Bioinformatics analysis of the transcriptomic profiles of Duchenne muscular dystrophy patients / narrative competence and cognitive mapping as a culturally sustaining pedagogy in the education of emergent bilinguals

Bianca Gonzalez

University at Albany, State University of New York, bgonzalez3@albany.edu

5-1-2022

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Bioinformatics Commons, and the Biology Commons

Recommended Citation

This Master's Thesis is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
Bioinformatics analysis of the transcriptomic profiles of Duchenne Muscular Dystrophy Patients

by

Bianca G. Gonzalez

A Thesis

Submitted to the University at Albany, State University of New York

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science

College of Arts & Sciences

Department of Biological Sciences

Spring 2022
Abstract

Duchenne Muscular Dystrophy (DMD) is a muscle wasting disease that primarily affects boys characterized by loss of the dystrophin gene. Dystrophin is necessary for proper muscle growth, muscles that lack dystrophin are more prone to damage. In this thesis, a bioinformatics approach was used to evaluate the transcriptomic profiles of two datasets obtained from DMD patients. The two datasets were analyzed separately, as well as combined to identify differentially expressed genes. The first dataset, GSE6011, has 23 DMD samples and 14 controls. The second, GSE38417, consists of 16 DMD samples and six controls. Differential gene expression of both datasets combined found that GSE6011 had 1586 significant genes and GSE38417 had 7107; overall meta-analysis identified 3321 significant genes. Gene Ontology analysis showed enrichment of genes related to inflammation. Differential gene analysis of the datasets separately identified upregulation of ECM components and the developmental isoforms of heavy myosin heavy chains. The upregulation of MYH8 and downregulation of dystrophin are both considered hallmarks of Duchenne muscular dystrophy and were identified in my analysis.

Gene ontology analysis also demonstrated an upregulation of immune response-related gene sets.

In conclusion my analysis showed that many genes are being differentially expressed in Duchenne muscular dystrophy, particularly those related to immune system response.
ACKNOWLEDGMENTS

Thankyou Dr. Bijan Dey, Dr. Melinda Larsen and Dr. Andy Berglund.

Special thank you to Dr. Pauline Carrico for helping with edits and your professional advice. You became a great friend and mentor in a short period of time.
TABLE OF CONTENTS

Abstract  -------------------------------------------------------------------------------------------------------------

Chapter 1: Introduction

1.1 Epidemiology of Duchenne muscular dystrophy---------------------------------------------------------------

1.2 Dystrophin and the DAPC  -----------------------------------------------------------------------------------------------------------------------------

1.3 Impact of dystrophin loss in muscle cells-----------------------------------------------------------------------------------------------------------------

1.4 Inflammation in Duchenne muscular dystrophy--------------------------------------------------------------------------------------------

1.5 Fibrosis in Duchenne Muscular dystrophy --------------------------------------------------------------------------------------------------------------------------------

Chapter 2: Thesis Objective

Chapter 3: Materials and Methods

3.1 Datasets  -------------------------------------------------------------------------------------------------------------

3.2 Analysis using GEO2R-------------------------------------------------------------------------------------------

3.3 Meta-analysis using ImaGEO ----------------------------------------------------------

Chapter 4: Results and Discussion

4.1 Table 1: GSE38417 Top ten upregulated genes-----------------------------------------------------------------------------------------

4.2 Table 2: GSE38417 Top Ten downregulated genes------------------------------------------------------------------
Chapter 5: Conclusions and Future Directions

References
Chapter 1: Introduction

Epidemiology of Duchenne Muscular dystrophy

Duchenne muscular dystrophy is an X-linked recessive disorder that primarily affects boys with a prevalence of about 10 in every 100,000 males. It is possible for girls to be afflicted with muscular dystrophy but at a significantly lower rate of less than one per million females. Women with DMD are asymptomatic carriers of the disorder in one of their X-chromosomes. (Duan, Goemans, Takeda, Mercuri, & Aartsma-Rus, 2021) Patients with DMD typically start to show symptoms at a young age, around 2-3 years of age. Children inflicted with DMD will experience difficulties climbing stairs, a waddling gait, and frequent falls. DMD is a progressive disease, with muscle wasting growing more severe as the child ages. By the ages of 10-12 most become wheelchair bound and by 20 years of age will likely require ventilation to some degree. With proper treatment a DMD patient may up to 40 years, usually succumbing to respiratory or heart failure.

