Differential Gene Expression and Alternative Splicing Analysis of Myotonic Dystrophy Patient RNA Samples

PROJECT 1

CHERRON GRIFFITH

JULY 9, 2024

Myotonic Dystrophy (DM) Background

- a multi-systemic autosomal dominant disorder that affects most of the tissues throughout the body

- DM1 is a more severe form of DM, compared to DM2

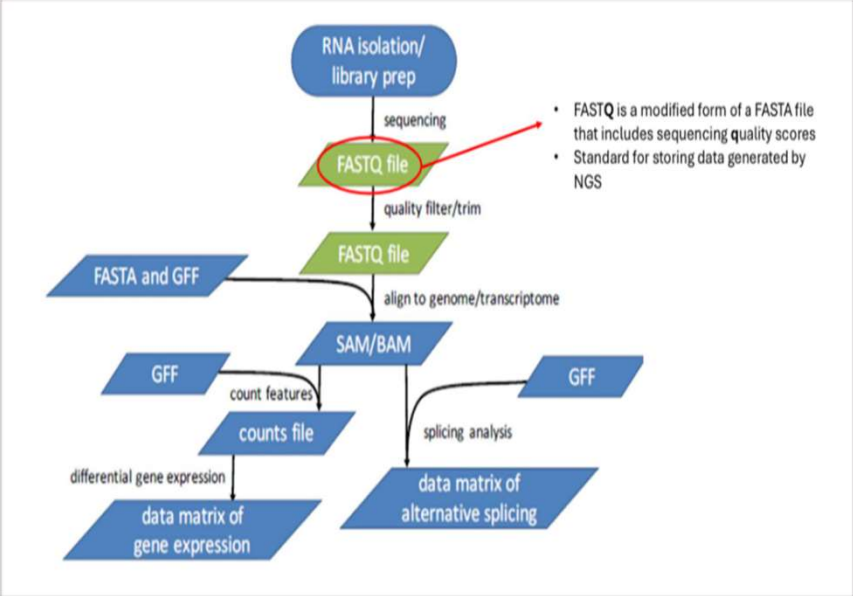
- caused by varying lengths of simple repeat expansions, or microsatellites, that lead to shifts within the patient's genome
  - DM1 show CUG RNA repeat expansion
  - DM2 show CCUG RNA repeat expansion

- Expansions result in muscleblind (MBNL) proteins being sequestered by the toxic RNA
  - MBNL proteins responsible for activation and suppressing exon inclusion in pre-mRNA decrease in number as toxic RNA continues to build up
  - Decreased MBNL protein concentration leads to splicing changes, leading to muscle defect symptoms

Methods and Materials

1) FASTQ was provided from containing sample RNA sequences and Phred quality scores
2) The program STAR was use as an alignment tool to map the FASTQ reads to the human reference genome, resulting in SAM and BAM files
3) DESeq program was used for differential gene expression analysis in RStudio
   a. PCA plot, MA plot, Volcano plot, Heatmap, and Normalized Counts plot were each generated using the same program within RStudio using counts files for each sample
4) rMATS program was used for alternative splicing analysis in the HPCC terminal
   a. PCA plots (filtered and unfiltered), Sashimi plot, ad Violin plot were each generated using the BAM files for each sample

RNA-Seq Experimental Overview



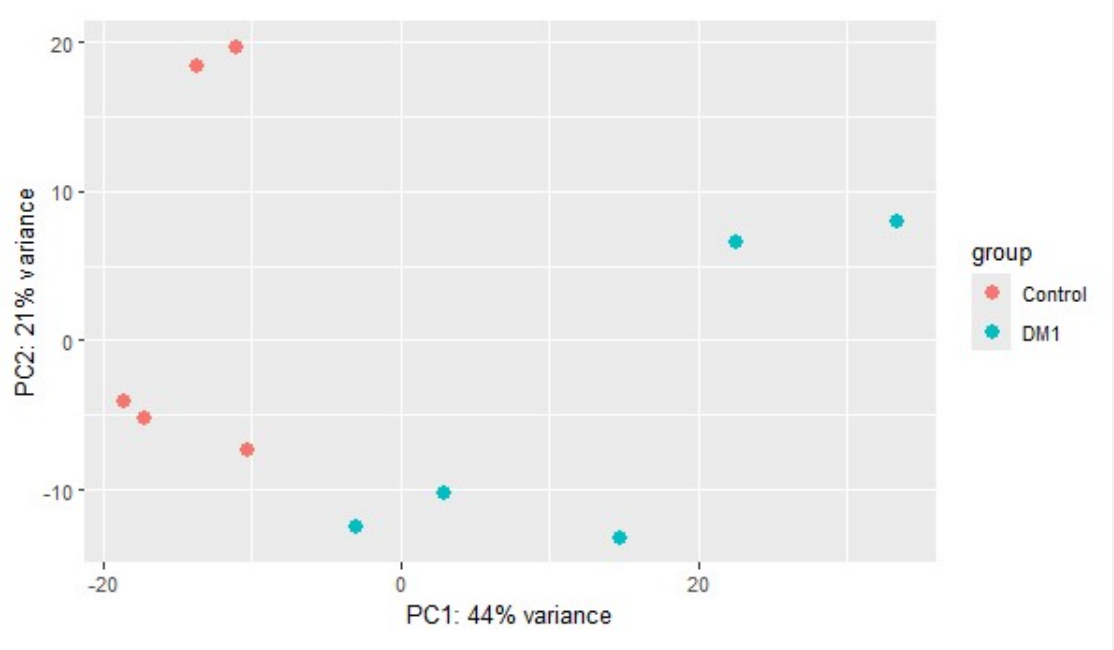
https://2024rnainstit-xd74059.slack.com/files/U0736MK16KD/F079UG88BM3/lecture-3_fastq-dmseq-bioinformatics-_-hpcc.pptx

