsn)

#variance stabilizing transformation
vsd = vst(dds)
plotPCA(vsd, intgroup = c("SampleGenotype", "SampleTime"))
```



Gene cluster heatmap

```
library(genefilter)
```

```
##
## Attaching package: 'genefilter'
```

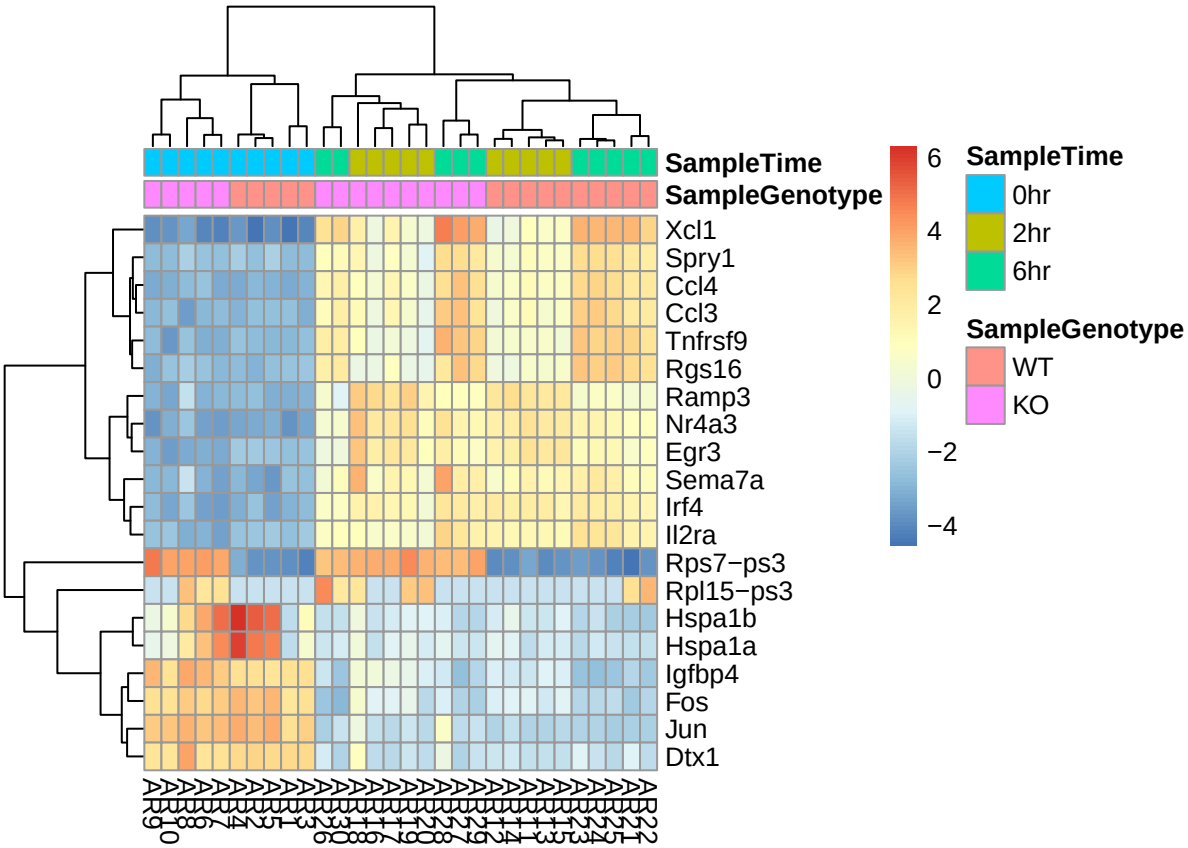
```
## The following objects are masked from 'package:MatrixGenerics':
##
##     rowSds, rowVars
```

```
## The following objects are masked from 'package:matrixStats':
##
##     rowSds, rowVars
```

```
topVarGenes = head(order(rowVars(assay(vsd)), decreasing = T), 20)
mat = assay(vsd)[ topVarGenes, ]

rownames(mat) <- mapIds(org.Mm.eg.db,
                    keys = rownames(mat),
                    column = "SYMBOL",
                    keytype = "ENSEMBL",
                    multiVals = "first")
```

```
## 'select()' returned 1:1 mapping between keys and columns
```

```
mat = mat - rowMeans(mat)
anno = as.data.frame(colData(vsd)[, c("SampleGenotype", "SampleTime")])



# png(filename = "h1eatmap.png", width = 6, height = 6, units = "in", res = 300, type = "cairo")
pheatmap::pheatmap(mat, annotation_col = anno)
```




```
# while (!is.null(dev.list()))  dev.off()
```

Processing data as time course: Start with likelihood ratio test to remove genotype specific differences - the remaining genes with small p values showed genotype specific effects after time 0hr.

