

class17

Muhammad Tariq

2025-06-02

```
BiocManager::install("tximport")
```

```
## Bioconductor version 3.20 (BiocManager 1.30.25), R 4.4.3 (2025-02-28 ucrt)
```

```
## Warning: package(s) not installed when version(s) same as or greater than current; use  
## 'force = TRUE' to re-install: 'tximport'
```

```
## Installation paths not writeable, unable to update packages  
## path: C:/Program Files/R/R-4.4.3/library  
## packages:  
## cluster, foreign, lattice, MASS, Matrix, mgcv, nlme
```

```
## Old packages: 'cli', 'rlang'
```

```
library(ggplot2)  
library(jsonlite)  
library(BiocManager)
```

```
## Bioconductor version '3.20' is out-of-date; the current release version '3.21'  
## is available with R version '4.5'; see https://bioconductor.org/install
```

```
library(tximport)
```

Import Kallisto Output from abundance.h5 Files

```
files <- c(  
  "SRR2156848" = "SRR2156848_quant/abundance.h5",  
  "SRR2156849" = "SRR2156849_quant/abundance.h5",  
  "SRR2156850" = "SRR2156850_quant/abundance.h5",  
  "SRR2156851" = "SRR2156851_quant/abundance.h5")
```

Read and Combine Kallisto abundance.tsv Files (Manual Import)

```

# Read each abundance.tsv file
s1 <- read.table("SRR2156848_quant/abundance.tsv", header = TRUE, sep = "\t")
s2 <- read.table("SRR2156849_quant/abundance.tsv", header = TRUE, sep = "\t")
s3 <- read.table("SRR2156850_quant/abundance.tsv", header = TRUE, sep = "\t")
s4 <- read.table("SRR2156851_quant/abundance.tsv", header = TRUE, sep = "\t")

# Combine into one counts matrix
counts <- data.frame(
  row.names = s1$target_id,
  SRR2156848 = s1$est_counts,
  SRR2156849 = s2$est_counts,
  SRR2156850 = s3$est_counts,
  SRR2156851 = s4$est_counts
)

```

Filter Non-Zero and Variable Transcripts

```

counts_filtered <- counts[rowSums(counts) > 0, ]
counts_filtered <- counts_filtered[apply(counts_filtered, 1, sd) > 0, ]

```

Principal Component Analysis

```

pca <- prcomp(t(counts_filtered), scale. = TRUE)
summary(pca)

```

```

## Importance of components:
##              PC1      PC2      PC3      PC4
## Standard deviation 183.6353 177.3561 171.3005 4.877e-11
## Proportion of Variance 0.3568 0.3328 0.3105 0.000e+00
## Cumulative Proportion 0.3568 0.6895 1.0000 1.000e+00

