

Class17

Celia L Ochoa Medina

```
# BiocManager::install("tximport")
# Load the package
library(rhdf5)
library(tximport)

# setup the folder and filenames to read
folders <- list.files(pattern = "_quant")
files <- file.path(folders, "/abundance.h5")

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

1 2 3 4

```
names(files) <- folders
files
```

```
          SRR2156848_quant          SRR2156849_quant
"SRR2156848_quant//abundance.h5" "SRR2156849_quant//abundance.h5"
          SRR2156850_quant          SRR2156851_quant
"SRR2156850_quant//abundance.h5" "SRR2156851_quant//abundance.h5"
```

```
head(files)
```

```
          SRR2156848_quant          SRR2156849_quant
"SRR2156848_quant//abundance.h5" "SRR2156849_quant//abundance.h5"
          SRR2156850_quant          SRR2156851_quant
"SRR2156850_quant//abundance.h5" "SRR2156851_quant//abundance.h5"
```

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

	[,1]	[,2]	[,3]	[,4]
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)
```

```
[1] 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,]
```

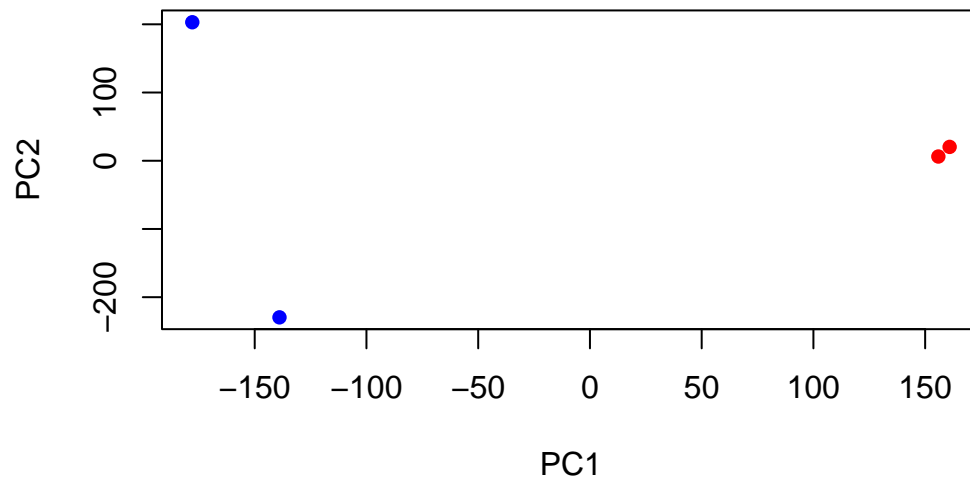
#now make PCA

```
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

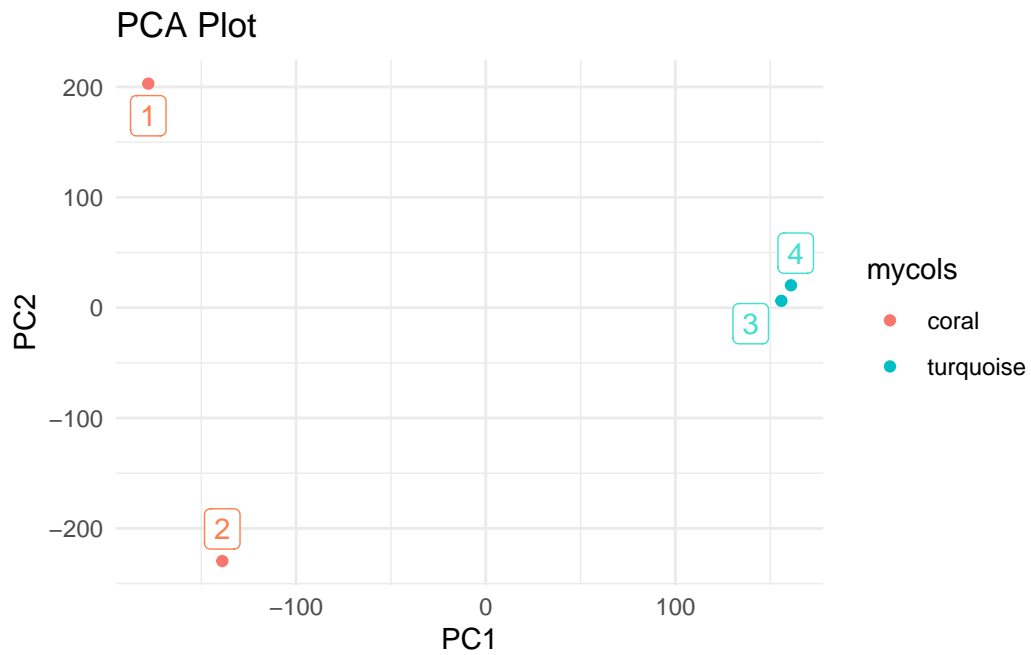
```
plot(pca$x[,1], pca$x[,2],  
     col=c("blue","blue","red","red"),  
     xlab="PC1", ylab="PC2", pch=16)
```



```
library(ggplot2)
library(ggrepel)
```

```
pca_df <- data.frame(pca$x)
mycols <- c("coral", "coral", "turquoise", "turquoise")
```

```
ggplot(pca_df, aes(x = PC1, y = PC2)) +
  geom_point(aes(color = mycols)) +
  xlab("PC1") + ylab("PC2") +
  ggtitle("PCA Plot") +
  geom_label_repel(label = rownames(pca_df), col = mycols) +
  theme_minimal()
```



```
ggplot(pca_df, aes(x = PC3, y = PC4)) +  
  geom_point(aes(color = mycols)) +  
  xlab("PC3") + ylab("PC4") +  
  ggtitle("PCA Plot") +  
  geom_label_repel(label = rownames(pca_df), col = mycols) +  
  theme_minimal()
```