```
library(tidyverse)
```

```
## Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
## had status 1
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v stringr 1.4.0
## v tidyr   1.1.4     v forcats 0.5.1
## v readr   2.1.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x IRanges::collapse()     masks dplyr::collapse()
## x Biobase::combine()      masks BiocGenerics::combine(), dplyr::combine()
## x matrixStats::count()    masks dplyr::count()
## x IRanges::desc()         masks dplyr::desc()
## x tidyr::expand()         masks S4Vectors::expand()
## x dplyr::filter()         masks stats::filter()
## x S4Vectors::first()      masks dplyr::first()
## x dplyr::lag()            masks stats::lag()
## x ggplot2::Position()     masks BiocGenerics::Position(), base::Position()
## x purrr::reduce()         masks GenomicRanges::reduce(), IRanges::reduce()
## x S4Vectors::rename()     masks dplyr::rename()
```

```
## x AnnotationDbi::select() masks dplyr::select()
## x IRanges::slice()        masks dplyr::slice()
## x readr::spec()           masks genefilter::spec()
```

```
ddsTC = DESeqDataSet(gse, design = ~ SampleGenotype + SampleTime + SampleGenotype:SampleTime)
```

```
## using counts and average transcript lengths from tximeta
```

```
## Warning in DESeqDataSet(gse, design = ~SampleGenotype + SampleTime +
## SampleGenotype:SampleTime): some variables in design formula are characters,
## converting to factors
```

```
#likelyhood ratio test
ddsTC = DESeq(ddsTC, test="LRT", reduced = ~ SampleGenotype + SampleTime)
```

```
## estimating size factors
```

```
## using 'avgTxLength' from assays(dds), correcting for library size
```

```
## estimating dispersions
```

```
## gene-wise dispersion estimates
```

```
## mean-dispersion relationship
```

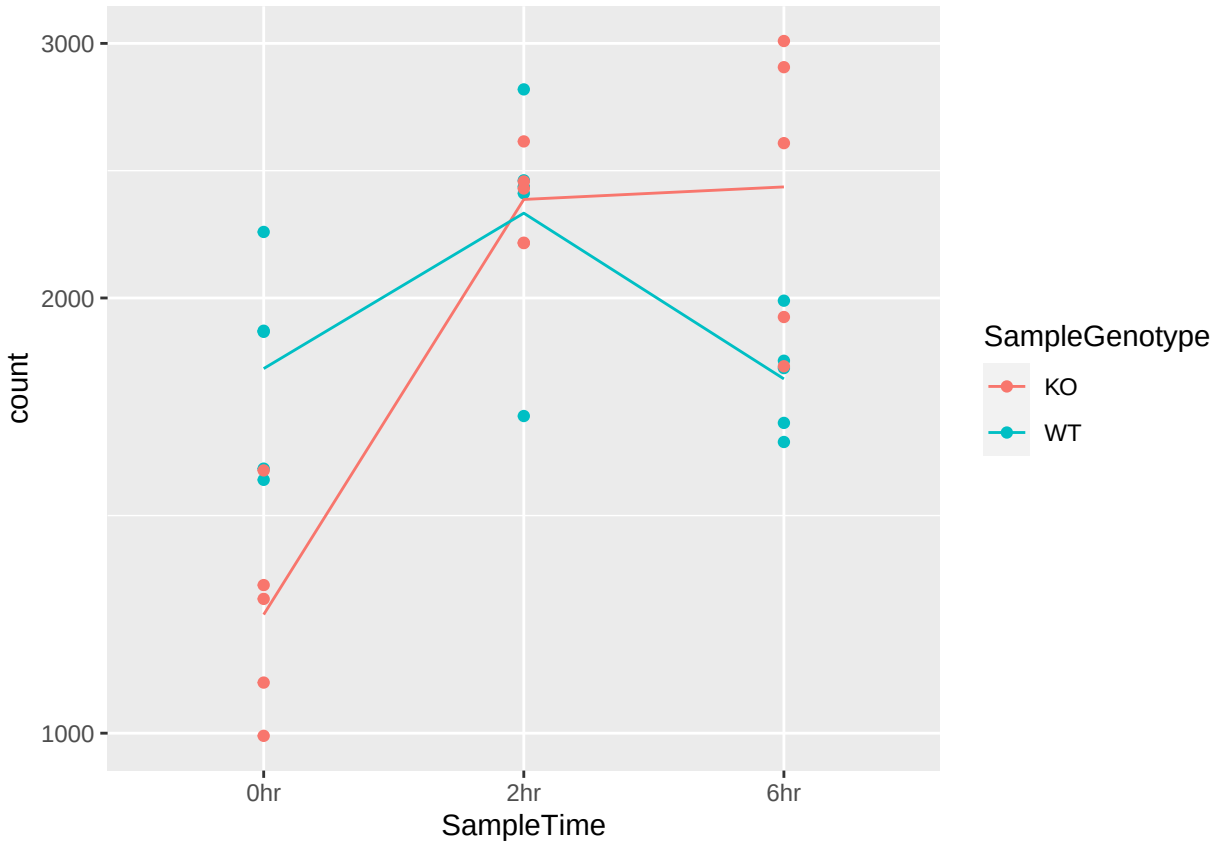
```
## final dispersion estimates
```

