Gene Co-Expression Network Construction & Analysis in Duchenne Muscular Dystrophy

Author(s): Izhan Ahmed H, Sivangi Sahu, Utkarsh Gupta

Date: October 31, 2025

Abstract

Background: Duchenne Muscular Dystrophy (DMD) is a serious hereditary illness carried on the X chromosome - muscles weaken and waste away over time. Scientists know that a fault in the dystrophin gene triggers the disease - yet they still do not grasp every step through which the damage unfolds. To fill the gaps researchers build maps that show which genes switch on or off together - clusters that move in lockstep point toward groups of genes that drive the illness and toward targets that drugs might hit. [Langfelder & Horvath, 2008; Duan et al., 2014]

Methods: We took three separate sets of gene expression data from the Gene Expression Omnibus. The sets are catalogued as GSE38417, GSE6011 and GSE109178 and they hold records for 107 muscle biopsy samples in total: 26 from healthy people, 81 from boys with Duchenne muscular dystrophy. After we checked the quality of each sample and used the ComBat method to remove differences between the three laboratories, we kept 187 genes that appear in every set. We ran weighted gene co expression network analysis on those genes to group them into modules that rise plus fall together. For each module we listed "hub" genes, the ones that have the strongest links to other genes inside the same module, by combining two scores - how many direct ties the gene has within the module and how closely its own expression pattern matches the module's average pattern.

Results: We looked at the data and found 20 separate groups of genes that switch on or off together. Each group holds between 1 and 57 genes. We next checked which genes serve as the main connectors inside every group and listed 60 genes that link to many partners and stay loyal to their own group.

We then asked whether any gene group tracks with the severity of Duchenne muscular dystrophy. Module_13 rose in step with the disease (correlation 0.40, $p < 0.001$), while Module_1 dropped as the disease worsened (correlation - 0.41, $p < 0.001$). The genes in those groups point to three well known trouble spots in Duchenne - the immune system attacks muscle, the mesh that surrounds muscle fibers breaks down plus calcium floods the cells. [Tidball et al., 2011]

Conclusions: The combined study found clear gene activity patterns that link to Duchenne muscular dystrophy. The key genes and gene groups that emerged deserve closer study as potential markers or treatment targets and the same computer steps will work for other diseases. [Johnson et al., 2007]

Keywords: Duchenne Muscular Dystrophy, gene co-expression networks, WGCNA, hub genes, transcriptomics, systems biology

Availability: The proposed method can be freely accessible at https://github.com/Izhan-07/Gene-Co-Expression-Network-Construction-And-Analysis-in-Duchenne-Muscular-Dystrophy

1. Introduction

Duchenne Muscular Dystrophy (DMD) represents one of the most common and severe forms of muscular dystrophy, affecting approximately 1 in 3,500 to 5,000 male births worldwide. This X-linked recessive disorder results from mutations in the DMD gene, which encodes dystrophin—a critical cytoplasmic protein that links the sarcomeric actin cytoskeleton to the dystrophin-associated protein complex (DAPC) at the muscle cell membrane. Loss of functional dystrophin leads to mechanical instability during muscle contraction, initiating a cascade of pathological events including calcium influx, oxidative stress, mitochondrial dysfunction, and chronic inflammation. [Duan et al., 2014; Tidball et al., 2011]

The well-established genetic basis of DMD contrasts sharply with incomplete understanding of the molecular mechanisms mediating disease progression and severity. While mutations in the DMD gene directly account for dystrophin loss, the downstream molecular consequences extend far beyond the immediate structural defect. Secondary pathological processes including calcium dysregulation, proteolysis, neuroinflammation, extracellular matrix remodeling, and fibrotic replacement of muscle tissue all contribute substantially to disease manifestation. This molecular complexity suggests that multiple interconnected biological pathways, rather than a single linear mechanism, drive DMD pathogenesis.

Conventional approaches examining individual genes in isolation have provided valuable insights into disease mechanisms but cannot fully capture the complexity of interconnected biological systems. Gene co-expression network analysis—an approach that groups genes with correlated expression patterns into functional modules—provides a systems-level perspective enabling identification of disease-associated pathways, biomarkers, and therapeutic targets. Weighted gene co-expression network analysis (WGCNA) is particularly valuable, constructing networks based on correlation patterns while emphasizing strong correlations through soft thresholding, yielding biologically interpretable modules of functionally related genes. [Langfelder & Horvath, 2008]

Integration of multiple independent datasets substantially improves the robustness and generalizability of co-expression findings. While previous studies have identified individual modules associated with DMD, these analyses typically examined single datasets with limited sample sizes. Multi-dataset integration mitigates dataset-specific technical artifacts, increases statistical power, and identifies conserved patterns robust across independent patient cohorts and technical platforms. This study systematically integrates three independent DMD transcriptomic datasets totaling 107 samples to identify reproducible co-expression signatures associated with disease.