Dystrophin and the DAPC

Duchenne muscular dystrophy results from mutations of the dystrophin gene, that prevent the production of Dp427m, the isoform specific to muscle tissue. DMD also encodes two other isoforms Dp427c, expressed in cortical neurons, and Dp427p, expressed in cerebellar Purkinje cells. (Duan, Goemans, Takeda, Mercuri, & Aartsma-Rus, 2021) Within muscle cells, dystrophin links the cytoskeletal F-actin with the extracellular matrix via both its N and C-terminal domains (Duan, Goemans, Takeda, Mercuri, & Aartsma-Rus, 2021). The carboxyl terminal of dystrophin binds the dystrophin associated complex (DAPC). DAPC consist of dystroglycans, sarcoglycans, integrins and caveolin. Removal of dystrophin destabilizes DAPC, weakening the important link
between the cytoskeleton and extracellular matrix. (Nowak & Davies, 2004) Subsequently
muscle fibers of Duchenne muscular dystrophy patients are prone to muscle damage every time
they contract their muscles. As previously mentioned, deletions are not the only mutations seen,
frameshifting and nonsense mutations of dystrophin also are seen in DMD. Such mutations result
in early stopping of protein translation, the truncated dystrophin may be non-functional and/or
unstable. (Duan, Goemans, Takeda, Mercuri, & Aartsma-Rus, 2021)

**Impact of dystrophin loss in muscle cells**

When dystrophin deficient muscles contract, mechanical stressed is placed on the
sarcolemma, due to the destabilized DAPC. In healthy tissue dystrophin binds to the sarcolemma
along with the DAPC, linking the cytoskeleton, extracellular matrix, and sarcolemma. (Duan,
Goemans, Takeda, Mercuri, & Aartsma-Rus, 2021) This linkage helps maintain the integrity of
the sarcolemma. But when dystrophin is loss in DMD patients the sarcolemma is damaged by
muscle contraction. Damage to the membrane allows calcium ions to enter the muscle cell,
subsequently triggering degradation pathways and eventually ending in muscle cell death.
(Brusa, Magri, Bresolin, Comi, & Corti, 2019) (Lovering, Porter, & Bloch, 2005). About 60-70%
of the DMD mutations are deletions. Other types of mutations are possible: 5-15% are
duplications, and 20% are point mutations. All these mutations result in lack of functional
dystrophin in muscle cells. Loss of dystrophin results in muscle is damaged every time it
contracts, causing a progressive loss in muscle mass.

**Inflammation and Duchenne muscular dystrophy**
When there are high levels of inflammatory cytokines such as, TNFa, IL-6, and TWEAK, muscle regeneration is inhibited. DMD muscle tissue exhibits chronic inflammation, the consistent inhibition of myogenesis is a major driver for muscle wasting. (Howard, Pasiako, Blesso, Fussell, & Rodriquez, 2020). But inflammation isn’t all bad; it’s a necessary response to muscle injury that helps with regeneration. Initially the pro-inflammatory cytokines that are released from M1 macrophages stimulate myoblast proliferation. Next anti-inflammatory cytokines from M2 macrophages promote the differentiation of myoblast. Myoblast proliferation and differentiation are the first steps towards a mature myofiber. (Arnold, et al., 2007) But when inflammation is chronic, like that of DMD, muscle regeneration capacity is loss. In dystrophin deficient muscles the damage to the sarcolemma activates necrosis. Dying myofibers are then removed by M1 macrophages while M1 cells secrete proinflammatory cytokines (Cruz-Guzman, Rodriquez-Cruz, & Cedillo). Muscle biopsy of DMD patients have previously shown an upregulation of proinflammatory cytokines such as TNF-a, IL-1, and IL-6. (Cruz-Guzman, Rodriquez-Cruz, & Cedillo).

The body’s innate immune response has pattern recognition receptors such as toll like receptors (TLRs) that can detect both materials from microbes and materials coming from damage to cells. Toll-like receptors are a family of receptors that recognize invading pathogens and other molecules that may pose a threat to the body. When ligands bind to and activate this family of receptors intracellular signaling cascades initiate the innate immune response. Damaged cells release different molecules known as DAMPs, which will activate TLRs when released from the cell. Specifically, in Duchenne muscular dystrophy DAMPs are single stranded myofiber-derived RNA molecules that arise from the damaged dystrophin deficient muscles have
been identified. DMD patients see an upregulation of TLR7 (Rosenberg, et al., 2018) (Chen Y.W., et al., 2005) of the NF-kB signaling pathway which induces the expression of inflammatory cytokines and other inflammation mediators.

*Fibrosis in Duchenne muscular dystrophy patients*

The chronic inflammation and muscle injury that occurs in Duchenne muscular dystrophy patients creates a vulnerability to fibrosis, which is the hardening and formation of scar in tissues. Resulting from an uncontrolled wound-healing process triggered by chronic inflammation and injury. Evaluation of fibrosis in Duchenne muscular dystrophy muscle and immune stained for Collagen IV. Their results demonstrated a significant increase of collagen deposition within the biopsied muscle (Zhou & Lu, 2010)
Chapter 2: Thesis Objective

In this thesis I performed meta-analysis to examine the expression profiles of Duchenne muscular dystrophy samples from two different datasets. My goal was to identify what genes are being differently expressed in these samples. I focused on the top ten upregulated and downregulated genes from each dataset. Then used current literature to determine which of these genes have been identified as differentially expressed in Duchenne muscular dystrophy. GEO2R was used to analyze the log-fold change of gene expression, ranking by p-value. A meta-analysis done on both datasets together was performed using ImaGEO. Lastly, gene ontology enrichment analysis was performed on GSE38417 using ReactomePA, the results of which were compared to a previously published gene ontology analysis of GSE6011. Additionally, gene ontology analysis was also done on both datasets together using ImaGEO. Some processes of interest for Duchenne muscular dystrophy are myogenesis and inflammation. It was interesting to determine what if any genes related to these processes are being differently expressed in the two datasets.

