Investigating the impact of TDP-43 loss in FTD-ALS using bulk RNA-seq analysis

Siavash Raissi

Introduction

Amyotrophic lateral sclerosis (ALS) is characterized as a fatal neurodegenerative disease caused by the loss of motor neurons, often resulting in muscle weakness and overall brain degeneration (Suk, 2020). Adding to these symptoms, approximately 10-15% of ALS patients are also diagnosed with frontotemporal dementia (FTD), which is affiliated with frontal and anterior temporal lobe degeneration (Masrori, 2020). FTD patients often demonstrate reduced ability in executive function, as well as language impairment, suggesting a similar neuronal origin. Both conditions are thought to be tightly correlated due to TDP-43, an RNA-binding protein responsible for RNA transcription, splicing, and transport (Ma, 2022). In FTD-ALS, it is hypothesized that the abnormal neurological phenotype may originate from TDP-43 aggregating in the cytoplasm of neurons, causing the resulting degradation in cognitive and motor degradation (Ferrari, 2013). Currently, it is estimated that mutations in TDP-43 are observed in about 97% of all ALS patients (Suk, 2020). Thus, it is crucial for researchers to further investigate TDP-43's role as a protein responsible for maintaining normal neurological function.

TDP-43's vast range of mutations and their close correlation with FTD-ALS symptoms thus lends credibility to the background of Ma, et al.'s original basis for the start of their investigation. In their paper, the authors Ma, et al. explore the role of TDP-43 inhibition on promoting cryptic exon inclusion in UNC13A and investigate its downstream impacts on FTD-ALS prognosis (Ma, 2022). To accomplish this, the researchers utilize RNA-seq techniques to document the transcriptomes present in TDP-43 positive and negative nuclei. After performing computational analyses to identify alternative splicing events, they are able to highlight UNC13A as a promising target of investigation. They further reinforce the integrity of this gene as an important factor in FTD-ALS development by referring to similar results across genome-wide association studies and a frontal cortex brain bank, where they identify genetic variants that correlate with the cryptic exon inclusion in UNC13A, inactivating the protein and statistically linking it to various forms of FTD-ALS. Thus, the study makes use of various computational methods and available databases to form a basis for the rest of the investigation which involved wet-lab experiments to support the hypotheses presented. For instance, after identifying UNC13A as a target of alternative splicing events, the researchers perform an immunoprecipitation assay for TDP-43 to demonstrate that it is binding directly to the relevant introns in the genes. Similarly, in situ hybridization assays were utilized to detect when and where the UNC13A CE was present in human brain cells.

The ultimate findings of this study thus highlight the importance of UNC13A and its linked variants as a possible biomarker for diagnosing FTD-ALS and identifying those at risk in the future. Additionally, the authors discuss the potential for therapeutic targets to utilize this finding

as a means to restore some functionality to those who suffer from the downstream effects of TDP-43 loss.

However, despite the paper's thorough investigation of UNC13A, there are opportunities to further contribute to their initial analyses and expand new methodologies to their data. On page 125, the authors indicate that "there were 55 alternatively spliced genes in common between the two analyses" (Ma, 2022). However, this only analyzed aberrant differential splicing, and not the differential expression of any of these genes, nor their resulting enrichment. As a result, this course project sought to apply and expand the author's methods as a means to investigate genes that are differentially expressed across TDP-43 positive and negative samples and interpret their annotations. Are genes involved in the nervous system being differentially expressed? What function do these genes serve in the nervous system and how may their misexpression contribute to the development of FTD-ALS?

To accomplish this task, many of the methods mentioned by the authors were implemented, as well the resources available in the paper. In the Methods section, the authors mention that prior to performing aberrant splicing analysis, they first performed their own RNA-seq differential expression analysis and used the mapping profile they generated as a means to determine the future exon splicing events. This course project replicates the first half of their workflow and expands upon the differential expression analysis. Bulk RNA-seq count data was found on the GEO database (GSE126542) for gene expression in 7 samples for cells that are TDP-43 positive and negative. Further differential expression analysis was performed using the DESeq2 package utilizing the methods presented in the DESeq vignette, while referencing the code published in the Ma, et al.'s GitHub (Love, 2014). Once the differential expression analysis for each gene and the respective adjusted p-values were generated, genes with significant log fold changes between the two TDP-43 genotypes were analyzed using the Gene Ontology Resource (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). Several visualizations were produced, including the dispersion of the p-values, a PCA classification of the different samples, MAplots with different log fold shrinkages applied to the data, and several barplots with counts for the identified GO and KEGG annotations. As a result, the original paper and its in-depth methods section provided heavy guidance for the expansion of its differential expression analysis.

Methods

The primary bulk RNA-seq count dataset for this originated from a previous flow cytometry (FACS) experiment separating cells with and without TDP-43 from 7 patients with FTD-ALS (Liu et al., 2019). This dataset, which is available under GSE126542 in the Gene Expression Omnibus, consists of 16 FASTQ files for each patient (Liu et al., 2019). These patients consist of equal parts males and females. Serving as the primary variable of investigation, 7 of the samples are from TDP-43 positive cells, 7 are from TDP-43 negative cells from the same patients, and 2

are unclassified. Additionally, though not the primary variable being analyzed in the differential expression analysis, it is important to also note that 6 cases have been diagnosed as FTD-ALS, while 10 are classified as FTD. This dataset measures 57,865 genes across all 16 patients.

To extract genes with significant changes in expression across the TDP-43 positive and negative cell groups, RNA-seq differential analysis was performed using DESeq2 with code to analyze this count data available on Ma, et al.'s GitHub (Ma, 2022). However, to enhance the value of the project, analysis was performed independently guided with the official DESeq vignette, which is publicly available online, while referencing the authors methods if any additional guidance was needed (Love, 2014).

This workflow involved the following steps. First, RNA count data and its respective annotations were loaded into R. Due to the accessible count data, initial steps involving FASTQ read mapping with STAR were skipped. Next, the datasets were provided as input to DESeq to form a DESeqDataSet, with gender and genotype (TDP positive/negative status) included as variables in the design formula. Genes were pre-filtered, excluding genes with a count less than 7, the number of samples in either the TDP-43 positive and negative groups. TDP-43 positive genotypes were established as the reference level and the DESeq algorithm was run. Log fold change shrinkage, a technique utilized to inform a prior distribution by shrinking log fold changes in genes with low counts, was performed to remove noise from the resulting data using the "apeglm," "normal," and "ashr" methods (Anqi, 2019; Stephens, 2016). Independent filtering, a method utilized to increase detection power by removing genes with low counts, was also performed and DESeq results were calculated separately using an alpha of .05, thus optimizing the number of genes with a p-value less than .05. This decreased the number of significant genes (padj < .05) present from 2091 to 1892, demonstrating the stricter filtering criteria established on the dataset and the final results extracted.