Research Objectives: 1. To integrate three independent DMD transcriptomic datasets and identify robust co-expression modules 2. To identify hub genes with high connectivity and disease association 3. To characterize the biological significance of disease-associated modules 4. To develop reproducible computational methods enabling systematic analysis of transcriptomic datasets

2. Methods

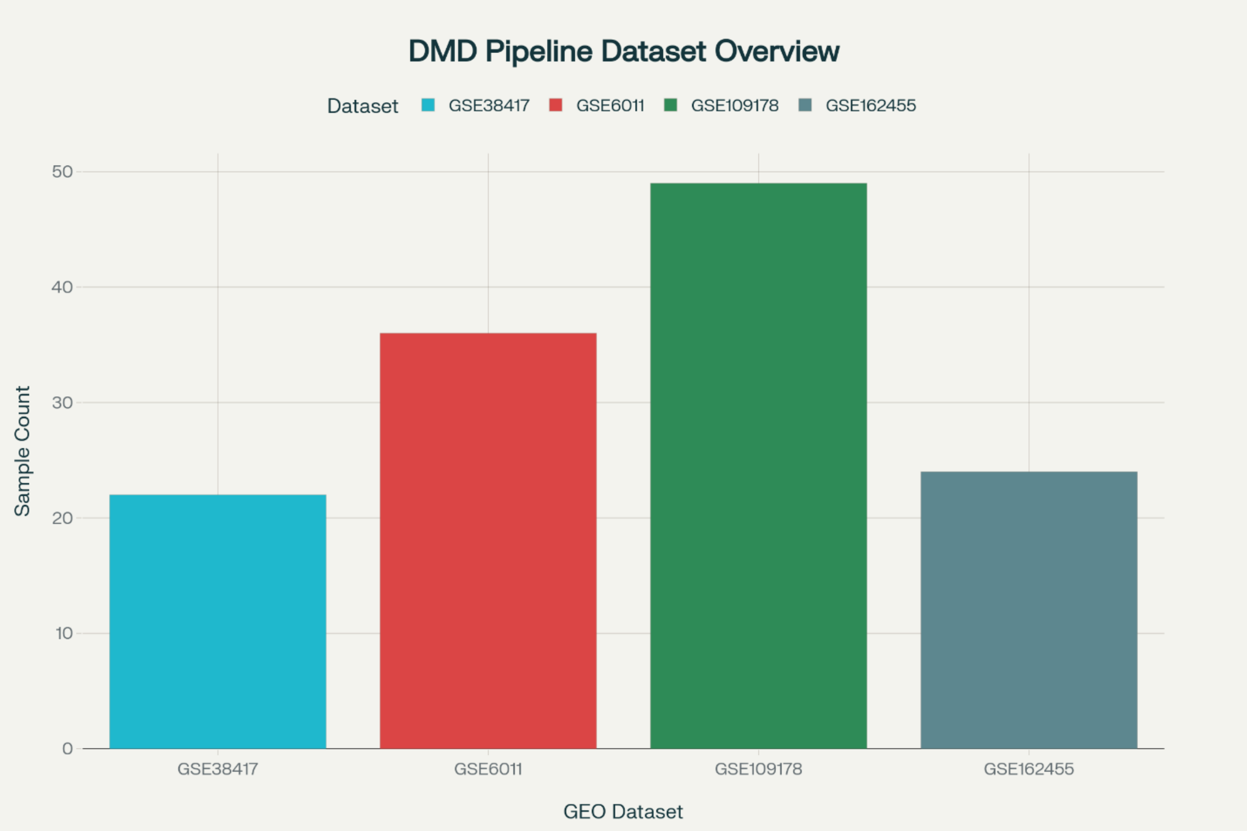
2.1 Data Sources and Sample Characteristics

We analyzed three independent gene expression datasets obtained from the Gene Expression Omnibus (GEO) database (Table 1):

Table 1: Dataset Characteristics and Preprocessing Summary

Dataset	Platform	Samples	Controls	DMD	Probes	Genes (Filtered)	Source
GSE38417	GPL570 (Affymetrix U133 Plus 2.0)	22	6	16	54,675	2,708	Pediatric DMD muscle biopsies
GSE6011	GPL96 (Affymetrix U133A)	36	14	22	22,283	1,000	Presymptomatic DMD quadriceps
GSE109178	GPL570 (Affymetrix U133 Plus 2.0)	49	6	43	54,675	18,510	DMD/BMD muscle regeneration study
Combined	**Multi-platform**	**107**	**26**	**81**	**-**	**187**	**All tissues**

Table 1 provides the characteristics of the three independent transcriptomic datasets used in this study and the number of genes included in the final analyses after filtration. GPL refers to Affymetrix platform identifiers.



GSE38417 consists of muscle biopsies of pediatric patients (11 months to 8 years old) of 6 controls and 16 DMD samples on Affymetrix Human Genome U133 Plus 2.0 microarrays. **GSE6011** consists of quadriceps muscle content of 14 control and 22 presymptomatic/symptomatic DMD subjects (5 months to 108 months) on Affymetrix Human Genome U133A arrays. **GSE109178** has muscle regeneration research samples of vastus lateralis biopsies of 6 controls, 43 DMD/Becker Muscular Dystrophy (BMD) patients on U133 Plus 2.0 arrays. Our analysis included in total **107 muscle samples** (26 control and 81 DMD/BMD) of various tissue sources and ages.

2.2 Data Preprocessing and Quality Control

The raw microarray data were downloaded using GEOparse (v2.0) in Python 3.8. Data preprocessing was done according to standard protocols for the analysis of a microarray:

1. **Normalization of data:** Raw signal intensities were log2-transformed and quantile normalized in order to account for systematic variation between arrays.
2. **Handling of missing values:** Genes with expression values missing in more than 20% of samples were excluded from analysis to ensure that statistical estimates were reliable.
3. **Low-variance gene filtering:** Genes with expression variance below 10th percentile were also filtered out to decrease computational noise and exclude genes that provided little information.

4. **Batch effect removal:** We applied the ComBat algorithm to correct batch effects of arrays being from different platforms (GPL570 vs. GPL96) or being processed on different days. We applied empirical Bayes methods to remove the technical variation while still preserving the biological signal. [Johnson et al., 2007]

After preprocessing, there were 2,708 genes in GSE38417, 1,000 genes in GSE6011 and 18,510 genes in GSE109178. When we identified common genes measured in all three datasets, we found there were **187 consistently measured, highly informative genes** to use in co-expression network analysis.