My overall goal was to cast light on what genes are being differently expressed in DMD. Analysis of each dataset separately allows for evaluation of how differential gene expression changes in DMD patients from different datasets. Meta-analysis of the two datasets combined allows for evaluation of differential gene expression in DMD patients on a broad scale. One dataset is from pre-symptomatic patients, the other is from older symptomatic patients. Duchenne muscular dystrophy is a progressive disease so age may be a factor in how differential gene expression profiles changes between the two datasets. Gene ontology analysis evaluates what processes and signaling pathways are differentially expressed in Duchenne muscular dystrophy patients.
Chapter 3 Materials and Methods

Datasets

The microarray dataset used is GSE38417 and was generated from RNA extracted from muscle biopsy samples from boys with Duchenne muscular dystrophy. The biopsies were taken and examined by Dr. Eric P. Hoffman at the children’s national medical center. Additionally, pathologically normal samples were taken as controls. The original researchers performed a western blot analysis to confirm the absence of dystrophin in DMD samples, as well as the presence of it in control samples. There was a total of sixteen Duchenne muscular dystrophy samples and six control samples. Each sample originated from a different individual. The ages of the control samples were not aged matched. Of the DMD patients, five were less than two years of age, eight between the ages of 3-5, and three were between the ages of 6 and 8. The purpose of this study was to evaluate what genes are differentially expressed in the muscles of Duchenne muscular dystrophy patients versus those of the controls. To examine the gene profiles of the samples, the original researchers used Human U133 2.0 arrays. For their analysis they set the FDR p value for significance at 0.05 and set the fold change to at least 1.5 in DMD/CTRL. For my analysis I will be using the same p-value of 0.05 and a fold difference of at least 1.5, the same as the previously published paper.

The second dataset is GSE6011 which contains the expression data from the quadriceps from young DMD patients and age matched controls. There are expression profiles from 19 DMD patients characterized as pre-symptomatic (1-2 years old). There are 14 age-matched control samples from boys not afflicted with DMD. None of the patients were reported to be undergoing any anti-inflammatory corticosteroid treatments. The chip type used was the HG-
U133A genechipCRNA, RNA was isolated from the samples using protocols from the Affymetrix gene Expression Manual. In the original publication the researchers looked at the overall genome to see what genes are differently expressed in DMD patients.

The original analysis was done using p-value of 0.05 and a fold change difference of at least 1.5. My analysis used the same p-value of 0.05 and my fold change threshold will be 1.5. (Pescatori, et al., Gene Expression profiling in the early phases of DMD: a constant molecular signature characterizes DMD muscle from early postnatal life throughout disease progression, 2007).

Methods

To analyze the individual datasets separately GEO2R was used. The interactive web tool uses the Bioconductor packages from R studio to compare two or more groups of samples in datasets uploaded to gene expression omnibus database. The program ranks the differentially expressed genes by p-value, the smallest p-value are the most statistically significant. By default, the Benjamini and Hochberg false discover rate method is used to adjust p-values, this method is commonly used for microarray datasets.

ImaGEO is a meta-analysis web tool that accepts GSE datasets directly and does an analysis using R-studio. The analysis was performed as a fixed effect model, which the program states is better for studies in which differences are thought to be due to an underlying effect, such as DMD. The p-value threshold was set to 0.05 which is the same value used by previously published papers. The generated heatmap demonstrates differences in gene expression of DMD patients compared to the controls (Figure 1) (Toro-Dominguez, et al., 2018).

Up-regulated genes had a fold change of at least 1.5 for GSE38417 and 1.0 for GSE6011.
Down-regulated genes did not exceed fold change value of 1.5 for GSE38417 and 1.0 for GSE6011. For the combined analysis above a fold change of 1.5 were considered up-regulated, those with fold changes below 1.5 were considered down-regulated.
Chapter 4 Results and Discussion

GSE6011 and GSE38417 was previously used in a published study focusing on the MIF network. Their statistical analysis used linear models for microarray data, the cutoffs for p-value were <0.05 and log fold change of > 2. Their analysis reported that Macrophage Migration Inhibitory Factor network and the receptors CD74, CD44, and CXCR4 were all upregulated in Duchenne muscular dystrophy patients. (Lombardo, et al., 2019) Macrophage inhibitor factor is a proinflammatory cytokine that is secreted by activated T-cell and macrophages in response to stress as a regulator of the resulting immune response. MIF activates the ERK1/ERK2-mitogen activated protein kinase pathway. (Calandra & Roger, 2003)

The two datasets differ in age. GSE38417 has older patients while GSE6011 has younger pre-symptomatic patients. The fold changes of GSE6011 were much lower than those of GSE38417. Duchenne muscular dystrophy is a progressive muscle wasting disease so the transcriptomic profiles of patients may differ as they grow older, and the disease gets more severe, which may explain the differences in fold changes observed. For each dataset the top ten up-regulated and down-regulated genes were focused on to pin-point differences in how differential gene expression in DMD patients differs between the two datasets. (table 1-4). The patients in GSE6011 are all pre-symptomatic and the upregulated and downregulated genes were different than those of GSE38417, which will be discussed in greater detail in the following paragraphs.