```

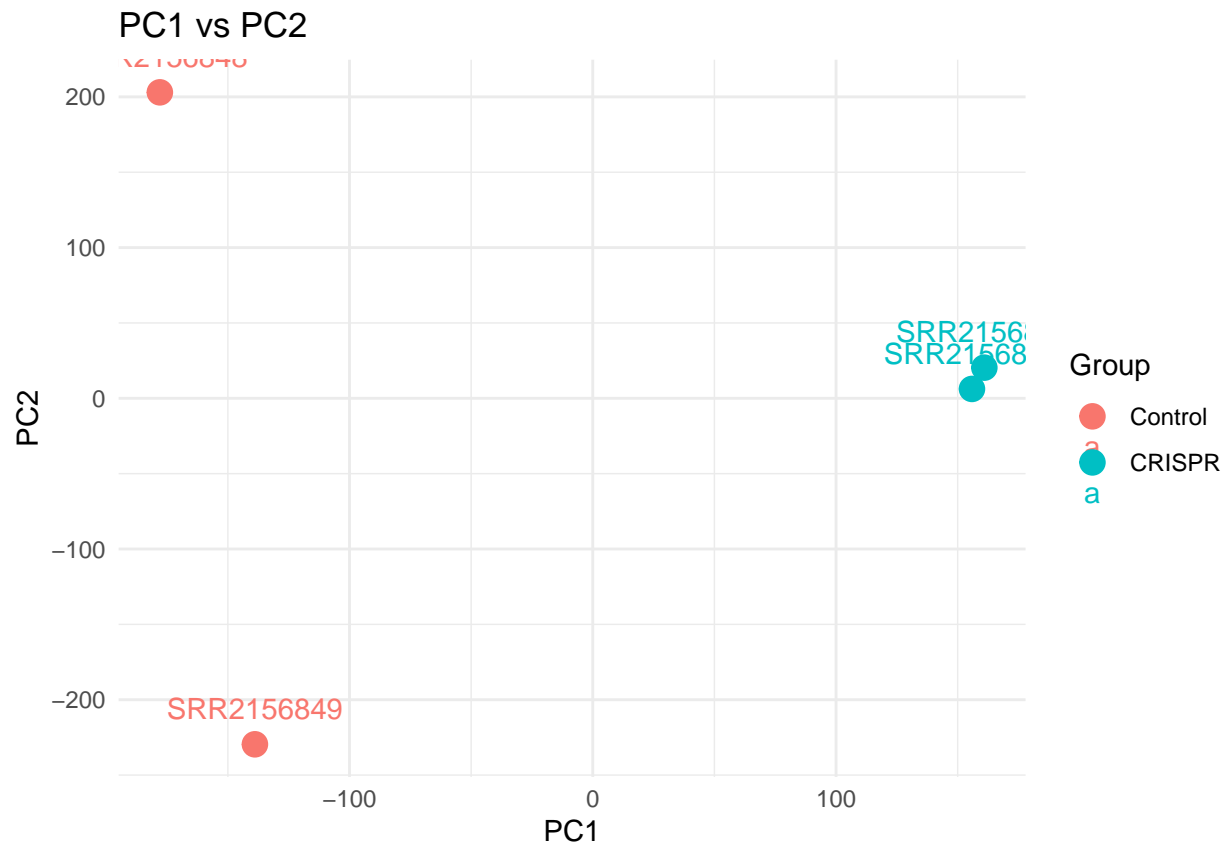
PCA Plotting with ggplot2

```

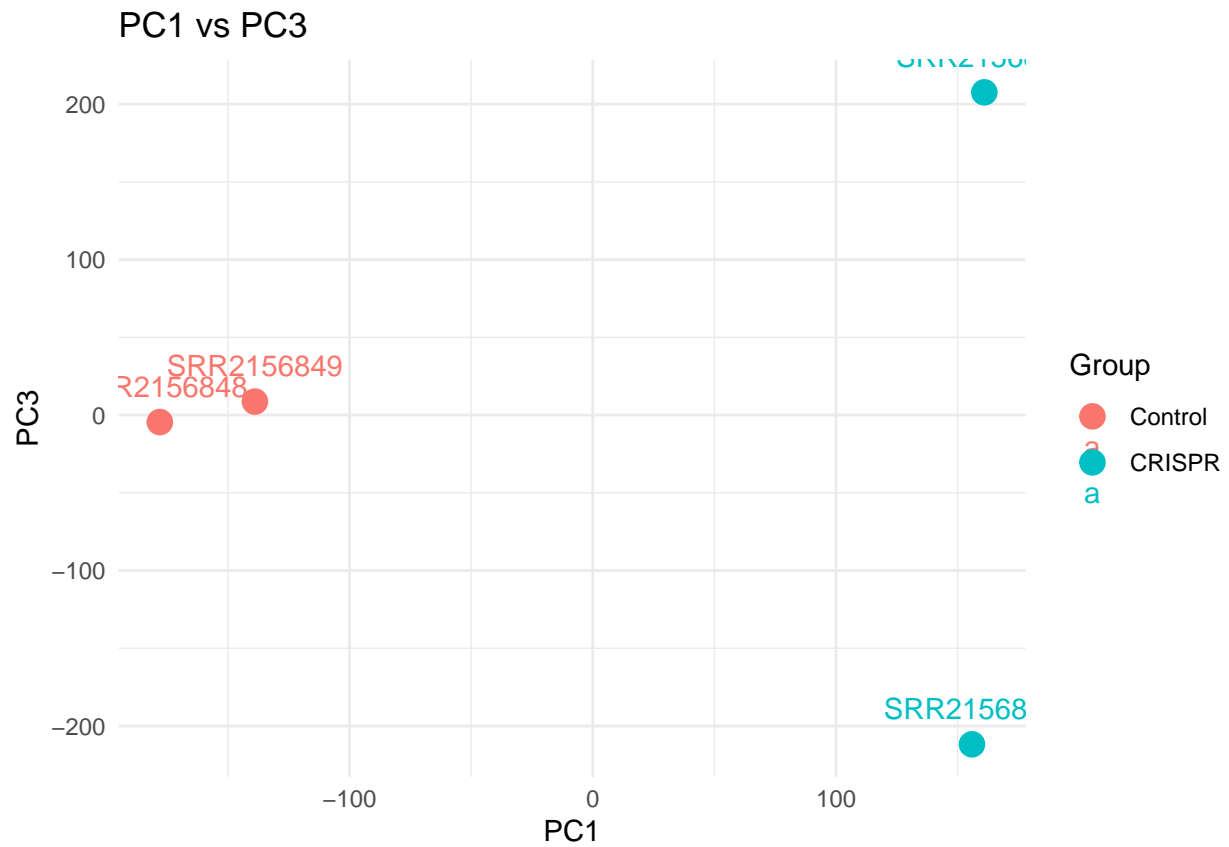
pca_df <- as.data.frame(pca$x)
pca_df$Sample <- rownames(pca_df)
pca_df$Group <- c("Control", "Control", "CRISPR", "CRISPR")

# PC1 vs PC2
ggplot(pca_df, aes(x = PC1, y = PC2, color = Group, label = Sample)) +
  geom_point(size = 4) +
  geom_text(vjust = -1.2) +
  labs(title = "PC1 vs PC2") +
  theme_minimal()

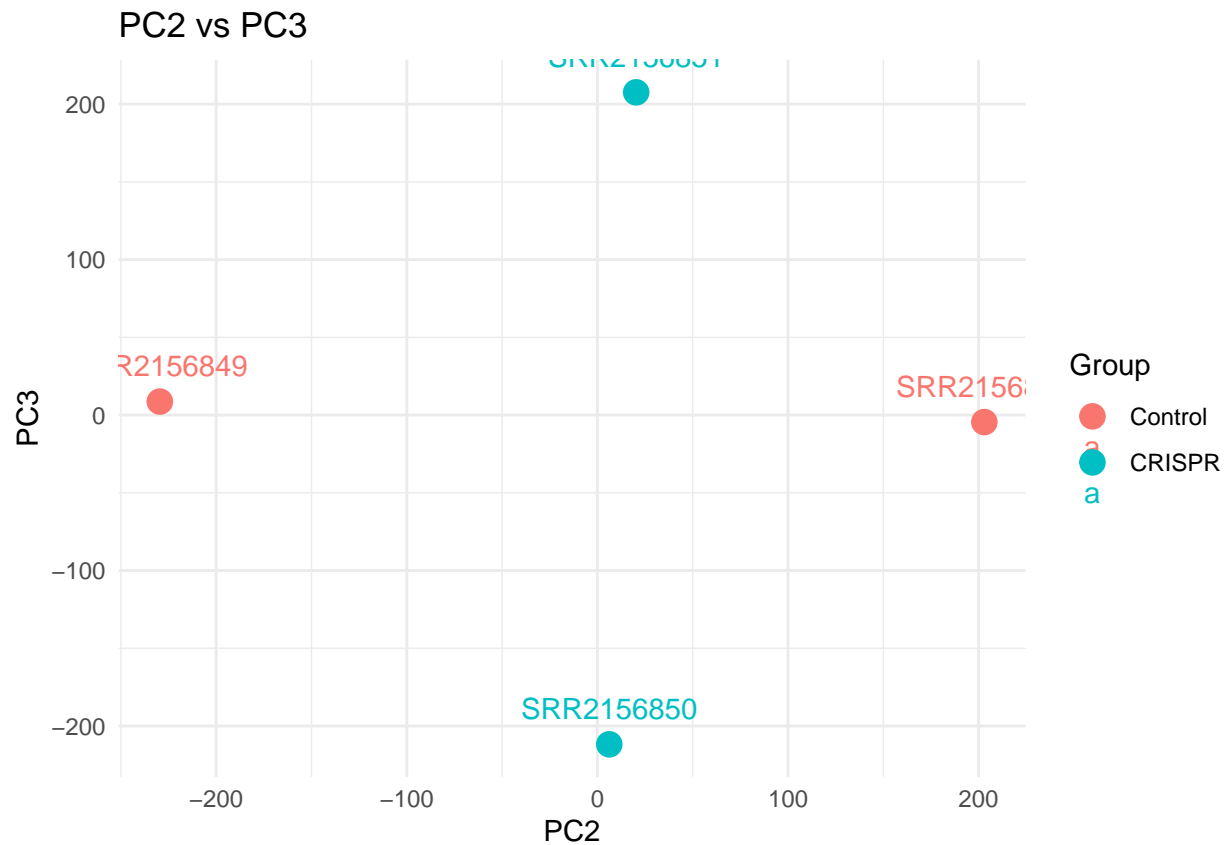
```



```
# PC1 vs PC3
ggplot(pca_df, aes(x = PC1, y = PC3, color = Group, label = Sample)) +
  geom_point(size = 4) +
  geom_text(vjust = -1.2) +
  labs(title = "PC1 vs PC3") +
  theme_minimal()
```



```
# PC2 vs PC3
ggplot(pca_df, aes(x = PC2, y = PC3, color = Group, label = Sample)) +
  geom_point(size = 4) +
  geom_text(vjust = -1.2) +
  labs(title = "PC2 vs PC3") +
  theme_minimal()
```



Optional: View Kallisto Run Statistics

```
library(jsonlite)
info1 <- fromJSON("SRR2156848_quant/run_info.json")
info2 <- fromJSON("SRR2156849_quant/run_info.json")
info3 <- fromJSON("SRR2156850_quant/run_info.json")
info4 <- fromJSON("SRR2156851_quant/run_info.json")

c(
  SRR2156848 = info1$p_pseudoaligned,
  SRR2156849 = info2$p_pseudoaligned,
  SRR2156850 = info3$p_pseudoaligned,
  SRR2156851 = info4$p_pseudoaligned
)
```

```
## SRR2156848 SRR2156849 SRR2156850 SRR2156851
##          86.6          87.1          88.9          89.1
```