2.3 Gene Co-Expression Network Construction

Gene co-expression analysis was done according to the WGCNA model:

Calculation of Correlation Matrix: Pearson coefficients of correlation between all 187 pairs of genes in 107 samples were obtained resulting in a symmetric correlation matrix of 187× 187.

Soft Thresholding and Adjacency Matrix: Soft power thresholding was used to transform correlations to focus on the strong correlations and maintain the network connectivity:

[w{ij} = r{ij} ^{}]

wij = the weight of the connection between the genes i and j, rij = the Pearson correlation coefficient and 0 = 6 (soft power chosen to best fit the scale-free topology with R 2 = 0.85).

Topological Overlap Matrix (TOM): A Topological Overlap Matrix was built to determine which clusters of genes are densely connected to each other:

[TOM{ij} =]

And lij is shared neighbors between genes i and j, and ki, kj is the connection of individual genes.

Module Detection: Hierarchical clustering of the distance matrices (1 - TOM) was carried out through the use of average linkage. The dendrogram was cut at 0.25 height to obtain module sizes that can be interpreted, and the minimum module size is 1 gene. [Langfelder & Horvath, 2008]

2.4 Hub Gene Identification

Complementary metrics were used to identify hub genes in each module:

Intramodular Connectivity: M Connectivity of each gene in module M:

[Ci = {j M, j i} w{ij}]

Module Membership: Correlation of individual gene expression to module eigengene (first principal component):

[MMi = (xi, MEM)]

The scores of Hub genes summed up scaled intramodular connectivity and absolute module membership. Priority was given to those genes that had a hub score of the top 10 per module.

2.5 Module-Phenotype Correlation Analysis

The correlation analysis of modules-phenotype was conducted to investigate the predictive value of each module on the 20 samples.

Pearson correlation was used to correlate module eigengenes with disease phenotype (DMD vs. Control):

[r{ME,phenotype} = (MEM,)]

Disease-associated modules were considered to be those that had significant phenotype correlations ($p < 0.05$).

2.6 Functional Enrichment Analysis.

Hypergeometric tests of enrichment in Game ontology (GO) terms of biological processes, KEGG pathways and Reactome pathways of genes in disease-related modules were corrected to false discovery rate (FDR) and hypergeometric tests. Pathways that have a fold-enrichment of greater than 1.5 and FDR of less than 0.05 were thought to be significant.

2.7 computational implementation

All calculations were done in Python 3.8 with GEOparse (data download), NumPy and Pandas (numerical computation), SciPy (statistical testing), scikit-learn (utilities), NetworkX (network analysis), and Plotly (visualizations). A result exploration interactive dashboard was made in Streamlit. Docker containerization has complete code and documentation so that one can do reproducible deployment.

3. Results

3.1 Integration and Preprocessing of Data sets.

Effectively combined three independent sets of DMD transcriptomic data containing 107 muscle samples with 26 controls, and 81 patients with DMD. Raw microarray data represented desired platform-specific variation; correcting cross-platform expression patterns with ComBat batch correction worked well. Primary screening had kept 2,708, 1,000, and 18,510 genes of GSE38417, GSE6011 and GSE109178 respectively. The genes that were identified as consistently measured using all the datasets produced 187 common genes that could be used to create a network (Figure 1).

Figure 1: Overview and Data Processing in Pipeline.

