Class16

Zixuan Zeng(A16142927)

```
library(tximport)
folders <- dir(pattern="SRR21568*")
samples <- sub("_quant", "", folders)
files <- file.path( folders, "abundance.tsv" )
names(files) <- samples

txi.kallisto <- tximport(files, type = "kallisto", txOut = TRUE)
```

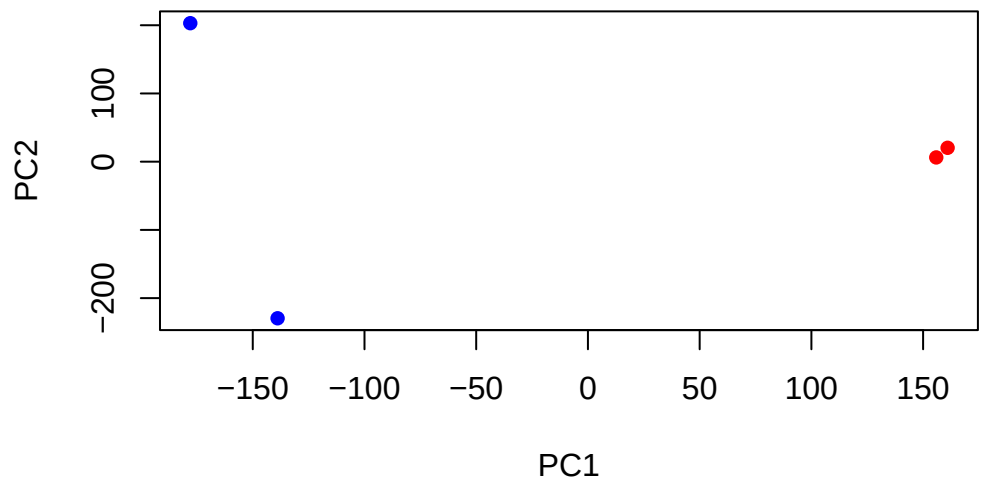
```
Note: importing `abundance.h5` is typically faster than `abundance.tsv`

reading in files with read_tsv

1 2 3 4
```

```
to.keep <- rowSums(txi.kallisto$counts) > 0
kset.nonzero <- txi.kallisto$counts[to.keep,]
keep2 <- apply(kset.nonzero,1,sd)>0
x <- kset.nonzero[keep2,]
```

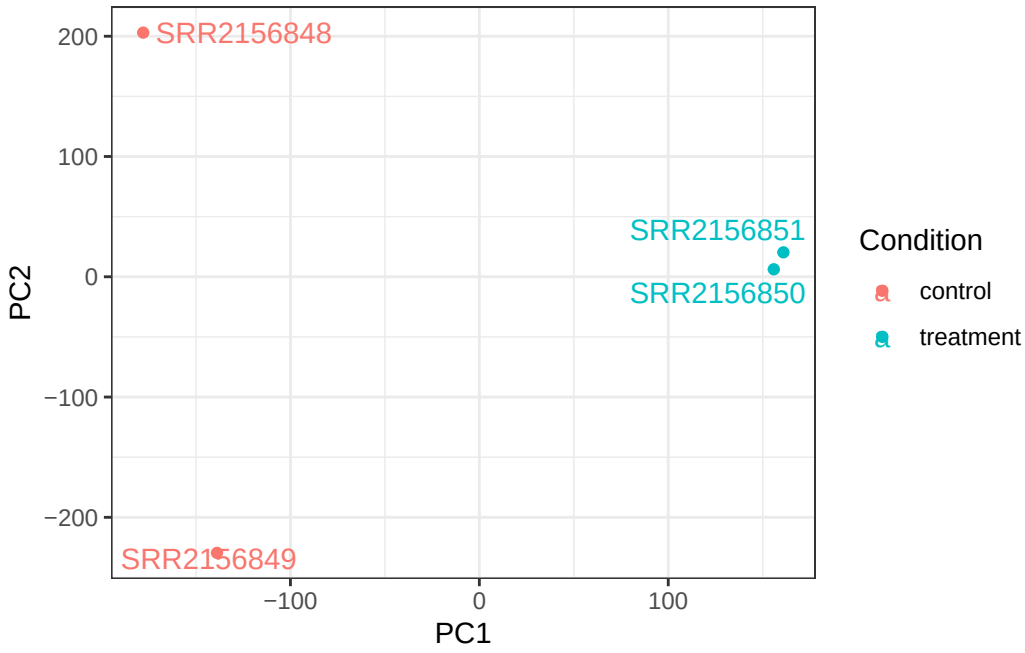
```
pca <- prcomp(t(x), scale=TRUE)
summary(pca)
```

```
Importance of components:
                            PC1      PC2      PC3      PC4
Standard deviation     183.6353 177.3561 171.3005 4.91e-11
Proportion of Variance   0.3568   0.3328   0.3105 0.00e+00
Cumulative Proportion    0.3568   0.6895   1.0000 1.00e+00
```

