Project 1 Presentation Script

DM Background

- For project 1, I did the differential gene expression and alternative splicing analysis for myotonic dystrophy patient RNA samples
- myotonic dystrophy or DM is a multi-systemic autosomal dominant disorder that affects most of the tissues throughout the body
- it is caused by varying lengths of simple repeat expansions or microsatellites that lead to shifts in the genome
- this project focuses on the more sever DM1 that shows a CUG RNA repeat expansion
- these expansions result in muscleblind or MBNL proteins, which are responsible for suppressing exon inclusion in pre-mRNA, and these are sequestered by the now toxic RNA
- therefore muscle defect symptoms occur due to splicing changes as the MBNL concentration decreases

Methods and Materials

- for this project the FASTQ files were provided including their Phred quality scores
- these files were then used to produce SAM and BAM files with the STAR alignment program and a human reference genome
- the DESeq program was used to perform differential gene expression analysis in RStudio to create the plots seen in the following slides using count files for each sample
- the rMATS program in the HPCC terminal was used to perform alternative splicing analysis. However, the plots were generated using the maser program in RStudio

PCA Plot

- differential gene expression analysis involves searching for significant changes in the expression of genes within the experimental data
- this PCA plot shows the variance between the DM1 and Control samples with the controls in orange and the DM1 samples in green
- there is some similarity in the way that both groups cluster in that two samples cluster at the top and three at the bottom
- however there positions on the x axis show that the samples do cluster by condition

MA Plot

- the MA plot shows the overall trend of differential gene expression in regards to the relationship between the Log2Fold change and the mean normalized counts of the experimental data
- the Log2Fold change describes the change in expression of a specific gene in the control samples compared to the DM1 samples
- while the normalized counts is the result of the read counts of a specific gene divided by the total number of reads in that sample
- the position of majority of the blue points being in the positive log2fold region shows that majority of the genes are upregulated

Volcano Plot

- the volcano plot further supports how the genes are mostly upregulated since majority of the red points each representing a gene have a positive log2fold change
- the cutoffs for this plot were set to a log2base change greater than 1.5 and an adjusted p-value less than 0.05

Heatmap

- the heatmap displays the expansion level values of the top 20 most significantly expressed genes according to a cold to warm color tone scale
- with downregulation or low expression being represented as cold and upregulation or high expression as warm
- as seen here these genes behave similarly among the DM1 samples and display significant upregulation and high expression as compared to the control samples

Normalized Counts Plot

- MYH3 was the most statistically significant gene within the experimental data
- its adjusted p-value in the control samples are significantly lower than that of the DM1 samples

Filtered PCA Plot

- alternative splicing analysis involves the use of isoform quantification and pairwise comparison to analyze RNA splicing events
- this PCA plot was filtered to show the statistically significant genes with a false discovery rate or FDR less than 0.05 and a delta PSI greater than 0.1
- the plot displays more clustering among the control samples, supporting the possibility that there is a difference in which the genes were spliced

Violin Plot

- the violin plot specifically displays the five MBNL1 splicing events that fit the same cutoffs mentioned previously
- event 59859 shows the most variation between the samples with a significant difference in PSI values

Sashimi Plot

- this event was also represented in a modified version of a sashimi plot to get a closer look at the differential levels of exon inclusion in the DM1 samples
- it again displays a higher PSI value than the control samples, further proving that this spliced exon is promoted by DM1