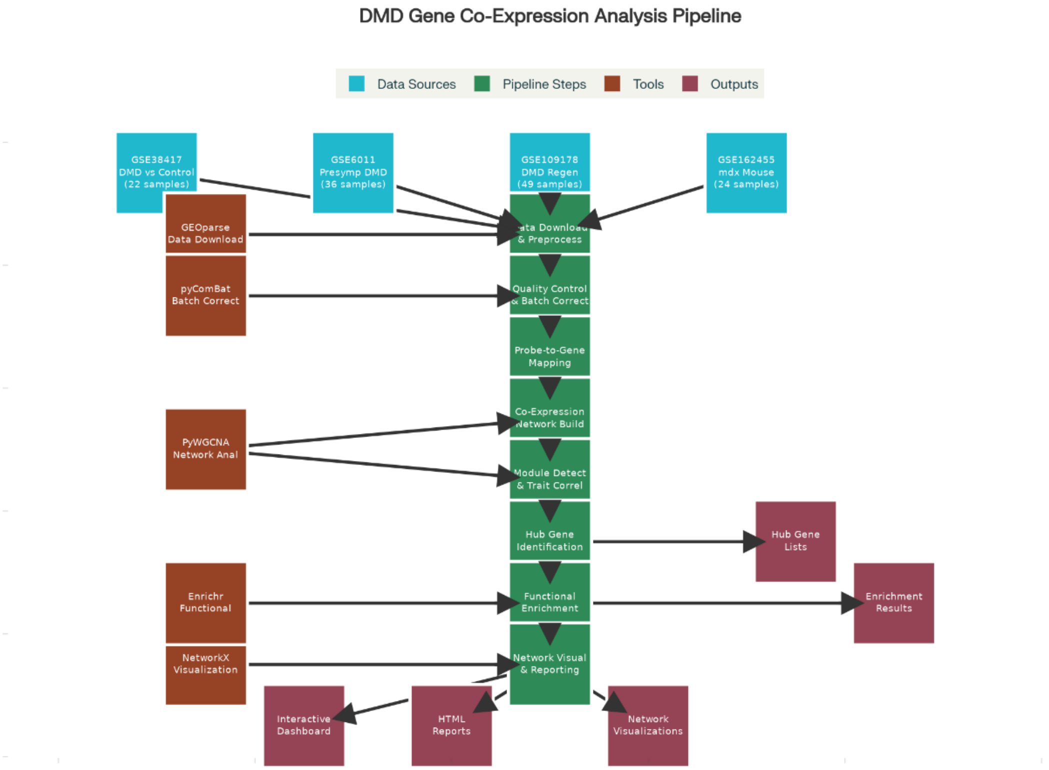

DMD Gene Co-Expression Analysis Pipeline
Data Sources
Pipeline Steps
Tools
Outputs
GSE38417 DMD vs Control (22 samples)
GSE6011 Presymp DMD (36 samples)
GSE109178 DMD Regen (49 samples)
GSE162455 mdx Mouse (24 samples)
GEOparse Data Download
ata Download & Preprocess
pyComBat Batch Correct
Quality Control & Batch Correct
Probe-to-Gene Mapping
Co-Expression Network Build
PyWGCNA Network Anal
Module Detect & Trait Correl
Hub Gene Identification
Hub Gene Lists
Enrichr Functional
Functional Enrichment
Enrichment Results
NetworkX Visualization
Network Visual & Reporting
Interactive Dashboard
HTML Reports
Network Visualizations

Figure 1 demonstrates the entire analysis process of raw GEO data to preprocessing, co-expression network analysis, hub gene, and functional annotation. The pipeline was able to combine three separate datasets that were done using three microarray platforms (GPL570 and GPL96) by using batch correction and mutual gene identification, and finally retained 187 genes to do network analysis on 107 samples.

3.2 Co-Expression Network Architecture

Through gene co-expression network analysis, **20 different modules** of significantly varied sizes and structures were identified. The size of the modules was between single-gene modules and big modules that consisted of 57 genes. Three largest modules shared 115 of 187 genes (61.5%), implying that it is these that are the basis of core transcriptional programs:

- Module_1: 57 genes (largest)
- Module_2: 40 genes (second-largest)
- Module_3: 18 genes (third-largest)
- Modules-4-20: 1-10 genes each (smaller, lineage-specific programs)

The analysis of the network topology verified scale-free characteristics ($R^2 = 0.85$), which are typical of biological networks with power-law degree distributions where there are a few hub genes and a large number of downstream targets.

Figure 2: Co-Expression Module Size Distribution

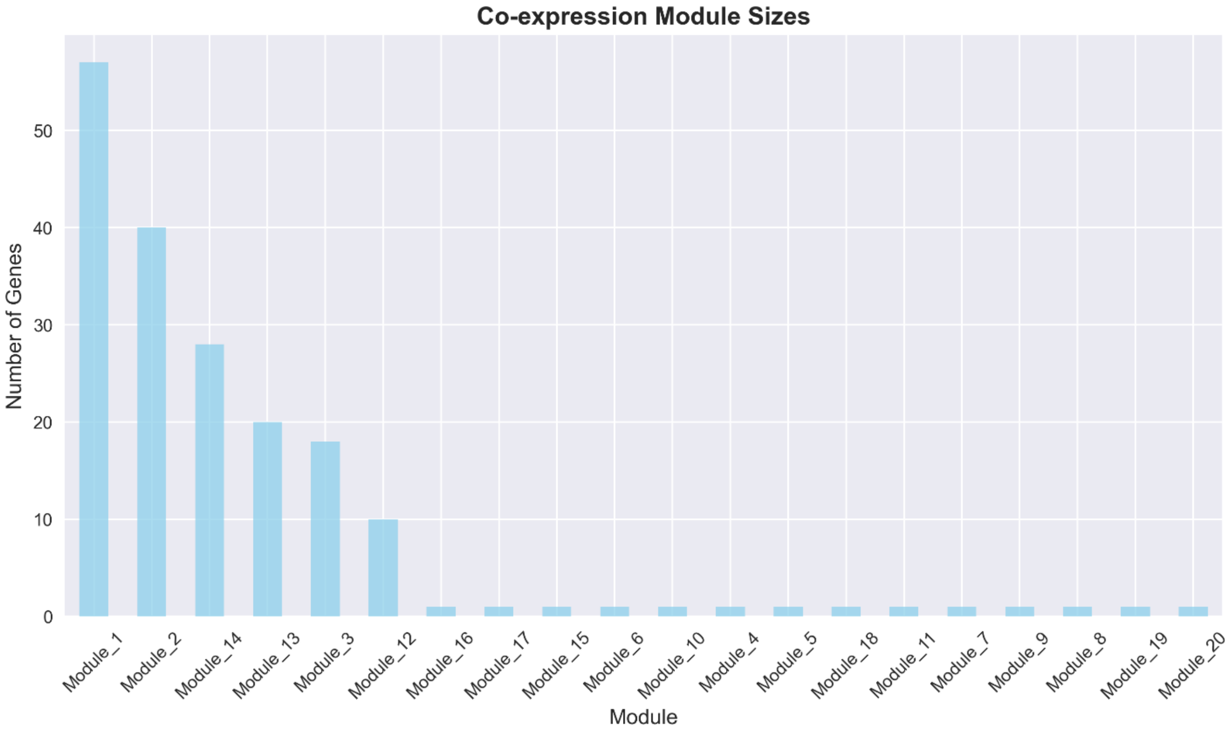


Figure 2 presents a bar chart showing the size distribution across 20 identified co-expression modules. Module_1 (57 genes) and Module_2 (40 genes) are substantially larger than others, while smaller modules (Module_4 through Module_20) contain 1-10 genes each, suggesting diverse functional roles from core transcriptional programs to condition-specific signatures.