A variety of transformations and visualizations were then performed using the filtered results. By first applying an *rlog* transformation to the data to reduce the dependence of the variance on the mean, the regionReport package from Bioconductor allowed for the easy creation of an html page with an adjusted p-value histogram, a PCA plot clustering samples by gender and genotype, a heatmap of the samples and expression matrix, and count plots for each of the most significantly expressed genes. These visualizations validated the overall process utilized to perform the analysis.

Afterwards, the results across samples with the greatest log fold changes with padj < .05 were selected to identify the most significantly differentially expressed genes. Significant genes were split into groups with positive log fold changes greater than 1, and negative log fold changes less than -1, to determine upregulated and downregulated genes respectively. To interpret the hundreds of genes present in each group, GO and KEGG were selected to produce gene

enrichment analysis. The resulting annotations generated from lists of all significant genes, significantly upregulated genes, and significantly downregulated genes were visualized in barplots based on the number of counts of genes in each group, where GO results represent biological processes, and KEGG results represent pathways.

Results

DESeq analysis was performed on the TDP-43 positive versus negative dataset provided by the authors as described on Bioconductor's official vignette. Following independent filtering, a list of the most significant genes across the two TDP-43 genotypes, with the lowest adjusted p-values, had been created (Figure 1). Genes like C1orf194 appear to have the most significant expression, with a positive log fold increase of 3.4 when TDP-43 is removed (Figure 1).

	baseMean	**log2FoldChange**	**lfcSE**	**stat**	**pvalue**	**padj**
C1orf194	1051.490549	3.477998	0.249384	13.94638	3.31E-44	8.06E-40
AC019048.1	2979.481592	2.879894	0.220319	13.07149	4.79E-39	5.84E-35
CFP	451.553861	3.525593	0.304841	11.56537	6.17E-31	5.01E-27
MVP	336.5109228	3.143964	0.276907	11.35385	7.10E-30	4.32E-26
LINC01456	2428.970791	3.601207	0.331642	10.8587	1.81E-27	8.83E-24
RP1-272E8.1	48.3487653	3.622798	0.347613	10.42193	1.97E-25	8.00E-22
RP11-97O12.6	179.546011	3.370901	0.325184	10.36613	3.54E-25	1.23E-21
FKSG62	565.2562505	2.604117	0.25531	10.19981	1.99E-24	6.05E-21
CTD-2135J3.3	513.4083017	-2.31802	0.243013	-9.53869	1.45E-21	3.92E-18
MEPE	392.3231065	-4.34223	0.457487	-9.49148	2.28E-21	5.55E-18

Figure 1: A table of the 10 genes of 1892 produced by DESeq with the lowest p-adj values after independent filtering with an $\alpha = .05$.

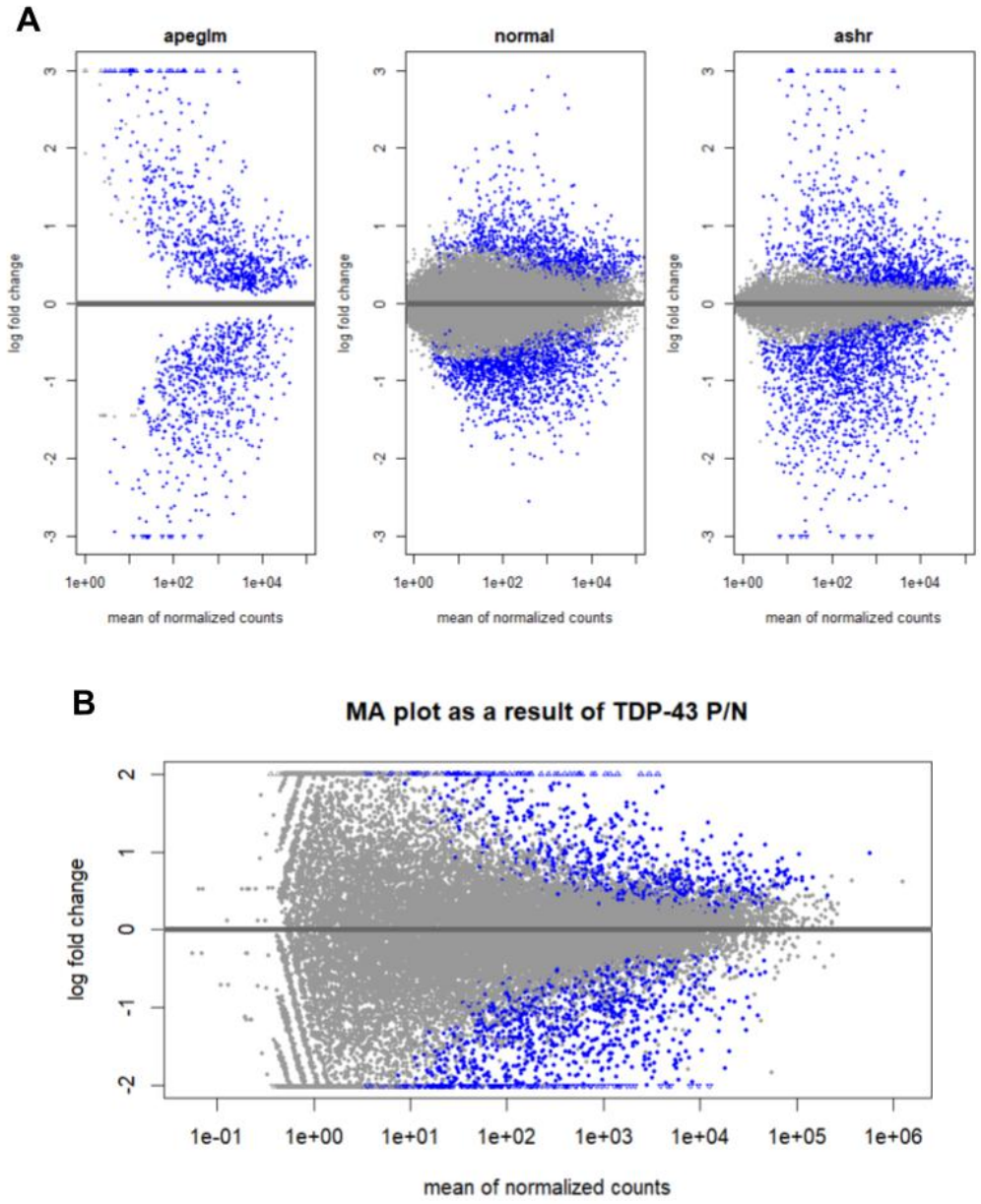


Figure 2: MA plots generated from the DESeq results including different methods for LFC shrinkage across TDP-43 positive and negative samples. **A.** MA plots with different LFC shrinkages applied to DESeq results. From left to right, apeglm shrinkage method, normal shrinkage, and ashr shrinkage. Points in blue represent genes with padj < 05. **B.** DESeq results MA plot without LFC shrinkage. As the mean of normalized counts increases, the intensity of log fold change appears to decrease.