Differential Gene Expression (DGE) Analysis

SEARCHING FOR DIFFERENTIALLY EXPRESSED GENES (DEGS) WITHIN THE EXPERIMENTAL DATA

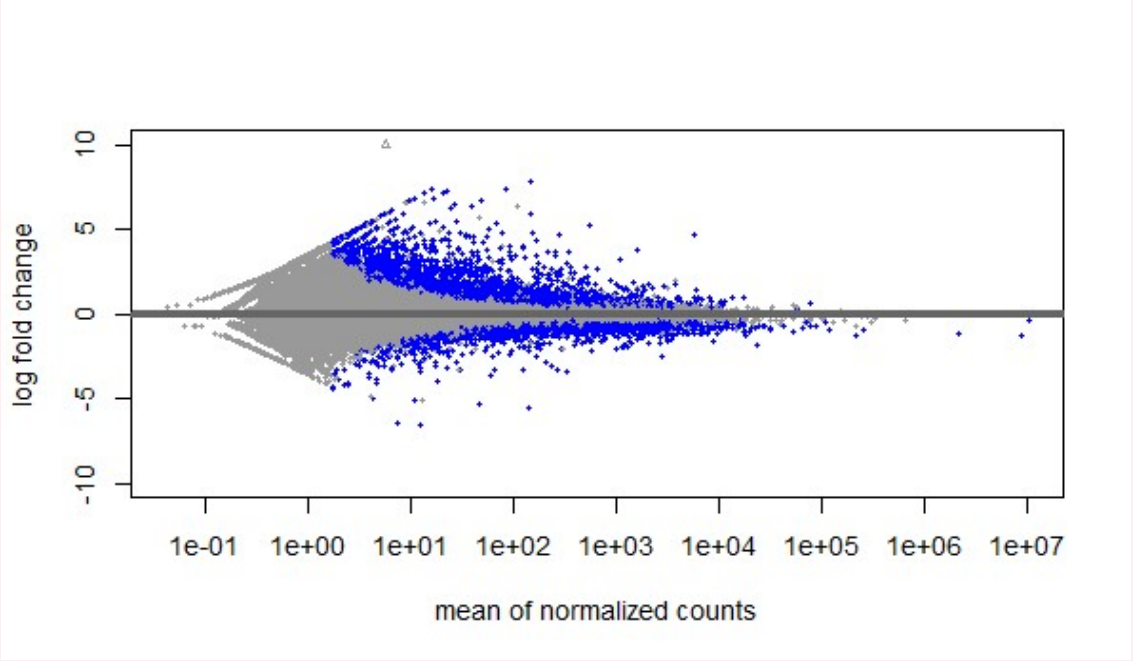
Principal Component Analysis (PCA) Plot

- Each sample is plotted separately to show variance between each sample
  - Control samples are in orange
  - DM1 patient samples are in green
- There is some variance within the control sample group and also within the DM1 sample group
  - Similarity in the way both groups cluster with one cluster in bottom left and another cluster at top right
  - The samples do cluster based on condition