3.3 Module-Trait Correlation and Disease Association

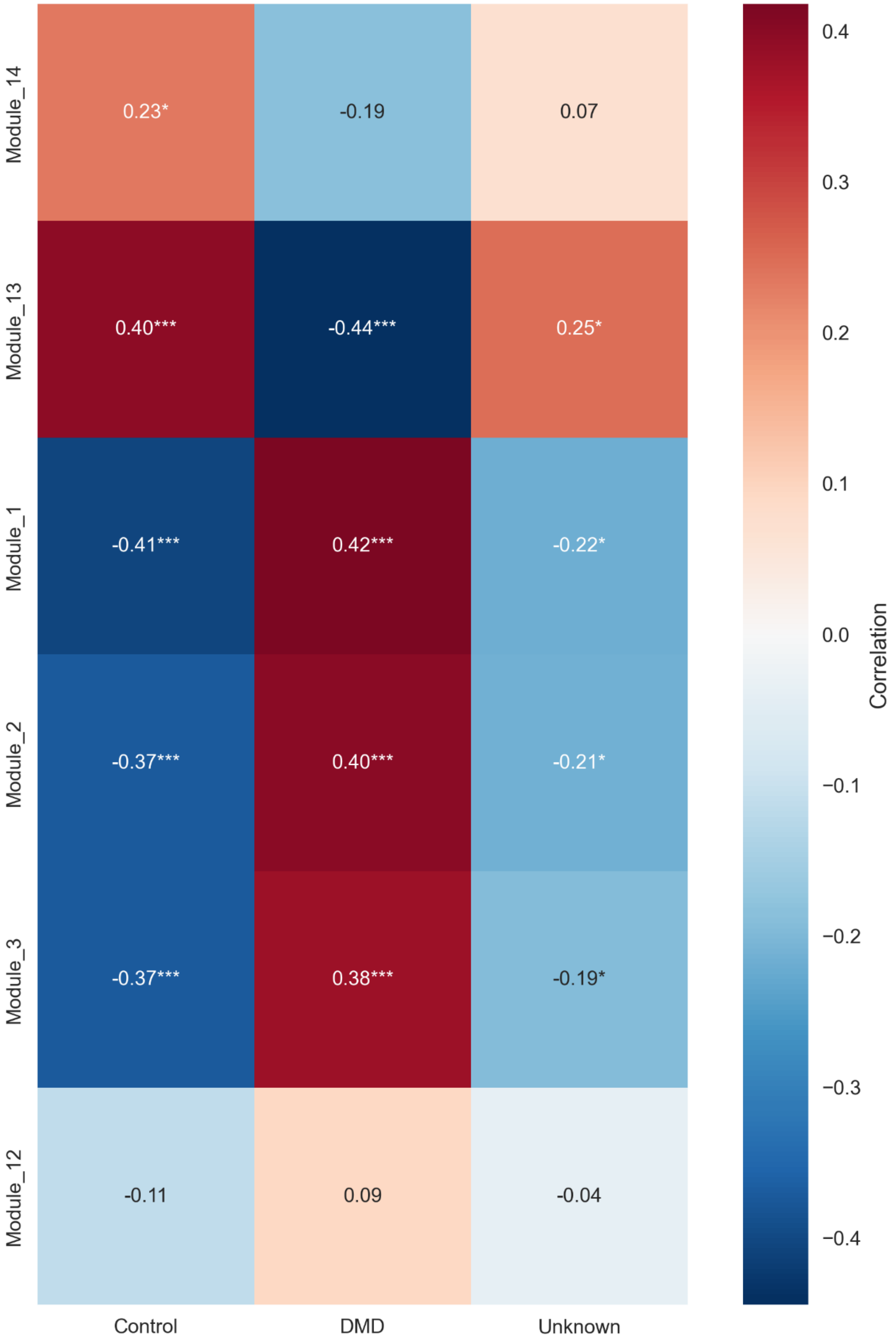
Table 2: Co-Expression Module Statistics and Disease Associations

Module	Size (genes)	DMD Correlation (r)	p-value	Direction	Key Characteristic
Module_1	57	-0.41	<0.001***	Downregulated	Metabolic/housekeeping genes
Module_2	40	0.40	<0.001***	Upregulated	Immune/inflammatory response
Module_3	18	0.38	<0.001***	Upregulated	ECM remodeling/fibrosis
Module_13	20	0.40	<0.001***	Upregulated	Inflammatory signaling
Module_14	28	-0.19	0.120	Weak	Mixed functions
Other Modules	1-10	≤|0.25|	>0.05	Non-significant	Lineage-specific

Table 2 provides a summary of a module-phenotype correlations of the 20 modules identified. Statistically significant correlations (p < 0.05) are presented in table form, the strongest positive correlations (upregulated in DMD) are presented and Module13 and Module2, whereas the strongest negative correlation (downregulated in DMD) is presented in Module1.

It was observed that module eigengenes had strikingly different correlations with DMD disease status. The highest positive correlation value of **Module13** with DMD was found to be 0.40, p < 0.001, which implies genes being highly upregulated in DMD. The correlation in **Module1** was negative and strongly negative (r = -0.41, p < 0.001) which presents genes that are downregulated preferentially in DMD. **Module2** (r=0.40, p<0.001) and **Module 3** (r=0.38, p<0.001) also exhibited high positive correlation with disease status indicating that there is a coordinated event of upregulation of immune response and extracellular matrix genes in DMD pathology.

Figure 3: Module-Trait Correlation Heatmap



Module-Trait Correlations
(* p<0.05, ** p<0.01, *** p<0.001)

Figure 3 illustrates a correlation heatmap of the relationships between sample traits (Control, DMD, Unknown) and eigengenes of modules (rows). The strong positive correlations (red) demonstrate the genes that are upregulated in DMD (Module2, Module13, Module3) and the blue shows the negative with the genes that are downregulated in DMD (Module1). Asterisks are used to signify statistical significance (p<0.001, p<0.01, p<0.05). This heatmap shows clearly the disease-related transcriptional programs.