```
## fitting model and testing
```

```
resTC = results(ddsTC)
resTC$symbol = mcols(ddsTC)$symbol

stim = plotCounts(ddsTC, which.min(resTC$padj),
                  intgroup = c("SampleGenotype", "SampleTime"), returnData = T)
stim$hour = as.numeric(as.character(stim$SampleTime))
```

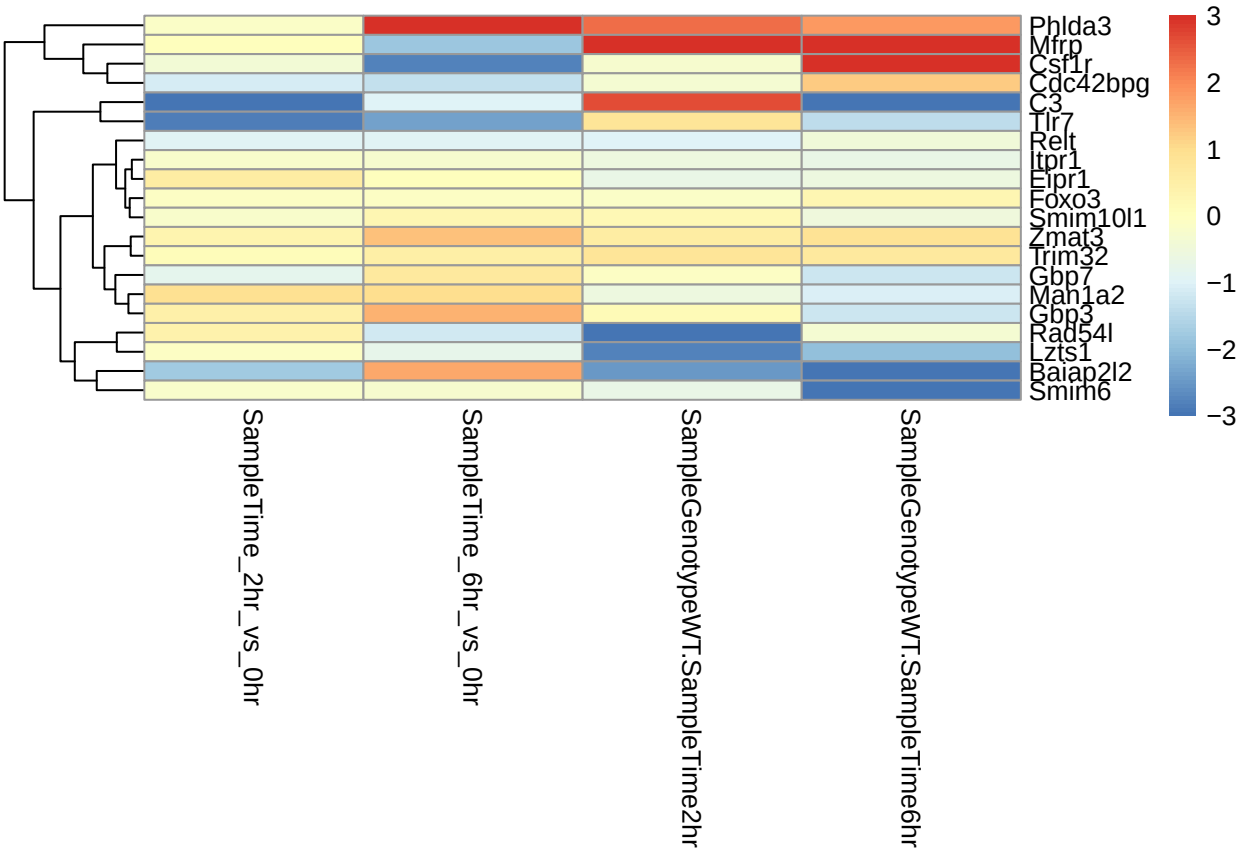
```
## Warning: NAs introduced by coercion
```

```
ggplot(stim,
       aes(x = SampleTime, y = count, color = SampleGenotype, group = SampleGenotype)) + geom_point() +
  stat_summary(fun = mean, geom = "line") + scale_y_log10()
```



Cluster significant genes by profile

```
## 'select()' returned 1:1 mapping between keys and columns
```


Phlda3
Mfrp
Csf1r
Cdc42bpg
C3
Tlr7
Relt
Itpr1
Elpr1
Foxo3
Smim10l1
Zmat3
Trim32
Gbp7
Mah1a2
Gbp3
Rad54l
Lzts1
Baiap2l2
Smim6
3
2
1
0
−1
−2
−3
SampleTime_2hr_vs_0hr
SampleTime_6hr_vs_0hr
SampleGenotypeWT.SampleTime2hr
SampleGenotypeWT.SampleTime6hr