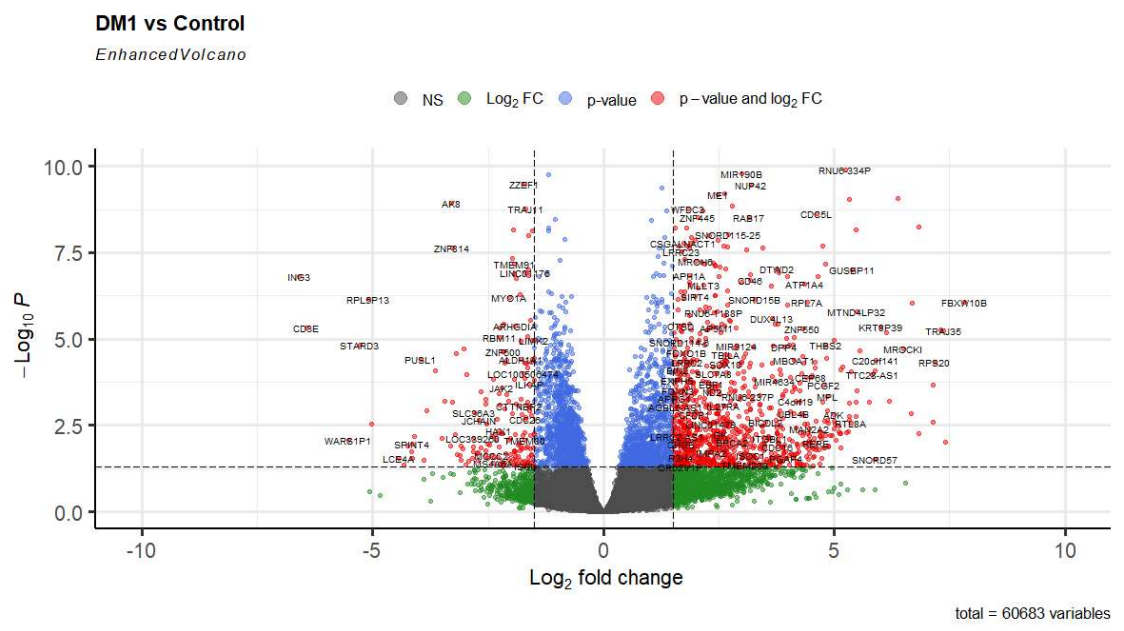
MA Plot

- Displays the overall trend of differential gene expression using the relationship between the log-fold change and the mean normalized counts of the data
- shows that majority of the genes are significantly upregulated as indicated by the position of the blue points in the positive Log2Fold range