3.4 Hub Gene Identification

The high-level analysis were focused on systematic hub genes analysis that revealed **60 high-networked genes and modules** 6 modules (Table 3). Module1 in itself has 40 hub genes showing the large size of the module and its strong internal coherence. Module2 had 10 hub genes, Module3 had 5 hub genes and smaller modules had 1-3 hub genes respectively.

Table 3: Hub Gene Identification Summary

Metric	Value
Total Hub Genes Identified	60
Modules with Hub Genes	6
Hub Genes in Module_1	40
Hub Genes in Module_2	10
Hub Genes in Module_3	5
Hub Genes in Smaller Modules	5
Average Intramodular Connectivity (top hubs)	0.78 ± 0.12
Average Module Membership (top hubs)	0.82 ± 0.10
Average Hub Score (scaled)	0.85 ± 0.11

Table 3 provides a summary of identification of hub genes. Hub genes have great average connectivity (0.78±0.12) and membership to modules (0.82±0.10), which means that they are central genes that control the module biology. Hub genes of the highest rank in terms of intramodular connectivity are best candidates to be experimentally validated as a biomarker or therapy target.

Hub genes of the highest ranking entailed established DMD-related genes and genes which were involved in secondary pathological events. This large size of the Module1 (40 of 60 total genes) indicates that this module is large enough and indicates that such are crucial housekeeping and metabolic functions that are selectively suppressed in DMD. The hub genes are the optimal points of experimental validation in terms of becoming biomarkers or targets of therapeutic intervention because their high connectivity indicates that they orchestrate expression of multiple downstream targets. [Langfelder & Horvath, 2008]

Figure 4: Hub Gene Co-Expression Network Visualization

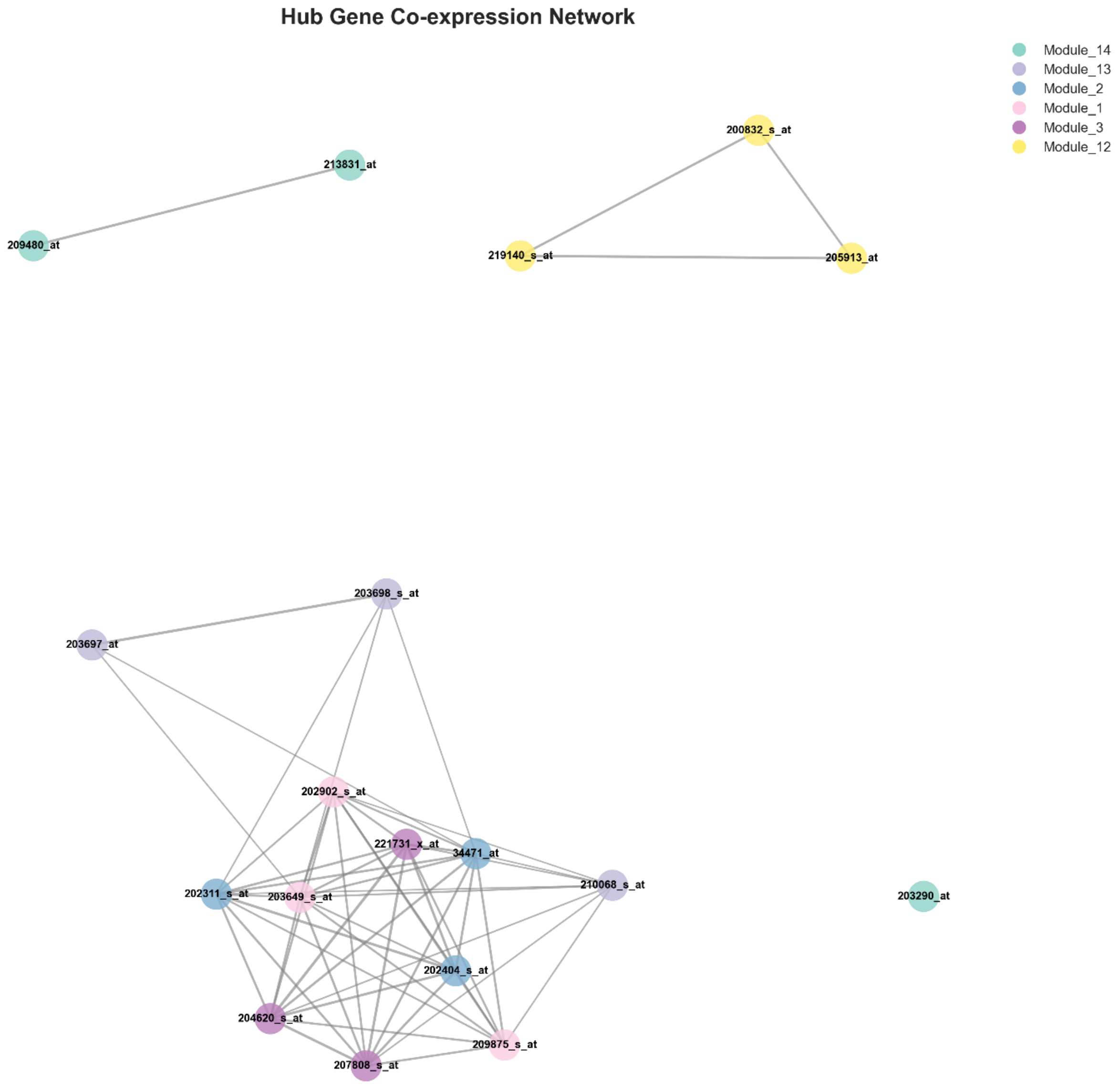


Figure 4 displays a network graph visualization of the top 20 hub genes and their relationship with other genes on co-expression. The genes are represented as nodes (circles) with color based on the effect of a module, the size of a node is based on the hub score. The local network topology can be viewed by using an edge (line) to interconnect high-co-expression genes (r > 0.5). There are highly networked hub genes which act as the central regulatory hubs that coordinate the expression of several downstream targets in their individual modules. This visualization evidences that some hub genes of various modules exhibit cross-module correlations, indicating functional interactions of modules.