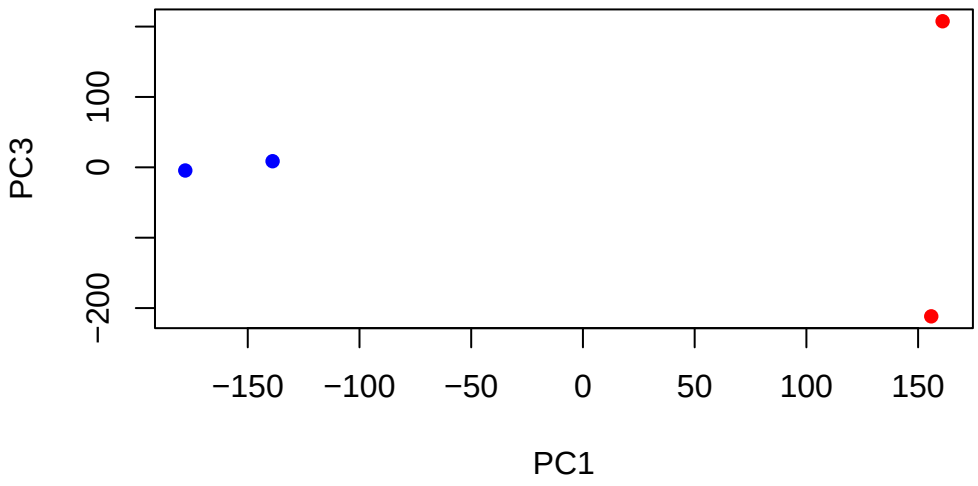
```
plot(pca$x[,1], pca$x[,2],
     col=c("blue","blue","red","red"),
     xlab="PC1", ylab="PC2", pch=16)
```



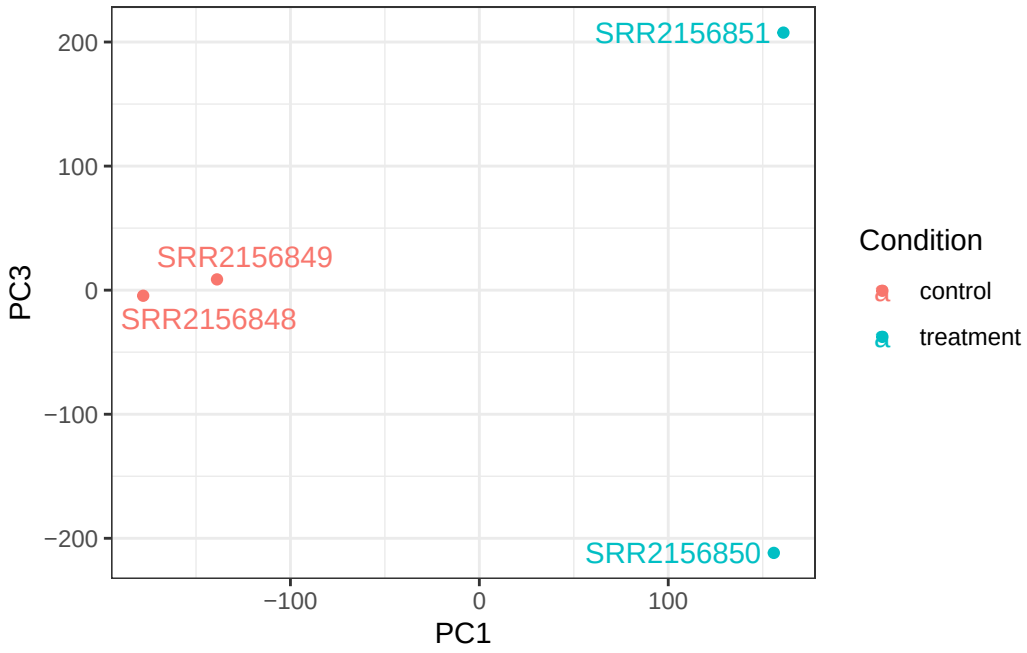

```
library(ggplot2)
library(ggrepel)

# Make metadata object for the samples
colData <- data.frame(condition = factor(rep(c("control", "treatment"), each = 2)))
rownames(colData) <- colnames(txi.kallisto$counts)

# Make the data.frame for ggplot
y <- as.data.frame(pca$x)
y$Condition <- as.factor(colData$condition)

ggplot(y) +
  aes(PC1, PC2, col=Condition) +
  geom_point() +
  geom_text_repel(label=rownames(y)) +
  theme_bw()
```




```
pca <- prcomp(t(x), scale=TRUE)
plot(pca$x[,1], pca$x[,3],
     col=c("blue","blue","red","red"),
     xlab="PC1", ylab="PC3", pch=16)
```