Four MA plots were generated featuring log fold shrinkage (LFC) (Figure 2). The MA plot with apeglm LFC appears to lose almost all of its noise, while normal and ashr shrinkage methods retain some noise, with ashr maintaining greater variance in LFC magnitude across genes (Figure 2A). Across all MA plots, including that with no shrinkage applied, it appears that significant

genes with the highest counts, and therefore expression, tend to demonstrate the lowest magnitude of LFC between the TDP-43 positive and negative samples (Figure 2B).

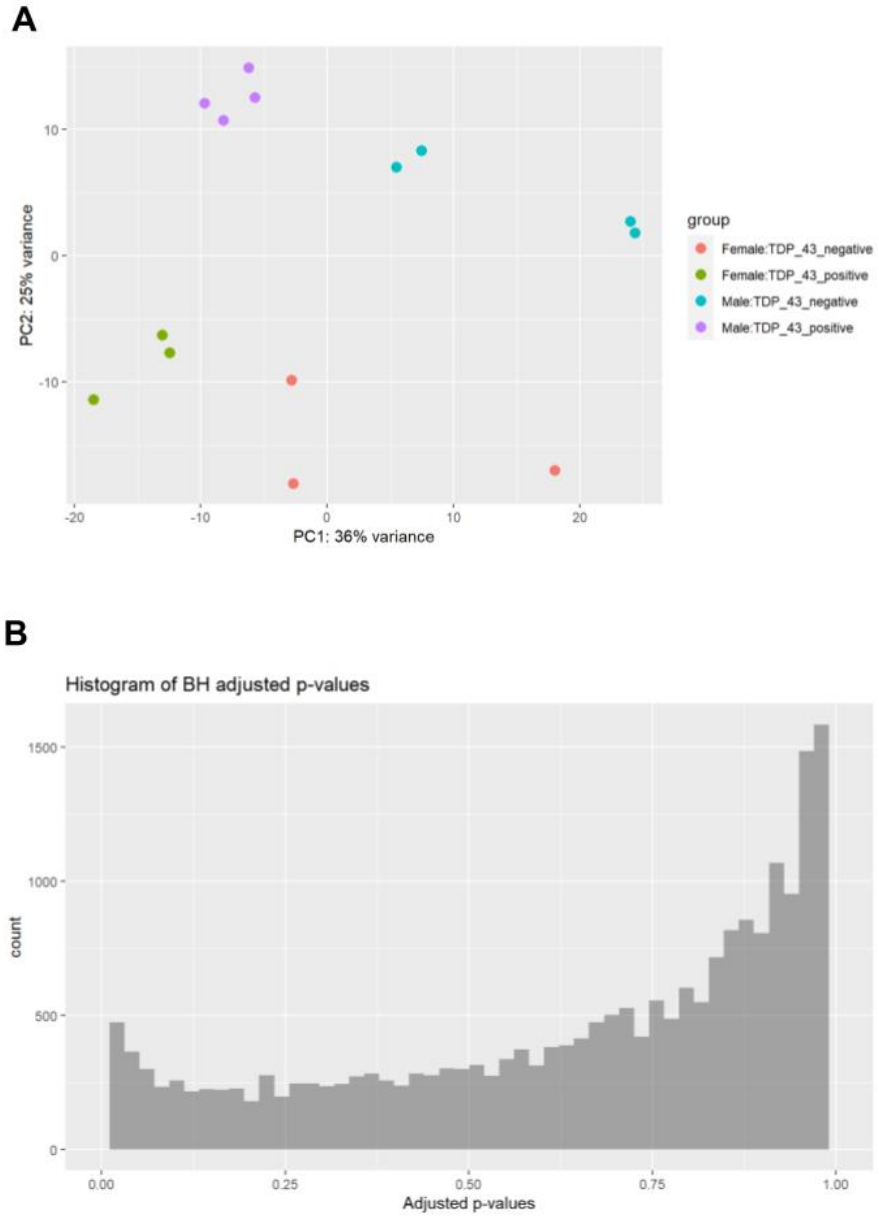


Figure 3: Visualizations of the DESeq results variance and the extent to which it is accounted for. **A.** A PCA plot reducing the dimensionality of the DESeq results based on genotype/sex. Space between subclusters in the TDP-43 negative groups are not accounted for. **B.** A histogram of the adjusted p-values generated for each gene from DESeq. A noticeable right-skew is present.

The adjusted p-values of all the genes also tended to skew to the right, with there being significantly more genes with higher padj-values near 1 than those less than .05 (Figure 3B). This reduces the overall confidence generated, as it is clear that the FDR method utilized, the Benjamini-Hochberg procedure, did not result in uniformly distributed adjusted p-values as

intended. This may be due to unaccounted variance, as identified in the PCA plot of the DESeq results, with a clear unlabeled space present between the TDP-43 negative samples (Figure 3A).

