

Class17

Aileen Andrade (PID A17033749)

2025-03-04

Instance ID

i-0b7b517a68441b279 (bimm143_aileenandrade) Open an SSH client.

Locate your private key file. The key used to launch this instance is bimm143_aileenandrade.pem

Run this command, if necessary, to ensure your key is not publicly viewable. `chmod 400 "bimm143_aileenandrade.pem"`

Connect to your instance using its Public DNS: `ec2-54-69-133-115.us-west-2.compute.amazonaws.com`

Example:

```
ssh -i keyfile ubuntu@ec2-54-69-133-115.us-west-2.compute.amazonaws.com
```

```
scp -i keyfile ubuntu@ec2-54-69-133-115.us-west-2.compute.amazonaws.com:~/*_quant .
```

Q1. File paths correct for data input.

```
library(tximport)
library(rhdf5)

# setup the folder and filenames to read
folders <- dir(pattern="SRR21568*")
samples <- sub("_quant", "", folders)
files <- file.path( folders, "abundance.h5" )
names(files) <- samples

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

1 2 3 4

```
head(txi.kallisto$counts)
```

| | SRR2156848 | SRR2156849 | SRR2156850 | SRR2156851 |
|-----------------|------------|------------|------------|------------|
| ENST00000539570 | 0 | 0 | 0.00000 | 0 |
| ENST00000576455 | 0 | 0 | 2.62037 | 0 |
| ENST00000510508 | 0 | 0 | 0.00000 | 0 |
| ENST00000474471 | 0 | 1 | 1.00000 | 0 |
| ENST00000381700 | 0 | 0 | 0.00000 | 0 |
| ENST00000445946 | 0 | 0 | 0.00000 | 0 |

```
colSums(txi.kallisto$counts)
```

| SRR2156848 | SRR2156849 | SRR2156850 | SRR2156851 |
|------------|------------|------------|------------|
| 2563611 | 2600800 | 2372309 | 2111474 |

```
sum(rowSums(txi.kallisto$counts)>0)
```

```
[1] 94561
```

```
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,]
```

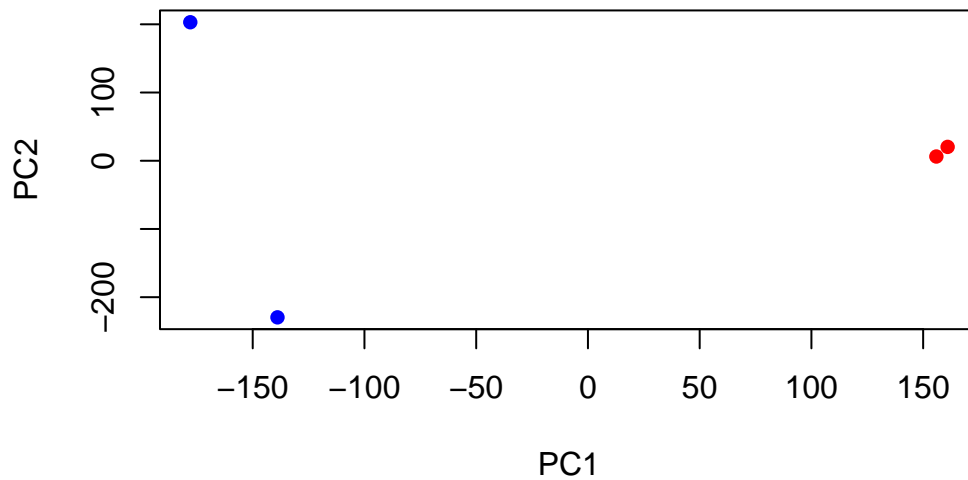
```
pca <- prcomp(t(x), scale=TRUE)
```