<table>
<thead>
<tr>
<th>Probe ID</th>
<th>Gene Symbol</th>
<th>logFC</th>
<th>FC</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>34471_at</td>
<td>MYH8</td>
<td>9.36</td>
<td>657.1141</td>
<td>1.3E-12</td>
</tr>
<tr>
<td>205940_at</td>
<td>MYH3</td>
<td>8.51</td>
<td>364.5569</td>
<td>1.6E-11</td>
</tr>
</tbody>
</table>
### Table 1: Top 10 upregulated genes of the GSE38417. FC threshold was set to greater than 1.5 and the p-value threshold was set to 0.05. FC = DMD/Controls. A total of 5608 genes were upregulated in GSE38417.

<table>
<thead>
<tr>
<th>Probe ID</th>
<th>Gene Symbol</th>
<th>logFC</th>
<th>FC</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>206717_at</td>
<td>MYH8</td>
<td>8.14</td>
<td>282.0877</td>
<td>2.75E-12</td>
</tr>
<tr>
<td>209875_s_at</td>
<td>SPP1</td>
<td>6.58</td>
<td>95.67035</td>
<td>6.07E-12</td>
</tr>
<tr>
<td>203649_s_at</td>
<td>PLA2G2A</td>
<td>6.45</td>
<td>87.42658</td>
<td>2.75E-08</td>
</tr>
<tr>
<td>205145_s_at</td>
<td>MYL5</td>
<td>6.19</td>
<td>73.00888</td>
<td>2.24E-10</td>
</tr>
<tr>
<td>230867_at</td>
<td>COL6A6</td>
<td>6.12</td>
<td>69.55103</td>
<td>1.1E-14</td>
</tr>
<tr>
<td>215389_s_at</td>
<td>TNNT2</td>
<td>5.85</td>
<td>57.68003</td>
<td>8.03E-11</td>
</tr>
<tr>
<td>238047_at</td>
<td>ARHGAP36</td>
<td>5.81</td>
<td>56.10277</td>
<td>2.37E-10</td>
</tr>
</tbody>
</table>

### Table 2: Top 10 downregulated Genes from dataset GSE38417. FC threshold was set to less than 1 and p-value threshold was set to 0.05. FC = DMD/controls. A total of 15757 genes were downregulated in GSE38417.

<table>
<thead>
<tr>
<th>Probe ID</th>
<th>Gene Symbol</th>
<th>logFC</th>
<th>FC</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>205694_at</td>
<td>TYRP1</td>
<td>-5.79</td>
<td>0.018073</td>
<td>1.3E-12</td>
</tr>
<tr>
<td>224588_at</td>
<td>XIST</td>
<td>-5.2</td>
<td>0.027205</td>
<td>0.000197</td>
</tr>
<tr>
<td>1561504_s_at</td>
<td>MYLK4</td>
<td>-5.19</td>
<td>0.027394</td>
<td>2.62E-10</td>
</tr>
<tr>
<td>1558322_a_at</td>
<td>PAQR9</td>
<td>-4.96</td>
<td>0.032129</td>
<td>7.31E-14</td>
</tr>
<tr>
<td>213880_at</td>
<td>LGR5</td>
<td>-4.9</td>
<td>0.033493</td>
<td>2.11E-08</td>
</tr>
<tr>
<td>205030_at</td>
<td>FABP7</td>
<td>-4.7</td>
<td>0.038473</td>
<td>1.33E-10</td>
</tr>
<tr>
<td>221008_s_at</td>
<td>ETNPPL</td>
<td>-4.65</td>
<td>0.03983</td>
<td>1.54E-15</td>
</tr>
<tr>
<td>227671_at</td>
<td>XIST</td>
<td>-4.65</td>
<td>0.03983</td>
<td>0.000187</td>
</tr>
<tr>
<td>242271_at</td>
<td>SLC26A9</td>
<td>-4.62</td>
<td>0.040667</td>
<td>5.52E-13</td>
</tr>
<tr>
<td>206349_at</td>
<td>LGI1</td>
<td>-4.61</td>
<td>0.04095</td>
<td>5.26E-15</td>
</tr>
</tbody>
</table>
In dataset GSE38417 a total of 5608 genes were found to be upregulated and a total of 15757 were considered downregulated. I have included the top 10 of both the upregulated and downregulated genes in my results section (table 1 and Table 2). The top ten upregulated genes were MYH8, MYH3, MYH8, SPP1, COL1A1, PLA2G2A, MYL5, COL6A6, TNNT2, and ARHGAP36. Past analysis of Duchenne muscular dystrophy patients has also reported upregulation of TNNT2 (Bakay, Zhao, Chen, & Hoffman, A web-accessible complete transcriptome of normal human and DMD muscle, 2002). This provides validation to my results.