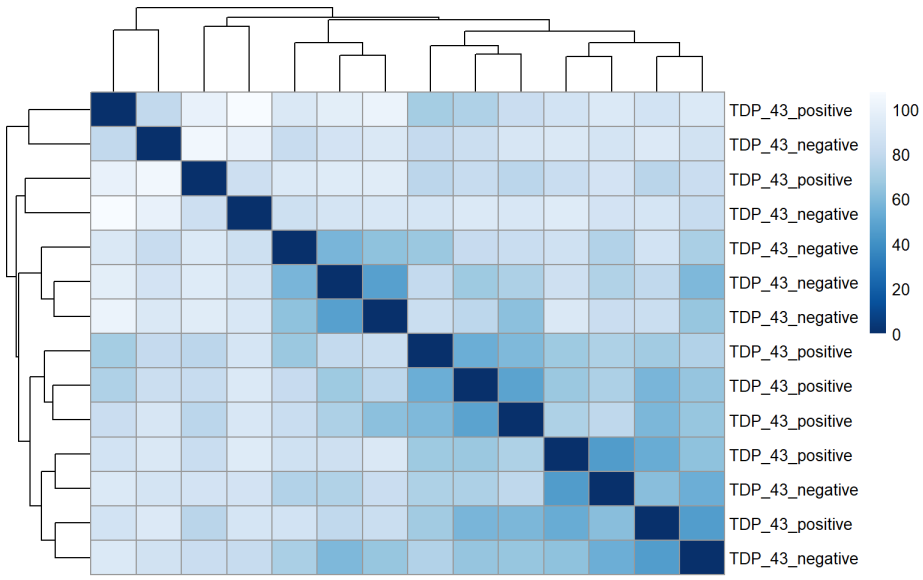
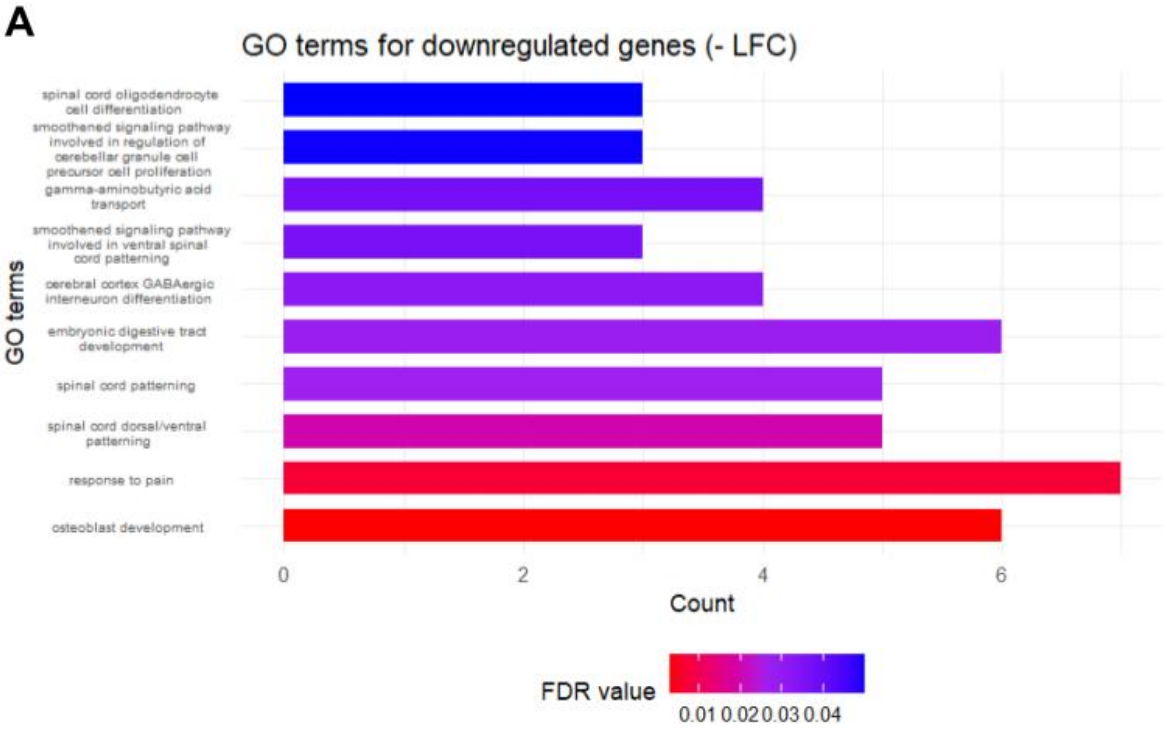


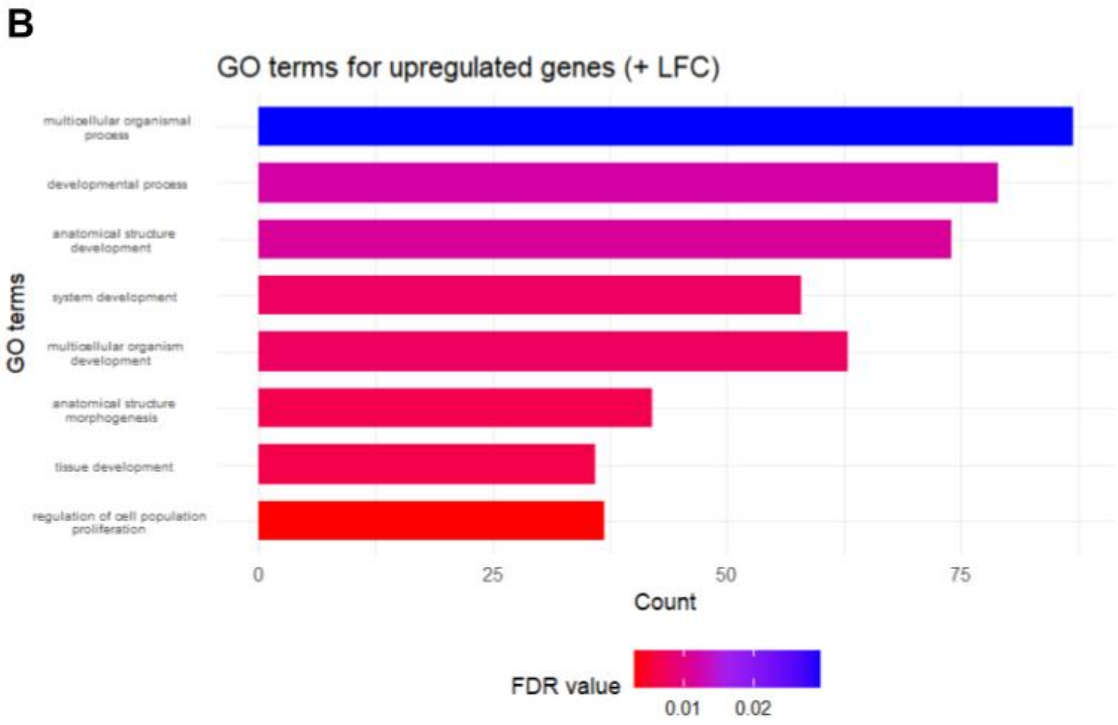
Figure 4: A correlation plot of combined gene expressions for TDP-43 positive and negative samples.

A generated correlation plot of the samples also accurately clusters the samples by their genotype (Figure 4). TDP-43 positive samples appear to be together in dark clumps, and the same is said for the majority of the negative samples, further validating the DESeq results expected from the count data.

Gene enrichment analysis was performed using GO, and was represented in bar plots with the gene count plotted against the most significantly present functional annotations (padj < .05) (Figure 5). Downregulated genes featured significant functional annotations directly related to nervous system development, such as “spinal cord oligodendrocyte differentiation” and “smoothened signaling pathway involved in regulation of cerebellar granule cell precursor cell proliferation” (Figure 5A). Upregulated genes failed to capture the same level of specificity, only using broad functional annotations such as “multicellular organismal processes” (Figure 5B).