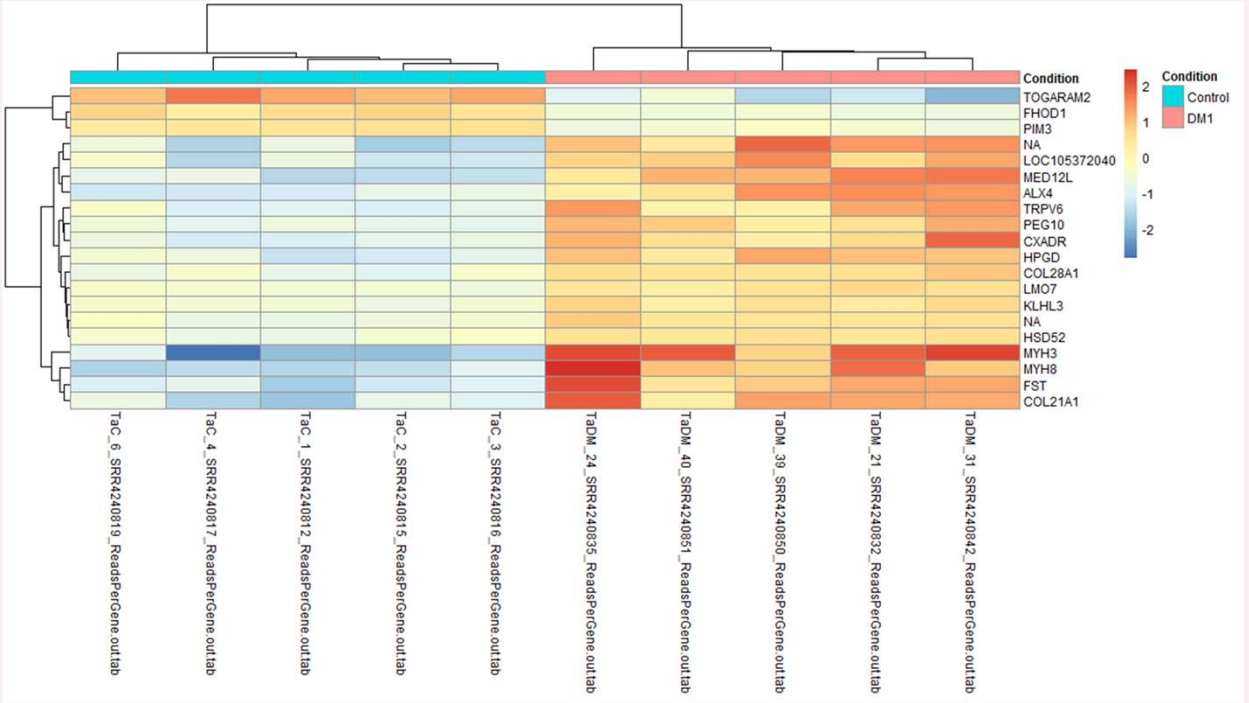
Volcano Plot

- Displays the overall trend of differential gene expression using the relationship between the Log2FoldChange and the adjusted p-value
- Majority of red points each representing a gene display a positive Log2Fold
  - shows that majority of the genes are upregulated



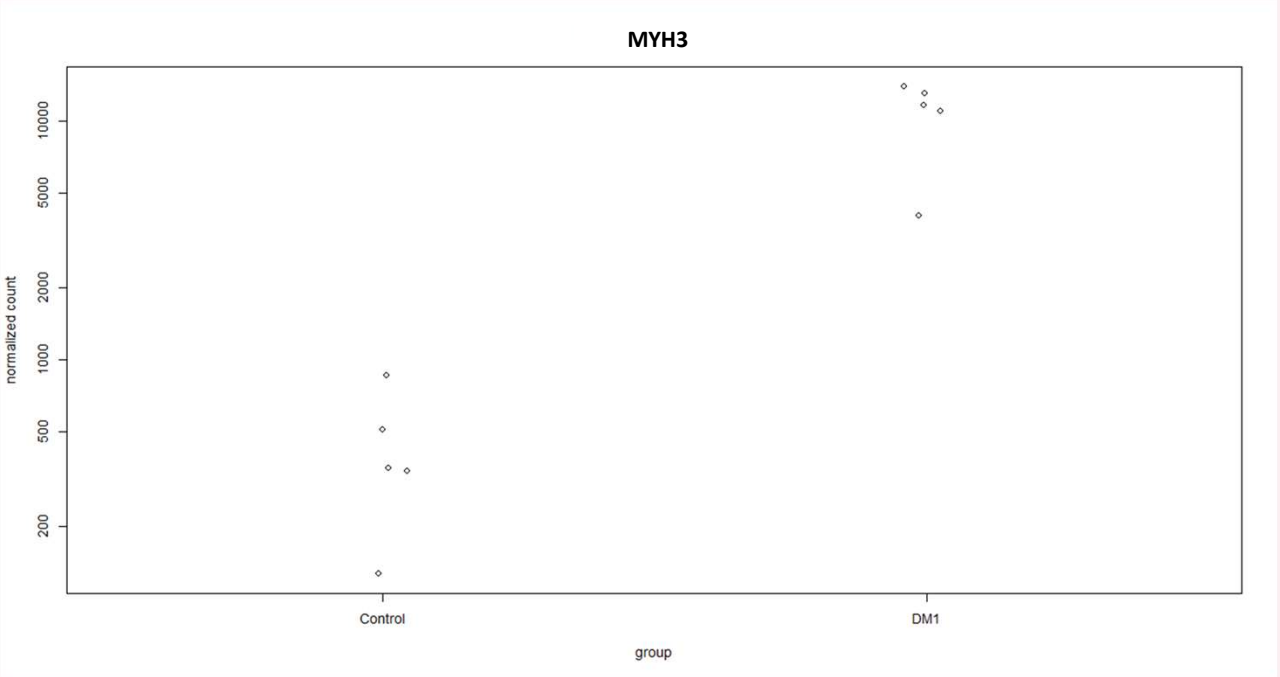
Heatmap

- Displays the 20 most differentially expressed genes and shows their count abundance in comparison to the other libraries
  - Expression level values are shown in a cold to warm color tone with down (low) expression as cold and up (high) expression as warm
- The top differentially expressed genes behave similarly in the DM1 samples while also being significantly distinct from the control samples
  - These genes are primarily up regulated in the DM1 samples, as shown by their warmer tone (negative on the scale)



Normalized Counts Plot

- Normalized counts for the most statistically significant gene were plotted based on adjusted p-value
- This gene for each sample is plotted individually to show the changes between control and DM1 samples
- This gene is significantly different between the control and DM1 samples
  - Control sample adjusted p-values are lower than the adjusted p-values of the DM1 samples