Upregulation of MYH8 is a reported hallmark of Duchenne muscular dystrophy. (Haslett, et al., 2002). The upregulation of MYH8 has been previously reported in Duchenne muscular dystrophy patients. Its overexpression is considered a hallmark of muscular dystrophy. MYH8, along with MYH3, are developmental isoforms of skeletal muscle myosin heavy chains. (Haslett, et al., 2002) Typically, developmental isoforms such as MYH8 are present during embryonic development and then disappear after birth. In adults, the developmental myosin's will become expressed during muscle regeneration. (Schiaffino, Rossi, Smerdu, Leinwand, & Reggiani, 2015) MYH8 upregulation is observed across the board in my analysis of both datasets separately (table 1 and table 3) as well as in the combined analysis (figure 1).

Collagen type III alpha 1 chain gene is responsible for providing the instructions to make type III collagen. Type III collagen is found in the skin, lungs, intestinal walls, and the walls of blood vessels. Collagen proteins provide strength and support of the many tissues. Collagen is also a major component of the extracellular matrix. Type I collagen is the most abundant type found in the body. Collagen type I alpha 2 chain was another collagen protein included in the top 10 upregulated genes. COL1A2 is one of the genes responsible for providing instructions to
make type I collagen. Additionally, Collagen type I and type III have been previously shown to have large log fold changes when looking at the entire transcriptome of Duchene muscular dystrophy patients. That study was looking at early-stage DMD to see what was happening genetically. They found that a 1.5-month-old patient was already experiencing an increase in extracellular matrix synthesis. (Pescatori, et al., Gene expression profiling in the early phases of DMD: a constant molecular signature characterizes DMD muscle from early postnatal life throughout disease progression, 2007) My analysis did not focus on the age of the patients, but I was able to confirm that in Duchene muscular dystrophy patients there is a clear collagen increase in Duchenne muscular dystrophy patients (table 1).

To understand why ECM components such as collagen may be upregulated, I looked to previous literature. Matrix metalloproteinase-1 (MMP-1) is a protein that hydrolyzes the components of the extracellular matrix (Visse & Nagase, 2003). MMP-1 was found to be downregulated in the fibroblasts of DMD patients, MMP-1 is responsible for degrading type I and type III collagen and limiting fibrosis. The same study also reported an upregulation of collagen type I. (Zanotti, Gibertini, & Mora, 2009). Another study explained that connective tissue infiltration is a secondary response to the absence of dystrophin. They used microarrays to study gene expression in DMD patients and found that genes encoding type I and type III collagen, also known as fibril forming collagens, were upregulated in DMD skeletal muscle. Such observations are not surprising considering injured tissue has been found to have higher levels of these collagens. One theory is that the proteoglycan’s biglycan and Lumigan interact with collagen and forming a connection with the ECM and DAPC. Additionally, in dystrophin deficient tissue the biglycan interaction may act as a greater stabilizer to compensate for the lack of dystrophin. (Haslett, et al., 2002).
Secreted phosphoprotein 1 is an important non-collagenous bone protein also known as Osteopontin. SPP1 has been identified as a modifier gene that impacts the phenotype of Duchenne Muscular Dystrophy. (Vo & McNally, 2015). Cultures of DMD muscles cells demonstrated that the absence of dystrophin induces Osteopontin (SPP1) expression. During the early stages of muscle cell regeneration Osteopontin aids the process. But chronic overexpression of Osteopontin is suspected to hinder regeneration (Vianello, et al., 2017) In my analysis SPP1 was identified as a gene being upregulated in DMD patient samples of GSE34817 (table 1). It was also reported as being upregulated in DMD patients in the combined analysis (figure 1). SPP1 has an impact on the severity of disease progression and how well patients respond to glucocorticoid treatments. SPP1 has many functions and is considered an inflammatory marker expressed in muscle fibers and muscle inflammatory cells. Interestingly in mdx mice, double knockout of SPP1 and dystrophin resulted in decreased fibrosis and less functional deficits. But the same study also demonstrated that decreased SPP1 expression in mdx mice resulted in increased muscle weakness. Indicating that SPP1 has a positive impact on muscle regeneration and a negative role in inflammation. (Pegoraro, et al., 2011).

A total of 15757 genes were reported as downregulated in the DMD patients of the GSE38417 study. The top ten downregulated genes were TYRP1, XIST, MYLK4, PAQR0, LGR5, FABP7, ETNPPL, XIST, SLC26A9 and LGI1. (Table 2).
<table>
<thead>
<tr>
<th>Probe ID</th>
<th>Gene Symbol</th>
<th>logFC</th>
<th>FC</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>205940_at</td>
<td>MYH3</td>
<td>4.492622</td>
<td>22.51199</td>
<td>5.38E-19</td>
</tr>
<tr>
<td>206717_at</td>
<td>MYH8</td>
<td>4.238497</td>
<td>18.8762</td>
<td>1.73E-21</td>
</tr>
<tr>
<td>34471_at</td>
<td>MYH8</td>
<td>4.124213</td>
<td>17.4386</td>
<td>1.52E-22</td>
</tr>
<tr>
<td>205132_at</td>
<td>ACTC1</td>
<td>2.745661</td>
<td>6.706968</td>
<td>6.62E-18</td>
</tr>
<tr>
<td>201744_s_at</td>
<td>LUM</td>
<td>2.287392</td>
<td>4.881729</td>
<td>1.16E-09</td>
</tr>
<tr>
<td>202404_s_at</td>
<td>COL1A2</td>
<td>2.282401</td>
<td>4.864868</td>
<td>2.82E-09</td>
</tr>
<tr>
<td>204122_at</td>
<td>TYROBP</td>
<td>2.126627</td>
<td>4.366952</td>
<td>1.66E-17</td>
</tr>
<tr>
<td>219087_at</td>
<td>ASPN</td>
<td>2.104746</td>
<td>4.30122</td>
<td>2.21E-10</td>
</tr>
<tr>
<td>202310_s_at</td>
<td>COL1A1</td>
<td>2.055216</td>
<td>4.156059</td>
<td>1.51E-09</td>
</tr>
<tr>
<td>211161_s_at</td>
<td>COL3A1</td>
<td>2.019218</td>
<td>4.05364</td>
<td>1E-09</td>
</tr>
</tbody>
</table>