Similar barplots were also generated using KEGG, an enrichment database dedicated to highlighting specific pathways overtly represented in a list of genes (Figure 6). Analogous trends were obtained. The downregulated genes contained pathway annotations for messengers directly related to the nervous system, like cyclic adenosine monophosphate (cAMP) (Figure 6A). Upregulated genes are less specific, yet point to general developmental processes.


A
GO terms for downregulated genes (- LFC)
spinal cord oligodendrocyte cell differentiation
smoothened signaling pathway involved in regulation of cerebellar granule cell precursor cell proliferation
gamma-aminobutyric acid transport
smoothened signaling pathway involved in ventral spinal cord patterning
cerebral cortex GABAergic interneuron differentiation
embryonic digestive tract development
spinal cord patterning
spinal cord dorsal/ventral patterning
response to pain
osteoblast development
GO terms
0
2
4
6
Count
FDR value
0.01 0.02 0.03 0.04


B
GO terms for upregulated genes (+ LFC)
multicellular organismal process
developmental process
anatomical structure development
system development
multicellular organism development
anatomical structure morphogenesis
tissue development
regulation of cell population proliferation
GO terms
0
25
50
75
Count
FDR value
0.01
0.02

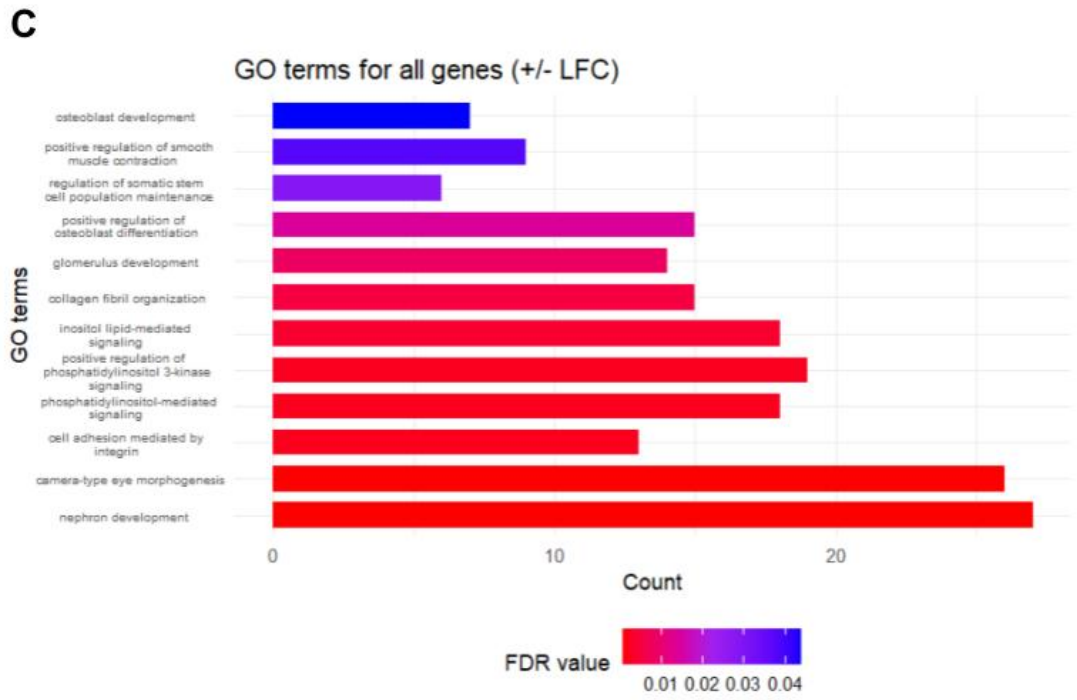
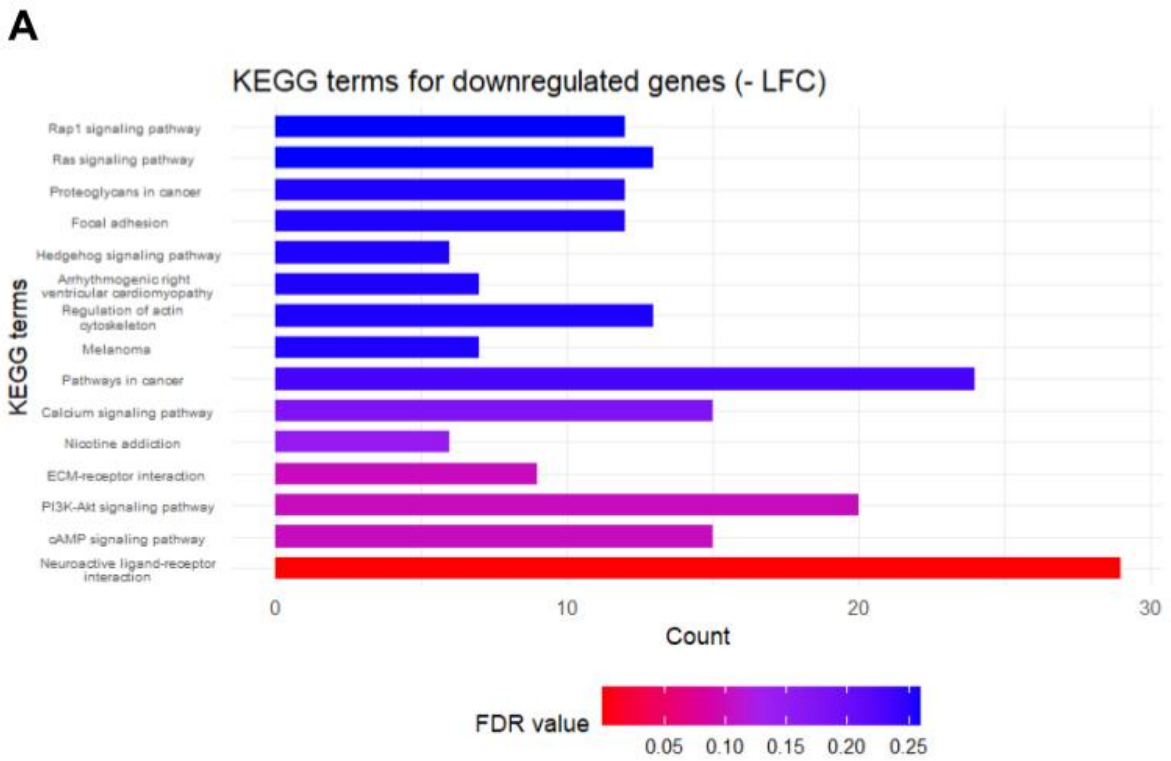
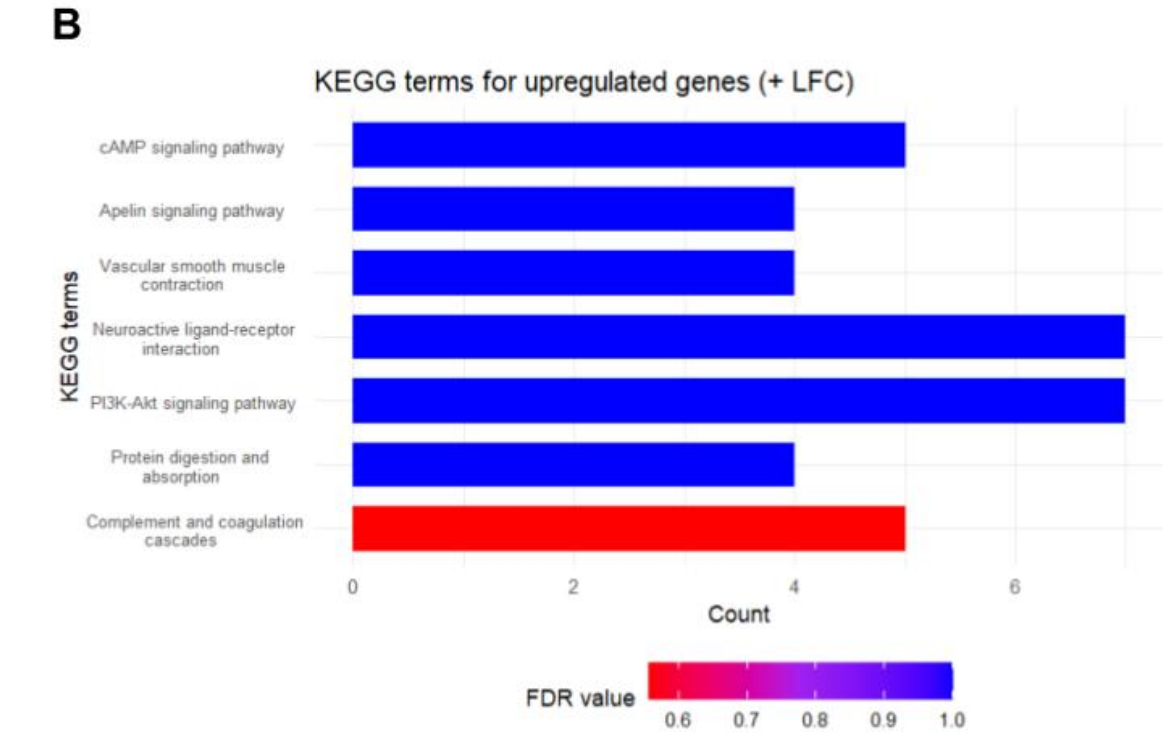