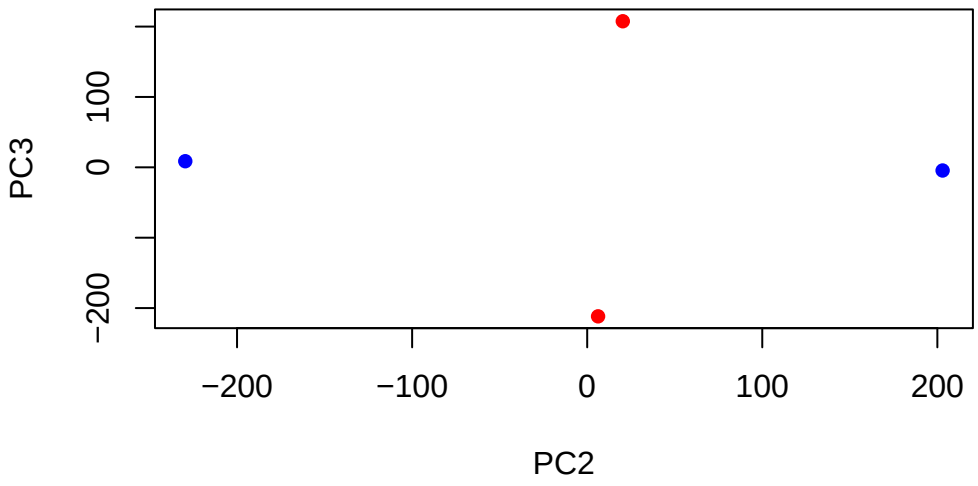

```
ggplot(y) +
  aes(PC1, PC3, col=Condition) +
  geom_point() +
  geom_text_repel(label=rownames(y)) +
  theme_bw()
```



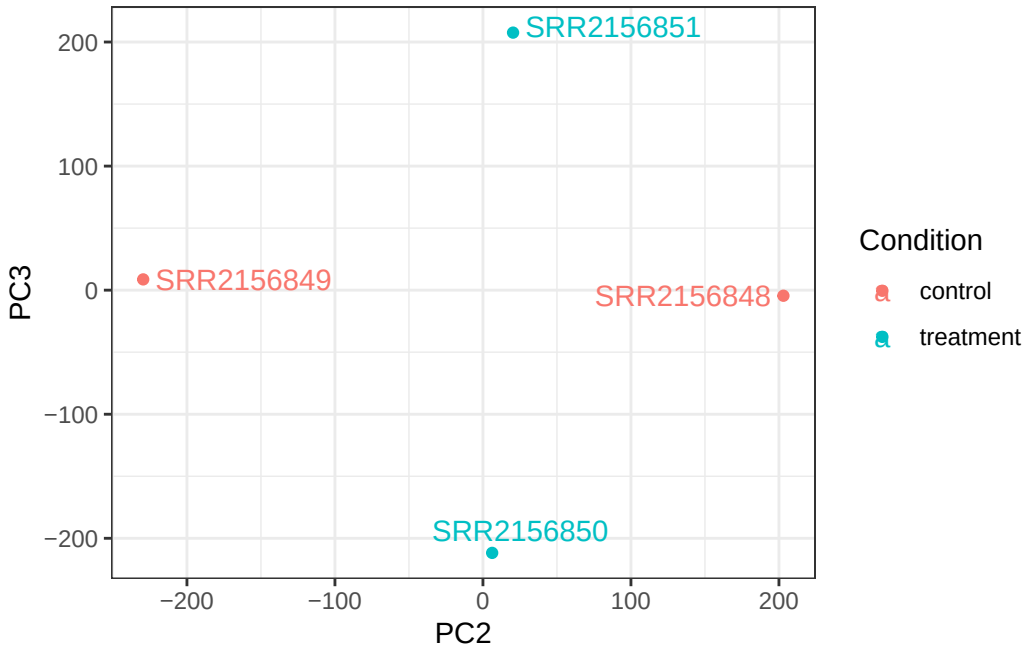

```
pca <- prcomp(t(x), scale=TRUE)
summary(pca)
```

```
Importance of components:
                            PC1      PC2      PC3      PC4
Standard deviation     183.6353 177.3561 171.3005 4.91e-11
Proportion of Variance   0.3568   0.3328   0.3105 0.00e+00
Cumulative Proportion    0.3568   0.6895   1.0000 1.00e+00
```

```
plot(pca$x[,2], pca$x[,3],
     col=c("blue","blue","red","red"),
     xlab="PC2", ylab="PC3", pch=16)
```




```
ggplot(y) +
  aes(PC2, PC3, col=Condition) +
  geom_point() +
  geom_text_repel(label=rownames(y)) +
  theme_bw()
```


SRR2156851
SRR2156849
SRR2156848
SRR2156850
200
100
0
−100
−200
PC3
−200
−100
0
100
200
PC2
Condition
control
treatment