Table 3 Top ten Upregulated genes from dataset GSE6011. FC threshold was set to greater than 1.5 and a p-value threshold was set to 0.05. FC values greater than 1.5 are considered upregulated. A total of 1637 genes were upregulated in GSE6011.
Table 4 Top 10 downregulated Genes from dataset GSE6011. FC threshold was set to less than 1.5 and p-value threshold was set to 0.05. FC values less than 1.5 are considered downregulated. Fold change = DMD/Controls. A total of 1043 genes were downregulated in GSE6011.

For GSE6011 top ten upregulated and downregulated genes the fold change threshold was 1.0. It was necessary to use a lower fold change threshold for this dataset because the fold changes were lower in this dataset than GSE38417. Likely due to the overall younger age of the Duchenne muscular dystrophy patients in this dataset.

In GSE6011, 1637 genes were upregulated in Duchenne muscular dystrophy patients. The top 10 upregulated genes were MYH3, MYH8, MYH8, ACTC1, LUM, COL1A2, TYROBP, ASPN, COL1A1, and COL3A1. (Table 1). Three of these genes were also upregulated in GSE38417, they are MYH3, MYH8 and COL1A1 (Table 1 and Table 3). 1043 genes were downregulated in DMD patients. The top ten downregulated genes were DMD, PFKB3, UCP3, PDK4, GLUL, LPL, ART3, LPL, PPP1R1A, and HMGCS2 (table 4). The
absence of dystrophin is the defining factor of DMD so it being downregulated provides validation to this analysis.

When comparing the two datasets, the fold changes in GSE34817 are much higher than those of GSE6011. (Tables 1-4). One possible explanation for this could be the age of the patients. Dataset GSE6011 consist of pre-symptomatic DMD patients (1-2 years old). GSE34817 on the other hand has samples from older patients ranging from less than 2 years to 8 years old. This matters because Duchenne Muscular Dystrophy is a progressive disease so as the patients age the genetic dysregulation, they experience gradually becomes more severe. Since GSE6011 covered a narrow age range the differential genes identified do not completely reflect what the transcriptomic profile of older patients.
Figure 1 Heat map generated by ImaGEO from datasets GSE GSE6011 and GSE38417. P-value cutoff 0.05. Uses max P value method. GSE6011 had a total of 1586 statistically significant genes (below the p-value threshold) GSE38417 had 7107 statistically significant genes. Combined meta-analysis showed a total of 3321 genes being statistically significant.
Next a combined analysis was done on both datasets together. In this portion of the analysis the focus is how gene expression changes in Duchenne muscular dystrophy patients compared to samples without DMD. All the gene expression values from both datasets will be used to pull out which genes are most significantly differentially expressed. Meta-analysis reported 3321 genes statistically differently expressed in the two datasets. MYH3, MYH8, SPP1, COL1A1 and TNNT2 are all reported as being upregulated (figure 1). Dystrophin is downregulated acting as a validator to the meta-analysis results. Combined analysis further confirmed the results of separate analysis.

<table>
<thead>
<tr>
<th>ID</th>
<th>Average FC</th>
<th>FC-GSE60</th>
<th>FC-GSE38</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGFBP7</td>
<td>0.233898</td>
<td>0.119038</td>
<td>0.348759</td>
<td>insulin like growth factor binding protein 7</td>
</tr>
<tr>
<td>IGF1</td>
<td>0.295255</td>
<td>0.06855</td>
<td>0.521961</td>
<td>insulin like growth factor binding protein 1</td>
</tr>
<tr>
<td>TNNT2</td>
<td>0.810197</td>
<td>0.308945</td>
<td>1.31145</td>
<td>troponin T2, cardiac type</td>
</tr>
</tbody>
</table>

Table 4: Combined analysis of both datasets using ImaGEO. P-value cutoff 0.05. Uses max P value method. TNNT2 is a marker for heart failure and is upregulated in GSE38417 with a fold change of 1.31145. IGFBP-7 is an inhibitor of IGF-I and -II which promote skeletal muscle growth. Both IGF-I and IGFBP-7 did not see any upregulated, which differs from the findings of a previously published study.