```
summary(pca)
```

Importance of components:

| | PC1 | PC2 | PC3 | PC4 |
|------------------------|----------|----------|----------|-------|
| Standard deviation | 183.6379 | 177.3605 | 171.3020 | 1e+00 |
| Proportion of Variance | 0.3568 | 0.3328 | 0.3104 | 1e-05 |
| Cumulative Proportion | 0.3568 | 0.6895 | 1.0000 | 1e+00 |

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



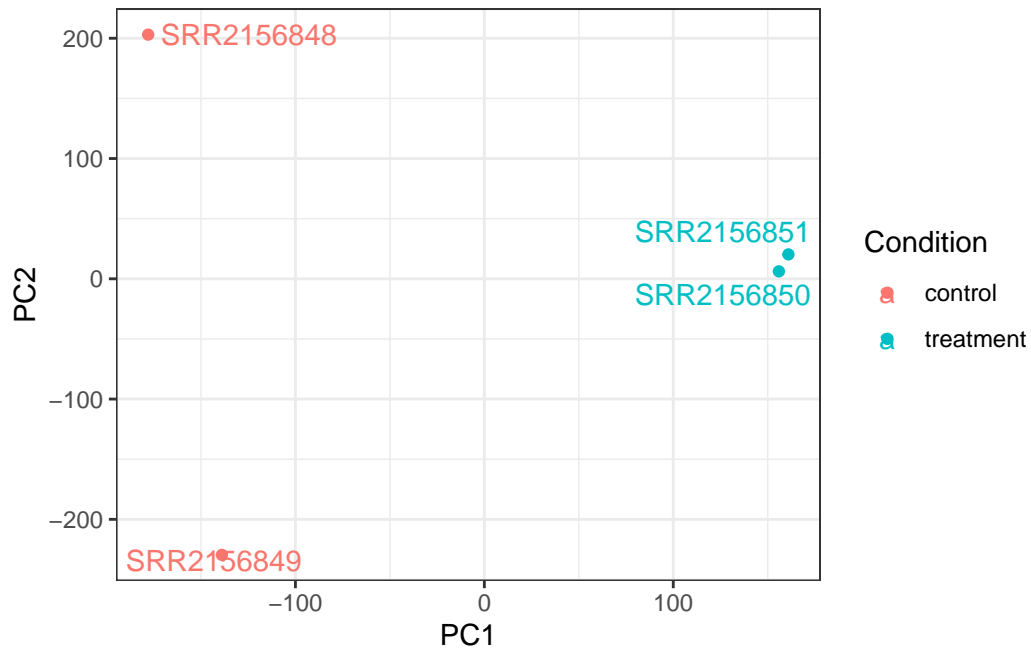
Q2. Use ggplot to make a similar figure of PC1 vs PC2 and a separate figure PC1 vs PC3 and PC2 vs PC3.

```
library(ggplot2)
library(ggrepel)

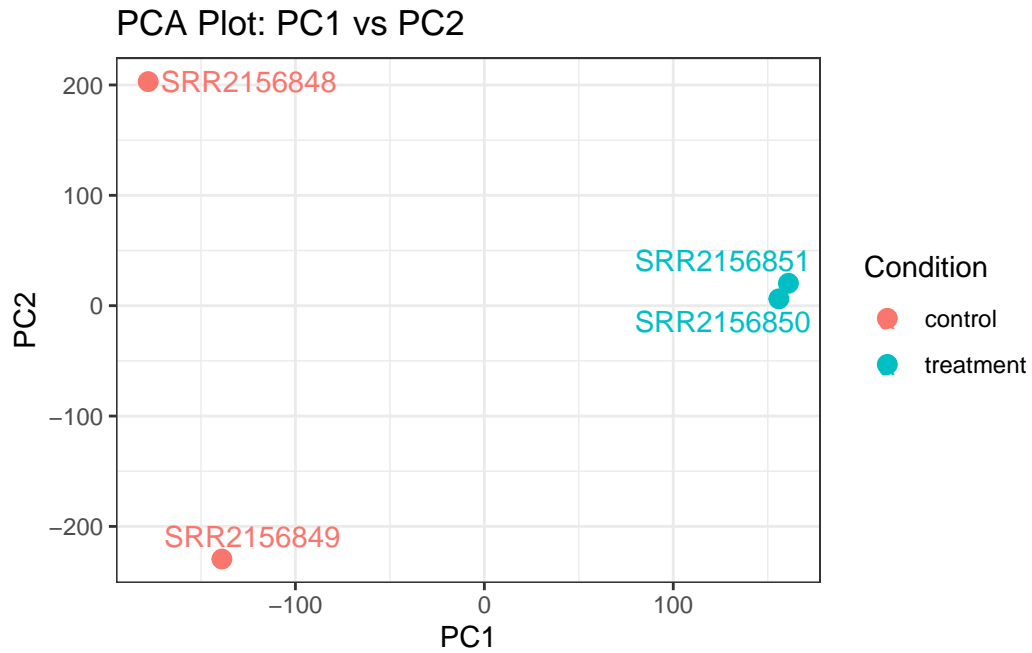
# Make metadata object for the samples
colData <- data.frame(condition = factor(rep(c("control", "treatment"), each = 2)))
rownames(colData) <- colnames(txi.kallisto$counts)