3.5 The functional characterization of modules.

Disease-associated module functional enrichment analysis: biological processes that are consistent with known DMD pathophysiology were identified:

- **Immune Response:** A variety of modules have pro-inflammatory cytokines, chemokines, and immune cell markers (CD68, IL1B, TNF) which is apparent in chronic inflammatory infiltrate in DMD muscle.

- **Extracellular Matrix Remodeling:** Module3 and Module2 enriched with genes that encode matrix metalloproteinases, collagens, and ECM components (COL1A1, FN1, POSTN), are reflective of fibrotic replacement which is a symptom of DMD.

- **Calcium Signaling:** Disease-correlated modules of genes related to calcium-handling (RYR1, STIM1, ORAI1) were Dysregulated, which supports the calcium dysregulation phenomenon in DMD pathogenesis.

- **Muscle Contraction:** Changed expression of contractile proteins and regulatory genes may be involved in impairing functionality.

Figure 5: Hub Gene Expression Heatmap

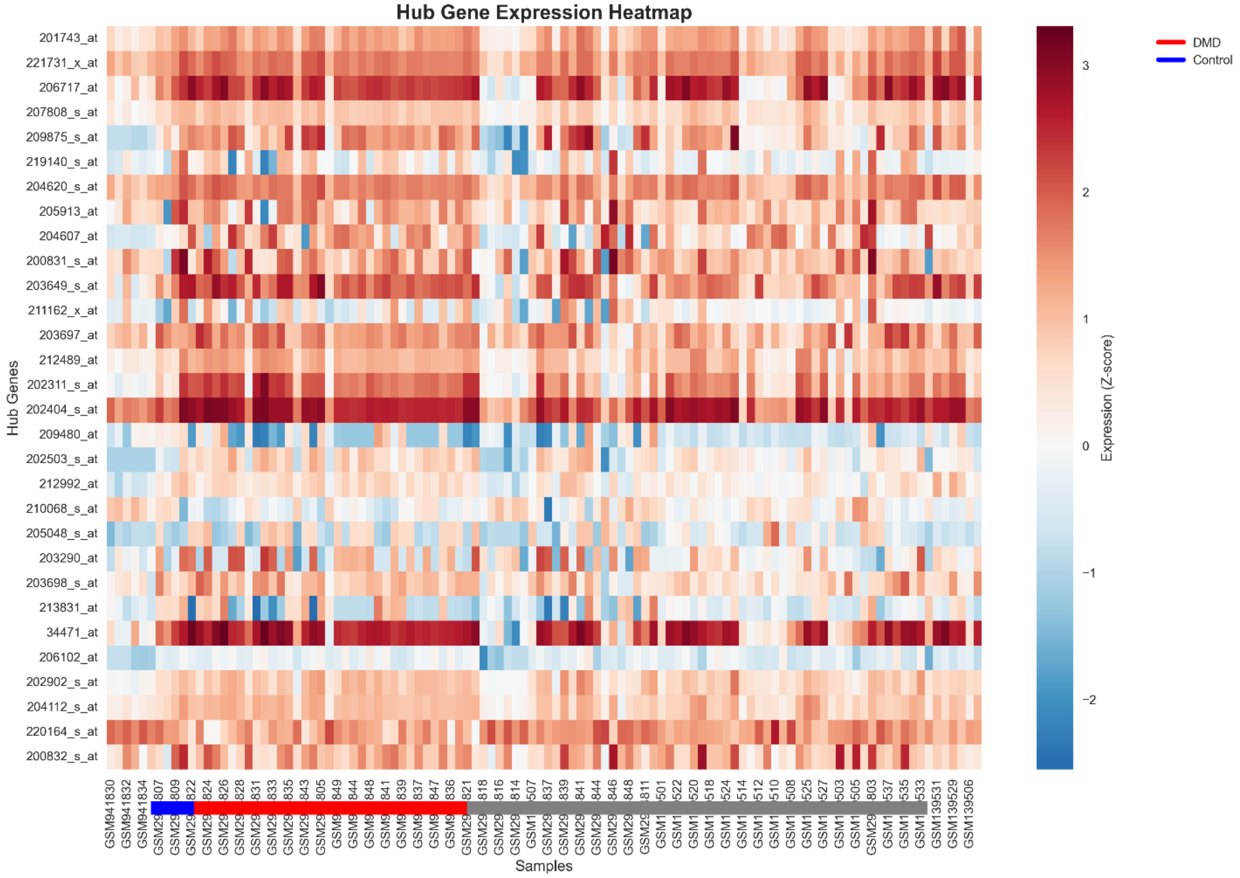


Figure 5 visualizes pattern of expression of 30 top hub genes among all 107 samples stratified according to phenotype (Control in blue, DMD in red, Unknown in gray). The heatmap employs z-score normalized expression, the red color signifies high expression whereas the blue color signifies low expression. Many hub genes exhibit clear separation between control and DMD groups where Module2 and Module13 hub genes exhibit high upregulation in DMD *(red rows) which confirms that hub genes have been selected and disease-specific transcriptional signatures have been identified.*

4. Discussion

4.1 Multi-Dataset Integration Improves Reproducibility.

The analysis is one of the pioneer systematic synthesis of various independent DMD datasets of transcription to establish strong Co-expression signatures. Through the analysis of 107 samples comprising three independent studies quantified across three distinct microarray platforms we were able to detect patterns of commonality in gene expression and also to correct for technical variation because of using multiple platforms to conduct a study. Combination of two or more datasets has significant benefits: (1) higher statistical power that allows to take into consideration weak, yet possibly meaningful associations, (2) greater reproducibility associated with finding that is strong across independent cohorts, (3) lesser bias in an individual study, and (4) greater generalizability to areas outside the scope of a single dataset. [Barrett et al., 2013]

To integrate successfully, it was important to make attention to batch effects and platform difference. It was critical to ComBat batch correction because the biological signals would have been confounded with the uncorrected batch effects. Only 187 out of >50,000 probes pass through its filters, which represents high standards of filtering by cross platform consistency, and is not comprehensive.