Figure 5: Barplots representing the gene counts for each biological functional annotation as determined using the GO Resource. **A.** Functional annotations for genes downregulated with a negative LFC in TDP-43 negative cells. Nervous system developmental processes dominate the annotations. **B.** Functional annotations for genes upregulated with a positive LFC in TDP-43 negative cells. Annotations fail to demonstrate specific processes related to FTD-ALS. **C.** Functional annotations for genes in TDP-43 negative cells, regardless of positive or negative LFCs.




B
KEGG terms for upregulated genes (+ LFC)
cAMP signaling pathway
Apelin signaling pathway
Vascular smooth muscle contraction
Neuroactive ligand-receptor interaction
PI3K-Akt signaling pathway
Protein digestion and absorption
Complement and coagulation cascades
KEGG terms
0
2
4
6
Count
FDR value
0.6
0.7
0.8
0.9
1.0

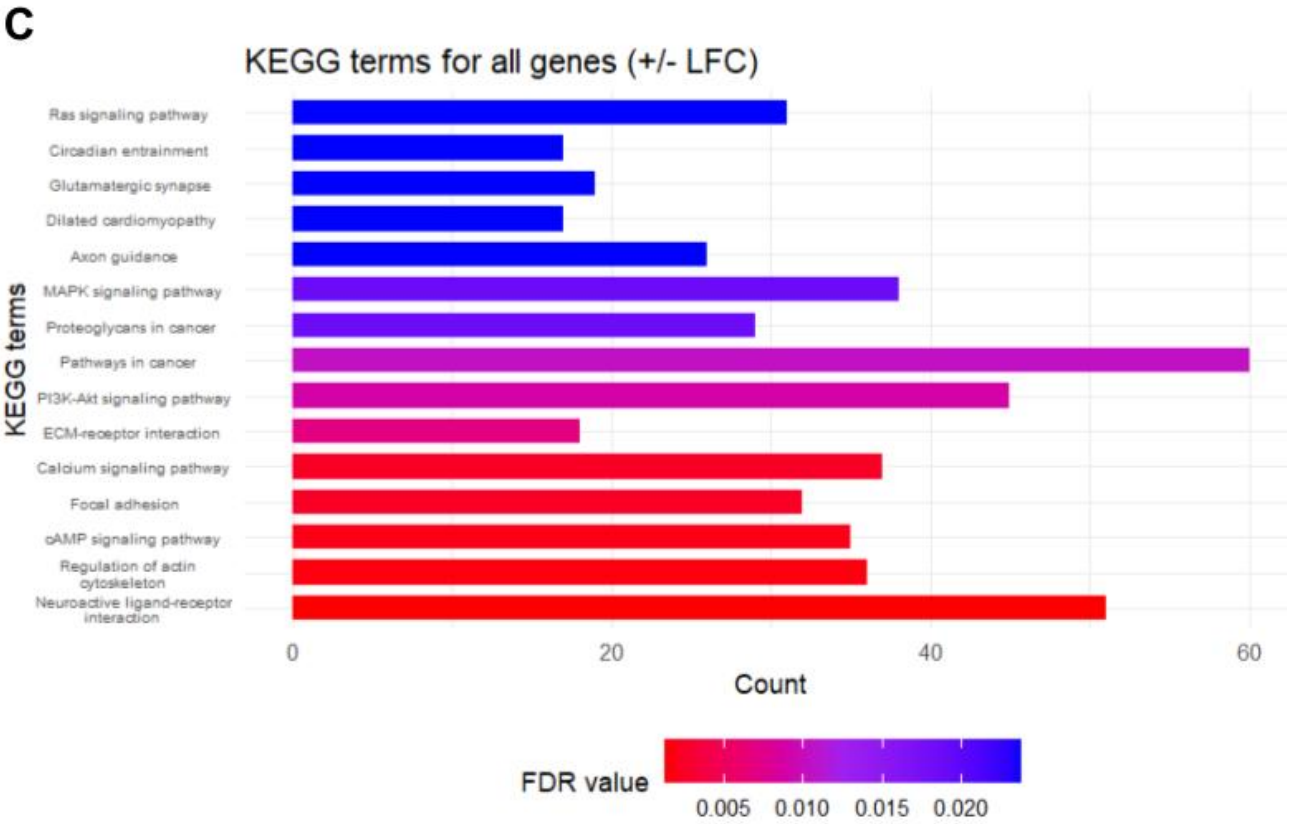

C
KEGG terms for all genes (+/- LFC)
Ras signaling pathway
Circadian entrainment
Glutamatergic synapse
Dilated cardiomyopathy
Axon guidance
MAPK signaling pathway
Proteoglycans in cancer
Pathways in cancer
PI3K-Akt signaling pathway
ECM-receptor interaction
Calcium signaling pathway
Focal adhesion
cAMP signaling pathway
Regulation of actin cytoskeleton
Neuroactive ligand-receptor interaction
KEGG terms
0
20
40
60
Count
FDR value
0.005
0.010
0.015
0.020

Figure 6: Barplots representing the gene counts for each KEGG pathway annotation as determined using KEGG. **A.** Pathway annotations for genes downregulated with a negative LFC in TDP-43 negative cells. Nervous system developmental processes dominate the annotations. **B.** Pathway annotations for genes upregulated with a positive LFC in TDP-43 negative cells. All annotations fail to reach significant levels of representation (padj < .05). **C.** Pathway annotations for genes in TDP-43 negative cells, regardless of positive or negative LFCs.

Discussion

Overall, the investigation into TDP-43 positive and negative gene expression patterns yielded expected results. When assigned functional and pathway annotations, the genes discovered to be significantly represented in the list of downregulated genes in the TDP-43 negative samples, while accounting for sex, result in annotations primarily focused on neurodevelopmental processes (Figure 5A, 6A). Specifically, terms found in the downregulated GO functions, like "osteoblast development" (padj=3.16E-03), correspond to previous research demonstrating that ALS results in impaired osteoblast properties (Zhu, 2015). Osteoblasts are responsible for regenerating bone, so the finding that their development is hindered in neuromuscular diseases like ALS, which result in muscular atrophy, is not surprising. Similarly, the finding that "cerebral cortex GABAergic interneuron differentiation" (padj=3.15E-02) is downregulated in TDP-43 negative cells correlates with the pathology of FTD, as the disease primarily causes frontotemporal lobe degeneration, which is a region located in the cerebral cortex (Murley, 2020). It is possible that in FTD patients, GABA-releasing neurons, which are often responsible for excitatory and inhibitory responses in the brain, fail to differentiate and specialize correctly, resulting in the adverse symptoms and effects of FTD. A relatively interesting insight, which is currently not accounted for in academic literature, is the finding that "spinal cord patterning" (padj=2.67E-02) genes are significantly downregulated in TDP-43 negative cells (Figure 5A). Spinal cord patterning involves the initial development of neurons along the rostral-caudal axis on the spinal cord (Andrews, 2019). But spinal cord patterning during early development also comprises the coordination of motor neurons and neuromuscular activity. Thus, the finding generated by DESeq and GO that proteins related to spinal cord patterning are downregulated later in life suggests that TDP-43 loss affects other instances of this pathway being activated, perhaps during periods of repair in that region. Thus, these malfunctioning processes may contribute to the dysfunction of motor neurons in ALS.