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

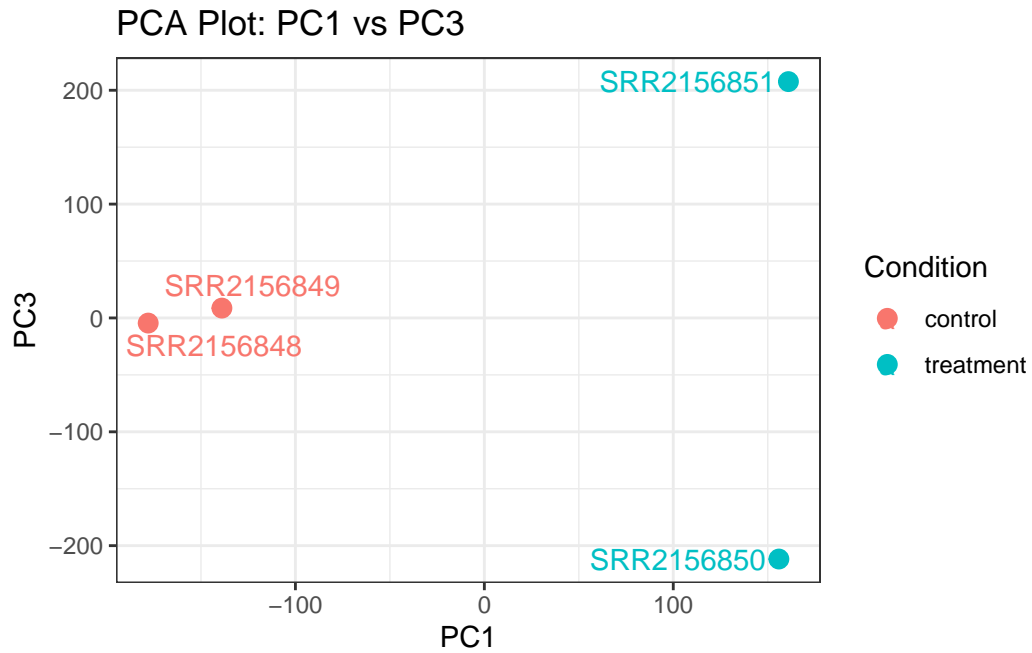
ggplot(y) +
  aes(PC1, PC2, col=Condition) +
  geom_point() +
  geom_text_repel(label=rownames(y)) +
  theme_bw()
```



```
# Plot PC1 vs PC2
p1 <- ggplot(y) +
  aes(x = PC1, y = PC2, col = Condition) +
  geom_point(size = 3) +
  geom_text_repel(label = rownames(y)) +
  labs(title = "PCA Plot: PC1 vs PC2") +
  theme_bw()
print(p1)
```



```
# Plot PC1 vs PC3
p2 <- ggplot(y) +
  aes(x = PC1, y = PC3, col = Condition) +
  geom_point(size = 3) +
  geom_text_repel(label = rownames(y)) +
  labs(title = "PCA Plot: PC1 vs PC3") +
  theme_bw()
print(p2)
```



```
# Plot PC2 vs PC3
p3 <- ggplot(y) +
  aes(x = PC2, y = PC3, col = Condition) +
  geom_point(size = 3) +
  geom_text_repel(label = rownames(y)) +
  labs(title = "PCA Plot: PC2 vs PC3") +
  theme_bw()
print(p3)
```