Analysis done on a similar dataset looked at the expression profiles of ten Duchenne muscular dystrophy patients and eight health controls. One aspect of late stage DMD is heart failure, so the expression of cardiac related genes is relevant when investigating DMD. The researchers in this study looked at genes specific to cardiac tissue. Five genes were reported as upregulated, TNNT2, CARP, CASQ2, CUGBP2 and connexin 43. (Bakay, Zhao, Chen, & Hoffman, A web-accessible complete transcriptome of normal human and DMD muscle, 2002). The paper hypothesized that the upregulation of CARP and connexin 43 was related to macrophage infiltration. TNNT2 is considered a marker for heart failure and its upregulation may be only due to induction of myoblast differentiation. Temporal profiling of myoblast indicated that TNNT2 upregulation is limited to the activated myoblasts of dystrophin deficient tissues. (Bakay, Zhao, Chen, & Hoffman, 2002)
From the combined analysis the fold changes of relevant cardiac related genes were evaluated, of which only TNNT2 had data available. In GSE6011 TNNT2 was slightly downregulated with a fold change of 0.31. GSE38417 demonstrated upregulation of TNNT2 with a fold change of 1.31 (table 5). The fold change of TNN2 is much less than that of the study. One possible explanation is that TNNT2 is very lowly expressed in non-differentiating myoblast and normal muscle (Bakay, Zhao, Chen, & Hoffman, A web-accessible complete transcriptome of normal human and DMD muscle, 2002). Interestingly the pre-symptomatic patients did not show up-regulation of TNNT2 like GSE38417 did. (Table 1 and 5). The combined analysis reported an upregulation of TNNT2 (figure 1). This suggests that TNNT2 is not initially dysregulated in DMD patients, instead dysregulation appears as the disease progresses.

The same study also found that IGF-I and IGF-II was upregulated in dystrophin deficient mice models. IGF-I and/or IGF-II can promote the growth of skeletal muscle. Inhibitory IGF binding proteins-2, -4, -6, and -7 and IGFBP-5 protease (PRSS11) were also upregulated. Their upregulation may prevent any promotion of skeletal muscle growth by IGF-I and -II in DMD patients (Bakay, Zhao, Chen, & Hoffman, A web-accessible complete transcriptome of normal human and DMD muscle, 2002). The datasets only had data available for IGF-I and IGGBP-7, both of which did not show any upregulation (Table 5). From this analysis it cannot be confirmed whether inhibition of IGF-I and -II by IGF inhibitory binding proteins is preventing skeletal muscle growth.

*Gene Ontology Analysis*
Table 6: ImaGEO generated gene ontology for biological processes. Includes datasets GSE6011 and GSE34817. The maximum p-value meta-analysis method was used with a p-value cutoff of 0.05.
Figure 2: ReactomePA gene ontology indicating what pathways are being enriched. The size of the dot is the number of genes within the given gene set that is associated with the given GO term. The color corresponds with the p-adjusted value, which gives information on how statistically significant a particular enrichment of the gene set is.

ReactomePA is a gene ontology package in R that was utilized on GSE34817. This analysis was done to identify gene sets that were being enriched in Duchenne muscular dystrophy patients. Some gene sets related to immune response were shown to be enriched. Such as platelet activation, signaling and aggregation and Class I MHC mediated antigen processing and presentation. Reactome also reported an enrichment of genes related to negative regulation of PI3K/AKT network and PI3K/AKT signaling in cancer. Rho-GTPase signaling showed the highest portion of genes being enriched within a given gene set, followed by diseases of signal transduction by growth factor receptors and second messengers. GSE34817 consisted of older patients and overall had shown a more dramatic degree of fold changes. (Table 1 and Table 2).
For this reason, individual gene ontology analysis was done on only GSE34817, rather than the younger pre-symptomatic patients found in GSE6011.

Gene Ontology was done on both datasets together to identify gene sets being enriched in DMD patients and what pathways and processes they were related to. In previously published literature Gene Ontology analysis was done on GSE6011. In this dataset, genes related to muscles were differently expressed. Dystrophin for example was downregulated, as expected. Their analysis found that enzymes related to glycogen metabolism, glycolysis, TCA cycle, lipid transport and β-oxidation are slightly downregulated. Additionally, genes related to inflammation and fibrosis are upregulated in Duchenne muscular dystrophy patients. (Pescatori, et al., Gene expression profiling in the early phases of DMD: a constant molecular signature characterizes DMD muscle from early postnatal life throughout disease progression, 2007). My gene ontology analysis included both GSE34817 and GSE6011, it was interesting to compare how this gene ontology compares to the one done on just GSE6011.