Similarly, when analyzing KEGG annotations focusing on pathways represented in the downregulated gene list, it was discovered that messengers, such as cAMP and the P13K-Akt signaling pathway, are included (Figure 6A). cAMP is a commonly referenced secondary messenger contributing to extracellular signal transduction in the nervous system (Duman et al., 1999). Thus, any downregulation in cAMP signaling may be indicative of some dysregulated neuronal function, whether that relates to the cognitive disabilities found in FTD or the motor functionality in ALS. Additionally, the P13K-Akt pathway has also been previously shown to be

inhibited in an ALS model, resulting in reduced motor neuron viability (Yin et al., 2015). Thus, it can be confidently concluded that downregulated genes and their resulting function and pathway enrichment methods generated results that did not differ from expectations of FTD-ALS's known pathology. FTD-ALS is a condition known to localize its impact on the brain and its motor and cognitive functions (Ferrari, 2013). Thus, the final annotations generated validate both this initial assumption and the analytical methodologies used to perform the differential expression analysis.

The upregulated genes and their corresponding annotations did not share the functional specificity generated by the downregulated genes. GO results for the upregulated genes only generated broad, umbrella terms like "tissue development" (Figure 5B). It is likely that though some proteins in general tissue development are being upregulated, there are no other instances of coordinated gene groupings within the upregulated gene list. The KEGG results for the upregulated genes provide some more specific annotations, such as "neuroactive ligand-receptor interaction" (padj=1) (Figure 6B). However, these annotations all demonstrate insignificant adjusted p-values over .05. Thus, the KEGG results likely lack any interpretative meaning, and support the previous insights that the genes with positive LFCs are missing any relevant, significant annotations that can be interpreted.

Although many of the annotations related back directly to FTD-ALS pathology, as previously stated, some findings were quite surprising. For instance, the KEGG analysis for both upregulated and downregulated genes returned a number of annotations relating to pathways in cancer (padj=.010), dilated cardiomyopathy (padj=.024), and focal adhesion (padj=0.002) in the extracellular matrix (Figure 6A, Figure 6C). At first, these results were ignored because despite their significant adjusted p-values, it was assumed that these annotations were the product of noise. However, it is possible that even these annotations reflect real trends seen in patients with ALS. Studies have found possible correlations between FTD-ALS and cardiomyopathy, as shown in the final KEGG analysis (Freedman et al., 2013; Gdynia et al., 2006). Additionally, as predicted by the downregulation of the "pathways in cancer" annotation, TDP-43 knockdown has also been shown to inhibit breast cancer progression (Ke, 2018). Though these two conditions have already been associated with ALS in the past, the fact that the corresponding DESeq results reflect this finding suggests that a considerable number of FTD-ALS comorbidities may also be the result of TDP-43 loss. Thus, similar RNA-seq gene enrichment methods can be possibly implemented to uncover previously-ignored conditions that also coexist with FTD-ALS. With more time, a more detailed literature review should be performed into every one of the annotations produced, regardless of direct relation to FTD-ALS, as a means to potentially discover common secondary pathways that are affected as a result of TDP-43 variants or loss. Additionally, in future work, these methods could possibly be implemented in a study with a greater sample size to further evaluate the statistical correlation between conditions like cardiomyopathy, cancer, and FTD-ALS with TDP-43 loss.