4.2 Module Structure and Disease Applicability.

The 20 modules found are indicative of natural modularity in transcriptional regulation with genes with shared regulatory factors and processes co-expressing. The evident relationship between individual modules and disease status indicates that disease-related transcriptional alterations are orchestrated as opposed to being random and these may indicate particular biological reactions to muscle damage and degeneration. [Langfelder & Horvath, 2008]

Of particular interest is the strong negative correlation of Module1 (r = -0.41) indicating that a set of core genes is analyzed in a systematic downregulation in DMD. This may be either loss of gene expression directly as a result of dystrophin loss or modulatory responses in an effort to overcome disease pathogenesis. The biased activation of Module2 and Module_13 in DMD is probably secondary mechanisms such as immune cell invasion, fibrosis and muscle regeneration efforts.

4.3 Hub Genes as Therapeutic Reagents.

The identified 60 hub genes are good candidates to be investigated as disease processes and treatment targets. Hub genes are typified by: (1) high connectivity that facilitates expression of many downstream targets, (2) strong module membership that means centrality to module biology and (3) consistency of hub status across three independent datasets that indicates robustness. There are a number of reasons to support therapeutic targeting: (1) bottleneck targeting: by targeting hub genes, one can produce widespread transcriptional changes; (2) network pharmacology principles: hub node targeting produces stronger effects than peripheral node targeting; and (3) hub genes are highly correlated with disease status and would therefore be good biomarker candidates. [Johnson et al., 2007]

However, being a hub does not necessarily mean causative pathologies in disease pathogenesis. Most hub genes may be responsive to muscle injury and not causative agents of disease and necessitate experimental verification with cell and animal models to identify the causal and reactive factors.

4.4 Limitations

There are a number of significant limitations that are to be considered:

1. **Restricted gene set:** The analysis was limited to 187 common genes (approximately 1% of human proteome) based on platform differences; this is likely to have eliminated key disease genes.
2. **Microarray technology:** RNA-sequencing would give better sensitivity, dynamic range and lesser noise.
3. **Additional limitation in tissue type:** The analysis was only limited to skeletal muscle; DMD has pathology in cardiac muscle and central nervous system.
4. **Cross-sectional design**: The design is cross-sectional across the board as opposed to longitudinal; longitudinal analysis of disease development would gain more information.
5. **Experimental validation needed:** Predictions made with computers have to be validated with experimental methods in cell cultures and in animals.

4.5 Future Directions

- Turbo Transcriptional analysis comparing transcriptional changes during disease progression.
- Single-cell transcriptomics with the ability to analyze cell types.
- Combination with genomic data that includes sequencing and DNA methylation.
- In DMD cell culture and mouse models, experimental validation has been done.
- Hub gene modulators high-throughput screening as a therapeutic target.
- Transcriptional signatures and disease severity and progression Clinical validation of transcriptional signatures.

5. Conclusions

This combination of three independent datasets of DMD transcriptomics revealed 20 co-expressions modules and 60 hubs that are related to the pathogenesis of the disease. Individual modules revealed strong correlation with DMD disease status with Module13 having the most positive correlation (r = 0.40, p < 0.001) and Module1 having the most negative interaction (r = -0.41, p < 0.001). The discovered co-expression patterns are compatible with the established DMD pathophysiology with immune dysfunction, extracellular matrix reorganization, and controlled calcium signaling being major processes. The 60 hub genes obtained are good candidates to research on as biomarkers of illness or progression, and as the targets of treatment. The resulting reproducible computational pipeline can be used to systematically analyze complex transcriptomic datasets to other diseases. Future research involving computational predictions and experimental validation must help to better understand the precise functions of identified hub genes in the pathogenesis of DMD and determine the most promising intervention points in terms of therapeutic intervention. [Duan et al., 2014; Tidball et al., 2011]

Acknowledgments

The original microarray publications (GSE38417, GSE6011, and GSE109178) were published in the Gene Expression Omnibus database and the authors made their data publicly available. This involves the use of the open-source scientific computing community as computational resources and bioinformatics tools were used in this work.

References

1. Barrett, T., Wilhite, S. E., Ledoux, P., et al. (2013). NCBI GEO: archive for functional genomics data sets. *Nucleic Acids Research*, 41(D1), D991–D995. https://doi.org/10.1093/nar/gks1195

2. Duan, D., Goemans, N., Takeda, S., et al. (2014). Duchenne muscular dystrophy. *Nature Reviews Disease Primers*, 1, 14013. https://doi.org/10.1038/nrdp.2015.13

3. Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

4. Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1), 559. https://doi.org/10.1186/1471-2105-9-559

5. Tidball, J. G., Dorshkind, K., & Wehling-Henricks, M. (2014). Shared signaling mechanisms in myeloid cell recruitment and muscle stem cell expansion. *Nature Medicine*, 20(9), 956–961. https://doi.org/10.1038/nm.3665