RESULTS

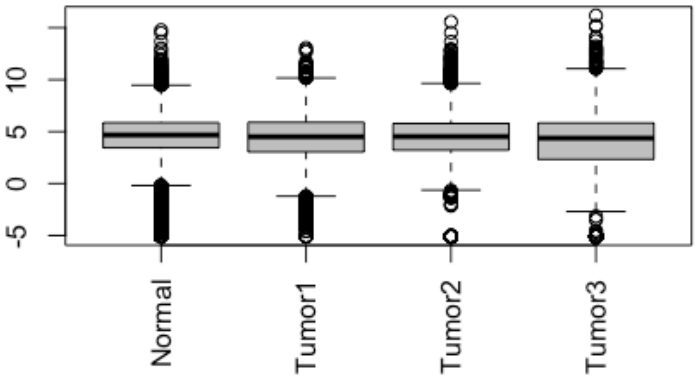

Figure 1: Box plots of pseudoNormCounts for each of the samples.

Script used : boxplot(pseudoNormCounts, col = "gray", las = 3, cex.names = 1)

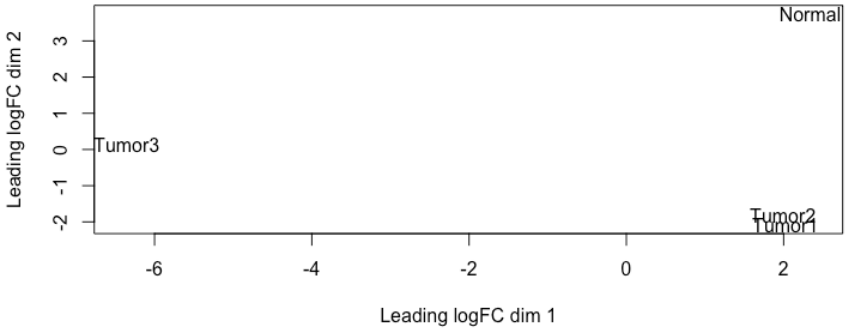

Figure 2: Data Exploration: The multidimensional scaling shows the relatedness among the samples. As expected, Tumor1 and Tumor2 are clustered together away from Normal sample. Tumor3 is far away which is not surprising since, the sequencing quality of Tumor3 was not optimal as I noticed after the alignment.

Script used:

```
plotMDS(y)
design <- model.matrix(~ 0 + group)
design
```

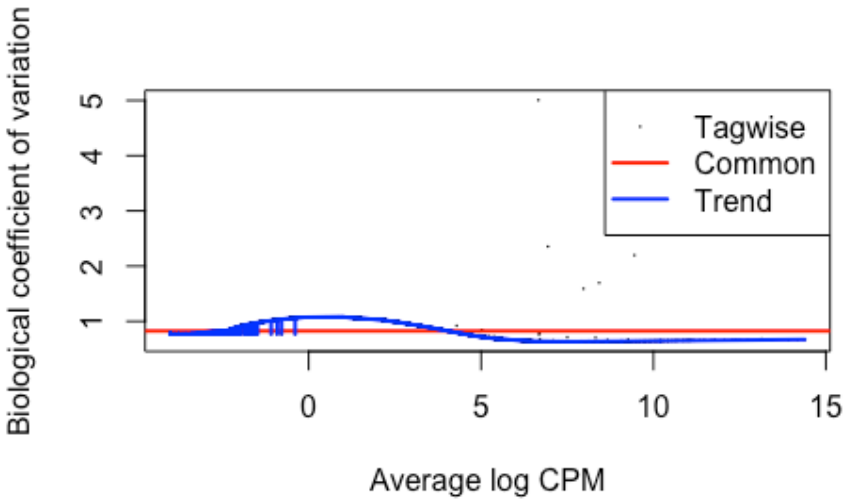

Figure 3: Biological coefficient of variation (BCV) which is the square root of the dispersion parameter in the negative binomial model.

Script used:

```
PlotBCV(y)
```

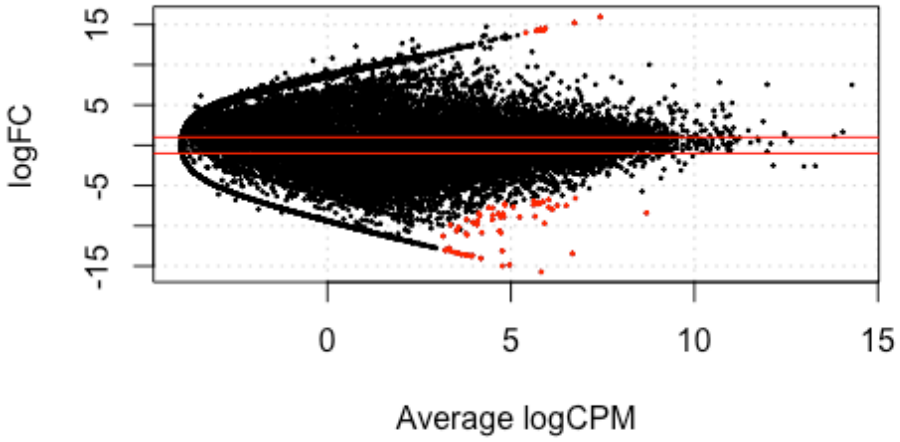

Figure 4: MA plot showing the Log Fold change VS Average Log of Counts Per Million (CPM). The statistically significant differentially expressed genes are shown in red.

Script used:

```
deGenes <- decideTestsDGE(lrt, p=0.001)
deGenes <- rownames(lrt)[as.logical(deGenes)]
plotSmear(lrt, de.tags=deGenes)
abline(h=c(-1, 1), col="2")
length(deGenes)
```

CONCLUSION/FUTURE DIRECTIONS:

Conclusions of my analysis: The RNA seq analysis I performed comparing the Normal VS all the Tumor samples, gave some very good information. First of all, it gave me an initial idea about the quality control of my samples and which genes are differentialyl expressed in all tumors VS the normal adjacent tissue. It also led to the first conclusion that the quality of Tumor3 is less than optimal. Moving forward, I will either try to fix the sample through normalization or transformation, throw out the sample entirely, or repeat the sequencing of tumor3.

Future directions of the project: Regarding the differential expression, I plan to move forward by running all four samples as their own group. For examples Normal VS Tumor1, Normal VS Tumor2 and probably I will exclude Tumor3 complete due to poor quality. This way I will look at things that are significantly different between each tumor sample vs. the control but that aren't significantly different between tumor samples. Also, once RNA seq analysis is complete I will perform gen set enrichment analysis or IPA trying to find which pathways are dysfunctional/downregulated.