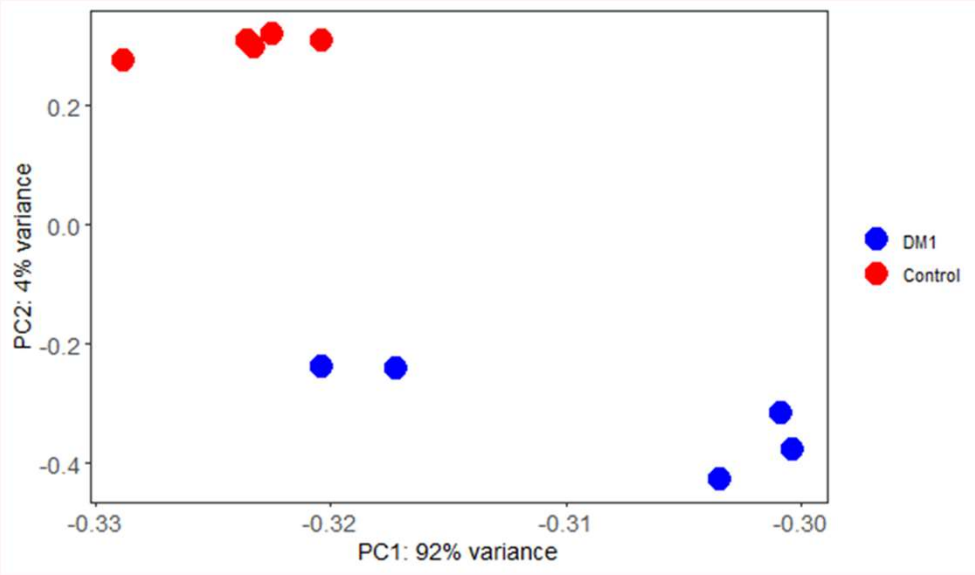


Alternative Splicing Analysis

ANALYSIS OF SPLICING EVENTS USING ISOFORM QUANTIFICATION AND PAIRWISE COMPARISON

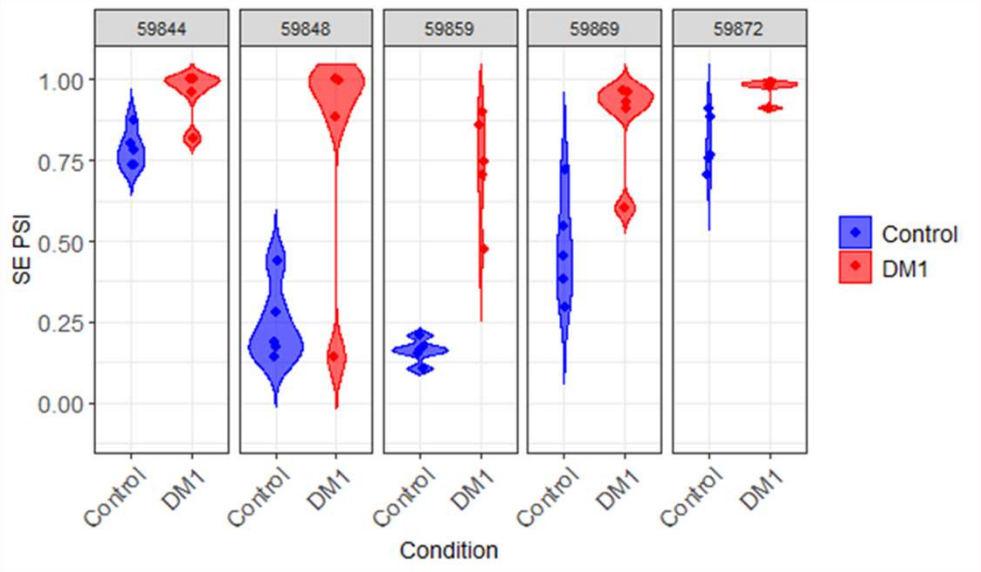
Filtered PCA Plot

- Plot is filtered to show the genes with a false discovery rate (FDR) less than or equal to 0.05 and a delta PSI greater than or equal to 0.1
- Shows more clustering among the control samples as compared to the DM1 samples
  - Supports the possibility that there is a difference in which the genes displayed are spliced



Violin Plot

- Plots the genes with MBNL1 splicing events with a delta PSI more than 0.1 and FDR less than 0.05
- Event 59859 shows the most variation between the DM1 samples and the control samples
  - Have a significant difference in their PSI values



Modified Sashimi Plot

- Visualizes the differential levels of exon inclusion in DM1 samples vs control samples
- this MBNL1 splicing event (event 59859 from the violin plot) has a higher PSI in the DM1 samples compared to that of the control samples
  - Supports the possibility that this spliced exon is promoted by DM1