Additionally, I performed a gene ontology analysis in R-studio on dataset GSE34817 to gain perspective on how the results would compare to both the published analysis done on GSE6011 and a combined analysis. ReactomePA gene ontology analysis identified some pathways that were being enriched, such as the class I MHC mediated antigen processing and presentation (figure 2). This gene set involves histocompatibility complex class I molecules. MHC is important in cell mediated immunity, by binding antigens and presenting them to CD8+ T-cells. (Loureiro & Ploegh, 2006).
In Duchenne muscular dystrophy patients MHC class I and II has been reported as being upregulated. This upregulation is activated by the increase in TLR7 expression. TLR7 is activated by myofiber-derived RNA molecules classified by DAMPs (Rosenberg, et al., 2018). It’s enrichment in the microarray datasets is further validated by the finding of previous studies. The chronic activation of innate immunity has been suspected as playing a role in preventing the regeneration of damaged myofibers (figure 2). RhoA-GTPase controls the expression of MyoD, meaning RhoA-GTPase signaling is crucial for myogenesis induction. The Rho kinase family are serine/threonine kinases, which include Rho, Rac, and cdc42 proteins RhoA GTPase is thought to play a role in utrophin upregulation. (Bonet-Kerrache, Fortier, Comunale, & Gauthier-Rouviere, 2005) Utrophin is very similar to dystrophin and can bind to the dystrophin associated complex when dystrophin is not available. One potential treatment for Duchene muscular dystrophy is to increase the expression of utrophin in place of dystrophin. My analysis did not provide any information on utrophin expression. However, the upregulation of the Rho-GTPase activity is evidence that myogenesis related pathways are still active in the DMD samples of GSE3417. Genes related to extracellular matrix organization are also shown to be enriched in Duchenne muscular dystrophy patients (figure 2). Upregulation of collagen was seen in the same dataset (table 1). Dataset 6011 didn’t show the same degree of collagen upregulation, possibly due to the patients being pre-symptomatic.

Duchenne muscular dystrophy patients are afflicted with elevated inflammatory responses with corticosteroid treatment being the standard of care for DMD. (Kourakis, et al., 2021) Inflammation occurs in response to pathogens or tissue injury that trigger a signaling cascade initiating the increase of proinflammatory cytokine genes. Such a process relies heavily on protein phosphorylation via protein kinase. Mitogen-activated protein kinases are a family of
protein kinases that play an important role in regulation the inflammatory response. The MAPK signaling cascade can be activated in response to a stress cue such as cytokines. Proinflammatory cytokines such as IL-1, IL-6 and TNF-a activate p38 kinases (Patterson, Nibbs, McInnes, & Siebert, 2014). Activated MAPKs will then translocate to the nucleus and phosphorylate different transcription factors to change gene expression. Gene ontology analysis reported an enrichment of genes involved in positive regulation of MAPK cascade. Further analysis using Reactome also reported an enrichment of the MAPK family signaling cascades. (Figure 2). It’s clear that the Duchenne muscular dystrophy patients included in the combined dataset meta-analysis are experiencing inflammation. As previously discussed, this is likely due to muscle tissue injury occurring when the dystrophin deficient muscles contract.
Conclusions and Future Directions

Differential gene expression analysis of two different Duchenne muscular dystrophy datasets done both separately and together shows a great many genes are differentially expressed in DMD patients. Doing them separately and together allowed me to compare the results and see how different two datasets might look using the same methods of meta-analysis. Overall, there were some differences, but both datasets separately showed expression of at least one hallmark of DMD. A major difference was seen in the log fold changes of the two datasets, this is likely due to the younger age of the patients in GSE6011. Duchenne muscular dystrophy is a progressive disease, and so younger patients wouldn’t be expected to show such severe differences in gene expression that older patients would. By combining the datasets and doing a differential gene analysis and gene ontology a broader picture was drawn about DMD differential gene expression.

Enrichment of gene sets related to inflammation is what stuck out to me the most. A previous study isolated satellite cells from young and aged mdx mice and determined that these cells regained their ability to regenerate once removed from the inflammatory environment of the dystrophin deficient mice. (Boldrin, Zammit, & Morgan, 2015) The researchers had that the inflammatory environment of dystrophin deficient mice is what inhibits regeneration and not any dysregulation of satellite cells. This study combined with the results of my data analysis led me to hypothesize that in DMD patients, myogenesis related genes are not dysregulated but instead the elevated inflammatory response caused by the absence of dystrophin contributes to the muscle’s inability to repair itself. To better explore this hypothesis more work would need to be done to evaluate gene expression of myogenesis related genes. Additionally, it would be
interesting in future work to determine what aspect of the inflammatory response is preventing muscle regeneration and how that can be targeted in treatment.

The datasets I used did not allow me to directly look at RNA expression. This limits the scope in which I could study myogenesis in these DMD patients. Non-coding RNAs and miRNAs have recently been found to have a big impact on myogenesis, and so it’s important to consider how their expression levels are impacted in DMD patients. Additionally, future studies should focus on determining what aspect of the immune system inhibits myogenesis. Treatments focused on this may be helpful in regaining muscle function in later stage patients of DMD and other related dystrophies. Additionally, management of inflammation is currently the standard of care of DMD patients. The glucocorticoids that are currently used clinically act non-selectively and cause a lot of complications. (Kourakis, et al., 2021) Determining what aspect(s) of inflammation in DMD patients contributes the most to muscle wasting as well as preventing muscle regeneration can help develop anti-inflammatory treatments that act specifically thus limiting the negative impacts.
Bibliography

Alexa, A., & Rahnenfuhrer, J. (2021, May 19). Gene set enrichment analysis with topGO.