Though the final results appear to generally reflect expectations, there are some elements of the analytical methods that could be improved. It is questionable whether the LFC shrinkage performed is considerably impacting the variance or general trends measured in the DESeq results (Figure 2). For instance, the apeglm shrinkage is preferred for small sample sizes, such as the ones in this dataset which only utilize 14 samples. However, apeglm also severely reduced the noise by eliminating almost all genes that were not differentially expressed between the two groups, inviting skepticism regarding the processes or bias present in the dataset, and whether it was configured correctly (Figure 2A). A literature review of the apeglm function used to perform the shrinkage could not point to a specific reason as to why this occurred, and thus, with no certainty in the benefit derived, the shrinkage was ignored for future steps (Anqi, 2019). Similar shrinkage methods also did not provide any clear differences compared to the data without LFC shrinkage (Figure 2B). Thus, with more time, the benefit and cost of performing LFC shrinkages should be further investigated as it relates to the context of this particular dataset.

Additionally, there is likely an additional source of variance which was not accounted for in the design formula or GSE sample metadata. The histogram of adjusted p-values shows a clear skew to the right, which indicates that the p-values are not uniformly distributed as intended (Figure 3B). This may be the result of additional variance that is not represented in the current design formula. In the PCA plot, there is a clear gap separating the TDP-43 negative samples into smaller clusters (Figure 3A). After testing all possible metadata variables, as well as assigning pseudo-batch values to each sample, no considerable improvement was made in the histogram's distribution. The skew of these adjusted p-values suggests that the detection of false positives is thus flawed. In future analyses, more effort should be dedicated towards uncovering possible sources of variation that accounted for this distribution.

In addition to these limitations, the largest challenges for completing the project involved technical issues. For the beginning stages of the project, it was anticipated that performing DESeq requires intense computational power, and thus necessitates the use of the Tufts cluster or compute server. This presented complications to otherwise simple methods, as X11 forwarding had to be correctly configured to enable the generation of graphics from the data. Similarly, IDEs utilized to connect to the compute server presented new complications, such as preventing custom configuration of chart dimensions, or occasionally corrupting graphics generated by DESeq. Ultimately, it was discovered that these analyses can be performed on a local PC, and thus, RStudio was opted to instead perform the DESeq workflow.

On a personal level, this project allowed me to develop valuable statistics skills by introducing me to the methodological mindset required to execute a biostatistics workflow on real patient data. I found that as I implemented each concept learned in class, such as how to interpret different outputs generated by software like DESeq, I found myself understanding the course material on a deeper level. At the same time, the project provided me with an opportunity to

exercise my background as a Biology student when interpreting the final annotations generated by gene enrichment software.

Works Cited

Andrews, Madeline G et al. “New perspectives on the mechanisms establishing the dorsal-ventral axis of the spinal cord.” *Current topics in Developmental Biology* vol. 132 (2019): 417-450. doi:10.1016/bs.ctdb.2018.12.010

Anqi Zhu, Joseph G Ibrahim, Michael I Love, Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences, *Bioinformatics*, Volume 35, Issue 12, June 2019, Pages 2084–2092, https://doi.org/10.1093/bioinformatics/bty895

Duman RS, Nestler EJ. Functional Roles for cAMP and cGMP. In: Siegel GJ, Agranoff BW, Albers RW, et al., editors. *Basic Neurochemistry: Molecular, Cellular and Medical Aspects.* 6th edition. Philadelphia: Lippincott-Raven; 1999. Available from: https://www.ncbi.nlm.nih.gov/books/NBK27915/

Ferrari, R et al. “FTD and ALS: a tale of two diseases.” *Current Alzheimer Research* vol. 8,3 (2011): 273-94. doi:10.2174/156720511795563700

Freedman, D Michal et al. “The association between cancer and amyotrophic lateral sclerosis.” *Cancer Causes & Control: CCC* vol. 24,1 (2013): 55-60. doi:10.1007/s10552-012-0089-5

Gdynia, H-J et al. “Cardiomyopathy in motor neuron diseases.” *Journal of Neurology, Neurosurgery, and Psychiatry* vol. 77,5 (2006): 671-3. doi:10.1136/jnnp.2005.078600

Ke, Hao et al. “Loss of TDP43 inhibits progression of triple-negative breast cancer in coordination with SRSF3.” *Proceedings of the National Academy of Sciences of the United States of America* vol. 115,15 (2018): E3426-E3435. doi:10.1073/pnas.1714573115

Liu, E. Y. et al. Loss of nuclear TDP-43 is associated with decondensation of LINE retrotransposons. *Cell Rep.* 27, 1409–1421 (2019).

Love, M.I., Huber, W., Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology,* 15:550. 10.1186/s13059-014-0550-8

Ma, X.R., Prudencio, M., Koike, Y. et al. TDP-43 represses cryptic exon inclusion in the FTD–ALS gene UNC13A. *Nature* 603, 124–130 (2022). https://doi.org/10.1038/s41586-022-04424-7

Masrori P, Van Damme P. Amyotrophic lateral sclerosis: a clinical review. Eur J Neurol. 2020 Oct; 27(10):1918-1929. doi: 10.1111/ene.14393. Epub 2020 Jul 7. PMID: 32526057; PMCID: PMC7540334.

Murley, Alexander G et al. "GABA and glutamate deficits from frontotemporal lobar degeneration are associated with disinhibition." *Brain : A Journal of Neurology* vol. 143,11 (2020): 3449-3462. doi:10.1093/brain/awaa305

Stephens, M. (2016) False discovery rates: a new deal. *Biostatistics*, 18:2. 10.1093/biostatistics/kxw041

Suk, T.R., Rousseaux, M.W.C. The role of TDP-43 mislocalization in amyotrophic lateral sclerosis. *Mol Neurodegeneration* 15, 45 (2020). https://doi.org/10.1186/s13024-020-00397-1

Yin, Xiang et al. "Downregulated AEG-1 together with inhibited PI3K/Akt pathway is associated with reduced viability of motor neurons in an ALS model." *Molecular and Cellular Neurosciences* vol. 68 (2015): 303-13. doi:10.1016/j.mcn.2015.08.009

Zhu, Ke et al. "Impaired bone homeostasis in amyotrophic lateral sclerosis mice with muscle atrophy." *The Journal of Biological Chemistry* vol. 290,13 (2015): 8081-94. doi:10.1074/jbc.M114.603985